Springer Oceanography

Manuel G. Velarde
Roman Yu. Tarakanov
Alexey V. Marchenko *Editors*

The Ocean in Motion

Circulation, Waves, Polar Oceanography

Springer

Springer Oceanography

The Springer Oceanography series seeks to publish a broad portfolio of scientific books, aiming at researchers, students, and everyone interested in marine sciences. The series includes peer-reviewed monographs, edited volumes, textbooks, and conference proceedings. It covers the entire area of oceanography including, but not limited to, Coastal Sciences, Biological/Chemical/Geological/Physical Oceanography, Paleoceanography, and related subjects.

More information about this series at http://www.springer.com/series/10175

Manuel G. Velarde · Roman Yu. Tarakanov
Alexey V. Marchenko
Editors

The Ocean in Motion

Circulation, Waves, Polar Oceanography

Springer

Editors
Manuel G. Velarde
Instituto Pluridisciplinar
Universidad Complutense de Madrid
Madrid
Spain

Alexey V. Marchenko
Department of Arctic Technology
University Center in Svalbard
Longyearbyen
Norway

Roman Yu. Tarakanov
Shirshov Institute of Oceanology
Russian Academy of Sciences
Moscow
Russia

ISSN 2365-7677 ISSN 2365-7685 (electronic)
Springer Oceanography
ISBN 978-3-319-71933-7 ISBN 978-3-319-71934-4 (eBook)
https://doi.org/10.1007/978-3-319-71934-4

Library of Congress Control Number: 2017962541

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Abstract

The book is dedicated to the 70th anniversary of Eugene Georgievich Morozov, a Russian observational oceanographer. His field of interests includes internal waves, ocean circulation, deep ocean currents, and Arctic oceanography. The book includes description of the most important features obtained from the observational studies of internal waves and especially tidal internal waves in the Global Ocean, augmented with their theoretical, numerical, and laboratory modeling. The book also includes contributions on physics of surface waves and their interaction with internal waves. There are also chapters with detailed discussions of the processes of interaction between internal waves and deep currents in the ocean, particularly currents of Antarctic Bottom water in abyssal fractures. The oceanic circulation and processes in fjords, including those that occur under ice as well as ice properties are also discussed in this book.

Part I
Personal Reminiscences

Honorary Note. Evgeny Georgievich Morozov: A Life at Sea as a Devoted Ocean Observer

Manuel G. Velarde, Roman Yu. Tarakanov and Alexey V. Marchenko

On March 15, 1946, Eugene G. Morozov, an outstanding field oceanographer, was born in Moscow. He is Doctor of Sciences in Physics and Mathematics. Since 1993 he is Head of the Hydrological Processes Laboratory at the Shirshov Institute of Oceanology (Russian Academy of Sciences) in Moscow. He has been President (2011–2015) of the International Association for the Physical Sciences of the Ocean (IAPSO, in 1999 he was elected member of the IAPSO Executive Committee). He is Laureate of the Shirshov Medal (2016). He was also a member of the executive committees of the International Union of Geodesy and Geophysics (IUGG) and Scientific Council for Oceanic Research (SCOR). He is the Chairman of the Oceanographic Section at the Russian National Geophysical Committee, which is a Russian link to the IUGG. Dr. Morozov is member of the editorial boards of four journals: Izvestia Atmospheric and Oceanic Physics, Russian Journal of Earth Sciences, Oceanological Researches, and Fundamental and Applied Hydrophysics.

He married in 1971 to Elena A. Morozova who is a scientist in laser physics. They have two sons: Nicolay (1977, a computer programmer) and Victor (1979, also a field oceanographer).

In the 1970s, a new generation of oceanographers began to work at the Shirshov Institute of Oceanology. Like Eugene, they were graduates from the Department of the Ocean Dynamics of the Moscow Institute of Physics and Technology. This department headed by the outstanding Russian theoretician in hydromechanics

M. G. Velarde (✉)
Instituto Pluridisciplinar, Universidad Complutense de Madrid, Madrid, Spain
e-mail: mgvelarde@pluri.ucm.es

R. Yu.Tarakanov
Shirshov Institute of Oceanoloy, Russian Academy of Sciences, Moscow, Russia
e-mail: rtarakanov@gmail.com

A. V. Marchenko
Department of Arctic Technology, University Center in Svalbard, Longyearbyen, Norway
e-mail: aleksey.marchenko@unis.no

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_1

Professor S. S. Voit (1920–1987) brought to science many talented explorers-researchers in the field of geophysical hydromechanics. Five decades later, those graduates became leading scientists in several laboratories of the Institute. Former students that started their work at the Institute gained experience from the previous generations of oceanographers in the conditions of the real ocean. They became acquainted with the knowledge of experienced scientists and tried to find new interesting scientific problems. At first glance it seems easy as there are quite many open problems yet to be solved by young physical oceanographers. They appear clear when participating in oceanographic cruises so to get in touch with the World Ocean. Its endless surface embraces a whole set of hydrophysical phenomena: waves, currents, eddies, ocean level fluctuations, and the interaction with the atmosphere.

Eugene started his marine activity in 1967 by participating in a coastal expedition on the Black Sea. His first oceanic expedition was in 1969 exploring the western boundary currents in the Atlantic. He published his first paper in 1971. His supervisors were Professors V. G. Kort, S. S. Voit, and Yu. A. Ivanov.

During his life at sea Eugene conducted research on many large scale currents in the World Ocean: the Gulf Stream, the Antarctic Circumpolar Current, the Kuroshio, the California Current, the subsurface undercurrents in the Indian (Tareev Current) and Atlantic (Lomonosov Current) oceans and the Falkland Current. He also studied mesoscale eddies and rings of the jet currents. Morozov has also been an Arctic oceanographer. He has explored the Barents and Kara seas. Further he has investigated oceanographic processes near glaciers that descend into the sea (collaboration with A. V. Marchenko).

Eugene has focused on exploring and understanding internal waves and ocean circulation. One major result of his research was establishing the fact that the main source of internal tide generation is interaction between the barotropic tide and submarine ridges. Previously it was considered that the significant generation occurs over continental slopes. He constructed a chart of internal tide amplitudes in the World Ocean. He demonstrated that approximately a quarter of the barotropic energy dissipation is spent in the generation of internal tides over submarine ridges. Other studies (some in collaboration with M. G. Velarde, Yu. A. Ivanov, E. N. Pelinovsky, S. V. Nikitin, and A. V. Marchenko) led to understanding internal waves in straits like the Strait of Gibraltar, the Kara Gates Strait, the Bab el Mandeb Strait, the Drake Passage, and also in the open ocean. In his most recent long-lasting oceanographic activity he has explored the flows in the abyssal channels of the Atlantic: Vema Channel, Romanche and Chain fracture zones, Kane Gap, Vema Fracture Zone, and others (collaboration with W. Zenk and R. Tarakanov). Under his expeditionary scientific leadership, four deep underwater cataracts in the fractures of the Mid-Atlantic Ridge were found.

Since 1969, Eugene made 47 voyages in the ocean including 34 in the Atlantic, 4 in the Indian, 4 in the Pacific, 4 in the Southern and one in the Arctic Ocean. The expeditions on oceanographic ships became his beloved work. Routes of his expeditions in the Global Ocean are shown in the attached figure/World Ocean chart (Fig. 1). He was the chief scientist in 15 trans-Atlantic expeditions. He also

participated in 15 coastal expeditions. These expeditions included underwater scuba diving in shallow waters, work on zodiacs and other small boats, work on ice, driving snow mobiles, research near glaciers descending to the fjords.

The following items are worth highlighting: 1969–1972: the western boundary currents in the Atlantic, 1973–1974: the subsurface equatorial undercurrent in the Indian Ocean, 1974: his first expedition to the Pacific Ocean, 1975: his first polar research in the Norwegian Sea and 1976: research of deep brines in the depressions of the Red Sea. During 1977–1978 he participated in two expeditions of the joint USSR-USA POLYMODE Experiment. He was a member of the USSR POLYMODE delegation to the USA in 1977 and 1978. In 1981: voyage to the typhoon region of the Pacific Ocean, 1982–1983: expedition in the Southern Ocean south of New Zealand, close to the Antarctic continent, 1985: the Soviet large-scale oceanographic program Mesopolygon (mesoscale eddy investigation experiment), in which 76 moorings were deployed simultaneously in the Atlantic, 1986: voyage through the East Pacific. In 1987 he was with the greatest Soviet oceanographic field experiment named Megapolygon (large scale study of the eddy field) with synchronous deploying of 173 moorings in the Northwest Pacific. In 1989 the exploration of the Gulf Stream rings with participation in the research on long-distance sound propagation in the deep acoustic waveguide in the Atlantic. In 1990 observation of the atmosphere-ocean interaction in the region of the Great Newfoundland Bank. In the 1990s, there was a break in field oceanography activities due to the dramatic financial position of science in the former USSR. Nevertheless, in 1995 he participated in the World Ocean Circulation Experiment working with T. Joyce at the Woods Hole Oceanographic Institution as an expert on the WOCE data quality evaluation of the CTD data collected in the WOCE expeditions. In 1999 the field oceanographic explorations of the Shirshov Institute of Oceanology resumed and then he participated in a number of voyages repeating some of the WOCE hydrographic sections in the Atlantic. In 2003 the Russian scientists resumed their oceanographic explorations in the Drake Passage. Since that time Dr. Morozov was the chief scientist in two oceanographic research studies

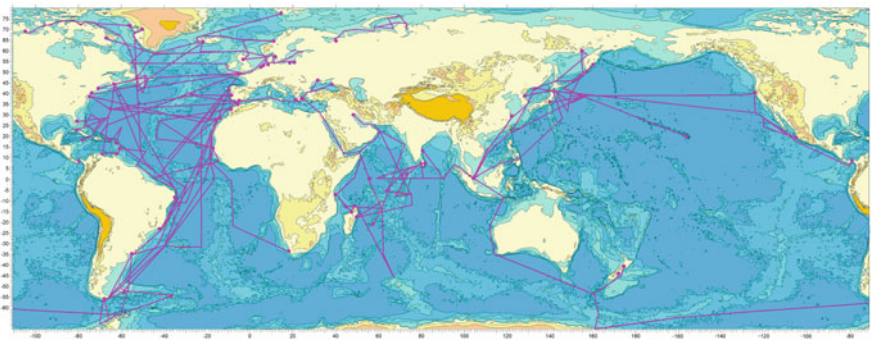


Fig. 1 Expedition routes followed by Eugene G. Morozov in the Global Ocean

in the Drake Passage. In 2004, Dr. Morozov made a voyage as a scientist in the Baltic and North seas on the tall sail ship "Kruzenshtern". In 2007 was a voyage in the Kara Sea. In the period 2002–2017 there were annual voyages to study Antarctic Bottom Water flows in the deep channels and fracture zones of the Atlantic. During these voyages he visited Antarctica. The results of his studies were published in several leading journals of the profession and in seven research frontier book: Oceanic Internal Waves, Nauka, Moscow, 1985; Surface and Internal Waves in the Artic Seas, Gidrometeoizdat, Leningrad, 2002 (editor); Waves in Geophysical Fluids, Springer, Berlin, 2006 (chapter); Abyssal Channels in the Atlantica Ocean, Water Structure and Flows, Springer, Berlin, 2010; A Touch of the Ocean, Nauka, Moscow, 2013 (editor); Without Bounds. A Scientific Canvas of Nonlinearity and Complex Dynamics, Springer, Berlin, 2013 (coeditor and author of a chapter); Oceanic Internal Waves: Observations, Analysis and Modeling. A Global View, Springer, Berlin, 2018.

Many colleagues from several countries have fruitfully collaborated with Eugene G. Morozov. Among them are M. G. Velarde (Spain, collaboration started in 1998), W. Zenk (Germany), G. Weatherly (USA), L. Maas, H. van Haren, and T. Gerkema (Holland), A. V. Marchenko (Norway), V. Vlasenko (UK), G. Nagy, and D. Severov (Uruguay) and I. Ansorge (South Africa).

Life at sea is the kind of experience very few are capable to endure for a long period in life. Piling up different periods, all together Eugene spent 10 years of his life at sea. Deploying buoys, recovering them a year or more later (with the frustration of a number of them being lost or significantly deteriorated), and obtaining reliable and valuable field data (data collecting, processing and analysis) not always leads to standard scientific publications. Sure they permit to produce useful technical reports and monographs. Eugene has coauthored over two hundred research papers and he is coauthor or scientific coordinator of six leading research monograph books, one of them mentioned above.

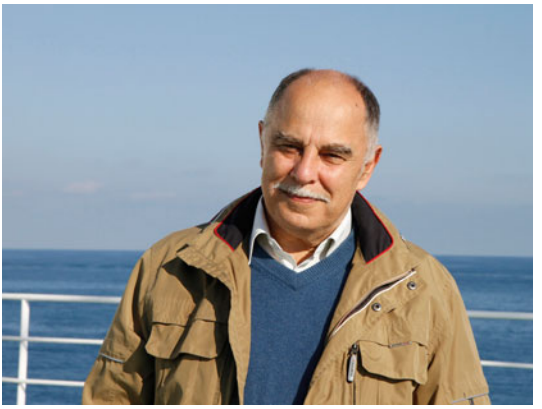
Eugene has made numerous scientific presentations to the Assemblies of IUGG and IAPSO, the European Geosciences Union and Asia Oceania Geosciences Society as well as presentations at research institutions in the USA, Canada, Brazil, Argentina, Uruguay, Chile, Japan, China, Australia, New Zealand, South Africa, Madagascar, Spain, Germany, the Netherlands, France, Norway, Sweden, the United Kingdom, Austria and Italy. He convened several symposia of International scientific assemblies.

Last but not least, Eugene has been active in yachting sports. During his sport activity he won several prizes for winning racing competitions on the yachts of Finn single handed Olympic class in the Moscow yachting competitions. It has been contended that the Finn is the most physical and tactical single hander sailboat in the world.

We wish, for our outstanding colleague and friend Evgeny Georgievich Morozov, many years of enjoyment of life, family and Science. In addition, this will be for the future benefit of the scientists of the World Ocean.

Gallery: An Ocean Scientist and His Life at Sea

Manuel G. Velarde, Roman Yu. Tarakanov and Alexey V. Marchenko



Eugene G. Morozov 2017

M. G. Velarde (✉)
Instituto Pluridisciplinar, Universidad Complutense de Madrid, Madrid, Spain
e-mail: mgvelarde@pluri.ucm.es

R. Yu. Tarakanov
Shirshov Institute of Oceanoloy, Russian Academy of Sciences, Moscow, Russia
e-mail: rtarakanov@gmail.com

A. V. Marchenko
Department of Arctic Technology, University Center in Svalbard, Longyearbyen, Norway
e-mail: aleksey.marchenko@unis.no

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_2

1953: Just before entering the elementary school

1968: Before graduating from the Moscow Institute of Physics and Technology

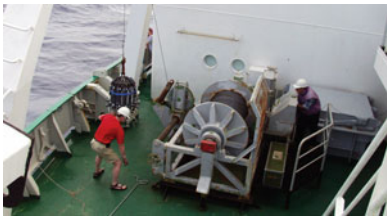
1969: Moscow, on the way to yachting racing. Yachting sport strongly influenced his decision to change specialty from gas dynamics (rockets) to hydrodynamics (oceanography)

1974: Colombo, Sri Lanka. Press conference onboard the R/V “Vityaz” together with his supervisor V. G. Kort (both standing)

1975: Norwegian Sea. International cooperation. Among the USSR and USA scientists before the POLYMODE experiment

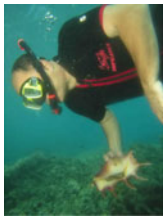
1977: Together with Yu. Ivanov, his supervisor on internal wave research with New York in the background

2002: Antarctica, Cuberville Island

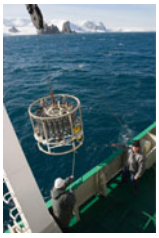


2002: Work in the Vema Channel. E. Morozov is operating the winch, and S. Pisarev is holding the rosette

2004: Sail ship Kruzenstern; at the yardarm end



2005: Underwater reasearch in the shallow tropical waters. Scuba diving was part of research in the shelf regions



2005: Antarctica, Bransfield Strait. Together with R. Tarakanov

2002: Drake passage, shooting an XBT sensor

2007: Perugia, Italy. SCOR/IAPSO work group on mixing

2009: Around SBE carousel; T. Demidova and V. Liapidevskii (front row), R. Tarakanov (left), E. Morozov (center)

2009: Crossing the equator in the Atlantic, Eugene is playing the role of Neptunus

2009: Spitsbergen, steering a zodiac together with A. Marchenko

2011: Arctic research of the interaction between oceanic water and glacier fresh water

2013: Five IAPSO presidents E. Morozov, L. Mysak, S. Imawaki, P. Rizzoli, R. Muench

2014: Scientists and crew onboard cruise 38 of the oceanographic ship « Akademik Sergey Vavilov»

2015: IAPSO Assembly in Prague. T. McDougal, H. Bryden, D. Smythe-Wright, T. Yamagata, W. Munk, E. Morozov, L. Mysak

2017: Together with Manuel G. Velarde

2017: Together with his son Nicol and grandsons Yaroslav (6) and Maxim (10)

2017: Together with his wife Elena, son Victor, and granddaughter Zoya

New Steps of the Modern Oceanography: Reminiscences of My Work with Evgeny Georgievich Morozov

Victor G. Neiman

Eugene G. Morozov knows not only how to perform field measurements in the most interesting locations, but also how to process and analyze these measurements applying all his expertise and knowledge of physics and mathematics. As a result, just right after each expedition, a draft of a scientific paper has been ready. The spectrum of the scientific coverage of his publications includes almost all aspects of the hydrological regime of the ocean. However, the main part of his publications is dedicated to his continuous interest and scientific passion: the internal tides. When we hear his presentations about the new results in this field of oceanography his colleagues and I feel involved in his scientific enthusiasm. Eventually, it led to a paper together with Eugene on internal waves (published in Doklady Earth Sciences in 2000), despite the fact that I was already an elderly scientist who has been studying oceanic currents during all my life. The title of the paper was "Energy decay of internal tidal waves generated near submarine ridges". The analysis in this paper was related to the fact that internal tides generated over large-scale underwater slopes can propagate over more than 1000 km before their energy decreases by one order of magnitude. Thus, this research clarified the spatiotemporal characteristics of the dissipation of internal tides, which appeared important for the correct estimates of the energy balance of the oceanic circulation needed for the global climate estimates. Not more than three years passed since this publication, and I was again involved be Eugene and his colleagues in the research of internal tides in the western region of the Arctic Ocean. This work was published in 2003 also in Doklady Earth Sciences under the title "Internal tidal waves in the Barents Sea". In this case my professional interest to the thermodynamic regime of the Barents Sea was not occasional. Not long before this publication we had finished a research on estimating the response of the Barents Sea to the El Niño events, which

V. G. Neiman (✉)
Shirshov Institute of Oceanology, Russian Academy of Sciences,
Moscow, Russia
e-mail: vneiman2007@yandex.ru

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_3

also influences the hydrodynamics of the ocean. The research performed by Eugene Morozov with his colleagues in the Barents Sea showed the existence of internal waves with extremely high amplitudes in the region, which were caused by the interaction of the barotropic tide with the bottom topography. This was a manifestation of the specific hydrophysical peculiarity of the polar ocean region, which does not require a steep bottom slope for the generation of strong internal tides as at low latitudes. Another research in the Arctic seas was based both on the numerical modeling and field data from the moorings in the Kara Gates Strait. The amplitudes of internal tides in this strait are as high as 50 m, while the depth of the sea is only 230 m. The physical properties of internal tides in the Kara Gates Strait are similar to the well known internal waves in the Strait of Gibraltar also analyzed by Eugene.

Our joint research on some regional problem of internal waves is only part of his fundamental long-term research of this important oceanographic problem. I am proud to use this opportunity and report of my participation in this scientific research together with Eugene. His personal contribution to the solution of the problem of internal tide generation is recognized in the world community of oceanographers. He was the first to show that the main cause of the generation of internal tides is the interaction between the currents of the barotropic tide and bottom topography, which are the regions of the major dissipation of the barotropic tide energy. He is close to the implementation of his long wish to publish a monograph on his researches of internal tides. In the end of 2017, he is planning to submit a manuscript of the book to the Springer publishers. Knowing his carefulness and professionalism I think that this monograph would become quite useful to the oceanographers, especially to those who study waves in the ocean.

Without doubts I agreed to the suggestion of Eugene to consult the analysis of new oceanographic data in the Southern Ocean and propagation of Antarctic Bottom Water. Fortunately, my scientific background included the knowledge of the processes around Antarctica, which I gathered during the Russian Antarctic expeditions in the 1950–1970s. In 2006, we published a paper together with Eugene on the spreading of Antarctic Bottom Water in the equatorial Romanche Fracture Zone. One of the deepest depressions of the Atlantic Ocean exceeding 7400 m is located in this region. We tested a hypothesis how Antarctic Bottom Water formed in the Weddell Sea conserves the climatic signal of the surface waters after propagating a distance over 10,000 km. In 1967, during the first cruise of the RV "Akademik Kurchatov", I participated in the hydrophysical measurements in the Romanche Fracture Zone. Almost in 40 years after this cruise, in 2005 in cruise 19 of the R/V "Akademik Ioffe" Eugene carried out similar measurements on a new level. After we analyzed these observations including the historical data of six other expeditions, in which the instruments were lowered in the Romanche Fracture Zone from 1948 to 1972, we found that the climate signal exists in the deep waters as a positive temperature trend in the thermal structure of Antarctic Bottom Water. The trend was not high, only a mere thousand fraction of Celsius degree per year. Certainly, it would be very exciting to attribute the deep water temperature trend to the recent global climate change. However, currently we do not even try to speculate on this theme understanding that such an approach needs much more

observations in the Atlantic Ocean deep. There I just asked myself, who among us could be able to organize it? I found nobody but Eugene.

In 2010, appeared his book "Abyssal Channels in the Atlantic Ocean: Water Structure and Flows". In this research monograph, the authors presented the unique experimental results of Antarctic Bottom Water transport from the Southern Hemisphere to the north through a series of deep channels in submarine ridges. The data collected in many expeditions allowed him to trace the pathway of Antarctic Bottom Water, from the Antarctic Weddell Sea to the North Atlantic. One can hardly imagine all the difficulties he had to overtake to organize these expeditions including finances, logistics, technical, and weather conditions that were on the way to his goal. For example, the choice of the locations of stations in individual fractures was sometimes similar to finding a needle in a haystack. Deployments of moorings in narrow fractures at strictly specified points also required mental and physical efforts of the scientists and crew. However, everything was done, and the moorings were recovered. The book on the flows in the abyssal channels for many years will be a complete description of the flows of the bottom water and Antarctic influence in these specific regions of the ocean.

No doubt that the success of Eugene's work on this logistically difficult problem of Antarctic water spreading was provided by his organizing talent and a unique ability to find the solution of the problems in any situation, even when other people could not do this. He always keeps the scientific line formulated in the 19th century by the first Russian oceanographer admiral S.O. Makarov: "We write what we see; we do not write what we do not see". In other words, Eugene in his personal publications or in the publications with coauthors never allowed presenting the desired result as the real one. He knows that facts but not fantasies are the scientific basis. Everything, which is not supported by facts, cannot be trusted.

The time when Eugene entered our oceanographic science became the period of one of the active and fruitful missions of scientists to study the features of the World Ocean that demand the multidisciplinary organizing research work. Eugene has selected professionalism, timeliness, and accuracy as the necessary qualities of his work. He always concentrates on the main features of research. His work at the Institute facilitated the appearance of new significant scientific ideas among the colleagues. Owing to his initiative the modern approaches and technologies of the scientific research were practically implemented. These qualities of Eugene were recognized within the Shirshov Institute and well beyond it. For instance, in 2011, he was elected President of the International Association for the Physical Sciences of the Ocean (IAPSO). We hope that his energy will help the promotion and progress of the new ideas that will help humanity in the research and understanding of the World Ocean.

Fifty Years of Collaboration with Evgeny Georgievich Morozov

Boris N. Filyushkin

I met Eugene for the first time in the summer of 1967. He was a student of the Moscow Institute of Physics and Technology on a marine practice in the Black Sea (Gelendzhik). We were conducting an experiment on the spatiotemporal variability of internal waves over the shelf of Blue Bay and in the open sea. Three moorings with temperature recorders were installed on the shelf. In 1970, he started his work at the Shirshov Institute of Oceanology. The data collected in 1967 on the shelf became the basis for his first publication together with his supervisor Yu. A. Ivanov.

It happened that the first acquaintance with internal waves during his marine practice became his dominating activity for many future years. In 1975, he defended his Ph.D. Thesis on internal waves. In 1985, he published a book on internal waves, and in 1989, he defended his doctoral dissertation on internal waves. In 1995, he published in the Deep-Sea Research Journal his well-known paper "Semidiurnal internal wave global field". In this research he used the available data in many regions of the ocean and a model to show the main role of submarine ridges in the generation of internal tides in the ocean. He studied the internal tides in many geographical regions: the Gibraltar, Bab-el-Mandeb, and Kara Gates straits. In 2008, he began his field studies of internal waves under ice in the fjords of Spitsbergen.

The next stage of his scientific activity became the research of the bottom currents in the Atlantic and propagation of Antarctic Bottom Water in the abyssal channels of the Atlantic. This program started in 2002 and has been continuing until present. A total of 18 expedition cruises to the North and South Atlantic has been made since then within this program. The measurements were conducted in the regions of the abyssal channels: the Vema Channel (31° S), Romanche Fracture Zone (equator), Vema Fracture Zone (11° N), Kane Gap (9° N), and many others.

B. N. Filyushkin (✉)
Shirshov Institute of Oceanology,
Russian Academy of Sciences, Moscow, Russia
e-mail: borisfil@yandex.ru

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_4

High level of his professionalism and experience as a field oceanographer made possible to perform unique observations in the deep narrow channels in these regions. The peculiarity of these experiments is that the temperature, salinity, and velocity profiles must be measured at depths of 4000–5000 m with an accuracy of positioning not exceeding 100 m, while the slopes of the channels are very steep. Sometimes the width of the channel at the bottom was as narrow as 200 m and the instruments were safely lowered almost to the bottom. The profiles stopped at a distance of 2–5 m over the seafloor.

He conducted these measurements together with his colleagues from the Institute of Oceanology and with the scientists from Germany and the Netherlands. His researches became fundamental in our understanding of Antarctic Bottom Water spreading in the Atlantic.

It is well known among oceanographers that the personality of a scientist is fully revealed in long expeditions at sea. We spent almost a year with Eugene in three long expeditions onboard the research ships in 1970, 1984, and 1989. We became good friends. Eugene is a very sociable and communicative person. At the same time, he is a very strict person with respect to himself and those people who work with him. He showed his professional qualities during the field works in the ocean on the deck and in the laboratory. He is a good manager of field works and professional specialist in operation.

His targeted wish to gain the planned result allowed him to make contributions to the solution of two important oceanographic problems. His authority is known by many oceanographers around the world.

This note is a contribution of the chronology of his activity in oceanography. Eugene is full of new plans. He is working to realize them. I am sure that new interesting studies and publications will appear.

In all his life and scientific activity he was always aimed at the specific result, which should be gained without any deviations aside.

In conclusion I would like to use the English saying: “I wish to have such a neighbour!”

Part II
Scientific Contributions: Internal and Surface Waves

Internal Undular Bores in the Coastal Ocean

Roger Grimshaw and Chunxin Yuan

Variable-Coefficient Korteweg-de Vries Equation

Large amplitude internal wave trains are commonly observed in the coastal ocean, see the reviews by [9, 10, 15, 19, 20, 25] and the book by [27]. Since these are long nonlinear waves it is now widely accepted that the basic paradigm for these waves is based on the Korteweg-de Vries (KdV) equation, first derived in this context by [2, 3] and subsequently by many others, see the aforementioned references. In the usual physical variables to describe internal waves in the coastal ocean the KdV equation is

$$\eta_t + c\eta_x + \mu\eta\eta_x + \lambda\eta_{xxx} = 0\,. \tag{1}$$

Here $\eta(x,t)$ is the amplitude of the modal function $\phi(z)$, defined by

$$\left\{\rho_0(c-u_0)^2\phi_z\right\}_z + \rho_0 N^2\phi = 0\,, \quad \text{for} \quad -h < z < 0\,, \tag{2}$$

$$\phi = 0 \quad \text{at} \quad z = -h\,, \quad (c-u_0)^2\phi_z = g\phi \quad \text{at} \quad z = 0\,. \tag{3}$$

This also serves to define the phase speed c. Here $\rho_0(z)$ is the background density field, stably stratified so that $\rho N^2 = -g\rho_{0z} > 0$, $u_0(z)$ is a background horizontal current and h is the undisturbed fluid depth. This modal system in general has an infinite set of solutions, ordered by the phase speeds, so that the lowest (zero) mode is the barotropic mode with the fastest speed $c \approx \sqrt{gh}$, followed by the first internal mode with a much slower speed, and so on. The coefficients μ, λ are given by

$$I\mu = 3\int_{-h}^{0} \rho_0 \left(c-u_0\right)^2 \phi_z^3\, dz\,, \tag{4}$$

R. Grimshaw (✉) · C. Yuan
Department of Mathematics, University College London, London, UK
e-mail: r.grimshaw@ucl.ac.uk

C. Yuan
e-mail: chunxin.yuan.14@ucl.ac.uk

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_5

$$I\lambda = \int_{-h}^{0} \rho_0 \left(c - u_0\right)^2 \phi^2 \, dz \,, \tag{5}$$

$$I = 2 \int_{-h}^{0} \rho_0 \left(c - u_0\right) \phi_z^2 \, dz \,. \tag{6}$$

The KdV equation (1) is integrable, and the outcome of a *localised* initial condition is a finite set of rank-ordered solitary waves and some small-amplitude dispersing radiation, see [1, 29]. However, here we are concerned with the undular bore solution, which can be found as the outcome of a step initial condition by using the Whitham modulation equations, see [18], or more generally is a representation of the development of a solitary wave train from a broad initial condition. Such internal undular bores are also generated by transcritical flow over topography, see [9, 11, 13].

When the depth h, and background current u_0, density ρ_0 vary slowly in the horizontal direction with x, the KdV equation (1) is replaced by a variable-coefficient KdV (vKdV) equation. first derived in the Boussinesq approximation in the absence of a background current by [26], and then in the general case by [8], see also [14, 15, 30]. It has the same form as (1) with two extra terms,

$$\eta_t + c\eta_x + \frac{cQ_x}{2Q}\eta + \mu\eta\eta_x + \lambda\eta_{xxx} + \sigma\eta = 0 \,, \tag{7}$$

$$Q = c^2 I \,, \quad I\sigma = -\int_{-h}^{\eta_0} \phi\phi_z F_{0z} dz \,, \quad F_0 = \rho_0(u_0 u_{0x} + w_0 u_{0z}) + p_{0x} \,. \tag{8}$$

Here the modal equation depends also on x parametrically, that is $\phi = \phi(z : x), c = c(x)$, and hence the coefficients μ, λ, Q also depend (slowly) on x. The coefficient Q ensures conservation of wave action flux $Q\eta^2$ at the linear long wave order, and σ arises due to the presence of a body force in the basic horizontal momentum equation, needed in general whenever the basic density field and basic current vary in the horizontal direction. In (8) p_0 is the basic pressure, such that $p_{0z} = -g\rho_0$, and w_0 is the basic vertical velocity such that $u_{0x} + w_{0z} = 0$.

It is convenient to transform this to the "spatial" evolution form, asymptotically equivalent to (7)

$$X = \int_{x_0}^{x} \frac{dx}{c} - t \,, \quad T = \int_{x_0}^{x} \frac{dx}{c} \,, \tag{9}$$

$$\eta_T + \frac{Q_T}{2Q}\eta + \nu\eta\eta_X + \delta\eta_{XXX} + \sigma\eta = 0 \,, \tag{10}$$

$$\nu = \frac{\mu}{c}\,, \quad \delta = \frac{\lambda}{c^3}\,. \tag{11}$$

A further simplification is to absorb the factor Q,

$$U = Q^{1/2}\eta\,, \quad U_T + \frac{\nu}{Q^{1/2}} UU_X + \delta U_{XXX} + \sigma U = 0\,. \tag{12}$$

Next a further transformation yields the canonical form with only one variable coefficient,

$$U_\tau + \alpha UU_X + U_{XXX} = 0\,, \tag{13}$$

$$\text{where} \quad U = RA\,, \quad R = \exp\,(-\int_0^\tau \beta\, d\tau')\,,$$

$$\tau = \int_0^T \delta\, dT\,, \quad \alpha = \frac{R\nu}{\delta Q^{1/2}}\,, \quad \beta = \frac{\sigma}{\delta}\,. \tag{14}$$

The coefficients R, α, β vary with τ. Equation (13) is the vKdV equation of interest, especially useful when there is a polarity change, that is the nonlinear coefficient μ and hence α change sign. It has two conservation laws

$$U_\tau + \{\frac{\alpha U^2}{2} + U_{XX}\}_X = 0\,, \tag{15}$$

$$\{\frac{U^2}{2}\}_\tau + \{\frac{\alpha U^3}{3} + UU_{XX} - \frac{U_X^2}{2}\}_X = 0\,. \tag{16}$$

In the conservative case when $\beta = 0, R = 1$, these represent conservation of mass and wave action flux respectively.

There is now a considerable literature on the application of the variable-coefficient KdV equation (7), or the transformed equation (13), to model the propagation of internal solitary waves propagating over variable topography, see the reviews by [14, 15]. However, observations of oceanic internal solitary waves rarely show instances of isolated internal solitary waves, which instead usually occur in a wave train resembling an internal undular bore. While it is the case that the leading waves in these wave trains can each be well approximated as a solitary wave, it is desirable to consider the dynamics of an internal undular bore as a whole, and so take account of possible interactions between the waves in the wave train. Hence, recently [16] used the variable-coefficient KdV equation (13) to simulate the behaviour of internal undular bores propagating over variable topography. The same model was used by [17] to model tsunami waves propagating up a slope. In both cases a special emphasis was placed on the front of the undular bore which can each be represented by a simplified model as a solitary wave train. In this article we review and supplement the results

obtained by [16] for internal undular bores. In section "Solitary Wave Train" we present the modulation equations for a solitary wave train, and describe briefly how these may be used when waves propagate in a region where $\alpha = \alpha(\tau)$ varies. Then in section "Numerical Simulations of a Process Model" we discuss how a solitary wave train behaves when there is a change of polarity, that is, there is a critical point where α changes sign. We conclude in section "Discussion".

Solitary Wave Train

The equations for the asymptotic description of a solitary wave train were initially developed by [7]. They can also be obtained from the Whitham equations for a modulated periodic wave train by taking the solitary wave limit, see [5, 16]. That approach is summarised in the Appendix, and here we just present the outcome.

$$U = a\,\mathrm{sech}^2(\gamma\theta; m)\} + d\,, \quad \theta = k(X - V\tau)\,, \quad V - \alpha d = \frac{\alpha a}{3} = 4\gamma^2 k^2\,, \tag{17}$$

with three parameters, the amplitude a, the pedestal d and the wavenumber k, to be determined. The pedestal d satisfies

$$d_\tau + \alpha d d_X = 0\,, \tag{18}$$

and so can be regarded as a known quantity. The equation for conservation of waves (30) provides one equation for k

$$k_\tau + (kV)_X = 0\,. \tag{19}$$

and the third equation is (37), that is

$$\{\frac{a^2}{k\gamma}\}_\tau + V\{\frac{a^2}{k\gamma}\}_X + \frac{a^2}{k\gamma}\alpha d_X = 0\,, \tag{20}$$

The pair (19, 20) form a nonlinear hyperbolic system for a solitary wave train, and can be solved explicitly. The wavenumber k can be eliminated from (19) and (20) to yield

$$\mathscr{A}_\tau + (\alpha d + \frac{\alpha a}{3})\mathscr{A}_X + \mathscr{A}\alpha d_X = 0\,, \quad \mathscr{A} = \{\frac{a^3}{\alpha}\}^{1/2}\,. \tag{21}$$

This is now an equation for the amplitude a alone, and is readily solved using characteristics. Then, with a, d and hence V also known, the wavenumber k can be found from (19) which is a linear hyperbolic equation for k.

Formally the modulation equations for a *single* solitary wave can be found by considering modulations in τ alone. Then (18) shows that d is a constant which can be set to zero without loss of generality. Next, it follows from (19, 21) that $k, \mathscr{A}$ are

constants, and the latter yields the well-known adiabatic expression $a^3 \propto \alpha$. However, this requires taking the limit $k \to 0$ which is inconsistent with the assumption that the width scale of the wave should be much less than that of the variable medium defined here by the variation in α. Instead, as is now well-known, see the reviews by [10, 14, 15], a multi-scale asymptotic expansion for the solitary wave should be used, which confirms the adiabatic expression $a^3 \propto \alpha$ due to conservation of wave action flux, but also reveals that the deforming solitary wave is accompanied by a trailing shelf needed to conserve the total mass. Assuming here that $\alpha > 0$, so that the solitary wave is one of elevation, the trailing shelf has an amplitude at the solitary wave location proportional to $\alpha^{-8/3}\alpha_\tau$ and hence is an elevation or depression shelf according as α is increasing or decreasing, $\alpha_\tau > 0$ or $\alpha_\tau < 0$. An analogous result holds when $\alpha < 0$. The essential difference between the solitary wave and the solitary wave train is that in the latter, the mass is represented by the independent parameter d, whereas the solitary wave has only one parameter, say the amplitude a whose variation is already determined, while the solitary wave mass is $2a/\gamma$ and is then not a constant and varies as $|\alpha|^{-1/3}$.

Of special interest is the case when there is a change of polarity, that is, there is a critical point $\tau = \tau_c$ where α changes sign, say from $\alpha > 0$ to $\alpha < 0$, implying that solitary waves are waves of elevation for $\tau < \tau_c$ and waves of depression for $\tau > \tau_c$. The behaviour of a single solitary wave as it passes through this critical point is now well understood, see the reviews by [10, 14, 15] and the recent study by [16]. As the solitary wave approaches the critical point, its amplitude decreases as $\alpha^{1/3}$ but at the same time the amplitude of the trailing shelf of depression grows as $\alpha^{-8/3}$. Close to the critical point, when the solitary wave and the trailing shelf have comparable amplitudes, the adiabatic behaviour breaks down. The whole structure passes through the critical point and in $\tau > \tau_c$ generates a depression rarefaction wave connected to the original zero level by an undular bore of elevation waves. The modulation theory for a solitary wave train presented above can be used to describe both processes before and after the critical point.

When α is a constant, taken here as positive without loss of generality, a full representation of an undular bore can be found by seeking a similarity solution of the Whitham modulation equations (30, 31, 32), where the modulus m depends only on $X/\alpha U_0 \tau$, see [6, 18, 29] and the review by [4]. This describes an expanding wave train connecting a zero level at the front where $m \to 1$ to a mean level $U_0 > 0$ at the rear where $m \to 0$. At the front the leading wave is a solitary wave of amplitude $2U_0$ and at the rear the waves are linear sinusoidal waves. However, in a variable medium, when as here $\alpha = \alpha(\tau)$, although the Whitham modulation equations are again available, see the Appendix, it would seem that no such simple wave solution is available to describe the evolution of an undular bore. A recent study by [5] of a water wave undular bore propagating up a slope (that is $\alpha > 0$ increases) demonstrated that the deformation at the front of the undular bore is essentially non-adiabatic. Briefly, it is argued that if the undular bore retains its structure as a single-phase wave train, then the jump U_0 is preserved, and so then the leading solitary wave would have a constant amplitude $2U_0$. But this is inconsistent with the result that the leading solitary wave amplitude should behave as $\alpha^{1/3}$. The resolution of this inconsistency is

the formation of a solitary wave train ahead of the undular bore. This solitary wave train can be described by the reduced set of modulation equations (18, 19, 21), and there is a region where the rear of the solitary wave train interacts with the undular bore, forming a two-phase wave interaction. The rear part of the undular bore retains its shape, where pertinently we note that in the similarity solution for the constant coefficient case, α occurs only in the variable $x/\alpha U_0 \tau$, implying that the effect of a variable α is in some sense equivalent to adjusting the time scale. This leading solitary wave train can be described by a similarity solution of the modulation equations (18, 19, 21),

$$d = 0\,, \quad a = \frac{3\alpha^{1/3} X}{\chi}\,, \quad X_m(\tau) < X < X_M(\tau)\,, \quad \chi = \int_0^\tau \alpha^{4/3}(\tau')\, d\tau'\,. \tag{22}$$

At the head $X = X_M(\tau)$ of the wave train, the amplitude is $a_M = 3\alpha^{1/3} X_M/\chi$, while at the rear $X = X_m(\tau) > 0$ of the wave train, the amplitude is $a_m = 3\alpha^{1/3} X_m/\chi$. The speed of each solitary wave is $V = \alpha a/3$ (17) which yields the asymptotic position of each wave as $X = C\chi$ where C is a "constant" which depends on the initial position of each solitary wave. This must be matched to the following undular bore to determine $X_{M,m}(\tau)$, and the details of that are described by [5]. But note that if it is assumed that the leading wave in the following undular bore has amplitude $2U_0$ then we may expect that $a_m \approx 2U_0$ and $a_M \approx 2U_0\alpha^{1/3}$, assuming that initially $\alpha = 1$. The dynamics of an undular bore as it passes through a critical point was examined recently by [16]. Before the critical point, the behaviour of the leading solitary wave train can again be described by the modulation theory as above, where now $\alpha > 0$ decreases towards zero. The behaviour after the critical point is described below.

Numerical Simulations of a Process Model

Numerical simulation of an internal undular bore propagating up a slope were reported by [16] using the transformed vKdv equation (13). Here we review and supplement those results. For the numerical simulation we use a pseudo-spectral method where the nonlinear term is evaluated in physical space.

We consider a process model to detect interesting dynamics, and choose the coefficient α in the transformed equation (13) to model internal waves propagating up a slope, in two cases, one when the nonlinear coefficient α increases, and one when it decreases with change of sign, causing a polarity change. Note there is no loss of generality in choosing the initial value of $\alpha > 0$, as otherwise one one can make the transformation $U \to -U$. Thus we set $\alpha = \alpha(\tau)$ varying monotonically from $\alpha = 1, \tau = 0$ to some constant value $\alpha = \alpha_a, \tau \geq \tau_a$. Then there are two cases, either $\alpha_a > 1$, or $\alpha_a < 0$ (a change of polarity). Specifically the coefficient α is given by

$$\alpha = 1 + (\alpha_a - 1)\tanh(K\tau)\,, \tag{23}$$

where $K, \tau_a, K\tau_a \geq 1$ are chosen so that α varies smoothly and slowly from 1 at $\tau = 0$ to α_a at $\tau = \tau_a$. There are two cases, either $\alpha_a > 1$ for propagation up a slope, or $\alpha_a < 0$ for propagation up a slope and through a critical point of polarity change.

The initial condition $U(X,0) = U_{ic}(X)$ is either (1) a KdV solitary wave or (2) a modulated cnoidal wave representation of an undular bore in the constant coefficient KdV equation evolving from a step of height $U_0 > 0$ at time $\tau = -\tau_1$, as developed by [6, 18, 29] and subsequently used by many authors.

$$(1)\,:\, U_{ic}(X) = U_0\,\mathrm{sech}^2(\kappa X)\,, \quad U_0 = 12\kappa^2\,, \tag{24}$$

$$(2)\,:\, U_{ic}(X) = U_0\,\mathrm{ENV}(X)\{2m\mathrm{cn}^2(\kappa(X - V\tau_1); m) + 1 - m\}\,, \tag{25}$$

$$-U_0\tau_1 < X < \frac{2U_0\tau_1}{3}\,, \quad V = \frac{U_0}{3}\{1+m\}\,, \quad U_0 = 6\kappa^2k^2\,,$$

$$X = \frac{U_0\tau_1}{3}\{1 + m - \frac{2m(1-m)K(m)}{E(m) - (1-m)K(m)}\}\,.$$

In case (1) the evolving solitary wave has a time scale of $(\kappa V)^{-1}$ where the speed $V = 4\kappa^2$, and so to be slowly varying we choose $K \ll \kappa V = 4\kappa^3$. In case (2) the initial undular bore occupies a domain of length $L_{ub} = 5U_0\tau_1/3$ and we choose τ_1 so that $L_{ub} \gg 1$. The envelope ENV(X) is chosen to be very close to a box of height 1, and of very long length $L > L_{ub}$, occupying a domain to contain the initial undular bore. The front end of the box is placed precisely at the front end of the undular bore $X = 2U_0\tau_1/3$, but the rear end is chosen far away from the rear end of the bore. As it evolves in the constant coefficient KdV equation (that is $\alpha = 1$) the leading wave is a solitary wave of amplitude $2U_0$ and so we again choose K so that $K \ll 4\kappa^3$ where here $U_0 = 6\kappa^2k^2$. Note that the wavelength $2\pi/k$ is a free parameter.

A numerical simulation of (13) using the solitary wave initial condition for the case when α increases is shown in Fig. 1. As expected, this exhibits adiabatic behaviour with the amplitude increasing from the initial value $U_0 = 4$ to $U_0\alpha_a^{-1/3} = 4.58$ at $\tau = \tau_a$. At same time a trailing shelf of very small amplitude can be seen. Then in Fig. 2 we show a case with the same solitary wave initial condition, but with a polarity change in which α varies from $\alpha = 1$ to $\alpha = -1$. As the solitary wave of elevation approaches the critical point, its amplitude decreases as $U = U_0\alpha^{1/3}$ but at the same time the amplitude of the trailing shelf of depression grows as $\alpha^{-8/3}$. Close to the critical point, when the solitary wave and the trailing shelf have comparable amplitudes, the adiabatic behaviour breaks down, and the whole structure passes through the critical point into the region $\tau > \tau_c$, generating a depression rarefaction wave connected to the original zero level by an undular bore of elevation waves. The waves in the undular bore section have a shorter length scale than the rarefaction wave, giving the appearance of a solitary wave train of depression solitary waves riding on an elevation pedestal. These solitary waves ride down the pedestal,

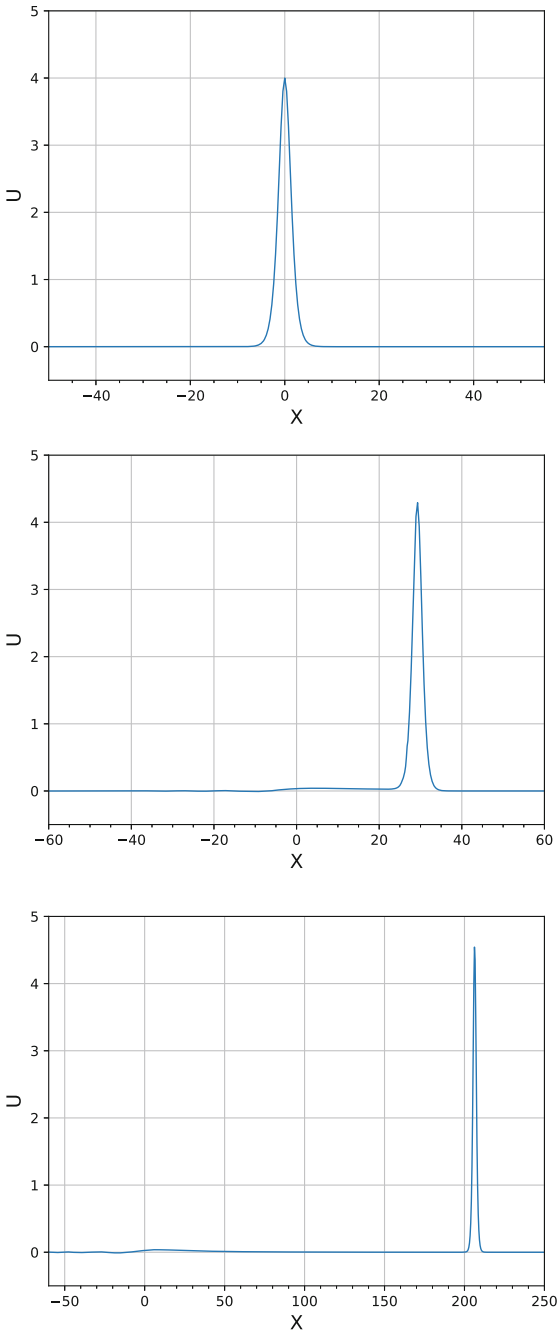


Fig. 1 A simulation of the vKdV equation (13) when $\alpha(\tau)$ varies from 1 to 1.5 as specified by (23) for the solitary wave initial condition (24) with $U_0 = 4$ and $K = 0.03$, $\tau_a = 100$ ($K\tau_a = 3$); the top panel is at $\alpha = 1$, the middle panel is at $\alpha = 1.25$ and the bottom panel is at $\alpha = 1.5$

and at $\tau = \tau_a$ the leading solitary wave has reached the zero level at the head of the rarefaction wave. In this region where $\alpha < 0$, the solitary wave train equations (18, 21) have a similarity solution

$$d = \frac{X - X_0}{\eta}\,, \quad \eta = \int_{\tau_c}^{\tau} \alpha(\tau')\, d\tau'\,, \quad X < X_0\,, \tag{26}$$

$$\mathscr{A} = -\frac{1}{\eta}\{\frac{3(X - X_0)}{\eta\xi}\}^{3/2}\,, \quad \xi = \int_{-\infty}^{\eta} \frac{|\alpha(\eta')|^{1/3}\, d\eta'}{|\eta'|^{5/3}}\,, \quad \mathscr{A} = \{\frac{a^3}{\alpha}\}^{1/2}\,. \tag{27}$$

Note that here $\alpha < 0$ and so $\eta < 0$, ensuring that the rarefaction wave $d > 0$ in $X < X_0$. The determination of X_0 requires a detailed matching with the solution at the critical point, beyond the scope of this present review. The rarefaction wave (26) can only extend to a point $X - X_0 = -L_r(\eta)$ where $L_r(\eta)$ is likewise undetermined. But the mass of the rarefaction wave is then $-L_r^2(\eta)/2\eta$ and this can be approximately equated to the initial solitary wave mass $2U_0/\kappa = 2(12U_0)^{1/2}$ (24), thus giving an approximate expression for $L_r(\eta)$. The expression (27) for the solitary wave amplitude a holds on the domain $-L_r(\eta) < X - X_0 < -L_s(\eta)$ where the upper bound $L_s(\eta)$ determines the amplitude of the leading solitary wave, that is $a_s = -3|\alpha|^{1/3}L_s/|\eta|^{5/3}\xi$. The values of a_s, L_s are undetermined and requires matching with the solution at the critical point, beyond the scope of this review article. However, an approximate estimate can be based on the assumption that since the emerging solitary wave train is the leading edge of an undular bore resolving the jump at the rear of the rarefaction wave, and then $a_s = 2L_r/\eta$, where in turn L_r is estimated from conservation of mass, as above. Note that when $\alpha < 0$ is a constant, then as $\tau \to \infty, \eta \sim \alpha\tau, \xi \sim 3/2|\alpha|^{1/3}\tau^{2/3}$ and then $a \sim -2(X - X_0)/\alpha\tau$, whose envelope is a rarefaction wave. The outcome from the numerical simulation in Fig. 2 shows qualitative agreement with all the above features. For the parameters in this simulation $U_0 = 4$, and so the initial mass is 13.86, which yields a value $L_r \approx 5.26|\eta^{1/2}|$ and then $a_s \approx 10.52|\eta|^{-1/2}$; at $\tau = \tau_a = 100$, $\eta = -62$ (26), and then $a_s \approx 1.34$, in good agreement with the numerical simulation.

Next we examine how an undular bore behaves when propagating on a slope. As for the case of a solitary wave, there are two main situations, one where the nonlinear coefficient α increases, and the other where α decreases and changes sign at a critical point. In the first case, which is similar to the study by [5] of a surface wave train propagating up a slope, we expect from the discussion in section "Solitary Wave Train" that the leading waves in the undular bore will form a solitary wave train ahead of the undular bore, with amplitudes deforming adiabatically as $\alpha^{1/3}$. A typical numerical simulation of (13) using the undular bore initial condition (25) is shown in Fig. 3. First, note that the effect of the front end of the enclosing envelope $ENV(X)$, which of numerical necessity is a smoothed out version of a rapid change, is to truncate the amplitudes of the leading waves in the initial undular bore. Without the envelope

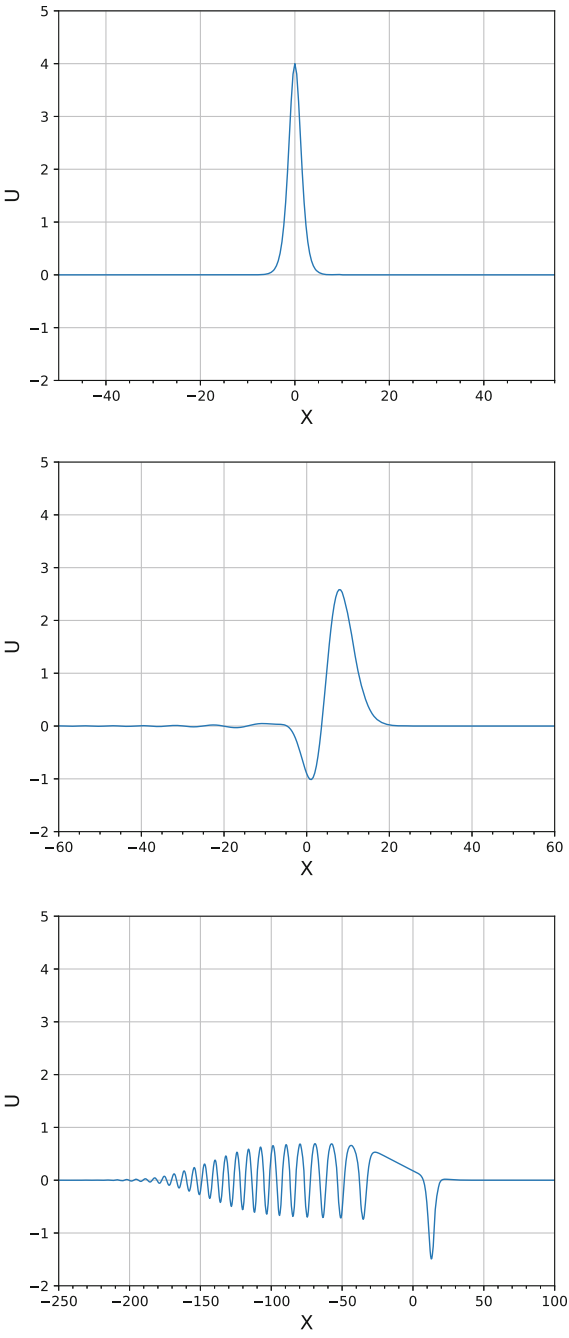


Fig. 2 A simulation of the vKdV equation (13) when $\alpha(\tau)$ varies from 1 to -1 as specified by (23) for the solitary wave initial condition (24) with $U_0 = 4$ and $K = 0.03$, $\tau_a = 100$ ($K\tau_a = 3$); the top panel is at $\alpha = 1$, the middle panel is at $\alpha = 0$ and the bottom panel is at $\alpha = -1$

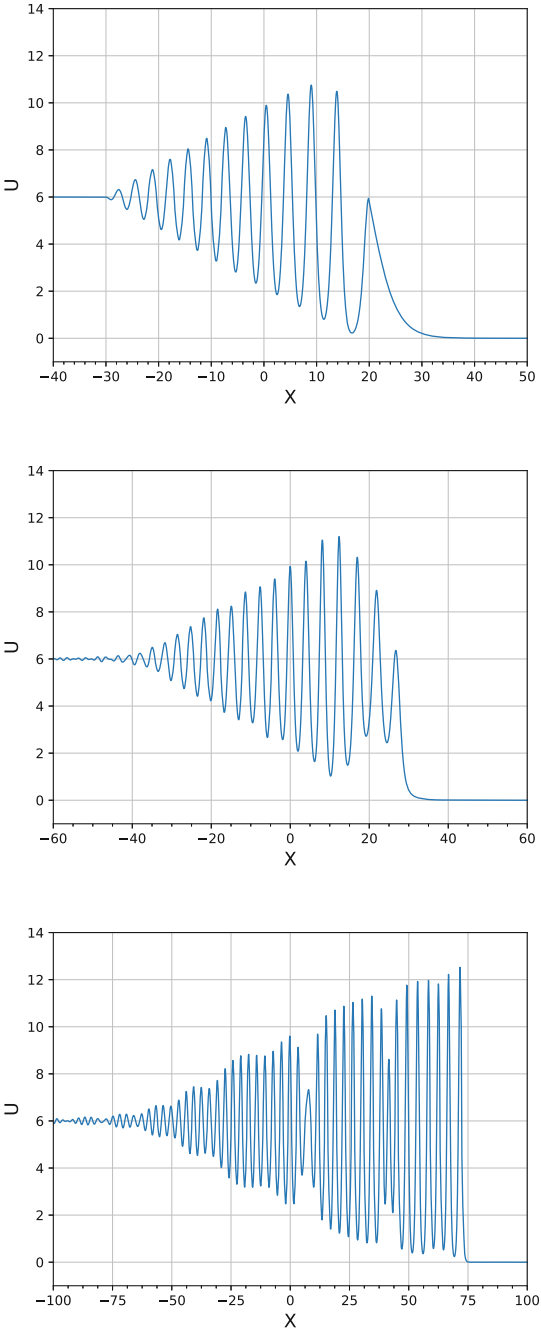


Fig. 3 A simulation of the vKdV equation (13) when $\alpha(\tau)$ varies from 1 to 1.5 as specified by (23) for the undular bore initial condition (25) with $U_0 = 6, \tau_1 = 5$ and $K = 0.3, \tau_a = 10$ ($K\tau_a = 3$); the top panel is at $\alpha = 1$, the middle panel is at $\alpha = 1.25$ and the bottom panel is at $\alpha = 1.5$

the leading wave would have an amplitude of $2U_0 = 12$ in this simulation, whereas we see that instead the largest wave in the undular bore has an amplitude of about 10. Nevertheless, when taking this complication onto account, we see the formation ahead of the undular bore of two solitary wave trains, and each has the expected linear profile (22). At the end of the simulation at $\tau - = \tau_a$ the leading wave has an amplitude about 12, consistent with the asymptotic estimate $10\alpha_a^{1/3} = 11.45$. We interpret the presence of two wave solitary wave trains as due to the truncation at the initial front causing some ambiguity about the which is the leading wave. But we note that the leading wave in the second solitary wave train has an amplitude of about 11, consistent with an asymptotic estimate of $9.5\alpha a^{1/3}$ where we see that 9.5 is the amplitude of the second highest wave in the initial undular bore.

In the second case, when α decreases to zero and changes sign after a critical point, we show a typical numerical simulation of (13) with the initial condition (25) in Fig. 4. As for the previous case the truncation at the head of the undular bore does have an effect. Nevertheless, up to the critical point (middle panel) we see similar behaviour to that shown in Fig. 3 except that now the amplitudes in the emerging solitary wave trains decrease. After the critical point, the leading wave in these emerging solitary wave trains has deformed in a manner similar to that described above in Fig. 2 for a single solitary wave passing through a critical point. We can clearly see the formation of an elevation rarefaction wave, and some still quite small solitary waves of depression riding on this pedestal. On the other hand the rear of the undular bore essentially retains its shape, and shows little evidence of any change in amplitude. Essentially it is behaving as a linear wave. A more detailed analysis of this and the previous case in Fig. 3 can be found in [16]. Hence the essential dynamics as described above on section "Numerical Simulations of a Process Model" is not changed when expressed in terms of the original physical variables, although in particular the amplitude magnitudes may be altered significantly.

Discussion

Our purpose in this brief review article is to show how the variable-coefficient Korteweg-de Vries equation (7) can be used to model the propagation of internal undular bores over the continental slope. As we noted in the Introduction there have been many studies on the application of the variable-coefficient KdV equation (7), or the transformed equation (13). to model the propagation of internal solitary waves propagating over variable topography, see the reviews by [14, 15]. However, since observations of oceanic internal solitary waves often show that they occur as part of a wave train, it is desirable to consider the dynamics of an internal undular bore as a whole. This was undertaken recently by [16] who used the variable-coefficient KdV equation (13) to simulate the behaviour of internal undular bores propagating over variable topography, and in this article we have reviewed this application in a briefer and less detailed manner.

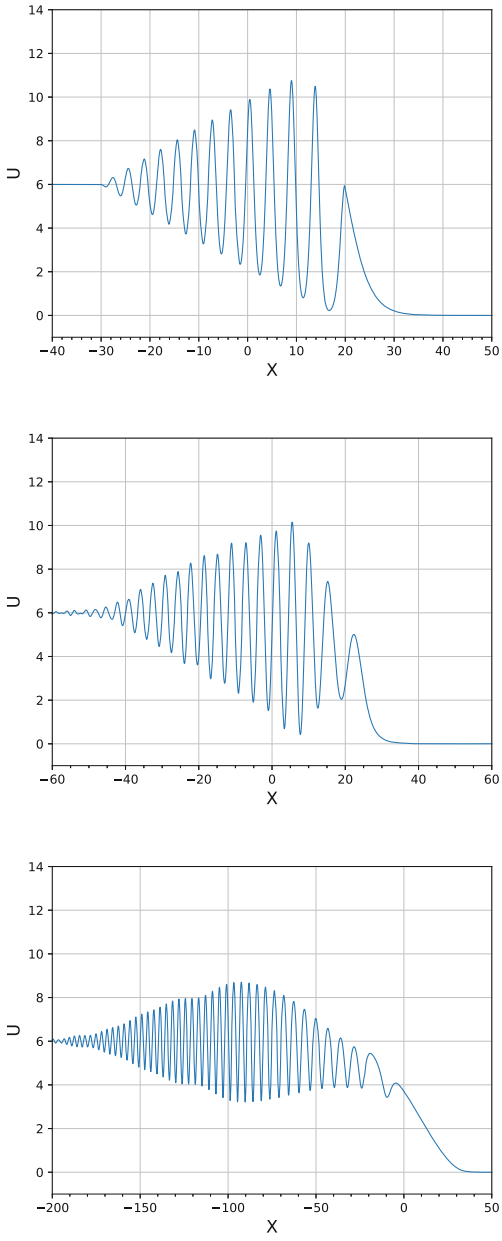


Fig. 4 A simulation of the vKdV equation (13) when $\alpha(\tau)$ varies from 1 to -1 as specified by (23) for the undular bore initial condition (25) with $U_0 = 6, \tau_1 = 5$ and $K = 0.3, \tau_a = 10$ ($K\tau_a = 3$); the top panel is at $\alpha = 1$, the middle panel is at $\alpha = 0$ and the bottom panel is at $\alpha = -1$

It is important to note that the theory and modelling is developed for the transformed equation (13) which has only one variable coefficient, namely α, rather than for the original equation (7) which has five variable coefficients, namely $c, Q, \mu.\lambda, \sigma$. The back transformation from (13) to (7) involves only amplitude factors and changing the time and space scales, see the sequence of equations (9, 10, 11, 12). Hence the essential dynamics is not changed when expressed in the physical variables. Although, importantly the amplitudes may be significantly altered during the back transformation, the polarity of the waves is not changed. We have focussed on two principal scenarios. In the first there is no polarity change, that is, the coefficient α does not change sign. Then as the undular bore propagates in the variable medium, the leading waves separate and form a solitary wave train, with subsequent adiabatic behaviour. In the second case there is a polarity change, that is the coefficient α changes at a critical point. The passage of the undular bore through the critical point does involve some non-adiabatic behaviour, which leads after the critical point to the formation of a rarefaction wave of the same polarity as the original waves in the undular bore, and the formation of a solitary wave train of the opposite polarity riding on the pedestal formed by the rarefaction wave. Application of simulations of the transformed equation (13) to actual oceanic situations seem to be quite rare, but we note the recent study by [23].

Acknowledgements RG was supported by the Leverhulme Trust through the award of a Leverhulme Emeritus Fellowship.

Appendix

Here we summarise the derivation of the equations (18, 19, 20) presented in [16]. When the coefficient α in (12) is a constant the KdV equation supports a periodic travelling wave, $U(X - V\tau)$, the well-known cnoidal wave,

$$U = a\,\{b(m) + \mathrm{cn}^2(\gamma\theta; m)\} + d\,, \quad \theta = k(X - V\tau)\,, \tag{28}$$

$$V - \alpha d = \frac{\alpha a}{3}\left\{\frac{2-m}{m} - \frac{3E(m)}{mK(m)}\right\} = 4\gamma^2 k^2\left\{2 - m - \frac{3E(m)}{K(m)}\right\}. \tag{29}$$

Here $cn(x; m)$ is the Jacobian elliptic function of modulus $m, 0 < m < 1$, and $K(m)$ and $E(m)$ are the elliptic integrals of the first and second kind, The expression (28) has period 2π in θ so that $\gamma = K(m)/\pi$, while the spatial period is $2\pi/k$. The (trough-to-crest) amplitude is a and the mean value over one period is d. It is a three-parameter family with parameters k, m, d say. As the modulus $m \to 1$, this becomes a solitary wave, since then $b \to 0$ and $\mathrm{cn}(x) \to \mathrm{sech}(x)$, while $\gamma \to \infty$, $k \to 0$ with $\gamma k = \Gamma$ fixed. As $m \to 0$, $b \to -1/2$, $\gamma \to 1/2$, $\mathrm{cn}(x) \to \cos(x)$, and it reduces to a sinusoidal wave $(a/2)\cos(\theta)$ of small amplitude $a \sim m$ and wavenumber k.

The Whitham modulation theory allows this cnoidal wave to vary slowly with τ, X, that is the wavenumber k, modulus m and mean level d vary slowly with τ, X. The Whitham modulation equations describing this variation can be obtained by averaging conservation laws, the original Whitham method, see [28, 29] or by exploiting the integrability of the constant-coefficient KdV equation, see [21] for instance. Because here we are concerned with the case when $\alpha = \alpha(\tau)$ varies slowly with τ, and so the variable-coefficient KdV equation (13) is not integrable, we will use the original Whitham method, readily adapted to this present case. A similar strategy was used by [24] for a frictionally perturbed KdV equation. An alternative method developed by [22] for a perturbed KdV equation is not available here because to use it one must make a change of variable in (13) $\tilde{U} = \alpha U$ to generate a KdV equation for $\tilde{U}$ with a perturbation term of the form $\alpha_\tau \tilde{U}/\alpha$. But as one of our concerns is with the situation when α passes through zero, this approach cannot be used here.

As three modulation equations are needed, we supplement (15, 16) with the equation for conservation of waves,

$$k_\tau + (kV)_X = 0\,. \tag{30}$$

The remaining two modulation equations are obtained by inserting the cnoidal wave solution into the conservation laws (15, 16) and averaging over the phase θ. The outcomes are

$$d_\tau + \alpha M_X = 0\,, \quad M = <\frac{U^2}{2}>\,, \tag{31}$$

$$M_\tau + P_X = 0\,, \quad P = <\frac{\alpha U^3}{3} - \frac{3U_X^2}{2}>\,, \tag{32}$$

where the $<\cdots>$ denotes a 2π-average over θ. The expression M is given by

$$M = \frac{d^2}{2} + \frac{a^2}{2}\{C_4 - b^2\}\,,$$

$$C_4 = \frac{1}{3m^2K(m)}\{3m^2K(m) - 5mK(m) + 4mE(m) + 2K(m) - 2E(m)\}\,, \tag{33}$$

while that for P is given by

$$P = \alpha\{-\frac{2d^3}{3} + 2dM + a^3\{-\frac{2b^3}{3} + \frac{(1-m)b}{2m} + (b + \frac{1-2m}{2m})\,C_4 + \frac{5}{6}C_6\}\,,$$

$$C_6 = \frac{1}{15m^3K(m)}\{15m^3K(m) - 34m^2K(m) + 23m^2E(m)$$

$$+27mK(m) - 23mE(m) - 8K(m) + 8E(m)\}\,. \tag{34}$$

Here the notation C_4, C_6 denote $<cn^4>, <cn^6>$ respectively, and like $b = -C_2 = -<cn^2>$ depend on the modulus m only.

To obtain the modulation equations for a solitary wave train, we take the limit $m \to 1$ and then $b \sim -1/K(m)$, $C_4 \sim 2/3K(m)$, and $C_6 \sim 8/15K(m)$. To leading order $M \sim d^2/2$ and $P \sim \alpha d^3/3$ and then both equations (31, 32) reduce to the same equation for d alone,

$$d_\tau + \alpha d d_X = 0\,, \tag{35}$$

and so d can be regarded as a known quantity. At the same time, the cnoidal wave expression (28) reduces to

$$U = a\,\mathrm{sech}^2(\gamma\theta; m)\} + d\,, \quad \theta = k(X - V\tau)\,, \quad V - \alpha d = \frac{\alpha a}{3} = 12\gamma^2 k^2\,, \tag{36}$$

with two parameters still to be determined. The equation for conservation of waves (30) provides one equation for k and the second equation is

$$\{\frac{a^2}{k\gamma}\}_\tau + V\{\frac{a^2}{k\gamma}\}_X + \frac{a^2}{k\gamma}\alpha d_X = 0\,, \tag{37}$$

This can be obtained by a more careful consideration of the limit $m \to 1$ in the modulation equations (31, 32) by retaining the terms in $1/K(m)$, or more directly by averaging the wave action conservation law (16) directly for a solitary wave, see [7] and the discussion in [5].

References

1. Ablowitz, M. J., & Segur, H. (1981). *Solitons and the inverse scattering transform*. Philadelphia: SIAM.
2. Benjamin, T. B. (1966). Internal waves of finite amplitude and permanent form. *Journal of Fluid Mechanics*, *25*, 241–270.
3. Benney, D. J. (1966). Long non-linear waves in fluid flows. *Journal of Mathematical Physics*, *45*, 52–63.
4. El, G. (2007). Kortweg-de Vries equation and undular bores. In R. Grimshaw (Ed.), *Solitary waves in fluids*. Advances in Fluid Mechanics (Vol. 47, pp. 19–53). WIT Press.
5. El, G. A., Grimshaw, R. H. J., & Tiong, W. K. (2012). Transformation of a shoaling undular bore. *Journal of Fluid Mechanics, 709*, 371–395.
6. Fornberg, B., & Whitham, G. B. (1978). A numerical and theoretical study of certain nonlinear wave phenomena. *Philosophical Transactions of the Royal Society A*, *289*, 373–404.
7. Grimshaw, R. (1979). Slowly varying solitary waves. I. Korteweg-de Vries equation. *Proceedings of the Royal Society*, *368A*, 359–375.
8. Grimshaw, R. (1981). Evolution equations for long nonlinear internal waves in stratified shear flows. *Studies in Applied Mathematics*, *65*, 159–188.
9. Grimshaw, R. (2001). Internal solitary waves. In R. Grimshaw (Ed.), *Environmental stratified flows* (pp. 1–27). Boston: Kluwer.
10. Grimshaw, R. (2007). Internal solitary waves in a variable medium. *Gesellschaft fur Angewandte Mathematik*, *30*, 96–109.

11. Grimshaw, R. (2010). Transcritical flow past an obstacle. *ANZIAM Journal*, *52*, 1–25.
12. Grimshaw, R. (2015). Change of polarity for periodic waves in the variable-coefficient Korteweg-de Vries equation. *Studies in Applied Mathematics*, *134*, 363–371.
13. Grimshaw, R. H. J., & Smyth, N. F. (1986). Resonant flow of a stratified fluid over topography. *Journal of Fluid Mechanics*, *169*, 429–464.
14. Grimshaw, R., Pelinovsky, E., & Talipova, T. (2007). Modeling internal solitary waves in the coastal ocean. *Surveys in Geophysics*, *28*, 273–298.
15. Grimshaw, R., Pelinovsky, E., Talipova, T., & Kurkina, A. (2010). Internal solitary waves: Propagation, deformation and disintegration. *Nonlinear Processes in Geophysics*, *17*, 633–649.
16. Grimshaw, R., & Yuan, C. (2016). The propagation of internal undular bores over variable topography. *Physica D*, *333*, 200–207.
17. Grimshaw, R., & Yuan, C. (2016). Depression and elevation tsunami waves in the framework of the Korteweg-de Vries equation. *Natural Hazards*, *84*, S493–S511.
18. Gurevich, A. V., & Pitaevskii, L. P. (1974). Nonstationary structure of a collisionless shock wave. *Soviet Physics JETP*, *38*, 291–297.
19. Helfrich, K. R., & Melville, W. K. (2006). Long nonlinear internal waves. *Annual Review of Fluid Mechanics*, *38*, 395–425.
20. Holloway, P., Pelinovsky, E., & Talipova, T. (2001). Internal tide transformation and oceanic internal solitary waves. In R. Grimshaw (Ed.), *Environmental stratified flows* (pp. 31–60). Boston: Kluwer.
21. Kamchatnov, A. M. (2000). *Nonlinear periodic waves and their modulations. An introductory course*. World Scientific.
22. Kamchatnov, A. M. (2004). On Whitham theory for perturbed integrable equations. *Physica D*, *188*, 247–281.
23. Liu, Z., Grimshaw, R., & Johnson, E. (2017). Internal solitary waves propagating through variable background hydrology and currents. *Ocean Modelling*.
24. Myint, S., & Grimshaw, R. (1995). The modulation of nonlinear periodic wavetrains by dissipative terms in the Korteweg-de Vries equation. *Wave Motion*, *22*, 215–238.
25. Ostrovsky, L. A., & Stepanyants, Y. A. (2005). Internal solitons in laboratory experiments: Comparison with theoretical models. *Chaos*, *28*, 037111.
26. Pelinovsky, E. N., Rayevsky, M. A., & Shavratsky, S. K. (1977). The Korteweg-de Vries equation for nonstationary internal waves in an inhomogeneous ocean. *Izvestiya, Atmospheric and Oceanic Physics*, *13*, 226–228.
27. Vlasenko, V. I., Stashchuk, N. M., & Hutter, K. (2005). *Baroclinic tides: Theoretical modelling and observational evidence*. Cambridge University Press.
28. Whitham, G. B. (1965). Nonlinear dispersive waves. *Proceedings of the Royal Society of London A*, *283*, 238–261.
29. Whitham, G. B. (1974). *Linear and nonlinear waves*. Wiley.
30. Zhou, X., & Grimshaw, R. (1989). The effect of variable currents on internal solitary waves. *Dynamics of Atmospheres and Oceans*, *14*, 17–39.

Calculating FRAM's Dead Water

John Grue

Dedicated to Eugene G. Morozov on his 70th Birthday

The internal wave (dead water) resistance on the Polar ship FRAM is obtained by two methods. The first is empirical, based on the original observations (Nansen, F.: Farthest North, Westminster: Archibald Constable and Company, 2 Whitehall Gardens, 1897. Vol. 1). The second is a strongly nonlinear interfacial method in three dimensions. The intersection between the empirical and theoretical resistances determines accurately the ship speed which is investigated varying the depth of the pycnocline, a quantity that was not measured by Nansen. A reduction to a fifth of the usual speed of the FRAM because of the dead water, as observed by Nansen, corresponds to a mid-depth of the pycnocline of slightly less than 4 m while FRAM's draught was 5 m. The wave wake at Froude number slightly above 0.5 is calculated by the nonlinear method. The linear ship wake and dead water resistance are found to be invalid.

Introduction

This paper concerns the dead water problem—the drag force due to the internal wave wake of a ship moving along the surface of a stratified fluid. The body may also be submerged. The phenomenon was first described by Nansen [1] during the Polar Expedition in 1893-96. The observations were later presented with more details by

J. Grue (✉)
Mechanics Section, Department of Mathematics, University of Oslo, Oslo, Norway
e-mail: johng@math.uio.no

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_6

Ekman [2]. Our purpose is to calculate as exactly as possible the internal wave (dead water) resistance, the wave wake and speed of the FRAM, for the conditions in the field. The computations illustrate and enhance the accuracy of the observations.

We first derive from the empirical descriptions of Nansen [1] and Ekman [2] the dead water resistance on FRAM. Then we compute by a strongly nonlinear interfacial model the same resistance on a model vessel of FRAM's dimensions. The intersection between the empirical and computed resistance curves provides a rather accurate estimate of the ship speed. This is then studied as function of the depth of the pycnocline, a quantity that was not measured when the phenomenon was observed. The wave wake at the small range of ship speeds is evaluated.

Ship internal wave wakes have commonly been analysed by linear theory assuming a source, pressure or thin-ship representation [3–6]. The dead water resistance due to the internal wave wake on a semi-submerged slender prolate spheroid was calculated in the linear regime comparing to experiments [7, 8]. Nonlinear models of the dead water problem have been requested because of the poor comparison between field measurements and linear theory [9]. The same lack was expressed in relation to two-dimensional model tank experiments [10].

Strongly nonlinear analysis and calculations have recently been developed for the dead water problem by Grue [11] and Grue et al. [12]. The main point is that the internal wave wake becomes rather prominent when the ship draught, here denoted by b_0, is comparable to the mid-depth of the pycnocline, here denoted by h_0, i.e. $b_0/h_0 \sim 1$. In the subcritical range, the nonlinear dead water resistance deviates from the linear theory by a large amount, when the Froude number Fr defined in (2) is less than 0.85. The differences are essential also for $0.85 < Fr < 0.95$. They are found to be minor at the critical speed. However, at $Fr = 1$ the nonlinear and linear wave fields differ fundamentally [11]. At large supercritical speeds there are essential differences between nonlinear and linear representation of the wave wake and the dead water resistance, where the nonlinear calculations for a large cargo ship cruising along a stratified subarctic fjord compared very well to a set of field observations, while the linear predictions were inferior [12]. Recent three-dimensional experiments with a bluff body in a moderately wide tank support the nonlinear calculations of the colossal internal wave resistance at the small Froude numbers [13].

The present nonlinear calculations very well represent the wave wake and internal wave resistance at the very small subcritical Froude numbers in question. The linear calculations exhibit rather the contrary where an observed asymmetry of the interfacial elevation at the ship is not found. Further, the linear wave wake has a vanishingly small amplitude. The dead water resistance is almost zero.

The paper is organised as follows: section "Nansen's Observations of the Dead Water" describes Nansen's observations of the dead water. The dead water resistance is calculated from the observations. The nonlinear interfacial method is outlined in section "Nonlinear Interfacial Model", and the internal wave resistance force obtained in section "Dead Water Resistance". The intersection between the empirical and computed resistance as function of the Froude number is then discussed. This

provides a rather accurate estimate of the ship speed in the observations. The wave wake at the small actual range of ship speeds is evaluated. Section "Conclusion" provides a conclusion.

Nansen's Observations of the Dead Water

The nonlinear calculations in section "Nonlinear Interfacial Model" and section "Dead Water Resistance" are directly fitted to the original observations made during the FRAM expedition while passing north of Sibiria. The observations of the dead water were described in Nansen [1, pp. 172–177]. Regarding the wave wake as observed by the view of the ocean surface, Nansen noted:

> Dead water manifests itself in the form of larger or smaller ripples or waves stretching across the wake, the one behind the other, arising sometimes as far forward as almost amidships.

The dead water wave wake corresponding to the observations is computed and visualised in Fig. 5 below. Regarding the speed of the FRAM, Nansen noted:

> Our speed was reduced to about a fifth part of what it would otherwise have been.

The descriptions in Ekman [2, pp. 9–11] provide more details and corrections of the original observations. The observations given in the two publications are summarised in Table 1.

The most important among the observations include:

1. The engine was working at full pressure without and with dead water;
2. The speed reduction to a fifth part;
3. The speed of the FRAM without dead water, of 4.5 or 5 knots;
4. Salt water at the level of the bottom cock, at 4 m depth.

The informations 1., 2., 3. are used to calculate, from the observational data, the dead water resistance on Fram, see section "Calculating the Dead Water Resistance from the Observations". These calculations are cross-compared to the nonlinear interfacial calculations of the deadwater resistance on a model of FRAM, see section "Calculations". The actual speed the FRAM had in the observations is subsequently calculated for comparison to the observational data.

The information 4. is used as an estimate of the mid-depth of the pycnocline in the observations. This depth is also used for the interface at rest in the interfacial model, where the upper layer depth is put to $h_0 = 4$ m. This enables an estimate of the internal wave reference speed (layer depths and densities, see Fig. 2):

$$c_0 = \left(\frac{g' h_0}{\rho_0/\rho_1 + h_0/h_1} \right)^{1/2} \simeq 1.02 \mathrm{ms}^{-1}, \tag{1}$$

where the densities corresponding to the field observations are $\rho_0 = 1.0\,\mathrm{kg\,dm}^{-3}$ and $\rho_1 = 1.028\,\mathrm{kg\,dm}^{-3}$, $g' = g\Delta\rho/\rho_1$, and g the acceleration of gravity. For comparison,

Table 1 Observations in Nansen [1, pp. 172–177] and Ekman [2, pp. 9–11]

	Nansen [1]	Ekman [2]
Position	Open water off Taimur Island	The sound between the isle of Taimur and Almqvist Islands
Speed reduction	A fifth part	Same (Ekman p. 11)
Speed in dead water		1.5 or 1 knot (Ekman p. 10)[a]
Speed without dead water		4.5 knots or perhaps 5 knots (Ekman p. 10)[a]
Speed in dead water		1 knot (Ekman p. 11)
Top layer	Drinking water	Fresh (drinking water)
Fresh water layer thickness		No measurement
Level of bottom cock in engine room		4 m[b,c]
Water quality at the at bottom cock	Far too salt to be used in the boiler	Perfect salt water
Ocean water, density		1.03 kg dm^{-3} (Ekman p. 43)[d]

[a]The engine was working at full pressure without and in dead water.
[b]The ship was in motion, thus the pycnocline at the rear of the ship was uplifted.
[c]The draught of FRAM was 5 m or more at that position.
[d]In the present calculations we use the more refined value of 1.028 kg/dm^3 of the density of the ocean water

an upper layer depth of $h_0 = 5$ m gives a reference speed of $c_0 \simeq 1.14$ ms^{-1}. An upper layer depth of $h_0 = 3$ m gives a reference speed of $c_0 \simeq 0.88$ ms^{-1}. In all calculations the layer depth ratio is put to $h_0/h_1 = 1/18$. Equation (1) is used to define the Froude number by (U the ship speed)

$$Fr = U/c_0. \tag{2}$$

Calculating the Dead Water Resistance from the Observations

The mean work (W) of FRAM's engine balancing the combined loss from the frictional and dead water drag is mathematised by Newman [14, p. 28]:

$$W = DU = \tfrac{1}{2}\rho_0 U_1^3 S C_{F_1} = \tfrac{1}{2}\rho_0 U_2^3 S(C_{F_2} + C_{dw}), \tag{3}$$

where D denotes the total drag force without and with the effect of the dead water, U_1 is the speed of the FRAM without dead water, U_2 the speed of the FRAM in dead water, C_{F_1} the frictional resistance coefficient at the speed U_1, C_{F_2} the frictional resistance coefficient at the speed U_2, C_{dw} the dead water resistance coefficient at the speed U_2, S the wetted surface area of the ship, and ρ_0 the water density of the upper layer. Obtaining the drag force by $D = \frac{1}{2}\rho U^2(C_F + C_{\text{ship waves}})$ is a classical

Table 2 Speed U of FRAM, Reynolds number $Re = Ul_0/\nu$, frictional resistance coefficient C_F obtained by the I.T.T.C. and A.T.T.C. lines

U	Re	C_F (I.T.T.C.)	C_F (A.T.T.C.)
5 knots	4.2×10^7	0.237×10^{-2}	0.236×10^{-2}
1 knots	0.84×10^7	0.309×10^{-2}	0.302×10^{-2}

convention. The balance (3) is obtained since the engine of FRAM was working at full speed, without and with the dead water. The dead water resistance coefficient may be obtained by

$$C_{dw} = (U_1/U_2)^3 \, C_{F_1} - C_{F_2}, \tag{4}$$

expressing C_{dw} by C_{F_1} times the inverse speed reduction ratio cubed minus C_{F_2}.

The frictional resistance coefficients $C_{F_{1,2}}$ are obtained from the I.T.T.C. (International Towing Tank Conference) and A.T.T.C. (American Towing Tank Conference) empirical curves [14, Fig. 2.12]. The I.T.T.C. line is given by $C_F = 0.075/(\log_{10} Re - 2)^2$ and the A.T.T.C. line by $0.242/\sqrt{C_F} = \log_{10}(Re \times C_F)$. Both curves depend on the Reynolds number $Re = Ul_0/\nu$, where U is the ship speed, $l_0 = 30$ m the ship length of FRAM and $\nu = 1.79 \cdot 10^{-6}$ m^{2}s^{-1} the kinematic viscosity of the water at 0 °C. Values of Re and C_F for the actual ship speeds are given in Table 2 where the average values of the I.T.T.C. and A.T.T.C. lines are used in the calculations.

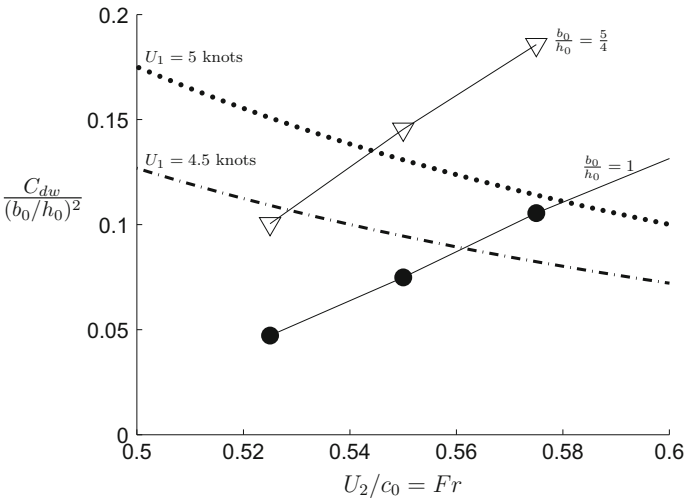


Fig. 1 Dead water resistance coefficient $C_{dw}/(b_0/h_0)^2$ vs. $U_2/c_0 = Fr$. Equation (4) with $U_1 = 5$ knots (dotted line) and $U_1 = 4.5$ knots (dash-dotted line). Nonlinear interfacial calculations with $b_0/h_0 = 1$ (• with solid line) and $b_0/h_0 = 5/4$ (∇ with solid line)

Figure 1 plots the resistance coefficient (4) as function of the speed U_2 in the dead water, for the two different ship speeds U_1 without dead water, of 5 knots or 4.5 knots, corresponding to the range given in Ekman [2]. The dead water resistance coefficient is divided by the ratio between the ship draught and upper layer depth squared, i.e. $(b_0/h_0)^2$. While the draught of the FRAM was $b_0 = 5$ m, the mid depth of the pycnocline was most possibly located at $h_0 = 4$ m depth (or somewhat shallower). In the calculations below we find it convenient to calculate the dead water resistance coefficient divided by $(b_0/h_0)^2$ where the nonlinear interfacial calculations in section "Nonlinear Interfacial Model" and section "Dead Water Resistance" are obtained for the range $0.8 \leq b_0/h_1 \leq 1$ of the nondimensional draught and where values for the actual draught of $b_0/h_0 = 5/4$ are obtained by extrapolation. The level of the pycnocline is further elaborated on in section "Upper Layer Depth. Draught of Ship Model" below.

Nonlinear Interfacial Model

The interaction between the ship moving at forward speed U and the stratified sea is here modelled assuming a two-layer fluid, with the pycnocline replaced by an interface. This is a good approximation when the internal wavelength is great compared to the pycnocline thickness [15]. The two-layer model favourably compares with a three-layer model regarding the wave speeds and wave patterns in the dead water problem [12]. The Froude number of the surface waves based on the ship length, of $U/\sqrt{gl_0} \ll 1$, justifies the rigid lid condition at the upper boundary of the fluid layer.

A definition sketch is found in Fig. 2. Let $\mathbf{x} = (x_1, x_2)$ denote horizontal coordinates in the plane that coincides with the interface at rest, y the vertical coordinate and t time. Following Grue [11] the nonlinear interfacial model asssumes an upper layer of density ρ_0, thickness h_0 at rest and is referred to as fluid 0. Below is a lower fluid 1 of density ρ_1 and depth h_1. Incompressible and irrotational motion in each of the layers are assumed. The motion is governed by the Laplacian potentials ϕ_0 and ϕ_1 in the two layers, where index 0 refers to layer 0 and index 1 to layer 1.

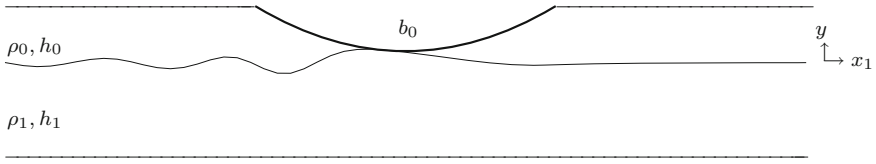


Fig. 2 Two-layer model. Ship geometry of draught b_0 in the upper layer. Layer depths at rest h_0 (upper) and h_1 (lower). Corresponding densities ρ_0 and ρ_1

The kinematic boundary condition at the ship, at position $y = h_0 + \beta(\mathbf{x}, t)$, is given by $\partial\beta/\partial t + W_F = 0$, where F denotes the upper boundary of the upper fluid 0. The shape $\beta(\mathbf{x}, t)$ of the vessel is defined in (18) below. The normal velocity $W_F = \mathbf{U}\cdot\nabla_H\beta$ is connected to the fluid motion by

$$W_F = \frac{\partial\phi_0}{\partial n}\sqrt{1+|\nabla_H\beta|^2} = -\frac{\partial\phi_0}{\partial y} + \nabla_H\beta\cdot\nabla_H\phi_0 \quad \text{at} \quad y = h_0+\beta. \tag{5}$$

Here $\nabla_H = (\partial/\partial x_1, \partial/\partial x_2)$ denotes horizontal gradient. The normal vector points into fluid 0. The potential evaluated at the position of the ship hull is introduced by $\phi_{0F}(\mathbf{x}, t) = \phi(\mathbf{x}, y = h_0 + \beta(\mathbf{x}, t), t)$.

Let the interface be denoted by I and its motion described by the elevation $y = \eta(x_1, x_2, t)$. Values of the potentials along I are introduced by $\phi_{0I}(\mathbf{x}, t) = \phi_0(\mathbf{x}, y = \eta, t)$ and $\phi_{1I}(\mathbf{x}, t) = \phi_1(\mathbf{x}, y = \eta, t)$ at I. Scaled normal velocities along I are introduced by

$$W_I = \frac{\partial\phi_0}{\partial n}\sqrt{1+|\nabla_H\eta|^2} \quad \text{and} \quad V_I = \frac{\partial\phi_1}{\partial n}\sqrt{1+|\nabla_H\eta|^2} \quad \text{at } I, \tag{6}$$

for the upper and lower fluid, respectively, where n points into the upper fluid 0.

Solution of the Laplace Equation

Following Grue [11] the solution of the Laplacian potentials ϕ_0 and ϕ_1 is expressed by a set of integral equations. The velocities W_I, V_I along I and potential ϕ_{0F} along F are obtained using the method of successive approximations, where $W_I = W_I^{(1)} + W_I^{(2)} + W_I^{(3)} + ...$, $V_I = V_I^{(1)} + V_I^{(2)} + V_I^{(3)} + ...$, $\phi_{0F} = \phi_{0F}^{(1)} + \phi_{0F}^{(2)} + \phi_{0F}^{(3)} + ...$ The leading, linear approximation of the set of equations is obtained by the triple $(W_I^{(1)}, V_I^{(1)}, \phi_{0F}^{(1)})$. The quadratic approximation, obtained by $(W_I^{(1)} + W_I^{(2)}, V_I^{(1)} + V_I^{(2)}, \phi_{0F}^{(1)} + \phi_{0F}^{(2)})$, includes the leading coupling between the variables ϕ_0 and $\partial\phi_0/\partial n$, and ϕ_1 and $\partial\phi_1/\partial n$, as well as the excursions η along I and β along F, where β is prescribed. The cubic approximation is obtained similarly.

The quadratic and cubic contributions have been compared to reference solutions in the two-dimensional case. The comparison documents that the expansions converge very rapidly. Thus, for $\eta/h_0 \sim 1$, $|W_I^{(3)}|/|W_I^{(1)} + W_I^{(2)} + W_I^{(3)}| \simeq |W_I^{(3)}|/|W_I| \ll 1$, $|V_I^{(3)}|/|V_I^{(1)} + V_I^{(2)} + V_I^{(3)}| \simeq |V_I^{(3)}|/|V_I| \ll 1$. This means that $W_I^{(1)} + W_I^{(2)}$ and $V_I^{(1)} + V_I^{(2)}$ represent the full nonlinearity in the calculations for elevations up to $\eta/h_0 \sim 1$.

The linear and quadratic contributions are obtained by use of Fourier transform [11]:

$$\mathcal{F}(W_I^{(1)}) = -kT_0\mathcal{F}(\phi_{0I}) - \frac{\mathcal{F}(W_F)}{C_0}, \tag{7}$$

$$\mathcal{F}(W_I^{(2)}) = kT_0\mathcal{F}(\eta W_I^{(1)}) - \mathrm{i}\mathbf{k}\cdot\mathcal{F}(\eta\nabla_H\phi_{0I}) + \frac{\mathrm{i}\mathbf{k}\cdot\mathcal{F}(\beta\nabla_H\phi_{0F})}{C_0}, \tag{8}$$

$$\mathcal{F}(V_I^{(1)}) = kT_1\mathcal{F}(\phi_{1I}), \tag{9}$$

$$\mathcal{F}(V_I^{(2)}) = -kT_1(\eta V_I^{(1)}) - \mathrm{i}\mathbf{k}\cdot\mathcal{F}(\eta\nabla_H\phi_{1I}), \tag{10}$$

$$\mathcal{F}(\phi_{0F}^{(1)}) = \frac{\mathcal{F}(\phi_{0I})}{C_0} - \frac{T_0\,\mathcal{F}(W_F)}{k}, \tag{11}$$

$$\mathcal{F}(\phi_{0F}^{(2)}) = -\frac{\mathcal{F}(\eta W_I^{(1)})}{C_0} + \frac{T_0\,\mathrm{i}\mathbf{k}}{k}\cdot\mathcal{F}(\beta\nabla_H\phi_{0F}) - \mathcal{F}(\beta W_F). \tag{12}$$

Here, $\mathcal{F}$ denotes Fourier transform, $\mathcal{F}^{-1}$ inverse transform, $\mathbf{k} = (k_1, k_2)$ wavenumber vector in Fourier space and $k = |\mathbf{k}|$. Further, $T_0 = \tanh(kh_0)$, $T_1 = \tanh(kh_1)$ and $C_0 = \cosh(kh_0)$. In (8) and (12) $\phi_{0F} = \phi_{0F}^{(1)} + \phi_{0F}^{(2)}$.

Time-Integration

The interfacial motion is expressed by the difference and sum potentials along the interface:

$$\Psi(\mathbf{x},t) = \phi_{1I}(\mathbf{x},t) - \mu\phi_{0I}(\mathbf{x},t),\ \ \Phi(\mathbf{x},t) = \phi_{1I}(\mathbf{x},t) + \phi_{0I}(\mathbf{x},t),\ \ \text{at } I, \tag{13}$$

where $\mu = \rho_0/\rho_1$ denotes the density ratio. Note that the jump condition in the former of the equations in (13) accounts for interactions between the density jump and ambient pressure fields. The interfacial elevation and jump in potential along I are integrated forward in time using the kinematic and dynamic boundary conditions at the interface, giving

$$\eta_t = V_I = W_I, \quad \Psi_t + g'\eta = \mathcal{NL}_2, \quad \text{at} \quad I. \tag{14}$$

The r.h.s. of the latter equation in (14) is:

$$\mathcal{NL}_2 = -\frac{|\nabla_H\phi_{1I}|^2 - \mu|\nabla_H\phi_{0I}|^2 - (1-\mu)W_I^2}{2+2|\nabla_H\eta|^2} - \frac{-2W_I\nabla_H\eta\cdot\nabla_H(\phi_{1I}-\mu\phi_{0I}) + |\nabla_H\eta\times\nabla_H\phi_{1I}|^2 - \mu|\nabla_H\eta\times\nabla_H\phi_{0I}|^2}{2+2|\nabla_H\eta|^2}, \tag{15}$$

and is evaluated in the nonlinear calculations. In the linear calculations, $\mathcal{NL}_2$ is put to zero. An RK4-scheme is used for the time integration of the Fourier transformed versions of (14).

Dead Water Resistance

The pressure force on the ship along the motion direction due to the internal wave wake is given by $F_1 = \int_F -pn_1 dS$, where integration is over the wetted body surface and the pressure obtained by the Bernoulli equation, $p = -\rho_0(\partial\phi_0/\partial t + \frac{1}{2}|\nabla\phi_0|^2 + gy) + const$. The time dependent force is obtained by use of Gauss' theorem and the transport theorem [11]:

$$F_1 = -\rho_0 \frac{d}{dt} \int_F \frac{\partial \phi_{0F}}{\partial x_1} \beta \, d\mathbf{x} - \rho_0 U \int_F \frac{\partial \phi_0}{\partial x_1} \frac{\partial \beta}{\partial x_1} \, d\mathbf{x} + \frac{\rho_0}{2} \int_F |\nabla \phi_0|^2 \frac{\partial \beta}{\partial x_1} d\mathbf{x}. \tag{16}$$

The internal wave dead water resistance coefficient is defined by

$$C_{dw} = \frac{F_1}{\frac{1}{2}\rho_0 S U^2}, \tag{17}$$

where S denotes the wetted surface area of the ship.

Upper Layer Depth. Draught of Ship Model

The ratio between the ship draught b_0 and the upper layer depth h_0 is an important parameter of the dead water resistance where a value of b_0/h_0 close to unity produces a strong wave wake and force. A large ship volume relative to h_0^3 is another important nonlinearity parameter in the dead water problem [12].

Returning to Nansen [1], regarding the level and vertical extent of the pycnocline, the water was described as salt at the level of the bottom cock of the engine room, at 4 m depth. This is shallower than the draught of FRAM's keel of 5 m. The stratification of the sea was not measured. We note: First, the FRAM was in motion during the observations, where the pycnocline at the subcritical speed becomes uplifted at the rear of the ship. Second, the pycnocline in the observations eventually had some vertical extension.

The averaged depth of the pycnocline at rest might have been at 4 m, shallower, or at 5 m. In the present two-layer calculations we shall assume that the depth of the interface is at $h_0 = 4$ m. We include calculations from [11] using $h_0 = 5$ m.

The FRAM has a length of $l_0 = 30$ m and a width of $w_0 = 11$ m which in nondimensional terms become: $l_0/h_0 = 7.5$ and $w_0/h_0 = 2.75$, respectively ($h_0 = 4$ m). The layer depth ratio in the calculations is put to $h_1/h_0 = 18$ and the density ratio put to $\rho_0/\rho_1 = \mu = 1 - \epsilon$ with $\epsilon \to 0$. The shape of the model ship is given by

$$\beta(x_1, x_2) = -b_0\Big((1 - (2x_1/l_0)^2 - (2x_2/w_0)^2\Big). \tag{18}$$

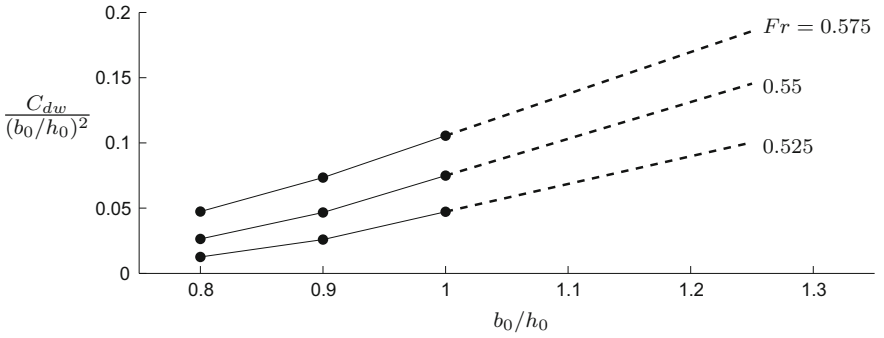


Fig. 3 Calculated (•) and extrapolated (− − −) $C_{dw}/(b_0/h_0)^2$ vs. b_0/h_0 for $0.525 < Fr < 0.575$, $h_0 = 4$ m. Numerical vessel given in (18) with $l_0/h_0 = 7.5$, $w_0/h_0 = 2.75$

Calculations

The calculations are obtained for horizontal domains of length of $L_1 = 240h_0$, width of $L_2 = 100h_0$, and of corresponding resolution of 720 by 400 computational points. The upper layer depth and reference length is $h_0 = 4$ m. The time simulations lasting for $300\, h_0/c_0$ have a gentle ramp-up phase of $100\, h_0/c_0$. The time step is $0.05\, h_0/c_0$. The resistance force attains a steady value.

The coefficient $C_{dw}/(b_0/h_0)^2$ is obtained for nondimensional ship draughts of b_0/h_0 between 0.8 and 1, see Fig. 3. Values for the actual draught of FRAM of 5 m ($b_0/h_0 = 5/4$) are obtained by linear extrapolation. The calculations with a reference depth of $h_0 = 4$ m are compared to the similar calculations with $h_0 = 5$ m in [11] for Froude numbers in the range $0.5 < Fr < 1.1$ where the important differences are observed for the small subcritical speeds of $Fr < 0.8$, see Fig. 4. The dead water resistance coefficients in the linear calculations are much smaller than the nonlinear counterpart when $Fr < 0.9$.

Comparison to the Observations

Returning to the main questions in this paper, what were the conditions of the dead water observations regarding

1. the speed;
2. the mid-level of the pycnocline;
3. the wave wake?

Regarding the ship speed this appears at the intersection between the sets of resistance coefficients, the first derived from the observations, the second from the nonlinear interfacial calculations where both are shown in Fig. 1. If we may assume

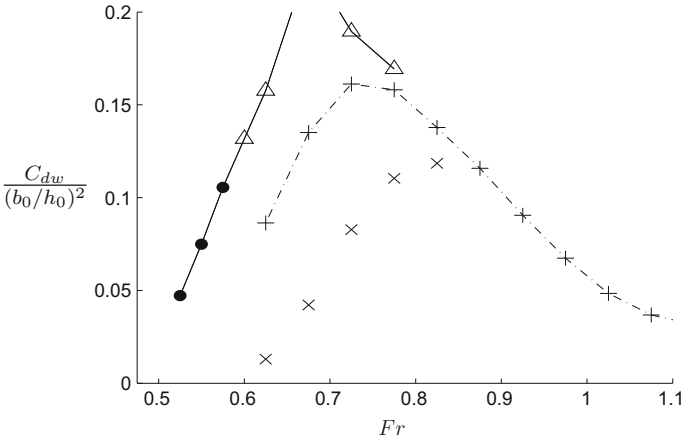


Fig. 4 Dead water resistance coefficient $C_{dw}/(b_0/h_0)^2$ vs. Fr and b_0/h_0, for $h_0 = 4$ m, $b_0/h_0 = 1$ (•, Δ, solid line), $h_0 = 5$ m, $b_0/h_0 = 1$ (+, dash-dotted line). Linear calculations for $h_0 = 5$ m, $b_0/h_0 = 1$ (×)

that i) the extrapolations of the nonlinear calculations are valid for $b_0/h_0 = 5/4$ ($h_0 = 4$ m), and ii) the speed of FRAM without dead water was 5 knots, Fig. 1 estimates the Froude number to be $Fr \simeq 0.54$ corresponding to 0.55 ms^{-1} (1.1 knots). If FRAM's speed without the dead water was 4.5 knots the corresponding estimate gives $Fr \simeq 0.53$ corresponding to 0.54 ms^{-1} (1.08 knots). In the first case the speed reduction is 1/4.5 and in the second case, of 1/4, where the former is closer to Nansen's [1] original estimate (of 1/5) compared to the other. The results in Fig. 1 indicate that a mid-level of the pycnocline at rest of 4 m is reasonable estimate, corresponding to the depth of the bottom cock of the FRAM. For comparison, the nonlinear resistance coefficients with a deeper upper layer level of $h_0 = 5$ m gives an intersection at $Fr \simeq 0.605$ for $U_1 = 5$ knots, obtaining $U_2 \simeq 1.4$ knots and a corresponding speed reduction factor of $1/3.6$. Similarly, assuming an upper layer depth of ∼3 m moves the theoretical curves to the left in Fig. 1 obtaining the intersection at a still smaller Froude number, of approx. 0.5, giving a ship speed U_2 of 0.88 knots. The speed reduction factor then becomes 1/5.7 and is larger compared to Nansen's original estimate.

Regarding the wave wake this is calculated for the two low subcritical speeds of $Fr = 0.55$ and 0.525, see Fig. 5. The nonlinear interfacial elevation of $-0.5 < \eta/h_0 < 0.5$ exhibits an important asymmetry below the ship along the speed direction giving rise to the strong resistance force. The corresponding linear elevations are a factor of 1/10 smaller with $-0.05 < \eta/h_0 < 0.05$. The linear elevation below the ship is symmetrical giving a force that is zero.

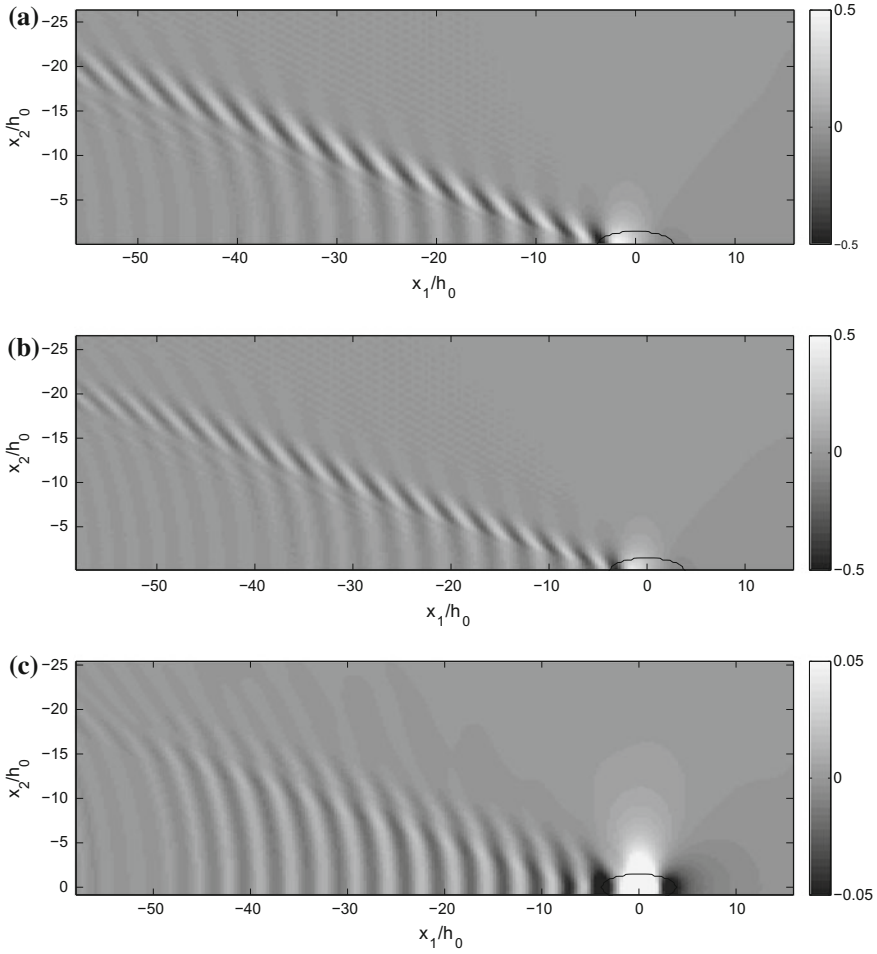


Fig. 5 Nonlinear interfacial elevation η/h_0 (grey scale). Ship draught $b_0/h_0 = 1$. a) $Fr = 0.55$, b) $Fr = 0.525$, c) same as b) but linear calculation. $h_0 = 4$ m

Conclusion

The dead water resistance on the Polar ship FRAM is by two different methods analysed for the conditions fitting to the original observations. The empirical method based on the observational data obtains the resistance as a decaying function of the ship speed in the dead water, while the theoretical resistance increases with the low subcritical Froude number. The intersection between the empirical and computed resistances determines accurately the ship speed. This is investigated as function of the mid-depth of pycnocline where the latter was an unknown parameter in the observations of Nansen [1]. The speed reduction of FRAM in the dead water, to one fifth

of its usual speed, is perhaps one of the most accurate measures of the observations. This speed reduction is found to occur for a mid-depth of the pycnocline of slightly less than 4 m. Calculations illustrate the internal wave wake of a model of FRAM moving at $Fr = 0.525$ and 0.55. Linear theory is incapable of obtaining the internal wave wake and resistance force at the very small Froude numbers.

References

1. Nansen, F. (1897). Farthest North, Westminster: Archibald Constable and Company, 2 Whitehall Gardens (Vol. 1).
2. Ekman, V. W. (1904). XV. On dead-water. In Nansen. F. (Ed.) *The Norwegian north polar expedition* 1893–1896. Scientific results, Brögger, Christiania
3. Hudimac, A. A. (1961). Ship waves in a stratified ocean. *Journal of Fluid Mechanics*, *11*, 229–243.
4. Crapper, G. D. (1967). Ship waves in a stratified ocean. *Journal of Fluid Mechanics*, *29*, 667–672.
5. Keller, J. B., & Munk, W. H. (1970). Internal wave wakes of a body moving in a stratified fluid. *Physics of Fluids*, *13*, 1425–1431.
6. Yeung, R. W., & Ngyuen, T. C. (1999). Waves generated by a moving source in a two-layer ocean of finite depth. *Journal of Engineering Mathematics*, *35*, 85–107.
7. Miloh, T., Tulin, M. P., & Zilman, G. (1993). Dead-water effects of a ship moving in stratified seas. *Journal of Offshore Mechanics and Arctic Engineering*, *115*, 105–110.
8. Tulin, M. P., Yao, Y., & Wang, P. (2000). The generation and propagation of ship internal waves in a generally stratified ocean at high densimetric Froude numbers, including nonlinear effects. *Journal of Ship Research*, *44*(3), 197–227.
9. Watson, G., Chapman, R. D., & Apel, J. R. (1992). Measurements of the internal wave wake of a ship in a highly stratified sea loch. *Journal of Geophysical Research*, *97*(C6), 9689–9703.
10. Mercier, J. M., Vasseur, R., & Dauxious, T. (2011). Resurrecting dead-water phenomenon. *Nonlinear Processes in Geophysics*, *18*, 193–208.
11. Grue, J. (2015). Nonlinear dead water resistance at subcritical speed. *Physics of Fluids*, *27*, 082103. https://doi.org/10.1063/1.4928411.
12. Grue, J., Bourgault, D., & Galbraith, P.S. (2016). Supercritical dead water: Effect of nonlinearity and comparison with observations. *Journal of Fluid Mechanics*, *803*, 436–465. https://doi.org/10.1017/jfm.2016.518.
13. Gou, Y., Xu, W. B., Zhang, X. W., Teng, B. Experiment study on the towing resistance of a barge in a two-layer fluid. In *Proceedings of the 32nd International Workshop on Water Waves and Floating Bodies*, Dalian, China. Retrieved April 23–26, 2017, from http://www.iwwwfb.org.
14. Newman, J. N. (1977). *Marine hydrodynamics*. MIT Press
15. Grue, J., Jensen, A., Rusås, P.-O., & Sveen, J. K. (1999). Properties of large-amplitude internal waves. *Journal of Fluid Mechanics*, *447*, 257–278.

Internal Solitary Waves in a Layered Weakly Stratified Flow

Nikolay Makarenko, Janna Maltseva, Roman Tarakanov and Kseniya Ivanova

Introduction

Internal waves play a significant role in the energy transformation and mass transport in the oceanic stratified flows [9]. In many cases, large-amplitude waves are generated due to the interaction of internal tides with irregular bottom topography near underwater ridges [16]. Our contribution has been inspired by recent field observations [24] related to extremely long series of internal waves and Kelvin–Helmholtz billows in the equatorial Romanche Fracture Zone of the Mid-Atlantic Ridge. We consider a theoretical model intended to describe internal solitary waves in a two-layer fluid flow with the density depending exponentially on the thickness of both layers. This model generalizes the non-linear models early suggested by [2, 15, 19] for a system with constant densities in both layers, as well as the latest "2.5-layer" models considered by [11–13, 25]. The method of derivation involves the analysis of the non-linear Dubreil-Jacotin—Long equation that results from the stationary Euler equations of stratified fluid. Long-wave scaling procedure uses a small Boussinesq parameter which characterizes gentle slope of the density profile in the layers and small density jump at their interface. This asymptotic procedure combines the approaches, applied formerly to pure two-fluid system, with perturbation technique

N. Makarenko (✉) · J. Maltseva
Lavrentyev Institute of Hydrodynamics, Novosibirsk State University, Novosibirsk 630090, Russia
e-mail: makarenko@hydro.nsc.ru

J. Maltseva
e-mail: maltseva@hydro.nsc.ru

R. Tarakanov
Shirshov Institute of Oceanology, Moscow 117997, Russia
e-mail: rtarakanov@gmail.com

K. Ivanova
Aix-Marseille Universite, Marseille 06, France
e-mail: ivanova.kseniya15@gmail.com

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_7

developed in [1, 10] for continuous stratification. The parametric range of solitary wave is considered in the framework of the constructed mathematical model. It is demonstrated that these wave regimes can exist close to the parametric domain of the Kelvin–Helmholtz instability. Such a marginal stability of long internal waves could explain the formation mechanism of very long billow trains, which intensify mixing of the abyssal waters.

Basic Equations

We consider 2D motion of inviscid two-layered fluid, which is weakly stratified due to gravity in each layer. It is assumed that the flow is confined between the flat bottom $y = -h_1$ and the rigid lid $y = h_2$ (see Fig. 1).

The layers are separated by the interface $y = \eta(x, t)$ with the equilibrium level $y = 0$. The fully nonlinear Euler equations describing the flow are

$$\begin{aligned} &\rho(u_t + uu_x + vu_y) + p_x = 0, \\ &\rho(v_t + uv_x + vv_y) + p_y = -\rho g, \\ &\rho_t + u\rho_x + v\rho_y = 0, \quad u_x + v_y = 0, \end{aligned} \tag{1}$$

where ρ is the fluid density, u and v are the velocity components, p is the pressure and g is the gravity acceleration. Non-disturbed parallel flow has no vertical velocity and elevation (i.e. $v = 0$, $\eta = 0$), but the horizontal velocity $u = u_0(y)$ may be piecewise constant,

$$u_0(y) = \begin{cases} u_1 \ (-h_1 < y < 0), \\ u_2 \ (0 < y < h_2). \end{cases} \tag{2}$$

In this stationary case, fluid density $\rho = \rho_0(y)$ and the pressure $p = p_0(y)$ should be coupled by the hydrostatic equation $dp_0/dy = g\rho_0$. We consider the upstream density profile depending exponentially on height,

$$\rho_0(y) = \begin{cases} \rho_1 \exp\left(-N_1^2 y/g\right) & (-h_1 < y < 0), \\ \rho_2 \exp\left(-N_2^2 y/g\right) & (0 < y < h_2), \end{cases} \tag{3}$$

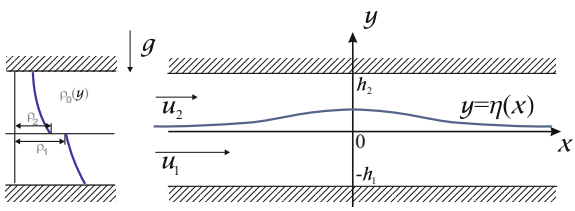


Fig. 1 Scheme of the flow

where $N_j = \text{const}$ is the Brunt—Väisälä frequency in j-th layer, and constants ρ_1 and ρ_2 are such that $\rho_2 < \rho_1$. The special case $N_j = 0$ is related to the ordinary two-fluid system with a piecewise constant density $\rho = \rho_j$ in j-th layer, but we specify $N_j \neq 0$ for the wave model to be constructed.

Further, we consider a steady non-uniform flow, hence we have $\eta_t = 0$ and $u_t = v_t = \rho_t = 0$ in Eq. (1). We introduce the stream function ψ by standard formulae $u = \psi_y$, $v = -\psi_x$, so the mass conservation implies the following dependence $\rho = \rho(\psi)$, and the pressure p can be found from the Bernoulli equation

$$\frac{1}{2}\,(\psi_x^2 + \psi_y^2) + \frac{1}{\rho(\psi)}\,p + gy = b(\psi). \tag{4}$$

We are seeking solitary-wave solutions, we require that fluid velocity (u, v) to attain the upstream velocity $(u_j, 0)$ as $x \to -\infty$. In this case, kinematic boundary conditions at the bottom, at the interface and at the lid take the form

$$\psi = -u_1 h_1 \ \ (y = -h_1), \qquad \psi = 0 \ \ (y = \eta), \qquad \psi = u_2 h_2 \ \ (y = h_2), \tag{5}$$

respectively.

It is known [26] that stationary system (1) can be reduced to the non-linear Dubreil-Jacotin—Long (DJL) equation for stream function:

$$\rho(\psi)\,(\psi_{xx} + \psi_{yy}) + \rho'(\psi)\left(gy + \frac{1}{2}\,\psi_x^2 + \frac{1}{2}\,\psi_y^2\right) = H'(\psi). \tag{6}$$

Here, function $H(\psi) = \rho(\psi) b(\psi)$ involves the Bernoulli function $b(\psi)$ and the density function $\rho(\psi)$, so that H is specified by the upstream condition. More exactly, the density function is determined by relation $\rho(\psi) = \rho_0(\psi/u_j)$ in j-th layer, and the Bernoulli function is defined by the formula

$$b(\psi) = \begin{cases} \dfrac{1}{2}\,u_1^2 + g\dfrac{\psi}{u_1} + \dfrac{g^2}{N_1^2}\left(1 - e^{\frac{N_1^2\psi}{gu_1}}\right), & -h_1 < y < \eta(x), \\ \dfrac{1}{2}\,u_2^2 + g\dfrac{\psi}{u_2} + \dfrac{g^2}{N_2^2}\left(1 - e^{\frac{N_2^2\psi}{gu_2}}\right), & \eta(x) < y < h_2. \end{cases}$$

As a consequence, we can rewrite the DJL equation (6) as follows:

$$\psi_{xx} + \psi_{yy} = \frac{N_j^2}{gu_j}\left\{ g\left(y - \frac{\psi}{u_j}\right) + \frac{1}{2}\left(\psi_x^2 + \psi_y^2 - u_j^2\right)\right\}, \tag{7}$$

where $j = 1$ should be taken in the lower layer, and $j = 2$ in the upper layer. Similarly non-linear terms also appear in the boundary condition

$$[\rho(\psi)(\psi_x^2 + \psi_y^2 + 2gy - 2b(\psi)] = 0, \qquad y = \eta(x), \tag{8}$$

where square brackets denote the discontinuity jump at the interface. According to Eq. (4), condition (8) provides the continuity of pressure p everywhere in the flow domain. Hence, we take into account all the nonlinearities from the exact Euler equations (1). Finally, we reformulate boundary condition (8) on the basis of conservation of the total horizontal momentum in a steady two-layer flow,

$$\int_{-h_1}^{h_2} (p + \rho u^2)\, dy = \text{const}.$$

Excluding the pressure p from this relation under the Bernoulli equation (4) leads to the integral relation

$$\rho_1 \int_{-h_1}^{\eta(x)} e^{-\frac{N_1^2\psi}{gu_1}} \left[(\psi_y^2 - \psi_x^2 + u_1^2 + 2g\left(\frac{\psi}{u_1} - y\right) - \frac{2g^2}{N_1^2}\left(e^{\frac{N_1^2\psi}{gu_1}} - 1\right)\right] dy + \tag{9}$$

$$+ \rho_2 \int_{\eta(x)}^{h_2} e^{-\frac{N_2^2\psi}{gu_2}} \left[(\psi_y^2 - \psi_x^2 + u_2^2 + 2g\left(\frac{\psi}{u_2} - y\right) - \frac{2g^2}{N_2^2}\left(e^{\frac{N_1^2\psi}{gu_2}} - 1\right)\right] dy = C,$$

where, the constant C depends on parameters of upstream flow as follows:

$$C = 2\rho_1 g \left[\left(e^{\frac{N_1^2 h_1}{g}} - 1\right)\left(\frac{u_1^2}{N_1^2} + \frac{g^2}{N_1^4}\right) - \frac{g h_1}{N_1^2}\right] +$$

$$+ 2\rho_2 g \left[\left(1 - e^{-\frac{N_2^2 h_2}{g}}\right)\left(\frac{u_2^2}{N_2^2} + \frac{g^2}{N_2^4}\right) - \frac{g h_2}{N_2^2}\right].$$

Indeed, it is not obvious that combersome integral relation (9) is equivalent to boundary condition (8), which is rather simple. However, this equivalence can be checked immediately by differentiating the relation (9) with respect to variable x, so the integrals become to be evaluated explicitly due to Eq. (7). We note in this context that Eq. (9) is used instead of (8) and produces more effectively the model differential equation for the function $\eta(x)$ describing strongly nonlinear waves.

Non-dimensional Formulation

Now we introduce scaled independent variables x, y and scaled unknown functions η, ψ in order to reformulate the Eqs. (5), (7) and (9) in a dimensionless form.

Namely, the ratio h_1/π is used as an appropriate length scale for x, y, η, and normalized volume discharges $u_j h_j/\pi$ serve as the units for the stream function ψ considered separately in lower- ($j = 1$) or upper layer ($j = 2$). The value π is introduced here only due to the specific form of trigonometric modal functions, which are typical for the exponential density distribution (3). Scaling procedure with this density profile uses the Boussinesq parameters σ_1, σ_2 and μ defined by the formulae

$$\sigma_j = \frac{N_j^2 h_j}{\pi g} \quad (j = 1, 2), \qquad \mu = \frac{\rho_1 - \rho_2}{\rho_2}. \tag{10}$$

Here, constants σ_j characterize the slope of density profile in continuously stratified layers, and the parameter μ determines the density jump at the interface.

Following [22], we introduce densimetric (or internal) Froude number

$$F_j = \frac{u_j}{\sqrt{g_j h_j}} \qquad (j = 1, 2)$$

which presents scaled fluid velocity u_j in j-th layer, defined with reduced gravity acceleration $g_j = (\rho_1 - \rho_2)g/\rho_j$. In addition to the Froude numbers F_j, it is also convenient to use the pair of the Long's numbers λ_j given by the following formula:

$$\lambda_j = \frac{N_j h_j}{\pi u_j} \quad (j = 1, 2).$$

The Long's numbers λ_j are coupled with the Boussinesq parameters σ_1, σ_2, μ and the Froude numbers F_j by the following relations

$$\lambda_1^2 = \frac{\pi\sigma_1(1+\mu)}{\mu F_1^2}, \quad \lambda_2^2 = \frac{\pi\sigma_2}{\mu F_2^2}.$$

Finally, we introduce the ratio of undisturbed thicknesses of the layers $r = h_1/h_2$. By that notation, the bottom is located at $y = -\pi$, and relation $y = \pi/r$ defines the rigid lid. Thus, we obtain the equations for scaled stream function ψ and wave elevation η as follows:

$$\psi_{xx} + \psi_{yy} + \lambda_1^2(\psi - y) = \frac{1}{2}\sigma_1\left(\psi_x^2 + \psi_y^2 - 1\right) \qquad (-\pi < y < \eta(x)) \tag{11}$$

$$\psi_{xx} + \psi_{yy} + \lambda_2^2 r^2(\psi - ry) = \frac{1}{2}\sigma_2\left(\psi_x^2 + \psi_y^2 - r^2\right) \qquad (\eta(x) < y < \pi/r). \tag{12}$$

The kinematic boundary conditions (5) can be rewritten now as follows:

$$\psi(x, -\pi) = -\pi, \qquad \psi(x, \eta(x)) = 0, \qquad \psi(x, \pi/r) = \pi, \tag{13}$$

and the dimensionless version of the integral relation (9) takes the form

$$\int_{-\pi}^{\eta} e^{-\sigma_1\psi}\left\{\frac{\mu F_1^2}{2}\left(\psi_y^2-\psi_x^2+1\right)+\frac{1+\mu}{\pi}\left(\psi-y-\frac{e^{\sigma_1\psi}-1}{\sigma_1}\right)\right\}dy+ \qquad (14)$$

$$+\int_{\eta}^{\pi/r} e^{-\sigma_2\psi}\left\{\frac{\mu F_2^2}{2r^3}\left(\psi_y^2-\psi_x^2+r^2\right)+\frac{1}{\pi r}\left(\psi-ry-\frac{e^{\sigma_2\psi}-1}{\sigma_2}\right)\right\}dy=C$$

with constant

$$C=\pi\mu\left(F_1^2+\frac{F_2^2}{r^2}\right)+(1+\mu)\,\frac{e^{\sigma_1\pi}-1-\sigma_1\pi}{\pi(\lambda_1^2+\sigma_1^2)}+\frac{1-\sigma_2\pi-e^{-\sigma_2\pi}}{\pi r^2(\lambda_2^2+\sigma_2^2)}.$$

Constant C is chosen here so that the upstream horizontal flow is described by the solution $\eta=0$ and $\psi=y$ $(-\pi<y<0)$, $\psi=ry$ $(0<y<\pi/r)$.

The Non-linear Long-Wave Model

The Boussinesq parameters σ_1, σ_2 and μ are small in the case of extremely weak stratification in the abyssal water. We assume here that these parameter are of the same order, so we can use a single small parameter σ by setting

$$\sigma=\sigma_1=\sigma_2=\mu. \qquad (15)$$

In accordance with this hypothesis, the derivation procedure of non-linear long-wave model should involve slow horizontal variable $\xi=\sqrt{\sigma}\,x$, as it was demonstrated by [1] in the case of weak linear stratification. We expand the stream function into power series with respect to σ as

$$\psi=\psi^{(0)}(\xi,y)+\sigma\,\psi^{(1)}(\xi,y)+\cdots \qquad (16)$$

where the leading-order term $\psi^{(0)}$ defines the hydrostatic mode, and coefficient $\psi^{(1)}$ provides the correction due to non-linear dispersion. All these coefficients $\psi^{(k)}$ can be uniquely determined from equations (11) and (12) (with fixed Long's numbers λ_1 and λ_2) under kinematic boundary condition (13). Thus, we obtain

$$\psi^{(0)}=y-\eta\,\frac{\sin\alpha_1(y)}{\sin\alpha_1(\eta)} \qquad (-\pi<y<\eta),$$

and

$$\psi^{(0)} = ry - r\eta\,\frac{\sin\alpha_2(y)}{\sin\alpha_2(\eta)} \qquad (\eta < y < \pi/r),$$

where is denoted

$$\alpha_1(y) = \lambda_1(\pi + y), \qquad \alpha_2(y) = \lambda_2(\pi - ry).$$

The dispersive term $\psi^{(1)}$ is much more complicated, it has the form

$$\psi^{(1)} = \frac{\eta(\eta-y)}{2}\,\frac{\sin\alpha_1(y)}{\sin\alpha_1(\eta)} +$$

$$+\frac{\sin\alpha_1(y)}{2\lambda_1}\left(\frac{\eta}{\sin\alpha_1(\eta)}\right)_{\xi\xi}\left\{(\pi+\eta)\cot\alpha_1(\eta) - (\pi+y)\cot\alpha_1(y)\right\} +$$

$$+\frac{\eta^2}{6}\left\{\frac{\sin\lambda_1(y-\eta)-\sin\alpha_1(y)}{\sin^3\alpha_1(\eta)} + \frac{1+\sin^2\alpha_1(y)}{\sin^2\alpha_1(\eta)} - \frac{\sin\alpha_1(y)}{\sin\alpha_1(\eta)}\right\}$$

in lower layer, and

$$\psi^{(1)} = \frac{r^2\eta(\eta-y)}{2}\,\frac{\sin\alpha_2(y)}{\sin\alpha_2(\eta)} +$$

$$+\frac{\sin\alpha_2(y)}{2\lambda_2}\left(\frac{\eta}{\sin\alpha_2(\eta)}\right)_{\xi\xi}\left\{(y-\pi/r)\cot\alpha_2(y) - (\eta-\pi/r)\cot\alpha_2(\eta)\right\} +$$

$$+\frac{r^2\eta^2}{6}\left\{\frac{\sin\lambda_2 r(\eta-y)-\sin\alpha_2(y)}{\sin^3\alpha_2(\eta)} + \frac{1+\sin^2\alpha_2(y)}{\sin^2\alpha_2(\eta)} - \frac{\sin\alpha_2(y)}{\sin\alpha_2(\eta)}\right\}$$

in the upper layer. Now we substitute power expansion (16) of function ψ into integral relation (14) and truncate the terms with powers higher than the first degree of σ. Hence, system (11)–(14) reduces to the first-order ordinary differential equation for the wave elevation $\eta(x)$ having the following form

$$\left(\frac{d\eta}{dx}\right)^2 = \eta^2\,\frac{D(\eta; F_1, F_2)}{Q(\eta; F_1, F_2)}. \tag{17}$$

Here, function D is given by the formula

$$D(\eta; F_1, F_2) = \sqrt{\pi}F_1\cot\alpha_1(\eta) + \sqrt{\pi}F_2\cot\alpha_2(\eta) + \frac{1}{3}(1-r)\eta - 1$$

where α_1 and α_2 should be taken as

$$\alpha_1(\eta) = \frac{\pi + \eta}{\sqrt{\pi} F_1}, \quad \alpha_2(\eta) = \frac{\pi - r\eta}{\sqrt{\pi} F_2}$$

since we have at the leading order in σ the relations $\lambda_j = 1/\sqrt{\pi} F_j$ $(j = 1, 2)$ resulted under the condition (15). Denominator Q in (17) has the following form

$$2Q(\eta; F_1, F_2) =$$

$$= \left(\pi F_1^2 - 2\sqrt{\pi} F_1 \eta \cot \alpha_1(\eta) + \eta^2 \cot^2 \alpha_1(\eta) \right) \left(\frac{\eta + \pi}{\sin^2 \alpha_1(\eta)} - \sqrt{\pi} F_1 \cot \alpha_1(\eta) \right) +$$

$$+ \left(\frac{\pi F_2^2}{r^2} - 2\frac{\sqrt{\pi} F_2}{r} \eta \cot \alpha_2(\eta) + \eta^2 \cot^2 \alpha_2(\eta) \right) \left(\frac{\pi - r\eta}{\sin^2 \alpha_2(\eta)} - \sqrt{\pi} F_2 \cot \alpha_2(\eta) \right).$$

Small-amplitude waves can be modelled by simplified weakly nonlinear version of the Eq. (17) which is written as

$$\left(\frac{d\eta}{dx} \right)^2 = \eta^2 \frac{D_0 + D_1 \eta + D_2 \eta^2}{Q(0; F_1, F_2)} \tag{18}$$

where the coefficients D_0 and D_1 are

$$D_0 = D(0; F_1, F_2) = \sqrt{\pi} F_1 \cot \frac{\sqrt{\pi}}{F_1} + \sqrt{\pi} F_2 \cot \frac{\sqrt{\pi}}{F_2} - 1,$$

$$D_1 = D'_\eta(0; F_1, F_2) = -\cot^2 \frac{\sqrt{\pi}}{F_1} + r \cot^2 \frac{\sqrt{\pi}}{F_2} + \frac{2}{3}(r - 1),$$

and the explicit form of coefficient D_2 is not important here.

Solitary Waves

Solitary-wave solutions of Eq. (17) are given in the implicit form by the following formula

$$x = \pm \int_a^\eta \sqrt{\frac{Q(s; F_1, F_2)}{D(s; F_1, F_2)}} \frac{ds}{s} \tag{19}$$

where parameter a determines non-dimensional amplitude of the wave. Parametric range of solitary waves is formed by the domain on the (F_1, F_2)-plane where radical function Q/D is ensured to be non-negative. It is easy to check that $Q(0; F_1, F_2) > 0$,

so function $Q(s; F_1, F_2)$ is positive in the vicinity of point $s = 0$. Therefore, function D plays determining role here. This function depends on F_1 and F_2 and can change the sign even at small s, where the leading-order coefficient D_0 from formula (18) dominates. As a consequence, the map of solitary-wave regimes is formed by the Froude numbers (F_1, F_2) so that inequality $D_0(F_1, F_2) > 0$ is true (see Fig. 2). Indeed, this inequality defines the range of non-linear waves, which are supercritical with respect to the phase speed of linear harmonic wave-packets.

Figure 3 compares the profiles of solitary waves calculated by formula (19) for two different pairs of Froude numbers $(F_1, F_2) = (1.43, 1.18)$—red line, $(F_1, F_2) = (1.64, 0.97)$—blue line (corresponding colored points are marked on the spectrum

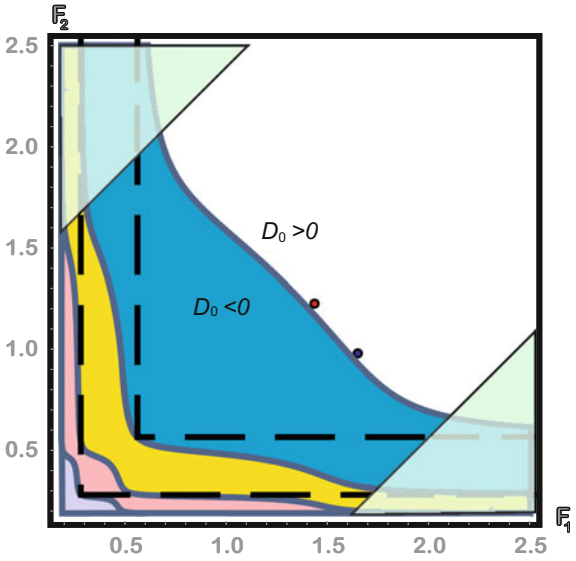


Fig. 2 Spectrum of linear waves (colored modes 1–4) and parametric domain of solitary waves (non-colored)

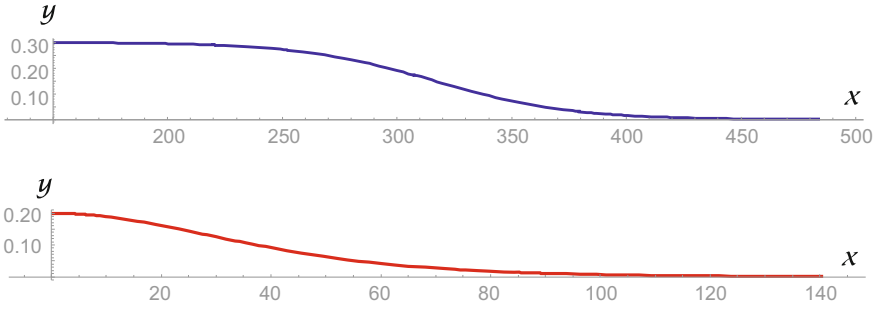


Fig. 3 Profiles of interfacial solitary waves

map (Fig. 2)). These symmetric waves have almost the same amplitude but one of them seems to be extremely broad. Such a broadening occurs when the coefficient D_1 from (18) nearly vanishes, so the higher-order nonlinearity becomes important. This is the same effect as the balance of quadratic and cubic nonlinearities in the weakly nonlinear KdV–mKdV—Gardner model [6, 7, 20]. Broadening of internal waves was investigated theoretically by [8, 14, 23], and studied experimentally by [5].

In addition to the other features, internal Froude numbers F_1 and F_2 also characterize the magnitude of velocity jump at the interface in the upstream flow. The shear $u_1 \neq u_2$ between the layers can trigger the develop of the Kelvin–Helmholtz instability which provides non-stationary formation of billow trains [3, 21]. Constant two-layer flow is linearly stable under long-wave perturbations if the inequality

$$|u_1 - u_2| < \sqrt{\frac{g(\rho_1 - \rho_2)(\rho_1 h_2 + \rho_2 h_1)}{\rho_1 \rho_2}}$$

is true, and this flow is unstable in the opposite case. Exactly the same bound for a *variable* difference $|u_1 - u_2|$ and *variable* layer thicknesses h_1, h_2 follows from the *non-linear* stability criteria predicted by the shallow water theory [4, 19]. As a consequence, we have the stability domain

$$|\sqrt{r}F_1 - F_2| < \sqrt{1 + r}$$

shown for $r = 1$ on the Fig. 2 as an inclined strip confined between shadowed triangles in the quarter-plane (F_1, F_2). Figures 4 and 5 demonstrate fragments of quasi-steady shear flow recorded on a 350 m mooring station located at a depth of 4720 m at the entrance to the Romanche Fracture Zone [24]. Trains of internal waves modulated by tide propagate here along with the sharp temperature gradient near isotherm 0.85 °C, which separates the lower layer of cold Antarctic Bottom Water (AABW) from the overlying warmer water. Owing to this fact, the moored CTD/LADCP data indicate permanently marginal stability of the flow with the Richardson number $0.25 < Ri < 1$. Tidal amplification of the shear triggers the formation of small-scale overturns which create long trains of the Kelvin–Helmholtz billows. Bold curves in Figs. 4 and 5 show overlapped profiles of solitary waves calculated by the solution (19). The solitary wave shown on Fig. 4 is relatively short, and the flow is apparently non-symmetric due to intense breaking, which localizes the sharp wave-crest downstream. In contrast, Fig. 5 demonstrates a long series of weaker overturns, which are distributed uniformly along with gently sloped wave top. It is interesting that similar overturning near the middle part of the broad solitary wave was observed in laboratory experiments [5].

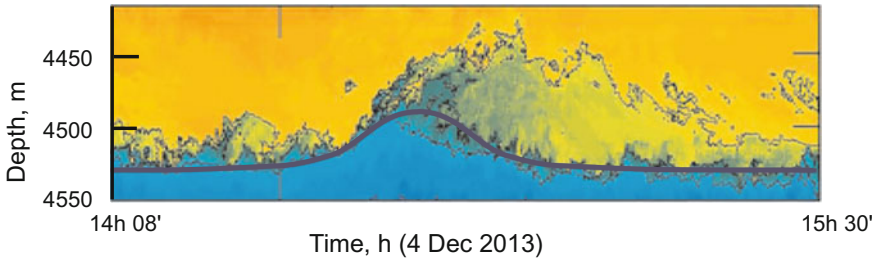


Fig. 4 Interfacial solitary wave affected by the Kelvin – Helmholtz instability

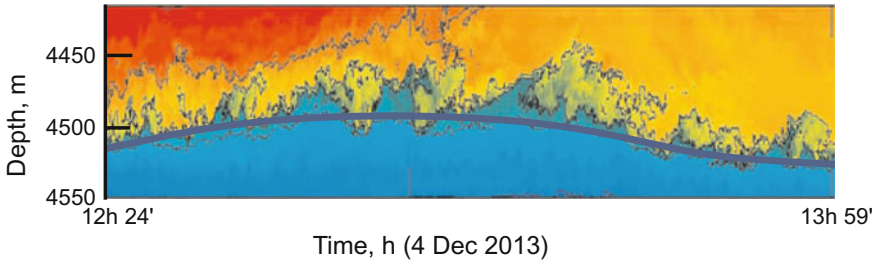


Fig. 5 Broad solitary wave with K–H billows

Conclusion

In this contribution we have considered the problem on stable internal waves at the interface between the exponentially stratified fluid layers. An ordinary differential equation describing large amplitude solitary waves has been obtained using the long-wave scaling procedure. Parametric range of the solitary waves is characterized including extreme regimes such as broad plateau-shape solitary waves. It is demonstrated that these solitary wave regimes can be affected by the Kelvin–Helmholtz instability generated due to the velocity shear at the interface.

Acknowledgements This work was supported by Russian Foundation for Basic Research (grant No 15-01-03942). RYuT acknowledges the support by Russian Sciences Foundation (grant No 16–17-10149).

References

1. Benney, D. J., & Ko, D. R. S. (1978). The propagation of long large amplitude internal waves. *Studies in Applied Mathematics*, *59*, 187–199.
2. Choi, W., & Camassa, R. (1999). Fully nonlinear internal waves in a two-fluid system. *Journal of Fluid Mechanics*, *396*, 1–36.

3. Drazin, P. (2002). *Introduction to hydrodynamic stabilitry*. Cambridge, UK: Cambridge University Press.
4. Gavrilyuk, S. L., Makarenko, N. I., & Sukhinin, S. V. (2017). *Waves in continuous media*. In: Lecture Notes in Geosystem Mathematics and Computing. Cham, Switzerland: Birkhäuser/Springer.
5. Grue, J., Jensen, A., Rusås, P. O., & Sveen, J. K. (2000). Breaking and broadening of internal solitary waves. *Journal of Fluid Mechanics*, *413*, 181–217.
6. Helfrich, K. R., & Melville, W. K. (2006). Long nonlinear internal waves. *Annual Review of Fluid Mechanics*, *38*, 395–425.
7. Kakutani, T., & Yamasaki, N. (1978). Solitary waves on a two-layer fluid. *Journal of the Physical Society of Japan*, *45*, 674–679.
8. Lamb, K., & Wan, B. (1998). Conjugate flows and flat solitary waves for a continuously stratified fluid. *Physics of Fluids*, *10*, 2061–2079.
9. LeBlond, P. H., & Mysak, L. A. (1978). *Waves in the ocean*. Amsterdam: Elsevier.
10. Long, R. R. (1965). On the Boussinesq approximation and its role in the theory of internal waves. *Tellus*, *17*(1), 46–52.
11. Makarenko, N. I., & Maltseva, J. L. (2008). An analytical model of large amplitude internal solitary waves. In E. Pelinovsky & C. Kharif (Eds.), *Extreme Ocean Waves* (pp. 179–189). Dordrecht: Springer.
12. Makarenko, N. I., & Maltseva, J. L. (2009a). Phase velocity spectrum of internal waves in a weakly-stratified two-layer fluid. *Fluid Dynamics*, *44*(2), 278–294.
13. Makarenko, N. I., & Maltseva, J. L. (2009b). Solitary waves in a weakly stratified two-layer fluid. *Journal of Applied Mechanics and Technical Physics*, *50*(2), 229–234.
14. Makarenko, N. I., Maltseva, J. L., & Kazakov, A. Y. (2009). Conjugate flows and amplitude bounds for inernal solitary waves. *Nonlinear Processes in Geophysics*, *16*, 169–178.
15. Miyata, M. (1985). An internal solitary wave of large amplitude. *La Mer*, *23*(2), 43–48.
16. Morozov, E. G. (1995). Semidiurnal internal wave global field. *Deep Sea Research*, *42*(1), 135–148.
17. Morozov, E., Demidov, A., Tarakanov, R., & Zenk, W. (2010). *Abyssal channels in the Atlantic Ocean: Water structure and flows*. Dordrecht: Springer.
18. Morozov, E. G., Tarakanov, R Yu., Lyapidevskii, V Yu., & Makarenko, N. I. (2012). Abyssal cataracts in the Romanche and Chain fracture zones. *Doklady Earth Sciences*, *446*, 1211–14.
19. Ovsyannikov, L. V., Makarenko, N. I., Nalimov, V. I., et al. (1985). *Nonlinear problems of the theory of surface and internal waves*. Novosibirsk: Nauka (in Russian).
20. Pelinovsky, E. N., Polukhina, O., Slyunaev, A., & Talipova, T. (2007). Internal solitary waves. In R. Grimshaw (Ed.), *Solitary waves in fluids* (pp. 85–110). Southhampton: WIT Press.
21. Thorpe, S. A. (1985). Laboratory observations of secondary structures in Kelvin–Helmholtz billows and consequences for ocean mixing. *Geophysical & Astrophysical Fluid Dynamics*, *34*, 175–190.
22. Turner, J. S. (1973). *Buoyancy effects in fluid*. Cambridge UK: Cambridge University Press.
23. Turner, R. E. L., & VandenBroeck, J.-M. (1988). Broadening of interfacial solitary waves. *Physics of Fluids*, *31*, 2486–2490.
24. Van Haren, H., Gostiaux, L., Morozov, E., & Tarakanov, R. (2014). Extremely long Kelvin–Helmholtz billow trains in the Romanche Fracture zone. *Geophysical Research Letters*, *44*(23), 8445–8451.
25. Voronovich, A. G. (2003). Strong solitary internal waves in a 2.5- layer model. *Journal of Fluid Mechanics*, *474*, 85–94.
26. Yih, C. S. (1980). *Stratified Flows*. New York: Academic Press.

Surface Manifestations of Internal Waves Induced by a Subsurface Buoyant Jet (Experiment and Theory)

Valerii G. Bondur, Yuliya I. Troitskaya, Ekaterina V. Ezhova, Vasiliy I. Kazakov, Alexandr A. Kandaurov, Daniil A. Sergeev and Irina A. Soustova

Introduction

Wastewater disposal from submerged outfalls is one of the major sources of pollution in coastal waters [4, 6, 8, 9, 33]. The discharge stresses and transforms coastal ecosystems and influences the quality of life in the coastal zone, making the monitoring of such systems an important issue. One of the prominent methods for the monitoring of submerged collectors is remote sensing [3, 7, 4, 6, 8, 9, 33]. Understanding dynamics of effluents in a stratified ocean is relevant for the evaluation of the efficiency of remote methods. Currently, physical mechanisms of surface manifestations of wastewaters are not well understood. In addition to the direct outcome of disposed wastewaters to the surface, surface deformation by buoyant vortices or internal waves has been discussed in [5, 3, 7, 4, 6, 8, 9, 33], while another mechanism due to the complex interaction between turbulence, internal waves, tidal currents, and bottom topography has been proposed in [7]. The aim of this contribution is to summarize the results of recent investigations of vertical buoyant plumes in stratified fluids, and modeling wastewater effluents, focusing on the generation of internal waves and their surface manifestations.

Particular attention is given to the major series of experiments performed in the Large Thermally Stratified Tank (LTST) of the Institute of Applied Physics of the Russian Academy of Sciences (IAP RAS), Nizhny Novgorod, Russia, in 2007–2012 [10, 21, 39]. A safe regime of wastewater disposal should minimize the load on the ambient coastal waters, which means that the wastewaters should be diluted

V. G. Bondur (✉)
AEROCOSMOS Research Institute for Aerospace Monitoring, Moscow, Russia
e-mail: vgbondur@aerocosmos.info

Y. I. Troitskaya · E. V. Ezhova · V. I. Kazakov · A. A. Kandaurov · D. A. Sergeev
I. A. Soustova
Institute of Applied Physics, Russian Academy of Sciences, Nizhni Novgorod, Russia
e-mail: yuliya@appl.sci-nnov.ru

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_8

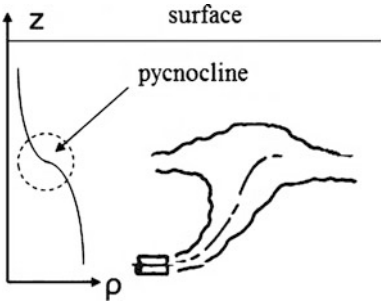


Fig. 1 Schematic of sewage disposal from submerged collectors

most efficiently and trapped in the water column. A typical diffuser of a modern outfall system contains several outlet nozzles allowing generally positively buoyant effluent to be discharged horizontally into the denser oceanic waters [34]. The effluent forms turbulent buoyant plumes rising towards the surface. More effective dilution is achieved by placing the outlet nozzles at a considerable distance from each other to avoid plume coalescence, since turbulent entrainment by separate plumes is more effective (e.g., [16]. In addition, stratification of coastal waters is characterized by the seasonal pycnocline, and outfall constructions are designed so that wastewater plumes are arrested by the pycnocline (Fig. 1). Note that this type of stratification has been reproduced in all the experiments in the LTST.

The outcome of the plume to the surface or its trapping by the ambient fluid depends on the plume parameters as well as on the peculiarities of the stratification. This problem has been extensively investigated employing classical Morton-Taylor-Turner theory [36] and laboratory experiments [25, 34], while recently more complicated "escaping" criteria have been formulated in [17, 18]. Laboratory experiments most often have been performed with two-layer fluids separated by a pycnocline, while "escaping" criteria have been formulated for arbitrary stratification.

However, almost all investigations regarding jets and plumes in a stratified fluid leave aside an interesting feature pertaining to fountains, it is a dynamical flow. A wastewater plume propagating towards the surface effectively entrains surrounding fluid and the density of the plume fluid becomes close to that of the ambient fluid. Thus, a plume in the pycnocline is negatively buoyant and almost vertical, in other words, a fountain is formed. Oscillations of the plume top near the mean penetration height were first mentioned by Turner [38]. In the experimental studies [27, 40], the oscillations of submerged fountains observed in both turbulent, [26] and laminar regimes [40] are described. A comprehensive work on the oscillations of turbulent fountains in a homogeneous fluid has been reported in [14, 28]. These authors have developed a classification of fountains based on the Froude number at the outlet, $Fr = U/\sqrt{(g'R)}$, where U is the fountain velocity at the inflow, R is the source radius, and g' is reduced gravity. It has been shown that weak fountains characterized by the Froude numbers of order one display vigorous vertical oscillations with an amplitude up to 50% of its total mean height. Similar to

the above-mentioned work [10] our experimental data reported here demonstrate that the buoyant plume top oscillates near the plume mean penetration height. In a stratified fluid, these oscillations may lead to the generation of internal waves. A similar phenomenon has been described by Karlikov and Trushina in Karlikov [31, 32], where the jet impinging on the air-water interface generates surface waves.

The mechanisms of plume oscillations are discussed in section "Generation of Internal Waves by Turbulent Buoyant Plumes". We give an overview of the laboratory experiments and numerical works devoted to this problem and summarize the results of the theoretical analysis by [10] and by Ezhova and Troitskaya [21]. Section "Structure and Energetics of Internal Waves" presents the results regarding the structure and energetics of internal waves generated by the plumes from the collector, including experimental data from LTST, results of numerical simulations, and a theoretical description. Finally, Surface Manifestations of Internal Waves focuses on the surface manifestations of internal waves (IW).

Generation of Internal Waves by Turbulent Buoyant Plumes

Oscillations of fountains in a stratified fluid, i.e. in a system similar to a plume interacting with a pycnocline are able to generate internal waves. This effect has been first demonstrated experimentally in the experiments modelling wastewater outfalls by Troitskaya et al. [39], Bondur et al. [10] (note that the Froude number of a fountain in the pycnocline is of the first order of magnitude).

The flow from the collector of the disposal system has been modelled in the LTST. The overall LTST dimensions are as follows: 20 m long, 4 m wide, and 2 m deep [2]. The density stratification in the LTST is generated by means of liquid heating and cooling by heat exchangers installed along the lateral boundaries of the tank. The results on the formation of nonuniform vertical distribution of temperature in the LTST [10] were performed in two series of experiments in the LTST with different thermocline depth. The distance from the middle of the thermocline to the surface was 40–50 cm in one series of experiments (standard stratification) and 13–15 cm in the other series (shallow thermocline). In the series of experiments with a shallow thermocline, horizontal velocity at the surface was measured by Particle Tracking Velocimetry (PTV). The experimental setup is shown in Fig. 2. A collector was modeled by a metallic pipe, blocked at one end, with 5 holes on its lateral surface. The pipe was located horizontally across the tank in the middle of it at depth H from the surface. The axes of the holes were oriented horizontally. The free end of the pipe was connected by a hose to a reservoir filled with alcohol solution.

The temperature oscillations in the thermocline were measured using an antenna of 13 temperature sensors (thermistors). The antenna was fixed on a special portable track and could be placed at different distances from the collector model. The

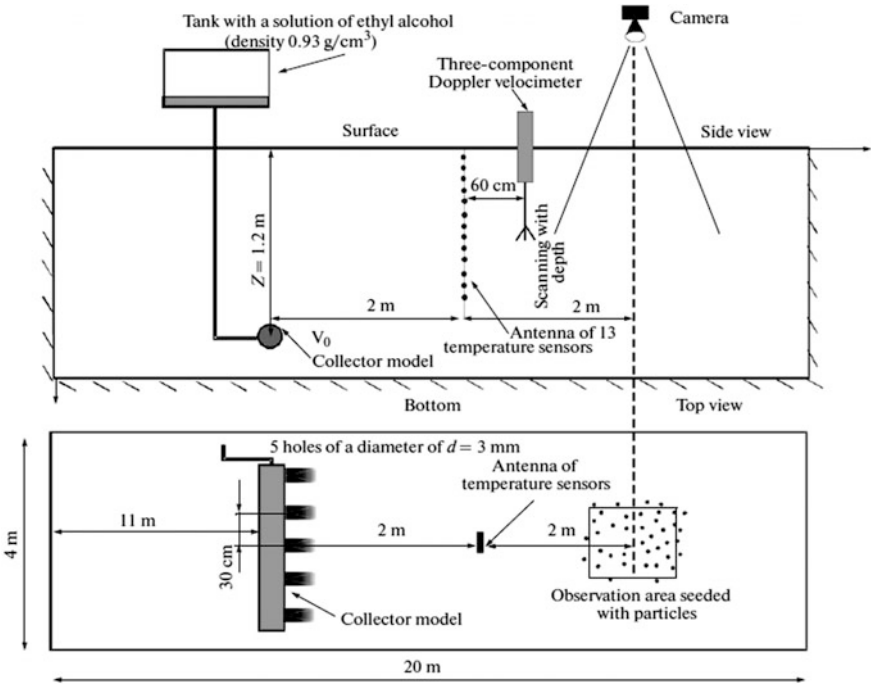


Fig. 2 Experimental setup in the LTST. [1] reservoir with alcohol solution (density 0.93 g/cm^3), [2] ultrasound three-component Doppler velocity profiler, the collector model (pipe with 5 holes of 3 mm diameter), [3] array of 13 thermistors. Depth $H = 160$ cm in the experiments with standard stratification, $H = 120$ cm in the experiments with a shallow thermocline. In the experiments with a shallow thermocline: [5] area of the measurements of surface velocities, [7] camera

profiles of the horizontal flow under the thermocline were measured by a three-component Doppler velocity profiler.

Internal waves with a pronounced spectral peak at the frequency close to the maximum buoyancy frequency were detected in a wide range of the outflow parameters (outflow velocity varied from 40 to 200 cm/s). Later, this effect was observed in the numerical simulations by Ezhova et al. [20, 21] part 1, 2. Note, that [10] measured internal waves in the experiments modelling discharge from submerged wastewater outfalls. This means that IW were generated by several identical oscillating plumes not interacting or weakly interacting with each other (the distance between the nozzle outlets was approximately twice as large as the plume diameter at the inflow of the jet to the pycnocline). Later in the works by Ezhova et al. [20, 21], part 1, 2 the authors considered a single oscillating jet with the similar parameters as a separate plume at the pycnocline, but did not reveal the difference in the frequencies of generated IW (see Fig. 3 for the illustration of IW and their spectra).

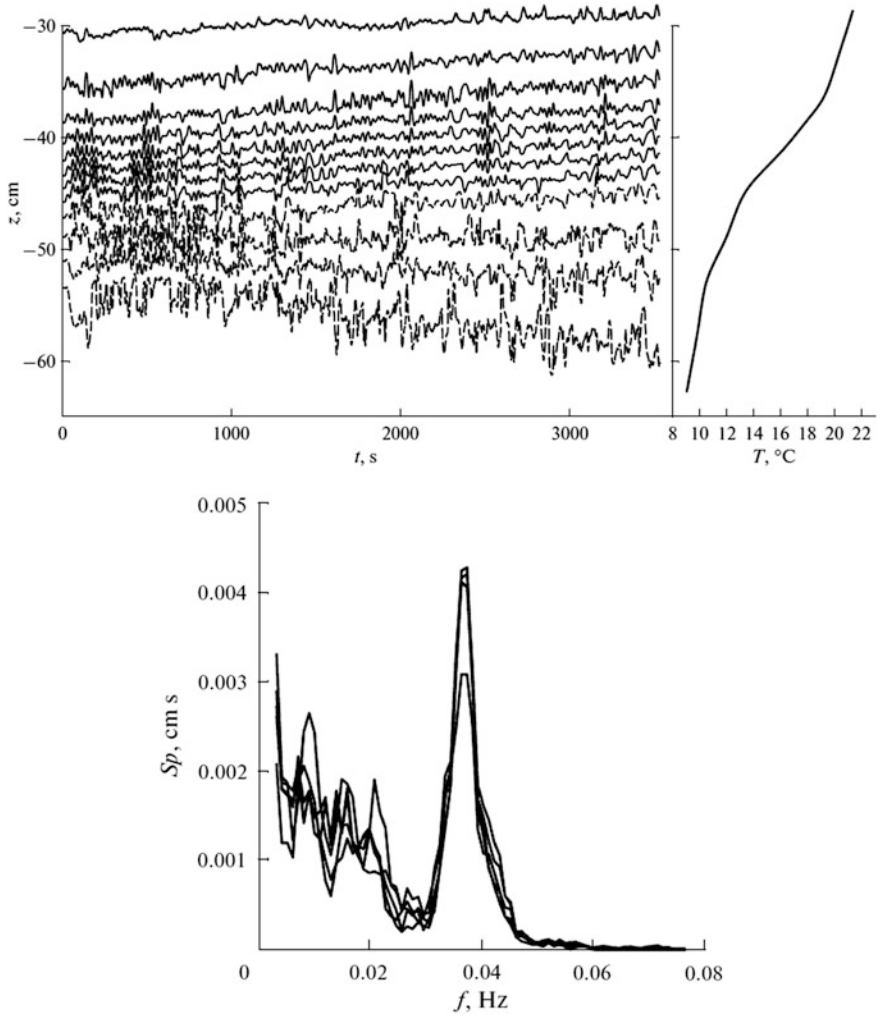


Fig. 3 Top panel: Cross section of isotherm oscillations at an outflow velocity of 150 cm/s. Bottom panel: Spectra of isotherms in the thermocline region at an outflow velocity of 150 cm/s. Different curves correspond to different depths

The mechanism of IW generation by a turbulent plume has been proposed by Troitskaya et al. [39], while a detailed theoretical analysis can be found in Bondur et al. [10] and Ezhova et al. [20, 21] part 1, 2. To explain the observed phenomenon a hypothesis of self-sustained oscillations was suggested. The amplitude of the self-sustained oscillations is known to be described by the Landau equation [20, 28]:

$$\frac{da}{dt} = a\left(\mu(R - R_c) - \nu\,|a|^2\right)$$

where R is a control parameter, R_c is its critical value, ν is a parameter of nonlinear decay. Amplitude of steady state oscillations «a» satisfies the following equation:

$$\left((R - R_c) - \frac{\nu}{\mu}|a|^2\right)a = 0.$$

Experimentally measured dependence of parameter R proves the excitation of a globally unstable mode on a jet (Fig. 4) [21].

The analysis shows that IW can be generated as a result of self-sustained oscillations of a plume in the pycnocline. Self-sustained oscillations develop in a system characterized by absolute instability. It is well-known that free shear flows are convectively unstable. It has been shown in [29] that the type of instability can change to absolute if the hydrodynamic flow is lighter than the surrounding fluid or due to the presence of a counterflow. Thus, a plume in the pycnocline can be absolutely unstable due to the counterflow.

Linear stability analysis of the fountain in the pycnocline was performed in [10, 21]. The earlier analysis was based on the experiment in a small stratified tank using Particle Image Velocimetry [10]. This experiment was intended to demonstrate the effect, while the parameters did not provide scale modeling of wastewater outfall. Assuming the flow in the pycnocline locally quasi-parallel, the authors performed a theoretical linear stability analysis assuming that the mean velocity profiles in the subsequent cross-sections present the basis flow (Fig. 5).

They have demonstrated that there is a region of absolute instability, while self-sustained oscillations can develop in the plume. Moreover, theoretically estimated frequencies of an absolutely unstable axisymmetric mode showed good correspondence with experimentally measured frequency of plume oscillations and

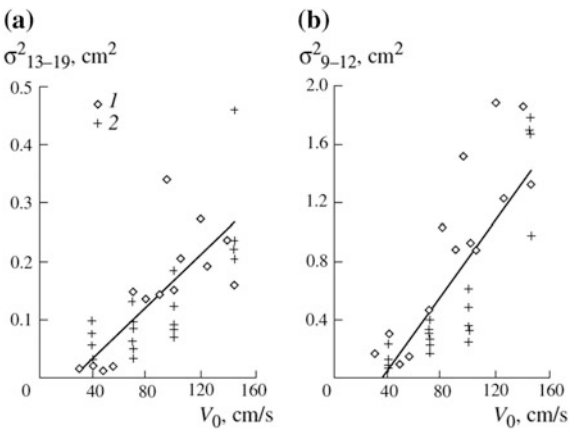


Fig. 4 Isotherm displacement variance in the thermocline versus discharge rate: **a** in the upper part of the thermocline (isotherms 13–19 °C) and **b** in the lower part of the thermocline (isotherms 9–12 °C). Notations 1 and 2 correspond to two series of experiments

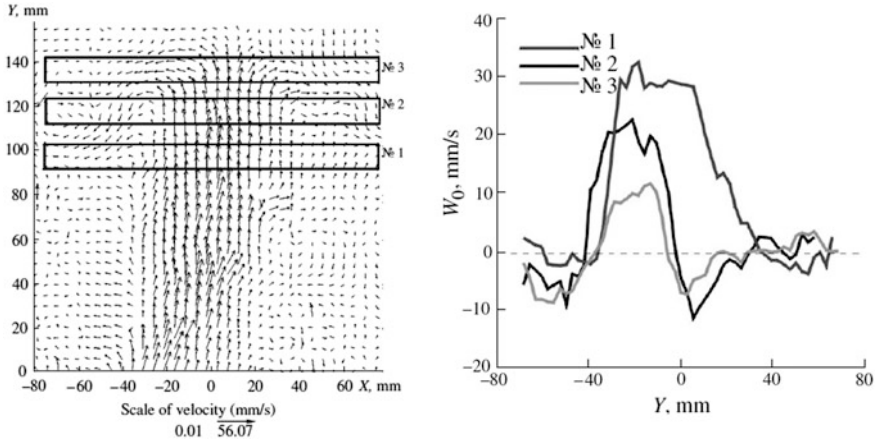


Fig. 5 Vector velocity field of turbulent buoyant plume (left) and vertical velocity profiles in cross "Introduction", "Generation of Internal Waves by Turbulent Buoyant Plumes" and "Structure and Energetics of Internal Waves" (right)

with the range of theoretically calculated frequencies (see Fig. 6 displaying a spectrum of plume oscillations and the range of frequencies obtained from the linear stability analysis).

The results of more complicated experiments in the LTST are described in [20]. The experiments were set up to model a turbulent fountain interacting with a thermocline with the parameters similar to those of a modeled wastewater plume (Fig. 7). The fountain was formed by a turbulent vertical jet providing vertical velocity profiles at the inflow of the jet to the thermocline similar to those in the experiments modeling wastewater plumes.

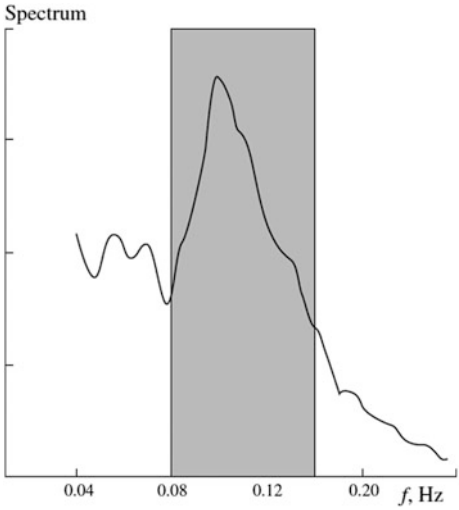


Fig. 6 Spectrum of oscillations of the upper boundary of the jet. The marked area corresponds to the calculated limits of variations of the possible frequency of self-sustained oscillations

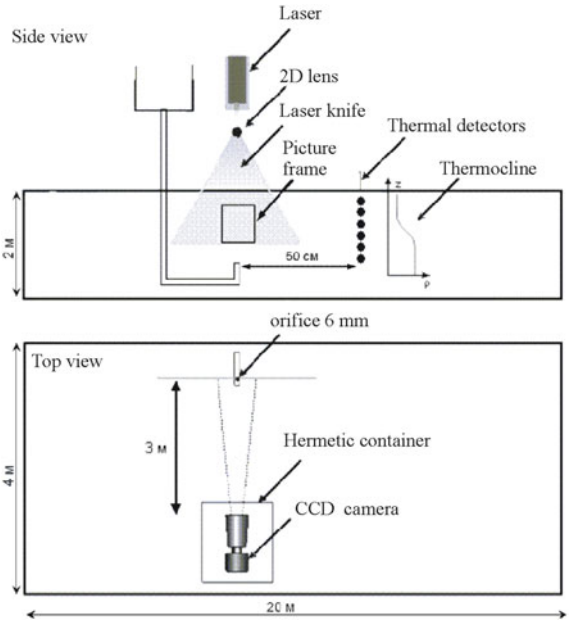


Fig. 7 Experimental setup in the LTST for investigating the mechanism of internal wave generation by axisymmetric turbulent jets

A camera was located in a waterproof box to record plume oscillations simultaneously with IW measurements. This allowed us to compare spectra of the plume oscillations and IW and confirm that IW were generated by the plume oscillations (Fig. 8). The structure of plume oscillations has been investigated in detail; it was concluded that IW are generated at a frequency corresponding to an axisymmetric

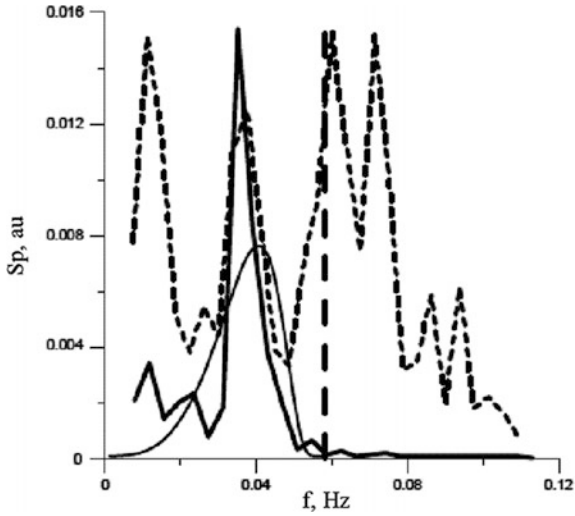


Fig. 8 Spectra of the jet oscillations (dashed line) and internal waves (solid line). The straight dashed line corresponds to the maximum buoyancy frequency; the thin solid curve is the frequency dependence of the excitation coefficient of the lowest internal wave mode

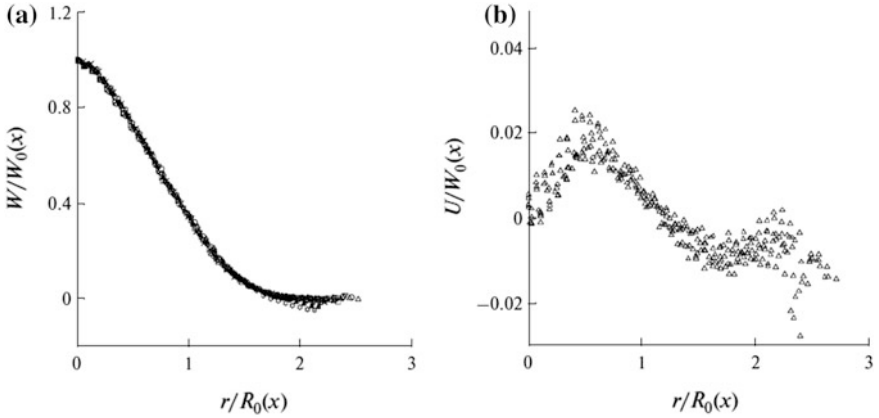


Fig. 9 Profiles of the average vertical (left) and radial (right) jet velocity in self-similar coordinates [20]

unstable mode. Furthermore, velocity in the plume was measured by the PIV and a theoretical linear stability analysis was performed.

Note that the procedure of measuring mean velocity profiles was different in the experiments by Bondur et al. [10], Ezhova et al. [20]. Linear stability analysis requires a basic flow. In the first case, the authors averaged the instantaneous profiles in the reference frame moving with the plume, while in the second case the standard averaging procedure was applied in a fixed reference frame. The average profiles in these reference frames are different. The velocity profiles in a reference frame moving with the plume resemble "top hat" profiles (see Fig. 5), while the standard procedure in a fixed reference frame leads to the Gaussian-shaped profiles (Fig. 9). In the latter case, unstable axisymmetic mode cannot develop in the quasi-parallel framework, because the necessary condition of instability is not satisfied for this mode. However, the analysis accounting that the flow is not parallel shows that the axisymmetric mode is unstable even in the case of the Gaussian-shaped profiles; addition of a counterflow causes absolute instability similar to the previous case. Theoretically predicted frequencies of an axisymmetric mode are close to those experimentally measured and to the frequencies of IW.

Additional confirmation of self-sustained regimes of oscillating fountains in two-layer stratified fluid was obtained in the numerical simulations by Druzhinin and Troitskaya [19]. The authors demonstrated that laminar fountains oscillate in a two-layer stratification, emitting internal waves in the form of concentric circles at small Froude numbers replaced by spirally propagating internal waves at larger Froude numbers. The result was explained in the framework of the mode competition (competition of modes developing on the fountain in the pycnocline).

Finally, we comment on the influence of stratification on the generation of internal waves. Ansong and Sutherland [1] did not find correlation between the frequency of plume oscillations and internal waves in the experiments with the

vertical turbulent plumes at the stratification consisting of a linearly stratified layer over a homogeneous layer. Ezhova et al. [24] performed a numerical experiment with a turbulent jet at the stratification with a pycnocline of a finite thickness. Two different pycnocline thicknesses have been considered, one is a narrow pycnocline as compared to the plume diameter in the pycnocline and another one is a wider pycnocline equal to the plume diameter. The frequencies of jet oscillations in different pycnoclines were remarkably similar, while the IW frequencies depended on the ambient stratification. However, peaks in the spectra of IW coincided with the spectra of plume oscillations (when the peaks in the spectra of jet oscillations were lower than the maximum buoyancy frequency). Another numerical investigation by Ezhova et al. (submitted) considers buoyant plumes in a stratified fluid with the complex stratification: weakly linear with a pycnocline of a finite width, modeling subglacial discharge in the Greenland fjords. Plume oscillations at complicated stratifications result in noisier spectra. The peaks can be still detected when the stratification resembles two layers separated by a pycnocline, while the measurements at stratifications more similar to that reported by Ansong and Sutherland (1990) show correspondence with their results.

Structure and Energetics of Internal Waves

The structure of internal waves generated by the vertical turbulent plumes in a two-layer stratification with a pycnocline of finite thickness were investigated in detail by Bondur et al. [11, 12]. The work describes the results of laboratory scale modeling of submerged wastewater outfalls in the LTST (see Fig. 7) and the corresponding theoretical analysis. Experimental measurements were performed using an antenna of thermistors located at a significant distance from the collector model in the thermocline (approximately a dozen of plume diameters measured at the inflow of the plume to the pycnocline) (Fig. 10).

The antenna was located vertically in the thermocline, thus measuring temperature at several fixed points in the vertical direction (see the description of the experiment in section "Introduction"). The isotherms corresponding to different fixed temperatures were retrieved from the collected data and the spectra of isotherm displacements from the equilibrium depth were assumed as the spectra of IW. As mentioned in in section, "Genaration of Internal Waves by Turbulent Buoyant Plumes" all spectra, regardless the outflow velocity, have a pronounced peak close to 0.6–0.7 N_max, where N_max is the maximal buoyancy frequency (Figs. 11).

The vertical buoyant plume penetrates through the pycnocline, then bounces back and propagates horizontally as a gravity current at the level of the neutral buoyancy. The velocity of the gravity current in the thermocline was measured in the above mentioned experiments by a three-component Doppler velocity profiler. These measurements allowed us to develop a theoretical model of internal waves based on the mean temperature profile with the account for the horizontal flow in the thermocline. The velocities of these horizontal flows are comparable with the

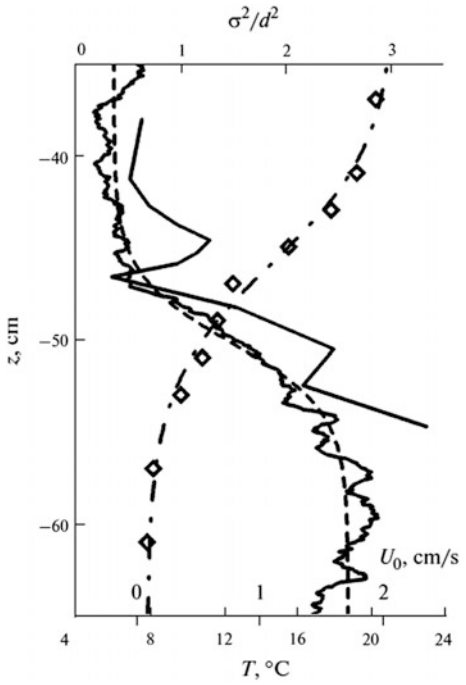


Fig. 10 Average flow characteristics: operating profile of stratification (rhombs), and its approximation (dashed and dotted line); profile of the mean jet velocity (solid line) and its approximation (dotted line): the rms profiles

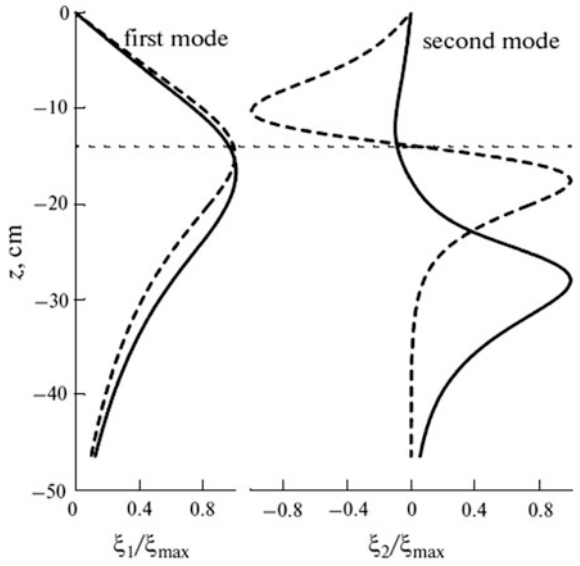


Fig. 11 Vertical profiles of the first and second modes in the presence of a horizontal flow under the thermocline (solid line) and in stationary stratified liquid (dotted line)

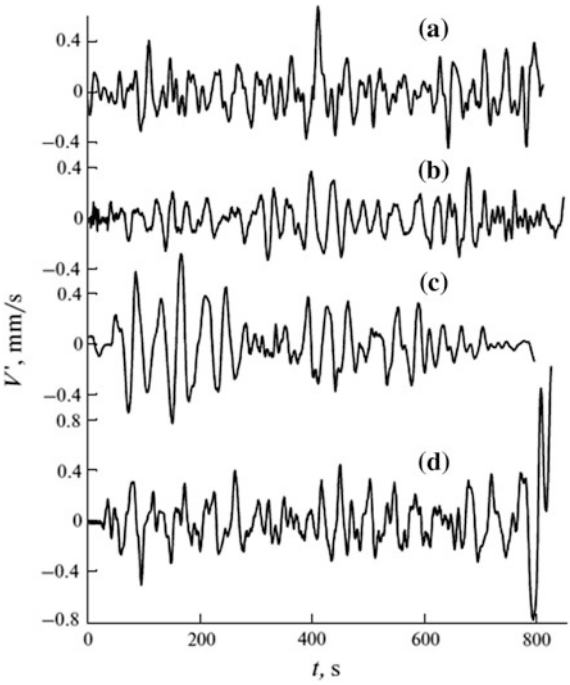


Fig. 12 Rates of disturbances at the basin surface as measured by the PTV method for the following rates of liquid outflow from the collector model: 40 cm/s **a**, 40 cm/s **b**, 100 cm/s **c**, 150 cm/s (**d**)

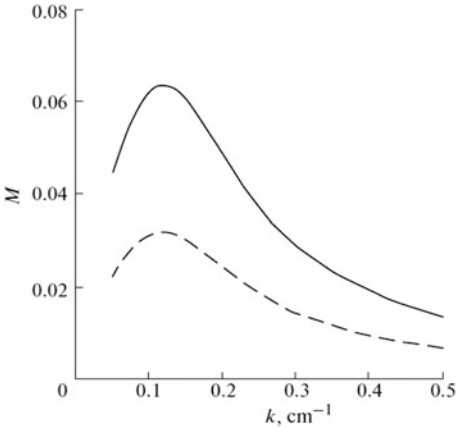


Fig. 13 Graph of the dependence of the modulation coefficient of short waves $M(k)$ on the wavenumber for the following parameters of internal waves at the surface: $\omega/q = 9.6$ cm/s, $u^* = 18$ cm/s, = 0.6 cm/s (solid line), and = 0.3 cm/s (dashed line)

phase velocities of internal waves; hence, they should be taken into account in the theoretical model. The model, described in detail in Bondur et al. [11, 12] is based on the Taylor-Goldstein equations

$$\frac{\partial^2 y}{\partial^2 z} - k^2 y + \frac{N^2}{(U_0 - c)^2} y - \frac{U_0''}{(U_0 - c)} y = 0, \tag{3}$$

where ψ is the stream function, N(z) is the profile of the buoyancy frequency, $U_0(z)$ is the velocity profile of a gravity current, k is the wavenumber, ω is the wave frequency, c is the phase velocity.

The eigenvalue problem for this equation was solved numerically for the experimentally measured profiles of velocity and temperature (see Fig. 11). Dispersion curves for the two basic modes and the modal structure (vertical profile of internal waves) at the generation frequency were calculated. The results of modeling explain the measured vertical structure of internal waves with two peaks located in the thermocline and in the horizontal flow.

In the absence of a gravity current in the thermocline, two basic modes of internal waves at the generation frequency of IW (0.6–0.7 Nmax) can be described as follows. Figure 11 demonstrates two fundamental modes of IW for the displacement of liquid particles from the equilibrium ξ, which can be defined using the stream function:

$$\begin{aligned} w &= -\tfrac{\partial^2 \Psi}{\partial x} = -ik\Psi = ik(U_0 - c)\xi \\ y &= -(U_0 - c)\xi \end{aligned}, \tag{4}$$

where w is the vertical velocity.

The first mode has one maximum located in the centre of the pycnocline, while the second mode has two maxima with opposite phases. In the presence of a gravity current in the thermocline, the peak of the first mode is displaced towards the gravity current, while the second mode has a pronounced peak in the current while the second peak located in the thermocline is significantly reduced. Thus, the combination of these two modes allows us to explain the properties of the experimental profile of isotherm displacements. However, though a satisfactory agreement has been achieved, the minimum between the peaks is usually better pronounced in the experimental profiles than it was predicted by the theoretical model in Bondur et al. [12].

Inclusion of the turbulent viscosity associated with the shear flow and a dependence on the vertical coordinate in the modal analysis, similarly to [35] might result in a better agreement with the experimental data. Furthermore, [23] have recently estimated the energetics of internal waves based on the numerical experiments modeling a vertical turbulent jet in a stratified fluid. They have calculated the energy flux of internal waves at a distance of ~5 and ~6.25 diameters from the jet (measured at the inflow to the thermocline) and compared it to the jet energy flux at the inflow to the thermocline. The estimates have been obtained for two stratifications including a relatively thin thermocline as compared to the jet diameter and a thermocline thickness of the same order as the jet diameter. The calculations

showed that approximately 5% of the jet energy is spent to generate IW in a thin thermocline (assuming the IW decay with the radial distance from the jet, this estimate can be somewhat increased), while only 2–3% of the jet energy flux is transported away from the jet by IW in the thick thermocline.

The difference in the energy fluxes can be explained as follows. The amplitudes of jet oscillations are similar in the thin and thick thermoclines. However, the temperature gradient in the thick thermocline is weaker and thus, the maximal buoyancy frequency is lower. One of the spectral peaks of plume oscillations is higher than the maximal buoyancy frequency and the waves at this frequency cannot propagate. Moreover, numerical simulations by [23] revealed a horizontal counterflow in the upper thermocline attributed to the turbulent entrainment by the jet top in the thick thermocline. This counterflow results in the transfer of the part of IW energy towards the plume, thus decreasing the total energy flux transferred away from the jet by IW.

Another interesting feature that has been revealed by [23] is that the ratio of the energy flux of IW to the jet at the inflow to the thermocline remains approximately constant in the simulations in the thin thermocline and almost does not depend on the control parameter, which in this case is the Froude number. This effect can be explained considering simplified energy conservation similar to the estimates made by [15] for fountains. Note that Burridge and Hunt considered a different type of oscillations, namely the vortices appearing on the inclined surface formed by the top of the fountain, and their propagation as interfacial waves. Here, we consider internal waves generated by the vertical oscillations of a jet top as a whole.

A thin thermocline can be treated similar to the interface between two fluids of different temperatures, so as a first approximation to the dispersion relation we take $w = \sqrt{g'k}$, where g′ is reduced gravity. Note that it can be deduced from the dispersion curves calculated in the presence of a horizontal current and without it [12] that the dispersion curves for the first mode of IW at the generation frequency are close to each other. Thus, an estimate for the group velocity is $c_{gr} \sim \sqrt{g'\lambda}$, where λ is the wavelength. The estimates of the oscillation frequency from [20] is $\omega \sim$ 0.4 U/D, where U is the jet velocity at the inflow to the thermocline and D is the jet diameter. Thus, an estimate for the wavelength is $\lambda \sim g' D^2/U^2$. The energy flux of internal waves at distance R from the jet centre can be estimated as $F_{iw} \sim c_{gr} A^2 g' R$ [37]. We obtain the following estimate for the ratio of IW energy flux to the jet energy flux in the thermocline

$$E_{iw} \sim c_{gr} A^2 g' R / U^3 D^2$$

The amplitude of IW is proportional to the jet penetration height h, while $h \sim Fr^2$ [23]. Thus, one obtains from these simple estimates that E does not depend on Fr which is consistent with the numerical results in [23].

Note, that the dependence of IW amplitudes A^2 on the outflow velocity U, presented in [10] was approximated by a linear function similarly to the dependence of the jet oscillations amplitude A^2 on the control parameter. However, the results

of numerical simulations and our estimates reported here suggest that it is proportional to Up^4, where Up is the jet velocity at the inflow to the thermocline. This might provide a better approximation of the experimental data by [10].

Surface Manifestations of Internal Waves

Surface manifestations of internal waves were investigated both experimentally and theoretically by [13]. The experimental part consists of two series of experiments in the LTST with different arrangement of the measuring equipment. The experiments were performed at the stratification with a shallow thermocline in order to make the surface currents caused by internal waves detectable. The horizontal velocities at the water surface were measured by a modified Particle Tracking Velocimetry technique (PTV), see [13] for details. The measured velocities are shown in Fig. 12.

As a first step, the time before the signal detected by the PTV at a sufficient distance from the source (a collector model) was compared to the theoretically estimated time of the signal arriving to the measurement area. The latter time was derived from the theoretical group velocity of internal waves at the generation frequency. The velocity has been obtained, as described in section "Generation of Internal Waves by Turbulent Buoyant Plumes", from the dispersion relation following from the solution of the eigenvalue problem for the Taylor-Goldstein equation (with measured profiles of buoyancy and mean horizontal velocity in the thermocline). Theoretically predicted times for different outflow velocities correspond to the experimentally measured times, thus confirming that a signal detected at the surface could have been related to internal waves.

As a second step, theoretically predicted amplitudes of the horizontal velocity at the surface were compared to the experimentally measured velocities. Theoretical estimates were as before based on the solution of the eigenvalue problem for the Taylor-Goldstein equation. The authors used the vertical profile of the stream function to calculate the displacements of the liquid particles from the equilibrium due to IW at the frequency of generation [3]. The theoretical estimates of surface velocities were calculated on the basis of these considerations:

$$u(z,t) = (c - U_0)\frac{\partial \xi}{\partial z} - \xi \frac{\partial U_0}{\partial z}. \tag{5}$$

The theoretically predicted velocities significantly exceeded the experimentally measured data. This discrepancy was explained by the presence of surfactant film in the LTST. The films were taken into account in the theoretical analysis, and an agreement between the measured and theoretical velocities was obtained for the experimentally measured elasticity coefficients.

The third step of the analysis included a theoretical estimate of the hydrodynamic contrasts, which surface currents can induce in the spectra of surface wind

waves due to IW, thus providing a mechanism for internal waves to be manifested at the surface. Hydrodynamic contrast was defined as M(k) = N1/N0, where N0 is the spectral action density of an undisturbed surface wave spectrum, and N1 corresponds to the perturbations in the spectral action density due to the horizontal surface currents induced by IW. Bondur et al. [11] solved the spectral action balance equation [30], using the estimates of IW parameters for the field conditions (since the authors performed scale modeling of a submerged collector, it allowed us to use the data from the laboratory experiments to get the estimates for the field conditions employing similarity coefficients). The wind speed taken for the estimates was 5 m/s. The estimated phase velocities of IW were rather low as compared to the group velocity of the surface waves at this wind speed, while the synchronism condition was not satisfied. According to the estimates, the hydrodynamic contrasts reached 7% (Fig. 13), which could be detected by the application of the special methods to the optical satellite images. Ermakov and Salashin [24] reported that hydrodynamic contrasts can be increased 7–8 times for the bound waves (harmonics of the free waves); thus, the estimates of the hydrodynamic contrasts for such waves can increase up to ~60%.

Taking into account the presence of the free waves, the probability of finding a bound wave is 50% at the wind speed 5 m/s, thus, the final estimate of hydrodynamic contrasts for bound waves is 25%. These contrasts can be detected by remote sensing.

Conclusions

In this contribution, we summarized the main results of the work on internal waves generated by vertical turbulent plumes in stratified fluids, including the mechanisms of IW generation, the structure of IW, and their surface manifestations.

The majority of results were obtained on the basis of modeling of submerged wastewater outfalls in the ocean, and thus, they are applicable for the monitoring of the coastal zone. The other potential implications include buoyant plumes generated by subglacial discharge in the Greenland fjords [22]. Fresh water from the surface of the marine-terminated glaciers can percolate through the channels inside the glacier and come out at the base of the glacier forming a buoyant plume. Those turbulent plumes propagating along the ice face can significantly increase the rates of glacier melting, and account for their presence could improve the predictions of the melt rates. However, the locations of the plumes are difficult to identify. The IW in this case could serve as an indication of the submerged plume. The proper analysis, however, should include the sediment load in such a flow, and more important, the presence of an ice wall.

The challenges related to this topic can be formulated as follows:

More complicated stratifications have to be considered in order to understand the effect of stratification on the internal waves generated by the vertical oscillations of

a turbulent plume/jet in the pycnocline. Transition from two-layers separated by the pycnocline to the linear stratification of the upper layer and a homogeneous lower layer is of special interest, since in the linear stratification the generation of IW by buoyant plumes has not been identified [1].

The origin of the vortex structures forming at the surface of the plume top and their role in the entrainment processes has to be clarified. These vortex structures have been mentioned by Burridge and Hunt [15] and can be seen in the instantaneous temperature visualizations of the jet top in the thermocline in [23]. These structures can serve as an additional source of internal waves in the thin pycnoclines providing sufficiently high buoyancy frequencies.

The further investigations of the mechanisms of plume oscillations are needed, since linear stability analysis is not fully justified for a turbulent flow. Application of the modern numerical techniques to the vertical turbulent jets and plumes in a stratified fluid could provide a deeper insight in the structure of the flows and confirm/discard the present conclusions pertaining self-sustained oscillations.

Acknowlgedments This study was supported by the Russian Foundation for Basic Research (projects 18-05-00292, 16-55-52022 MHT-a, 17-05-41117 RGS) and the Ministry of Education and Science of the Russian Federation (project RFMEFI57716X0234).

Numerical code development and numerical modeling were supported by the Russian Science Foundation (№ 15-17-20009). The basic salary of the IAP RAS authors was obtained from FASO (project № 0035-2014-0032).

References

1. Ansong, J. K., & Sutherland, B. R. (2010). Internal gravity waves generated by convective plumes. *Journal of Fluid Mechanics, 648*, 405–434.
2. Arabadzhi, V. V, Bogatyrev, S. D., Bakhanov, V. V. et al. (1999). Laboratory modeling of hydrophysical processes in the upper ocean layer (the large thermostratified tank, Institute of Applied Physics, Russian Academy of Sciences). In V. I. Talanov, & E. N. Pelinovskii (Eds.), Near-surface ocean layer. physical processes of remote probing, (Vol. 2, pp. 231–251) (in Russian).
3. Bondur, V. G. (2005). Complex satellite monitoring of coastal water areas. In *31st International Symposium on Remote Sensing of Environment, ISRSE*.
4. Bondur V. G. (2011). Satellite monitoring and mathematical modelling of deep runoff turbulent jets in coastal water areas. *Waste Water-Evaluation and Management* (pp. 155–180). InTech: Croatia. ISBN 978-953-307-233-3.
5. Bondur, V. G., & Grebenyuk, Y. V. (2001). Remote indication of anthropogenic impacts on the marine environment caused by deep-water sewage discharge: Modeling and Experiment. *Issled. Zemli Kosmosa 6*,, 1–19.
6. Bondur V, Tsidilina M. (2005). Features of formation of remote sensing and Sea truth databases for the monitoring of Anthropogenic impact on ecosystems of coastal water areas. In *31st International Symposium on Remote Sensing of Environment, ISRSE* (pp. 192–195).
7. Bondur, V. G., Keeler, R., Gibson, C (2005). Optical satellite imagery detection of internal wave effects from a submerged turbulent outfall in the stratified ocean. *Geophysical Research Letters 32*, L12610. https://doi.org/10.1029/2005GL022390.

8. Bondur, V. G., Zhurbas, V. M., Grebenyuk, Yu. V. (2006). Mathematical modeling of turbulent jets of deep-water sewage discharge into coastal basins. *Oceanology, 46*(6), 757–771.
9. Bondur, V. G. Grebenjuk, Yu. V. Sabinin, K. D. (2008). Variability of internal tides in the coastal water area of Oahu Island (Hawaii). *Oceanology, 48*(5), 611–621.
10. Bondur, V. G., Grebenyuk, Y. V., Ezhova, E. V., Kazakov, V. I., Sergeev, D. A., Soustova, I. A., et al. (2009). Surface manifestations of internal waves investigated by a subsurface buoyant jet: 1. The mechanism of internal waves generation. Izvestiya, Atmosph. *Oceanic Physics, 45*(6), 779–790.
11. Bondur, V. G., Grebenyuk, Y. V., Ezhova, E. V., Kazakov, V. I., Sergeev, D. A., Soustova, I. A., et al. (2010). Surface manifestations of internal waves investigated by a subsurface buoyant jet: 2. Internal waves field. *Izvestiya, Atmosph. Oceanic Physics, 46,* 347–359.
12. Bondur, V. G., Grebenyuk, Y. V., Ezhova, E. V., Kazakov, V. I., Sergeev, D. A., Soustova, I. A., et al. (2010). Surface manifestations of internal waves investigated by a subsurface buoyant jet: 3. Surface manifestations of internal waves. *Izvestiya, Atmosph. Oceanic Physics., 6*(4), 482–491.
13. Bondur, V., Grebenyuk, Y., Ezhova, E., Kandaurov, A., Sergeev, D., & Troitskaya, Y. (2012). Applying of PIV/PTV methods for physical modeling of the turbulent buoyant jets in stratified fluid. In book The Particle Image Velocimetry—Characteristics, Limits and Possible Applications. Ed. By G. Cavazzini. In Tech: ISBN 978-953-51-0625-8. Chapter 13: 345–366.
14. Burridge, H. C., & Hunt, G. R. (2013). The rhythm of fountains: the length and time scales of rise height fluctuations at low and high Froude numbers. *Journal of Fluid Mechanics, 728,* 91–119.
15. Burridge, H. C., & Hunt, G. R. (2016). Entrainment by turbulent fountains. *Journal of Fluid Mechanics, 790,* 407–419.
16. Cenedese, C., & Linden, P. F. (2014). Entrainment in two coalescing axisymmetric turbulent plumes. Journal of Fluid Mechanics, *752*(R2). https://doi.org/10.1017/jfm.2014.389.
17. Caulfield, C. P., & Woods, A. W. (1998). Turbulent gravitational convection from a point source in a non-uniformly stratied environment. *Journal of Fluid Mechanics, 360,* 229–248.
18. Camassa, R., Lin, Z., McLaughlin, R. M., Merten, K., Tzou, C., & Walsh, J. (2016). White optimal mixing of buoyant jets and plumes in stratified fluids: theory and experiments. *Journal of Fluid Mechanics, 790,* 71–103.
19. Druzhinin, O. A., & Troitskaya, Y. I. (2010). Regular and chaotic dynamics of a fountain in a stratified fluid. *Chaos., 22,* 023116. https://doi.org/10.1063/1.4704814.
20. Ezhova, E. V., Sergeev, D. A., Kandaurov, A. A., & Troitskaya, Y. I. (2012). Nonsteady dynamics of turbulent axisymmetric jets in stratified fluid: Part 1. Experimental study. *Izvestiya, Atmosph. Oceanic Physics., 48*(4), 409–417.
21. Ezhova, E. V., & Troitskaya, Y. I. (2012). Nonsteady dynamics of turbulent axisymmetric jets in stratified fluid: Part 2. Mechanism of excitation of axisymmetric oscillations in a submerged jet. *Izvestiya, Atmosph. Oceanic Physics, 48*(5), 528–537.
22. Ezhova, E., Cenedese, C., & Brandt, L. Dynamics of a turbulent buoyant plume in a stratified fluid: an idealized model of subglacial discharge in Greenland fjords// submitted to JPO.
23. Ezhova, E., Cenedese, C., Brandt, L. (2016). Interaction between a vertical turbulent jet and thermocline. Journal of Physical Oceanography, *46*, 3415–3437.
24. Ermakov, S. A., & Salashin, S. G. (1994). Effect of strong modulation of gravity–capillary waves by internal waves. *Doklady Earth Sciences, 337*(1), 108–111.
25. Fischer, H. B., List, E., Koh, R., Imberger, J., & Brooks, N. (1979). *Mixing in inland and coastal waters*. California, USA: Academic Press.
26. Friedman, P. D. (2005). Oscillation height of a negatively buoyant jet. *Journal of Fluids Engineering, 128*(4), 880–882.
27. Friedman, P., Vadokoot, V. D., Meyer, W. J., & Carey, S. (2007). Instability threshold of a negatively buoyant fountain. *Experiments in Fluids, 42,* 751–759.
28. Hunt, G. R., & Burridge, H. C. (2015). Fountains in industry and nature. *Annual Review of Fluid Mechanics, 47,* 195–220.

29. Huerre, P., & Monkewitz, P. A. (1990). Local and global instabilities in spatially developing flows. *Annual Review of Fluid Mechanics, 22,* 473–537.
30. Hasselmann, K. (1962). On the nonlinear energy transfer in a gravity wave spectrum Part 1 General theory. *Journal of Fluid Mechanics, 12,* 481–500.
31. Karlikov, V. P., & Trushina, O. V. (1998). Self-oscillation of flat deep-water fountains. *Doklady Earth Sciences, 361*(3), 340–344.
32. Karlikov, V. P., Trushina, O. V. (2006). Self-oscillatory regimes of flowing of flat deep-water jets, In *Proceedings of IX All-Russia Congress on Theoretical and Applied Mechanics, Nizhnii Novgorod*.
33. Keeler R., Bondur V., Vithanage D. (April 2004). Sea truth measurements for remote sensing of littoral water. *Sea Technology,* 53–58.
34. Koh, C. Y., & Brooks, H. N. (1975). Fluid mechanics of waste-water disposal in the ocean. *Annual Review of Fluid Mechanics, 8,* 187–211.
35. Li, L., Smith, W. D., & Thorpe, S. A. (2015). Destabilization of a stratified shear layer by ambient turbulence. *Journal of Fluid Mechanics, 771,* 1–15.
36. Morton, B. R., Taylor, G. I., & Turner, J. S. (1956). Turbulent gravitational convection from maintained and instantaneous sources. *Proceedings of the Royal Society of London, Series A: Mathematical and Physical Sciences, 234,* 1–23.
37. Sutherland, B. R. (2010). *Internal gravity waves*. Cambridge University Press.
38. Turner, J. S. (1966). Jets and plumes with negative or reversing buoyance. *Journal of Fluid Mechanics, 26,* 779–792.
39. Troitskaya, Y. I., Sergeev, D. A., Ezhova, E. V., Soustova, I. A., & Kazakov, V. I. (2008). Self-induced internal waves excited by buoyant plumes in a stratified tank. *Doklady Earth Sciences, 419*(2), 506–510.
40. Williamson, N., Srinarayana, N., & Armsfield, S. W. (2008). Low-Reynolds-number fountain behavior. *Journal of Fluid Mechanics, 608,* 297–317.

Large Internal Solitary Waves in Shallow Waters

Valery Liapidevskii and Nikolay Gavrilov

Introduction

Nonlinear internal waves generated by internal tides play the important role in the energy transfer in shelf zones of seas. In coastal zones, shoreward propagation of surface and internal waves generally leads to breaking. The turbulence, generated by breaking and mixing processes at the wave fronts, induces the very effective mechanism of energy dissipation and momentum exchange, leading to intensive sediments suspension and transport in the shelf zone. Such high-energetic mechanisms of the shelf ventilation can effectively intensify the biological and hydrological processes in coastal waters. In particular, they can redistribute the waste waters and influence the water quality in near shore area. The run up of internal waves in near shore waters is very similar to the run up of surface long waves, but the process of internal wave breaking and dissipation is not quite understood. The main difference between internal and surface waves is that large internal waves can propagate for a long distance without breaking. In contrast to the energy dissipation mechanism for surface solitary waves in a homogeneous fluid, the energy dissipation in internal waves is closely connected with the entrainment and mixing in stratified shear flows. Very often, the wave fronts take the form of solitary wave trains with the large ratio of the wave amplitude to the upper layer depth [14]. Propagating to shore, they transform into the large amplitude internal waves of elevation. The large amplitude internal waves can be identified by their ability to carry trapped fluid horizontally for long distances [4, 15, 22, 24, 26, 27]. The transition from wave-like motion to the separate moving

V. Liapidevskii (✉) · N. Gavrilov
Lavrentyev Institute of Hydrodynamics, Lavrentyev ave. 15, Novosibirsk, Russia
e-mail: liapid@hydro.nsc.ru

N. Gavrilov
e-mail: gavrilov@hydro.nsc.ru

V. Liapidevskii
Novosibirsk State University, Pirogova str. 2, Novosibirsk, Russia

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_9

soliton-like waves ("solibores") containing trapped dense core is the common feature of the run-up process of internal waves. It can be observed in any shelf zone with high internal wave activity as well as in laboratory experiments. Breaking of internal solitary waves is closely connected with shear-induced instability at interfaces, which is an object of recent intense investigations in laboratory and field observations [1, 6, 16, 20, 23].

In the paper the multi-layer shallow water model describing the nonstationary interaction and decaying of large internal solitary waves is presented. The equations are the direct extension of the Green–Naghdi equations developed for open channel flows of multi-layer stratified fluids. It is shown that the two- and three-layer shallow water equations taking into account the dispersion effects, can describe large amplitude internal wave evolution for different types of flows (intrusions, bottom and subsurface internal waves in field and laboratory conditions).

Laboratory Experiments

Here we describe briefly the laboratory experiments performed in the Lavrentyev Institute of Hydrodynamics, which will be used in next sections to validate the mathematical models. More detailed description of the experiments can be found in [7–11].

Experiments were carried out in a test tank of length 3.2 m, width 0.2 m and depth 0.35 m, the walls of the test tank were made of Perspex. The test tank was divided by a vertical removable wall in two parts as it is shown in Fig. 1. The geometrical set-up of the experiments is clear from the sketches shown in the figure. Figure 1a illustrates the experiments on the symmetric solitary wave evolution along the interface. It shows the special installation (inclined bottom and lid in the flume) to provide the shoaling of symmetric solitary waves of the second mode over a shelf ($\alpha = \beta$) or the shoaling of the subsurface waves of depression ($\alpha > 0, \beta = 0$). Various applications of solitary wave dynamics revealed in laboratory experiments to the shoaling of large internal waves in a shelf zone have been discussed in [11, 12]. Here we focus our attention mostly on the special features of internal solitary wave propagation over the flat bottom ($\alpha = \beta = 0$). Figure 1b sketches the nonsymmetric solitary wave generation in the lock-exchange problem.

The experimental setup is shown in Fig. 2. For visualization of the flow pattern LIF-technique (Laser Induced Fluorescence) is used. The method is based on the fact that at low concentrations of fluorescein its luminosity during laser irradiation is directly proportional to the concentration. This allows not only to get qualitative information (for example, about mixing processes), but also quantitative information about the density of the liquid in any part of the investigated area. To create the light sheet, the diode-pumped solid-state laser "Mozart" is used, it provides a powerful continuous radiation at a wavelength of 532 nm. By changing thermostat temperature the radiation power could be varied from a few mW up to 5 W. The control system allows to change the width of the sheet and the direction of the light, also the mirror

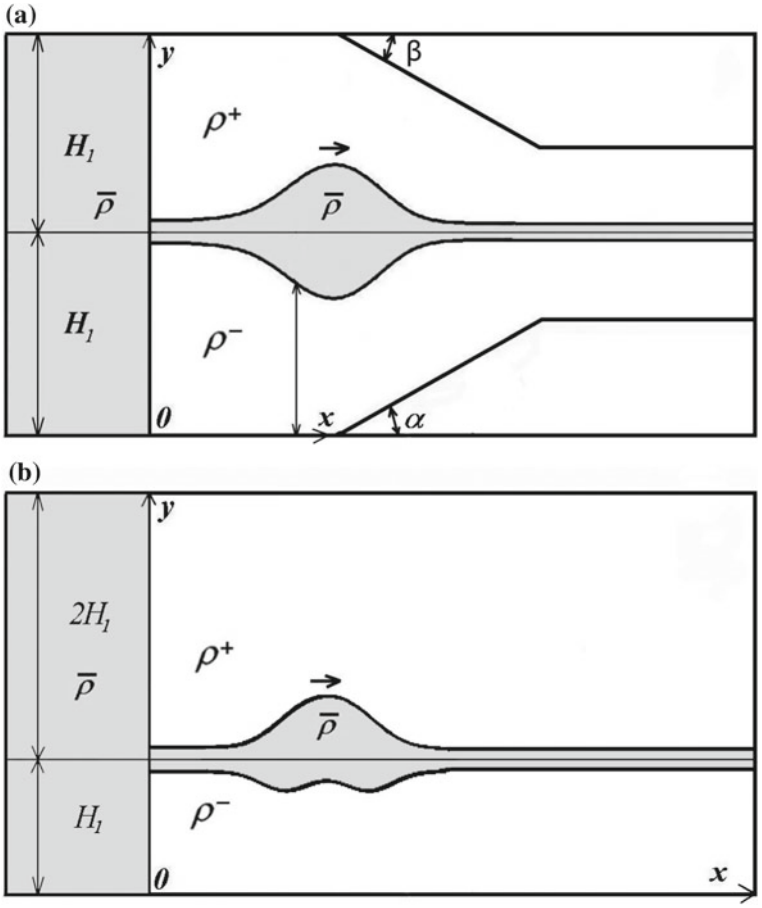


Fig. 1 Sketch of the lock problem: **a** symmetric solitary wave at the interface; **b** nonsymmetric solitary wave generation

vibration frequency can be varied in a wide range of parameters. In combination with laser power control it allows to get a uniform illumination of investigated area.

A weak solution of sugar or NaCl in water was used to create the stratification. Gravity current propagation was recorded with a digital video camera. Often in the field and laboratory experiments, the fixed probes are used. In this study new method is developed, that permits to transform the spatial pattern of the flow (Fig. 3a), obtained by video and photo records, to the temporal pattern, which could be fixed by stationary probes. This new method can be described as follows: a narrow strips, which consist of few pixels, are cut from each photo so that in new file containing all such fragments we record a temporary flow pattern (Fig. 3b). It allows to compare field data on large amplitude internal wave evolution in a coastal zone with laboratory experiments.

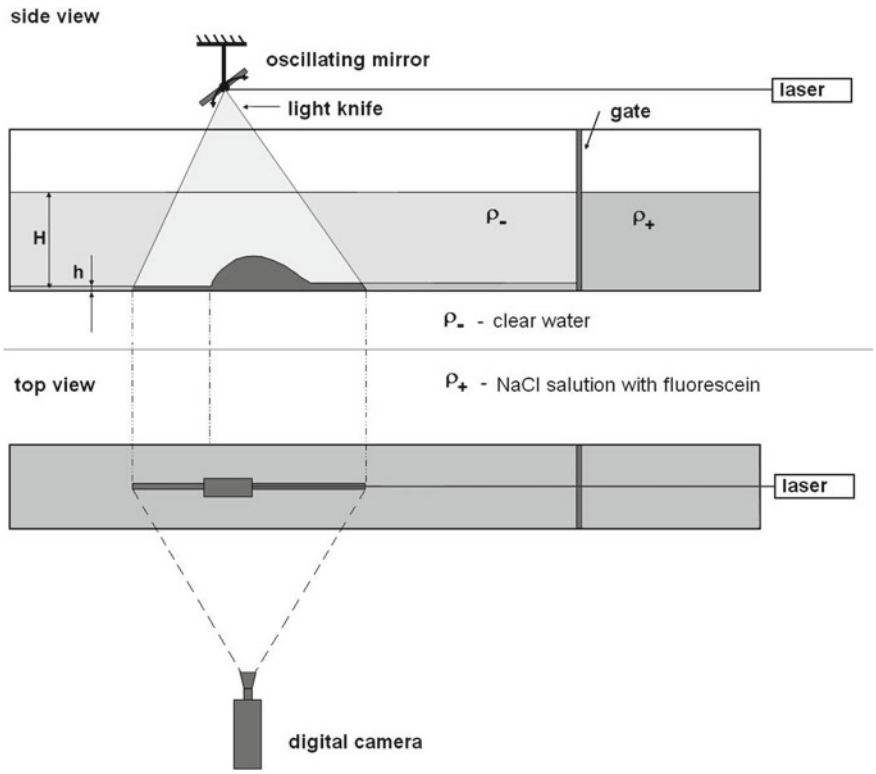


Fig. 2 Experimental setup

Another type of the flow visualization is portrayed in Fig. 4, where we show photographs of the generated solitary waves against the background of the luminous screen with a grid of inclined lines imposed on it. The optical inhomogeneities visualizing the perturbation propagation are distinctly seen. In high density gradient regions, we observed the characteristic distortion of lines, whereas the optical transparency of the fluid changes in the mixing areas [5]. In addition, dense fluid, which propagates along the interface in the form of a symmetric solitary wave, is slightly tinted with an ink solution for visualizing the mass transfer processes. The external wave boundaries are determined by the break and change in thickness of the inclined lines.

Bottom friction has no effect on the propagation of subsurface internal waves in oceans. But in laboratory experiments the bottom friction effects are replaced by the surface tension effect and the free surface can be considered as a rigid lid in experiments on large amplitude internal wave propagation. To avoid the bottom and surface effects on large amplitude waves, the series of laboratory experiments were conducted for short intrusions at the interface of two miscible fluids of different densities (Fig. 1a). The tank geometry was symmetric relative to the plane $y = H_1$

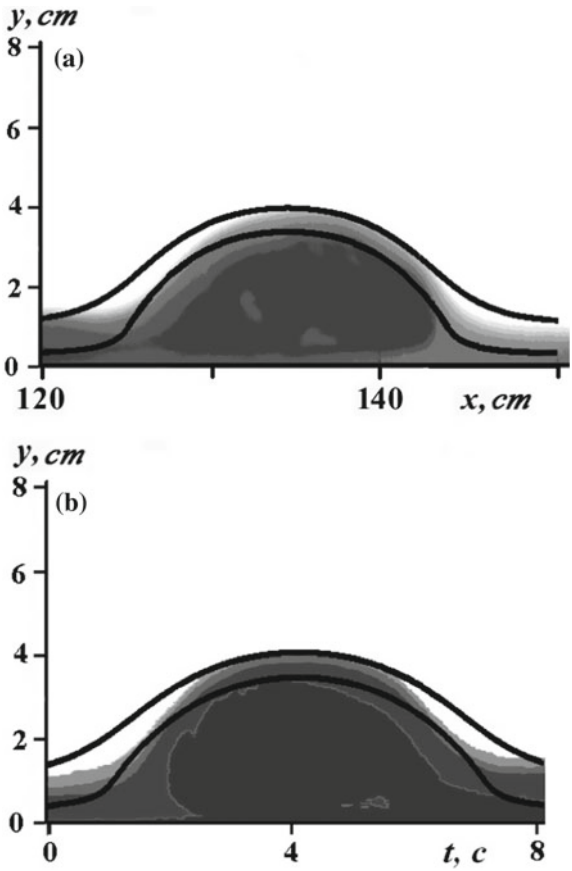


Fig. 3 Bottom solitary wave in the laboratory experiment [12]: **a** spatial patterns; **b** temporal patterns. The solid lines represent the results of calculations by the model ULM with the parameters: $h_0 = 0.3\,\text{cm}, \eta_0 = 0.8\,\text{cm}, \zeta_0 = 6.9\,\text{cm}, b = 5\,\text{cm/s}^2, \bar{b} = 0.5b, Fr = 0.476$

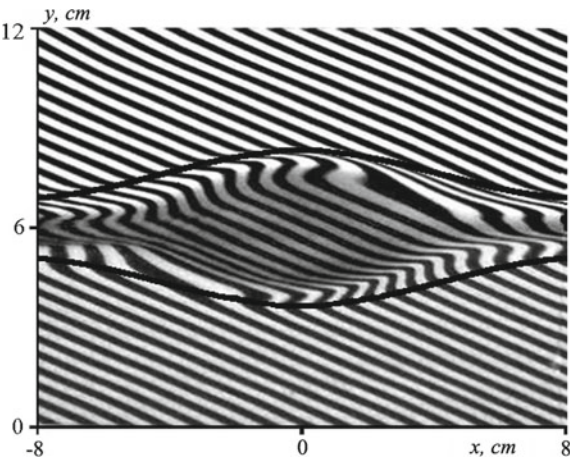


Fig. 4 Snapshot of a symmetric solitary wave of the second mode propagating from left to right at the interface. The solution of (14) for $Fr_s = 0.48, h_0 = 0.15$ is shown by solid curves

with the ordinate y chosen in the vertical direction. So the bottom and the lid of the tank had the same angles of inclination to the horizon. The length of the compartment with the mixed fluid of density $\bar{\rho} = 0.5(\rho^- + \rho^+)$ was chosen so that the only one solitary wave at the interface was produced. It was also symmetric relative to the undisturbed interface. Due to the flow symmetry, the corresponding component of the Reynolds stress at $y = H_1$ vanished. The velocities generated in the homogeneous layers by the intrusion were small enough and the friction at the bottom could be neglected too, but the friction between the surrounding fluid and the intrusion should be included in the mathematical model to describe properly the decay rate of solitary waves propagating along the interface. We do not specify in the paper all details of the experiments on the decay of symmetric solitary waves. Some approaches to the problem are discussed in [7] and is illustrated below.

When the symmetry of the experiment is broken (Fig. 1b), the interfacial intrusion starts to generate intense trailing waves of the first mode and loses its energy rather

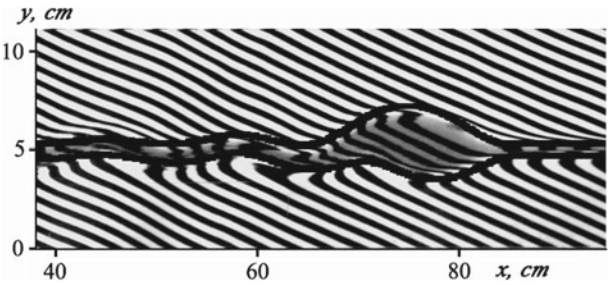


Fig. 5 Generation of intense internal waves of the first mode by the short intrusion at the interface: background is the experimental form of the intrusion generated in the lock problem with $h_0 = 5/12H, \bar{b} = 5/12b$; thick lines are the results of nonstationary calculations by the basic model (BM)

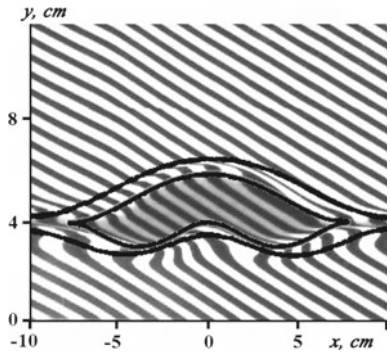


Fig. 6 Nonsymmetric solitary wave of the second mode generated in the lock problem with special choice of stratification shown in Fig. 1b ($h_0 = H/3, \bar{b} = b/3$): internal solid lines are constructed by the exact solution of SM and correspond to the boundaries of the dark (colored) fluid carried by the wave along the pycnocline, outer solid lines represent the steady-state solutions of the basic model (BM)

quickly (Fig. 5). Nevertheless, for special choice of stratification the nonsymmetric solitary wave can be generated, which is able to propagate along the channel keeping its permanent form (Fig. 6). Solid lines shown in Figs. 3, 4, 5 and 6 represent the solutions of the mathematical model developed in next sections.

Mathematical Models

Basic Model

As a basic model for simulation of nonlinear internal waves, we consider the three–layer shallow water equations for large amplitude, weakly nonhydrostatic long waves introduced in [19]. The model is a version of multi-layer strongly nonlinear equations derived in [2] with the following simplifications:

1. Boussinesq approximation $|\rho - \rho_0| \ll \rho_0$ is used, where ρ, ρ_0 are the density and reference density, correspondingly [14].
2. The pressure distribution in the intermediate layer is supposed to be hydrostatic.

The first condition is well justified due to small density variation in the ocean. The second one will be used in two cases:

(i) to describe large amplitude internal waves of the first mode with relative small thickness of the interlayer
(ii) to simulate interface solitary waves of the second mode, in which wave velocity and particle velocity differ only slightly.

In dimensional variables the governing equations take the form

$$
\begin{aligned}
h_t + (hu)_x = 0, \quad \eta_t + (\eta v)_x = 0, \quad \zeta_t + (\zeta w)_x = 0,\\
u_t + (\frac{1}{2}u^2 + b(h+z) + \bar{b}\eta + p)_x + \frac{\beta^-}{3h}(h^2\frac{d_1^2 h}{dt^2})_x = f^-,\\
v_t + (\frac{1}{2}v^2 + \bar{b}(h+\eta+z) + p)_x = \bar{f},\\
w_t + (\frac{1}{2}w^2 + p)_x + \frac{\beta^+}{3\zeta}(\zeta^2\frac{d_2^2\zeta}{dt^2})_x = f^+,\\
\frac{d_1}{dt} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x}, \quad \frac{d_2}{dt} = \frac{\partial}{\partial t} + w\frac{\partial}{\partial x},\\
b = (\rho^- - \rho^+)g/\rho^+, \quad \bar{b} = (\bar{\rho} - \rho^+)g/\rho^+.
\end{aligned}
\tag{1}
$$

Here g is the gravity acceleration, p is the modified pressure, $\rho^+, \bar{\rho}, \rho^-$ are the densities, ζ, η, h are the thicknesses, w, v, u are the mean horizontal velocities in the upper, intermediate and lower layers, correspondingly. The function $z = z(x)$ describes the bottom profile. We suppose that the bottom changes very smoothly, so we don't

include in (1) the terms containing the first and the second derivatives of the function z [13, 25]. The friction terms $f^{\pm}$ and $\bar{f}$ will be discussed later.

In Boussinesq approximation we also have

$$h + \eta + \zeta + z = H \equiv const. \tag{2}$$

In view of (1) the total flow rate

$$Q = Q(t) = hu + \eta v + \zeta w \tag{3}$$

can be found from boundary conditions. Finally, Eqs. (1)–(3) with $b \equiv const$, $\bar{b} \equiv const$ after excluding pressure p represent the closed system of equations for unknown variables h, ζ, u, w.

The small shallow water parameter $\beta = (H/L)^2$, where H is the channel depth and L is a typical wave length, in dimensional variables is taken equal to one, so the coefficients $\beta^{\pm}$ in (1) are chosen equal to 1 or 0 depending on the model consideration. Below we will use the following notations

- Basic Model (BM) for $\beta^+ = \beta^- = 1$.
- Upper Layer Model (ULM) for $\beta^+ = 1, \beta^- = 0$.
- Bottom Layer Model (BLM) for $\beta^+ = 0, \beta^- = 1$.
- Hydrostatic Model (HM) for $\beta^+ = \beta^- = 0$.

Remark 1 Without Boussinesq approximation and for $\eta \equiv 0$ BM coincides with the strong nonlinear two-layer model derived in [3].

Remark 2 The special cases of (1)–(3), namely, ULM and BLM, in which one of the outer layers is hydrostatic, can be applied effectively to large amplitude internal waves of elevation or depression, since in corresponding layers the mean particle velocity approaches the wave velocity and the second derivative in the pressure term vanishes.

Remark 3 Hydrostatic Model (HM) represents well-known three-layer shallow water equations in Boussinesq approximation [17].

Symmetric Model (SM)

Consider flows symmetric about the central line $y = H_1 = H/2$ with $\bar{\rho} = (\rho^+ + \rho^-)/2, \zeta \equiv h, w \equiv u$ (Fig. 1a). In this case, it is sufficient to consider flows only in low part of the channel ($0 < y < H_1$) and use the two-layer flow scheme. The governing equations for SM take the form

$$\begin{aligned} h_t + (hu)_x &= 0, \\ u_t + uu_x + \bar{b}h_x + p_x + \frac{1}{3h}(h^2 \frac{d_1^2 h}{dt^2})_x &= f^-, \\ v_t + vv_x + p_x &= \bar{f}, \end{aligned}$$

$$hu + (H_1 - h)v = Q_1(t). \tag{4}$$

Note that SM describes an important class of symmetric intrusions at the interface and, in particular, the symmetric solitary waves [7].

Travelling Waves

For the models BM, ULM, BLM, SM we construct travelling waves, i.e., solutions of Eqs. (1)–(3) or Eq. (4), which depend on the variable $\xi = x - Dt (D \equiv const)$. Such solutions exist for nondissipative models over a flat bottom ($f^{\pm} = 0, \bar{f} = 0, z \equiv 0$). All considered models are invariant under the Galilean transformation, therefore we can consider steady-state solutions as a representatives of travelling waves.

Steady-State Solutions of (1)–(3)

The following relations can easily be drawn for steady-state solutions of (1)–(3)

$$\begin{gathered} hu = Q^-, \quad \eta v = \bar{Q}, \quad \zeta w = Q^+, \\ \frac{1}{2}v^2 + \bar{b}(h+\eta) + p = \bar{J}, \quad h + \eta + \zeta = H \end{gathered} \tag{5}$$

and two more integrals can be derived after some calculations

$$\begin{gathered} \frac{1}{2}u^2 + bh + \bar{b}\eta + p - \frac{\beta^-}{3}u(h^2u')' - \frac{\beta^-}{6}h^2(u')^2 = J^-, \\ \frac{1}{2}w^2 + p - \frac{\beta^+}{3}w(\zeta^2 w')' - \frac{\beta^+}{6}\zeta^2(w')^2 = J^+. \end{gathered} \tag{6}$$

We express the unknown variables h, η, ζ, p, v as functions of u and w and reduce (1)–(3) to the system of ordinary differential equations

$$\begin{gathered} \frac{\beta^-}{3}h^2uu'' = \frac{1}{2}u^2 + bh + \bar{b}\eta + p + \frac{\beta^-}{2}h^2(u')^2 - J^-, \\ \frac{\beta^+}{3}\zeta^2 ww'' = \frac{1}{2}w^2 + p + \frac{\beta^+}{3}\zeta^2(w')^2 - J^+, \end{gathered} \tag{7}$$

where

$$\begin{gathered} h = Q^-/u, \quad \zeta = Q^+/w, \quad \eta = H - h - \zeta, \quad v = \bar{Q}/\eta, \\ p = \bar{J} - \frac{1}{2}v^2 - \bar{b}(h+\eta). \end{gathered}$$

By choosing the corresponding values $\beta^{\pm}$, Eq. (7) describe steady-state flows for all mentioned above models BM, ULM, BLM and HM.

Solitary Waves

Consider the special class of solutions for ULM, BLM, SM, which satisfy for $|x| \to \infty$ the following conditions

$$\begin{aligned} &h \to h_0, \quad \zeta \to \zeta_0, \quad u \to u_0, \quad w \to u_0, \\ &h' \to 0, \quad h'' \to 0, \quad \zeta' \to 0, \quad \zeta'' \to 0 \end{aligned} \tag{8}$$

(here and below, the prime denotes differentiation with respect to the variable x). Solutions of the problem (7)–(8) describe solitary waves moving with the velocity u_0 in the laboratory system of coordinates.

Solitary Waves in SM

For SM the solitary waves have been investigated in [7, 8]. The soliton-like solution for SM can be found in quadratures. To see that we rewrite (7) for SM in dimensionless variables

$$\begin{aligned} h_* = h/H_1, \quad u_* = u/\sqrt{\bar{b}H_1}, \quad p_* = p/\bar{b}H_1, \\ t_* = t\sqrt{\bar{b}/H_1}, \quad x_* = x/H_1, \quad Fr_s = u_0/\sqrt{\bar{b}H_1}. \end{aligned}$$

("star" is omitted below).

The governing equations for (7)–(8) take the form

$$\begin{aligned} hu = h_0 Fr_s, \quad (1-h)v = (1-h_0)Fr_s, \\ \frac{1}{2}v^2 + p = \frac{1}{2}Fr_s^2 + p_0, \\ \frac{1}{2}u^2 + h + p + \frac{1}{3}hu^2h'' - \frac{1}{6}u^2(h')^2 = \frac{1}{2}Fr_s^2 + h_0 + p_0. \end{aligned} \tag{9}$$

Equation (9) can be reduced to ODE [9]

$$(h')^2 = G(h) = \frac{3(h-h_0)^2(Fr_s^2 - h + h^2)}{Fr_s^2 h_0^2(1-h)}. \tag{10}$$

It follows from (10) that the solitary waves exist for

$$Fr_s^2 - h + h^2 > 0.$$

Therefore, the admissible intervals for h are

$$0 < h_0 \le h \le h^-, \quad h^+ \le h \le h_0 < 1, \tag{11}$$

where

$$h^{\pm} = \frac{1 \pm \sqrt{1 - 4Fr_s^2}}{2}. \tag{12}$$

Naturally, the necessary condition for solitary wave existence reads

$$Fr_s \le \frac{1}{2}. \tag{13}$$

The profile of the symmetric solitary wave can be found from quadratures

$$x = x^{\pm}(h) = \pm \int_{h_m}^{h} \frac{ds}{\sqrt{G(s).}} \tag{14}$$

Here $h_m = h^+$ for a wave of depression and $h_m = h^-$ for a wave of elevation. The solution of (14) for $Fr_s = 0.48, h_0 = 0.15$ is shown by solid curves in Fig. 4 (in dimensional variables).

Remark 4 It is important to underline that solutions (14) can be effectively applied to large amplitude internal wave simulation for different scales of flow not only at interfaces (the second mode symmetric waves), but also to subsurface waves of depression and to bottom waves of elevation.

Solitary Waves in BLM and ULM

In similar manner, solitary wave solutions for BLM and ULM may be represented in quadratures. Consider (7) with $\beta^- = 1, \beta^+ = 0$. The governing equations describing the long subsurface waves of depression (BLM) take the form

$$\begin{aligned} \frac{1}{2}u^2 + bh + \bar{b}\eta + p - \frac{1}{3}h^2uu'' + \frac{1}{2}h^2(u')^2 = \frac{1}{2}u_0^2 + bh_0 + \bar{b}\eta_0 + p_0 = J^-, \\ \frac{1}{2}w^2 + p = \frac{1}{2}u_0^2 + p_0 = J^+, \\ \frac{1}{2}v^2 + \bar{b}(h+\eta) + p = \frac{1}{2}u_0^2 + \bar{b}(h_0 + \eta_0) + p_0 = \bar{J}, \\ hu = h_0u_0 = Q^-, \quad \zeta w = \zeta_0 u_0 = Q^+, \quad \eta v = \eta_0 u_0 = \bar{Q}, \quad h + \eta + \zeta = H. \end{aligned} \tag{15}$$

In view of (15) we have

$$P(\eta,\zeta)=\frac{\bar{Q}^2}{2\eta^2}-\frac{(Q^+)^2}{2\zeta^2}+\bar{b}(H-\zeta)=\bar{J}-J^+=\bar{b}(h_0+\eta_0)$$

or

$$\eta^2=\frac{\bar{Q}^2}{2(\bar{J}-J^+)+(Q^+)^2/\zeta^2-2\bar{b}(H-\zeta)}. \tag{16}$$

It follows from (16)

$$\eta=\eta(\zeta),\quad h=h(\zeta)=H-\zeta-\eta(\zeta). \tag{17}$$

We may find dependencies $\zeta=\zeta_1(h),\eta=\eta_1(h)$ from (17) and rewrite (15) in the form

$$hh''-\frac{1}{2}(h')^2=\frac{3}{u^2}(J^--\frac{1}{2}u^2-bh-\bar{b}\eta+\frac{1}{2}w^2-J^+)=\Phi(h). \tag{18}$$

Finally, (18) reduces to ODE

$$(h^{-1/2}h')'=h^{-3/2}\Phi(h)$$

or

$$(h')^2=2h\Psi(h) \tag{19}$$

with

$$\Psi(h)=\int_{h_0}^{h}\frac{\Phi(s)}{s^2}ds. \tag{20}$$

The wave profile $h=h(x)$ is calculated from (19), (20). Other unknown variables may be found from (17)–(19). For given dimensionless parameters $\bar{b}/b, h_0/H, \eta_0/H$ solitary waves represent the one-parameter family depending on the Froude number

$$Fr=\frac{u_0}{\sqrt{bH}}. \tag{21}$$

Remark 5 For ULM ($\beta^-=0,\beta^+=1$) describing the bottom waves of elevation, the one-parameter family of solitary waves may be constructed in a similar way or just by "inversion" solutions of BLM relative to the midline of the channel.

Nonsymmetric Solitary Waves

In contrast to the models ULM, BLM, SM considered above, the solitary waves for BM can be found only for special values of the dimensionless parameters $\bar{b}/b, h_0/H,$

η_0/H and Fr. The particular case of BM corresponding to the symmetric solitary waves of the second mode (SM) is considered in section "Solitary Waves in SM".

An interesting example of a nonsymmetric solitary wave was found experimentally for special initial data in the lock problem shown in Fig. 1b, namely, for $h_0 = H/3$, $\bar{b} = b/3, \eta_0 \ll H$. The length of the left compartment has been chosen so that the only one wave moving on the right was generated. The experimental photo is shown in Fig. 6. From the symmetry consideration of the flow in the regions $0 \le y \le h_0$ and $h_0 \le y \le H$, it may be concluded that the intrusion from the compartment (the wave core with dark fluid) consist of the wave of elevation and two waves of depression. Such waves have been found as solutions of SM for the case $\eta_0 \ll H$ in [11] (Fig. 4).

Here we consider the case $\eta_0 > 0$. To construct a soliton-like solution, we must start with the data $h = h_0, \zeta = \zeta_0, u = u_0, w = u_0, u' = 0, w' = 0$ for $x \to -\infty$ and use the asymptotics

$$\vec{U}(x) = \vec{U}_0 + \hat{U} exp(\lambda x), \quad \lambda > 0, \tag{22}$$

where $\vec{U}(x) = (h, \zeta, u, w)^{Tr}$. Omitting the standard calculations connecting with the linearization of (7), we have

$$\lambda = \sqrt{\frac{3}{2A}(\sqrt{B^2 - 4AC} - B)}, \tag{23}$$

where

$$\begin{aligned} A &= h_0^2 \zeta_0^2 u_0^4, \\ B &= h_0^2 u_0^2 (\bar{b}\zeta_0 - (1 + \zeta_0/h_0)u_0^2) + \zeta_0^2 u_0^2 ((b - \bar{b})h_0 - u_0^2) - h_0 \zeta_0^2 u_0^4/\eta_0, \\ C &= ((b - \bar{b})h_0 - u_0^2)(\bar{b}\zeta_0 - (1 + \zeta_0/h_0)u_0^2) - h_0 u_0^2 (\bar{b}\zeta_0 - u_0^2)/\eta_0. \end{aligned} \tag{24}$$

The asymptotics (22)–(24) is developed for arbitrary value of the Froude number $Fr = u_0^2/\sqrt{bH}$ and the initial layer parameters $h_0/H, \zeta_0/H, \bar{b}/b$, but the steady-state solutions of (7) satisfying (2) would represent solitary waves with (8) only for special combinations of the initial layer parameters. In the next section the example of the nonsymmetric solitary wave will be considered.

Model Validation and Nonstationary Calculations

Consider first steady-state solutions of the developed models and their application to experimental data, then nonstationary solutions of (1) will be discussed.

Internal Solitary Waves in Laboratory and Field Experiments

Internal Waves of Depression

The constructed solutions for ULM and BLM can be compared with numerous laboratory and field experimental data. We consider first the laboratory experiments on internal solitary waves of depression performed in [6]. These results are the most appropriate to the three-layer flow scheme of BLM, since the thickness of the intermediate layer was controlled in experiments. Solitary wave profiles calculated by (19) are shown in Fig. 7 together with the experimental data from [6] (Fig. 3c, d). Analogously to the laboratory experiments described above in section "Laboratory Experiments", the solitary waves of depression were generated by removing the vertical gate between two stratified fluids at rest (the lock-exchange problem). The dimensional parameters determining the waves are $h_0 = 30\,\text{cm}$, $\eta_0 = 5.5\,\text{cm}$, $\zeta_0 = 5\,\text{cm}$, $b = 20\text{cm/s}^2$, $\bar{b} = \frac{1}{2}b$, $Fr = 0.485$ (Fig. 7a) and $h_0 = 29.5\,\text{cm}$, $\eta_0 = 5.5\,\text{cm}$, $\zeta_0 = 1.5\,\text{cm}$, $b = 20\,\text{cm/s}^2$, $\bar{b} = \frac{1}{2}b$, $Fr = 0.462$ (Fig. 7b) (the runs No. 1 and No. 24 in [6]). Solid black lines on Fig. 7 are the results of calculations, solid white lines are the experimentally found positions of pycnocline between the layers, the backgrounds are photos of the experiment.

In oceans and seas the density stratification is rather complicated. Nevertheless, we may apply the three-layer shallow water models for large internal waves in shelf zones. In Fig. 8a the time-dependence of isopycnal deformation during the subsurface solitary wave passage in the shelf zone of the South China sea is shown [20] (Fig. 3e). Thick black lines are the result of calculation by BLM with the following dimensional data: $h_0 = 395\,\text{m}$, $\eta_0 = 100\,\text{m}$, $\zeta_0 = 30\,\text{m}$, $b = 4 \times 10^{-2}\text{m/s}^2$, $\bar{b} = 0.3b$, $Fr = 0.447$. The total depth is 525 m.

Internal Waves of Elevation

As was mentioned above, UBL is developed for large internal waves of elevation. In laboratory experiments it can be used for calculations as the spatial patterns so the temporal patterns of bottom internal waves (Fig. 3a and b). The solid black lines are the results of calculations by ULM with the parameters: $h_0 = 0.3\,\text{cm}$, $\eta_0 = 0.8\,\text{cm}$, $\zeta_0 = 6.9\,\text{cm}$, $b = 5\,\text{cm/s}^2$, $\bar{b} = 0.5b$, $Fr = 0.476$. UBL also can be applied for field data taken from [18] (Fig. 8b). The measurements of the vertical temperature distribution were performed 09.09.2011 in nearshore waters of the Sea of Japan at the depth 18 m. In Fig. 8b the contour plot of temperature (thin lines) is shown together with the calculations by ULM (thick lines). The upper layer was practically homogeneous with the temperature 18 °C. The "bolus" of cold water with the temperature 10–12 °C in the wave core passed through the bottom measurement station in 5 min. The wave amplitude reached half of the total depth. The ULM parameters of the wave are: $h_0 = 0.9\,\text{m}$, $\eta_0 = 2\,\text{m}$, $\zeta_0 = 15.1\,\text{m}$, $b = 1.2 \times 10^{-2}\,\text{m/s}^2$, $\bar{b} = 0.3b$, $Fr = 0.47$.

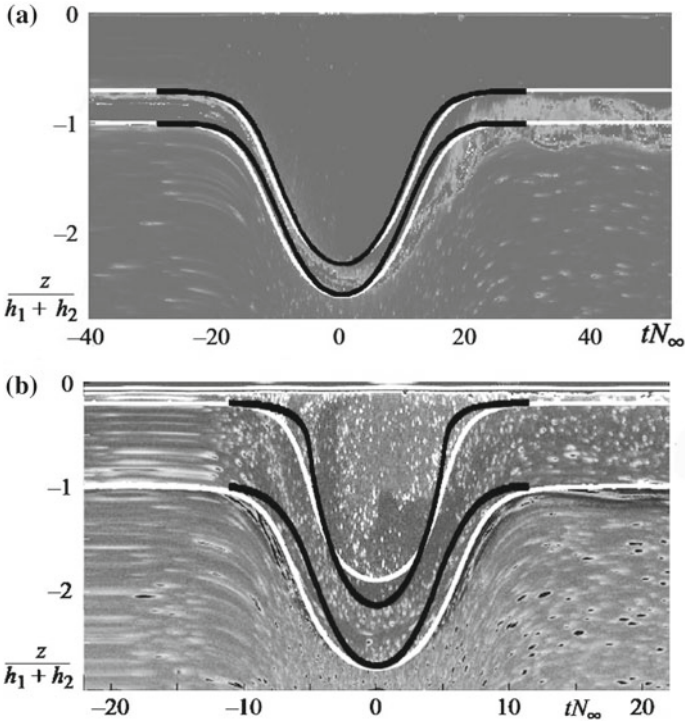


Fig. 7 Solitary wave profiles calculated by (19) (black solid lines) together with the experimental data from [6]: **a** $h_0 = 30\,\text{cm}$, $\eta_0 = 5.5\,\text{cm}$, $\zeta_0 = 5\,\text{cm}$, $b = 20\,\text{cm/s}^2$, $\bar{b} = \frac{1}{2}b$, $Fr = 0.485$ (Run 1); **b** $h_0 = 29.5\,\text{cm}$, $\eta_0 = 5.5\,\text{cm}$, $\zeta_0 = 1.5\,\text{cm}$, $b = 20\,\text{cm/s}^2$, $\bar{b} = b/2$, $Fr = 0.462$ (Run 24). White solid lines are experimental boundaries of pycnocline

Nonsymmetric Solitary Waves

An interesting example of the short intrusion spreading as the nonsymmetric internal solitary wave is shown in Fig. 6. The wave carries colored fluid from the left compartment after removing the vertical wall in the lock-exchange problem depicted in Fig. 1b. The boundaries of the trapped core practically coincide with the solution of (7) with vanishing interface thickness out of the wave. This rather simple solution (internal solid lines in Fig. 6) has been described in [11]. Note that the mean fluid velocity in the core is constant and equal to the wave velocity. Therefore, the wave profile in such wave may be found in explicit form [9].

Let us consider further the steady-state solution of (7)–(8) with the asymptotics (22)–(24). For the dimensional parameters $H = 12\,\text{cm}$, $\eta_0 = 0.4\,\text{cm}$, $h_0 = H/3 - \eta_0/2$, $\zeta_0 = 2H/3 - \eta_0/2$, $b = 5\,\text{cm/s}^2$, $\bar{b} = b/3$, which correspond to experimental data, the numerical solution represent the solitary wave (outer solid lines in Fig. 6). It is demonstrated experimentally [11] that such wave keeps its form for a long

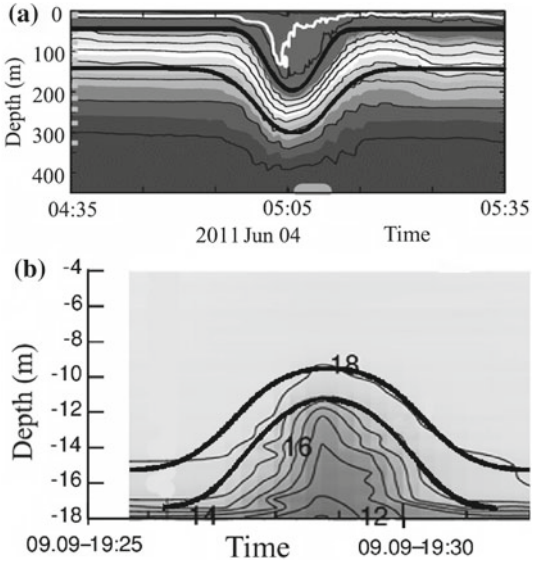


Fig. 8 Solitary waves in field data: **a** time-dependence of isopycnal deformation during the subsurface solitary wave passage in the shelf zone of the South China sea [20]. Thick black lines are the result of calculation by BLM ($h_0 = 395\,\text{m}$, $\eta_0 = 100\,\text{m}$, $\zeta_0 = 30\,\text{m}$, $b = 4 \times 10^{-2}\text{m/s}^2$, $\bar{b} = 0.3b$, $Fr = 0.447$); **b** bottom solitary wave in the Sea of Japan at the depth 18 m. The contour plot of temperature (thin lines) is shown together with the calculations by ULM (thick lines). The ULM parameters of the wave are: $h_0 = 0.9\,\text{m}$, $\eta_0 = 2\,\text{m}$, $\zeta_0 = 15.1\,\text{m}$, $b = 1.2 \times 10^{-2}\,\text{m/s}^2$, $\bar{b} = 0.3b$, $Fr = 0.47$

distance as it moves along the pycnocline in the flume. If the governing parameters in the lock problem is changed, the intrusion stops to be soliton-like and it starts to generate nonstationary internal waves of the first mode (see section "Nonsymmetric Solitary Waves").

Non-stationary Problem

Let us return to the basic model (1). For numerical realization of Eq. (1) it is convenient to rewrite the system in the form of conservation laws. Analogously to [12], Eq. (1) can be written as follows

$$h_t + (hu)_x = 0, \quad \zeta_t + (\eta w)_x = 0,$$
$$K_t + (Ku - \frac{1}{2}u^2 + b(h+z) + \bar{b}\eta + p - \frac{\beta^-}{2}h^2u_x^2)_x = f^-,$$
$$R_t + (Rw - \frac{1}{2}w^2 + p - \frac{\beta^+}{2}\zeta^2 w_x^2)_x = f^+. \qquad (25)$$

Here

$$K = u - \frac{\beta^-}{3h}(h^3 u_x)_x, \quad R = w - \frac{\beta^+}{3\zeta}(\zeta^3 w_x)_x,$$
$$\eta = H - h - \zeta, \quad v = \frac{Q(t) - hu - \zeta w}{\eta}. \tag{26}$$

We apply Eqs. (25)–(26) for numerical calculations of nonstationary problems, concerning the propagation of large amplitude internal waves in the frame of models BM, ULM, BLM, SM. In all cases we consider wave generation inside of the channel (the lock problem etc.), so we apply the "wall condition" at the left and right boundaries of the calculation domain ($Q \equiv 0$). The variables h, ζ, K, R are considered as dependent variables, describing the time evolution of the flow, and the velocities $u = u(t, x)$ and $w = w(t, x)$ can be restored for given $h, \zeta, h_x, \zeta_x, K$ and R from the linear system of ODE (26). This numerical scheme have been developed in [21] for open channel flows and applied in [9, 11] for internal wave calculations. The numerical scheme is formally a version of Godunov's scheme, in which the fluxes through the lateral boundaries of the meshes are calculated by a Riemann solver. To find the appropriate time step for SM or USM, BSM, BM models we consider the equilibrium system without dispersion ($\beta^{\pm} = 0$), which coincides with common two or three-layer shallow water equations, correspondingly. In all calculations presented in the paper the boundary conditions are $u = 0, w = 0$ at the side walls of the simulating domain. As the initial conditions, we use the exact solution (10) for SM or the step-wise functions $h = h_0(x), \zeta = \zeta_0(x)$ together with $u = u_0(x) \equiv 0, w = w_0(x) \equiv 0, K = K_0(x) \equiv 0, R = R_0(x) \equiv 0$. The topography $y = z(x)$ is also included, when the problem on the solitary wave shoaling is considered.

Decay of Solitary Waves

In real stratified fluids solitary waves do not propagate at a constant velocity. Friction and entrainment of an ambient fluid result in the wave deceleration. Experiments on the symmetric solitary wave propagation along the pycnocline together with numerical calculations by SM have demonstrated the ability of the two-layer flow scheme to simulate the dissipation processes in the flow [7, 9]. It was shown that the main input in the energy dissipation of solitary waves was given by the interfacial friction term in (4)

$$f^- = -\frac{c_i(u - v)|u - v|}{h}, \quad \bar{f} = \frac{c_i(u - v)|u - v|}{H_1 - h} \tag{27}$$

and the bottom friction could be neglected. The friction coefficient $c_i = 10^{-2}$ was chosen considerably larger then the usual value of the bottom friction coefficient $c_b \sim 4 \times 10^{-3}$ for turbulent flows, because the mechanisms of the bottom and interlayer

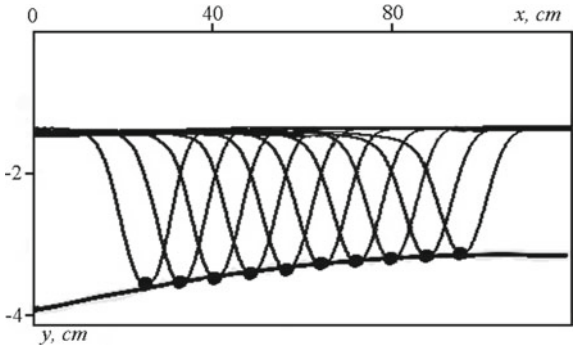


Fig. 9 Decay of the symmetric solitary wave. Bold points and bold line represent the experimental positions of wave crests. Thin solid lines show the calculated wave profiles through the fixed time interval (4 s). The exact solution (10), (14) with $H_1 = 5\,\text{cm}$, $h_0 = 3.65\,\text{cm}$, $\bar{b} = 2.5\,\text{cm/s}^2$, $Fr = 0.48$ was used as initial data

friction are quite different. If the bottom friction in turbulent flows is a rather classic problem, the dissipation due to instability generation at the interfaces caused by the large internal wave propagation is the object of intensive investigations in the last decades [1, 6].

The results of the numerical calculation of the solitary wave propagation in a flat horizontal channel is shown in Fig. 9 (the model SM). The exact solution (10), (14) with $H_1 = 5\,\text{cm}$, $h_0 = 3.65\,\text{cm}$, $\bar{b} = 2.5\,\text{cm/s}^2$, $Fr = 0.48$ was used as initial data. The calculated wave profiles are plotted through the fixed time interval $\Delta t = 4$ s (thin lines). The bold points in Fig. 9 together with the bold line passing through them are the positions of the solitary wave crests found from experiments [9]. One can see from Fig. 9 that the mathematical model (4), (27) with the friction coefficient $c_i = 10^{-2}$ describes the decay of the solitary wave along the channel. It is important to note that a solitary wave lost its initial symmetry during propagation due to friction effects and the shedding rate of mass from the wave is closely related to the friction rate.

Shoaling Solitary Waves

In the section we compare the ability of the two- and three-layer models (SM and BM) to simulate the shoaling subsurface solitary waves up to the moment of an internal bore formation. Transformation of the internal solitary wave of depression in continuously stratified fluid over a slope is shown in Fig. 10. The 2D nonlinear numerical model with continuous vertical stratification and turbulent exchange has been developed in [28]. The model contains a number of parameters and have been applied to different types of stratified flows. The calculated wave profiles at different positions over the slope are taken from Fig. 2a and b in [28]. The thin lines show the

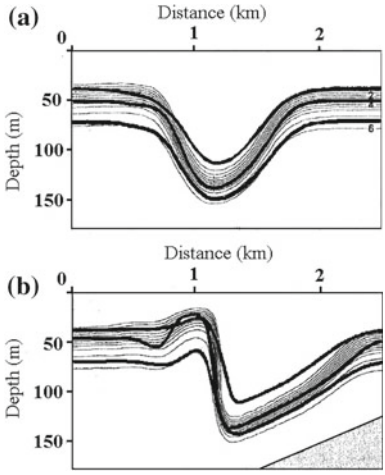


Fig. 10 Shoaling internal wave over a slope ($\tan\alpha = 0.05$). The wave moves to the right. Comparison of numerical results from [28] (thin lines, Fig. 2a and b) with the numerical solutions by the basic model (BM). The outer thick curves show the boundaries of the layers in BM and the internal thick curve shows the interface in the two-layer flow described by SM [10]. Corresponding wave parameters are given in the text

deformation of the density level with the density contour interval of $0.5\,\mathrm{kg/m^3}$. The comparison with calculations by the two-layer model SM has been performed in [10] (Fig. 7). In both cases the calculations have started with soliton-like initial profile of the wave. In Fig. 10 we combine these calculations with calculations by the three-layer model BM. Note that for BM calculations the initial solitary wave profile is constructed by the steady-state solution of BLM (19)–(20). The initial stratification for the three-layer scheme of flow is chosen as follows: $H = 300\,\mathrm{m}$, $h_0 = 230\,\mathrm{m}$, $\eta_0 = 30\,\mathrm{m}$, $b = 6 \times 10^{-2}\,\mathrm{m/s^2}$, $\bar{b} = 0.25b$, $Fr = 0.495$. The wave moves to the right over the slope ($z_x \equiv 0.05$). The outer thick curves show the boundaries of the layers in BM and the internal thick curve shows the interface in the two-layer flow described by SM. We may draw two main conclusions from Fig. 10:

1. the models derived above may be applied to simulate large internal waves in continuously stratified fluids too;
2. shoaling large internal waves can be described adequately by the models SM, BLM, which don't take into account nonhydrostatic effects in the upper layer.

Nonsymmetric Solitary Waves

In section "Nonsymmetric Solitary Waves" the steady-state solution of BM describing a nonsymmetric solitary wave of the second mode was constructed and compared with the experiment (Fig. 6). To investigate the stability of the solution, we must calculate the corresponding nonstationary problem. In Fig. 11 the result of the nonsymmetric wave evolution calculated by the three-layer model (BM) is shown. As the initial data for calculations, the wave of permanent form shown in Fig. 6 is chosen. It may be concluded from Fig. 11 that after passing more then ten initial lengths

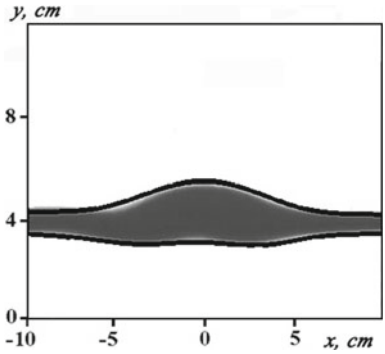


Fig. 11 Evolution of the nonsymmetric solitary wave shown in Fig. 6. The wave profile (solid lines) is calculated by the three-layer model (BM). The dark (colored) domain is the fluid carried by the wave along the pycnocline in the experiment

the wave keeps its symmetry, though the amplitude of the wave slightly decreases. Moreover, the calculated wave boundaries correspond to the boundaries of the dark (colored) fluid carried by the wave along the pycnocline in the corresponding experiment taken from [11] (Fig. 4b) and sketched in Fig. 1b.

In the same experiment shown in Fig. 1b, but with the initial depth h_0 slightly changed, the wave symmetry breaks. In Fig. 5 the experimental form of the intrusion generated in the lock problem shown in Fig. 1b is presented together with the numerical calculations by the model BM. The only distinction from the previous case of the symmetric wave is that the initial depth of the lower layer was taken $h_0 = 5/12H$ instead of $h_0 = 4$ cm and, correspondingly, $\bar{b} = 5b/12$. We can see that the wave lost its permanent form and it is generating the trailing waves of the first mode. It may be concluded also from Fig. 5 that BM reproduces well the main features of the nonstationary wave interaction.

Conclusions

In the paper the hierarchy of multi-layer shallow water equations describing large internal wave dynamics is developed. The main feature of considered models is that the nonhydrostatic effects are taken into account, but not in all layers. Such models allow us to find analytically soliton-like solutions representing internal waves of the first and the second modes. Laboratory experiments as well as the field data presented in the paper show that the shallow water approximation is adequate for large internal waves, which amplitude is comparable with the total depth of the channel. The three-layer shallow water equations (1) with the intermediate hydrostatic layer is chosen as the basic model (BM). The equations in the outer layers contain the additional nonhydrostatic pressure terms analogous to that in Green-Naghdi equations developed for open channel flows. In fact, Eq. (1) are a variant of multi-layer equations derived in [2]. The basic model can be applied effectively to simulate intrusions propagating along the interfaces (Fig. 5). Moreover, it may be used for simulation of

shoaling internal waves and solibore formation (Fig. 10). The steady-state solution of BM describes also the nonsymmetric solitary wave of the second mode shown in Fig. 6.

The submodels ULM, BLM, SM of the basic model BM allow us to find the soliton-like waves in two- or three-layer stratified fluids. One-parameter family of solitary solutions for ULM, BLM or for SM depends on the Froude number Fr or Fr_s, correspondingly. It is worth to note that BM doesn't have such solutions for arbitrary initial flow parameters.

Decay of internal waves due to breaking and entrainment processes is very important for wave dynamics. It is shown for the symmetric solitary waves of the second mode that the decay of internal waves can be simulated by the corresponding friction terms (27) (Fig. 9). Nevertheless, future development of the multi-layer shallow water approach must include the modelling of mixing and entrainment processes at the interfaces.

Acknowledgements This work was supported by the Russian Foundation for Basic Research (Grant No. 15-01-03942).

References

1. Carr, M., Franklin, J., King, S. E., Davies, P. A., Grue, J., & Dritschel, D. G. (2017). The characteristics of billows generated by internal solitary waves. *Journal of Fluid Mechanics*, *812*, 541–577.
2. Choi, W. (2000). Modeling of strongly nonlinear internal gravity waves. In: Y. Goda, M. Ikehata and K. Suzuki (Eds.), *Proceedings of the Fourth International Conference on Hydrodynamics* (pp. 453–458).
3. Choi, W., & Camassa, R. (1999). Fully nonlinear internal waves in a two-fluid system. *Journal of Fluid Mechanics*, *386*, 1–36.
4. Derzho, O. G., & Grimshaw, R. (2007). Asymmetric internal solitary waves with a trapped core in deep fluids. *Physics of Fluids*, *19*, 096601.
5. Ermanyuk, E. V., & Gavrilov, N. V. (2007). A note on the propagation speed of a weakly dissipative gravity current. *Journal of Fluid Mechanics*, *574*, 393–403.
6. Fructus, D., Carr, M., Grue, J., Jensen, A., & Davies, P. A. (2009). Shear-induced breaking of large internal solitary waves. *Journal of Fluid Mechanics*, *620*, 1–29.
7. Gavrilov, N. V., & Lyapidevskii, V Yu. (2009). Symmetric solitary waves in a two-layer fluid. *Doklady Physics*, *54*(11), 508–511.
8. Gavrilov, N. V., & Liapidevskii, V. Yu. (2010). Finite-amplitude solitary waves in a two-layer fluid. *Journal of Applied Mechanics and Technical Physics*, *51*(4), 471–481.
9. Gavrilov, N., Liapidevskii, V., & Gavrilova, K. (2011). Large amplitude internal solitary waves over a shelf. *Natural Hazards and Earth System Sciences*, *11*, 17–25.
10. Gavrilov, N., Liapidevskii, V., & Gavrilova, K. (2012). Mass and momentum transfer by solitary internal waves in a shelf zone. *Nonlinear Processes in Geophysics*, *19*, 265–272.
11. Gavrilov, N. V., Liapidevskii, V. Y., & Liapidevskaya, Z. A. (2013). Influence of dispersion on the propagation of internal waves in a shelf zone. *Fundam. prikl. gidrofiz.*, *6*, 25–34 (in Russian).
12. Gavrilov, N. V., Liapidevskii, V Yu., & Liapidevskaya, Z. A. (2015). Transformation of large amplitude internal waves over a shelf. *Fundam. prikl. gidrofiz.*, *8*(3), 32–43 (in Russian).

13. Gavrilyuk, S. L., Liapidevskii, V. Yu., & Chesnokov, A. A. (2016). Spilling breakers in shallow water: Applications to Favre waves and to the shoaling and breaking of solitary waves. *Journal of Fluid Mechanics*, *808*, 441–468.
14. Helfrich, K. R., & Melville, W. K. (2006). Long nonlinear internal waves. *Annual Review of Fluid Mechanics*, *38*, 395–425.
15. Lamb, K. G. (2003). Shoaling solitary internal waves: On a criterion for the formation of waves with trapped cores. *Journal of Fluid Mechanics*, *478*, 81–100.
16. Lamb, K. G., & Farmer, D. (2011). Instabilities in an internal solitary-like wave on the Oregon Shelf. *Journal of Physical Oceanography*, *41*, 67–87.
17. Liapidevskii, V. Yu., & Teshukov, V. M. (2000). *Mathematical models of long-wave propagation in an inhomogeneous fluid*. Novosibirsk: Publishing House of the Siberian Branch of the Russian Academy of Sciences (in Russian).
18. Liapidevskii, V. Yu., Kukarin, V. F., Navrotsky, V. V., & Khrapchenkov, F. F. (2013). Evolution of large amplitude internal waves of in a swash zone. *Fundam. prikl. gidrofiz.*, *6*, 35–45 (in Russian).
19. Liapidevskii, V. Yu., Novotryasov, V. V., Khrapchenkov, F. F., & Yaroshchuk, I. O. (2017). Internal undular bore in a shelf zone. *Journal of Applied Mechanics and Technical Physics, 58*(5), 809–818.
20. Lien, R Ch., Henyey, F., Ma, B., & Yang, Y. J. (2014). Large-amplitude internal solitary waves observed in the northern South China sea: Properties and energetic. *Journal of Physical Oceanography*, *44*(4), 1095–1115.
21. Le Metayer, O., Gavrilyuk, S., & Hank, S. (2010). A numerical scheme for the Green-Naghdi model. *Journal of Computational Physics*, *229*, 2034–2045.
22. Maxworthy, T. (1980). On the formation of nonlinear internal waves from the gravitational collapse of mixing regions in two and three dimensions. *Journal of Fluid Mechanics*, *96*, 47–64.
23. Moum, J. N., Farmer, D. M., Smyth, W. D., Armi, L., & Vagle, S. (2003). Structure and generation of turbulence at interfaces strained by internal solitary waves propagating shoreward over the continental shelf. *Journal of Physical Oceanography*, *33*, 2093–2112.
24. Moum, J. N., & Nash, J. D. (2005). River plumes as a source of large amplitude internal waves in the coastal ocean. *Nature*, *437*, 400–403.
25. Serre, F. (1953). Contribution a l'etude des ecoulements permanents et variables dans les canaux. *Houille Blanche*, *8*(3), 374–388 (in French).
26. Stamp, A. P., & Jacka, M. (1995). Deep-water internal solitary waves. *Journal of Fluid Mechanics*, *305*, 347–371.
27. Tung, K. K., Chan, T. F., & Kubota, T. (1982). Large amplitude internal waves of permanent form. *Studies in Applied Mathematics*, *66*, 1–44.
28. Vlasenko, V., & Hutter, K. (2002). Numerical experiments on the breaking of solitary internal waves over a slope-shelf topography. *Journal of Physical Oceanography*, *32*, 1779–1793.

Internal Gravity Waves in Horizontally Inhomogeneous Ocean

Vitaly V. Bulatov and Yury V. Vladimirov

Introduction

The dynamics of wave motion in the ocean is currently of great interest because they are important in geophysics and oceanology. As a rule, theoretical analysis of these phenomena is based on the asymptotic methods, because the study of unperturbed hydrodynamic equations leads to asymptotic expansions (ansatzs, which is a German term for a type of solutions). These expansions permit solving the problems of perturbed equations, which can be used to describe the effects of nonlinearity, inhomogeneity, and non-stationary behavior of the real ocean. To obtain a detailed description of a wide range of physical phenomena related to wave dynamics of the stratified horizontally inhomogeneous unsteady ocean, it is necessary to start from the sufficiently developed mathematical models, which are usually quite complicated, nonlinear, and multi-parametric. They can be investigated completely only using efficient numerical methods. However, there are several cases, in which a preliminary qualitative concept of the phenomena under study can be obtained on the basis of simper asymptotic models and analytic methods for studying these models. These models then enter a set of "blocks" used to construct the complete pattern of wave dynamics, which permits discovering the correlation between different wave phenomena and their relationship. Sometimes, despite the seeming simplicity of the model assumptions, a successive choice of the solution form allows one to obtain physically interesting results [1, 2, 5, 14].

The propagation of internal gravity waves (IGW) in the ocean is strongly affected by the horizontal inhomogeneity and unsteady behavior of the basic

V. V. Bulatov (✉) · Y. V. Vladimirov
Institute for Problems in Mechanics, Russian Academy of Sciences, Moscow, Russia
e-mail: internalwave@mail.ru

Y. V. Vladimirov
e-mail: vladimyura@yandex.ru

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_10

hydrophysical parameters. In this contribution, we generalize a method of geometrical optics, i.e., the space-time ray method, which permits solving the problem of mathematical modeling of IGW dynamics in the horizontally inhomogeneous and vertically stratified ocean. The ray representations agree well with the intuitive and empirical concepts of IGW propagation in the real ocean. This method is sufficiently universal, and in many cases, this is the only possible method for approximate calculations of wave fields in the ocean. The most typical horizontal inhomogeneities of the real ocean are the variations in the bottom topography of the ocean, horizontal inhomogeneities of the density field, and unsteady ocean currents. An exact analytic solution can be obtained, for example, using the method of separation of variables only if the density distribution and the bottom topography can be described by sufficiently simple model functions. If the bottom topography and the ocean stratification are arbitrary, then one can construct only the asymptotic representations of the solution or solve the problem numerically. But the numerical solution does not permit obtaining and analyzing the qualitative characteristics of the wave field at large distances, which is necessary, for example, when solving the IGW detection problem by remote methods including, for example, radar imaging [8, 10, 12, 13, 15].

The mathematical modeling of IGW wave dynamics in the horizontally inhomogeneous and vertically stratified ocean is possible on the basis of a modified version of the space-time ray method (a method of geometrical optics). The specific form of asymptotic representations can be determined by solving the problems, which describe the IGW dynamics in the vertically stratified, horizontally homogeneous, and steady-state ocean. As a rule, when studying the evolution of IGW packets in the ocean with slowly varying and unsteady parameters, it is assumed that this wave packet is locally harmonic. In contrast to the majority of works, in which this problem has been studied, the proposed modified method of geometrical optics allows one to describe the structure of wave packets near singular surfaces such as caustics and wave fronts [3, 4, 6, 7].

The term “geometrical optics” has different meanings in the scientific literature. The geometrical optics understood in the narrow (or ray) sense deals only with the methods for constructing images by using the rays, while the geometrical optics understood in the wider (or wave) sense is a method for obtaining approximate descriptions of wave fields. In the wave interpretation, which is used in this paper, the rays, as a rule, form only the geometric skeleton, on which the wave filed is “sewn on”. According to the two previous interpretations of the geometrical optics, two periods in its development exist. The first ray period was ideologically completed by Hamilton’s fundamental works, which significantly influenced the development of the classical mechanics. The construction of rays underlies the instrumental optics, which is mainly oriented to design various optical devices. The contemporary wave period originates from the Debye’s works, which decisively influenced the formation of ray concepts in the wave theory [5].

The asymptotic representation of the solutions of wave packet propagation in the ocean with horizontally inhomogeneous density and numerical computations at the typical oceanic parameters testify that the horizontal inhomogeneity significantly

affects the real IGW dynamics in the ocean. All results of wave dynamics modeling presented in this contribution can be used for arbitrary density distributions and other parameters of the stratified ocean. It is necessary to consider them in the context of consistency with the available data of IGW full-scale measurements in the ocean. Such methods for analyzing the wave fields are important not only because they are illustrative, universal, and efficient in various problems, but also because they can serve as a semi-empirical basis for the other approximate methods in the theory of wave packet propagation in the ocean.

The waves in media with slowly varying parameters have been studied in many publications, while the amount of works dealing with the problem of studying IGW in the media with variable parameters is quite rare (mainly because of significant mathematical difficulties encountered in these problems). In the first section of this paper, we present the basics of the space-time ray method (a method of geometrical optics) with regard to the special characteristics of IGW, which permits studying the wave dynamics in the horizontally inhomogeneous and vertically stratified ocean. In the second section, we discuss the problems of IGW propagation in the stratified ocean of variable depth.

IGW Fields in the Horizontally Inhomogeneous Ocean

Our analysis starts from a linear system of hydrodynamic equations [8, 10, 13, 15]

$$\begin{gathered}\rho_0 \frac{\partial u_1}{\partial t} = -\frac{\partial p}{\partial x}, \quad \rho_0 \frac{\partial u_2}{\partial t} = -\frac{\partial p}{\partial y}, \quad \rho_0 \frac{\partial w}{\partial t} = -\frac{\partial p}{\partial z} + g\rho, \\ \frac{\partial u_1}{\partial x} + \frac{\partial u_2}{\partial y} + \frac{\partial w}{\partial z} = 0, \quad \frac{\partial \rho}{\partial t} + u_1 \frac{\partial \rho_0}{\partial x} + u_2 \frac{\partial \rho_0}{\partial y} + w \frac{\partial \rho_0}{\partial z} = 0.\end{gathered} \tag{1.1}$$

Here (u_1, u_2, w) are components of the IGW velocity vector; p and ρ are perturbations of the pressure and density; g is the acceleration of gravity (the z axis is directed downwards). Using the Boussinesq approximation, which means that the unperturbed density $\rho_0(z, x, y)$ in the first three equations in system (1.1) is assumed to be constant, we reduce system (1.1) to the form:

$$\begin{gathered}\frac{\partial^4 w}{\partial z^2 \partial t^2} + \Delta \frac{\partial^2 w}{\partial t^2} + \frac{g}{\rho_0} \Delta \left(u_1 \frac{\partial \rho_0}{\partial x} + u_2 \frac{\partial \rho_0}{\partial y} + w \frac{\partial \rho_0}{\partial z} \right) = 0, \\ \frac{\partial}{\partial t} \left(\Delta u_1 + \frac{\partial^2 w}{\partial z \partial x} \right) = 0, \quad \frac{\partial}{\partial t} \left(\Delta u_2 + \frac{\partial^2 w}{\partial z \partial y} \right) = 0, \quad \Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.\end{gathered} \tag{1.2}$$

We use the "rigid lid" condition at the surface and zero velocity at the bottom: $W = 0$, $(z = 0, -H)$, where H is the ocean depth as the boundary conditions. We assume that, in the media with horizontally inhomogeneous density field, the steady-state flows due to this field can be neglected. Indeed, it follows from the hydrodynamic equations that if the unperturbed density is a function of horizontal coordinates, then the existence of the steady-state density distribution $\rho_0(z, x, y)$ implies the existence of steady-state flows. These flows are rather slow, and they can be neglected in the first approximation. Therefore, it is usually assumed that

$\rho_0(z,x,y)$ is the background density field formed under the action of mass forces and non-adiabatic sources, and this field is given a priori, for example, by experimental data [2, 5].

Now we consider harmonic waves $(u_1, u_2, w) = \exp(i\omega t)(U_1, U_2, W)$. System (2) cannot be solved by the method of separation of variables, and therefore it is necessary to use asymptotic methods. The scales of horizontal variations in the ocean parameters can be greater than the scales of vertical variability [8, 10, 13, 15]. Further we introduce the dimensionless variables: $x^* = x/L$, $y^* = y/L$, $z^* = z/h$, where L is the characteristic scale of horizontal variations of density ρ_0 and h is the characteristic scale of vertical variations in ρ_0 (for example, the width of the thermocline). In the dimensionless coordinates, system (1.2) becomes (hereinafter, the asterisk in the indices is omitted)

$$\begin{gathered} -\omega^2\left(\frac{\partial^2 W}{\partial z^2}+\varepsilon^2\,\Delta W\right)+\varepsilon^2\,\frac{g_1}{\rho_0}\left(\varepsilon\,U_1\,\frac{\partial\rho_0}{\partial x}+\varepsilon\,U_2\,\frac{\partial\rho_0}{\partial y}+W\,\frac{\partial\rho_0}{\partial z}\right)=0,\\ \varepsilon\,\Delta U_1+\frac{\partial^2 W}{\partial z\partial x}=0,\quad \varepsilon\,\Delta U_2+\frac{\partial^2 W}{\partial z\partial y}=0,\quad \varepsilon=\frac{h}{L}<<1,\quad g_1=\frac{g}{h}. \end{gathered} \tag{1.3}$$

We seek for the asymptotic solution of (1.3) in the form typical for the method of geometrical optics [7].

$$\begin{gathered} \mathbf{V}(z,x,y)=\sum_{m=0}^{\infty}(i\varepsilon)^m\mathbf{V}_m(z,x,y)\exp(S(x,y)/i\varepsilon),\\ \mathbf{V}(z,x,y)=(U_1(z,x,y),U_2(z,x,y),W(z,x,y)), \end{gathered}$$

where function $S(x,y)$ and vector function $\mathbf{V}_m$, $m = 0, 1\ldots$, are sought. As a rule, below, we determine only the leading term of this asymptotic expansion for the vertical velocity component $W_0(z,x,y)$. We obtain the following from the two last equations in (1.3)

$$U_{10}=-\frac{i\partial S/\partial x}{|\nabla S|^2}\frac{\partial W_0}{\partial z},\quad U_{20}=-\frac{i\partial S/\partial y}{|\nabla S|^2}\frac{\partial W_0}{\partial z},\quad |\nabla S|=\left(\frac{\partial S}{\partial x}\right)^2+\left(\frac{\partial S}{\partial y}\right)^2.$$

Equating the terms of order O(1), we obtain the equation for function $W_0(z,x,y)$. This equation is written as

$$\begin{gathered} \frac{\partial^2 W_0(z,x,y)}{\partial z^2}+|\nabla S|^2\left(\frac{N^2(z,x,y)}{\omega^2}-1\right)W_0(z,x,y)=0,\\ W_0(0,x,y)=W_0(-H,x,y)=0, \end{gathered} \tag{1.4}$$

where $N^2(z,x,y)=\frac{g_1}{\rho_0}\frac{\partial\rho_0}{\partial z}$ is the Brunt–Väisälä frequency depending on the vertical and horizontal coordinates. It is well known that the basic boundary-value vertical spectral problem for internal waves (1.4) has countably many eigenfunctions W_{0n} and eigenvalues $K_n(x,y,\omega)\equiv|\nabla S_n|$. Functions $W_{0n}(z,x,y)$ and $K_n(x,y,\omega)$ are assumed to be known; index n is omitted because we assume that all calculations

are carried out for a separate wave mode. We use the eikonal equation $(\partial S/\partial x)^2 + (\partial S/\partial y)^2 = K^2(x,y)$ to determine function $S(x,y)$. In the plane case, the initial conditions for eikonal S are posed on line L: $x_0(\alpha), y_0(\alpha)$, $S(x,y)|_L = S_0(\alpha)$. To solve the eikonal equation, we construct the rays, i.e., the characteristics of this equation, which have the following form

$$\frac{dx}{d\sigma} = \frac{p}{K(x,y)}, \quad \frac{dp}{d\sigma} = \frac{\partial K(x,y)}{\partial x}, \quad \frac{dy}{d\sigma} = \frac{q}{K(x,y)}, \quad \frac{dq}{d\sigma} = \frac{\partial K(x,y)}{\partial y}, \tag{1.5}$$

where $p = \partial S/\partial x$, $q = \partial S/\partial y$, $d\sigma$ is the ray length element. The initial conditions of p_0 and q_0 for solution (1.5) are determined by solving the following system

$$p_0 \frac{\partial x_0}{\partial \alpha} + q_0 \frac{\partial y_0}{\partial \alpha} = \frac{\partial S_0}{\partial \alpha}, \quad p_0^2 + q_0^2 = K^2(x_0(\alpha), y_0(\alpha))$$

whose solution and the initial conditions $x_0(\alpha), y_0(\alpha), p_0(\alpha), q_0(\alpha)$ determine the ray $x = x(\sigma,\alpha)$, $y = y(\sigma,\alpha)$. After the rays are constructed, eikonal S can be determined by integrating along the ray: $S = S_0(\alpha) + \int_0^\sigma K(x(\sigma,\alpha), y(\sigma,\alpha))d\sigma$. Eigenfunction $W_0(z,x,y)$ is calculated up to multiplication by arbitrary function $A_0(x,y)$: $W_0(z,x,y) = A_0(x,y) f_0(z,x,y)$, where $f_0(z,x,y)$ is the solution of the basic vertical spectral problem with normalization $\int_0^H (N^2(z,x,y) - \omega^2) f_0^2(z,x,y)dz = 1$. Then, after rather cumbersome analytic calculations, we obtain the conservation law along the eikonal characteristics: $\frac{d}{d\sigma}\left(\ln \frac{A_0^2(x,y)I(x,y)}{K^2(x,y)}\right) = 0$, where $I(x,y)$ is the geometric divergence of the rays (characteristics). We note that the wave energy flux is proportional to $A_0^2 K^{-1} R$, where R is the width of an elementary ray tube; therefore, the quantity equal to the wave energy divided by the modulus of the wave vector is preserved in this case.

The long-range IGW fields in the real ocean are, as a rule, non-harmonic wave packets. Indeed, at a far distance form perturbation sources, the complete wave field is a sum of separate wave modes whose asymptotics, depending on the stratification, depth, and other parameters of the ocean, can be expressed in terms of the Airy function or the Fresnel integrals. Therefore, to study the problem of wave packet evolution in a horizontally smoothly inhomogeneous and unsteady stratified medium, it is necessary to use another ansatz [2, 5, 7].

We introduce slow variables $x^* = \varepsilon x$, $y^* = \varepsilon y$, $t^* = \varepsilon t$ (since z is not assumed to be a slow variable, we omit the asterisk in the index), where $\varepsilon = \lambda/L \ll 1$ is a small parameter characterizing the smoothness of the medium variations along the horizontal line (λ is the characteristic wave length, and L is the scale of horizontal inhomogeneity). Then system (1.2) for determining the velocity components (U_1, U_2, W) in these slow variables becomes

$$\frac{\partial^4 W}{\partial z^2 \partial t^2} + \varepsilon^2 \frac{\partial^2 W}{\partial t^2} + \frac{g}{\rho_0}\Delta\left(\varepsilon U_1 \frac{\partial \rho_0}{\partial x} + \varepsilon U_2 \frac{\partial \rho_0}{\partial y} + W \frac{\partial \rho_0}{\partial z}\right) = 0,$$
$$\varepsilon \Delta U_1 + \frac{\partial^2 W}{\partial z \partial x} = 0, \quad \varepsilon \Delta U_2 + \frac{\partial^2 W}{\partial z \partial y} = 0. \tag{1.6}$$

Further we consider the superposition of harmonic waves (in slow variables x, y, t) $W = \int \omega \sum_{m=0}^{\infty} (i\varepsilon)^m W_m(\omega, z, x, y) \exp\left(\frac{i}{\varepsilon}[\omega t - S_m(\omega, x, y)]\right) d\omega$, where functions $S_m(\omega, x, y)$ are assumed to be odd with respect to ω and $\min_{\omega} \partial S/\partial\omega$ is attained at $\omega = 0$ (for all x and y). We substitute this representation into (1.6) and see that function $W_m(\omega, z, x, y)$ for $\omega = 0$ has a pole of order m. Therefore, the model integrals, or phase functions $R_m(\sigma)$, for some terms of the asymptotic series are expressions $R_m(\sigma) = \frac{1}{2\pi}\int_{-\infty}^{\infty} (i/\omega)^{m-1} \exp(i(\omega^3/3 - \sigma\omega)) d\omega$, where the contour of integration bypasses point $\omega = 0$ from above, which ensures the exponential decay of functions $R_m(\sigma)$ for $\sigma \gg 1$. Functions $R_m(\sigma)$ have the following property $\frac{dR_m(\sigma)}{d\sigma} = R_{m-1}(\sigma)$, where $R_0(\sigma) = Ai'(\sigma)$, $R_1(\sigma) = Ai(\sigma)$, $R_2(\sigma) = \int_{-\infty}^{\sigma} Ai(u) du$, etc. Obviously, starting from the corresponding properties of the Airy integrals, we can conclude that functions $R_m(\sigma)$ are related as $R_{-1}(\sigma) + \sigma R_1(\sigma) = 0$, $R_{-3}(\sigma) + 2R_0(\sigma) - \sigma^2 R_1(\sigma) = 0$. For the model integrals $R_m(\sigma)$ describing the long-range IGW fields in the deep ocean, one can use the following expressions $R_0(\sigma) = \mathrm{Re}\int_0^{\infty} \exp(-it\sigma - it^2/2)\, dt \equiv \mathrm{Re}\Phi(\sigma)$; in this case, functions $R_m(\sigma)$ satisfy the recurrence relations $R_{-3}(\sigma) - 2iR_{-1}(\sigma) - i\sigma R_{-2}(\sigma) = 0$ and $R_{-1}(\sigma) + i\sigma R_0(\sigma) = 0$ [2, 5, 7]. It follows from the above and the structure of the first term of the uniform asymptotics (Airy or Fresnel wave) in a stratified and horizontally homogeneous medium that the solution of system (1.6) can be sought in the following form (index n is omitted for a separate wave mode)

$$W = \varepsilon^0 W_0(z, x, y, t) R_0(\sigma) + \varepsilon^a W_1(z, x, y, t) R_1(\sigma) + \varepsilon^{2a} W_2(z, x, y, t) R_2(\sigma) + \cdots,$$
$$\mathbf{U} = \varepsilon^{1-a} \mathbf{U}_0(z, x, y, t) R_1(\sigma) + \varepsilon \mathbf{U}_1(z, x, y, t) R_2(\sigma) + \varepsilon^{1+a} \mathbf{U}_2(z, x, y, t) R_3(\sigma) + \cdots,$$

where $\mathbf{U}$ is the vector of IGW horizontal velocity and the phase function argument $\sigma = (S(x, y, t)/a)^a \varepsilon^{-a}$ is assumed to be of the order of unity. This expansion agrees well with the general approach of the method of geometrical optics and the space-time ray method. Its generalization is used to study the dynamics of IGW fields in the horizontally inhomogeneous stratified ocean.

We also note that this structure of the solution implies that, in a horizontally inhomogeneous medium, the solution depends on both the "fast" (vertical coordinate) and "slow" (horizontal coordinates) variables. As a rule, the solution is sought in "slow" variables, and the structure elements depending on "fast" variables are obtained as integrals of some functions slowly varying along the space-time rays. This choice of the solution permits describing the uniform asymptotics of IGW fields propagating in the stratified ocean with slowly varying parameters, which is true both near and far from the wave fronts of a separate wave mode. If it is necessary to describe the behavior of the field only near the wave front, then one

can use one of the methods of the geometrical optics, i.e., the "traveling wave" method, and the weakly dispersion approximation in the form of the corresponding local asymptotics to seek the representation of the phase function argument σ in the form $\sigma = \alpha(t,x,y)(S(t,x,y) - \varepsilon t)\varepsilon^{-a}$; here function $S(t,x,y)$ describes the wave front position. It is found by solving the eikonal equation $\nabla^2 S = c^{-2}(x,y,t)$, where $c(t,x,y)$ is the maximal IGW group velocity of the corresponding wave mode, i.e., the first term in the expansion of the dispersion curve at zero. Function $\alpha(t,x,y)$ (the second term of the dispersion curve expansion) describes the space-time evolution of the pulse width of non-harmonic Airy or Fresnel waves and can be found from some conservation laws along the eikonal equation characteristics whose specific form is determined by the physical conditions of the problems under study.

Further we compare the analytic results with the results of the analysis of measurements of IGW variability in a real medium with horizontally varying characteristics, namely, in the Northwest Pacific, according to the data recorded by moorings in the "Megapolygon" experiment in the Northwest Pacific. The measurements of the currents and the temperature recorded by the "Megapolygon" moorings allowed us to study the variability of tidal internal waves over the area of 460 × 520 km. The length of the tidal internal wave was calculated by integration of the basic IGW spectral equation with the real depth distribution of the Vaisala-Brunt frequency and with zero boundary conditions at the ocean surface and the ocean floor taking into account the Earth's rotation. The wave length of the first mode in the "Megapolygon" area is equal approximately to 130 km, the wave length near the Emperor Ridge is greater (167 km), and it is equal to 156 km at a distance of 2000 km to the east. The wave propagation direction is also very stable and varies from 240° to 300°, which corresponds to the actual wave propagation to the west and northwest from the Emperor Ridge. Some diffraction of tidal internal waves was observed in the "Megapolygon" study site, i.e., the direction of wave propagation varied from the northwest in the southeast of the site and to the west in its northwest part [5].

Let us consider the amplitude variations of the internal tide in the course of its propagation to the west and to the east from the Emperor Ridge. The IGW amplitudes were calculated from the deviations of the temperature values measured on moorings; then, the values were divided by the average vertical gradient of temperature. Figure 1 illustrates the variations in the tidal internal wave amplitude versus distance. The calculations show that the IGW amplitude decreases

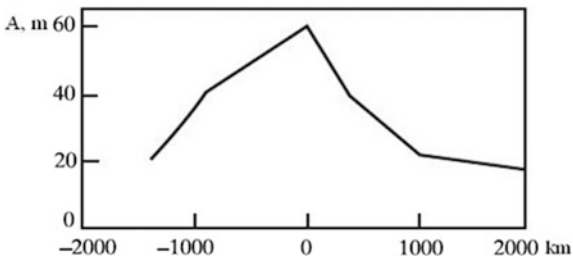


Fig. 1 The tidal IGW amplitude A versus the distance to the Emperor Seamounts

approximately by 10% at the distance equal to the length of the tidal internal wave (130–150 km) [11].

We can also estimate the influence of different factors, including the horizontal inhomogeneity of density, on the IGW decay. In the framework of the theory discussed above, we consider the evolution of IGW frequency ω corresponding to the semidiurnal period T = 12 h, which also admits slow variations in the stratification along the wave propagation path. The real geometry of the experiment allows us to assume that the problem under study is two-dimensional, which means that the stratification depends only on two variables: depth z and distance x along the wave propagation path.

Now we consider the case of constant depth H and stratification N linearly depending only on x: $N(x) = N_1 + (N_2 - N_1)\,x/L$, where L is the distance between the two observation points, $x = x_1 = 0$ is the initial point, $x = x_1 = L$ is the end point, and $N_{1,2} = N(x_{1,2})$. We consider only the first mode $\eta_1(z,x)$ of the amplitude of the vertical displacement of particles and omit its index. We seek the amplitude $\eta(z,x)$ in the form $\eta(z,x) = A(x)\,f(z,x)$, where $f(z,x)$ is the normalized eigenfunction of the standard boundary-value problem for the equation of internal waves with the normalization $\int_0^H (N^2(x) - \omega^2) f^2(z,x)\,dz = 1$, which has the form $f(z,x) = \sqrt{\frac{2}{H(N^2(x) - \omega^2)}}\sin(\pi z/H)$. Amplitude A(x) depending only on x is determined from the conservation law: $\frac{A^2(x_1)}{k^2(x_1)}da(x_1) = \frac{A^2(x_2)}{k^2(x_2)}da(x_2)$, where $k(x)$ is the absolute value of the horizontal wave vector, and $da(x)$ is the width of an elementary wave tube. Since the problem is two-dimensional, the width of the ray tube does not vary along the ray and the conservation law is simpler: $A(x)/k(x) = const$. Since we consider small values of ω, the velocity of wave propagation is close to the maximum group velocity $c(x) = N(x)\,H/\pi$; hence, the wave number is equal to $k(x) = \omega\pi/N(x)\,H$ and the corresponding wave length is equal to $\lambda(x) = 2N(x)\,H/\omega$. Then, under the assumption that the observation points are at the same depth, it follows from the conservation law $(A_{1,2} = A(x_{1,2}))$ that $A_1N_1 = A_2N_2$ or $A_2 = A_1\lambda_1/\lambda_2$. Then the total amplitude attains the following values $W_{1,2} = A_{1,2}\sqrt{\frac{2}{H(N_{1,2}^2 - \omega^2)}}$, which implies $W_2 = W_1\frac{N_1}{N_2}\sqrt{\frac{(N_1^2 - \omega^2)}{(N_2^2 - \omega^2)}}$ or $W_2 = W_1\lambda_1^2/\lambda_2^2$, because $\omega \ll N$, i.e., the amplitude of the internal gravity wave is inversely proportional to the squared wave length. The wave travel time τ along the horizontal ray is determined from the equation of characteristics $\frac{dx}{dt} = c(x)$, where $c(x) = (N_1 + ax)\,H/\pi$ and $a = (N_2 - N_1)/L$. Integrating this equation, we obtain the wave travel time $\tau = \frac{\pi}{aH}\ln\left(\frac{N_2}{N_1}\right) = \frac{TL}{(\lambda_2 - \lambda_1)}\ln\left(\frac{\lambda_2}{\lambda_1}\right)$. The available data of full-scale tests give the following values of the basic parameters of the problem: $\lambda_1 = 167$ km, $\lambda_2 = 156$ km, L = 2000 km. The wave attenuation coefficient without the wave length variations taken into account, which describes the amplitude decrease versus wave length denoted by β, gives the value of β: $\beta = 0.2^{167/2000} = 0.874$ with regard to relation $W_2 / W_2 = 0.2 \equiv \beta^{t/T} = \beta^{L/\lambda}$ derived from the observation results. The attenuation with regard to the wave length variations along the ray, $W_2/W_1 = \beta^{\tau/T}$,

with the theoretically calculated time of the wave travel time τ gives the following value $\beta = 0.878$. Thus, these estimates allow us to conclude that the influence of the density field inhomogeneities, which is taken into account in the above-described method for asymptotic representation of the wave fields, is one of the factors determining the scales of the space attenuation of IGW fields observed in field measurements.

Fields of IGW in the Ocean of Variable Depth

We consider one of the problems of IGW propagating in the stratified ocean of variable depth. In the framework of the linear theory, we study the non-viscous incompressible inhomogeneous medium with unperturbed density $\rho_0(z)$, which is bounded by surface $z = 0$ and ocean floor $z = \gamma y$ (the z axis is directed upwards, γ is the ocean floor slope). At point $x = x_0, y = y_0, z = z_0$ at the slope, there is a point mass source of power Q depending on time as $\exp(-i\omega t)$. The system of hydrodynamic equations for small perturbations of density ρ^*, pressure p^*, and velocity components (u_1, u_2, w) is written as [2–6]

$$\begin{gathered}\rho_0 \frac{\partial u_1}{\partial t} = -\frac{\partial p^*}{\partial x}, \rho_0 \frac{\partial u_2}{\partial t} = -\frac{\partial p^*}{\partial y}, \rho_0 \frac{\partial w}{\partial t} = -\frac{\partial p^*}{\partial z} + g\rho^* \\ \frac{\partial u_1}{\partial x} + \frac{\partial u_2}{\partial y} + \frac{\partial w}{\partial z} = Q \exp(-i\omega t)\delta(x - x_0)\delta(y - y_0)\delta(z - z_0), \\ \frac{\partial \rho^*}{\partial t} + w\frac{\partial \rho_0}{\partial z} = 0,\end{gathered} \tag{2.1}$$

where g is the acceleration of gravity. As the boundary conditions we pose the "rigid lid" condition at the ocean surface and zero mass flux at the ocean bottom

$$w = 0 \text{ at } z = 0,\ w + u_2\gamma = 0 \text{ at } z = -\gamma y \tag{2.2}$$

Under the assumption that the time-dependence of all solutions is harmonic $(p^*, \rho^*, u_1, u_2, w) = \exp(-i\omega t)(p, \rho, U_1, U_2, W)$, we obtain the following system of equations with boundary conditions (2.2)

$$\begin{gathered}i\omega\rho_0 U_1 = \frac{\partial p}{\partial x}, \quad i\omega\rho_0 U_2 = \frac{\partial p}{\partial y}, \quad i\omega\rho_0 W = -\frac{c^2 \partial p}{\partial z}, \\ \frac{\partial U_1}{\partial x} + \frac{\partial U_2}{\partial y} + \frac{\partial W}{\partial z} = Q\,\delta(x - x_0)\delta(y - y_0)\delta(z - z_0), \quad i\omega\rho = \frac{W\partial\rho_0}{\partial z},\end{gathered} \tag{2.3}$$

where $c^2 = \omega^2/(N^2 - \omega^2)$ and $N^2(z) = -\frac{g}{\rho_0}\frac{\partial \rho_0}{\partial z}$ is the Brunt-Väisälä frequency which is assumed constant: $N(z) = N = \text{const}$. These assumptions can be used to study the IGW fields in many regions of the World Ocean [12]. In the Boussinesq approximation, system (2.3) reduces to a single equation, for example, for pressure perturbations p with the corresponding boundary conditions

$$\frac{\partial^2 p}{\partial z^2} - \frac{1}{c^2}\left(\frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial x^2}\right) = -i\omega Q\rho_0\delta(x-x_0)\delta(y-y_0)\delta(z-z_0)/c^2, \tag{2.4}$$

$$\frac{\partial p}{\partial z} = 0 \text{ at } z = 0, \ \frac{\partial p}{\partial z} - \frac{\gamma}{c^2}\frac{\partial p}{\partial y} = 0 \text{ at } z = -\gamma y. \tag{2.5}$$

Since the variations in $\rho_0(z)$ are relatively small in the ocean, the value of ρ_0 in the right-hand side of (2.4) is understood, for example, as the value of the sea water density at the surface, i.e., we set $\rho_0 = \rho_0(0) = \text{const}$. Solution $p(x, y, z)$ must tend to zero as $\sqrt{x^2 + y^2 + z^2} \to \infty$. After function $p(x, y, z)$ is determined, velocity components (U_1, U_2, W) can be found from the first three equations of system (2.3), and density ρ is determined from the fifth equation in this system.

We change the variables as

$$y = r\text{ch}\varphi, \ z = -cr\text{sh}\varphi, \ r = \sqrt{y^2 - z^2/c^2}, \varphi = \frac{1}{2}\ln\frac{cy - z}{cy + z} \tag{2.6}$$

We perform the Fourier transform with respect to variable x (without loss of generality, we can set $x_0 = 0$). Since the absolute value of the Jacobian of transition from the coordinates (y, z) to (r, φ) is equal to cr, problem (2.4), (2.5) implies the following plane boundary-value problem for the Fourier transform $P(r, \varphi, l)$ of function $p(r, \varphi, x)$

$$\frac{\partial^2 P}{\partial r^2} + \frac{\partial P}{r\partial r} - \frac{1}{r^2}\frac{\partial^2 P}{\partial \varphi^2} - l^2 P = \frac{q}{r}\delta(r - r_0)\delta(\varphi - \varphi_0), \tag{2.7}$$

$$\frac{\partial P}{\partial \varphi} = 0 \quad \text{at} \quad \varphi = 0, \quad \frac{\partial P}{\partial \varphi} = 0 \quad \text{at} \quad \varphi = \varphi_r, \tag{2.8}$$

$$r_0 = \sqrt{y_0^2 - z_0^2/c^2}, \quad \varphi_0 = \frac{1}{2}\ln\frac{c\,y_0 - z_0}{c\,y_0 + z_0}, \quad \varphi_r = \frac{1}{2}\ln\frac{c + \gamma}{c - \gamma}, \quad q = i\omega Q\rho_0/c. \tag{2.9}$$

The solution of three-dimensional boundary-value problem (2.4), (2.5) with respect to variables (r, φ, x) is obtained from the solution of the plane problem (2.7), (2.8) by using the inverse Fourier transform

$$p(r, \varphi, x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} P(r, \varphi, l)\ \exp(ilx)dl. \tag{2.10}$$

We assume that the ocean floor slope γ is less than c or, in the trigonometric terminology, we assume that the ocean bottom slope is subcritical (the critical slope is $\gamma = c$) [3–6].

The homogeneous Eq. (2.7) with zero right part has real solutions $P(r, \varphi, l) = K_{i\mu}(lr)\cos(\mu\varphi)$ decreasing at infinity, where μ is an arbitrary real

number and $K_{i\mu}(lr)$ is the Macdonald function with imaginary index satisfying the modified parametric Bessel equation $LK_{i\mu}(lr)=0$, where $L=r^2\frac{\partial^2}{\partial r^2}+r\frac{\partial}{r\partial r}+(\mu^2-r^2l^2)$. We note that function $K_{i\mu}(lr)$ is real if the values of μ are real and argument lr is positive. Hence, we write the delta function $\delta(r-r_0)$ using a pair of direct and inverse Kantorovich-Lebedev transformations [5, 6]

$$F(\mu)=\int_0^{+\infty} K_{i\mu}(x)\frac{f(x)}{x}dx,\ f(x)=\frac{2}{\pi^2}\int_0^{+\infty}\mathrm{sh}(\pi\mu)K_{i\mu}(x)F(\mu)\mu d\mu\,.$$

This implies the expansion of the delta function (completeness condition) in the form

$$\delta(r-r_0)=\frac{2}{r\pi^2}\int_0^{+\infty}\mathrm{sh}(\pi\mu)K_{i\mu}(lr)\,K_{i\mu}(lr_0)\mu d\mu. \tag{2.11}$$

We seek the solution of problem (2.7) in the form

$$P(r,\varphi,l)=\frac{2q}{\pi^2}\int_0^{+\infty}\mathrm{sh}(\pi\mu)K_{i\mu}(lr)\,K_{i\mu}(lr_0)\Phi_\mu(\mu)\mu d\mu, \tag{2.12}$$

where the function of the angular variable $\Phi_\mu(\varphi)$ is still unknown. Substituting (2.11) and (2.12) into (2.7), we obtain the boundary-value problem for determining this function

$$\begin{gathered}\frac{d^2\Phi_\mu(\varphi)}{d\varphi^2}+\mu^2\Phi_\mu(\varphi)=-\delta(r-r_0),\\ \frac{d\Phi_\mu(0)}{d\varphi}=\frac{d\Phi_\mu(\varphi_r)}{d\varphi}=0.\end{gathered} \tag{2.13}$$

It follows from (2.13) that $\Phi_\mu(\varphi)$ is the angular Green function of the form

$$\Phi_\mu(\varphi)=-\frac{1}{\mu^2\varphi_r}-\frac{2}{\varphi_r}\sum_{n=1}^{\infty}\frac{\cos(\varphi\mu_n)\,\cos(\varphi_0\mu_n)}{\mu^2-\mu_n^2},\quad \mu_n=2\pi n/\ln\left(\frac{c+\gamma}{c-\gamma}\right),\quad n\geq 1. \tag{2.14}$$

In the expression for $P(r,\varphi,l)$ in (2.12), we consider a single wave mode $(n\geq 1)$

$$P_n(r,\varphi,l)=-\frac{4q\,\cos(\varphi\mu_n)\,\cos(\varphi_0\mu_n)}{\varphi_r\pi^2}\int_0^{+\infty}\frac{\mathrm{sh}(\pi\mu)\,K_{i\mu}(lr)K_{i\mu}(lr_0)\mu d\mu}{\mu^2-\mu_n^2}. \tag{2.15}$$

Here, the integral is understood in the sense of the principal value. Formula (2.15) can also be used for $n=0$ if we set $\mu_0=0$ and decrease the coefficient of the integral by a factor of two. First, we consider the case $r>r_0$. To deform the contour of integration over μ in expression (15), we use formula $K_\nu(t)=\pi(I_{-\nu}(t)-I_\nu(t))/(2\sin(\pi\mu))$ which, in this case with $\nu=i\mu$ for function $K_{i\mu}(lr_0)$, becomes

$$K_{i\mu}(lr_0)=-\pi\mathrm{Im}\,(I_{i\mu}(lr_0))/\mathrm{sh}(\pi\mu), \tag{2.16}$$

because functions $I_{i\mu}(x)$ and $I_{i\mu}(-x)$ are complex conjugate function. The intergrand in (2.15) is even with respect to μ; hence, we can use (2.16) to obtain

$$P_n(r,\varphi,l)=\frac{2q\,\cos(\varphi\mu_n)\,\cos(\varphi_0\mu_n)}{\pi\varphi_r}\mathrm{Im}\int\limits_{-\infty}^{+\infty}\frac{K_{i\mu}(lr)I_{i\mu}(lr_0)\mu d\mu}{\mu^2-\mu_n^2}. \tag{2.17}$$

Now the contour of integration in (2.17) can be closed in the lower half-plane. To verify this, we use the asymptotic expansions of $K_{i\mu}(x)$ and $I_{i\mu}(x)$ for $\mu=-i\nu$ as $\nu\to\infty$: $K_\nu(lr)\approx\sqrt{\pi/2\nu}(2\nu/elr)^\nu$, $I_\nu(lr_0)\approx\sqrt{\pi/2\nu}(2\nu/er_0l)^\nu/2\sqrt{2}$. Then we can obtain $K_\nu(lr)I_\nu(lr_0)\approx\pi\,\exp\,(-\nu(\ln r-\ln r_0))/4\nu\sqrt{2}$. This implies that the integrand is exponentially small in the lower half-plane for $r>r_0$. Then, taking into account the residues at points $\mu=\pm\mu_n$, we have

$$P_n(r,\varphi,l)=-\frac{2q\,\cos(\varphi\mu_n)\,\cos(\varphi_0\mu_n)}{\varphi_r}\,\mathrm{Re}(K_{i\mu_n}(lr)I_{i\mu_n}(lr_0)). \tag{2.18}$$

In the case $r<r_0$, we represent function $K_{i\mu}(lr)$ in the form (2.16) and closing the contour of integration in the lower half-plane we obtain expression (2.18), where it is necessary to interchange r and r_0. These expressions can be written as a single expression if we introduce notations $r_-=\min(r,r_0)$, $r_+=\max(r,r_0)$

$$P_n(r,\varphi,l)=-\frac{2q\,\cos(\varphi\mu_n)\,\cos(\varphi_0\mu_n)}{\varphi_r}\,\mathrm{Re}(K_{i\mu_n}(lr_+)I_{i\mu_n}(lr_-)). \tag{2.19}$$

In the case $n=0$, we similarly have

$$P_0(r,\varphi,l)=-\frac{q}{\varphi_r}\mathrm{Re}\,(K_0(lr_+)I_0(lr_-)). \tag{2.20}$$

Now we calculate the inverse Fourier transform (2.10) for the n-th mode $(n\geq 0)$ with regard to the fact that the steady-state standing wave is an odd function of variable x; as a result, we obtain $p_n(r,\varphi,x)=\frac{1}{\pi}\int_{-\infty}^{+\infty}P_n(r,\varphi,l)\,\cos(lx)dl$. This integral can be expressed in the terms of the hypergeometric function

$$p_n(r,\varphi,x) = -\frac{q\varepsilon_n \cos(\varphi\mu_n)\cos(\varphi_0\mu_n)}{\sqrt{\pi r\, r_0\varphi_r}}\mathrm{Re}Z,$$
$$Z = \frac{\Gamma(i\mu_n+1/2)}{\Gamma(i\mu_n+1)}(\tau/2)^{i\mu_n+1/2}F\left(\frac{i\mu_n+1/2}{2},\frac{i\mu_n+3/2}{2},i\mu_n+1,\tau^2\right), \tag{2.21}$$

where $\Gamma(z)$ is the gamma function, $F(\alpha,\beta,\gamma,z)$ is the hypergeometric function, $\tau = 2r\, r_0/(r^2 + r_0^2 + x^2)$, $\varepsilon_n = 1/2$ for $n = 0$, and $\varepsilon_n = 1$ for $n \geq 1$. The complete solution is obtained as a sum of all modes: $p(r,\varphi,x) = \sum_{n=0}^{\infty} p_n(r,\varphi,x)$, where r and φ are determined from (2.6), and r_0, φ_0, φ_r are determined from (2.9). We note that small values of τ correspond to the far field distance from the perturbation source, i.e., to the large values of r and x; a separate mode $p_n(r,\varphi,x)$ can be approximated by the expansion of the hypergeometric function in a series for $0 \leq z < 1$

$$F(\alpha,\beta,\gamma,z) = 1 + \frac{\alpha\beta}{\gamma}z + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)2!}z^2 + \cdots, \tag{2.22}$$

where, $\alpha = \frac{i\mu_n+1/2}{2}$, $\beta = \frac{i\mu_n+3/2}{2}$, and $\gamma = i\mu_n + 1$. However, as the mode number n increases at fixed z, it is required to take even greater number of terms in expansion (2.22) (the number of terms is $m \approx \mu_n z$), which hampers the calculation of wave modes with large numbers. For the further summation of the series (2.22), we use the WKB asymptotics of the hypergeometric function in (2.21)

$$F(\tau^2) \approx \exp\left(-\frac{i\mu_n}{2}\left(\ln\frac{\tau^2}{4} + \ln\frac{1+\sqrt{1-\tau^2}}{1-\sqrt{1-\tau^2}}\right)\right)/\sqrt{41-\tau^2}. \tag{2.23}$$

We use the asymptotics of the gamma function in (2.21) for large values of μ_n: $\frac{\Gamma(i\mu_n+1/2)}{\Gamma(i\mu_n+1)} \approx \exp(-i\pi/4)/\sqrt{\mu_n}$. Finally, we obtain the following expression for the WKB asymptotics of a separate wave mode at large μ_n

$$p_n(r,\varphi,x) \approx -\frac{q\sqrt{\tau}\cos(\varphi\mu_n)\cos(\varphi_0\mu_n)}{\sqrt{2\mu_n\pi\, r\, r_0}\,\varphi_r}\cos\left(\frac{\mu_n}{2}\ln\frac{1+\sqrt{1-\tau^2}}{1-\sqrt{1-\tau^2}} + \pi/4\right). \tag{2.24}$$

It is interesting to note that if we formally set $\mu_n \to \infty$ in expansion (2.22), let $z \to 0$ in the WKB asymptotics (2.23) for $F(z)$, and take into account that $z\mu_n \approx O(1)$, then, in both cases, we obtain the same value equal to $\exp(-iz\mu_n/4)$. Thus, expansion (2.22) and WKB asymptotics (2.23) are mutually consistent, i.e., there is a domain of z, μ_n, where these expressions coincide. It follows from (2.24) that the amplitude of the n-th mode decreases as $((x^2+y^2)n)^{-1/2}$ for large x, y. Expanding the phase in (2.24) for small τ, we see that, for large y, the half-wave length along axis y increases as $\pi y/\mu_n$, and along axis x, as $\pi x/2\mu_n$. The numerical calculations with the real parameters of the ocean show that the exact and asymptotic solutions agree well, except for the immediate vicinity of the

perturbation source, where the argument of the hypergeometric function tends to unity, which follows from the construction of the asymptotic solution. We note that expression (2.24) formally requires that $\mu_n \to \infty$, but already for the first mode $n=1$, asymptotic formula (2.24) gives a qualitatively true description of the exact solutions. The asymptotics of the zero mode can be calculated from (2.21) by setting $\mu_n = 0$. Then, taking into account that $F\left(\frac{1}{4}, \frac{3}{4}, 1, \tau^2\right) = \frac{2}{\pi\sqrt{1+\tau}} K\left(\frac{2\tau}{1+\tau}\right)$, where $K(x) = \int_0^{\pi/2} \left(1 - (x \sin \varphi)^2\right)^{-1/2} d\varphi$ is an elliptic integral of the first kind, we obtain the following expression

$$p_0(r, \varphi, x) \approx - \frac{q\sqrt{\tau}}{\pi\sqrt{1+\tau}\sqrt{2r\, r_0}\varphi_r} K\left(\frac{2\tau}{1+\tau}\right). \tag{2.25}$$

We use the asymptotics of $K(x)$ as $x \to 1$ with the leading term $K(x) \approx \ln 4 - \ln(1-x)/2$ and finally obtain the expression for the asymptotics of the zero mode

$$p_0(r, \varphi, x) \approx - \frac{q\sqrt{\tau}}{\pi\sqrt{1+\tau}\,\sqrt{2r\, r_0}\varphi_r} \left(\ln\left(\frac{1-\tau}{1+\tau}\right)/2\right). \tag{2.26}$$

We note that the exact and asymptotic solutions coincide completely near the perturbation source. There is difference between them at far distances from the source. This is related to the fact that the asymptotics of the elliptic integral works well as the argument tends to unity. Nevertheless, in the far region, the asymptotics qualitatively true describes the exact solution with an error at most equal to a few percent. The obtained asymptotic representations of the solutions for separate wave modes, including the zero mode, permit calculating the complete wave field. The sum of asymptotics (2.24) of infinitely many wave modes ($n = 1, 2, \ldots$) is expressed in terms of semi-logarithmic function

$$Li_{1/2}(z) = \sum_{n=1}^{\infty} \frac{z^n}{\sqrt{n}}, \quad B_{\pm}^{\pm} = \exp(i\pi(\pm\varphi \pm \varphi_0 + A(\tau))/\varphi_r), \quad A(\tau) = \tfrac{1}{2} \ln \frac{1-\sqrt{1-\tau^2}}{1+\sqrt{1-\tau^2}},$$
$$\sum_{n=1}^{\infty} p_n(r, \varphi, x) = - \frac{q\sqrt{\tau}\, \exp(-i\pi/4)}{8\pi\sqrt{4}\sqrt{1-\tau^2}\sqrt{rr_0}\, \varphi_r} \left(Li_{1/2}(B_+^+) + Li_{1/2}(B_-^+) + Li_{1/2}(B_+^-) + Li_{1/2}(B_-^-)\right). \tag{2.27}$$

The complete wave field is the real part of expression (2.27) and the zero mode (2.25). The semi-logarithmic function in (2.27) becomes infinite at the points, at which the condition $\pi(\pm\varphi \pm \varphi_0 + A(\tau))/\varphi_0 = 2\pi m, m = 0, 1, 2, \ldots$ is satisfied. The locus of points (x, y, z) satisfying this condition determines a system of rays if one of the variables is fixed. On planes (y, z) and (x, z), these solutions determine a pair of ascending rays and a pair of descending rays, which are radiated from the source and then reflect from the sloping ocean floor. Figure 2 presents the shadow picture of the complete wave field (level lines) on plane (y, z) for $x = 40$ m; the other

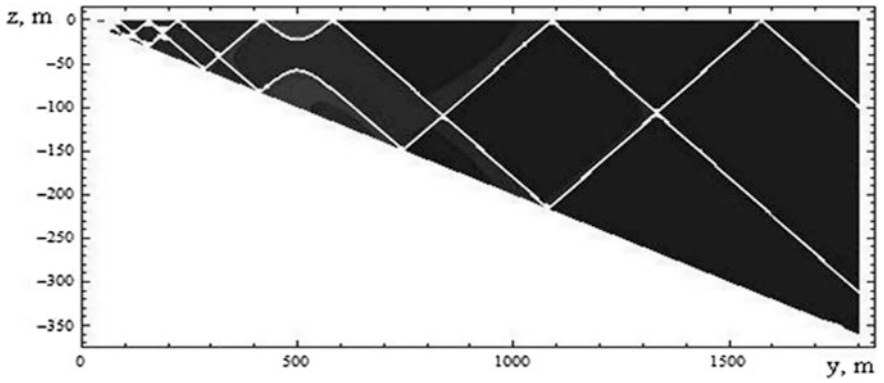


Fig. 2 Amplitude structure of IGW (pressure) in stratified ocean with non-uniform depth: analytical results

computational parameters are typical for the real ocean parameters: $N = 0.001\,\mathrm{s}^{-1}$, $\omega = 0.004\,\mathrm{s}^{-1}$, $\gamma = 0.2$, $c = 0.44$, $\rho_0 = 1000\,\mathrm{kg/m}^{-3}$, $Q = 1600\,\mathrm{m}^3/\mathrm{s}$, $y_0 = 500\,\mathrm{m}$, $z_0 = -4\,\mathrm{m}$.

These results clearly illustrate the ray structure of the constructed solutions, in particular, the set of incident and reflected rays; moreover, the cotangent of the angle between the incident ray and the vertical is approximately equal to 0.44, which agrees well with the ray theory. Indeed, according to this theory, the direction of group velocity Θ and the energy propagation direction are determined by expression $\mathrm{ctg}^2\Theta = c^2 = \omega^2/(N^2 - \omega^2)^2$ [8, 10, 13, 15]. The solutions are singular on the rays, because the model of ideal medium is used. The main contribution to the singularity is given by infinitely many short-wave modes with large numbers. In reality, to obtain the complete wave field, it is necessary to consider finitely many modes. This number is approximately determined by the Stokes characteristic scale $D = \sqrt{2\nu_0/N}$, where ν_0 is the kinematic viscosity and N is the Brunt-Väisälä frequency. Obviously, the wave modes with large numbers whose wave length is less that D do not contribute to the solution.

For comparison with the analytic results, in Fig. 3, we show the results of numerical simulation of the complete system of hydrodynamic equations, which describes the evolution of nonlinear wave perturbations over uneven ocean floor (Bay of Biscay, more than 60 wave modes were summed) [9].

The results show that the ray structure of the solution (Fig. 2) is clearly identified and, as the estimates show, the amplitude-phase structure of the wave fields is quite well described by asymptotic formulas (2.27).

Figure 4 illustrates the results of full-scale measurements of the amplitude structures of the tidal IGW in the same region of the World Ocean [9]. These full-scale data show that the wave patterns with profound ray structure can actually be observed in the real ocean, especially, when the IGW evolution over uneven ocean floor is investigated. In particular, the analytic, numerical, and full-scale data

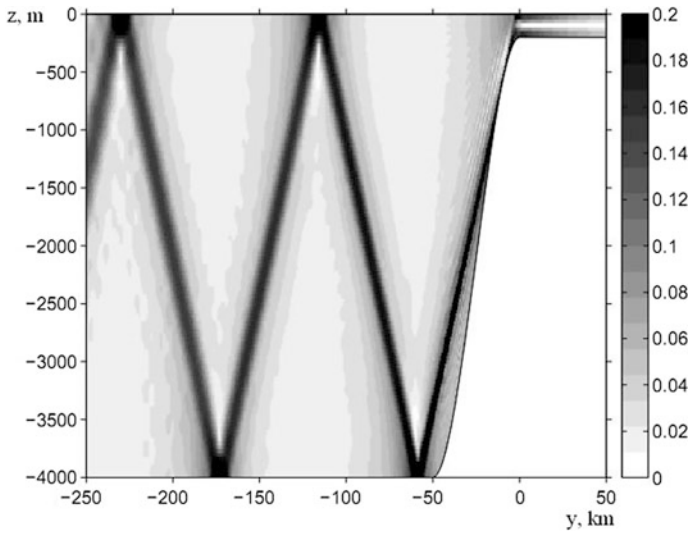


Fig. 3 Amplitude structure of IGW (velocity, m/s^{-1}) in stratified ocean with non-uniform depth: numerical simulation

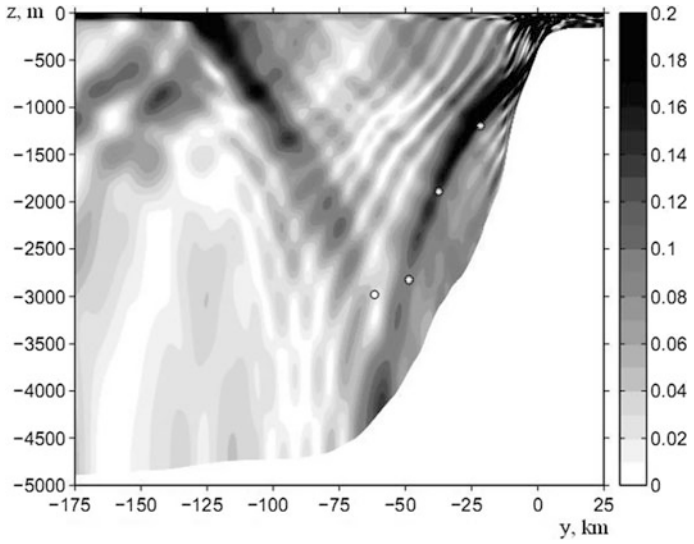


Fig. 4 Amplitude structure of IGW (velocity, m/s^{-1}) in stratified ocean with non-uniform depth: measurements results

show that the width of wave beams decreases as the shore is approached. Formally, in the linear statement, the width of the reflected IGW beam can be arbitrarily small for appropriate relations between the medium parameters (stratification, the ocean floor slope angle); hence, a significant local intensification of waves occurs near the ocean shore. It is clear that in the real ocean, the wave field energy remains finite in such spatial domains due to the action of nonlinear mechanisms of dissipation and turbulent mixing [1].

Conclusions

Thus, in the first section of the paper, a general method for calculating IGW fields in the horizontally inhomogeneous ocean is outlined, namely,

- for an arbitrary distribution of the Brunt-Väisälä frequency, the basic vertical spectral IGW problem is solved and the corresponding normalized eigenfunctions and eigenvalues are determined;
- the characteristic systems with appropriate initial conditions are solved numerically;
- after the characteristics (rays) are calculated, the eikonal (phase value) of the phase functions is determined by numerical integration along these rays;
- the geometric divergence of the ray tubes is determined, for example, by numerical differentiation of closely located characteristics;
- the IGW amplitude is calculated from the equations of the corresponding conservation laws along the rays (characteristics), in which the right parts of the relations are determined by using the locality principle, i.e., it is assumed that the ocean parameters remain horizontally unchanged over specific spatial intervals. Thus, it is assumed that the ocean is horizontally homogeneous on these space-time scales, and its density arbitrarily depends on the vertical coordinate.

The solutions obtained in the second section of the paper are exact and exhibit typical ray pattern of the IGW fields in the stratified ocean of variable depth obtained without using the mathematical methods of geometrical optics.

The universal character of the proposed asymptotic methods of modeling IGW fields in the ocean allows us to efficiently calculate the wave fields and, in addition, analyze qualitatively the solutions. This opens wide opportunities for investigating the wave fields in general, which is also important for formulating correct statements of mathematical models of wave dynamics and for obtaining express evaluations in the field measurements of internal waves.

References

1. Abdilghanie, A. M., & Diamessis, P. J. (2013). The internal gravity wave field emitted by a stably stratified turbulent wake. *Journal of Fluid Mechanics, 720,* 104–139.
2. Bulatov, V. V., & Vladimirov, Y. V. (2007). *Internal gravity waves: Theory and applications*. Moscow: Nauka Publishers.
3. Bulatov, V. V., & Vladimirov, Y. V. (2010). Far fields of internal gravity waves in stratified media of variable depth. *Russian Journal of Mathematical Physics, 17*(4), 400–412. https://doi.org/10.1134/S1061920810040035.
4. Bulatov, V. V., & Vladimirov, Y. V. (2011). The uniform asymptotic form of the internal gravity wave field generated by a source moving above a smoothly varying bottom. *Journal of Engineering Mathematics, 69*(2), 243–260. https://doi.org/10.1007/s10665-010-9388-6.
5. Bulatov, V. V., & Vladimirov, Y. V. (2012). *Wave dynamics of stratified mediums*. Moscow: Nauka Publishers.
6. Bulatov, V. V., & Vladimirov, Y. V. (2014). Exact and asymptotic solutions of the equation of internal gravity wave in a wedge-shaped region. *Journal of Applied Mathematics and Mechanics, 78*(5), 494–501. https://doi.org/10.1016/j.jappmathmech.2015.03.008.
7. Bulatov, V. V., & Vladimirov, Y. V. (2014). Asymptotical analysis of internal gravity wave dynamics in stratified medium. *Applied Mathematical Sciences, 8*(5), 217–240. https://doi.org/10.12988/ams.2014.311637.
8. Massel, S. R. (2015). *Internal gravity waves in the shallow seas*. Berlin: Springer.
9. Mauge, R., & Gerkema, T. (2008). Generation of weakly nonlinear non-hydrostatic internal tides over large topography: A multi-modal approach. *Nonlinear Processes in Geophysics, 15,* 233–244.
10. Miropolsky, Y. Z. (2001). *Dynamics of internal gravity waves in the ocean*. In O. V. Shishkina (Ed.) Boston: Kluwer Academic Publishers.
11. Morozov, E. G. (1995). Semidiurnal internal wave global field. *Deep Sea Research, 42*(1), 135–148.
12. Morozov, E. G. (2006). Internal tides. Global field of internal tides and mixing caused by internal tides. In J. Grue & K. Trulsen (Eds.), *Waves in geophysical fluids* (pp. 271–332). Wein: Springer.
13. Pedlosky, J. (2010). *Waves in the ocean and atmosphere: sntroduction to wave dynamics*. Berlin-Heidelberg: Springer.
14. Rees, T., Lamb, K. G., & Poulin, F. J. (2012). Asymptotic analysis of the forces internal gravity waves equation. *SIAM Journal of Applied Mathematics, 72*(4), 1041–1060.
15. Sutherland, B. R. (2010). *Internal gravity waves*. Cambridge: Cambridge University Press.

High-Resolution Observations of Internal Wave Turbulence in the Deep Ocean

Hans van Haren

Introduction

The dynamics of the ocean resembles that of the atmosphere in many respects. Both are basically driven by the heating of the sun and adjusted by the rotation of the Earth [1]. One crucial difference is that the atmosphere is cooled at a higher geopotential level than where it is heated, while the ocean is heated at its top. As warm air/water is generally less dense than cold air/water, the heating generates natural vertical turbulent convective motions in the atmosphere, as an effective heat engine [2], while the ocean is merely a heat transporter as its heat engine is very ineffective [3]. Thus, a large amount of potential energy is stored in the ocean, but it requires a mechanical energy source to transport the heat downward against the stable stratification. Although the stable stratification in the ocean interior is 1000 times less than the density difference between air and water across the ocean surface, vertical diapycnal exchange is considered weak. The weak exchange may seriously affect life in the ocean and the replenishment of nutrients to the photic zone, for example.

However, the turbulent exchange is not blocked but hampered. This is because the density stratification may support destabilizing shear to the point of marginal stability (e.g., [4]) and mechanical energy in the form of waves, just like at the ocean surface. With the 1000 times weaker stratification in the interior, 'internal waves' may grow to attain 100 m large amplitudes that go barely unnoticed at the surface. Although linear, sinusoidal waves transport momentum but not matter, nonlinear waves can break and generate irreversible turbulent diapycnal mixing. It is the paradox of Munk and Wunsch [3] that demonstrates the importance of the relatively small amount of kinetic energy put into internal waves that is crucial to

H. van Haren (✉)
Royal Netherlands Institute for Sea Research (NIOZ), Utrecht University,
Den Burg, The Netherlands
e-mail: hans.van.haren@nioz.nl

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_11

maintain the ocean stratification via turbulence exchange, i.e. to govern the 25,000 times larger potential energy input by the sun.

Sea floor topography is important in both the generation (e.g., [5–7]) and the breaking of internal waves (e.g., [8, 9]). Not only the topography around ocean basin's edges act as source/sink of internal waves but especially also the topography found in ridges, mountain ranges and seamounts distributed over the ocean floor. There is more underwater topography than on land. The main mechanical source interacting with the topography is the tidal motion, which carries about 70% of the internal wave energy on average [3]. Another prominent source is related with the rotation of the Earth and the geostrophic response to sudden changes, e.g., by the passage of atmospheric disturbances or by the collapse of fronts: Inertial motions [1, 10]. Although the general aspect ratio of large-scale vertical:horizontal ocean current components is 1:1000, or perhaps 1:100, an encounter with topography forces a relatively large vertical motion that sets the, initially flat and horizontal, layers of constant density into oscillatory movements. The thus generated internal waves can propagate in full three-dimensions 3D spatially. After interactions with waves at other frequencies, or currents, or topography, the dominant process still being unknown, the initially linear waves become nonlinear and break whereby dissipating their energy. The associated turbulence not only affects the ocean stratification but also the marine geology and chemistry, by resuspending sediment and redistributing suspended material, and marine biology by the redistribution of nutrients and oxygen.

As internal waves can propagate freely when their frequencies are between the inertial frequency f, of horizontal oscillatory motion with short vertical scale thus an important shear 'source', and the buoyancy frequency N, the natural frequency of vertical oscillatory motion governed by the stratification, they are rather slow waves compared to surface waves. Internal wave periods vary between about a day and hours (in the deep ocean) to minutes (in generally stronger stratified waters of shallow seas and near the ocean surface). Such a slow motion is associated with linear waves that go up and down regularly. Faster motions are found beyond the buoyancy scales where the transitions to turbulence are found. When turbulence is mainly induced by wave-breaking, fronts of deformed internal waves pass instrumentation fixed in space within minutes.

The wave-breaking is a complex process in which the two main mechanisms to stratified turbulence co-exist. Vertical current shear across thin interfaces leads to overturning of Kelvin Helmholtz billows or instabilities KHI (e.g., [11]). The wave-overturning leads to 'free' buoyancy driven convection turbulence of Rayleigh-Taylor instabilities of a collapse of unstable ('cold') fluid overlying less dense ('warm') water [12]. Both mechanisms relate in a complex manner, e.g., a convection column [13] leads to shear instabilities along its fringes [14] and shear-induced overturning leads to convection in its core [15]. In the ocean these processes are most developed above sloping bottom topography.

Although some topography-internal wave interaction processes are well described, such as internal hydraulic jumps or lee-wave formation over seamounts and ridges [16, 17] and frequent overturning from internal wave breaking over slopes steeper 'supercritical' than the internal tide slope (e.g., [18, 19], Sarkar 2016,

pers. comm.), others are puzzling. While laboratory experiments over uniform slopes suggest wave breaking on 'critical' bottom slopes that match the angle of internal wave propagation [20, 21], wave-breaking in the ocean is also observed away from critical slopes of the major internal wave frequency, mostly tides, e.g. [22]. Examples of the latter wave breaking is also observed to occur in areas where near-horizontal inertial motions are the dominant carrier wave like in the Mediterranean, or sub-inertial motions like in the Faeroe-Shetland Channel [23], or trapped waves like along the Rockall Bank. This suggests a sloshing motion importance for the generation of nonlinear waves leading to breaking [19].

Even if the precise process leading to nonlinear wave formation is unknown, or not known for every particular topographic region, the importance of internal wave breaking upon topography associates with the importance of 'boundary mixing' for the maintenance of ocean stratification, as suggested by Munk [24] and Munk and Wunsch [3]. Assuming a layer of 100 m of enhanced turbulence over topography in an ocean of average depth of 3,000 m, Armi [25] suggested a turbulent diffusivity of $K_z = 3 \times 10^{-3}$ m^2 s^{-1} sufficient to maintain the necessary heat flux in the ocean interior. This was debated by Garrett [26] as a boundary layer in the classic sense of Ekman [27], even over sloping bottoms [28], is homogeneous and thus very inefficient in its mixing, unless a complex restratification mechanism is invoked. The debate was based on steady flows having time scales much longer than the inertial period, or actually $t \gg 1/f$. Microstructure profiler observations by, e.g., [29, 30] to within 0.3 m from the bottom above sloping topography in shallow seas and [31] in the deep ocean showed that the turbulent flux ($\propto$turbulence dissipation rate) does increase towards the bottom and, indirectly, evidences effective mixing or a mechanism to restratify the sloping 'bottom boundary layer'. Detailed moored temperature sensor measurements (e.g., [32]) indicate that internal wave breaking dominates bottom friction in turbulence generation over sloping topography. It is also the key mechanism in rapidly restratifying the near-bottom area, by transporting the mixed waters into the interior and by pushing the stratification to within a meter or so from the bottom on time scales of the order of the inertial time scale, or shorter. They alternate with sudden upslope propagating frontal bores [23, 33] reaching up to 100 m above and actually touching the bottom thereby creating large sediment resuspension resembling a desert dust storm. In essence, a bottom boundary layer is not observed over sloping topography, at least not in the classic sense.

A challenge to this view is the requirement of a near-bottom reduction of the heat flux in order to have an upslope motion close to the bottom balancing the interior downward turbulent heat transport [34]. Whilst the concept is well-posed, it remains to be established how close to the bottom the heat flux should be reduced for the mechanism to work. Perhaps, the adoption of the theorem of Prandtl that the length-scales of the largest turbulent overturns can never be larger than the distance to the bottom in a mixing length theory may not be adequate for flows with more than one characteristic velocity [35]. Prandtl did not consider wave breaking over sloping topography, but frictional shear-driven turbulent overturning over a solid boundary.

A method has yet to be found to properly parameterize such wave-breaking motions, in the context of the large-scale ocean circulation. One of the unknowns is to relate to the proper scaling of topography. As the term 'rough topography' [9, 31] is somewhat imprecise in quantification, it would be better to consider the slope angle, in relation with the internal wave angles for given stratification, and the length scales of the slope-determination in association with the internal wave length scales. Apart from such modeling constraints, an encouraging recent tracer-release experiment near the Mid-Atlantic Ridge integrated over yearlong periods showed upslope motions that can only be balanced by diffusivities of about $K_z = 5 \times 10^{-3}$ m^2 s^{-1} (Ledwell 2017, pers. comm.). It will be shown in this paper that such values are within a factor of two similar to those determined from high-resolution temperature sensors moored in various deep-ocean topographic regions. Under conditions of very high Reynolds numbers in the ocean that are not achieved in laboratories and a tight temperature-density relationship, such observations prove adequate in quantifying internal wave turbulence because of many 1-Hz yearlong sampled realizations of true vertical profiles under very little mooring motion.

Instrumentation and Data

Quantifying Internal Wave Turbulent Mixing: Lowered Instruments

Most turbulence measurements in the ocean have been performed using free-falling (few free-ascending) microstructure profilers, either loosely tethered to the ship or freely operating. These profilers carry shear probes and fast-response temperature (sometimes also conductivity) sensors that more or less resolve the Kolmogorov scales of smallest turbulence dissipation overturns that are about 0.1–0.01 s in time and 0.001–0.01 m in space, under ocean conditions. For statistical reasons, the raw several 100 Hz sampled data are divided in blocks of a few seconds long to give estimates of turbulence dissipation rate and eddy diffusivity. The disadvantages of this instrumentation are its relatively slow vertical speed of about 0.7 m s^{-1}, its 1D profiling of essentially 3D physical processes and its costs: a full ocean depth apparatus requires financing over 0.5 million US dollars and not many exist in the community.

A simpler but more commonly useable method was proposed by Thorpe [36] to infer turbulence estimates from standard oceanographic Conductivity Temperature Depth CTD data. This shipborne equipment also collects vertical (1D) profiles of ocean quantities at a slightly higher speed of just under 1 m s^{-1} while sampling at a rate of 24 Hz for modern equipment. The turbulence estimation method is by reordering vertical density profiles into statically stable ones and keeping track of the displacements. Thus after some averaging to reduce noise the larger energy-containing turbulent eddies are resolved but not the (Kolmogorov) dissipation scales. In essence this is not a problem, provided some assumptions are made.

The first assumption is that the overturning displacements d are related to the Ozmidov scales L_O of largest overturns in stratified turbulence. As indicated by Thorpe [36], after suitable (rms) averaging this is true to within a constant ratio of $c_1 = L_O/d_{rms} = 0.8$. This value was empirically established by Dillon [37] as being a mean in a distribution of about one order of magnitude wide. The distribution reflects the various stages of turbulence development and the two principal turbulence processes of shear-induction and convection. Recently, the overturning method was challenged from numerical simulations showing that convection leads to an underestimate of the Ozmidov scale using the displacements. But, the high Reynolds number (10^4–10^6) turbulence in the stratified ocean is a mix of shear-induction and convection rendering the average ratio very close to 1 [38] and confirming Dillon's observations. The second assumption is to use a constant value $m_1 = 0.2$ for the mixing efficiency [39, 40]. As with the Ozmidov/Overturning scale ratio, this constant is a mean from a distribution of about one order of magnitude wide and thus requires some suitable vertical, horizontal and/or time averaging. The third assumption was not necessary to make for Thorpe [36] as he designed the method for use in lakes: It involves salinity and ship's motions. The disadvantages of using the overturning method in the ocean are the ship's motion by surface waves that are only partially compensated by a heave compensator and the contribution of salinity to density variations. Both require corrections to the raw data that approach the order of magnitude of turbulent overturns (e.g., [41–44]).

Quantifying Internal Wave Turbulent Mixing: Moored Instruments

Instead of lowering instruments from a ship one can use Eulerian moored instrumentation fixed in space under water and have the flow go past it. Under the assumption that one can separate via 'Reynolds decomposition' the turbulent fluctuations from the 'mean flow', one may directly estimate the turbulent momentum and/or heat (mass) fluxes with suitable current and temperature (conductivity) measurement devices. Although stand-alone instrumentation is heavily limited by sufficient power supply, this method is reasonably successful for estimating momentum fluxes in shallow seas by a moored acoustic point source measurement device like the Nortek Vector in shallow seas almost resolving the dissipation scales (e.g., [45]). Resolving larger scales, moored acoustic profiling instrumentation, especially operating in pulse-coherent mode, estimates statistically significant momentum fluxes above 'flat' shallow sea floors with dominant tidal friction (e.g., [46, 47]) and sloping topography [30]. However, the method has seldom produced significant results from the deep ocean [48], possibly due to the lack of sufficient scatterers. Few attempts have been made to actually estimate heat (mass) fluxes, above sloping topography [48, 30]. It appeared difficult to match the acoustic current measurements in slanted beams with data from a chain of

temperature 'T'-sensors along its central axis. More principally, the Reynolds decomposition requirement of a spectral gap between mean flow and turbulence is not found in the ocean. The method was thus also found inadequate for the deep ocean.

Quantifying Internal Wave Turbulent Mixing: Moored High-Resolution T-sensors

As an alternative, NIOZ developed high-resolution T-sensors to be moored with relatively many and supported by some current and pressure measurements. Initially T-sensors were connected via cables to a central data logger and power supply unit [22]. After considerable connection problems it was decided to work with stand-alone self-contained sensors [49]. The high-resolution <0.1 mK and high precision <0.5 mK sensors have a response time of 0.25 s in water and standard sample at a rate of 1 Hz (adjustable) for a duration of one year on one AA-battery. Typically, 100 sensors are taped to a plastic-coated steel cable of several 100 m long. This offers great flexibility and sensors have been taped at intervals between 0.04 and 16 m depending on the required configuration for study. This is adequate to resolve most of the stratified turbulence and internal wave overturning scales as mean Ozmidov scales are between 0.1 and a few 10's of meters while large-scale internal wave amplitudes reach 100 m. The steel cable is used to synchronize the sensor clocks so that a profile of up to 600 m long (tested) is measured within 0.02 s. Such a snapshot over this vertical range is not achievable using 0.7–1.0 m s^{-1} free-falling or lowered profilers. The large number of sensors in a profile is also adequate to correct for sensor drift, which amounts about 1 mK/mo after aging. For an averaging period longer than the buoyancy period, but commonly taken longer than the inertial period to be sure, turbulent overturns cannot last and the mean ocean profile is statically stably stratified. Any remaining unstable deviations are thus instrumental (or salinity driven, see below) and need to be corrected. This is achieved by forcing the mean observed profile to a smooth, stable profile using a constant correction per sensor. In practice, an averaging period of 2–7 days is chosen for this local correction and repeated for different times in the (up to yearlong) record.

Considerable effort is put in improving the mooring design by minimizing its motions to obtain almost truly Eulerian measurements with negligible artifacts in the frequency range of interest. At NIOZ it is preferred to have all buoyancy near the top so that recovery after successful release of the bottom weight is smooth without entanglement of cables. As large-scale ocean currents are almost horizontally, the associated drag forces by any object obstructing the flow are nearly perpendicular to gravity (negative buoyancy). A balance of forces is thus impossible and the mooring will inevitably be deflected by a non-zero flow. We impose the criterion that the vertical deflection may never exceed half the distance between any of the instruments along the mooring line, so typically <0.5 m for a mooring with a string of 1-m

separated T-sensors. This is achieved by using thin cables of 0.005 or 0.006 m diameter that have breaking strengths >20,000 N. These cables are tensioned by up to 5,000 N using a weight of >500 kg and net buoyancy of 300 kg where currents reach 0.3–0.4 m s^{-1}. Mooring design is verified using Richard Dewey's software (University of Victoria, BC, Canada). The mooring motion is always monitored by mounting at least one current meter and pressure and tilt sensors.

The sensors are calibrated at NIOZ using a thermostatic bath with constant temperature levels to within $\pm 10^{-4}$ °C of their preset values. The calibrated and drift-corrected data are transferred to Conservative (~potential) Temperature (Θ) values using the gsw-software described in [50]. As the ocean density is also determined by salinity besides temperature, several shipborne SeaBird-911 CTD-profiles are obtained near the mooring. The CTD-data are also used to establish the local temperature-density relationship for use of the moored T-sensor data as tracer for potential density anomaly variations $\delta\sigma_x$, referenced to the pressure level x/1000 dbar of the mooring (see for an example Fig. 1).

When the linear relationship is reasonably tight (relative error of about 1–10%), the constant gradient $\alpha = \delta\sigma_x/\delta\Theta$ indicates an apparent local thermal expansion

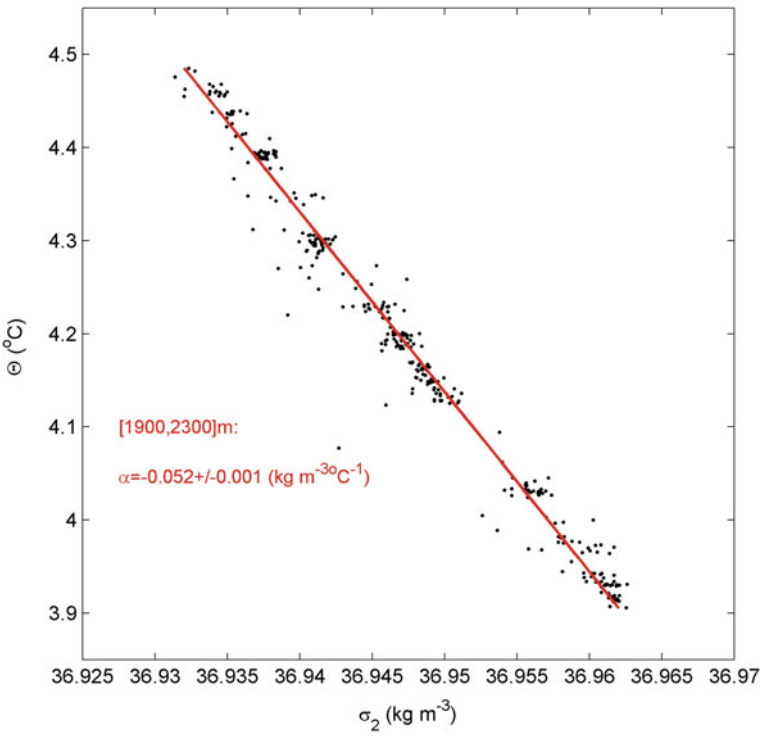


Fig. 1 Conservative temperature-density anomaly relationship, referenced to 2,000 dbar from CTD-data obtained between 1,900 and 2,300 m near a T-sensor mooring on a slope of Mount Josephine NE Atlantic Ocean. The slope of the linear best-fit denotes the local apparent expansion coefficient

coefficient. If this relationship is not tight a relatively large or dominant contribution results of salinity to density variations, as around Mediterranean outflow of relatively warm and salty waters into a cooler and fresher Atlantic environment and in polar near-surface regions where the cold waters lead to ambient low thermal expansion coefficients, with added ice melt. This leads to intrusions of apparent overturns that can last longer than the buoyancy period. If intrusions are thus detectable they may qualitatively demonstrate turbulent motions, but quantitative turbulent information is no longer obtainable [51]. In contrast, the above relationship is always tight for data from lakes, where temperature exclusively dominates density variations, with the notion that the natural thermal expansion coefficient changes sign at about 4 °C.

With a tight temperature-density relationship, turbulence dissipation rate $\varepsilon = c_1^2 d^2 N^3$ and vertical eddy diffusivity $K_z = m_1 c_1^2 d^2 N$ are estimated from the T-sensor data using the method of Thorpe [36]. Here, N is computed from the reordered profiles. Figure 2 demonstrates the post-processing of splitting original 1 Hz profiles

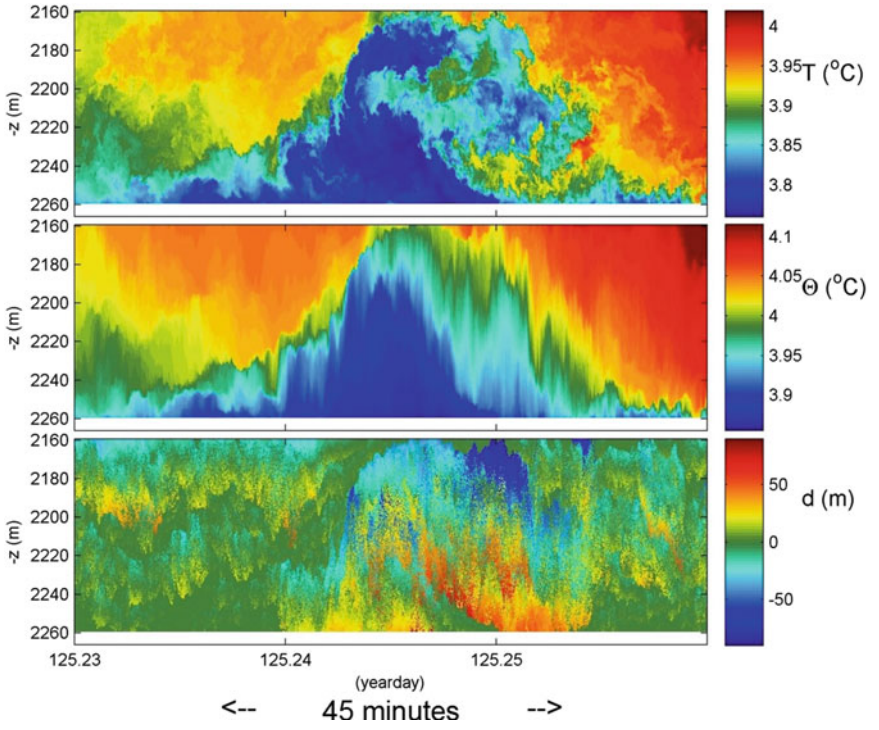


Fig. 2 Detailed time-depth series example showing the split of observed moored T-sensor data into stably reordered profiles and the necessary displacements following the reordering. Forty-five minutes of data from an arbitrary 100-m tall overturn observed above Mount Josephine (the local water depth is at the level of the x-axis). **a** Original 1-Hz sampled temperature data using 100 independent sensors at 1 m vertical intervals starting at 6 m from the bottom. **b** Data from **a**. transferred to conservative temperature 'CT' and reordered to stable 1-Hz profiles. **c** Displacements between original CT-data and data from **b**

into reordered stable profiles and displacements. The 3D appearance (Fig. 2a) so well captured by 17th century clouds and landscape painters is completely absent in the flat 2D image (Fig. 2b) from which the turbulent displacements (Fig. 2c) are removed. The raw turbulence parameter estimates for every sensor and every (1 s) profile are averaged over depth, hereafter indicated by <…>, and time, indicated by […].

Thorpe [9, 36] proposed to average over individual turbulent overturns, but as smaller overturns exist in larger ones that even may exceed the range of moored sensors (cf., Fig. 2), it was decided to average over the full vertical range of T-sensors when it is equal to or less than 100 m [32]. Commonly, about 5% of the T-sensors show electronic or calibration problems and their data are linearly interpolated between neighboring sensors. This affects the turbulence parameter estimates by less than 10%, well within the standard error in the estimates due to the procedure.

Observations

In the examples given in this section, areas are avoided where the temperature-density relationship is not tight. Although the locally established apparent thermal expansion coefficients vary considerably for different regions between 0.05 and 0.3 kg m^{-3} $°C^{-1}$ and their 'tightness' between 1 and 10% of the mean value, the standard errors in turbulence parameters are always about a factor of two. This is equivalent to errors in microstructure profiler data after considerable and careful post-processing (Oakey 1991, pers. comm.). The factor of two error is confirmed by comparing moored T-sensor overturn estimates above a seamount slope with nearby shipborne lowered ADCP/CTD data estimates using Gregg [52]'s shear-scaling [32].

In the Open Ocean Far Away from Topography

The stratified open ocean interior is permanently in motion, not merely by surface waves but especially also by internal waves with typical amplitudes of several tens of meters (e.g., [53–55]). The weakly turbulent motions are smooth, albeit not perfectly sinusoidal in shape, and have relatively large vertical coherent scales with near-zero phase difference over the range of thermistors ('local vertical mode-1'), at both tidal and near-buoyancy frequencies. Amplitudes vary continuously for motions within any particular finite frequency band; evidence of intermittency. Due to straining, the thickness of strongly stratified layers drops below 1 m, while weakly stratified layers can exceed tens of meters. In the Canary Basin [55], the mean Ozmidov scale $<L_O> \approx 0.4$ m (obtained from high-resolution shipborne CTD-data). The mean interior turbulence parameters are estimated as $[<\varepsilon>] \approx 10^{-9}$ m^2 s^{-3} and $[<K_z>] \approx 2 \times 10^{-5}$ m^2 s^{-1}, which is equivalent to values estimated for other, upper 2000 m, open ocean regions [52] and for, e.g., the summer stratified

North Sea [4]. Such values are still 100 times larger than molecular diffusivity values and sufficient to explain, at least for shallow seas, the gradual warming of the near-bottom waters and the flux of nutrients into the photic zone to entirely support the late-summer phytoplankton bloom. This was found to be due to the stratification being marginally stable, with gradient Richardson number $Ri = N^2/|S|^2$ values of about $Ri = 0.5$, where the shear magnitude $|S|$ was composed of a quasi-permanent large inertial/tidal shear across largest stratification added with occasional small-scale internal wave shear. These turbulence values are also typical for a saturated internal wave field allowing for sparsely distributed turbulence [56], similar to surface wave white capping of a puff here and a puff there. However, according to Munk [24] they are insufficient to maintain the ocean stratification.

Above Sloping Topography

When internal waves reach underwater topography like seamounts or ridges, the smooth motions turn into more irregular motions up to highly nonlinear and vigorous turbulent overturning, see for example Fig. 3. Although bottom slopes are on average just a few per cent and typical internal tide generating horizontal currents are 0.1 m s^{-1}, the impact of such currents and slopes is spectacular in their production of wave breaking. Turbulent overturning ranges from 10 to 100 m vertical scales (Figs. 2a and 3a).

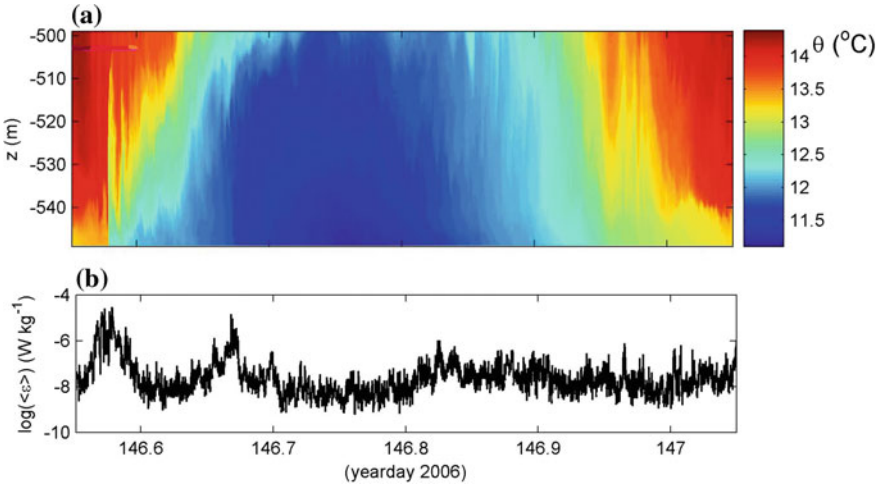


Fig. 3 One tidal period of temperature observations using 101 temperature sensors above Great Meteor Seamount NE Atlantic Ocean, sampled at a rate of 1 Hz at 0.5 m vertical intervals, starting at 0.5 m from the bottom. **a** Time-depth image shows a tidal wave well exceeding the 50 m range of sensors. **b** Time series of the logarithm of turbulence dissipation rate inferred from the data in (**a**)

Especially the transition from warming to cooling phase of the large-scale carrier wave, associated with an upslope moving frontal bore, provides large turbulence peaking at $\varepsilon = 10^{-5} - 10^{-4}$ m^2 s^{-3} and $K_z = 10^{-1} - 10^{0}$ m^2 s^{-1} (Fig. 3b). The reordered stable stratification is characterized by very thin layering down to the lowest vertical resolution, in this case $\Delta z = 0.5$ m and much smaller than $<L_O> \approx$ 20 m. The layering approaches the bottom to within a meter just prior to a frontal passage by large interior overturns associated with relatively strong (0.1 m s^{-1}) downward motions [32]. The large convective overturns preceding the front are not so efficient in their mixing ($m_1 \approx 0.05$), but the near-bottom stratified layer forming the frontal bore is more efficient ($m_1 \approx 0.3 \leq 0.36$) than average [57]. The downdraught seems part of sharpening the turbulent bore moving up the slope.

The frontal bore itself is basically the only large overturn extending from the bottom upward, thus being important for sediment resuspension governed by 0.1 m s^{-1} upward vertical current speeds [23]. Secondary instabilities are observed along its fringe, contributing to effective mixing [58]. Although turbulence associated with upslope propagating bores is always large, the intensity and precise extent vary every cycle of the carrier wave. The variations depend on the exact direction of bore propagation with respect to the main bottom-slope direction, the more upslope the stronger turbulent, and on the timing of arrival which varies over about 10% of the carrier wave period [59]. This variation in arrival time was attributed to variations in the background stratification which modify the paths of internal waves, and apparently also large-scale sub-inertial motions. Behind the front the trailing near-N waves also turn-over due to internal wave shear across the sharp interface, but in rapidly decreasing magnitude.

Averaged over time of a two-week period and over depth over the range of T-sensors gives $[<\varepsilon>] = 3 \pm 2 \times 10^{-7}$ m^2 s^{-3}, $[<K_z>] = 6 \pm 4 \times 10^{-3}$ m^2 s^{-1} for $[<N>] = 5 \pm 3 \times 10^{-3}$ s^{-1} above steep 'supercritical' slopes larger than the average internal tidal slope [32, 60]. These turbulence values are more than two orders of magnitude larger than estimated from ocean interior observations. Above less steep slopes, values are smaller, but even above slopes well less than the internal tide slope values are still about one order of magnitude larger than ocean interior values [60]. In the vertical above a supercritical slope, the latter low values are not observed up to 400 m above the bottom (Fig. 4) possibly due to the large extent of tidal motions that have top-trough excursions exceeding 100 m [61]. These observations tend to support microstructure profiler observations of the decrease of turbulence intensity to ocean interior values over a range of some 1000 m above sloping topography [31].

Interfacial Shear Instability Observations

Although near sloping bottoms moored high-resolution T-sensors show largest turbulent overturning commonly in the form of frontal bores around the transition between warming and cooling phases of the large-scale carrier wave, time series of

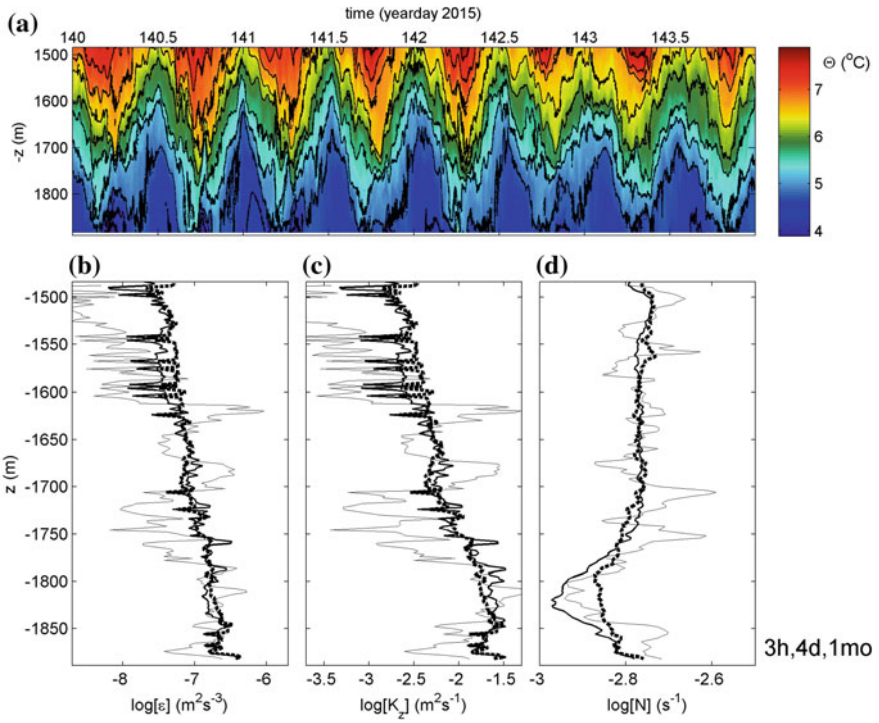


Fig. 4 Four-day detail of temperature and inferred turbulence measurements above a super-critical slope for internal tides of Mount Josephine. **a** Conservative temperature time-depth series. **b–d** Full depth profiles of logarithms of time-averaged parameter values. One month averages are indicated by thick-dashed lines, four day (Fig. 4a) averages by thick-solid lines and 3 h averages by thin-solid lines. **b** Turbulence dissipation rate. **c** Vertical turbulence diffusivity. **d** Buoyancy frequency. After [61]

vertically averaged turbulence parameters show a variation over four orders of magnitude (Fig. 3b; e.g., [32]). This is typical for ocean turbulence and is also observed in vertical microstructure profiles (e.g., [39, 52, 31]). Apparent spiky variations over two orders of magnitude are not instrumental noise but indicative of individual overturns. Typical examples are trains of shear-induced KHI.

Trains of 5–10 KHI-overturns of about 5 m high are observed during the warming phase of the carrier wave, when the large-scale shear irregularly coincides with small-scale shear of internal waves near the local buoyancy frequency, see for example Fig. 5. The average $<L_O> \approx 1.8$ m and some overturn-billows are short-lived $O(10^1 - 10^2$ s). A more prominent train of 275 KHI-overturns in one day has been observed in the deep ocean around 4,500 m [62]. There at the entrance of the equatorial Romanche Fracture Zone of the Mid-Atlantic Ridge the tidal flow interacts with the large-scale flow driven by Antarctic Bottom Water plunging

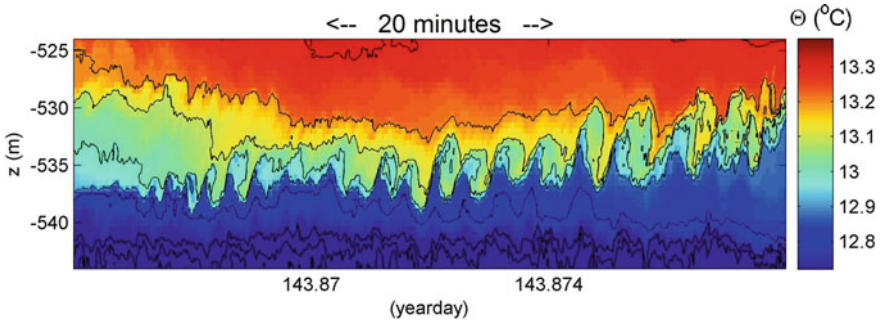


Fig. 5 Twenty minutes and 20 m detailed example of the downslope propagating, warming phase of the internal tide observed using NIOZ3 T-sensors at 0.5 m vertical intervals near the top of Great Meteor Seamount

underneath North-Atlantic Bottom Water on its trajectory from the (south) western to the (north) eastern Atlantic Basin. When the tidal phase is in the direction of the large-scale flow, KHI have a vertical extent of 10–20 m. Half a tidal period later when the phase is directed against the large-scale flow KHI exceed 100 m vertically.

In all these KHI observed in the deep ocean, the roll-up stage of initial KH billows seems to be limited to half a turn, resembling half a Rankine vortex [44], before collapse. This is somewhat different than found in laboratory experiments (e.g., [63]) and numerical modelling (e.g., [15, 64]). As a rare exception, near-surface ocean optic [65] and acoustic observations [66] show a roll-up. Although the lack of deep-ocean observations of multiple scale billow roll-up is presently unknown, it may have to do with the large Reynolds numbers associated with the turbulent environment above sloping topography, causing a rapid collapse of the billows.

Some Temperature Statistics

The largest overturns observed have a vertical length scale of about 100 m [62] and a time scale of just less than the buoyancy period [32]. The latter is commensurate the expectation that the buoyancy frequency is a natural scale separator between irreversible turbulent overturning and internal wave motions. As for the interpretation of (moored) temperature observations as a tracer for density variations, this information is useful to help distinguishing between genuine turbulent overturning and partially salinity compensated intrusions. The difference in shapes of apparent and turbulent temperature instabilities is also used for distinction.

A further distinction and determination is obtained from temperature scalar statistics [18]. In particular are we interested in the distinction between passive scalar, shear-dominated turbulence and active scalar, convection dominated

turbulence. The analysis confirms the differences sketched above between upslope cooling phase and downslope warming phase of the carrier wave. The former is consistent with passive scalar statistics but for the bottom half of a 100 m tall T-sensor array, while in the upper half also evidence is found of active scalar statistics. Likewise, the downslope phase is not exclusively dominated by passive scalar statistics. These observations confirm the notion that shear- and convection turbulence co-exist in the ocean, at least above sloping topography. In general strong intermittency is found during the downslope phase, while a clear inertial subrange of −5/3 spectral slope is observed during the upslope phase in particular in the lower 50 m above the bottom (cf., Fig. 6).

From extended statistics using nearly 500 sensors in a small-scale 3D mooring array of five lines 100 m long and 4 m apart horizontally, the inertial subrange is observed superposed by larger variability than expected from chi-squared statistics [67]. The observations suggest localized patches that are more coherent than their spectral surroundings (Fig. 6), and which are not found in open-ocean spectra. They may be due to intermittency, to convection in a shear-dominated environment, or to an interaction between turbulence and small-scale thin-layer internal wave motions: The precise mechanism is not known as the pattern is not yet resolved, except that their energy levels are found to be proportional to the tidal energy level as their peaks fall along a −2 spectral slope to within the 95% statistical significance level.

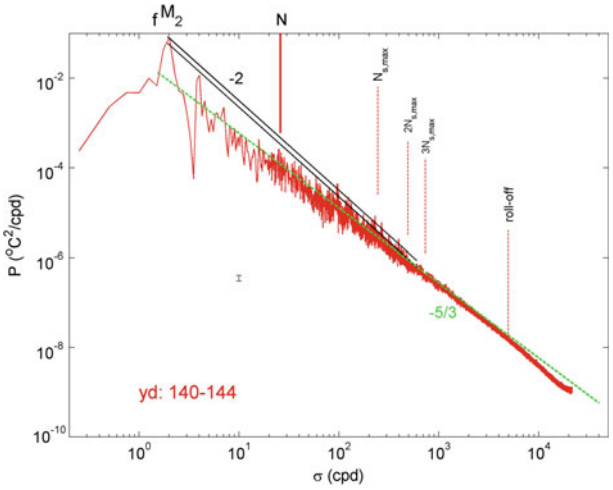


Fig. 6 Temperature variance spectrum, an average of 442 near-raw periodograms for the Kaiser window tapered time series of a four-day turbulent period observed by independent T-sensors at a small-scale 3D (five line) mooring array above a super-critical slope of Mount Josephine. The data are sub-sampled at 0.5 Hz for computational reasons. Besides inertial (f) and semidiurnal lunar tidal (M_2) frequencies several buoyancy frequencies are indicated including four-day large-scale mean N and the maximum small-scale $N_{s,max}$. The number −5/3 indicates the spectral slope $\sigma^{-5/3}$ for frequency σ, −2 the spectral slope σ^{-2}. The 95% significance level is indicated by the small vertical bar and by the width of the two −2 slopes. After [67]

Whatever, the 3D mooring reveals the transition from squashed stratified turbulence having an aspect ratio of ≤ 0.5 near the large-scale buoyancy frequency to fully developed turbulence of aspect ratio equal to 1 before coherence is lost around a frequency of about twice the maximum (resolved) small-scale buoyancy frequency.

Discussion and Conclusions

Turbulence values are high in the lower 100 m above sloping topography (e.g., [8, 32]). They are 100–1000 times larger than observed in the open ocean, away from boundaries (e.g., [52, 31). Considering the rapid restratification process by the (sloshing or propagating) carrier waves, the turbulent mixing is found to be effective. Thus, the interplay between the large-scale internal waves and the small-scale waves near the buoyancy frequency acts to create both the diapycnal turbulent mixing and the isopycnal transport of the mixed waters into the interior. The T-string observed high mean turbulent diffusivity values $O(10^{-3} - 10^{-2})$ m^2 s^{-1} are confirmed by various other Mid-Atlantic Ridge shipborne data (e.g., [68]) and Jim Ledwell's recent Mid-Atlantic Ridge tracer release experiments: These diffusivities are (observed to be) required to balance a near-bottom upslope motion.

Following Armi's [25] suggestion that about 3% of the ocean is occupied by such 100 m tall layer of large turbulence, the high mean turbulent diffusivities observed over sloping topography will yield an overall mean ocean-basin-interior diffusivity of $K_z \approx 1\text{–}3 \times 10^{-4}$ m^2 s^{-1}, sufficient to maintain the ocean stratification [3, 24]. Above particular steep seamount slopes of Mount Josephine (NE Atlantic Ocean) that are supercritical for internal tides, the local large-scale carrier wave, high local diffusivity values suggest a relatively narrow range of only 450 m depth interval sufficient to supply this overall ocean-basin-interior diffusivity if the turbulence observed over these particular slopes is extrapolated to other ocean topography [18]. This large turbulence above tidally supercritical slopes is confirmed by numerical modelling of Winters [19] and Sarkar (2016, pers. comm.). It supports earlier conjectures (e.g., [24, 25]) for the importance of sloping boundary mixing.

Previous and present data suggest that such mixing is quite universal, independent of forcing (carrier wave) mechanism (frequency), but strongly dependent on slope with respect to that of the carrier wave. The turbulence above sloping topography well (>300 m) below the summits of seamounts is determined by slope steepness and nonlinear wave evolution, but not by bottom-friction, and not necessarily by 'critical' internal tide reflection, internal hydraulic jumps or lee-wave generation. Lee-waves are estimated to carry about one-tenth of the energy of internal tides, although this may vary for certain locations [69]. Tides are not the only forcing mechanism, as turbulent overturns are also observed in seas like the Baltic and the Mediterranean Sea where tides are weak. Bottom slopes need not be 'critical' for sediment resuspension as suggested for the occurrence of intermediate nepheloid layers [70]: Vigorous bore-like motions have also been observed at a

variety of [tidally] non-critical slopes for sub-inertial motions (e.g., [23, 33]). Near-inertial shear, the largest internal wave shear outside internal tidal generation areas, is important for (high-frequency) internal wave breaking away from the bottom, in case shear-induced mixing is the dominant turbulence generation process. Thus, most internal wave—turbulence transitions seem associated with non-linear deformation of internal wave motions near local buoyancy frequencies.

The ocean interior is a much smoother linear-wave-like environment, but it too is in permanent motion rather than quiescent. This ocean in permanent motion is driven by surprisingly weak kinetic energy forcing of about 20% of the total energy presently used by mankind, about 500 times less than the heat flux transported in the ocean [3] and 25,000 times less than the solar energy received by the ocean. Man can maximally retrieve about 3% van the kinetic tidal energy or maximum 0.6% of his present-day demand with the notion that the present-day technology is not economically viable. In the hypothetical case that man were able to efficiently extract more tidal energy he would destroy the ocean stratification and with it ocean life, not only in the deep ocean but especially also in tidal inlets of nursery areas where currents are largest.

Of all the power of 100-m breaking waves above deep-ocean topography, while highly important for redistribution of sediment and for life e.g. by replenishing nutrients to cold-water coral mounds (e.g., [71]), which are found at depths where mineralization is relatively large as inferred from the seasonal oxygen minimum layer coinciding with nutrient maxima (e.g., [72]), very little is sensed at the surface. Therefore, in order to learn more about internal wave-turbulence processes it is mandatory to continue studying their characteristics in the deep sea and ocean, above sloping topography but especially also their extend into the interiors. Future instrumentation development should be directed to a truly four-dimensional, 3-D spatial plus temporal, resolution of the internal wave-turbulence motions to understand their intrinsic properties. Typical turn-over scales generated by KHI and frontal bores are between 1 and 100 m, whereas typical internal wave lengths are 100–1000 m. A small array with T-sensors a few meters apart would be useful to study internal wave—turbulence resolving the short spatial scales. A larger cubic hectometer array with T-sensors O(10 m) apart would be useful to study the statistical properties and internal wave propagation and will yield new insight in internal wave development. After all, deep-ocean sloping topography generates great surf!

Acknowledgements I greatly thank M. Laan, L. Gostiaux, A. Cimatoribus and F. Cyr for their discussions, collaboration in design and construction of NIOZ temperature sensors. NIOZ temperature sensors has been financed in part by NWO, the Netherlands Organization for Scientific Research.

References

1. Gill, A. E. (1982). *Atmosphere-ocean dynamics* (p. 662). Orlando Fl, USA: Academic Press.
2. Peixoto, J. P., & Oort, A. H. (1992). *Physics of climate* (p. 520). Melville NY, USA: AIP-Press.
3. Munk, W., & Wunsch, C. (1998). Abyssal recipes II: Energetics of tidal and wind mixing. *Deep Sea Research Part I: Oceanographic Research Papers, 45,* 1977–2010.
4. van Haren, H., Maas, L., Zimmerman, J. T. F., Ridderinkhof, H., & Malschaert, H. (1999). Strong inertial currents and marginal internal wave stability in the central North Sea. *Geophysical Reseach Letters, 26,* 2993–2996.
5. Bell, T. H. (1975). Topographically generated internal waves in the open ocean. *Journal Geophysical Research, 80,* 320–327.
6. LeBlond, P. H., & Mysak, L. A. (1978). *Waves in the Ocean* (p. 602). New York, USA: Elsevier.
7. Morozov, E. G. (1995). Semidiurnal internal wave global field. *Deep Sea Research Part I: Oceanographic Research Papers, 42,* 135–148.
8. Eriksen, C. C. (1982). Observations of internal wave reflection off sloping bottoms. *Journal Geophysical Research, 87,* 525–538.
9. Thorpe, S. A. (1987). Current and temperature variability on the continental slope. *Philosophical Transactions of the Royal Society of London A, 323,* 471–517.
10. Davies, A. M., Xing, J. (2005). The effect of a bottom shelf front upon the generation and propagation of near-inertial internal waves in the coastal ocean. *Journal of Physical Oceanography, 35*, 976–990.
11. Turner, J. S. (1979). *Buoyancy effects in fluids* (p. 368). Cambridge, UK: Cambridge University Press.
12. Sharp, D. H. (1984). An overview of Rayleigh-Taylor instability. *Physica D, 12,* 3–18.
13. Marshall, J., & Schott, F. (1999). Open-ocean convection: Observations, theory and models. *Reviews of Geophysics, 37,* 1–64.
14. Li, S., & Li, H. (2006). Parallel AMR code for compressible MHD and HD equations. T-7, MS B284, Theoretical division, Los Alamos National Laboratory, http://math.lanl.gov/Research/Highlights/amrmhd.shtml.
15. Matsumoto, Y., & Hoshino, M. (2004). Onset of turbulence by a Kelvin-Helmholtz vortex. *Geophysical Reseach Letters, 31,* L02807. https://doi.org/10.1029/2003GL018195.
16. Inall, M. E., Rippeth, T. P., Griffiths, C. R., & Wiles, P. (2005). Evolution and distribution of TKE production and dissipation within stratified flow over topography. *Geophysical Reseach Letters, 32,* L08607. https://doi.org/10.1029/2004GL022289.
17. Klymak, J. M., Legg, S., & Pinkel, R. (2010). A simple parameterization of turbulent tidal mixing near supercritical topography. *Journal of Physical Oceanography, 40,* 2059–2074.
18. Cimatoribus, A. A., & van Haren, H. (2015). Temperature statistics above a deep-ocean sloping boundary. *Journal of Fluid Mechanics, 775,* 415–435.
19. Winters, K. B. (2015). Tidally driven mixing and dissipation in the boundary layer above steep submarine topography. *Geophysical Reseach Letters, 42,* 7123–7130. https://doi.org/10.1002/2015GL064676.
20. Dauxois, T., Didier, A., & Falcon, E. (2004). Observation of near-critical reflection of internal waves in a stably stratified fluid. *Physics of Fluids, 16,* 1936–1941.
21. Ivey, G. N., & Nokes, R. I. (1989). Vertical mixing due to the breaking of critical internal waves on sloping boundaries. *Journal of Fluid Mechanics, 204,* 479–500.
22. van Haren, H., Groenewegen, R., Laan, M., & Koster, B. (2005). High sampling rate thermistor string observations at the slope of Great Meteor Seamount. *Ocean Science, 1,* 17–28.
23. Hosegood, P., Bonnin, J., & van Haren, H. (2004). Solibore-induced sediment resuspension in the Faeroe-Shetland channel. *Geophysical Reseach Letters, 31,* L09301. https://doi.org/10.1029/2004GL019544.

24. Munk, W. (1966). Abyssal recipes. *Deep-Sea Research, 13,* 707–730.
25. Armi, L. (1979). Effects of variations in eddy diffusivity on property distributions in the oceans. *Journal of Marine Research, 37,* 515–530.
26. Garrett, C. (1990). The role of secondary circulation in boundary mixing. *Journal Geophysical Research, 95,* 3181–3188.
27. Ekman, V. W. (1905). On the influence of the Earth's rotation on ocean-currents. *Arkiv för Matematik, Astronomi och Fysik, 2*(11), 1–52.
28. Weatherly, G. L., & Martin, P. J. (1978). On the structure and dynamics of the oceanic bottom boundary layer. *Journal of Physical Oceanography, 8,* 557–570.
29. Dewey, R. K., Crawford, W. R., Gargett, A. E., & Oakey, N. S. (1987). A microstructure instrument for profiling oceanic turbulence in coastal bottom boundary layers. *Journal of Atmospheric and Oceanic Technology, 4,* 288–297.
30. van Haren, H., Oakey, N., Garrett, C. (1994). Measurements of internal wave band eddy fluxes above a sloping bottom. *Journal of Marine Research, 52*, 909–946.
31. Polzin, K. L., Toole, J. M., Ledwell, J. R., Schmitt, R. W. (1997). Spatial variability of turbulent mixing in the abyssal ocean. *Science, 276*, 93–96.
32. van Haren, H., & Gostiaux, L. (2012). Detailed internal wave mixing observed above a deep-ocean slope. *Journal of Marine Research, 70,* 173–197.
33. Klymak, J. M., & Moum, J. N. (2003). Internal solitary waves of elevation advancing on a shoaling shelf. *Geophysical Reseach Letters, 30,* 2045. https://doi.org/10.1029/2003GL017706.
34. Ferrari, R., Mashayek, A., McDougall, T. J., Nikurashin, M., & Campin, J. (2016). Turning ocean mixing upside down. *Journal of Atmospheric and Oceanic Technology, 46,* 2239–2261.
35. Tennekes, H., & Lumley, J. L. (1972). *A first course in turbulence* (p. 293). Cambridge, MA USA: MIT Press.
36. Thorpe, S. A. (1977). Turbulence and mixing in a Scottish loch. *Philosophical Transactions of the Royal Society of London A, 286,* 125–181.
37. Dillon, T. M. (1982). Vertical overturns: A comparison of Thorpe and Ozmidov length scales. *Journal Geophysical Research, 87,* 9601–9613.
38. Mater, B. D., Venayagamoorthy, S. K., St. Laurent, L., & Moum, J. N. (2015). Biases in Thorpe scale estimates of turbulence dissipation. Part I: Assessments from large-scale overturns in oceanographic data. *Journal of Physical Oceanography, 45*, 2497–2521.
39. Oakey, N. S. (1982). Determination of the rate of dissipation of turbulent energy from simultaneous temperature and velocity shear microstructure measurements. *Journal of Physical Oceanography, 12,* 256–271.
40. Osborn, T. R. (1980). Estimates of the local rate of vertical diffusion from dissipation measurements. *Journal of Physical Oceanography, 10,* 83–89.
41. Galbraith, P. S., & Kelley, D. E. (1996). Identifying overturns in CTD profiles. *Journal of Atmospheric and Oceanic Technology, 13,* 688–702.
42. Gargett, A. E., & Garner, T. (2008). Determining Thorpe scales from ship-lowered CTD density profiles. *Journal of Atmospheric and Oceanic Technology, 25,* 1657–1670.
43. Stansfield, K., Garrett, C., & Dewey, R. (2001). The probability distribution of the Thorpe displacement within overturns in Juan de Fuca Strait. *Journal of Physical Oceanography, 31,* 3421–3434.
44. van Haren, H., & Gostiaux, L. (2014). Characterizing turbulent overturns in CTD-data. *Dynamics of Atmospheres and Oceans, 66,* 58–76.
45. Fer, I., & Paskyabi, M. (2014). Autonomous ocean turbulence measurements using shear probes on a moored instrument. *Journal of Atmospheric and Oceanic Technology, 31,* 474–490.
46. Lohrmann, A., Hackett, B., & Røed, L. P. (1990). High resolution measurements of turbulence, velocity and stress using a pulse-to-pulse coherent sonar. *Journal of Atmospheric and Oceanic Technology, 7,* 19–37.
47. Rippeth, T. P., Williams, E., & Simpson, J. H. (2002). Reynolds stress and turbulent energy production in a tidal channel. *Journal of Physical Oceanography, 32,* 1242–1251.

48. Gemmrich, J. R., & van Haren, H. (2002). Internal wave band eddy fluxes in the bottom boundary layer above a continental slope. *Journal of Marine Research, 60,* 227–253.
49. van Haren, H., Laan, M., Buijsman, D.-J., Gostiaux, L., Smit, M. G., & Keijzer, E. (2009). NIOZ3: Independent temperature sensors sampling yearlong data at a rate of 1 Hz. *IEEE Journal of Oceanic Engineering, 34,* 315–322.
50. IOC, SCOR, & IAPSO. (2010). The international thermodynamic equation of seawater—2010: Calculation and use of thermodynamic properties. Intergovernmental Oceanographic Commission, Manuals and Guides No. 56, UNESCO, Paris, France, p. 196
51. van Haren, H., Greinert, J. (2016). Turbulent high-latitude oceanic intrusions – details of non-smooth apparent isopycnal transport West of Svalbard. *Ocean Dynamics, 66*, 785–794.
52. Gregg, M. C. (1989). Scaling turbulent dissipation in the thermocline. *Journal Geophysical Research, 94,* 9686–9698.
53. Alford, M. H., & Pinkel, R. (2000). Observations of overturning in the thermocline: The context of ocean mixing. *Journal of Physical Oceanography, 30,* 805–832.
54. Marmorino, G. O. (1987). Observations of small-scale mixing processes in the seasonal thermocline. Part II: Wave breaking. *Journal of Physical Oceanography, 17,* 1348–1355.
55. van Haren, H., & Gostiaux, L. (2009). High-resolution open-ocean temperature spectra. *Journal Geophysical Research, 114,* C05005. https://doi.org/10.1029/2008JC004967.
56. Munk, W. (1981). Internal waves and small-scale processes. In B. A. Warren & C. Wunsch (Eds.), *Evolution of physical oceanography* (pp. 264–291). Cambridge MA, USA: MIT Press.
57. Cyr, F., & van Haren, H. (2016). Observations of small-scale secondary instabilities during the shoaling of internal bores on a deep-ocean slope. *Journal of Physical Oceanography, 46,* 219–231.
58. Geyer, W. R., Lavery, A. C., Scully, M. E., & Trowbridge, J. H. (2010). Mixing by shear instability at high Reynolds number. *Geophysical Reseach Letters, 37,* L22607. https://doi.org/10.1029/2010GL045272.
59. van Haren, H. (2006). Nonlinear motions at the internal tide source. *Geophysical Reseach Letters, 33,* L11605. https://doi.org/10.1029/2006GL025851.
60. van Haren, H., Cimatoribus, A. A., Gostiaux, L. (2015). Where large deep-ocean waves break. *Geophysical Research Letters, 42*, 2351–2357, https://doi.org/10.1002/2015GL063329.
61. van Haren, H. (2017). Exploring the vertical extent of breaking internal wave turbulence above deep-sea topography. *Dynamics of Atmospheres and Oceans, 77,* 89–99.
62. van Haren, H., Gostiaux, L., Morozov, E., & Tarakanov, R. (2014). Extremely long Kelvin-Helmholtz billow trains in the Romanche Fracture Zone. *Geophysical Reseach Letters, 41,* 8445–8451. https://doi.org/10.1002/2014GL062421.
63. Thorpe, S. A. (1973). Experiments on instability and turbulence in a stratified shear flow. *Journal of Fluid Mechanics, 61,* 731–751.
64. Klaassen, G. P., & Peltier, W. R. (1985). The onset of turbulence in finite-amplitude Kelvin-Helmholtz billow. *Journal of Fluid Mechanics, 155,* 1–35.
65. Woods, J. D. (1968). Wave-induced shear instability in the summer thermocline. *Journal of Fluid Mechanics, 32,* 791–800.
66. Moum, J. M., Farmer, D. M., Smyth, W. D., Armi, L., & Vagle, S. (2003). Structure and generation of turbulence at interfaces strained by internal solitary waves propagating shoreward over the continental shelf. *Journal of Physical Oceanography, 33,* 2093–2112.
67. van Haren, H., Cimatoribus, A. A., Cyr, F., & Gostiaux, L. (2016). Insights from a 3-D temperature sensors mooring on stratified ocean turbulence. *Geophysical Reseach Letters, 43,* 4483–4489. https://doi.org/10.1002/2016GL068032.
68. Thurnherr, A. M., Richards, K. J., German, C. R., Lane-Serff, G. F., & Speer, K. G. (2002). Flow and mixing in the Rift Valley of the Mid-Atlantic Ridge. *Journal of Physical Oceanography, 32,* 1763–1778.
69. Melet, A., Hallberg, R., Legg, S., & Nikurashin, M. (2014). Sensivity of the ocean state to lee-wave driven mixing. *Journal of Physical Oceanography, 40,* 900–921.
70. Cacchione, D. A., & Drake, D. E. (1986). Nepheloid layers and internal waves over continental shelves and slopes. *Geo-Marine Letters, 16,* 147–152.

71. Cyr, F., van Haren, H., Mienis, F., Duineveld, G., & Bourgault, D. (2016). On the influence of cold-water coral mound size on flow hydrodynamics, and vice-versa. *Geophysical Reseach Letters, 43,* 775–783. https://doi.org/10.1002/2015GL067038.
72. Tsuchiya, M., Talley, L. D., & McCartney, M. S. (1992). An eastern Atlantic section from Iceland southward across the equator. *Deep Sea Research Part A. Oceanographic Research Papers, 39,* 1885–1917.

Deep-Ocean Tides in the South-West Indian Ocean: Comparing Deep-Sea Pressure to Satellite Data

Leo R. M. Maas, Borja Aguiar-González and Leandro Ponsoni

Introduction

Tides have been known to humanity for thousands of years [6]. This knowledge was based on coastal observations of sea surface elevation and associated currents. It was not until the seventeenth century that tides were perceived as very long surface gravity waves that propagate not only along ocean boundaries but that also cross oceans [19]. Observations on the coherent nature of tides, over scales of many hundreds of kilometers or more, started in the nineteenth century [42]. These observations were fraught with surprise, as the tidal field in the North Sea turned out to be much more complex than expected. Its spatial amplitude and phase pattern could be rationalized only by anticipating the existence of amphidromic points: phase singularities at nodal points where the sea surface elevation field vanishes, and around which crest and troughs rotate [42]. Later, the appearance of amphidromes was recognized to be a generic feature in rotating fluids, owing their existence to the Coriolis force that acts transverse to the current. Amphidromic points are now a standard property in the

L. R. M. Maas (✉)
Institute for Marine and Atmospheric Research Utrecht, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands
e-mail: L.R.M.Maas@uu.nl

L. R. M. Maas
NIOZ Royal Netherlands Institute for Sea Research, Texel, The Netherlands

B. Aguiar-González
Department of Ocean Systems Sciences, NIOZ Royal Netherlands Institute for Sea Research, Utrecht University, P.O. Box 59, 1790 AB Den Burg, Texel, The Netherlands
e-mail: aguiar@nioz.nl

L. Ponsoni
Georges Lemaître Centre for Earth and Climate Research (TECLIM), Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium
e-mail: leandro.ponsoni@uclouvain.be

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_12

description of tidal patterns, both in the world oceans as well as in coastal regions. In fact, phase singularities, such as amphidromes, have been found in other settings too: e.g. in the inertial wave patterns that appear in homogeneous-density fluids contained in uniformly-rotating, fully-enclosed cubes [4, 5, 18, 21, 27], as well as in other branches of physics, like quantum mechanics [3].

In theoretical models, the simplest version of an amphidrome is obtained in a tidal channel when two counter-propagating Kelvin waves form nodal points [29, 38]. Amphidromes also appear in semi-analytical models of rotationally-modified surface gravity waves, both in half-infinite channels and in rectangular basins [32, 33, 38], as well as in numerical models of more realistically-shaped basins [15], although they are then mixed with Poincaré waves.

Due to the advent of satellite altimetry, however, spatial patterns of combined Kelvin and Poincaré waves could be directly observed in whole-field measurements of the Ocean [7], albeit at three kilometer resolution along and still more crudely across the satellite track. It revealed that on the open ocean, tides are of small-amplitude, a few decimeter at most, in line with the strength of the tidal forcing. Amplification is restricted to coastal regions, first of all due to the exponential character of Kelvin waves, induced by the Coriolis force, but occurring especially when a bay co-oscillates, as when in resonance with a basin mode [10].

Away from the equator, open ocean tides combine propagating, coastally-trapped Kelvin waves with Poincaré waves. Together these display amphidromic points that are traversed mostly in cyclonic (anti-clockwise) sense in the Northern Hemisphere, and *vice versa* in the Southern Hemisphere. However, at times, ocean tides instead display *sloshing*, and are then dominated by standing Poincaré waves that show a behaviour complementary to that of Kelvin waves, exhibiting off-coastal elevation maxima, betrayed by near-uniform phases. One example for the semidiurnal M2 tide is seen in Fig. 1 in the Mozambique Channel. It has a maximum and nearly uniform phase in the narrows. A more striking example, though, can be found in the central Indian Ocean, where the tidal maximum at 75° E, 19° S lies close to a phase saddle point at 71° E, 9.6° S, the point where *two* 260-degrees-phase-lines cross each other (see star in Fig. 1). The uniformity in phase near this local tidal maximum implies that over a large area the semidiurnal tide simply rises and sinks in unison. Notice the 4 amphidromes surrounding this mid Indian Ocean maximum, some lying actually on land (like on Sri Lanka and Madagascar).

The altimetry-derived barotropic tides are based on the Oregon State University Tidal Inversion Software (OTIS) [12]. In this work, we use the TPX08-ATLAS product, which is provided for 9 tidal constituents (M2, S2, N2, K2, K1, O1, P1, Q1 and M4), with a horizontal resolution of 1/30° (http://volkov.oce.orst.edu/tides), see Table 1, although we will not consider M4 tides here.

What will be relevant for the present study is the behaviour of the tide in the Mozambique Channel, as well as along a transect at 23° S, east of Madagascar, where our instruments are located. Taking a cursory look at these two regions, Fig. 1 shows that we may expect quite different behaviour. In the Mozambique Channel, the semidiurnal M2 tide—usually the strongest tidal component—has a standing character, testified by its near-uniform phase, and an amplified tidal elevation which

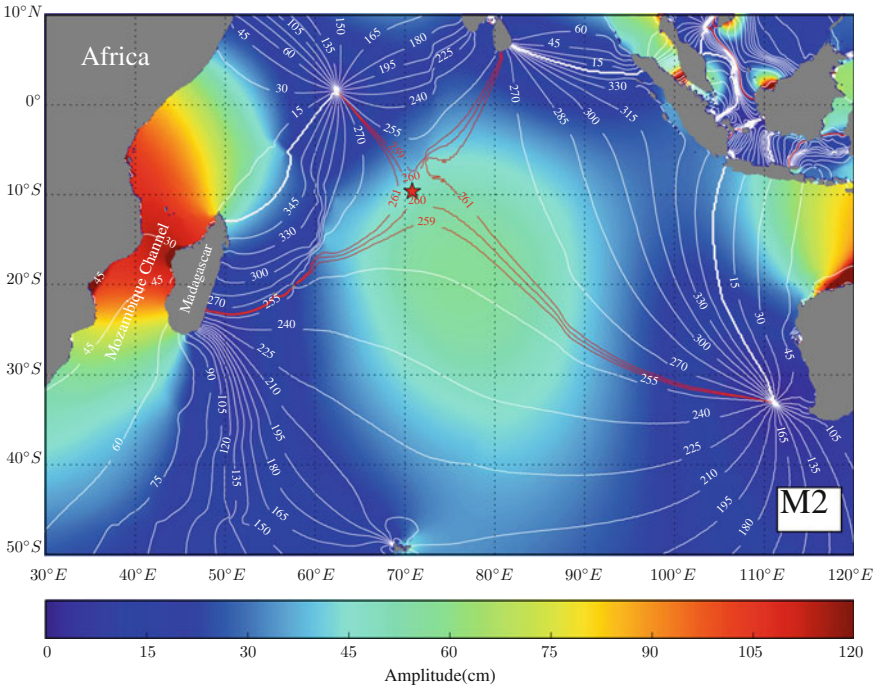


Fig. 1 M2 semidiurnal tidal map obtained from altimetry assimilated by OTIS showing amplitude (color, in cm) and phase with respect to Greenwich Mean Time (GMT) (lines, in °). Tides propagate around amphidromes in the direction of increasing phase. Notice tidal maximum and nearly uniform phase, both in the central Indian Ocean as well as in the Mozambique Channel

Table 1 List of tidal constituents with their corresponding period. Admittance for Mozambique Channel (MC) and East Madagascar (EM) is discussed in section "Coherent Surface and Internal Tides"

Tidal constituent	Period (h)	Admittance MC	Admittance EM
Q1	26.87	1.32	0.47
O1	25.82	1.20	0.28
P1	24.07	0.96	0.25
K1	23.93	0.92	0.26
N2	12.66	4.22	0.61
M2	12.42	5.07	0.51
S2	12	5.88	0.89
K2	11.97	5.85	0.80

reaches a magnitude of about 1.2 m. This betrays a local resonant phenomenon. In stark contrast, East of Madagascar, the M2 tide is close to the nodal point situated on Madagascar. Its surface elevation amplitude is diminished to 10 cm and it shows large phase differences along the coast.

Altimeter observations are assimilated into ocean tide models [37]. These models benefit from the additional constraints provided by the presence of 'sea truth' in the form of deep-ocean tidal observations of pressure. Such contact measurements are useful because altimeter observations do not fully cover the ocean surface. Deformations in the tidal potential, especially near coasts, as well as the presence of coherent yet weak internal tides cause inaccuracies in the determination of surface tidal amplitudes and phases.

For this reason, an effort has previously been put into comparing altimetry-derived tidal elevations to directly-measured tides [34], for which Bottom-Pressure Records (BPRs) have been used. The latter are insensitive to mooring line motions, and may thus provide reliable estimates of the surface tides. However, it turns out that long-term BPRs at depths in excess of 500 m are rare. There appear to be only about 80 deep-ocean pressure records that are at least 100 km apart and that last longer than one year [34]. In the South-West Indian Ocean there are none. Previous shelf and deep-sea bottom pressure measurements [24, 41] have also been used to investigate high-frequency contributions by linear and nonlinear (nonhydrostatic) internal waves, that fall outside the scope of the present investigation.

In the present chapter, we will determine tidal amplitudes and phases of BPRs at moorings deployed at depths in excess of 500 m along two transects in the South-West Indian Ocean, on either side of Madagascar. This region, described in section "Measurement Sites and Instrumentation", is not covered in the previous set of deep-ocean pressure measurements. Altimetry-derived tidal constants for this area are therefore possibly influenced by the presence of the Madagascar and Mozambique coasts and shelves.

The pressure recorders are situated at both a suite of 'bottom locations' (for the purpose of this study defined to be within 75 m from the bottom) as well as at 'mid-depth' sensors within the water column, typically positioned at 500 m below the surface or deeper. Details on these instruments are given in section "Measurement Sites and Instrumentation".

In section "Time Series of Pressure Measurements", all long-term pressure observations are converted to equivalent sea surface displacements and are subjected to a harmonic analysis, yielding tidal amplitudes and phases. Prior to performing the Harmonic Analysis, events showing blow-down of instruments, especially of those deployed far above the bottom, have been eliminated. In section "Coherent Surface and Internal Tides", the harmonic amplitudes and phases derived from BPRs are compared to each other and to altimetry-derived tidal constants. Subsequently, section "Deep Versus Mid-depth Pressure-Measurements of Tides" compares tidal constants derived from BPRs to those obtained from pressure recorders within the water column. In general, within the water column, surface tides explain less of the observed variance in the pressure time series. Part of the tidal variability is due to internal tides. Away from the bottom, amplitudes and phases of the coherent internal

tide—the internal tide that stays phase-locked to the surface tide over the analysed period—can be estimated by subtracting the surface tidal amplitudes and phases, found at BPRs along the same mooring. Incoherent internal tides change amplitude and phase due to changing environmental conditions (stratification and mean-flow structure). Low frequency features as eddies and currents, apart from being responsible for blow-down of mid-water column instruments, also make the determination of long-period tides difficult, which can therefore not be reliably extracted. It is for this reason that we here focus on the eight most prominent diurnal and semidiurnal tidal components only. Conclusions are drawn in section "Conclusions".

Measurement Sites and Instrumentation

The data we use come from two transects in the South-West Indian Ocean:

(1) the narrowest part of the Mozambique Channel (MC), studied in several 'Long-term Ocean Climate Observations' (LOCO) projects, [36, 40], and
(2) East of Madagascar (EM), at 23° S, studied in the 'INdian ATlantic EXchange in present and past climate' (INATEX) project [30, 31].

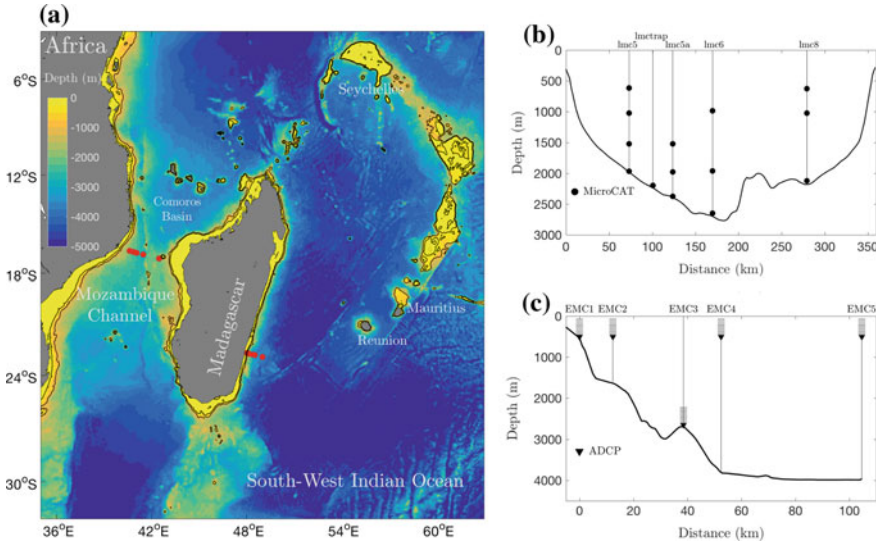


Fig. 2 **a** Bathymetric map (in meters depth) and geographical locations (red dots) of the mooring sites in the narrows of the Mozambique Channel (lmc5, lmctrap, lmc5a, lmc6, lmc8) and off South-East Madagascar (EMC1–EMC5). Detail of the mooring array in **b** the Mozambique Channel and **c** South-East Madagascar

Red dots in Fig. 2a indicate mooring locations in these two regions. Instruments on the moorings provide long-term records of velocity, pressure and hydrographic data. Previous analysis of MC data concerned long-term velocity and temperature records (instruments not shown in Fig. 2b). These were dominated by low-frequency variations due to the yearly, southward passage of four to seven MC-wide eddies [40]. These eddies seem to match the width of the Channel in size but are most pronounced on the Western side and could as well be interpreted as cross-channel propagating Rossby waves [13]. The velocity records also showed evidence of the persistent presence of highly intermittent internal tides with currents ranging from a few up to 10 cm/s [23].

East of Madagascar, the dominant motion is marked by a western boundary current, the East Madagascar Current (EMC). This along-coastal, southward-directed mean-flow eventually feeds the Agulhas Currents further south. At 23° S a steady train of westward propagating cyclones and anticyclones drive the EMC's main band of variability [31].

Here, we focus on data from pressure recording instruments, carried by SBE37-SM MicroCATs in the Mozambique Channel and by upward-looking Acoustic Doppler Current Profilers (ADCPs—RDI Workhorse Long Ranger 75 kHz) off Eastern Madagascar. The positions of these instruments on the mooring lines are indicated in Fig. 2b and c and in Tables 2 and 3.

Tidal amplitudes and phases are obtained from observed pressure time series with T_TIDE, a Harmonic Analysis package [28]. Using a least-squares method, this package determines amplitudes and phases for a given number of tidal constituents whose frequencies are known from celestial mechanics. Here we pay attention to the restricted set of frequencies listed in Table 1. We exclude the very long period tides (such as the Solar Semi-Annual, SSA, and monthly, MSM), as well as higher harmonics, such as M4, and will concentrate on the most prominent diurnal and semidiurnal frequencies, such as M2.

To eliminate the influence of low-frequency phenomena and associated blow-downs the pressure data were handled in the following way: (1) The mean pressure is subtracted. (2) Resulting pressures are divided by the product of density, ρ, and acceleration of gravity, g, yielding an equivalent surface 'displacement' time series, where a positive surface elevation is presented as a negative displacement. (3) The resulting displacement series was detrended. (4) A High-Pass (HP) filter was applied. (5) Blow-down events were eliminated to avoid estimating artificially large tidal amplitudes. (6) The resulting (gappy) time series was subjected to Harmonic Analysis, using T_TIDE applied to periods of one year length, attributing its resulting amplitudes and phases to its centre time. (7) To see the slow evolution of harmonic amplitudes and phases, this harmonic analysis was performed over subsequent one year periods, stepping forward by one day at a time, over a period of 9 to 21 months (a procedure that we call Moving T_TIDE, or MT_TIDE).

In step (2) an 'equivalent surface displacement' is defined for the following reason. While we measure pressure, both near the bottom, as well as within the water column, which varies due to hydrostatic effects (weight of fluid column) and non-hydrostatic effects (vertical accelerations), subtracting the mean pressure, its time-

Table 2 Summary of MicroCATs deployed in the narrows of the Mozambique Channel during LOCO 4, LOCO 5 and LOCO 6, and used in this study. See Fig. 2 for details of the instrument mooring design and location

Moor. site Depth (m)	Position	LOCO period	Instrument depth (m)	Meters above bottom	Total days	Dates mm/dd/yy
lmc5 1992 m	16° 38.86 S 40° 36.82 E	LOCO 4	617	1375	647	01/31/08–11/08/09
		LOCO 4	1020	972	613	01/31/08–10/05/09
		LOCO 4	1520	472	647	01/31/08–11/08/09
		LOCO 4	1966	26	647	01/31/08–11/08/09
lmctrap 2241 m	16° 42.40 S 40° 51.06 E	LOCO 5	2200	41	775	12/20/09–02/03/12
lmc5a 2402 m	16° 45.64 S 41° 3.76 E	LOCO 6	1520	882	863	02/06/12–06/18/14
		LOCO 6	1975	427	863	02/06/12–06/18/14
		LOCO 4	2375	27	220	01/31/08–09/07/08
lmc6 2691 m	16° 52.26 S 41° 28.72 E	LOCO 6	985	1706	863	02/06/12–06/18/14
		LOCO 6	1960	731	863	02/06/12–06/18/14
		LOCO 6	2650	41	863	02/06/12–06/18/14
lmc8 2199 m	17° 6.24 S 42° 28.7 E	LOCO 4	625	1574	689	01/31/08–12/20/09
		LOCO 4	1025	1174	689	01/31/08–12/20/09
		LOCO 4	2125	74	689	01/31/08–12/20/09

varying part may *increase* due to (1) suppression of the mooring's buoys and instruments, (2) surface elevations, or (3) isopycnal uplifts (internal waves). We eliminate suppression-related pressure increases, and divide the demeaned pressure by ρg, so as to be able to interpret it as surface displacement where, recall a positive 'displacement' corresponds to a free surface depression. However, these time-varying pressure perturbations may also be due to internal waves (isopycnal displacements). Thus, a given pressure perturbation, interpreted as an *equivalent* surface elevation (or surface displacement), could also be interpreted as an isopycnal displacement, but, due to the much weaker density contrast, of much larger vertical extent.

Table 3 Summary of the ADCPs deployed off South-East Madagascar during the INATEX project. See Fig. 2 for details

Mooring site	Position	Inst. mean depth (m)	Local depth (m)	Total days	Dates mm/dd/yy
EMC1	22° 41.879 S 48° 1.006 E	500	504	909	10/06/10–04/02/13
EMC2	22° 43.484 S 48° 8.013 E	500	1603	924	10/06/10–04/17/13
EMC3	22° 46.950 S 48° 22.860 E	2600	2654	924	10/06/10–04/17/13
EMC4	22° 48.689 S 48° 30.827 E	500	3799	924	10/06/10–04/17/13
EMC5	22° 55.483 S 49° 0.435 E	500	3959	924	10/05/10–04/16/13

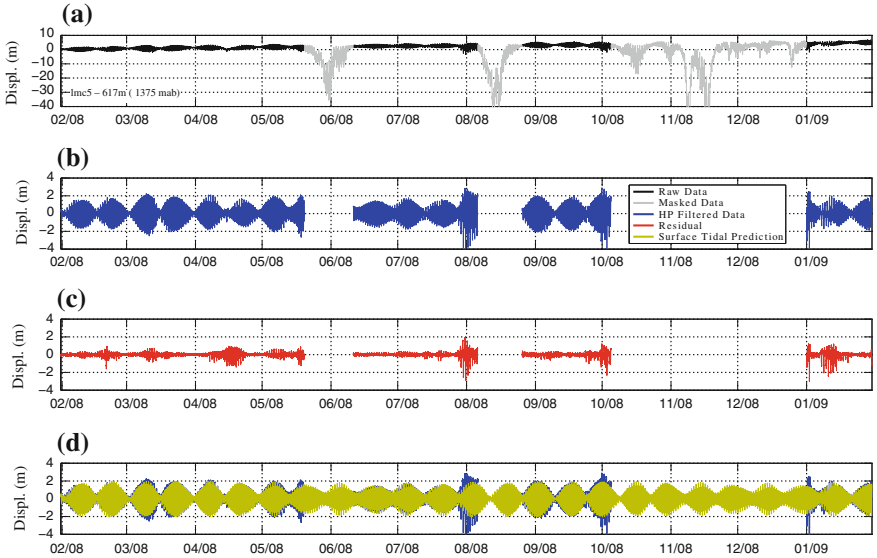


Fig. 3 Time series of pressure, converted to displacement, at lmc5 (at an average depth of 617 m), showing **a** demeaned and detrended time series of vertical displacements, **b** High-Passed filtered data excluding blow-downs, **c** residual field, not captured by Harmonic Analysis, **d** tidally back-predicted field (yellow) superimposed on HP filtered data (blue) and now also covering the gaps. Notice larger scale in (**a**)

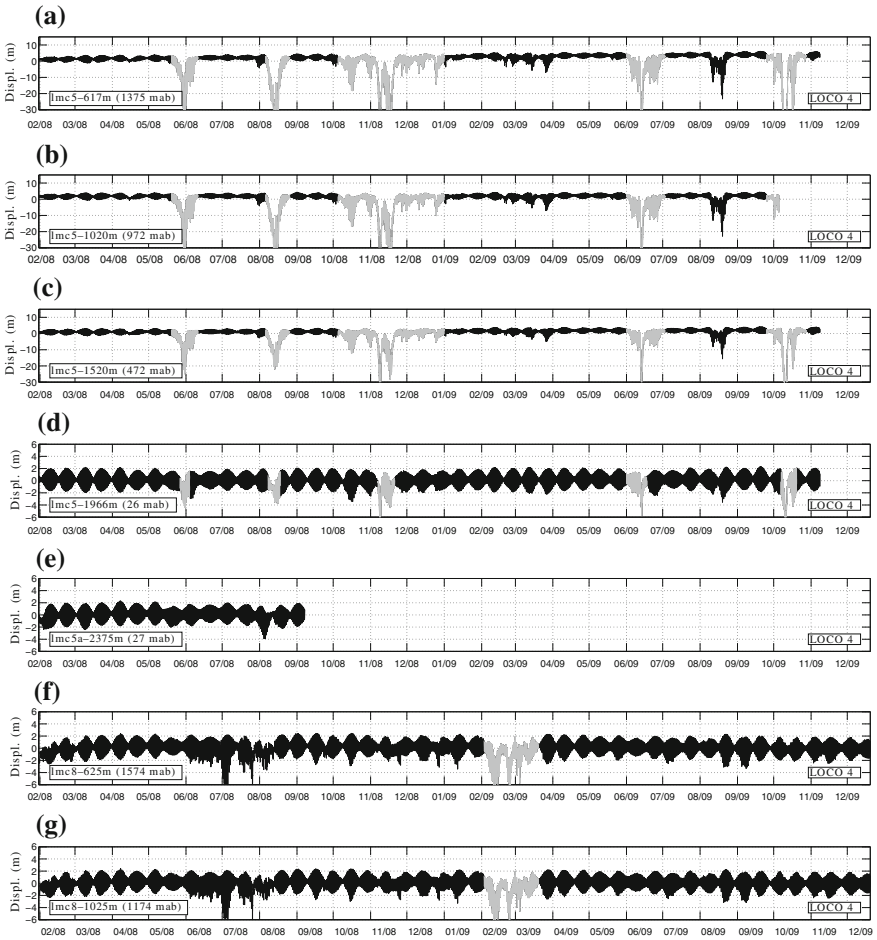


Fig. 4 Time series of equivalent surface displacement as measured by pressure recorders on MicroCATs during **a**–**h** LOCO 4, **i** LOCO 5, **j**–**n** LOCO 6. Time is given in months/year, denoted as mm/yy; November 2009 is e.g. indicated as 11/09. Notice differences in vertical scales and sampling periods. Labels indicate the site name, mean instrument depth and its height in meters-above-bottom (mab). Data in grey are identified as periods affected by blow-down and are, therefore, replaced by NaNs. Further details in Table 2

As High-Pass filter we use a fourth-order Butterworth filter with zero-phase response and cut-off frequency $c^{-1} \times \omega_c$, where $\omega_c = 0.83$ cpd, corresponding to a period of 29.07 h. The band width parameter, $c = 1.25$, was chosen to be narrow enough to capture closely the lowest frequency of interest, ω_c, while being wide enough to avoid filter 'ringing' [1].

An example, illustrating steps (2)–(6), is shown in Fig. 3. Note that the residual signal (red curve) still contains (near) tidal variations, also during periods when

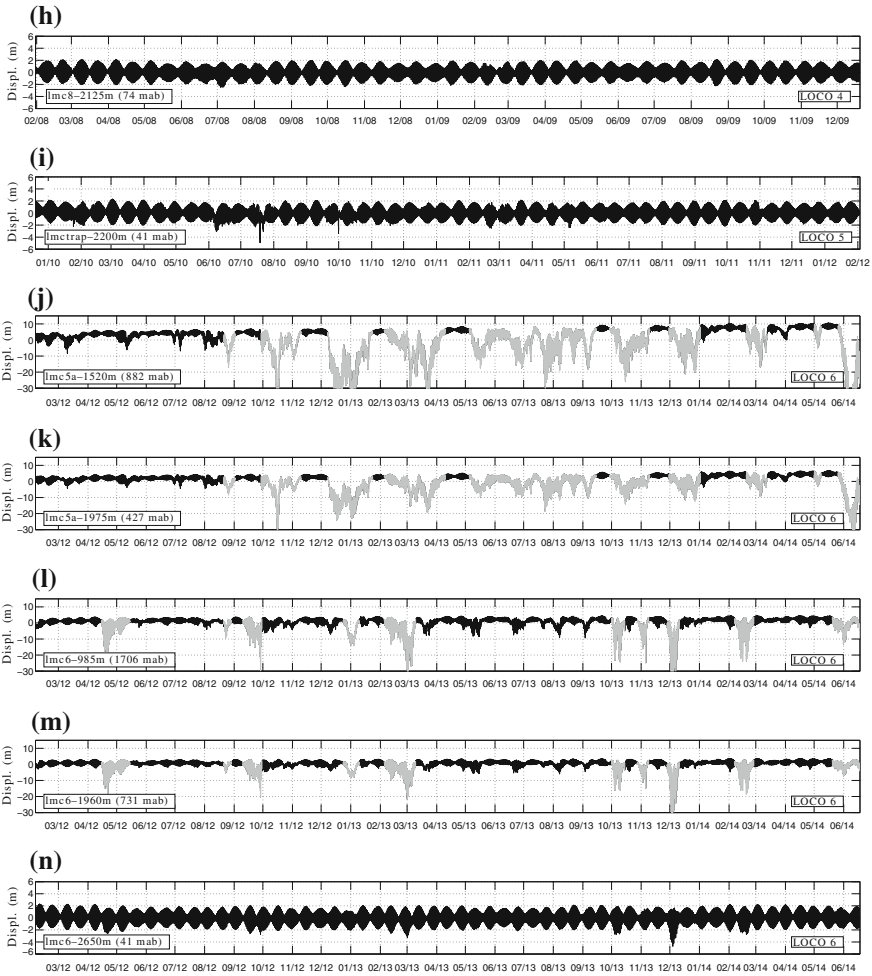


Fig. 4 (continued)

blow-downs do not occur. These come from the spectral cusps, unambiguously representing incoherent internal tides.

In the MC case we visually inspected all time-series, see Fig. 4. We compare neighbouring instruments to elucidate whether the disruption of the 'string of pearls' was actually due to a blow-down event or to some other phenomenon. We were particularly cautious in not removing those phenomena since they introduce natural variability we do not want to remove from our time-series. See for example panel (f) in Fig. 4. Compare events at mooring lmc8 during July and August of 2008 (not marked as NaN) with those during February and March of 2009 (marked as NaN). At first sight one could say both disruptions of the 'string of pearls' are due to a blow-down event, but taking a closer look one realises they are not. The marked events involve

a low-frequency, deep and noticeable excursion while during the non-marked events large-amplitude waves are of high-frequency.

In the EM data, blow-downs were removed by replacing downward displacements in excess of 10 m by NaNs.

Whether large tidal expressions during blow-downs are artificial, resulting from drag forces on buoys exerted by tidal currents superimposed on low-frequency, eddy-related currents, or genuine, expressing internal tides trapped within eddies [2, 17], is not clear and will not be settled here. We note, that HP-filtering also filters out the slow (large-scale) atmospheric pressure contribution to the BPRs. But, as [34] noticed, BPRs may still contain atmospheric tides that would not show up in altimetric measurements, and can be one of the causes for discrepancies between the two.

Time Series of Pressure Measurements

Mozambique Channel

At moorings located along the MC transect in Fig. 2b, Fig. 4 shows time series of both bottom as well as mid-water column pressure recorders. Panels a-h are from LOCO 4 (periods listed in Table 2), panel i from LOCO 5, and panels j-n from LOCO 6. We combine measurements from different LOCO periods to account for the best spatial and temporal coverage of instrumentation across the channel (both along the sea bottom and throughout the water column). The 'string of pearls', visible in all records, represents the spring-neap tidal cycle, produced by superposition of the two dominating surface tidal components, M2 and S2. This supports the notion that amplitude and phase of this tidal interference pattern is about the same for each of these records. Analysis of semidiurnal and diurnal tidal amplitudes and phases, discussed below, will confirm this in more detail.

Most instruments higher up in the water column frequently suffer from large vertical excursions, sometimes reaching apparent displacements of thirty meter or more (see Fig. 4, panels a–c and j–m). Interestingly, despite the long duration of these events (typically lasting a week or longer) these excursions decrease in size with depth. This is odd, as their long duration suggests them to be due to large scale features that should, if these were of a barotropic nature as the surface tide, penetrate throughout the whole water column. The weakening of these vertical displacements with depth suggests these events are instead due to frequent blow-down of buoys, caused by drag forces due to eddy and tide-related currents, that we aim to eliminate. The buoys apparently often fail to keep their mooring lines from pointing straight upwards. During previous observational periods in MC, tilt sensors indeed revealed tilts of over 30° that were moreover tidally modulated with another 5° [23]. Even BPRs sometimes suffer from vertical displacements of 6 m or more that are obviously not produced by surface elevations (see Fig. 4d).

Recall that the displacements defined above, present scaled pressure perturbations that can have different origins. BPRs, however, are not (or very marginally) subject to blow-downs and the interpretation of pressure variations as being due to surface tidal displacements is unambiguous, although we need to remember that an upward sea surface elevation corresponds to a negative displacement. As Fig. 4f–h show, blow-downs seem to be nearly absent at all depths on the most Eastern mooring, lmc8. Together with the bottom-pressure records from two other periods

Table 4 Mean and standard deviation of the variance of the total HP-filtered signal, the tidal signal, and the percent variance explained by the tides predicted from harmonic analyses of MC pressure sensor measurements. Values were computed by averaging all estimates from MT_TIDE. The standard deviation for each estimate is also indicated. Notice that the BPR at lmc5a lasts only seven months making it impossible to establish the evolution of its amplitude and phase. Recall that the mean and trend were removed before the harmonic analysis

Mooring site – Instrument depth (m) Meters above bottom (mab)	Total Var. (m^2)	Tidal Var. (m^2)	Tidal Var. (%)
lmc5 – 617 m (1375 mab)	1.4369 ± 0.3655	1.1491 ± 0.1711	81.9529
lmc5 – 1020 m (972 mab)	1.2545 ± 0.2763	1.0681 ± 0.1425	86.5204
lmc5 – 1520 m (472 mab)	1.0815 ± 0.1708	0.9809 ± 0.0986	91.4791
lmc5 – 1966 m (26 mab)	0.8823 ± 0.0100	0.8744 ± 0.0097	99.1022
lmctrap – 2200 m (41 mab)	0.7923 ± 0.0106	0.7710 ± 0.0166	97.3134
lmc5a – 1520 m (882 mab)	1.2189 ± 0.2256	1.0639 ± 0.1546	88.3847
lmc5a – 1975 m (427 mab)	1.0412 ± 0.1270	0.9825 ± 0.1076	94.6039
lmc5a – 2375 m (27 mab)	0.8669	0.8604	99.2438
lmc6 – 985 m (1706 mab)	1.5220 ± 0.1595	1.0794 ± 0.1243	70.8571
lmc6 – 1960 m (731 mab)	1.2055 ± 0.0929	1.0133 ± 0.0802	84.0482
lmc6 – 2650 m (41 mab)	0.9243 ± 0.0165	0.9160 ± 0.0132	99.1108
lmc8 – 625 m (1574 mab)	0.9321 ± 0.0546	0.8158 ± 0.0239	87.7736
lmc8 – 1025 m (1174 mab)	0.9160 ± 0.0474	0.8206 ± 0.0224	89.7537
lmc8 – 2125 m (74 mab)	0.8293 ± 0.0073	0.8236 ± 0.0076	99.3154

(panels d, e, i and n) these measurements confirm the dominance of surface tides in the time-varying pressure field. At other locations, however, mid-water column pressure signals frequently suffer from severe blow-down. This is a common feature for pressure measurements within the water column, and for this reason often only BPRs are used in the validation of altimetry-derived harmonic constants [34]. In section "Coherent Surface and Internal Tides", we will therefore first look at BPRs, but, since we also want to analyse pressure records higher up in the water column (section "Deep Versus Mid-depth Pressure-Measurements of Tides"), we subject all our measurements to the following procedure.

The amplitudes and phases resulting from MT_TIDE are assigned to the central time within each one-year time interval. Those of the most important semidiurnal and diurnal components are discussed in the next sections. The tidal variance estimate, obtained by averaging the variance of subsequent T_TIDE analyses, and its relative contribution to the total HP-variance, are shown in Table 4. The total and tidal variance are all O (1 m^2). For all moorings, a weak increase of the total HP-variance with height above the bottom is noticed. Except for the most Eastern mooring, lmc8, the tidal variance also increases with height above the bottom. This may be related to the stronger presence of coherent internal tides higher up in the water column, while being weaker or absent in the bottom boundary layer. Yet, the relative contribution of the tides to the total variance always decreases towards the surface. This may be attributed to the stronger (oceanic and atmospheric) mesoscale activity near the surface, introducing more phenomena influencing the pressure variability. Note that the tidal variance in Table 4 that T_TIDE computes for each one year period is based on many more (High-Passed) tidal components than those listed in Table 1, including higher harmonics, the so-called shallow water tides [28]. But in the following we will focus on just the major eight (semi)diurnal components.

East Madagascar

East of Madagascar, tides are much weaker than in the Mozambique Channel, at least near the bottom where they measure free surface displacements of O(10 cm) for the main tidal component, M2. Figure 5a, c show BPR (black line), tidal back-prediction (grey) and residual (red) at moorings EMC1 and EMC3, respectively (see Table 3). Notice that at these two instruments the total apparent surface displacement is restricted to about 1 m. The other three instruments in Fig. 5b, d, e, all far above the bottom, indicate unrealistically large vertical excursions that are a hundred times larger (notice the difference in scale). These are obviously due to mooring motions, resulting in occasional blow-down.

Harmonic Analysis of the BPRs at moorings EMC1 and EMC3 shows that the tides give a modest, yet genuinely tidal contribution of O(8–32%) to the total variance (Table 5). Indeed, this estimate of tidal variance looks reliable in Fig. 5, when comparing the tidal string of pearls (grey) to the variance carried by the residual (red). As to the question *why* these values are so low (compared to those in MC),

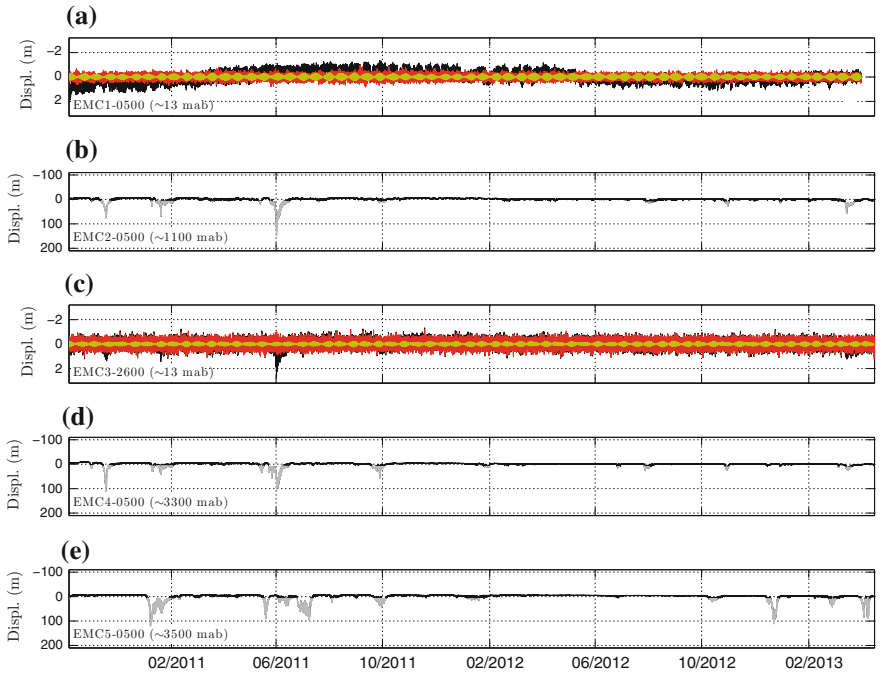


Fig. 5 EM time series (sampled from Oct/2010 to Apr/2013) converted to displacements (black line). Grey and yellow lines in panels **a** and **c** represent the tidal prediction and residual time series, respectively

Table 5 As Table 4 for the EM time series

Mooring site	Total Var. (cm^2)	Tidal Var. (cm^2)	% Tide
EMC1	0.048 ± 0.007	0.015 ± 0.002	31.97 ± 1.16
EMC2	1.372 ± 0.369	0.298 ± 0.090	21.80 ± 3.15
EMC3	0.083 ± 0.001	0.007 ± 0.001	8.54 ± 0.78
EMC4	1.103 ± 0.441	0.299 ± 0.136	26.58 ± 1.95
EMC5	2.317 ± 1.062	0.370 ± 0.205	15.47 ± 3.01

we are still guessing. Fourier spectra, show that after removal of the coherent tides there are still energetic peaks present in the cusps, surrounding the main tidal peaks. But, inspection of time series of BPRs clearly shows that EMC3 has many more supra-tidal oscillations than EMC1, which are corroborated by a higher level of the HF part of the Fourier spectrum (not shown). Whether this reflects the presence of shorter period surface waves, possibly related to the hill at which EMC3 is situated (see Fig. 2c), higher instrumental noise, internal waves, or has some other cause is not clear.

For the other instruments, tidal variance contributes a similar amount to the total variance, O(15–25%), but this does not necessarily represent surface or coherent internal tides. As testified by the much larger *relative* standard deviations in Table 5 of all variance estimates at EMC2, 4 and 5 when compared to those of the BPRs at EMC1 and 3, intermittent internal tides overshadow the surface and coherent internal tides. We will come back to this in section "Deep Versus Mid-depth Pressure-Measurements of Tides".

Coherent Surface and Internal Tides

Harmonic amplitudes and phases of surface tides should not depend on the particular observational time window from which they are determined (provided the observational period is long enough). By definition also the coherent internal tides have stable amplitudes and phases, albeit differing from those associated with the surface tide. Amplitude and phase stability can be used as a criterion to distinguish genuine from 'false' tidal signals (i.e. pressure variations due to tidal-current induced depression of instruments), or on a more positive note, can be used to separate the combined surface and coherent internal tides from incoherent internal tides. The moving version of the harmonic analysis performed by T_TIDE (MT_TIDE) allows us to make this separation.

Mozambique Channel Harmonic Constants

The results of running MT_TIDE for the Mozambique Channel pressure records are presented in Figs. 6 and 7 for semidiurnal and diurnal frequencies, respectively. The fluctuations in, especially amplitude, that occur for pressure records within the water column will be discussed in section "Deep Versus Mid-depth Pressure-Measurements of Tides", although it should be noticed here that tidal amplitudes are not necessarily always larger higher up in the water column, compared to the very stable bottom amplitudes. See for instance in Fig. 6 the M2 amplitude at lmc8. Phases are very stable at nearly all locations, amplitudes only near the bottom. It is significant that the least stable estimates and the largest M2 tidal amplitudes are found at the mid-channel mooring lmc5, especially at the pressure sensors that are highest above the bottom. At the three 'bottom' locations (less than 75 m away from the bottom), M2 amplitudes are around 113 cm. Phase differences are generally small. For M2 the phase centers around 37°. For the second most important component, S2, the amplitude is about 60 cm, and its phase about 79°. The other two semidiurnal components, N2 and K2, both reach amplitudes of O (17 cm).

Diurnal components (Fig. 7) also have quite stable amplitudes and phases. Amplitudes of O (2–7 cm) are smaller than those of the semidiurnal tides. The spreading in phase estimates is often larger than that of the semidiurnal components. Table 6

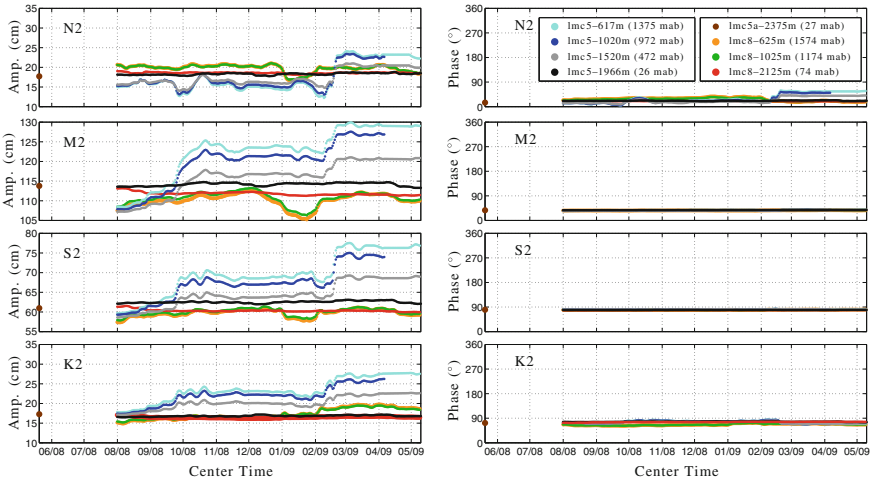


Fig. 6 Amplitudes (left, cm) and phases (right, °) of semidiurnal tidal constituents in MC during LOCO 4 obtained from MT_TIDE. Labels indicate the site name, mean instrument depth, and instrument height in meters-above-bottom (mab). The circle on the left vertical axis indicates the estimates for lmc5a obtained from T_TIDE applied to seven months of data only

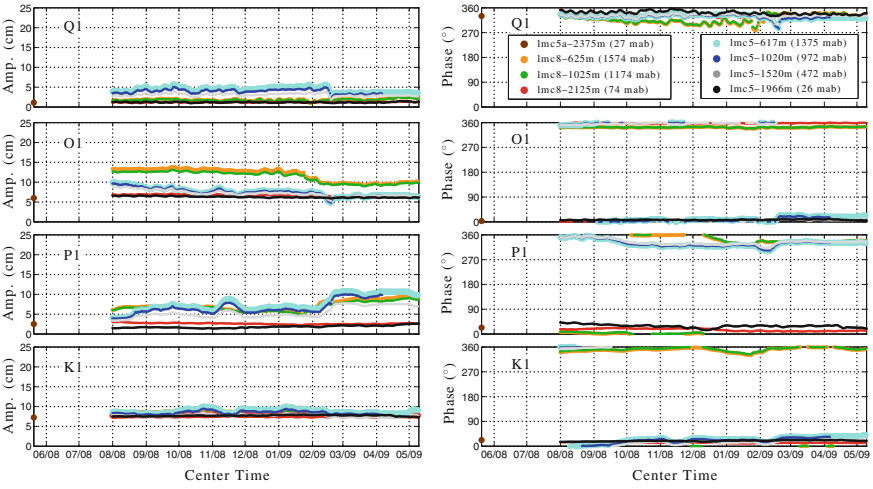


Fig. 7 Same as Fig. 6 for diurnal tidal constituents

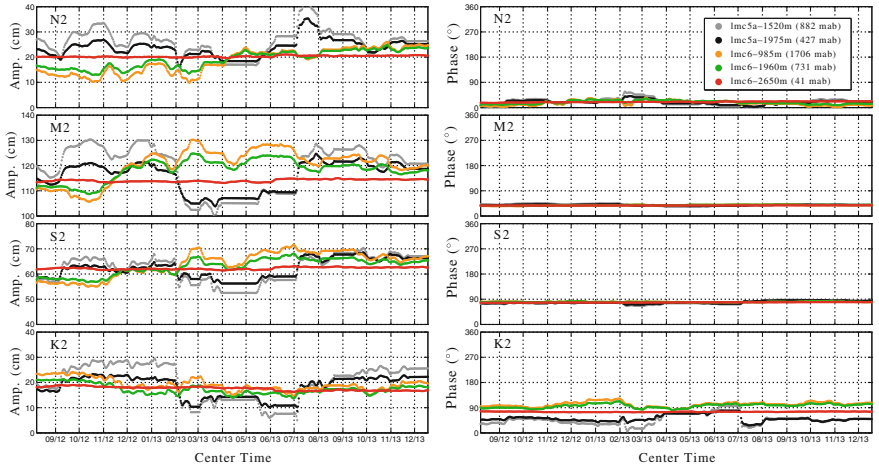


Fig. 8 Amplitudes (left, cm) and phases (right, °) of semidiurnal tidal constituents during LOCO 6 obtained from MT_TIDE. Labels indicate the site name, mean depth, and depth in meters-above-bottom (mab)

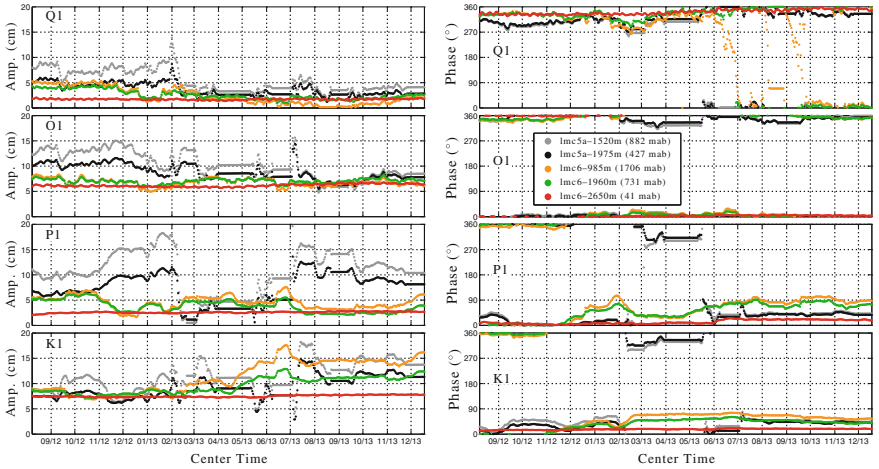


Fig. 9 As Fig. 8 for the diurnal frequencies

summarizes the MT_TIDE average amplitudes, phases and standard deviations of these eight tidal components.

Similar amplitudes and phases during the LOCO 6 period are shown in Figs. 8 and 9. The single BPR that was present in this period, 41 m above the bottom, at lmc6 (red curves), again has very stable tidal amplitudes and phases, in line with those obtained for the LOCO 4 period.

Table 6 Mean amplitude (cm) and phase (°) from MT_TIDE applied to the Mozambique Channel bottom pressure recorders. Instrument deployment height is indicated as meters-above-bottom (mab). Columns are ordered from West to East

Tidal Const.	lmc5–1966 m (26 mab)		lmctrap–2200 m (41 mab)		lmc5a–2375 m (27 mab)		lmc6–2650 m (41 mab)		lmc8–2125 m (74 mab)	
	Amp. (cm)	Phase (°)	Amp. (cm)	Phase (°)	Amp. (cm)	Phase (°)	Amp. (cm)	Phase (°)	Amp. (cm)	Phase (°)
Q1	1.2 ± 0.1	342.6 ± 6.1	1.6 ± 0.2	308.6 ± 13.9	1.1	330.2	1.7 ± 0.1	341.9 ± 8.6	1.2 ± 0.1	332.1 ± 4.7
O1	6.3 ± 0.2	7.2 ± 1.2	7.0 ± 0.2	352.1 ± 4.3	6.0	2.8	6.2 ± 0.3	2.5 ± 2.4	6.7 ± 0.2	358.3 ± 1.0
P1	1.8 ± 0.3	28.6 ± 5.3	3.0 ± 0.4	22.1 ± 7.3	2.5	23.2	2.6 ± 0.1	12.6 ± 7.5	2.6 ± 0.2	15.9 ± 4.4
K1	7.7 ± 0.1	19.5 ± 2	6.0 ± 0.2	349.5 ± 4.9	7.2	22.2	7.5 ± 0.2	17.4 ± 2.0	7.9 ± 0.1	12.5 ± 1.1
N2	18.3 ± 0.2	21.5 ± 0.9	17.1 ± 0.8	15.1 ± 1.4	17.7	14.9	20.3 ± 0.3	20.2 ± 1.6	18.6 ± 0.2	19.8 ± 0.6
M2	114.2 ± 0.4	38 ± 0.3	106.9 ± 0.9	36.4 ± 0.4	113.8	37.3	114.1 ± 0.5	37.7 ± 0.3	111.8 ± 0.4	37.4 ± 0.3
S2	62.5 ± 0.3	79.1 ± 0.2	57.7 ± 0.5	79.2 ± 0.7	61.0	80.0	62.2 ± 0.4	78.7 ± 0.7	60.3 ± 0.3	78.7 ± 0.3
K2	16.8 ± 0.2	77.1 ± 0.6	14.7 ± 0.8	73.3 ± 1.2	17.3	72.7	17.6 ± 0.8	75.3 ± 1.0	16.5 ± 0.2	76.0 ± 1.1

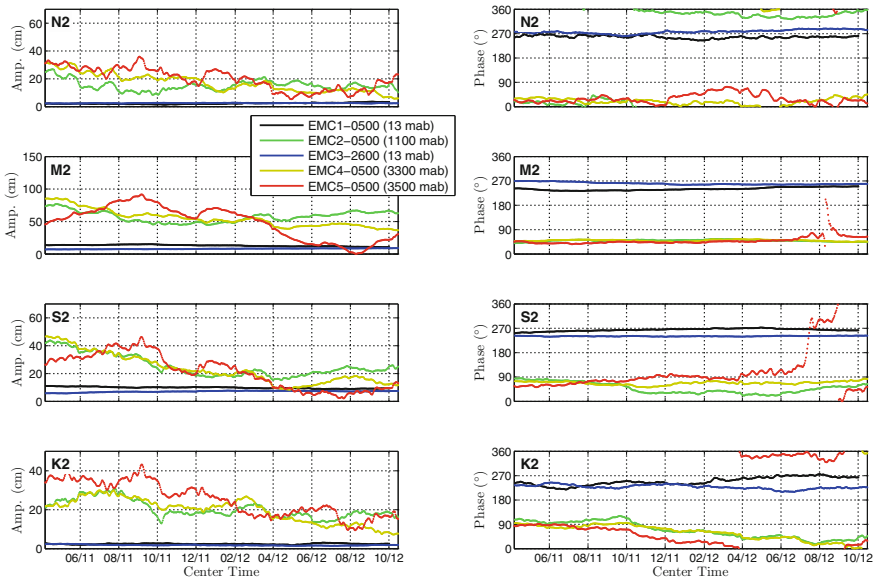


Fig. 10 Amplitudes (left, cm) and phases (right, °) of semidiurnal tidal constituents in EM, obtained from MT_TIDE. Labels indicate the site name and mean depth

East Madagascar Harmonic Constants

The slow-time evolution of semidiurnal amplitudes and phases for the East Madagascar moorings, after standard HP-filtering and blow-down removal, is shown for the four major semidiurnal frequencies in Fig. 10. These panels show that: (1) tidal amplitudes of O (10 cm) and phases are very stable at near-bottom instruments (EMC1 and EMC3), represented by a single average value (Table 7) together with a variance estimate (Table 5), (2) M2 tidal amplitudes at 'mid-depth' (EMC2, 4 and 5) are very high, O (50–100 cm), (3) tidal amplitudes at mid-depth depend on the precise computational interval for which T_TIDE is run—the time around which the one-year time series is centred, (4) for bottom instruments—representative of barotropic tides—phases of all semidiurnal components are nearly the same (approximately 250°), (5) compared to the phase of these BPRs, mid-depth pressure series are often roughly in anti-phase (on average, about 160° phase difference).

Similar observations can be made for the diurnal components, Fig. 11. Phases of bottom instruments (EMC1 and 3, black and blue lines respectively) are again stable despite very small O (1 cm) amplitudes, albeit now showing differences from one diurnal component to the next. Mid-depth instruments at EMC2, 4 and 5 still have very large amplitudes, but phases sometimes change rapidly. Phase changes of 180° (such as that for O1 in September 2012), are likely indicative of the passage of a node of an internal tide's vertical elevation field, that separates rising drom depressing isopycnal surfaces.

Table 7 Mean amplitude (cm) and phase (°) from MT_TIDE applied to the East Madagascar pressure recorders. Notice that recorders at EMC1 and EMC3 represent BPRs, while at the other moorings they are positioned 500 m below the surface. Columns are ordered from West to East

Tidal constituent	EMC1		EMC2		EMC3		EMC4		EMC5	
	Amp. (cm)	Phase (°)	Amp. (cm)	Phase (°)	Amp. (cm)	Phase (°)	Amp. (cm)	Phase (°)	Amp. (cm)	Phase (°)
Q1	0.6 ± 0.2	277.5 ±124.5	7.0 ± 4.1	175.9 ±29.6	0.7 ± 0.1	290.4 ± 11.0	4.3 ± 1.3	245.3 ± 76.0	5.1 ± 2.2	235.4 ± 93.4
O1	1.9 ± 0.6	35.0 ±13.9	4.1 ± 0.9	189.9 ±150.1	2.2 ± 0.1	23.5 ± 3.1	2.6 ± 1.4	133.9 ± 110.5	6.4 ± 2.3	114.4 ± 102.6
P1	1.2 ± 0.1	87.8 ±12.4	4.2 ± 2.8	185.1 ± 70.8	0.5 ± 0.3	98.4 ± 55.9	3.2 ± 1.7	229.5 ± 61.6	10.5 ± 5.7	84.5 ± 85.3
K1	2.7 ± 0.1	101.4 ± 4.1	11.3 ± 2.5	78.7 ± 18.8	2.6 ± 0.4	111.1 ± 7.7	9.3 ± 2.8	73.4 ± 8.1	16.0 ± 5.2	100.0 ± 13.9
N2	2.3 ± 0.5	257.8 ± 5.4	15.2 ± 3.6	233.3 ± 53.6	2.5 ± 0.1	277.2 ± 7.0	17.1 ± 6.5	38.5 ± 76.3	20.0 ± 7.9	34.1 ± 39.8
M2	13.0 ± 1.2	240.5 ± 5.4	57.5 ± 8.4	49.3 ± 3.6	8.0 ± 0.6	260.6 ± 4.5	54.8 ± 13.0	49.9 ± 2.9	48.6 ± 25.9	50.0 ± 17.1
S2	9.9 ± 0.6	264.3 ± 4.6	25.4 ± 7.7	49.0 ± 19.7	7.1 ± 0.5	240.9 ± 1.1	23.1 ± 10.5	67.8 ± 6.5	21.4 ± 12.4	99.7 ± 66.0
K2	2.4 ± 0.3	249.4 ± 14.9	20.3 ± 4.5	67.7 ± 30.1	1.8 ± 0.3	228.3 ± 7.9	19.6 ± 6.3	65.3 ± 47.4	25.3 ± 8.7	138.0 ± 135.1

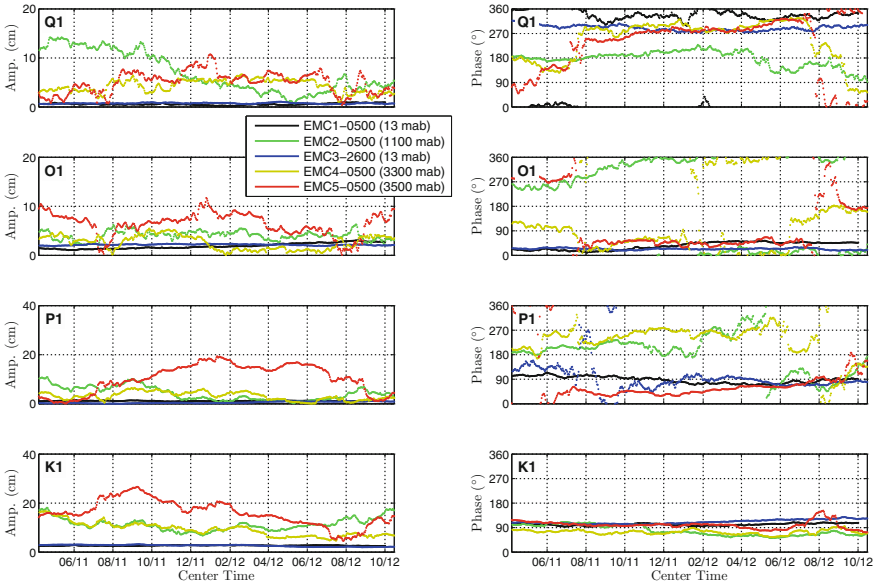


Fig. 11 Same as Fig. 10 for diurnal frequencies

Long-term averages, obtained from MT_TIDE, are shown in Table 7. They show the stability of the BPR-derived tidal amplitudes at EMC1 and 3: a signal-to-noise ratio (SNR), defined by the ratio of standard deviation to amplitude, of O (10%). This contrasts with the SNR for mid water column pressure records at EMC2, 4 and 5, which is O (30–50%). Of course, they also show that bottom pressure fluctuations, converted to surface displacement, have smaller amplitudes than those within the water column, were they again point at the presence of internal tides (see section "Deep Versus Mid-depth Pressure-Measurements of Tides").

Bottom-Pressure Versus Altimetry Derived Tides

The first and third rows of Fig. 12 display diurnal and semidiurnal tidal amplitudes, A, and phases, ϕ, from Table 6 of the five BPRs in MC. These are the amplitude and phases of an equivalent surface elevation field computed from the BPRs which are here presented in terms of the real and imaginary components $A(\cos\phi, \sin\phi)$ of the complex vectors $A\exp(i\phi)$. Similar 'harmonic vectors', obtained from satellite altimetry (OTIS), are shown in the second and fourth rows.

The comparison between BPRs and altimetry (OTIS) vectors in the Mozambique Channel, where tides are large, is excellent, especially for the semidiurnal components (last two rows). The difference vectors are O (5%) in magnitude compared to those of the BPRs, for which reason, in section "Deep Versus Mid-depth

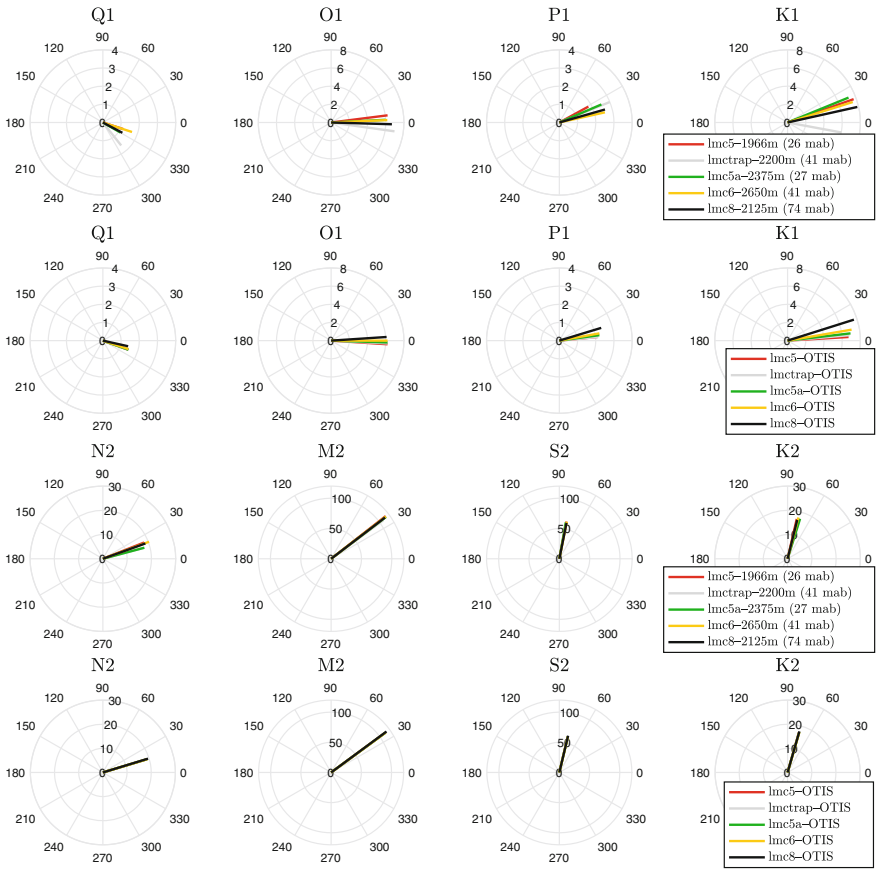


Fig. 12 Polar plots showing horizontal variation of amplitudes (cm) and phases (°) of diurnal (top two rows) and semidiurnal tidal constituents (bottom two rows) in MC as seen from bottom instruments and satellite altimeter (OTIS) data. The legend indicates the site name and mean depth. Notice differences in scale

Pressure-Measurements of Tides", BPR-derived harmonic vectors have been subtracted from those higher up in the water column. We subtract these BPR signals because, as they agree with altimetric tides, we can take them as representative of surface tides. Therefore, its subtraction from harmonic signals obtained from pressure records higher up in the water column leads to the signal of coherent internal tides. Semidiurnal tides appear spatially synchronised, and thus appear as standing waves. The diurnal components (first two rows) show more variations at the bottom. Those derived from altimetry (OTIS) show a clear and consistent eastward phase increase (from lmc5 towards lmc8), and also a slight increase in amplitude. By contrast, the BPRs show in general a slight westward phase propagation, although somewhat more erratic. Especially the bottom pressure recorder at lmctrap behaves, as we shall see later, somewhat differently for most frequencies. Comparing to harmonic

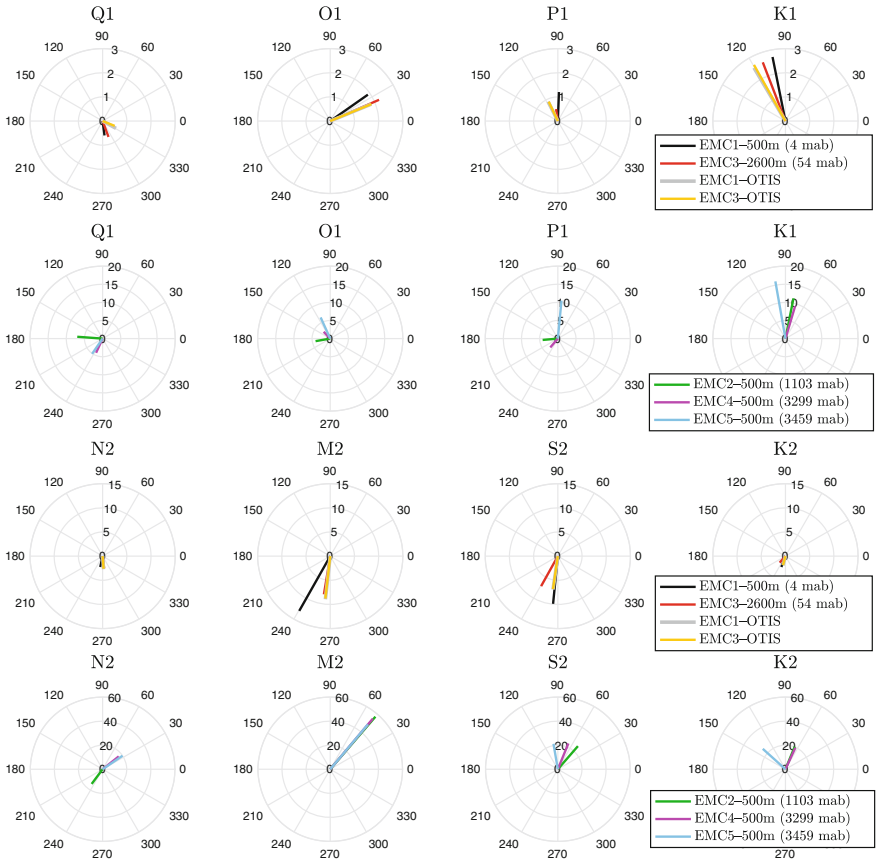


Fig. 13 Amplitudes (cm) and phases (°) of diurnal (top row) and semidiurnal (third row) tidal constituents at sites EMC1 and EMC3 as seen from BPRs and OTIS data. Amplitudes (cm) and phases (°) of diurnal and semidiurnal tidal constituents within the water column, at sites EMC2, EMC4 and EMC5 (second and fourth rows). The legend indicates the site name, mean depth and meters-above-bottom. Note that here polar plots share the same scale only within the same row

vectors at lmctrap during other LOCO periods (and using other pressure sensors—not shown), however, confirms the amplitudes and phases found at lmctrap during LOCO 5, although all seem to differ from those obtained at other MC locations. This may bear perhaps evidence of a relatively strong coherent internal diurnal tide that varies over shorter spatial scales than the forcing.

On the transect East of Madagascar, differences of the BPR-derived semidiurnal tides with the altimetric vectors, shown in the third row of Fig. 13, are larger than in MC, possibly because the background, low-frequency fields are relatively stronger and hence the determination of tidal amplitudes and phases less accurate. At EMC1, the OTIS magnitude of the M2 tide differs by O (50%) of that in bottom pressure. The spatial uniformity of the altimetry-derived harmonic constants in EM, that was

mentioned above, may however have been artificial, due to the extrapolation applied to altimetry-derived tidal constants on approach of the shoreline. Coastal proximity inhibits direct satellite altimeter measurements of surface elevations. BPRs show a decrease of surface tidal amplitude with increasing distance to the coast (compare EMC1 to EMC3) of the largest two semidiurnal frequencies (M2 and S2). This is in line with their presence as a Kelvin wave, trapped along the East Madagascar slope, which is expected to propagate northwards.

East of Madagascar, surface diurnal tides shown by BPRs in the first row of Fig. 13 reach amplitudes of O (1–3 cm) and compare well with altimetry derived (OTIS) harmonic vectors. The harmonic vectors at the other, mid-water column pressure sensors, however, are all much larger than those at the bottom, indicating their presence as a strong coherent diurnal internal tide that will be further discussed in section "Deep Versus Mid-depth Pressure-Measurements of Tides".

Credo of Smoothness

In the open ocean and in coastal regions, tides represent the response to body or boundary forces, respectively. In the open ocean, tides are due to the gravitational attraction by sun and moon, modulated by the motion of these celestial bodies, together with the motion and rotation of the earth [29]. Coastal regions (and inland seas and lakes) are too small to be able to respond to tidal forces directly. Coastal tides result due to the tides present at their ocean boundaries, via co-oscillation [10, 14]. Of course, the nature of this response does not only depend on the strength with which tides are present in the tidal potential or at the sea's boundary. Tidal response also depends on geometrical aspects of ocean or coastal basins, which may lead to tidal amplification when resonating with a basin's eigenfrequency, or to its suppression due to choking when in anti-resonance [20, 39]. Viewing the ocean or coastal response to tidal forcing as being determined mainly by the proximity of any of the tidal frequencies to any of the eigenfrequencies of these fluid basins, a certain similarity in the spatial distribution of tides of nearly similar frequency should not come as a surprise. This view was epitomized in the catchy phrase 'credo of smoothness' [26] and is clearly born out in the tidal altimetry fields, both for the semidiurnal and diurnal frequency bands, see Figs. 14 and 15.

At each tidal frequency, the response of the ocean to tidal forcing is given by an amplitude ratio and a phase difference between its locally observed elevation amplitude and Greenwich phase with those present in the tidal potential. Viewing the response of the ocean to tidal forcing as that of an oddly-shaped, damped mechanical resonator, one expects to see grossly similar features for slightly differing tidal frequencies, frictional effects smoothing out any sharp changes. During passage of such a resonance, 180 degree phase changes may still occur, but not abruptly. Looking at

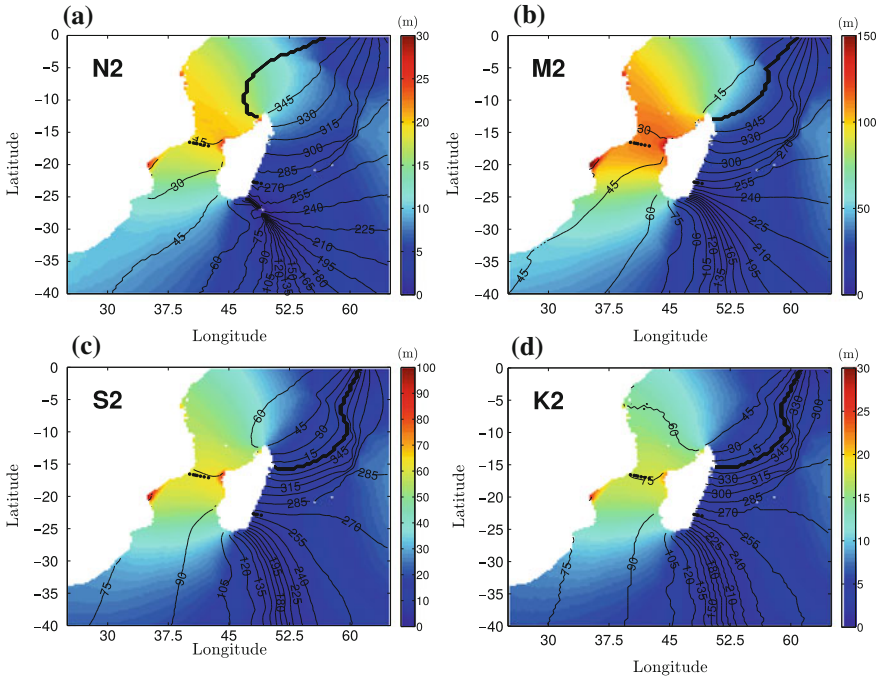


Fig. 14 Spatial amplitude (cm) and phase (°) pattern in the South-West Indian Ocean determined from OTIS by means of satellite altimetry data for semidiurnal tidal components. Mooring locations are indicated by dots

the altimetry-derived amplitudes and phases, Figs. 14 and 15, it is indeed striking to see the resemblance of the amplitude and phase distributions of the four semidiurnal tidal frequencies considered. The pressure observations, presented here, fully support these findings, except that they show that in the vicinity of Madagascar, a Kelvin wave enhancement may take place that may not be directly visible in the altimetry-derived East Madagascar amplitude fields.

By contrast, the tide in the Mozambique Channel is nearly uniform in phase, and suggests a local, geometry-induced resonance. Satellite altimetry (Fig. 1) indeed shows the M2-phase to become uniform a little North of the Mozambique Channel transect, at a phase of about 30°.

A time series of observed sea surface elevation, $\zeta(t)$, can be written as a sum, $\zeta = \sum_j Z_j \sin(\Omega_j t + \phi_j)$, of contributions at tidal frequency, Ω_j, of amplitude Z_j and phase ϕ_j. The ability of a particular geographical region to resonate can be measured by computing the admittance, i.e. the ratio of the observed free surface tidal amplitudes, Z_j, at a certain location, to their amplitudes, $\bar{Z}_j$, as present in the tidal potential. Here we ignore the phase shift involved and evaluate this amplitude ratio only. The tidal potential, W, is for over 98% captured by the first non-vanishing, second-degree term, W_2 [29]. In the equations of motion, this potential is usually expressed as an

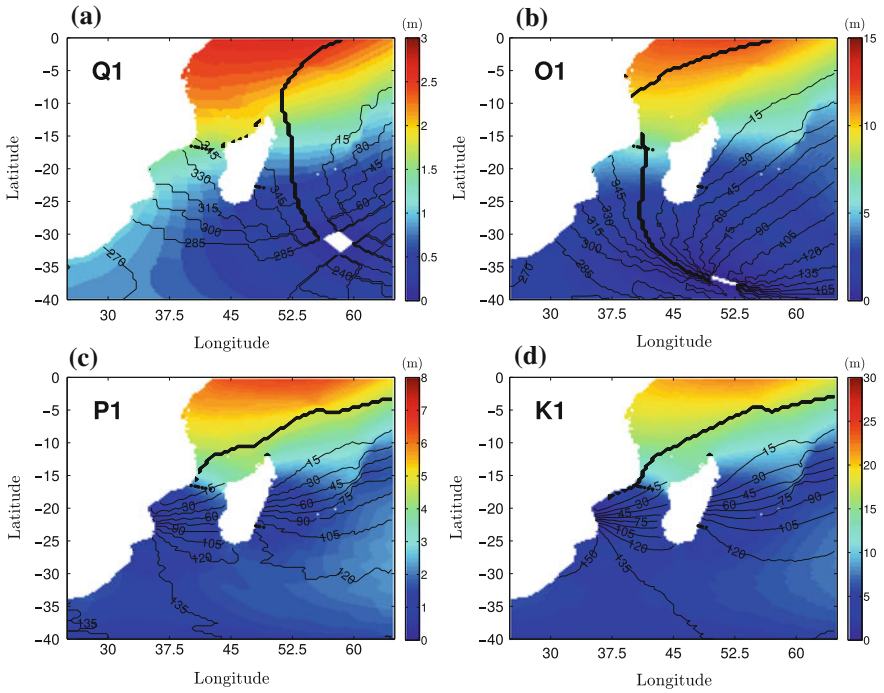


Fig. 15 Same as Fig. 14 for diurnal tidal components

equilibrium sea level, towards which the ocean tends, $\bar{\zeta} \equiv W_2/g$. This equilibrium tide is a nonlinear function of space and time, determined by declinations and rotation rates of Earth and celestial body, and their distance. It can likewise be expanded in a Fourier series and, at a certain geographical location, is expressed as the sum over tidal frequencies of a product of functions of latitude θ, $G_i(\theta)$, of diurnal ($i = 1$) and semidiurnal ($i = 2$) origin, the Doodson constant—an overall amplitude factor—a tidal potential coefficient, C_j, and a trigonometric function of time t and phase angle χ_j (dependent on the location's longitude and orbital parameters of moon and sun):

$$\bar{\zeta} = \sum_j \bar{Z}_j \ \sin(\Omega_j t + \chi_j), \ \text{where} \ \ \bar{Z}_j \equiv C_j G_i D/g.$$

Here, the Doodson constant, $D = 3GMR^2/4d^3$, where G denotes the universal gravitational constant, M the mass of the celestial body, R the radius of the earth, and d the distance of the centre of the Earth to the center of the gravitating body. For the moon, the Doodson constant, divided by gravitational acceleration (g), yields $D/g = 26.75$ cm. The latitude functions of diurnal and semidiurnal tides are given by $G_1 = \sin(2\theta)$ and $G_2 = \cos^2\theta$, respectively. Since the Mozambique Channel and East Madagascar transects have fairly uniform tidal

amplitudes each, we take their average values as observed tidal amplitudes, Z_j, in computing their admittance, $Z_j/\bar{Z}_j$. The tidal potential coefficients, C_j, of the eight tidal frequencies Ω_j: (Q1, O1, P1, K1, N2, M2, S2, K2), are given by [29] and read $C_j = (0.072, 0.377, 0.176, 0.530, 0.174, 0.908, 0.423, 0.115)$. With these coefficients, along the transects in the Mozambique Channel ($\theta \approx 16.75°$ S) and East of Madagascar ($\theta \approx 22.90°$ S), the admittance of these tidal components, $Z_j/\bar{Z}_j$, has been determined, shown in the second and third column of Table 1, respectively. Admittances indeed vary fairly smoothly, particularly in the Mozambique Channel where semidiurnal amplitudes are all resonantly amplified, peaking with an amplification factor of about 5.9, at a frequency in between S2 and K2.

The response at the diurnal frequencies in the Mozambique Channel is about the same as in the tidal potential (admittance close to 1), albeit picking up at frequencies below the lowest diurnal frequency considered here, Q1.

The tides derived from BPRs along the transect East of Madagascar are all suppressed compared to their presence in the tidal potential. The semidiurnal components are weak because of the proximity of an amphidromic point close to the transect; the diurnal components vary at a larger scale, and amplify towards the equator.

Deep Versus Mid-depth Pressure-Measurements of Tides

BPRs of tides show harmonic constants to be very stable. This is surprising as internal tides might well be present at the bottom too. These could lead to a modulation of tidal amplitudes and phases. Consider e.g. a two-layer ocean with upper and lower layer depths h_1 and h_2 and corresponding densities ρ_1 and ρ_2, respectively. Then the hydrostatic pressure at the bottom is given by a stationary part $\rho_1 g h_1 + \rho_2 g h_2$ and a variable part due to elevations of free surface, ζ, or interface, η, given by $g\rho_1\zeta + g(\rho_2 - \rho_1)\eta$. Expressed as elevation, by dividing by $g\rho_1$, this perturbation pressure reads $\zeta + \delta\eta$. It shows that pressure perturbations due to large interface elevations $\eta \gg \zeta$ might still be moderate due to weakness of the density contrast, $\delta \equiv \rho_2/\rho_1 - 1 \ll 1$. When the ocean is continuously-stratified, higher internal modal structures may appear within the sea. The net pressure perturbation at the bottom due to internal displacements can then be less than those within the water column as isopycnal elevations at one depth can be compensated by depressions at another. Judging from the near-bottom stability of the tides, internal tides are weak at the bottom either for the above reason, or because their presence is precluded by the presence of a thick bottom boundary layer at the top of which internal waves reflect, higher up in the water column.

Observations from pressure recorders, mounted higher in the water column do show the presence of internal tides. In fact, they often dominate the pressure signal, especially East of Madagascar, see the second and fourth row of Fig. 13, showing amplitudes at intermediate depths four times those found at the bottom. Mooring lmc8, on the East side of Mozambique Channel, is exceptional in showing hardly any

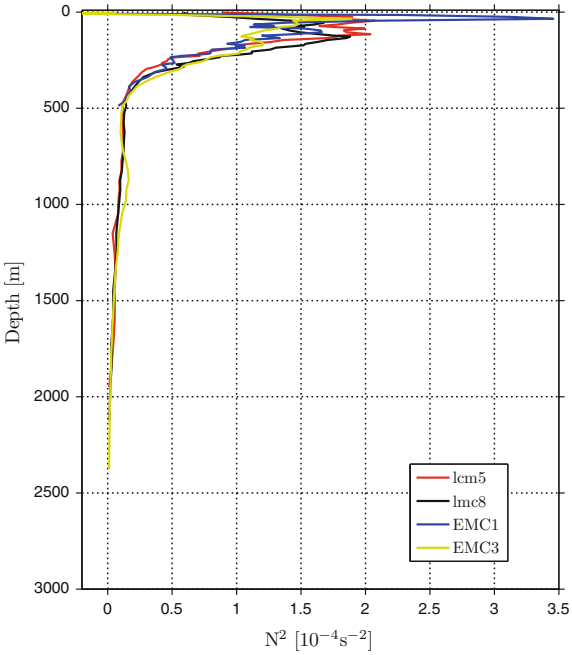


Fig. 16 Buoyancy-squared profiles obtained from CARS09 for Mozambique Channel (lmc5 and lmc8) and East of Madagascar (EMC1 and EMC3)

modulation of the harmonic constants, to the effect that the 'string of pearls'—the spring-neap tidal cycle—is retrieved also at mid-depth levels (see Fig. 4f–h). Other sites in the Mozambique Channel, further to the West, are more strongly affected by eddies, internal tides and blow-down of instruments. These mask the barotropic signal within the water column, and make the surface tide visible and reliably detectable near the sea bottom only.

Stratification

The ocean is a stratified medium, and so is the region on either side of Madagascar. Buoyancy frequency squared (N^2) profiles, were extracted for the regions of interest from the CSIRO Atlas of Regional Seas (CARS2009). This merges data from both research vessel instrument profiles and autonomous profiling (e.g., Argo floats), and is distributed on a 1/2° degree horizontal grid. For every grid point, the product also provides a correction for the annual cycle which allows users to create profiles for any date/season of the year. Figure 16 shows four typical N^2-profiles on either side of Madagascar. Both regions are characterised by a strong, shallow (seasonal) pycnocline (with a sharp maximum at a depth of about 50m, especially at EMC1). EM displayed also a deeper, second permanent pycnocline, at 850 m depth. Both

density profiles are stably-stratified down to the bottom, thus allowing for internal tides to propagate.

Mozambique Channel Coherent Internal Tides

Internal tides are ubiquitous [25], and in the Indian Ocean reach vertical amplitudes of 80 m or more, especially near Mascarene Ridge, close to the Seychelles (Fig. 2), to the North-East of Madagascar [16, 25]. There, internal tides take the shape of solitons, having large vertical isopycnal excursions of O (100 m), that find marked surface expressions [9]. In Mozambique Channel, internal solitons have also been observed south of our transect, near Sofala bank at 20.5° S where tides are stronger than in the MC narrows [8]. The clear patterns of waves in their satellite observations have been attributed to horizontally-propagating solitons, trapped to the pycnocline, that were generated either at the shelf edge or by upward-propagating internal wave beams impinging on the seasonal pycnocline from below. The latter beam itself was identified as the bottom reflection of a previously downward-propagating internal wave beam that was also produced at the shelf edge.

Along our Mozambique Channel transect, indications of mainly incoherent tides were extracted from ADCP and current meter records during the Agulhas Current Source EXperiment (ACSEX), a precursor of the LOCO projects used in the present study [23]. That study shows the internal tides to be quite intermittent. Their spatial structure, however, seems to accord with internal tides taking the shape of relatively broad beams, O (100 km) wide, also found in their numerical model. These beams follow nearly parabolic paths owing to the decrease in buoyancy frequency with increasing depth and corresponding steepening of internal wave paths, seen in Fig. 16.

In the present study, BPR-derived harmonic amplitudes and phases are the only ones that can confidently be attributed to surface tides. The magnitude of the coherent internal tides, displayed in Fig. 17, were obtained by subtracting these BPR-derived surface tidal harmonic vectors. As the semidiurnal surface tide is much larger, O (113 cm), than the coherent internal tide, having 'equivalent surface displacements' O (8 cm), only the surface tidal harmonic vector directions are shown in this figure (black sticks), their magnitudes, A_{surf}, being listed in each panel's legend. Recall that the apparently small magnitudes of the coherent internal tidal vectors represent a pressure perturbation, here expressed as an equivalent surface displacement, that can also be interpreted as a much larger internal, isopycnal displacement. The latter will magnify the surface displacement, crudely speaking by a factor δ^{-1}. For δ of O (10^{-3}), this brings us back to an estimated internal isopycnal displacement of O (80 m), as reported above for solitary wave displacements in the region.

One outstanding feature is the coherency of the surface tide along the Channel cross-section (different rows), a feature already addressed in section "Coherent Surface and Internal Tides". As mentioned, the coherent internal tides in Figs. 17 and 18 (coloured vectors) are obtained by subtracting at each level the surface tide

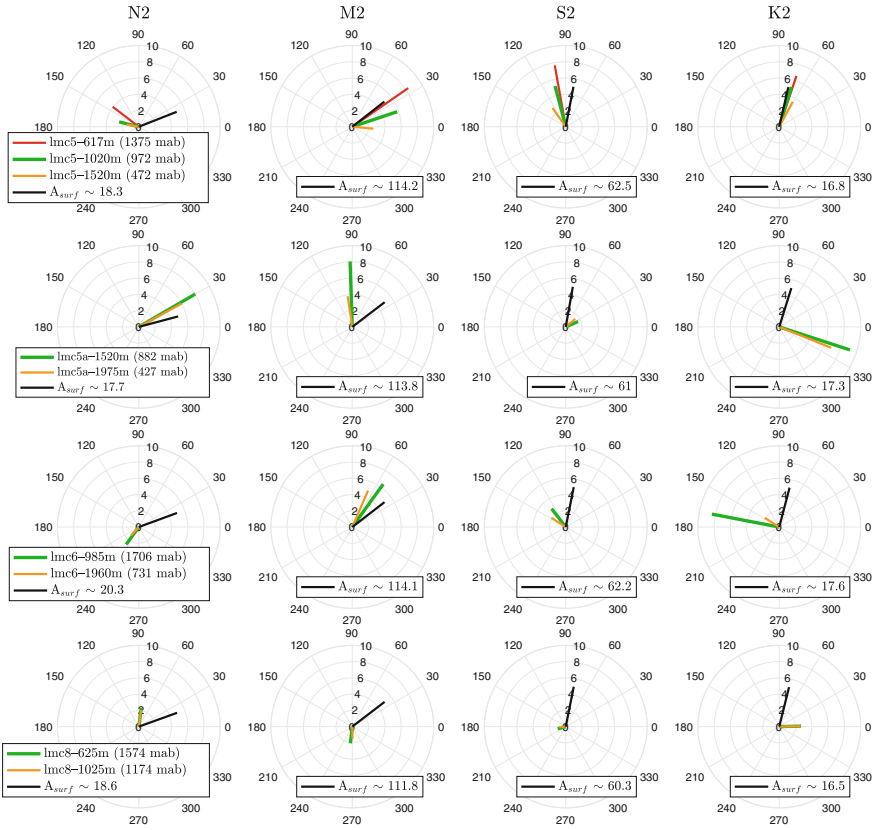


Fig. 17 Vertical variation of MC amplitudes (cm) and phases (°) of displacements by semidiurnal tidal constituents at sites arranged in rows from West (lmc5, top row), via the Middle (lmc5a and lmc6) to East (lmc8, bottom row). Coloured bars represent *coherent internal* tides. Black bar gives *direction* of surface tide (as measured at bottom). Amplitude of the surface tide, A_{surf}, is given in the legend (see Table 6 for further details). From left to right, columns represent tidal constituents N2, M2, S2 and K2, respectively. Here polar plots share the same scale only within the same column, i.e. tidal constituent

(Table 6) from the original harmonic vectors, i.e. the MT_TIDE average of the curves shown in Figs. 6 and 7. The existence of similar long-term, phase-locked internal tides were for instance observed near Hawaii [35]. The phase differences with depth that one obtains along a mooring in the vertical show direct evidence of oblique—beam-wise—internal tide propagation. In continuously-stratified fluids, upward phase propagation of internal waves (such as seen for M2 at lmc5) corresponds to downward energy propagation. Notice that at other locations (e.g. at lmc6) for M2, the phase decreases from the bottom upwards. Hence it displays a downward phase and upward energy propagation. Apparently, part of the multiply-reflected internal tidal beam is captured. Phase equality over depth (such as for all

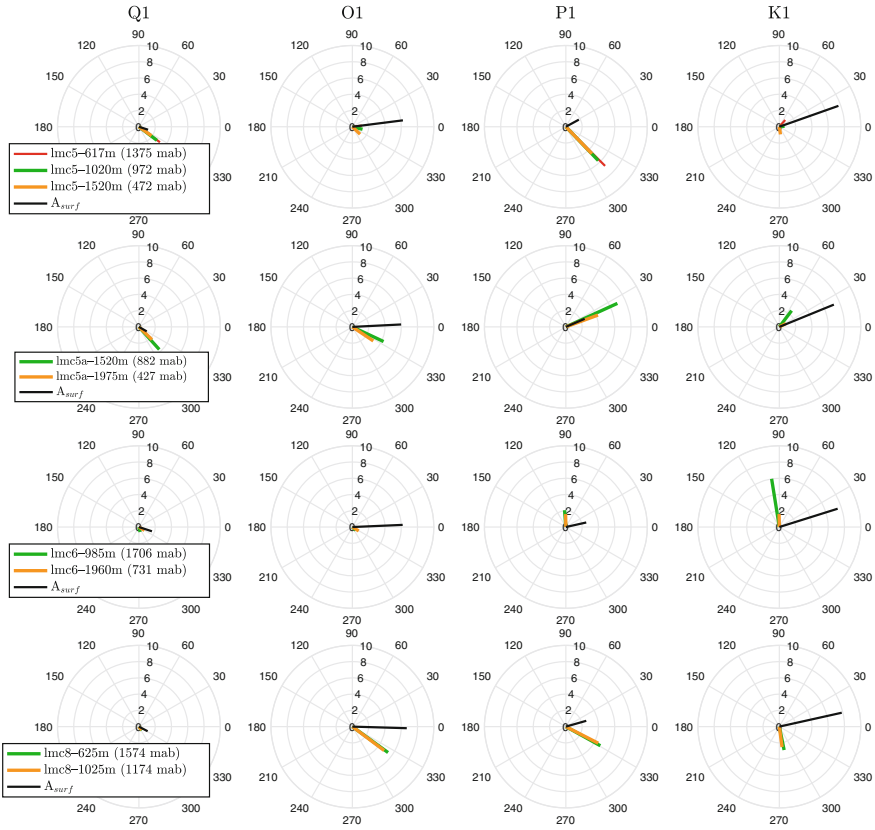


Fig. 18 As Fig. 17 for diurnal frequencies, from left to right: Q1, O1, P1 and K1, respectively, except that surface tidal amplitude (A_{surf}, black bar) here corresponds to bar length

frequencies at lmc5a and lmc8) implies either dominance of the surface tide or a horizontally-propagating, coherent internal tide. However, one has to keep in mind the very small aspect ratio of MC, and the ocean in general, so that low-frequency internal tides propagate at very small angles to the horizontal. Phase differences of baroclinic tides that occur in transverse channel direction (seen along columns in Figs. 17 and 18) are indicative of their shorter length scales.

Looking at the coloured vectors in Fig. 17, one notices the overall variability in amplitude and phase of the internal tide harmonic vectors, although it seems they are strongest in the center of the Mozambique Channel, near lmc5, lmc5a and lmc6. It is also remarkable that the coherent internal K2 tide seems to be at least as strong as the coherent internal M2 tide, despite the free surface K2 tide being seven times weaker than the surface M2 tide.

Figure 18 shows the diurnal surface (black vectors) and internal tides (coloured). In this case, since amplitudes of the surface and internal tides lie in the same range, the surface tide's vector length corresponds its magnitude.

A remarkable property of these observations is that diurnal baroclinic tidal vertical displacements always seem stronger and sometimes much stronger than those produced by the surface tide, see e.g. in Fig. 18 the amplitude increases with increasing height above the bottom at lmc5 and lmc5a for frequencies Q1 and P1. This requires an amplification mechanism. This may be because internal tides, forced by cross-isobath surface tidal motions near the shelf edges of the adjacent coasts are focused near the center of the Mozambique Channel. At the shelf edge, its amplitude will be set by the local surface tide, which, due to a reduced shelf depth, will be stronger than over the deep-sea, hence leading to stronger internal tides. Another reason for elevated amplitude levels of internal tides may be that once internal tides propagate beam-wise into the abyss, they reflect from sloping bottoms, at which point they will often be focused, leading to amplification of their energy density.

It is also remarkable that nearly all diurnal internal tide vectors are co-aligned with one another. This means they are in phase, with vectors increasing in magnitude higher up in the water column. This suggests them to be propagating horizontally, varying in magnitude along the vertical.

East Madagascar Coherent Internal Tides

For EM moorings we do not have collocated bottom and mid-depth instruments. Therefore, estimates from mid-depth instruments might contain the signal of both surface and internal tides. In Fig. 13, the harmonic vectors within the water column, at EMC2, 4 and 5, however, dwarf those at EMC1 and 3, related to the surface tide at the bottom. Evidently coherent internal tides are remarkably strong on the transect East of Madagascar. Looking at the semidiurnal components (lowest two rows), curiously, tidal phases within the water column are generally in opposition to those of the surface tide (as measured by BPRs). Even though moorings are up to 100 km apart, for the M2-tide the phase relationship within the water column is very tight. Subtracting the surface M2 tide (measured at EMC3 or OTIS) by eye from the harmonic vectors within the water column, leads to equivalent surface displacements due to internal tides of O (50 cm). With the conversion employed in section "Mozambique Channel Coherent Internal Tides" this would suggest isopycnal displacements of O(500 m)! Such large magnitude is suspicious and should be treated with caution. It is quite unexpected because of the weakness of the surface tide along the EM transect and surroundings. Near sloping topography, surface and coherent internal tides are not necessarily scale separated [22, 43]. This suggests the internal tide to have a barotropic length scale, such as for a forced internal tide. For S2 there appear some horizontal phase differences, suggesting the S2 coherent internal tide to have a propagating character.

Incoherent Internal Tides

So far we have discussed the surface tides (based on BPRs, section "Bottom-Pressure Versus Altimetry Derived Tides") and the coherent internal tides (which are long-period averages from MT_TIDE). For this, in MC we first subtracted the surface tide (section "Mozambique Channel Coherent Internal Tides"), while in EM the surface tide is much smaller than the coherent internal tide so that the long-period average immediately give an estimate for the coherent internal tides (section "East Madagascar Coherent Internal Tides"). Figures 6, 7, 8, 9, 10 and 11, however, also show large variations of tidal amplitudes and phases around their averages. These represent *incoherent* internal tides. These can, crudely speaking, be of comparable magnitude as the coherent internal tide. The large, mid-water column tidal amplitudes in e.g. Fig. 8 display gradual as well as sudden changes. The sudden jumps might be related to large-scale (blow-down) events, despite the applied filters. The data gaps that have been replacing these events, that enter or leave the analysed one-year time interval, may still have a disruptive effect. We notice that such jumps do not occur for the bottom instrument. The mid water-column pressure records at lmc5a are remarkable in showing correlating drops in M2, S2, K2 and P1 amplitudes, taking place from February to July 2013. This drop below the average levels given in Table 6 suggests that internal tides at instrument levels more than 425 m above the bottom have a significant presence during these periods. Notice that such internal tide variations also showed up in the mid water column pressure records of Figs. 6 and 7. N2, M2 and S2 amplitudes, as well as K2 phases, suggest the internal tides at lmc5a to be often anti-correlated with those at lmc6, their values fluctuating around those representative of the surface tide. This change in internal tide amplitude on relatively nearby moorings suggests that internal tidal beams change path due to variations in mean density and flow fields.

Conclusions

New bottom pressure records for the South-West Indian Ocean have been presented. These can be reliably interpreted as providing measurements of the surface tides and compare in general well with the satellite altimetry derived tides. As such, these BPRs might be useful in future world-wide assimilation of deep pressure data in numerical models of ocean tides when combined with altimetric data. According to Ray's [34] separation criteria, the moorings may not be separated well enough (by about 100 km) to be treated as independent, so that MC and EM transects may deliver two independent data points only. At the same time, spatial variations in BPRs in the vicinity of the coast may be realistic, and may prompt the development of a better resolution in near-coastal altimetry measurements.

Pressure measurements within the water column regularly suffer from blow-down of moorings by eddy, buoyancy, tidal and wind-driven flows. Blow-down leads to

virtual excursions much in excess of those realistically possible. These affect also the estimates of (internal) tides, due to tidal currents amplifying this blow-down—an unwanted effect—or due to trapping of internal tides within eddies, which we have not been able to unravel here. Accordingly, we simply eliminated blow-down events. In future work, one could compute actual instrument displacements using observed currents along the moorings, combined with a knowledge on the shape of the mooring line, the weights of instruments and cable, and the buoy's shape, size and buoyancy [11].

The reason to nevertheless study mid water-column pressure is the presence of both coherent as well as incoherent (i.e. intermittent) internal tides [23]. Intermittency is due to the varying paths taken by internal tides from generation sites (near shelf edges and other bottom irregularities, at critically sloping bathymetry) to measurement locations. These paths vary due to slow changes in density and velocity fields, associated with low-frequency motions.

Performing harmonic analysis (T_TIDE) over consecutive overlapping intervals we obtained multiple estimates of harmonic amplitudes and phases (MT_TIDE). The coherent surface and internal tidal constants are independent of the measurement period and are thus represented by the average of MT_TIDE. The variable part defines the incoherent internal tide. This allows separation of the coherent and incoherent tides (determined by their robustness and variance, respectively) in a manner that may be more reliable than the error band provided by T_TIDE from a single (year long) time series [28].

Both the coherent as well as incoherent tides are fairly strong, especially when the equivalent surface displacement is reconverted into isopycnal displacement. The latter can attain excursions of tens to hundreds of meters. The phase variation of the coherent and incoherent internal tides along a mooring line is usually weak. The amplitude variations are stronger. Horizontal variations are, obviously much larger, clearly displaying an internal tidal length scale that is much shorter than that of the corresponding surface tide.

Acknowledgements This work is dedicated to Eugene Morozov on the occasion of his seventieth birthday. The authors are thankful to Gary Egbert, Lana Erofeeva and Oregon State University for providing the altimetry-derived barotropic tides and to the Royal Netherlands Institute for Sea Research for providing the pressure data. Pressure data were collected in the INATEX program, 'INdian-ATlantic EXchange in present and past climate', funded by the Netherlands Organization for Scientific Research (NWO), section Earth and Life Sciences (ALW), through its ZKO Grant 839.08.43. The authors acknowledge Commonwealth Scientific and Industrial Research Organisation (CSIRO) for making available the Atlas of Regional Seas 2009 (CARS2009).

References

1. Alford, M. H. (2003). Redistribution of energy available for ocean mixing by long-range propagation of internal waves. *Nature*, *423*, 159–162.

2. Badulin, S. I., Shrira, V. I., & Tsimring, L. S. (1985). The trapping and vertical focusing of internal waves in a pycnocline due to the horizontal inhomogeneities of density and currents. *Journal of Fluid Mechanics*, *158*, 199–218.
3. Berry, M. (1987). The Bakerian lecture 1987: quantum chaology. *Proceedings of the Royal Society of London A*, *413*, 183–198.
4. Bewley, G. P., Lathrop, D. P., Maas, L. R. M., & Sreenivasan, K. R. (2007). Inertial waves in rotating grid turbulence. *Physics of Fluids*, *19*(7), 071701.
5. Boisson, J., Lamriben, C., Maas, L. R. M., Cortet, P. P., & Moisy, F. (2012). Inertial waves and modes excited by the libration of a rotating cube. *Physics of Fluids*, *24*(7), 076602.
6. Cartwright, D. E. (2000). *Tides: A scientific history*. Cambridge: Cambridge University Press.
7. Cartwright, D. E., & Ray, R. D. (1990). Oceanic tides from Geosat altimetry. *Journal of Geophysical Research: Oceans*, *95*(C3), 3069–3090.
8. da Silva, J. C. B., New, A. L., & Magalhães, J. M. (2009). Internal solitary waves in the Mozambique Channel: observations and interpretation. *Journal of Geophysical Research*, *114*, C05001. https://doi.org/10.1029/2008JC005125.
9. da Silva, J. C. B., New, A. L., & Magalhães, J. (2011). On the structure and propagation of internal solitary waves generated at the Mascarene Plateau in the Indian Ocean. *Deep-Sea Research I*, *58*, 229–240.
10. Defant, A. (1961). *Physical oceanography* (Vol. 2).
11. Dewey, R. K. (1999). Mooring design and dynamics: A Matlab package for designing and analyzing oceanographic moorings. *Marine Models*, *1*(1), 103–157.
12. Egbert, G. D., & Erofeeva, S. Y. (2002). Efficient inverse modeling of barotropic ocean tides. *Journal of Atmospheric and Oceanic Technology*, *19*, 183–204.
13. Harlander, U., Ridderinkhof, H., Schouten, M. W. & de Ruijter, W. P. M. (2009). Long-term observations of transport, eddies, and Rossby waves in the Mozambique Channel, *Journal of Geophysical Research*, *114*(C02003), 1–15.
14. Hendershott, M. C. (1981). Long waves and ocean tides. In *Evolution of physical oceanography* (pp. 292–341).
15. Hutter, K., Wang, Y., & Chubarenko, I. P. (2011). *Physics of lakes*. Springer.
16. Konyaev, K. V., Sabinin, K. D., & Serebryany, A. N. (1995). Large-amplitude internal waves at the Mascarene Ridge in the Indian Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, *42*(11–12), 20752083–20812091.
17. Kunze, E. (1985). Near-inertial wave propagation in geostrophic shear. *Journal of Physical Oceanography*, *15*, 544–565.
18. Lamriben, C., Cortet, P. P., Moisy, F., & Maas, L. R. M. (2011). Excitation of inertial modes in a closed grid turbulence experiment under rotation. *Physics of Fluids*, *23*(1), 015102.
19. LeBlond, P. H., & Mysak, L. A. (1981). *Waves in the ocean*. Elsevier.
20. Maas, L. R. M. (1997). On the nonlinear Helmholtz response of almost-enclosed tidal basins with sloping bottoms. *Journal of Fluid Mechanics*, *349*, 361–380.
21. Maas, L. R. M. (2003). On the amphidromic structure of inertial waves in a rectangular parallelepiped. *Fluid Dynamics Research*, *33*, 373–401.
22. Maas, L. R. M. (2011). Topographies lacking tidal conversion. *Journal of Fluid Mechanics*, *684*, 5–24.
23. Manders, A. M. M., Maas, L. R. M., & Gerkema, T. (2004). Observations of internal tides in the Mozambique Channel. *Journal of Geophysical Research: Oceans*, *109*(C12).
24. Moum, J. N., & Nash, J. D. (2008). Seafloor pressure measurements of nonlinear internal waves. *Journal of Physical Oceanography*, *38*(2), 481–491.
25. Morozov, E. G. (1995). Semidiurnal internal wave global field. *Deep Sea Research Part I: Oceanographic Research Papers*, *42*(1), 135–148.
26. Munk, W. H., & Cartwright, D. E. (1966). Tidal spectroscopy and prediction. *Philosophical Transactions of the Royal Society of London*, *259*(1105), 533–581.
27. Nurijanyan, S., Bokhove, O., & Maas, L. R. M. (2013). A new semi-analytical solution for inertial waves in a rectangular parallelepiped. *Physics of Fluids*, *25*(12), 126601.

28. Pawlowicz, R., Beardsley, B., & Lentz, S. (2002). Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE. *Computers and Geosciences*, *28*(8), 929–937.
29. Platzman, G. W. (1971). Ocean tides and related waves. *Mathematical Problems in the Geophysical Sciences*, 14(Part 2), 239–291.
30. Ponsoni, L., Aguiar-González, B., Maas, L. R. M., van Aken, H. M., & Ridderinkhof, H. (2015). Long-term observations of the East Madagascar Undercurrent. *Deep-Sea Research I*, *100*, 64–78.
31. Ponsoni, L., Aguiar-González, B., Ridderinkhof, H., & Maas, L. R. M. (2016). The East Madagascar Current: Volume transport and variability based on long-term observations. *Journal of Physical Oceanography*, *46*(4), 1045–1065.
32. Proudman, J. (1917). On the dynamic equation of the tides. Parts 1–3. *Proceedings of the London Mathematical Society*, Series 2, *18*, 168.
33. Rao, D. (1966). Free gravitational oscillations in rotating rectangular basins. *Journal of Fluid Mechanics*, *25*, 523–555.
34. Ray, R. D. (2013). Precise comparisons of bottom-pressure and altimetric ocean tides. *Journal of Geophysical Research: Oceans*, *118*(9), 4570–4584.
35. Ray, R. D., & Mitchum, G. T. (1997). Surface manifestation of internal tides in the deep ocean: Observations from altimetry and island gauges. *Progress in Oceanography*, 35–162.
36. Ridderinkhof, H., van der Werf, P. M., Ullgren, J. E., van Aken, H. M., van Leeuwen, P. J. & de Ruijter, W. P. M. (2010). Seasonal and interannual variability in the Mozambique Channel from moored current observations. *Journal of Geophysical Research: Oceans*, *115*(C06010), 1–18.
37. Schrama, E. J. O., & Ray, R. D. (1994). A preliminary tidal analysis of TOPEX/POSEIDON altimetry. *Journal of Geophysical Research: Oceans*, *99*(C12), 24799–24808.
38. Taylor, G. I. (1921). Tidal oscillations in gulfs and basins. *Proceedings of the London Mathematical Society*, Series 2, 148–181.
39. Terra, G. M., Doelman, A., & Maas, L. R. (2004). Weakly nonlinear cubic interactions in coastal resonance. *Journal of Fluid Mechanics*, *520*, 93–134.
40. Ullgren, J. E., van Aken, H. M., Ridderinkhof, H., & de Ruijter, W. P. M. (2012). The hydrography of the Mozambique Channel from six years of continuous temperature, salinity, and velocity observations. *Deep Sea Research Part I: Oceanographic Research Papers*, *69*, 36–50.
41. van Haren, H. (2013). Bottom-pressure observations of deep-sea internal hydrostatic and non-hydrostatic motions. *Journal of Fluid Mechanics*, *714*, 591–611.
42. Whewell, W. (1836). Researches on the tides. Sixth series. On the results of an extensive system of tide observations made on the coasts of Europe and America in June 1835. *Phillosophical Transactions of the Royal Society of London*, *126*, 289–341.
43. Zaron, E. D. & Egbert, G. D. (2007). The impact of the M2 internal tide on data-assimilative model estimates of the surface tide. *Ocean Modelling*, *18*, 210–216.

Internal Tides West of the Iberian Peninsula

Eugene G. Morozov and Manuel G. Velarde

Introduction

Semidiurnal internal tides exist almost everywhere in the ocean. It is generally accepted that internal tides are generated due to the interaction of the barotropic tide with the slopes of bottom topography. The largest amplitudes are usually found near underwater ridges and continental slopes [1, 2]. Tidal currents flow over the slopes of the bottom topography and obtain a vertical component, which induces periodic vertical displacements of the isopycnals. Hence, internal waves with tidal periods are generated that propagate from the areas of uneven bottom topography to the open ocean. As a result, the energy of the barotropic tide is converted into the energy of the internal tide. The maximum fluxes of energy lost by the barotropic tide are found in the regions of submarine ridges and continental slopes such as the South China Sea, the Mascarene Ridge, the Aleutian and Hawaii islands [3, 4]. Extremely strong internal tides are generated in the straits when the currents of the barotropic tide flow over the sills that cross the straits [5–8].

The maximum amplitudes of the internal tides are found in the regions with shallow ridge crests or shelves adjacent to deep waters if the inclination of the slopes is close to the inclination of the characteristic curves of the equation for internal waves [9]. In this contribution, we analyze internal tides generated over the Atlantic continental slope of the Iberian Peninsula.

Moored measurements with thermistor chains were carried out on the shelf break west of Portugal at 41° N [10]. The authors recorded high-frequency internal waves

E. G. Morozov (✉)
Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
e-mail: egmorozov@mail.ru

M. G. Velarde
Instituto Pluridisciplinar, Madrid, Spain
e-mail: mgvelarde@pluri.ucm.es

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_13

generated by the internal tide with an amplitude of 45 m. The estimated energy fluxes were 2 kW m^{-1}. Note for comparison that the energy fluxes associated with the strongest internal tides in the South China Sea are as high as 40 kW m^{-1} [11].

Very strong internal tides are generated over the continental slope on the northern coast of Biscay Bay [12]. Our analysis is based on the archived moored measurements gathered in the Iberian Basin during many expeditions within different oceanographic programs. We use the clusters of moorings as an antenna to estimate the direction of the internal tides and their wavelength. The method of calculation is based on the calculation of the spatiotemporal spectrum. A line of moorings normal to the slope can be helpful to estimate the decay of internal tides. Solitary moorings distributed over large regions can be used to evaluate the geographical distribution of the internal tide energy.

Analysis of the internal tides in the Strait of Gibraltar was carried out in [5]. The flow in the strait is characterized by a two-layer system of opposite currents resulting from the difference in sea level and water density (salinity and temperature) between the Atlantic Ocean and the Mediterranean Sea. The flow changes its direction approximately at a depth of 100 m. The barotropic tidal wave with mean velocities in the range 70–80 cm/s propagates through the strait. The barotropic tide generates an internal tidal wave, when the tidal currents flow over the Camarinal Sill located in the middle of the Strait of Gibraltar. Extreme tidal internal waves (with an amplitude reaching 160 m) are observed only over the Camarinal Sill. The vertical displacement of the 13 C isotherm ranges from 100 to 300 m. When the internal tide wave reaches the western part of the Strait of Gibraltar, its amplitude in the open ocean decreases to 30 m, which is comparable with the amplitudes of internal tides generated on the continental slopes of Iberia and Africa.

The wavelength of the internal tide propagating to the east ranges from 90 to 140 km, while the wavelength of the internal tide propagating to the west ranges from 45 to 60 km. The wavelength of the internal tide propagating in the eastern direction is greater and the difference is not surprising because the depths east of the Camarinal Sill are greater than in the western part of the Strait.

The Data

Many moorings were deployed in the region in the last 30 years. During the World Ocean Circulation Experiment (WOCE), two lines of moorings were deployed in 1993–1994 on the continental slope of the Iberian Peninsula as part of the Multidisciplinary Oceanographic Research in the Eastern Boundary of the North Atlantic project (MORENA).

The lines included three moorings ACM-27 (along 42° 16′ N) and four ACM-28 moorings along 41° 00′ N from 9° 30′ W to 11° 00′ W (Fig. 1). These measurements were intended to study the processes on the continental slope.

In 1988, the oceanographers onboard the Russian R/V "Vityaz" studied the spreading of Mediterranean Water from the Strait of Gibraltar to the Atlantic Ocean

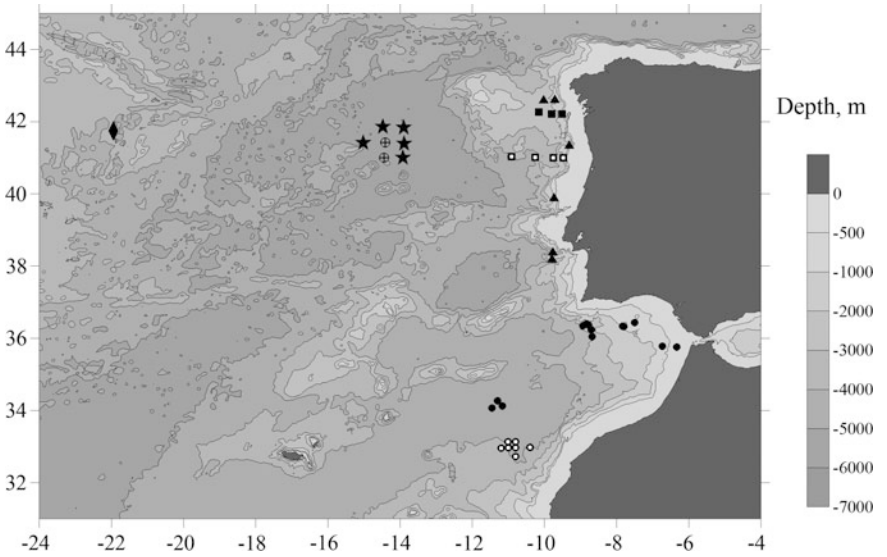


Fig. 1 Bottom topography (meters) west of the Iberian Peninsula. Depth contour lines are shown at depths of 500, 1000, 2000, 3000, 4000, and 5000 m. Locations of moorings are shown with black symbols. The following experiments are indicated with different symbols: WOCE ACM-27, ACM 28 (squares, the moorings used to calculate the energy decay of the internal tide are indicated with white dots in the middle); "Vityaz-88" (black dots), the moorings used to calculate spatiotemporal spectrum are indicated with white dots in the middle); OMEX experiment (triangles); moorings in 1984–1985 used to calculate the spatiotemporal spectrum (pentagon stars); two other moorings in 1984–1985 (circle with a cross), and a standalone mooring in 1980–1981 (diamond). Land is shown with gray color

and formation of the Mediterranean lenses. During the cruise, 19 moorings were deployed for a period of 15 days each west of the Strait of Gibraltar to study the formation of Mediterranean lenses of high salinity. The locations of moorings are shown in Fig. 1.

Moored measurements west of the Iberian Peninsula were carried out in 1996–1999 within the OMEX Program (Ocean Margin Exchange Project; mooring numbers 6988–6993 and 7016–7042). Several moorings were deployed on the continental slope: 42° 39′ N, 9° 42′ W; 42° 37′ N, 10° 02′ W; 39° 37′ N, 9° 44′ W; 41° 22′ N; 9° 18′ W; 38° 13′ N, 9° 47′ W; 38° 24′ N, 9° 46′ W.

The experiments in the 1970–1980s were aimed to study mesoscale eddies of the open ocean, which were discovered in the beginning of the 1970s. The researchers tried to investigate the dynamic properties of the oceanic eddies with a horizontal size of 200–300 km. The energy of such eddies is comparable with the energy of quasi-stationary currents. In 1984–1985, moored measurements were carried out on a cluster of five moorings located close to 41° 30′ N, 15° 00′ W. The depths of measurements were 1500 and 2600 m. The data were downloaded from the open

access CMDAC database at the Oregon State University (CMDAC numbers are acc00739–acc00747). Figure 1 shows the locations of moorings.

Other archived data were also used in our research. These experiments were also aimed to study mesoscale eddies. Two moorings were deployed in 1978–1979 at 41° 00′ N, 24° W. The measurements were made at 2300, 3300, and 4000 m. The CMDAC numbers are acc00705–acc00707. A mooring was deployed in 1980–1981 at 41° 45′ N, 22° 00′ W. The measurements were made at 600, 1500, 3000, and 3790 m. The CMDAC numbers are acc00522–acc00525.

We used the archived data of these measurements west of the Iberian Peninsula to study the properties of internal tide whose spatial scales (100–200 km) are close to the scales of mesoscale eddies (200–300 km). The available data with a time sampling of one hour are quite applicable to study internal tides with periods of 12 h.

Spatiotemporal Spectrum

We used the method developed for seismological problems and applied by Barber to the ocean waves to estimate the wavelength and direction of the internal tides [13]. We assume that the sensors are located randomly over the observational sites. The method is based on the calculation of the cross spectra for each pair of the possible combinations of sensors with further convolution at the frequency of the waves under study. The spectra are calculated using the Fourier transformation of the correlation function as described by Blackman and Tukey [14]. The amplitude and phase cross-characteristics of the oscillation are used to calculate the spatiotemporal spectra at the wave frequency and estimate the components of the horizontal wave number.

The method basically accounts for the statistical phase difference between each pair of the wave sensors. In our case we used the temperature sensors, which indicate the vertical motion induced by internal waves assuming that the vertical gradients of temperature are significant. It is important that the distance between the moorings should be comparable with the wavelength of the oscillations under study. Otherwise uncertainty appears in the interpretation of the phase differences if the distance between the sensors is too large. If the distance between the sensors is two small and the span of the array of sensors is smaller than the wavelength due to a limited number of sensors it is impossible to resolve the wavelength correctly.

If moored temperature measurements at the same depth at several points in the ocean are available, we can calculate cross spectra of fluctuations P and Q (co-spectrum and quadratic spectrum). Next, we perform a transformation at the M_2 semidiurnal tidal frequency f_0 to determine the distribution of mutual spectral energy at this frequency with respect to wavenumbers k_x and k_y.

So far we do not have a continuous spectrum of distances, but instead, we have a finite set of definite distances corresponding to the distances between the moored

stations, the calculation of a spatiotemporal spectrum at f_0 frequency was carried out using the following formula:

$$E(k_x, k_y, f_0) = 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} [P_{ij}(f_0) \cos 2\pi(k_x x_{ij} + k_y y_{ij}) - Q_{ij}(f_0) \sin 2\pi(k_x x_{ij} + k_y y_{ij})],$$

where $k_x^2 + k_y^2 = k^2$ is the spatial wave number, while the wavelength L is: $L = 1/k$; i, j are the sequential numbers of the sensors, n is the total number of the sensors; P_{ij} and Q_{ij} are the real and imaginary parts of the cross spectrum between the sensors with numbers i and j; $x_{ij} = X_i - X_j$ and $y_{ij} = Y_i - Y_j$ are the projections of the distances between the sensors on the horizontal axes x and y.

Usually, the number of sensors is small. Each sensor is installed on a mooring. The number of moorings deployed in the study sites rarely exceeds 10. Hence, the shape of the spectral peak is not a delta-function but has a finite width. Usually, the spatiotemporal spectra of internal tides with semidiurnal frequency (or possibly, other frequencies) are presented as contour lines of percentage normalized by the maximum value.

Analysis

We studied internal tides in the region west of the Iberian Peninsula on the basis of moored temperature and velocity measurements. The moorings deployed in the region allowed us to analyze the properties of internal tides.

Two lines of WOCE moorings in 1993–1994 on the continental slope of the Iberian Peninsula extending to the deep basin along 42° 16′ N and along 41° 00′ N allowed us to determine the decay of internal tide amplitude with the distance from the continental slope. A scheme of the region, bottom topography, and locations of moorings along the line normal to the continental slope are shown in Fig. 1. Locations of the instruments related to the continental slope are shown in Fig. 2.

The eastern mooring of the ACM 28 line was located at a distance of 60 km from the shore on the continental slope at a depth of 1300 m. The line of moorings extended to the depths of 4200 m. The length of the line of moorings was 150 km. All moorings returned the temperature data at 700–800 m. Intense internal tides were recorded on all of the moorings. The mean peak to peak temperature variations on different moorings were 0.25–0.50 °C, which corresponds to the amplitude of internal tides within 38–78 m. The amplitudes of semidiurnal internal tides were decreasing from 78 m over the continental slope to 38 m in the deep basin over a distance of 150 km. A graph of the variations of internal tide amplitudes at 700–800 m is shown in Fig. 2.

The locations of instruments over the background of depth are shown in the top panel of Fig. 2. The amplitudes of internal tides are shown in the bottom panel. One can see a decrease in the amplitudes up to the depths of 4200 m as the distance from

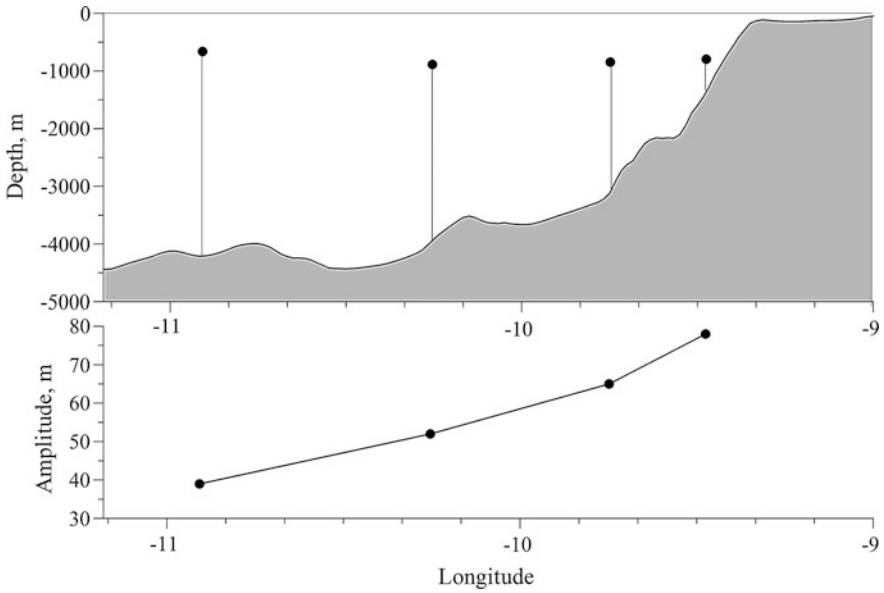


Fig. 2 Scheme of the moorings with current and temperature measurements over the bottom topography on the section along 41° N (top panel); Peak-to peak amplitudes of the semidiurnal internal tide versus western longitude (bottom panel)

the continental slope increases. The decrease in the amplitude is almost linear. The rate of decay is generally close to the estimates in [15, 16].

We also calculated energy decay of the internal tide with the distance from the continental slope. The energy densities of the internal tides averaged over a wave period were calculated as in [16]:

$$E_{TW}(z) = 0.25\rho\left(\overline{u_{IT}^2(z)} + \overline{v_{IT}^2(z)} + N^2(z)\overline{\varsigma_{IT}^2(z)}\right),$$

where the amplitudes of the semidiurnal internal tidal components are u_{IT}, v_{IT} (zonal and meridional currents) and ς_{IT} (vertical displacements); N is the Brunt-Väisälä frequency calculated from the vertical profiles of CTD data in the region. Tidal components and vertical displacements associated with the internal tide were calculated after band filtering of the velocity and temperature data from moorings. Then the temperature fluctuations were divided by the mean vertical temperature gradient. In both regions, the energy decay is approximately the same.

Energy decay versus normalized distance is shown in Fig. 3. Similarly to the energy decay studied in [16] the energy decay west of the continental slope of Iberia is governed by a power law approximately proportional to e^{-2}. A similar decay was found for the energy decay of internal tide propagating to the southeast from the Mozambique coast (Fig. 3).

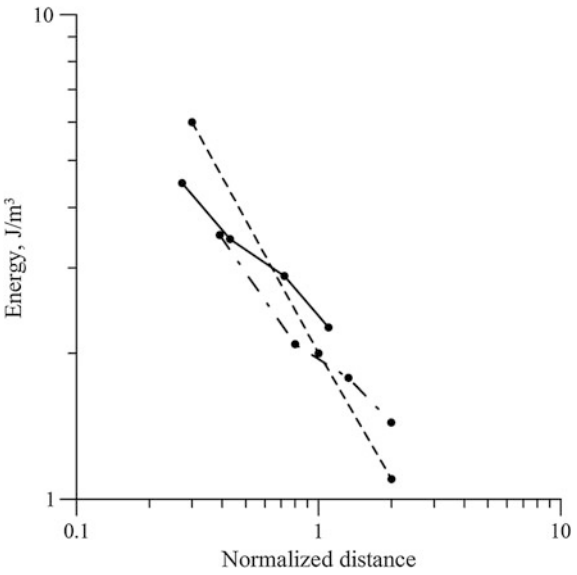


Fig. 3 Energy density of the semidiurnal internal tides (J/m^3) versus distance from the continental slope of the Iberian Peninsula (solid line). The horizontal axis shows the distance from the slope in the direction normal to the slope normalized by the wavelength of the first mode (145 km). A similar decay of the internal tide energy versus distance near the Mascarene Ridge in the Indian Ocean is shown with a dashed line [16] and southeast of Mozambique (dashed-dotted line). The dots on the graphs indicate the energy values based on the measured velocities and temperature on moorings

The moorings deployed in 1998 from the Russian R/V "Vityaz" close to the Strait of Gibraltar were organized in several clusters. The goal of the experiment was to study the formation of the Mediterranean lenses of the saline water flowing from the strait as the lower current. The horizontal scales of the clusters were selected to analyze the Mediterranean intrathermocline lenses (Meddies). These scales are close to the scales of internal tides, thus these measurements allowed us to estimate the amplitude, wavelength, and direction of the internal tides. The amplitudes were within 30–40 m. On a mooring cluster (33° N, 11° W), in which the moorings operated synchronously for 15 days, it was possible to estimate the direction and wavelength of the semidiurnal internal tides. We expected that the waves would propagate from the continental slope in the region of Morocco, but the wave arrived in the direction to the south-southwest (200) from the southwestern coast of the Iberian Peninsula. The wavelength was 100 km (Fig. 4).

The moored measurements in 1984–1985 in the middle of the Iberian Basin (41° 30′ N, 15° 00′ W) at depths of 2600–2700 m on five moorings over the ocean depths of 5300–5500 m resulted in the mean amplitude of the semidiurnal internal tide equal to 30–40 m, while at a depth of 1500–1600 m the amplitudes were 40–50 m.

The spatiotemporal spectrum based on the data at 2600 m using the Barber's method [13] is shown in Fig. 5. The wavelength of the semidiurnal internal tides was estimated at 105 km. The waves propagated from the northeast from a region of shallower depths less than 850 m. A bank exists here at 42° 30′ N, 12° 00′ W.

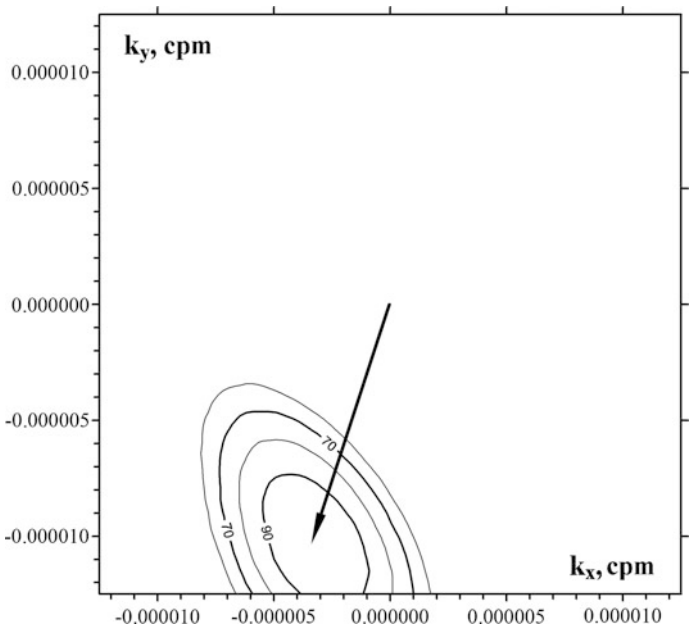


Fig. 4 Spatiotemporal spectrum at the semidiurnal frequency based on the data of six moorings shown in Fig. 1 (depth 1000 m) as black dots with a white circle in the middle ("Vityaz-88"). The contour lines on the spatiotemporal spectrum correspond to 90, 80, and 70% of the main maximum. The arrow shows the wave vector corresponding to a wavelength of 100 km

The moored measurements along the continental slope of the Iberian Peninsula in 1996–1999 within the OMEX Program allowed us to estimate the amplitudes of internal tides. The mean amplitude at depths of 600–1000 m was estimated at 30–40 m.

The amplitude of internal tides at a depth of 600 m measured on a standalone mooring in the middle of the Iberian Basin at point 41° 45′ N, 21° 57′ W in 1980–1981, was estimated at 30 m.

Conclusions

We analyzed the moored measurements in the Iberian Basin west of the Iberian Peninsula. The data were collected from many sources. All measurements confirm the generation of internal tides over continental slopes due to the interaction of the

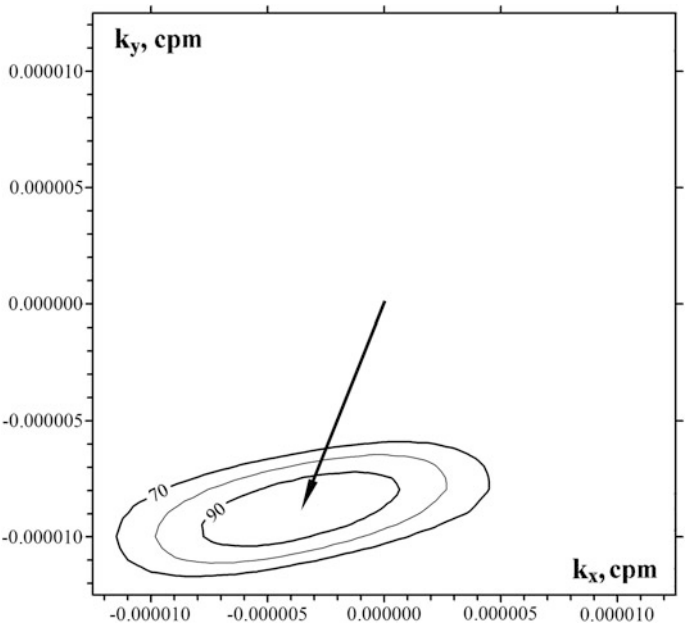


Fig. 5 Spatiotemporal spectrum at the semidiurnal frequency based on the data of five moorings in the middle of the Iberian Basin shown in Fig. 1 (depth 2600 m) as squares with black dots and a white circle in the middle (moorings in 1984–1985). The contour lines correspond to 90, 80, 70, and 60% of the main maximum. The arrow shows the wave vector corresponding to a wavelength of 105 km

barotropic tide with the sloping bottom topography. The amplitudes (peak to peak displacements) of internal tides over the continental slope are sometimes as high as almost 80 m. They decrease in the open ocean to 30 m.

Acknowledgements This work was supported by the Russian Science Foundation (project no. 14-50-00095).

References

1. Baines, P. G. (1982). On internal tide generation models. *Deep-Sea Research, 29*(3), 307–338.
2. Morozov, E. G. (1995). Semidiurnal internal wave global field. *Deep-Sea Research, 42*(1), 135–148.
3. Li, Q., & Farmer, D. M. (2011). The generation and evolution of nonlinear internal waves in the deep basin of the South China Sea. *Journal of Physical Oceanography, 41,* 1345–1363.
4. Morozov, E. G., & Vlasenko, V. I. (1996). Extreme tidal internal waves near the Mascarene Ridge. *J. Marine Systems, 9*(3–4), 203–210.

5. Morozov, E. G., Trulsen, K., Velarde, M. G., & Vlasenko, V. I. (2002). Internal tides in the Strait of Gibraltar. *Journal of Physical Oceanography, 32,* 3193–3206.
6. Morozov, E. G., Parrilla-Barrera, G., Velarde, M. G., & Scherbinin, A. D. (2003). The straits of Gibraltar and Kara Gates: A comparison of internal tides. *Oceanologica Acta, 26*(3), 231–241.
7. Morozov, E. G., Demidov, A. N., & Tarakanov, R. Y. (2008). Transport of Antarctic waters in the deep channels of the Atlantic Ocean. *Doklady Earth Sciences, 423*(1), 1286–1289.
8. Morozov, E. G., Paka, V. T., & Bakhanov, V. V. (2008). Strong internal tides in the Kara Gates Strait. *Geophysical Reseach Letters, 35*(16), L16603.
9. Morozov, E. G., Nechvolodov, L. V., & Sabinin, K. D. (2009). Beam propagation of tidal internal waves over a submarine slope of the Mascarene Ridge. *Oceanology, 49*(6), 745–752.
10. Jeans, D. R. G., & Sherwin, T. J. (2001). The evolution and energetics of large amplitude nonlinear internal waves on the Portuguese shelf. *Journal of Marine Research, 59,* 327–353. https://doi.org/10.1357/002224001762842235.
11. Alford, M. H., Peacock, T., MacKinnon, J. A., Nash, J. D., Buijsman, M. C., Centuroni, L. R., et al. (2015). The formation and fate of internal waves in the South China Sea. *Nature, 521* (7550), 65–69. https://doi.org/10.1038/nature14399.
12. Pingree, R. D., & New, A. L. (1991). Abyssal penetration and bottom reflection of internal tidal energy in the Bay of Biscay. *Journal of Physical Oceanography, 21,* 28–39.
13. Barber, N. F. (1963). The directional resolving power of an array of wave detectors. In *Ocean wave spectra* (pp. 137–150). NY, Engelwood Cliffs: Prentice Hall.
14. Blackman, R. B., & Tukey J. W. (1958). *The measurements of power spectra from the point of view of communications engineering*. New York: Dover.
15. Morozov, E. G., Vlasenko, V. I., Demidova, T. A., & Ledenev, V. V. (1999). Tidal internal wave propagation over large distances in the Indian Ocean. *Oceanology, 39*(1), 42–46.
16. Lozovatsky, I. D., Morozov, E. G., & Fernando, H. J. S. (2003). Spatial decay of energy density of tidal internal waves. *Journal Geophysical Research, 108*(C6), 3201–3216.

Asymmetric Baroclinic Response to Tidal Forcing Along the Main Sill of the Strait of Gibraltar Inferred from Mooring Observations

Jesús García-Lafuente, Simone Sammartino, José C. Sánchez-Garrido and Cristina Naranjo

Introduction

The Atlas of Internal Waves [1] highlights the Strait of Gibraltar as a region where short-period, large-amplitude internal wave trains progressing into the Mediterranean Sea occur regularly throughout the year. Indeed they are renowned and paradigmatic features of this area that have been known since long ago by fishermen and sailors navigating the Strait. They identified the tidal origin of the phenomenon that happens phase-locked with the barotropic tide (sea level oscillation) and exhibits intensity quite well related to the spring-neap tidal cycle. Pioneer field experiments carried out during the 1960s supported this empirical knowledge by suggesting that the flow in the Strait does not move as a continuous or smoothly-varying current, but as tidal-induced pulses [2, 3]. However, the severe technical difficulties caused by the strong currents of the region, the safety restrictions associated with the heavy maritime traffic, and the harsh meteorology, prevented the acquisition of reliable in situ observations until the late 1960s [4].

The basic theory of internal wave generation and propagation involves the combination of water stratification, bathymetry, and transcritical flows. These conditions are often met in coastal regions such as continental shelves, inlets, fjords and straits, where these waves are more frequently observed [1]. Wherever stratification, seabed irregularities or currents are enhanced, the generated internal wave train will also be. Equally important is the angle of incidence of the flow with the isobaths, normal incidence being the most favorable situation. The Strait of Gibraltar has a year-round notable stratification (compared with the seasonal stratification in other places of the ocean) due to the encounter of Mediterranean and Atlantic waters. At the same time, the presence of the protruding sill of

J. García-Lafuente (✉) · S. Sammartino · J. C. Sánchez-Garrido · C. Naranjo
Physical Oceanography Group, University of Málaga, Málaga, Spain
e-mail: glafuente@ctima.uma.es

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_14

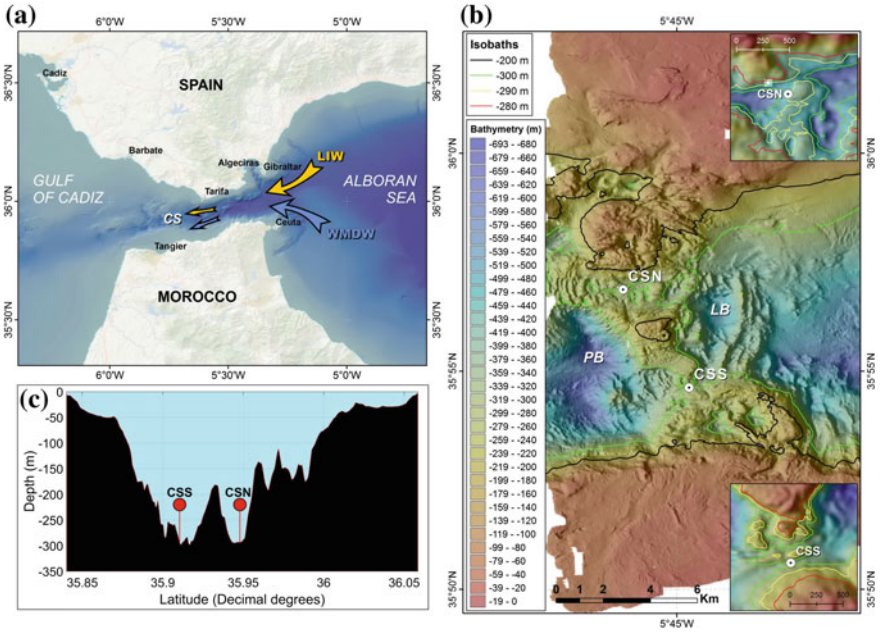


Fig. 1 **a** Map of the Strait of Gibraltar and approaches sketching the path of the two main Mediterranean water masses entering the strait. Camarinal Sill is indicated by the acronym CS. **b** Detailed shaded bathymetry of CS area from [30] with indication of the accurate position of the two deployed mooring lines (CSN and CSS, see also up and bottom insets). The two depressions at the east (LB for Levante Basin) and west (PB for Poniente Basin) of CS mentioned in the text are also indicated. **c** Detailed cross-section of CS showing the position of both mooring lines (which are not to scale)

Camarinal (CS, see Fig. 1), that intercepts the Mediterranean outflow, and the intense tidal currents [5–9] make this spot to gather optimal conditions for the occurrence of this spectacular phenomenon, as it has been widely recognized among the oceanographic community. This promoted the accomplishment of the Gibraltar Experiment during the 1980s.

This international cooperative program provided further experimental evidences of the small-scale structure of the internal waves and a theoretical frame to interpret their life cycle. The generation was explained within the internal hydraulic theory of the exchange through the Strait [10–12]. The comprehensive dataset collected during the experiment suggested that the internal wave train was the time evolution of a hydraulic jump regularly released on the west slope of CS that progresses towards the Mediterranean Sea [11]. The phenomenon is triggered by the flooding of the hydraulic control formed over CS [10, 11, 13], which occurs by the end of the flood tide when tidal currents weaken [7, 11]. All this happens shortly before the local high water due to the standing wave nature of the tide in the area [14]. The spring-neap modulation of semidiurnal currents and the importance of diurnal tidal currents [7], bring about fortnightly and diurnal variability on the flooding of the

hydraulic control and the subsequent internal bore release, as well as on the very intensity of the bore, which can even not be released [15].

The time evolution of the internal wave train has been addressed in Vlasenko et al. [16] using a simplified numerical model, whereas the generation mechanism was investigated in detail in Sánchez-Garrido et al. [13] by means of a high-spatial resolution numerical model implemented in a domain of very accurate topography. The model reproduces the generation and eastward propagation of the wave train very satisfactorily (Fig. 2a–e). Interestingly, it also indicates the formation of a smaller hydraulic jump to the east of CS during the ebb tide (Fig. 2f), when the Atlantic water reaches deeper over CS, and the westward propagation of weaker internal wave packets when the jump decays. However, these features occur much less regularly than its west-of-CS counterpart described previously and, therefore,

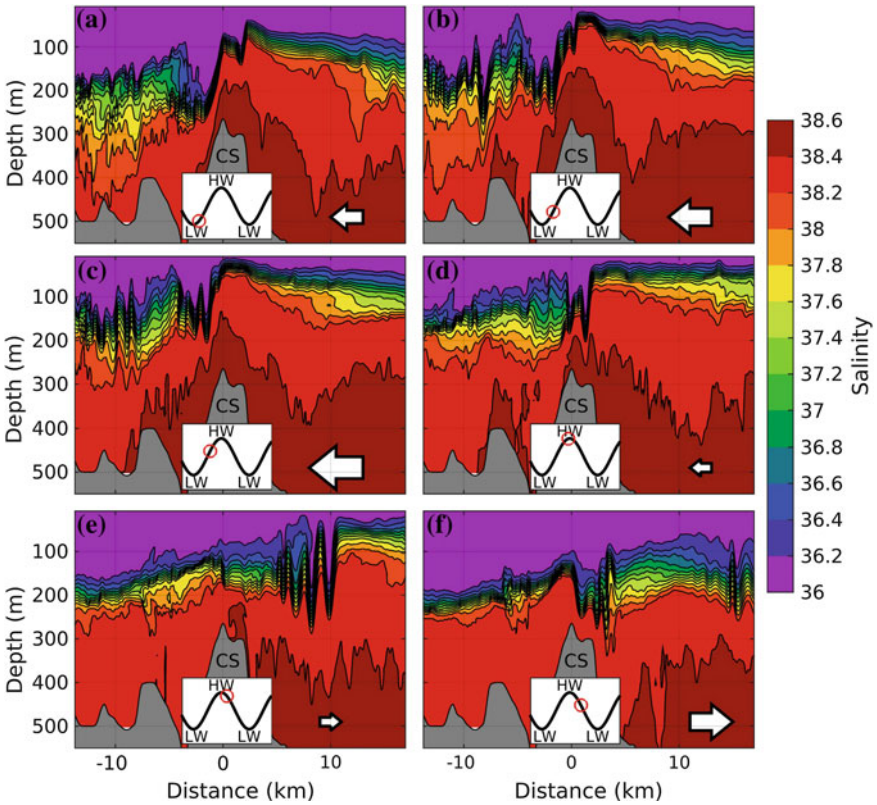


Fig. 2 Contours of salinity in an along-strait section of the Strait of Gibraltar that intersects CS, which is taken as the axis origin, at six selected times of the tidal cycle (see insets). They correspond to a situation of strong tidal forcing (spring tide) and have been adapted from Sánchez-Garrido et al. [13]. The arrows indicate the direction of the barotropic tidal current, their size attempting to represent its instantaneous strength

have little observational support, the Gibraltar Experiment dataset showing some hints of them [11].

The great asymmetry of the occurrence of eastwards versus westwards progressing wave trains lies in the background flow mean state. Figure 3 sketches two possible states, the first one (left panels) with a motionless stratified water column over a rough topography (sill), the second one (right panels) with a steady baroclinic exchange. When strong enough oscillating tidal currents are superimposed to the background state, the flow-topography interaction gives rise to the formation of symmetric hydraulic jumps in the first case, but not in the second one. When the jumps decay, they in turn originate wave trains progressing away from the sill in

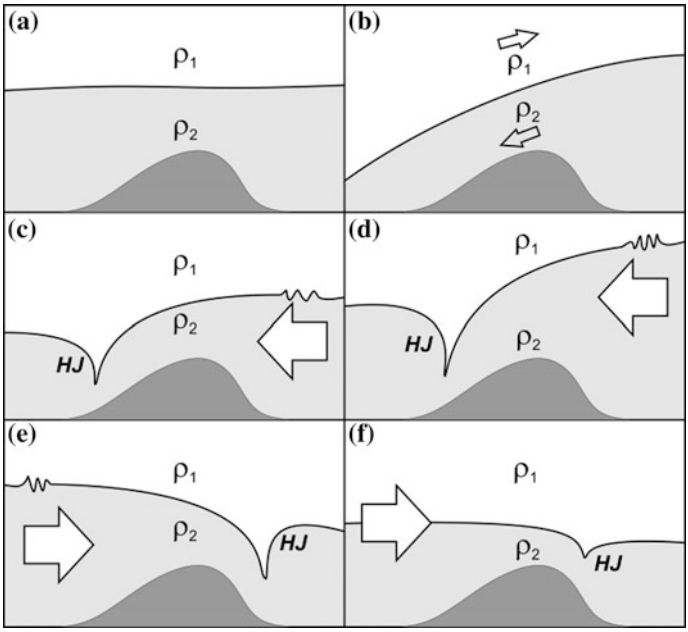


Fig. 3 Simplified sketch of the formation of hydraulic jumps and internal waves in a stratified (two-layer) flow over marked bathymetry (sill, deep grey shade). Panels **a** and **b** show the background situation in cases of no flow and steady baroclinic exchange, respectively. Panels **c** and **d** illustrate the result of (the same) barotropic tidal forcing on both background states when tidal currents point to the left. They lead to the eventual formation of hydraulic jumps (*HJ*) downstream of the sill, with the jump corresponding to the baroclinic mean exchange (panel d) showing greater amplitude. Panels **e** and **f** do the same for tidal currents pointing to the right. Again, an asymmetry appears with the jump formed in the no-flow background state (panel **e**) being larger than the counterpart of the mean baroclinic exchange (panel **f**). In both cases, the asymmetry arises from the different background states, which enhances (reduces) the total deep current when tidal currents point to the left (right) under steady baroclinic exchange. The short-wave packets propagating on the interface would represent the internal waves generated during the previous tidal cycle. These wave packets have similar probability of occurrence in either direction in the no-flow background state (panels **c** and **e**) but not in the baroclinic mean exchange (panels **d** and **f**) where the right-going waves are formed much more frequently

either direction with similar regularity in the first case, but obviously not in the second one.

Curiously, examples of both situations are met in the Mediterranean Sea: the Messina and Gibraltar straits. The Messina strait, roughly sketched in Fig. 3a, falls in the first class. It sustains strong tidal currents due to the different phase of the tide between the Tyrrhenian and Ionian seas [17]. In summer-autumn, when the seasonal thermocline is formed and the water column becomes stratified, internal wave packets moving northwards or southwards are regularly observed, with nearly equal rate of occurrence [1, 18]. On the contrary, the baroclinic background state sketched in Fig. 3b, which applies to Gibraltar with the Mediterranean Sea to the right, overwhelmingly breaks the balance toward the side of west-of-CS (left-of-sill in Fig. 3) formed jump and subsequent eastwards propagation of internal waves. Additionally, the year-round stratification in Gibraltar, driven by salinity rather than temperature differences, also results in a fairly even distribution of the observed wave packets over the year that contrasts with the summer-autumn concentration in Messina [1].

The formation of hydraulic jumps east of CS is reproduced by the numerical model of Sánchez-Garrido et al. [13], despite its apparent uncommonness (Fig. 2f). The west-going wave packets resulting from the decay of the east-of-CS jump are likewise elusive to observe, the few reported cases being offered by Alpers et al. [18] or Morozov et al. [9]. Interestingly, the former authors associated the events with the seasonal thermocline rather than with the permanent pycnocline between Atlantic and Mediterranean waters, which is the interface the east-going internal wave packets are linked to.

Even when the understanding of the whole process is very satisfactory, small-scale details (O(1 km)) of the formation processes of the hydraulic jumps on either side of CS, their flooding and subsequent release of internal bores, their cross-strait spatial structure, etc., are not so well known, let alone from an experimental point of view. This work addresses some of these issues from an observational approach, taking advantage of a field experiment primarily designed to investigate the cross-strait structure of the Mediterranean outflow in CS at subinertial time-scales [19]. Next section shows the detailed bathymetry of the CS area and the essential topographic features, section "Data and Data Processing" presents the data and data-processing, section "General Description" analyzes the data and section "Discussion and Conclusions" discusses and summarizes the conclusions drawn from this study.

A Close-Up to the Bathymetry of Camarinal Sill

Figure 1b displays an updated and detailed bathymetry of CS. The topography is very uneven with marked sub-kilometric features that steer the deep water flow. Of special interest is the presence of two larger-scale channels in the section, denoted by Camarinal North and Camarinal South Channels (CSN and CSS, hereinafter, Fig. 1),

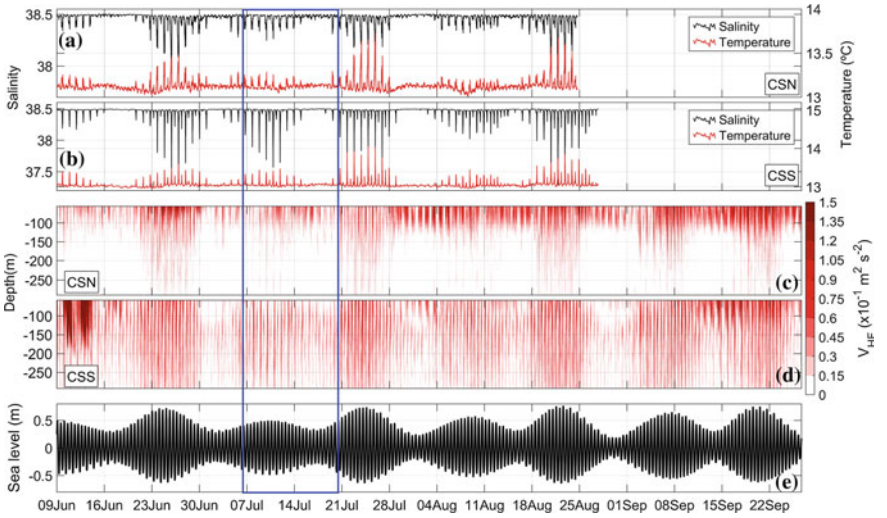


Fig. 4 **a** Temperature and salinity observed in CSN. **b** Same as **a** for CSS. In both panels, spikes are almost exclusively observed during spring tides and are more alike every two semidiurnal periods, that is, every 24 h approximately (see text). **c** Smoothed contours of the high-frequency variance of the ADCP velocity field ($\boldsymbol{V_{HF}}(z,t)$, see text) at CSN. **d** Same as **c** for CSS. **e** Sea level in Tarifa (see Fig. 1 for location). The blue rectangle indicates the portion of the series displayed in Fig. 5

through which the densest Mediterranean waters must flow. CSN channel is shallower (thalweg at −300 m, see inset in Fig. 1b), narrower (∼2 km at 200 m-depth isobaths), has a smooth U-shaped form and favors the evacuation of waters flowing along the northern half of the Strait, whereas CSS has a more V-like profile, is slightly deeper (thalweg at −310 m, Fig. 1) and broader (∼4.5 km at −200 m). The thalwegs distance is ∼5 km. The ratio of cross-areas (CSN/CSS) below 200 m depth is 1:2 approximately, which suggests similar fractions of the outflow across each channel. The Mediterranean waters flowing through them have slightly different characteristics due to the tendency of the relatively warm and salty Levantine Intermediate water (LIW) to flow closer to the Spanish shore and the Western Mediterranean Deep water (WMDW), colder and fresher, to flow attached to the Moroccan coast [20–22], see sketch in Fig. 1a. Therefore, LIW and WMDW will flow in greater proportion across CSN and CSS, respectively (see Fig. 4 in García-Lafuente et al. [19]).

Of particular interest are the closed depressions at either side of CS (LB and PB in Fig. 1b), which display erosive features. Quite probably, they have been carved out by strong flows that have had to pour down from the top of the sills of Camarinal in the flood (PB) and ebb (LB) tides, that is, by the supercritical flows associated with the alternating formation of hydraulic jumps at either side of CS. The asymmetry imposed by the background baroclinic exchange, which reinforces flood and weakens ebb tidal currents in the deeper layer, would accordingly explain

the dissimilarities in extension and depth of each depression, the western one (PB) being noticeably greater than the eastern one (LB). Therefore, the little intuitive presence of LB to the east of CS would actually be an indirect proof of east-going supercritical flow downhill of CS, which is a hypothesis on the focus of the present study.

Data and Data Processing

Two twin mooring lines equipped with a Conductivity-Temperature (CT) probe and an upward-looking Acoustic Doppler Current Profiler (ADCP) located at 10 and 12 m above the seafloor, respectively, were deployed at the thalwegs of CSN and CSS channels in bottom depths of 306 and 310 m, respectively (Fig. 1b, c). The ADCP observations spanned the period 9th of June to 25th of September 2013, and operated at a sampling interval of 4 min. The instruments were configured to sample 40 vertical bins, each one 6 m thick, so that the velocity profile does not reach the sea surface. However, it samples the Mediterranean water layer whose upper boundary (interface) was estimated from in situ salinity observations at a time-averaged depth of 120 m in CSN and 134 m in CSS by Bryden et al. [5], that is, ∼180 m above the seafloor at the mooring sites. The CT observations were collected every 2 min, starting on the 9th June as well, but they finished earlier due to battery run out (on 24th and 27th of August at CSN and CSS, respectively). In both cases, the sampling rate was enough to have several samples within the time scale of tens of minutes typical of the internal waves [15].

Figure 4a, b show sudden and remarkable spike-shaped changes on the time series of temperature and salinity that happen quite regularly during the spring phase of the fortnightly cycle (see Fig. 4e). These series are, therefore, good indicators of the occurrence of intense hydraulic features over CS. Velocity series, however, are no so straightforward indicators because the high frequency velocity fluctuations associated with the internal waves are superposed to very strong tidal flows that make the former appear noise-like.

In order to extract more useful indicators of the high frequency processes from the ADCP observations, the velocity profiles have been high-pass filtered with a filter of 2 cph cut-off frequency. The three components of the high-passed velocity have been combined to compute the variable $V_{HF}(z,t) = \frac{1}{2}\sum_{i=1}^{3} u_i(z,t)^2$, where $u_i(z,t)$ is the i-th component of the velocity at instant t and depth z. Since the variable is rather noisy in time and space, moving average operators with windows width of 45 min in time (11 data) and 30 m in space (5 depth-levels) have been applied to $V_{HF}(z,t)$ in order to smooth out the noise. These operations are equivalent to calculating the variance of the series within the established time and depth windows. The result of this procedure can be seen in Fig. 4c, d, which display the

expected spring-neap tidal modulation in both CSN and CSS sites, more neatly in the latter.

Finally, the location of the interface between Atlantic and Mediterranean waters is relevant in this study. A proxy for this interface is the surface of maximum vertical shear of the horizontal velocity [5, 23, 24], which has been estimated from the low-passed along-strait velocity, using a 5 cph cut-off frequency filter, and the same moving average window of 5 vertical bins in the vertical. The time-averaged locations of the interface so computed have been found at 110 (36) m and 153 (52) m at CSN and CSS respectively (bracketed values indicate the standard deviation), in good agreement with those provided by Bryden et al. [5]. A good deal of the interface variability is tidally-induced with semidiurnal oscillations of several tens of meters centered at M_2 frequency.

General Description

The spikes in Fig. 4a, b indicate the arrival to the bottom of CS of Mediterranean water mixed up with fresher and warmer overlying Atlantic water (the CT probes are hardly 10 m above the seafloor). As mixing depends on the strength of the tidal currents, the good matching between the periods of spikes appearance and spring tide is not surprising at all. Likewise, the diurnal inequality in the internal hydraulics of the strait driven by the important diurnal tidal currents [7, 9, 15] is also mirrored by the spike size, which are larger every two peaks. The size is greater in CSS (notice the different scale in Fig. 4a, b), surely as a result of the stronger currents across this channel. More importantly, the shape of the spikes differs noticeably from one site to the other, a fact that is better appreciated in Fig. 5.

Figure 4c and d display the footprint on the velocity series of the high-frequency motions associated with the internal hydraulics. Again, the contour intensity follows the spring-neap cycle, a fact that is particularly evident in the case of CSS. The cycle is less apparent in CSN, yet identifiable, because the variability tends to concentrate in the uppermost part of the water column, contrary to what happens in CSS where a mid-depth maximum is weakly suggested (see also Fig. 8b).

A quick inspection of Fig. 5, which enlarges the series inside the rectangle of Fig. 4, confirms the above mentioned features, but it also reveals new ones of interest. First, the temperature and salinity peaks always occur during the falling tide that coincides with the ebb current due to the standing-wave nature of the tidal wave in the strait [14]. The direction of the ebb current is to the east and, if strong enough (a situation met in spring tides), reverses the deep Mediterranean layer that momentarily will flow towards the Mediterranean Sea (Fig. 6b). A second noteworthy feature is the regular semidiurnal periodicity of the high frequency currents in CSS (Fig. 5b), with the expected greater similarity every two semidiurnal cycles due to the diurnal inequality. The pattern is not as clear in CSN (Fig. 5a), although it is still detectable. The contours, however, are now found during the

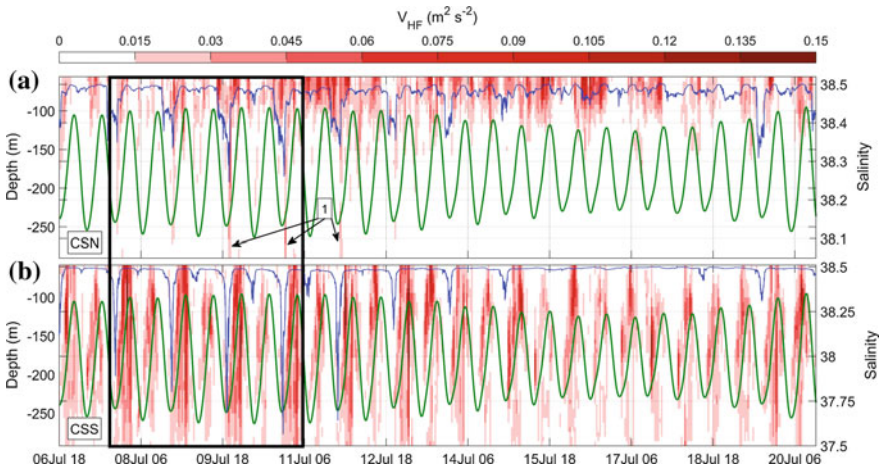


Fig. 5 Zoom of the fragment of the series indicated by the blue rectangle in Fig. 4, which coincides with a period of spring tides. With small variations, this period is representative of spring-tide conditions during all events recorded in the field experiment. Panels **a** and **b** present contours of the variance of the high-frequency velocity ($\boldsymbol{V_{HF}}(z,t)$, see text) along with the salinity (blue lines) at CSN and CSS, respectively. Sea level in Tarifa (green line, no units) is superposed in both panels for reference. Dates in the time axis are in the format ddMMMhh. The black rectangle marks the fragment zoomed in Fig. 8

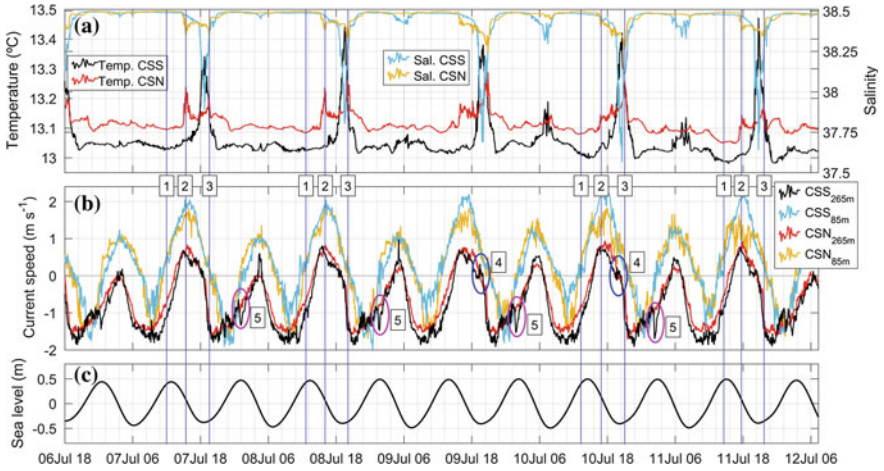


Fig. 6 **a** Salinity and temperature at CSN and CSS during a short fragment of the series displayed in Fig. 5. **b** Along-strait velocity at two selected depths (see legend). The along-strait axis is 5° and 20° anticlockwise from east in CSN and CSS, respectively, these values coming from the vertically-averaged orientation of the M_2 tidal ellipses obtained from the harmonic analysis of the velocity profiles [31]. **c** Sea level at Tarifa. Numbers inside squares refer to features discussed in the text

rising tide when the associated flood (westward) tidal currents enhance the velocity in the Mediterranean layer.

These features linked to the ebb and flood tides are addressed separately in the following sections. The discussion is just limited to spring tide situations, which is when the energetic high-frequency motions we are interested in are observed. Data within the spring tide period marked by the rectangle in Fig. 4 will be used in the analysis, for it is representative of all other spring tide cycles.

Ebb Tide

Shortly before the slack tide corresponding to high water, the water column above CS consists mainly of the Mediterranean water that has been flowing westwards during the previous flood tide (Fig. 2c is a good reference for this situation). Accordingly, the salinity registered by the CT probes at CSN and CSS is at their maxima (labels (1) in Fig. 6), suggesting that it is at these moments when the less mixed, purer Mediterranean waters are flowing out at those depths. A similar result is inferred from temperature data, which are at their minima. At that moment the observations illustrate the known cross-strait distribution of Mediterranean water masses in CS, with colder WMDW flowing preferably through CSS and warmer LIW doing it over CSN [25, 26]. The total current in the lower part of the water column still points westwards, whereas it has already reversed (or it is about to) in the upper column (labels (1), Fig. 6).

About 3 h after the high water, the ebb tide becomes maximum and so does the eastward current in the upper layer (labels (2), Fig. 6). The Atlantic water transport into the Mediterranean Sea also reaches a maximum, which in turn implies a thickening of the Atlantic layer that deepens the interface [23, 27]. In CSN, the deepening is perceived as a sudden increase of temperature and decrease of salinity, which are not detected in CSS (labels (2), Fig. 6). The ebb current is strong enough to reverse the flow in the lower Mediterranean layer, which now moves to the Mediterranean Sea. Under these circumstances, the establishment of hydraulic control and the eventual formation of a weak hydraulic jump east-of-CS, such as the one sketched in Fig. 3f or illustrated in Fig. 2f, emerges as a realistic possibility, in which case these peaks should be caused by processes related to the evolution and fate of this jump. During neap tides, the reduced mixing along with the diminished tidal transports prevent the mixed water to reach so deep and it is barely detected by the CT probes, which measure rather constant values of temperature and salinity (Fig. 4a, b).

During the following 2 h approximately (from labels (2) to (3) in Fig. 6) the thermo-haline properties of water registered at CSN keep on differing from pure Mediterranean water. After the mentioned sudden change, both variables display irregular high frequency fluctuations and, occasionally, gradual changes towards characteristics of more mixed-with-Atlantic water. The pattern has strong dependence on the diurnal inequality, since it is only distinctly displayed every two tidal cycles.

These high frequency temperature and salinity oscillations appear to be correlated to the perturbations of the current, which are visible at shallower depths in CSN (labels (2)–(3) in Fig. 6, see also Fig. 5a). The correlation suggests certain proximity of the instruments to the deeper part of the interfacial layer, whose small-scale vertical oscillations would leave correlated high-frequency signatures in all series. This period comes to an end around low water, when the ebb tide ceases and the flood cycle begins. Salinity and temperature series in CSN show another sudden change that brings the observed values back to the expected for Mediterranean waters (labels (3) in Fig. 6). This change is greater than the previous one (cf. labels (2) in Fig. 6), it happens more regularly and it is usually accompanied by a pulse of current that affects the whole water column in the variance series (see labels (1) in Fig. 5a and also labels (4) in Fig. 8a later on). We reason that these features are caused by the westward progression of an internal bore and the tied in internal wave packet that follows the release of the hydraulic jump formed east-of-CS.

The situation in CSS appears somewhat different. Here, the deepening of the interface during the ebb cycle does not reach the CT probe for as long as in CSN, and salinity and temperature fluctuations just consist of a single peak up to four times greater than those observed in CSN (Fig. 6a). It requires larger fraction of Atlantic water in the mixed water sampled by the CSS probe than by the CSN probe, which can only be explained if the interface goes deeper in the former site, although this deep excursion lasts shorter in the latter.

The peak at CSS occurs during the last hour of the ebb cycle and shortly before than the already mentioned second peak in CSN is detected (label (3) in Fig. 6). Actually, a lagged cross-correlation between the salinity (or temperature) series at both sites indicates that the second peak in CSN lags the sharp peak in CSS by $\sim$50 min. Despite being so noticeable in temperature or salinity, the process that causes these fluctuations in CSS does not leave such a clear signature in the series of high-frequency velocity as it does in CSN (see labels (1) in Fig. 5a). Instead of affecting the whole water column, as it is the case in CSN, peaks in CSS only leave a weak visible signature in the lower layer velocity consisting of short reversals (less than half an hour, see labels (4) in Fig. 6b) of the total current, which is about starting flowing westward. And the reversal does not even happen on a regular basis, but rather occasionally.

Flood Tide

The flood tide cycle begins after the occurrence of the sharp temperature and salinity spikes, which signpost the end of the ebb. During this cycle and under spring tide conditions, the east-going flow in CS reverses and the whole water column moves westwards (Fig. 6b). Following the progressive acceleration of the Mediterranean water flow, the interfacial layer rises up to even a few meters from the free surface. In fact, it is not unusual to detect cold sea surface signatures in the eastern approaches of CS in remote sensed images during the flood tide (see Fig. 7,

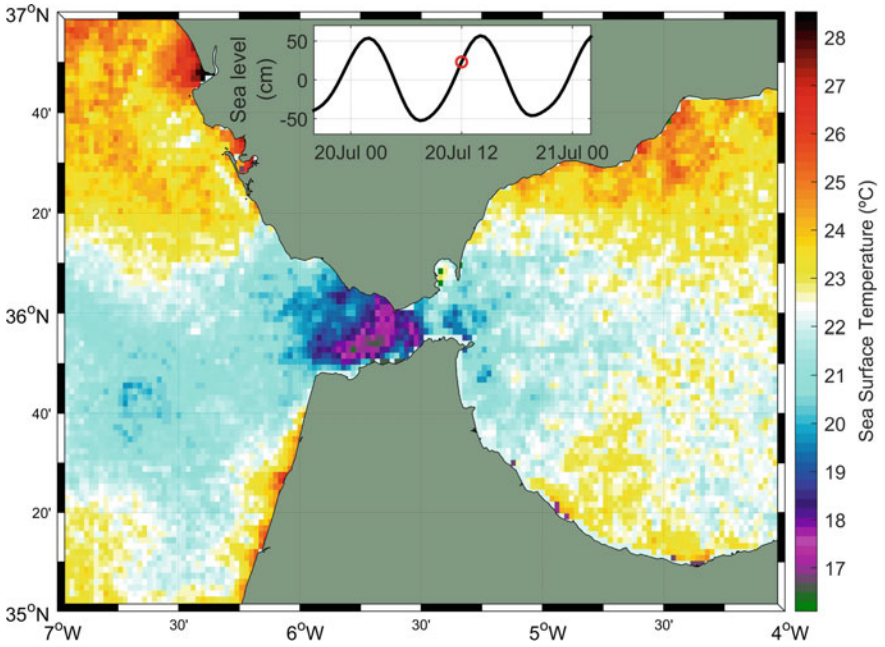


Fig. 7 Sea surface temperature from L4 SST three-hourly satellite observations by Météo-France-IFREMER (distributed by CMEMS-Copernicus) during the flood tide on 20th of July 2016. The image is a 3 h mean around the time indicated by the red circle in the inset, which displays the sea level in Tarifa

for instance). Its origin is caused by the uplift of the interface upstream of the sill, in this case to the east, forced by the marked increase of the outflow of Mediterranean waters during this phase of the tidal cycle. As a result, most of the water column over CS consists of Mediterranean water that leaves a rather flat signal in CSN and CSS temperature and salinity records (Fig. 6a).

During this tidal phase, a hydraulic jump on the westwern side of CS is regularly formed [10, 11, 13], whose subsequent release when the current slackens originates the well-known large amplitude internal wave train that progresses into the Mediterranean Sea. Such an energetic process must leave recognizable footprints in CS velocity observations, regardless of whether or not it leaves any in temperature and/or salinity. The enhanced high-frequency variability of the ADCP series during this tidal phase confirms the expectations, especially in CSS where the contrast with the little variability registered during the ebb tide is striking (Fig. 5b). The high-frequency variability in CSN is less organized with a pattern that remains partially concealed by the noise. Most of the relevant information comes, thus, from CSS where the following description focuses.

Figure 8 shows a three-day zoom to highlight some relevant features, some of them quite obvious as the mentioned ebb-flood dissimilarity or, again, the diurnal inequality.

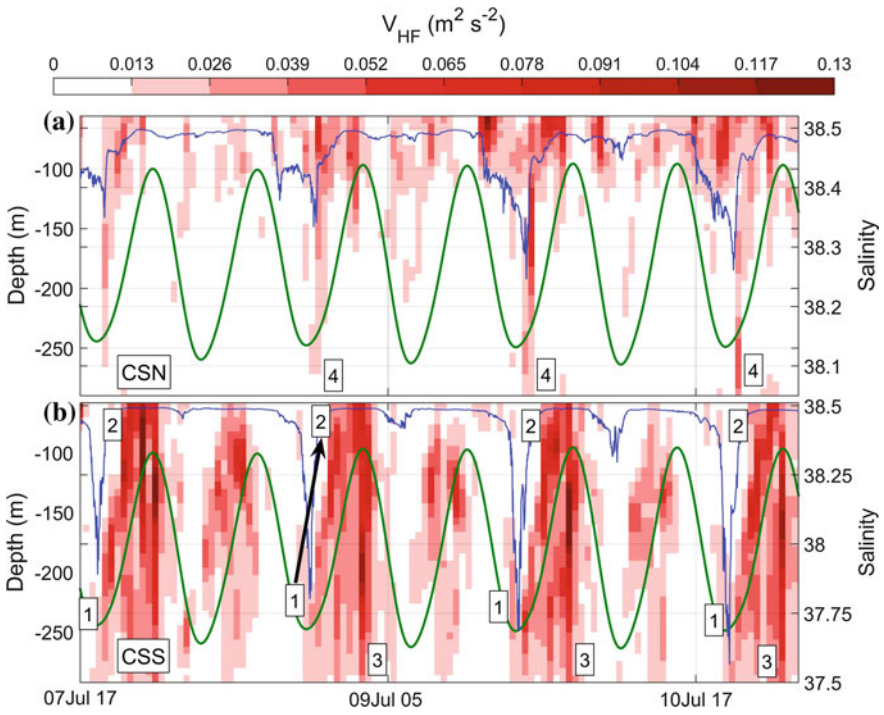


Fig. 8 Zoom of the fragment of the series inside the black rectangle in Fig. 5. Numbers refer to features discussed in the text

Scrutinizing in more detail, when the flood begins after the low water, the high-frequency variance field increases, depicting a maximum located at great depth (labels (1), Fig. 8) that gradually moves upwards [labels (2)] as the flood tide progresses (arrow in Fig. 8). And it occurs regardless of the semidiurnal cycle. Interestingly, the upper bound of the high-frequency variance stripes (labels (1)–(2) in Fig. 8) coincides with the depth of the maximum vertical shear of horizontal velocity that we have identified with the interface, which ascends as the volume of Mediterranean water, forced to overflow CS by the flood tide, increases. The fact that both features appear coupled to each other in their upward displacement suggests that the latter is linked to the establishment of the critical section over CSS and the consequent formation of the hydraulic jump on the lee side (westward in this case) of the sill.

By the end of the flood, shortly before the high water, the hydraulic control in CS is flooded [11, 13], the hydraulic jump develops into a propagating internal bore that in turn evolves in the well-known internal wave packets as the accumulated potential energy is partially converted into kinetic energy (see the successive snapshots in Fig. 2). Temperature and salinity in CSN and CSS are blind to this process as the interface during the flood tide is located well above the position of

the CT probes. Velocity records, on the other hand, exhibit a sharp pulse involving the whole water column in the high-frequency series at CSS (labels (3) in Fig. 8) with greater amplitude at middle depths. The pulse can be also identified in the total current, which is waning now, by a sudden westward increase of the deep layer velocity of up to 70 cm s^{-1} and short duration (labels (5) in Fig. 6b). It almost certainly corresponds with the passing of the bore over the sill. Therefore, the velocity observations in CSS would be suitable not only to follow the formation and subsequent evolution of the hydraulic control and associated west-of-CS hydraulic jump but also to detect its further release and eastwards propagation. Intriguingly, the velocity series at CSN are rather insensitive to these processes, the high-frequency contribution appearing somewhat blurred even though it tends to increase its energy during the flood tide too (Figs. 5a, 8a). It appears as if the east-of-CS hydraulic processes are better seen in CSN while the west-of-CS ones leave clearer signatures in CSS.

Discussion and Conclusions

The temperature, salinity and velocity data collected at CSN and CSS sites in the sill of the Strait of Gibraltar have proven to be useful to observe and follow the successive hydraulic states that take place in this particular environment along the tidal cycle. The data illustrate the different periodicities in the occurrence and strength of the internal features (Fig. 4), the most obvious one being the fortnightly cycle, which is fairly intuitive, but also a diurnal inequality arising from the relatively important diurnal tidal currents [7]. Being interested in high-energy internal processes, a specific spring-tide period has been chosen as representative in order to carry out an in-depth analysis (Fig. 5) that can be extended to all others cycles.

In addition to the alikeness of the data recorded in either site, the analysis has also highlighted differences in the observed patterns. In general terms, it could be said that CSN is better suited to observe internal processes associated with hydraulic transitions occurring over and on the eastern side of the main sill (i.e. the east-of-CS hydraulic jumps) while CSS is more sensitive to the processes taking place over and on the west side, which are the most representative and, consequently, the most addressed in the literature. The bottom topography tends to support this differentiation in the sense that the configuration of the closed depression east of CS (LB in Fig. 1b) seems to be more the erosion outcome of eventual strong eastward deep flow through CSN rather than through CSS, whereas the erosion of the western depression (PB in Fig. 1b) suggests a more efficient contribution from the flow through CSS, even though it appears to be contributed by both channels.

An obvious physical process that has influence on the cross-strait flow structure is the Earth's rotation. It gives rise to a semi-geostrophic flow (across-strait geostrophic balance) that holds even at tidal frequencies [5–7], with the interface being shallower in the north side for the mean exchange. However, the interface slope

responds to the tidally-induced fluctuations of the total flow, which modify its position and can even change its sign. Actually, during the ebb tide the whole water column may be flowing eastwards, particularly in spring tides (Fig. 6), which reduces the interface slope due to the cross-strait geostrophy. At the same time, the interface sinks to make room for the large volume of Atlantic water displaced towards the Mediterranean during this tidal cycle. Both processes act jointly to bring mixed water deeper in CSN than in CSS and explain why it reaches the CSN probe earlier and quits it later (labels (2) and (3), respectively, in Fig. 6). It is only by the end of the ebb tide, when the interface is at its maximum depth, that CSS probe registers the mixed water.

This description would suggest a gradual change rather than the pronounced temperature and salinity peaks that echo the arrival of this water at CSS, profusely shown in the figures of the paper. The drawback is overcome by letting the internal hydraulics act out. The peaks at this site have been related in section "Ebb Tide" to the sudden arrival of the internal bore released when the east-of-CS hydraulic jump decays, a circumstance that happens by the end of the ebb tide when the thickness of the Atlantic layer is maximum. If this jump is formed over the eastern rim of the sill elevation (see Fig. 1b), the bore will take a little longer to get to CSN than to CSS, since the former is ∼2.5 km farther from the rim than the latter. Converted into time, it means 40–45 min for a typical first baroclinic mode speed of 1 ms^{-1}, CSS leading CSN in satisfactory agreement with the delay inferred from cross-correlating the series of temperature (or salinity) at both sites discussed in section "Ebb Tide". The former analysis relies on a cross-strait orientation of the jump and on its simultaneous decay throughout its extension, which appear as reasonable assumptions. Notice that the interpretation of temperature and salinity peaks at CSS in terms of the passing of internal bores is further supported by the short-living deep current reversals observed in CSS (labels (4) in Fig. 6) that accompanies the occurrence of the spikes.

The considerably more studied hydraulic jump formed westwards of CS during the flood tide [10, 11, 13] leaves clearer signatures at CSS than at CSN, but only in the velocity profiles and not in the temperature or salinity series. The profiles reveal high-frequency fluctuations that shoal along with the interface in CSS (Fig. 8b). This uplift of the interface is a distinctive upstream response of hydraulically controlled flows, in which the interface upstream of the control section (the sill crest in this case) must shoal in order to accommodate larger (tidal) volume transport (e.g., [28]; Sect. 1.8). Such a response during the flood tide is therefore indicative of the recovery of hydraulic control over the sill that prevents the propagation of internal disturbances from the Atlantic to the Mediterranean.

The enhancement of the high-frequency velocity fluctuations at CSS can be due to either the presence of shear instabilities or to the trace of a transient upstream internal hydraulic jump evident in the high-resolution non-hydrostatic numerical model of Sánchez-Garrido et al. [13] (see Fig. 2a). We put forward the latter as their origin. This secondary jump would be associated with a second control section arising east of the sill as a result of the shoaling of the interface there and the concomitant reduction of the phase speed of internal disturbances. This control

section is usually referred to as "approach control" in the literature of internal hydraulics ([28, 29], Sect. 5), and would be formed over the very sill a short distance upstream of the crest of CS (Fig. 2a). Moreover, the associated upstream jump is much more apparent in CSS as revealed by the referred model simulation and satellite images (see Fig. 1 of Sánchez Garrido et al. [13] for a nice illustration of the situation).

It is hypothesized here that the high-frequency stripes of marked variance that shift upwards with the interface (Fig. 8b) are connected to the growing of this second jump. When the flood current weakens and the hydraulic controls are lost, the released internal bores eventually merge together and move eastwards as a unique feature (Fig. 2b, c). However, within this scenario, it appears that only the very energetic bore released by the main jump is specifically noticed at CSS, its footprint being the sudden westward pulse of the deep already mentioned (labels (5) in Fig. 6b). Similar traces are not seen so clearly in CSN, where the signal is much more blurred, though still weakly discernible, and concentrates in the upper water column in CSN (Figs. 5a and 8a).

Acknowledgements Data were collected in the frame of the Spanish Government-funded "INGRES-3" (CTM2010-21229/MAR) and "ENCIBA" (CTM2013-40886P) projects. We are particularly grateful to the Instituto Español de Oceanografía, which allocated ship-time of the research vessel Ángeles Alvariño for carrying out the field experiment. This experiment is also part of the PE12-RNM-1540 Regional Government of Junta de Andalucía Excellence project. CN acknowledges a research contract associated with CTM2013-40886P project. JCSG was partially supported by the "Juan de la Cierva" program (contract JCI-2012-13451) and SS acknowledges a post-doc contract linked to PE12-RNM-1540 project.

References

1. Jackson, C. R. (2004). *An Atlas of internal solitary-like waves and their properties* (2nd ed.). Alexandria, VA: Global Ocean Associates.
2. La Violette, P., & Lacombe, H. (1988). Tidal-induced pulses in the flow through the strait of Gibraltar. *Oceanologica Acta* (Special issue, Édn. scientifiques et médicales Elsevier SAS).
3. Lacombe, H., & Richez, C. (1982). The regime of the strait of Gibraltar. In: Jacques CJN (Ed) *Elsevier oceanography series* (pp 13–73). Elsevier.
4. Ziegenbein, J. (1970). Spatial observations of short internal waves in the strait of Gibraltar. *Deep Sea Research and Oceanographic Abstracts, 17,* 867–875. https://doi.org/10.1016/0011-7471(70)90004-5.
5. Bryden, H. L., Candela, J., & Kinder, T. H. (1994). Exchange through the Strait of Gibraltar. *Progress in Oceanography, 33,* 201–248. https://doi.org/10.1016/0079-6611(94)90028-0.
6. Candela, J., Winant, C., & Ruiz, A. (1990). Tides in the strait of Gibraltar. *Journal of Geophysical Research: Oceans, 95,* 7313–7335. https://doi.org/10.1029/JC095iC05p07313.
7. García Lafuente, J., Vargas, J. M., Plaza, F., et al. (2000). Tide at the eastern section of the strait of Gibraltar. *Journal of Geophysical Research: Oceans, 105,* 14197–14213. https://doi.org/10.1029/2000JC900007.
8. Morozov, E. G., Parrilla-Barrera, G., Velarde, M. G., & Scherbinin, A. D. (2003). The straits of Gibraltar and Kara gates: a comparison of internal tides. *Oceanologica Acta, 26,* 231–241. https://doi.org/10.1016/S0399-1784(03)00023-9.

9. Morozov, E. G., Trulsen, K., Velarde, M. G., & Vlasenko, V. I. (2002). Internal tides in the strait of Gibraltar. *Journal of Physical Oceanography, 32,* 3193–3206. https://doi.org/10.1175/1520-0485(2002)032<3193:ITITSO>2.0.CO;2.
10. Armi, L., & Farmer, D. M. (1985). The internal hydraulics of the strait of gibraltar and associated sills and narrows. *Oceanologica Acta, 8,* 37–46.
11. Armi, L., & Farmer, D. M. (1988). The flow of Mediterranean water through the Strait of Gibraltar. *Progress in Oceanography, 21,* 1–103. https://doi.org/10.1016/0079-6611(88)90055-9.
12. Helfrich, K. R. (1995). Time-dependent two-layer hydraulic exchange flows. *Journal of Physical Oceanography, 25,* 359–373. https://doi.org/10.1175/1520-0485(1995)025<0359:TDTLHE>2.0.CO;2.
13. Sánchez Garrido, J. C., Sannino, G., Liberti, L., et al. (2011). Numerical modeling of three-dimensional stratified tidal flow over Camarinal Sill, Strait of Gibraltar. *Journal of Geophysical Research: Oceans, 116,* C12026. https://doi.org/10.1029/2011JC007093.
14. García Lafuente, J., Almazán, J. L., Castillejo, F., et al. (1990). Sea level in the strait of Gibraltar: tides. *International Hydrographic Review, 47,* 111–130.
15. Sánchez Garrido, J. C., García Lafuente, J., Criado Aldeanueva, F., et al. (2008). Time-spatial variability observed in velocity of propagation of the internal bore in the Strait of Gibraltar. *Journal Geophysical Research, 113,* C07034. https://doi.org/10.1029/2007JC004624.
16. Vlasenko, V., Sanchez Garrido, J. C., Stashchuk, N., et al. (2009). Three-dimensional evolution of large-amplitude internal waves in the strait of Gibraltar. *Journal of Physical Oceanography, 39,* 2230–2246. https://doi.org/10.1175/2009JPO4007.1.
17. Alberola, C., Rousseau, S., Millot, C., et al. (1995). Tidal currents in the western mediterranean-sea. *Oceanologica Acta, 18,* 273–284.
18. Alpers, W., Brandt, P., Rubino, A., & Backhaus, J. O. (1996). Recent contributions of remote sensing to the study of internal waves in the straits of Gibraltar and Messina. In *Dynamics of mediterranean straits and channels*. CIESM Science Series n°2, Briand F (pp. 21–40). Bulletin de l'Institut océanographique, Monaco.
19. García-Lafuente, J., Naranjo, C., Sammartino, S., et al. (2017). The Mediterranean outflow in the strait of Gibraltar and its connection with upstream conditions in the Alborán Sea. *Ocean Science, 13,* 195–207. https://doi.org/10.5194/os-13-195-2017.
20. Bryden, H. L., & Stommel, H. M. (1982). Origin of the Mediterranean outflow. *Journal of Marine Research, 40,* 55–71.
21. Naranjo, C., García Lafuente, J., Sánchez Garrido, J. C., et al. (2012). The Western Alboran Gyre helps ventilate the western Mediterranean deep water through Gibraltar. *Deep Sea Research Part I: Oceanographic Research Papers, 63,* 157–163. https://doi.org/10.1016/j.dsr.2011.10.003.
22. Parrilla, G., Kinder, T. H., & Preller, R. H. (1986). Deep and intermediate mediterranean water in the western Alboran Sea. *Deep Sea Research Part A. Oceanographic Research Papers, 33,* 55–88. https://doi.org/10.1016/0198-0149(86)90108-1.
23. Sánchez Román, A., Sannino, G., García Lafuente, J., et al. (2009). Transport estimates at the western section of the Strait of Gibraltar: A combined experimental and numerical modeling study. *Journal of Geophysical Research: Oceans, 114,* C06002. https://doi.org/10.1029/2008JC005023.
24. Tsimplis, M. N., & Bryden, H. L. (2000). Estimation of the transports through the strait of Gibraltar. *Deep Sea Research Part I: Oceanographic Research Papers, 47,* 2219–2242. https://doi.org/10.1016/S0967-0637(00)00024-8.
25. Naranjo, C., Sammartino, S., García-Lafuente, J., et al. (2015). Mediterranean waters along and across the strait of Gibraltar, characterization and zonal modification. *Deep Sea Research Part I: Oceanographic Research Papers*. https://doi.org/10.1016/j.dsr.2015.08.003.
26. García-Lafuente, J., Sammartino, S., Naranjo, C., & Sánchez Garrido, J. C. (2017). Potential temperature and practical salinity collected at Espartel Sill (Strait of Gibraltar), from 2004-09-30 to 2016-09-21 at 348 meters depth.

27. Sánchez Román, A., Criado Aldeanueva, F., García Lafuente, J., & Sánchez Garrido, J. C. (2008). Vertical structure of tidal currents over Espartel and Camarinal sills, strait of Gibraltar. *Journal of Marine Systems, 74,* 120–133. https://doi.org/10.1016/j.jmarsys.2007.11.007.
28. Pratt, L. L. J., & Whitehead, J. A. (2007). *Rotating hydraulics: nonlinear topographic effects in the ocean and atmosphere*. New York: Springer.
29. Lawrence, G. A. (1993). The hydraulics of steady two-layer flow over a fixed obstacle. *Journal of Fluid Mechanics, 254,* 605. https://doi.org/10.1017/S0022112093002277.
30. Luján, M., Crespo-Blanc, A., & Comas, M. (2011). Morphology and structure of the Camarinal Sill from high-resolution bathymetry: evidence of fault zones in the Gibraltar strait. *Geo-Marine Letters, 31,* 163–174. https://doi.org/10.1007/s00367-010-0222-y.
31. Pawlowicz, R., Beardsley, B., & Lentz, S. (2002). Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE. *Computers & Geosciences, 28,* 929–937. https://doi.org/10.1016/S0098-3004(02)00013-4.

Mode 2 Internal Waves in the Ocean: Evidences from Observations

Andrey N. Serebryany

Introduction

Until recently, it was believed that the higher modes of internal waves predicted by the theory are not very common in the real ocean. Indeed, it was reported about ubiquitous observations in the ocean of internal waves of the first mode, while observations of internal waves of higher modes in the ocean were rare. In recent years, the situation changed and reports began to appear about the observations of internal waves of mode 2. In this relation we mention the South China Sea where extensive internal wave studies have been conducted in the past decade, which have revealed the frequent occurrence of the second mode in intense internal waves, especially in winter. In this contribution, we present experimental evidences of internal wave mode 2 from our observations that we have performed in various regions of the ocean. In particular, we will consider examples from the observations over the Mascarene Ridge, in the Luzon Strait, and on the Black Sea shelf.

Mode 2 Waves Over the Mascarene Ridge

Oceanic ridges are special places in the ocean from the point of view of internal waves: internal waves of maximum amplitudes are observed over submarine ridges [1, 2]. A classic example of such location is the Mascarene Ridge in the Indian Ocean, where many measurements of internal waves were conducted [3, 4]. During the research cruise of the R/V "Academician Nikolai Andreyev" in November–December 1990 we collected the data on the generation of large-amplitude internal waves of mode 2 in this region [3].

A. N. Serebryany (✉)
Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
e-mail: Serebryany@hotmail.com

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_15

The measurement site over the Mascarene Ridge was in the region that encompasses the southern part of the Saya de Malya bank and the northern part of the Nazareth bank (Fig. 1). In this region, there is a strong alternating current, which consists of two main components: a quasi-constant mean flow directed from the east to the west and an alternating tidal current. The velocity of the quasi-constant flow reached 1 m/s, and the amplitude of the current of the semidiurnal tide was 0.5 m/s. In the course of the work, a region with intense internal waves above the underwater sill was discovered.

The observations of internal waves were carried out using line temperature sensors by means of towing of the line sensor in the upper thermocline. During the ship drift, a series of rapid vertical soundings ("yo-yo") of the CTD profiler were also conducted.

The measurement cycle over the bank included continuous measurements over several runs oriented across the underwater sill (Fig. 1). The towing was carried out at a ship speed of 7.5–8.5 knots; the lower end of the line sensor was located at a depth of 100–106 m. The sensor then covered the interval of the well-pronounced subsurface thermocline. A record of the line sensor during the passage above the sill is shown in Fig. 2.

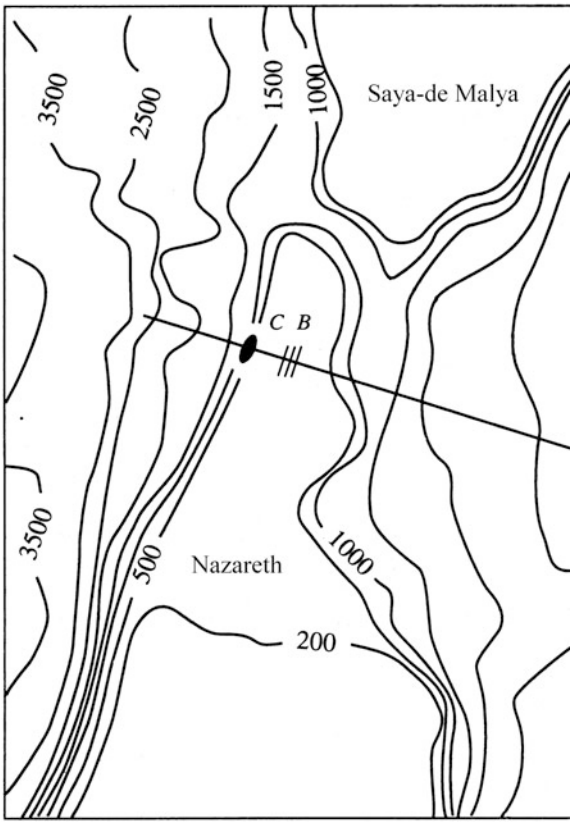


Fig. 1 Bottom topography in the measurement site and the track of the vessel during towing. Letters C and B denote the locations of internal waves of large amplitudes (C is related to the solitary wave mode 2, B is related to the train of soliton-like waves)

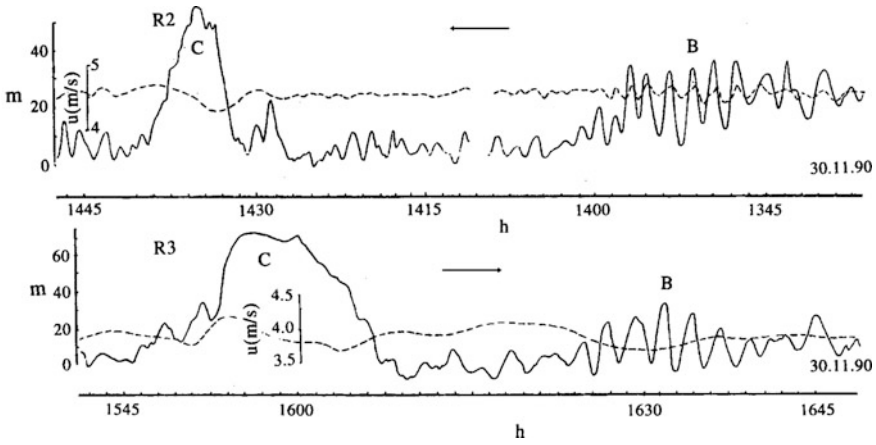


Fig. 2 A solitary elevation (C) and train of short-period waves (B) recorded by the towed line sensor on the opposite tacks above the sill. The log record is shown in dotted lines. Arrows show the ship's heading

During the first run of towing, when we approached the shallow part of the sill on the western side (left side in Fig. 1) we encountered a solitary elevation of a thermocline with a height of 79 m and a horizontal scale of about 3.8 km. This solitary wave was above the western slope of the sill over the depths of 700–900 m. A train of intense short-period waves (soliton-like) with characteristic heights of about 20 m was recorded in the subsequent section of the run after the passage of the solitary wave. Since the towed line sensor covered the upper part of the thermocline, it recorded a solitary elevation of the thermocline. In the lower part of the thermocline this wave disturbance was a solitary depression. This wave patter was clarified by the measurements from the drifting ship above the sill. On subsequent tacks we several times crossed the wave formations (solitary disturbance and a packet of short waves) encountered on the first tack, which allowed us to follow their evolution and understand the generation mechanism.

The vertical structure of the waves is clearly seen in Fig. 3. The structure was recorded using fast vertical casts with a CTD probe in the upper 200-m layer of the ocean, conducted every 5 min during the drift. The vertical cross sections relative to the low-frequency waves (Fig. 2, B and C) demonstrate the nature of oscillations of the second mode with the transition of the sign of vertical displacements below a depth of 100 m. The short-period waves in the train are of the similar nature, i.e. they show anti-phase oscillations in the upper and lower layers.

The numeric modeling carried out in the region of the Mascarene Ridge [5] confirmed the formation of a solitary disturbance in the form of an internal wave of the second mode on the western side of a submarine sill.

The collected data allow us to present the pattern of the internal wave field and its evolution in the form of a scenario, which to our opinion in general coincides with the known mechanism of the formation of lee waves. A significant wave

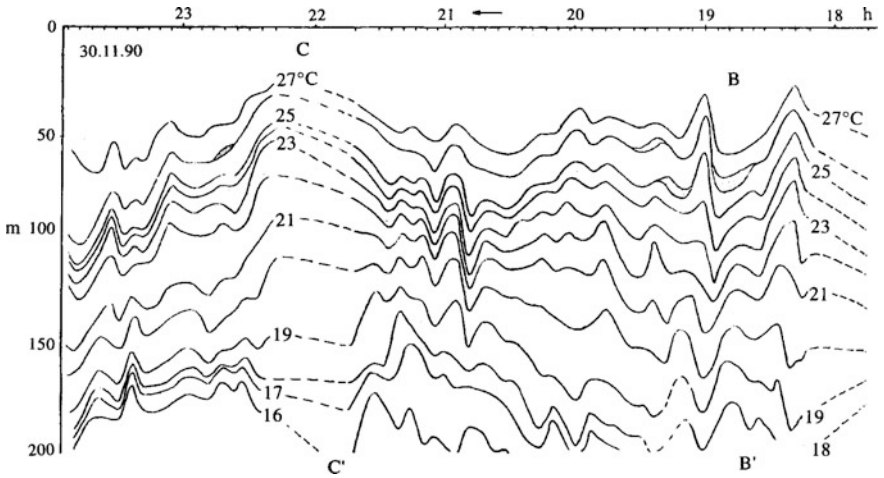


Fig. 3 Temperature section (isotherms) of a solitary disturbance of mode 2 (C) and a region of short-period waves (B), measured by "yo-yo" soundings during the drift above the sill

perturbation in the thermocline (solitary internal wave of mode 2) is formed on the western (leeward) side of the sill when the directions of the tidal current and the quasi-constant flow coincide. At the moment of weakening of the tidal current directed to the west, the solitary wave disturbance begins to move to the east towards the quasistationary mean flow. It moves to the shallow part of the sill, where it transforms and decays into a train of short internal waves resembling solitons. Then, the short-period internal waves remain within the bank and eventually collapse here failing to overcome a sufficiently strong counterflow.

Thus, in the central part of the ridge we detected intense internal waves: a single perturbation of the second mode in the form of a hydraulic internal jump formed by the current behind the underwater sill and the region above the sill occupied by short internal soliton-like waves. The process of transformation of a solitary disturbance propagating over a shallow bank during the ebb phase was investigated with its decay into the soliton-like waves. We emphasize especially that the short internal waves generated by a solitary wave also retained the features of belonging to the oscillations of the second mode. It should also be noted that these facts represented the first observation of the internal hydraulic jump of the second mode in the ocean.

Mode 2 Waves in the Luzon Strait

Over the past 15 years, regular studies of internal waves in the South China Sea, which is the World Ocean region with the largest internal waves (height more than 100 m) have been carried out [6, 7]. These studies revealed frequent occurrence of

the second mode in intense internal waves of the South China Sea, especially in winter [8]. Internal waves of the second mode have been recorded here many times over the continental slope and on the shelf using autonomous moored stations equipped with vertical chains of thermistors and current meters [8, 9].

In 2006, while conducting studies on the underwater ambient noise associated with internal waves in the Luzon Strait, we encountered there a large-amplitude solitary wave of the second mode [10]. The measurements were carried out from the R/V "Ocean Researcher 1", which was drifting. The vessel recorded a wave of mode 2 at the point with coordinates 21.112°N, 119.999°E north-west of the underwater ridge Heng Chun on May 23 at 11:40 local time. The passage of the wave was accompanied by sea surface manifestations in the form of a solitary band of choppy surface waves, clearly seen on the radar screen (Fig. 4). The wave propagated to the northwest (318°N) with a speed of about 3 m/s. The wave parameters were recorded using radar images, 150 kHz ADCP, as well as an EK 500 echo sounder at a frequency of 38 MHz. Figure 5 shows a sonar record, in which one can see manifestation of the wave. The wave was of the "convex type" of mode 2 with an amplitude of about 50 m. The propagation of the internal wave was accompanied by an increase in the underwater acoustic noise recorded by omnidirected hydrophone. After the observations, the ship moved to the eastern part of the Luson Strait and until the end of the cruise we no longer recorded any the solitary internal waves. At the same time, numerous cases of packets of short-period internal waves were observed.

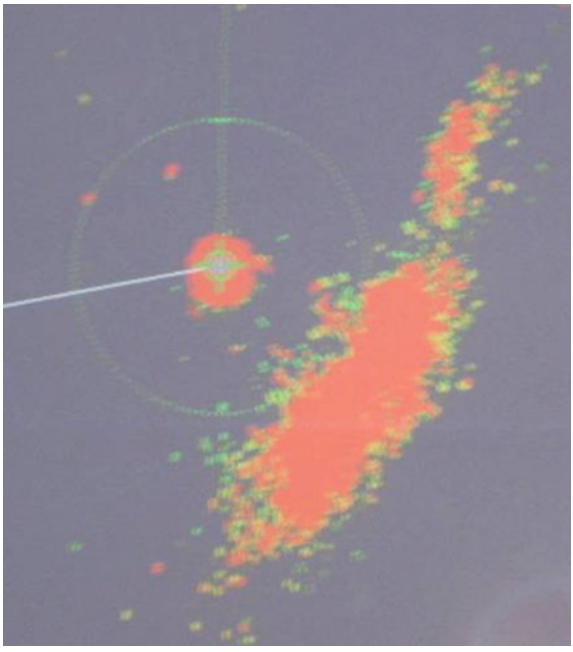


Fig. 4 Ship radar image of the solitary wave taken at 11:42 on May 23, 2006. The radius of circle is 1 mile. Rip band width is close to 1300 m

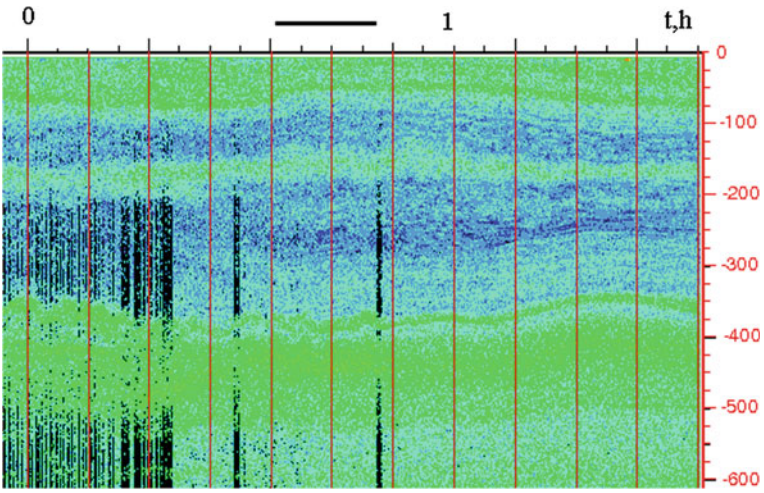


Fig. 5 Acoustic 38-kHz record visualizing the solitary internal wave based on the measurements with the EK 500 echosounder in the Luzon Strait experiment on May 23, 2006. The existence of the second-mode-type of the 50-m wave in the upper 600-m layer of the sea is evident. The horizontal line shows the time interval, within which the rips appeared

We conclude on the basis of observations that the waves of the second mode (at least some of them) observed over the continental slope and the shelf of the South China Sea are long-living waves. They were generated as waves of mode 2 in the Luzon Strait, and then they crossed the entire northwestern part of the sea.

Mode 2 Waves in the Black Sea

In 2011, the internal inertial waves of mode 2 were for the first time recorded on the Black Sea shelf [11]. Observations were carried out from the stationary oceanographic platform of the Marine Hydrophysical Institute from 9 to 18 July. The depth of the sea at the place of observations (shelf of the southern coast of Crimea) is 30 m. During the entire measurement period the weather was calm. There was a gradual warming of the water column temperature, which was accompanied by a monotonous depression of the thermocline. Thus, in the beginning of the measurements, the thermocline was at a depth of 7–8 m, and by the end its location was at the bottom at a depth of 25–26 m. In the first few days of measurements, the thermocline performed oscillations with a period close to 17 h (which corresponds to the local inertial period). On July 14, the thermocline displacements changed when its position was in the middle of the water column. Subsequently, quasi-inertial oscillations were recorded in antiphase, which was characterized as a manifestation of inertial internal waves of the second mode. Figure 6 illustrates the

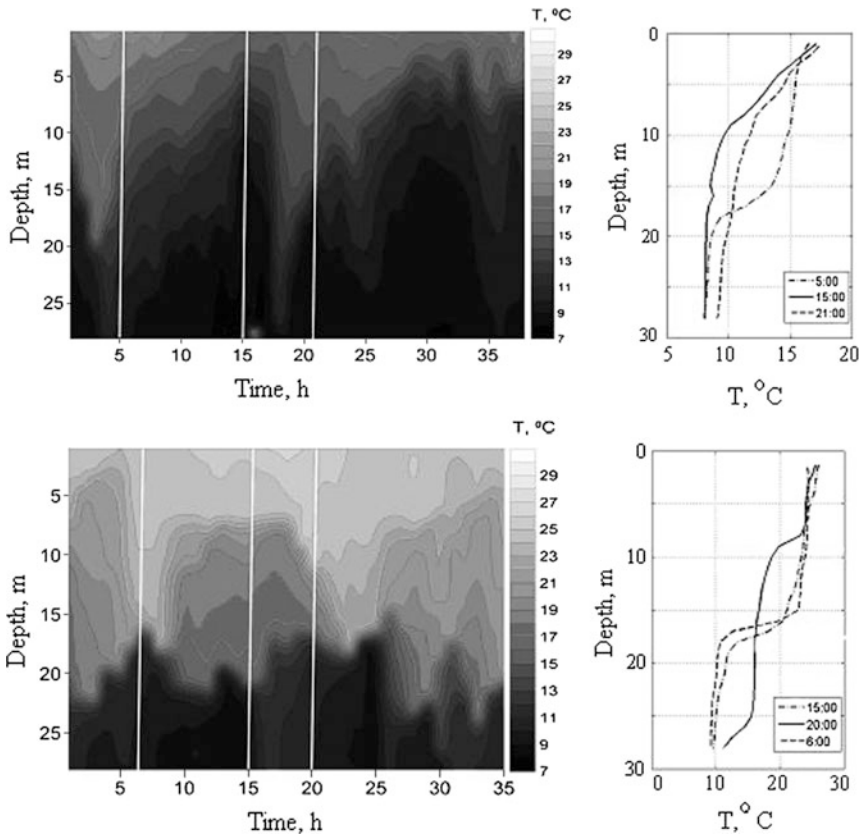


Fig. 6 Manifestation of internal inertial waves of mode 1 (top panel) and mode 2 (bottom panel). The 17-h temperature oscillations of the water column were recorded on July 9–10 (mode 1) and July 16–17 (mode 2). In addition, temperature profiles made at specific hours are shown (marked by vertical lines in the sections)

characteristic manifestations of internal waves of the inertial period of mode 1 and mode 2. The records show time sections of temperature in depth, performed according to hourly CTD soundings. During the passage of the internal wave of mode 1 (Fig. 6a), 17-h vertical displacements of the entire water column are clear seen, while the sharpened troughs of long waves are visible, which indicates their nonlinearity. The range of oscillations of the thermocline was 15–20 m. During the observations in 2011, five cycles of similar synchronous oscillations of the thermocline were recorded. Subsequently, the character of the inertial oscillations changed to oscillations in the antiphase. The most pronounced character of these oscillations was observed on July 16–17, which is illustrated in Fig. 6. It is seen from the figure that during the 17-h period the water layers rise from a depth of 16 m to a depth of 7–8 m, after which they again occupy the initial position. At the same time, the layers of the water column are displaced down from 16 m to a depth

of 21 m. Such motions can be attributed to internal wave mode 2 of the convex type. The amplitude of these waves was as high as 7 m.

Conclusions

We presented experimental evidences of the existence of internal wave mode 2 in different regions of the ocean on the basis of our observations. In addition to the geographical prevalence of this phenomenon, these examples demonstrated the variety of types of internal waves of mode 2. We observed a tide-generated solitary wave of large amplitude near the underwater ridge in the Indian Ocean. Near the underwater ridge in the Luzon Strait, we observed a 50-m solitary wave of mode 2. Its generation was associated with the tidal currents. In the shelf zone of the non-tidal Black Sea, internal inertial waves of both mode 1 and mode 2 were recorded. Thus, we have demonstrated the occurrence of mode 2 internal waves in the real oceanic conditions, although we must recognize that mode 1 is a more widespread phenomenon in the field of oceanic internal waves.

Acknowledgements This work was supported by the Russian Science Foundation (project no. 14-50-00095).

References

1. Morozov, E. G. (1995). Semidiurnal internal wave global field. *Deep Sea Research, 42*(1), 135–148.
2. Sabinin, K. D., & Serebryany, A. N. (2005). Intense internal waves in the ocean. *Journal of Marine Researh*, *63*(1), 227–261.
3. Konyaev, K. V., Sabinin, K. D., & Serebryany, A. N. (1995). Large-amplitude internal waves at the Mascarene Ridge in the Indian Ocean. *Deep Sea Research, 42*(11–12), 2075–2091.
4. Morozov, E. G., & Vlasenko, V. I. (1996). Extreme tidal internal waves near the Mascarene Ridge. *Journal of Marine Systems, 9*(3–4), 203–210.
5. Vlasenko, V. I., & Morozov, E. G. (1993). Generation of semidiurnal internal waves near a submarine ridge. *Oceanology, 33*(3), 282–286.
6. Alford, M. H., Peacock, T., MacKinnon, J. A., et al. (2015). The formation and fate of internal waves in the South China Sea. *Nature, 521*(7550), 65–69. https://doi.org/10.1038/nature14399.
7. Duda, T., Lynch, J., Irish, J., Beardsley, R., Ramp, S., Chiu, C.-S., et al. (2004). Internal tide and nonlinear internal wave behavior at the continental slope in the Northern South China Sea. *IEEE Journal of Oceanic Engineering, 29*(4), 1105–1130.
8. Yang, Y. J., Fang, Y. C., Chang, M.-H., Ramp, S. R., Kao, C.-C., Tang, T. Y. (2009). Observations of second baroclinic mode internal solitary waves on the continental slope of the northern South China Sea. *Journal of Geophysical Research: Oceans*, *114* (C10003). https://doi.org/10.1029/2009jc005318.
9. Yang, Y. J., Fang, Y. C., Tang, T. Y., & Ramp, S. R. (2010). Convex and concave types of second baroclinic mode internal solitary waves. *Nonlinear Processes in Geophysics, 17,* 605–614. https://doi.org/10.5194/npg-17-605-2010.

10. Serebryany, A., Liu, C.-T. (2012). Observations of large-amplitude internal wave of the second mode in Luzon Strait. *Geophysical Research Abstracts*, *14*: EGU2012–6431.
11. Serebryany, A. N., & Khimchenko, E. E. (2014). Observations of internal waves at Caucasus and Crimean shelves of the Black Sea in summer 2013 (in Russian). *Current Problems in Remote Sensing of the Earth from Space, 11*(3), 88–104.

Abyssal Mixing in the Laboratory

T. Dauxois, E. Ermanyuk, C. Brouzet, S. Joubaud and I. Sibgatullin

Motivation

The continuous energy input to the ocean interior comes from the interaction of global tides with the bottom topography yielding a global rate of energy conversion to internal tides of the order of 1TW [1, 2]. The subsequent mechanical energy cascade to small-scale internal-wave motion and mixing is a subject of active debate in view of the important role played by abyssal mixing in existing models of ocean dynamics. The oceanographic data support the important role of internal waves in mixing, at least locally: increased rates of diapycnal mixing are reported [3] in the bulk of abyssal regions over rough topography in contrast to regions with smooth bottom topography. A question remains: how does energy injected through internal waves at large vertical scales induce the mixing of the fluid?

Using laboratory experiments and numerical simulations, we suggest the energy cascade in internal wave attractors as a novel laboratory model of a natural cascade. We show that energy transfer from global to small scales in attractors operates via a hierarchy of triadic resonant interactions producing a complex internal wave

T. Dauxois (✉) · C. Brouzet · S. Joubaud
Univ Lyon, ENS de Lyon, Univ Claude Bernard Lyon 1, CNRS,
Laboratoire de Physique, 69342 Lyon, France
e-mail: thierry.dauxois@ens-lyon.fr

C. Brouzet
e-mail: Christophe.brouzet@ens-lyon.fr

S. Joubaud
e-mail: Sylvain.Joubaud@ens-lyon.fr

E. Ermanyuk
Lavrentyev Institute of Hydrodynamics, Novosibirsk, Russia
e-mail: ermanyuk@hydro.nsc.ru

I. Sibgatullin
Faculty of Mechanics and Mathematics, Moscow State University, Moscow, Russia
e-mail: sibgat@ocean.ru

I. Sibgatullin
Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_16

field with a rich multi-peak discrete frequency spectrum embedded into a continuous spectrum of weaker magnitude. Convincing evidences of a wave turbulence framework [4] for internal waves are also provided. Spontaneous summation of the wave-field components produces moreover a statistically significant amount of extreme overturning events which eventually lead to a well-measurable mixing. This suggests that such a set-up is appropriate to study abyssal mixing in the laboratory.

Let us consider a stratified fluid with an initially constant buoyancy frequency $N = [(-g/\bar{\rho})(\mathrm{d}\rho/\mathrm{d}z)]^{1/2}$, where $\rho(z)$ is the density distribution over the vertical coordinate z, $\bar{\rho}$, and g the gravity acceleration. The dispersion relation for linear internal gravity waves is given by

$$\theta = \pm \arcsin \Omega, \tag{1}$$

where θ is the slope of the wave beam to the horizontal, and Ω (resp. $\omega = \Omega N$) the non-dimensional (resp. dimensional) frequency of oscillations. This anisotropic dispersion relation requires preservation of the slope of the internal wave beam upon reflection at a rigid boundary. As we will recall below, in the case of a sloping boundary, this property gives a purely geometric reason for a strong variation of the width of internal wave beams (focusing or defocusing) upon reflection. Internal wave focusing provides a necessary condition for large shear and overturning, as well as shear and bottom layer instabilities at slopes.

In a confined fluid domain, focusing usually prevails, leading to a concentration of wave energy on a closed loop, the internal wave attractor [5]. At the level of linear mechanisms, the width of the attractor branches is set by the competition between geometric focusing and viscous broadening. (see [6] for a precise and interesting of this question). High concentration of energy at attractors make them prone to triadic resonance instability which sets in as the energy injected into the system increases [7]. Note that the particular case for which both secondary waves have a frequency equal to half of the forcing frequency is of special interest in the oceanographic context where viscosity is negligible. In that case, the appropriate name is parametric subharmonic instability and abbreviated as PSI. By abuse of language, some authors have sometimes extended the use of the name PSI to cases for which secondary waves are not corresponding to half of the forcing frequency. For the sake of terminological consistency, we propose to abbreviate triadic resonance instability using the acronym TRI.

The onset of instability is similar to the classic concept of triadic resonance, which is best studied for the idealized case, with monochromatic in time and space carrier wave as a basic state which feeds two secondary waves via nonlinear resonant interactions. The resonance occurs when temporal $\Omega_1 + \Omega_2 = \Omega_0$ and spatial $\mathbf{k_1} + \mathbf{k_2} = \mathbf{k_0}$ conditions are satisfied ($\mathbf{k}$ is the wave vector while subscripts 0, 1 and 2 refer to the primary, and two secondary waves, respectively). In a wave attractor, the wave beams serve as a primary wave, and the resonance conditions are satisfied with a good accuracy [7], providing a consistent physical framework for the short-term behavior of the instability.

The usual theory for the TRI does not take into account the finite width of the experimental beam. Qualitatively, the subharmonic waves can serve as an energy

sink for the primary wave if they do not leave the primary beam before they can extract substantial energy [8]. The group velocity of the primary wave is aligned with the beam, but the group velocity of the secondary waves is definitely not, and these secondary waves eventually leave the primary wave beam. This is a direct consequence of the dispersion relation (1), which relates the direction of propagation to the frequency: a different frequency, smaller for subharmonic waves, will lead to a shallower angle. The generalization to wave beams with a finite width is presented in detail in a recent review [9].

The Internal Wave Attractor to Enhance the Nonlinearities

In this section, we present the concept of internal wave attractor that is a key element to provide large amplitude internal wave beams produced thanks to the focusing properties of internal wave reflections.

Reflection of Internal Waves: A Focusing Mechanism

The dispersion relation of internal waves is very specific and leads to a very unusual reflection on a sloping boundary that has interesting properties, central for our objective as it will be immediately clear. To be more specific, let us consider an inviscid linearly stratified fluid of constant buoyancy frequency N and a sloping boundary, tilted with an angle α with respect to the vertical, as shown in Fig. 1. Note that, if this configuration does not seem natural for the reflection of internal waves on the topography at the bottom of the ocean, it corresponds by symmetry to a case with negative values of α, but is simpler for an experimental realization.

As the pulsation of the wave is conserved during the reflection, both the incident and reflected waves propagate with the same angle θ, according to the dispersion relation (1). It is worth to note that this is very different from the reflection in optics or acoustics where the electromagnetic or sound waves conserve the angle with respect to the normal to the sloping boundary, refereed usually as the classical Descartes reflection. For internal waves, this is the angle with the gravity that is conserved. This difference is illustrated in Fig. 1a. The reflected ray for optics or acoustics is the dashed arrow while the one for internal waves is the solid arrow.

This non-Descartes reflection is even more intriguing, and therefore interesting, when one considers a beam, and not only a ray. This is shown in Fig. 1b. The width of the reflected beam is thus reduced and one gets an energy focusing for these internal waves. It is important to emphasize that this phenomenon being a direct consequence of the linear dispersion relation, one has identified here a linear transfer toward smaller scales.

Note that for the two specific cases of a wall being vertical or horizontal ($\alpha = 0$ or $\pi/2$), the reflection of an internal wave is similar to a classical Descartes reflection, without any focusing effect.

The Focusing Parameter

It is of course possible to get more quantitative results. The linear theory of internal wave reflection has been developed first by Phillips [10] and is based on a well-known incident wave reflecting at a sloping boundary. Let assume that the incident wave is bi-dimensional (in the vertical plane) and can be described by the stream function

$$\psi_{\mathrm{i}}(x,z,t) = \psi_{0,\mathrm{i}} \exp\left[i\left(\omega_{\mathrm{i}}\, t - \mathbf{k}_{\mathrm{i}} \cdot \mathbf{r}\right)\right]. \tag{2}$$

The index i refers to the incident wave field, while ω_{i} and $\mathbf{k}_{\mathrm{i}}$ are the pulsation and wave vector of the incident wave and satisfy the dispersion relation (1). With u and w the horizontal and vertical velocity fields, the kinetic energy density is defined as

$$E_{c,\mathrm{i}} = \frac{1}{2}\bar{\rho}\left(|u|^2 + |w|^2\right) = \frac{1}{2}\bar{\rho}\left(|\partial_z \psi_{\mathrm{i}}|^2 + |\partial_x \psi_{\mathrm{i}}|^2\right) = \frac{1}{2}\bar{\rho} k_{\mathrm{i}}^2 |\psi_{0,\mathrm{i}}|^2. \tag{3}$$

When the incident wave hits the sloping boundary, a reflected wave is generated and can be expressed as follows

$$\psi_{\mathrm{r}}(x,z,t) = \psi_{0,\mathrm{r}} \exp\left[i\left(\omega_{\mathrm{r}}\, t - \mathbf{k}_{\mathrm{r}} \cdot \mathbf{r}\right)\right], \tag{4}$$

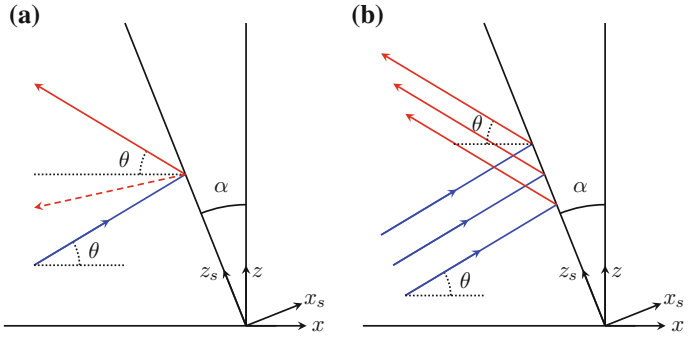


Fig. 1 **a** Reflection of an incident ray on the sloping boundary inclined with an angle α with respect to the vertical. In optics or acoustics, the reflected ray is along the dashed arrow, while for internal waves, the reflected ray is along the solid arrow. **b** Reflection of an incident internal wave beam on a sloping boundary for internal waves. The slope coordinates (x_s, z_s) are shown on both panels

where the index r refers to the reflected wave field. Its kinetic energy density corresponds to $E_{c,\mathrm{r}} = \frac{1}{2}\bar{\rho}k_{\mathrm{r}}^2|\psi_{0,\mathrm{r}}|^2$. The complete wave field is therefore $\psi = \psi_{\mathrm{i}} + \psi_{\mathrm{r}}$.

As the flow does not penetrate the sloping boundary, the total stream function field must vanish at the boundary defined by $x = -z\tan\alpha$. In order to simplify the boundary condition, one usually defines the coordinates attached to the slope (x_s, z_s), as shown in Fig. 1. The velocity fields in the slope coordinate system are $(u_s, w_s) = (-\partial\psi/\partial z_s, \partial\psi/\partial x_s)$ and the wave number is noted $\mathbf{k}_s = (k_{x_s}, k_{z_s})$. The non-penetration condition can be expressed as $u_s = 0$ at $x_s = 0$ and for all z_s and time t. On the total stream function field, this becomes

$$k_{z_s,\mathrm{i}}\psi_{0,\mathrm{i}}\exp\left[i\left(\omega_{\mathrm{i}}\,t - \mathbf{k}_{s,\mathrm{i}}\cdot\mathbf{r}_s\right)\right] + k_{z_s,\mathrm{r}}\psi_{0,\mathrm{r}}\exp\left[i\left(\omega_{\mathrm{r}}\,t - \mathbf{k}_{s,\mathrm{r}}\cdot\mathbf{r}_s\right)\right] = 0, \tag{5}$$

at $x_s = 0$ and for all z_s and time t. This leads to $\omega_{\mathrm{i}} = \omega_{\mathrm{r}} \equiv \omega$, $k_{z_s,\mathrm{i}} = k_{z_s,\mathrm{r}}$ and $\psi_{0,\mathrm{i}} = \psi_{0,\mathrm{r}} \equiv \psi_0$. Thus, the frequency and the wave vector component parallel to the sloping boundary are conserved during the reflection. The normal component of the wave vector can be determined using geometrical construction and the dispersion relation: one gets $k_{x_s,\mathrm{i}} = k_{z_s,\mathrm{i}}\tan(\theta - \alpha)$ and $k_{x_s,\mathrm{r}} = k_{z_s,\mathrm{r}}\tan(\theta + \alpha)$. Thus, the ratio between the norms of the two wave vectors is given by

$$\frac{k_{\mathrm{r}}}{k_{\mathrm{i}}} = \left|\frac{\cos(\theta - \alpha)}{\cos(\theta + \alpha)}\right| \equiv \gamma. \tag{6}$$

This defines the focusing parameter γ. Equation (6) confirms immediately that there is neither focusing, nor defocusing when the wall is vertical ($\alpha = 0$) or horizontal ($\alpha = \pi/2$): both cases lead indeed to $\gamma = 1$. One recovers indeed the case of the Descartes reflection, since keeping the angle with respect to the gravity for the internal waves, does correspond to keep the angle with respect to the normal of the wall (that is orthogonal or parallel to the gravity!).

However as soon as one considers a sloping boundary ($\alpha \neq 0$) or ($\alpha \neq \pi/2$), the focusing parameter is different from unity: it is for example greater than 1 in Fig. 1. The width of the reflected beam is thus reduced by the factor γ. This is a focusing reflection, the energy in the incident beam being concentrated. Indeed, Eq. (6) leads to $E_{c,\mathrm{r}} = \gamma^2 E_{c,\mathrm{i}}$ showing that the energy density is increased by a factor $\gamma^2 > 1$.

Interestingly, this parameter diverges when $\theta + \alpha$ tends to 90°. This corresponds to the case where the waves have a propagation angle very close to the slope of the wall. This situation is called critical reflection. Indeed, it is critical because γ diverges and thus, the wave length of the reflected wave tends to 0: nonlinear and dissipation effects cannot be overlooked and should be treated carefully. Using a weakly nonlinear theory, it has been shown [11] how to heal this singularity using matched asymptotic expansion. This is however not the case under study in the remainder of this work.

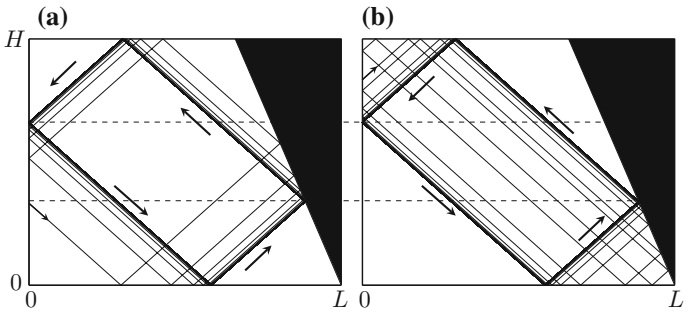


Fig. 2 Convergence towards the same attractor of two rays starting from two different points in the same geometry. The small arrows on the left vertical wall indicate the two positions where the rays start, while the large arrows show the direction of propagation of the periodic ray on the attractor. The two horizontal dashed lines show that the two attractors reached are exactly identical

An Internal Wave Billiard

If one considers now a closed basin as illustrated in Fig. 2, one realizes that the above focusing mechanism will lead to an extremely efficient focusing phenomenon. After the first reflection on the sloping wall, the beam depicted in Fig. 1b will reflect on the surface, then on the left vertical wall, and then on the bottom horizontal surface. These three Descartes-like reflections ($\alpha = 0$ or $\pi/2$) do not change neither the energy density nor the norms of the two wave vectors. However the following reflection on the sloping wall will again reduce the beam by a similar factor: it is straightforward to understand that after a few loops the beam will be extremely narrow, and indeed its width will inevitably vanish in the limit of infinitely many reflections, leading to a single ray bouncing on the walls.

Internal wave ray tracing in different closed basin shapes has been essentially studied by Leo Maas over the last twenty years. This can be viewed as an internal wave billiard [12]. The classical billiard studies the trajectories in a closed domain of a particle reflecting elastically and following the standard Descartes reflection. It can exhibit periodic motion, motion along an invariant curve or chaos [13].

In a trapezoidal domain, different attractors have been identified and carefully studied [5]. They are labelled using two indices: the number of reflections at the surface (or at the bottom) and the number of reflections on the vertical side wall (or on the slope). Figure 2 presents an attractor with only one reflection on the surface and on the vertical wall: a (1,1) attractor as the one we will use in the remainder of the paper.

However, the internal attractor is not the goal of our study, but rather the tool to drive strong instabilities within the fluid. As we have understood from the above discussion, already within the linear regime, such a wave attractor has an extremely efficient focusing power and nonlinearity will come into play, leading to triadic resonance instabilities that will drive efficiently the wave turbulence.

Experimental and Numerical Set-Up

We have combined numerical and experimental approaches to study the dynamics of stable and unstable internal wave attractors. The problem is considered in a classic trapezoidal set-up filled with a uniformly stratified fluid. Energy is injected into the system at global scale by the small-amplitude motion of a vertical wall.

Experimental Set-Up

The experimental set-up [7, 16] is sketched in Fig. 3. Experiments are conducted in a rectangular test tank of size $80 \times 17 \times 42.5$ cm^3 filled with uniformly stratified fluid using the conventional double-bucket technique. Salt is used as a stratifying agent. The density profile is measured prior and after experiments by a conductivity probe attached to a vertical traverse mechanism. The value of the buoyancy frequency N is evaluated from the measured density profile. The trapezoidal fluid domain of length L (measured along the bottom) and depth H is delimited by a sliding sloping wall, inclined at the angle α. The wall is slowly inserted into the fluid after the end of the filling procedure. The input forcing is introduced into the system by an internal wave generator [18, 19]. The time-dependent vertical profile of the generator is prescribed in the form

$$\zeta(z,t) = a \sin(\omega_0 t) \cos(\pi z/H), \tag{7}$$

where a and ω_0 are the amplitude and frequency of oscillations, respectively. In a horizontally semi-infinite domain, the motion of the generator would generate the first vertical mode of internal waves. The profile given in Eq. (7) is reproduced in discrete

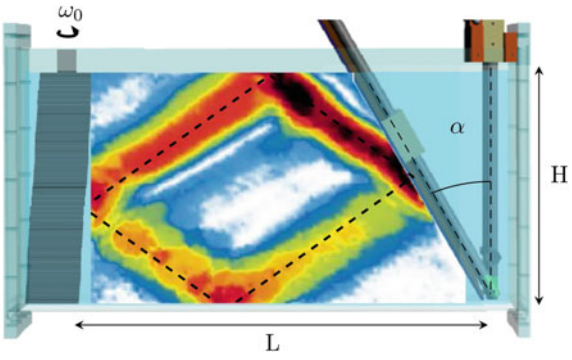


Fig. 3 The wave generator is on the left and the inclined slope on the right. A typical PIV snapshot showing the magnitude of the experimental two-dimensional velocity field obtained after 15 periods $T_0 = 2\pi/(N\Omega_0)$ of forcing is presented. Dashed lines show the billiard geometric prediction of the attractor

form by the horizontal motion of a stack of 47 plates. The whole-field velocity measurements are performed via the standard Particle Image Velocimetry (PIV) technique. To achieve this, the fluid is seeded with light-reflecting hollow glass spheres of size 8 μm and density 1100 kg/m^3. The sedimentation velocity of such particles is found to be very low, with negligible effect on results of velocity measurements. The longitudinal $(x, y = 0, z)$ mid-plane of the test section is illuminated by a vertical laser sheet coming through the side of the tank. The PIV acquisition leads to 2 velocity fields per second. This sampling rate is found to provide a sufficient resolution of the significant frequency components of the signal and the mesh of measurements is also found to be sufficient to resolve the small-scale details of the wave field.

Numerical Computations

Numerical computations were performed with a spectral element method [14, 15], using the modified open code nek5000. The geometry of the numerical set-up closely reproduces the experimental one. The full system of equations being solved consists of the Navier-Stokes equation in the Boussinesq approximation, the continuity equation and the equation for the transport of salt. Typical meshes used in calculations consist of 50 thousands to half-million elements, with 8 to 10-order polynomial decomposition within each element. Time discretization was 10^{-4} to 10^{-5} of the external forcing period. Comparisons of experimental and numerical results present a beautiful agreement, not only qualitative but also quantitative [16].

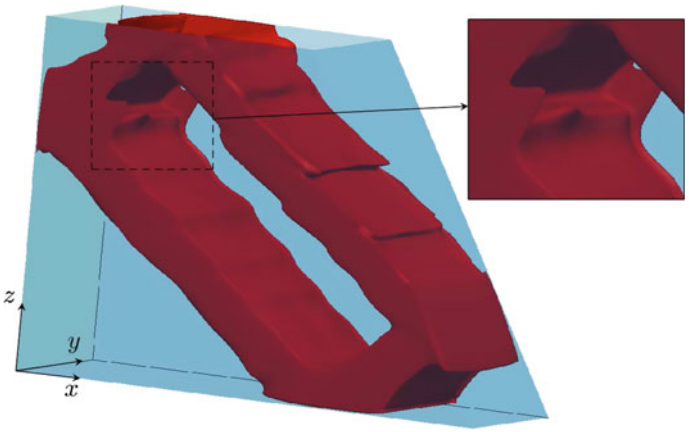


Fig. 4 Visualization of the three dimensional effects in the internal wave attractor. Snapshot of the instantaneous magnitude of the velocity field $(u^2 + w^2)^{1/2}$ produced by the 3D numerical simulations based on the spectral element method for $a = 2.4$ mm. The snapshot corresponds to a contour plot (level 0.25 cm/s) of the amplitude of the velocity field $(u^2 + w^2)^{1/2}$ at $t = 50\,T_0$. The inset presents a zoom to emphasize an example in which the flow field is dependent of the transverse direction

Figure 4 reveals the onset of TRI in the attractor. The numerical simulations clearly emphasize the importance of boundary layers close to the walls, and thus the importance of the three-dimensionality to recover experimental laboratory results quantitatively, nevertheless 2D simulations are fully sufficient for qualitative agreement. We checked [7, 16], that the temporal and spatial resonance conditions of TRI are satisfied experimentally and numerically.

Comparison

We have carefully compared the results obtained experimentally and numerically. Two examples are shown in Figs. 5 and 6, respectively for a stable and an unstable attractor. In both cases, the wave frequency is $\Omega_0 = 0.62 \pm 0.01$. Note that in the calculation, a piecewise linear approximation of the experimental density profile has been taken, with the lower layer of depth $H' = 30.8$ cm and buoyancy frequency N, and the upper layer of depth $\delta = 1.8$ cm with a density gradient 8 times smaller. The total depth of the fluid is therefore $H = H' + \delta = 32.6$ cm. The comparison emphasizes how precise are the two approaches, especially if one notes that the shade scale is the same in both panels.

More detailed comparisons can be found in [16, 17], where we showed that the results of three-dimensional calculations are in excellent qualitative and quantitative agreement with the experimental data, including the spatial and temporal parameters of the secondary waves produced by triadic resonance instability. Further, we explored experimentally and numerically the effect of lateral walls on secondary currents and spanwise distribution of velocity amplitudes in the wave beams.

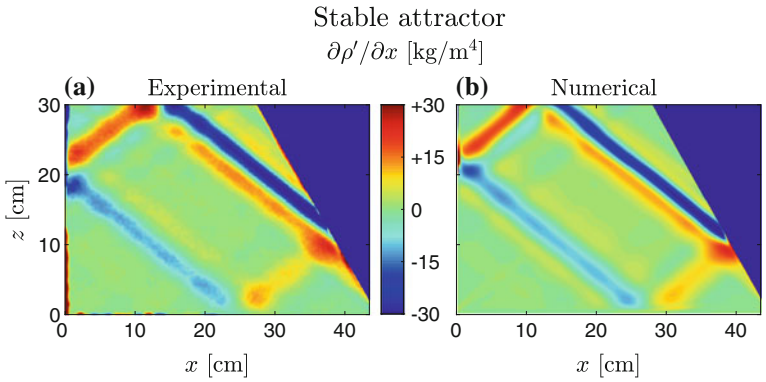


Fig. 5 Experimental (**a**) and numerical (**b**) snapshots of the horizontal density gradient at $t = 50\ T_0$ for a stable attractor. The amplitude of the wave maker is $a = 2$ mm for the experiment and $a = 1.8$ mm for the simulation

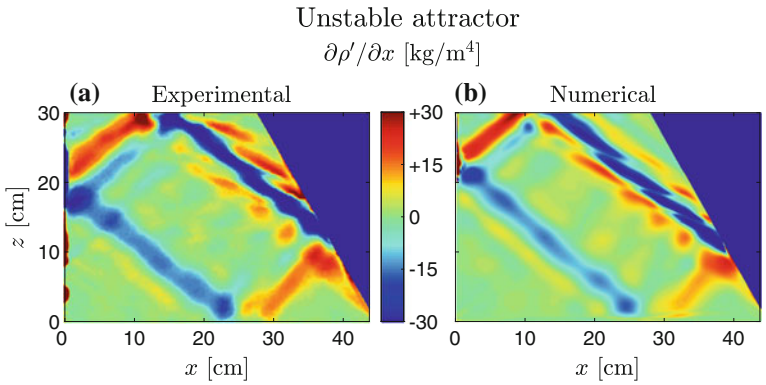


Fig. 6 Experimental (**a**) and numerical (**b**) snapshots of the horizontal density gradient at $t = 50\,T_0$ for an unstable attractor. The amplitude of the wave maker is $a = 2.5$ mm for the experiment and $a = 2.4$ mm for the numerical simulation

Finally, we tested the assumption of a bidimensional flow and estimated the error made in synthetic schlieren measurements due to this assumption.

The Energy Cascade Revealed by the Time-Frequency Diagram and the Bicoherence Plot

Using laboratory experiments and numerical simulations, we have shown that the internal wave attractor set-up, sketched in Fig. 3, provides an excellent energy cascade, emphasizing how internal wave attractors can be a novel laboratory model of a natural cascade [20].

Indeed, the internal wave attractor is the first step: the focalisation mechanism enhances the development of the triadic instability within the beams of the attractor. While the attractor is still visible, branches are progressively deformed by triadic resonance instability, leading to the presence of secondary waves.

An example of an experimental velocity field is shown in Fig. 7 at a late stage. The attractor is still visible, but branches are deformed by the presence of secondary waves. As it will be clear below, the internal wave frequency spectrum which was initially a Dirac function has been progressively enriched to give rise to a very complex spectrum, through a cascade of central interest.

Once the instability is well-developed, secondary waves are acting as primary waves for higher-order triadic interactions. If the focalisation is strong enough, this mechanism will of course repeat through the instability of the secondary waves. This is what is revealed by the time frequency diagram shown in Fig. 8a. Initially, only a signal around $\Omega_0 = 0.61$ is present, but almost immediately one distinguishes two secondary waves $\Omega_1 = 0.36$ and $\Omega_2 = 0.25$ whose sum gives Ω_0. However, again Ω_1 and Ω_2 are destabilized and this mechanism is pursued.

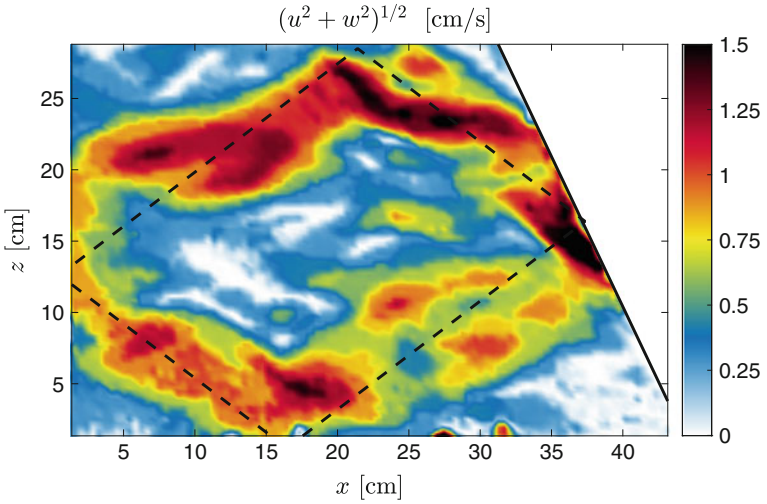


Fig. 7 Magnitude of the experimental two-dimensional velocity field at $t = 400\,T_0$ for $a = 5$ mm. Dashed lines show the billiard geometric prediction of the attractor, which is fully recovered when considering small forcing amplitude or at an earlier time when considering larger forcing

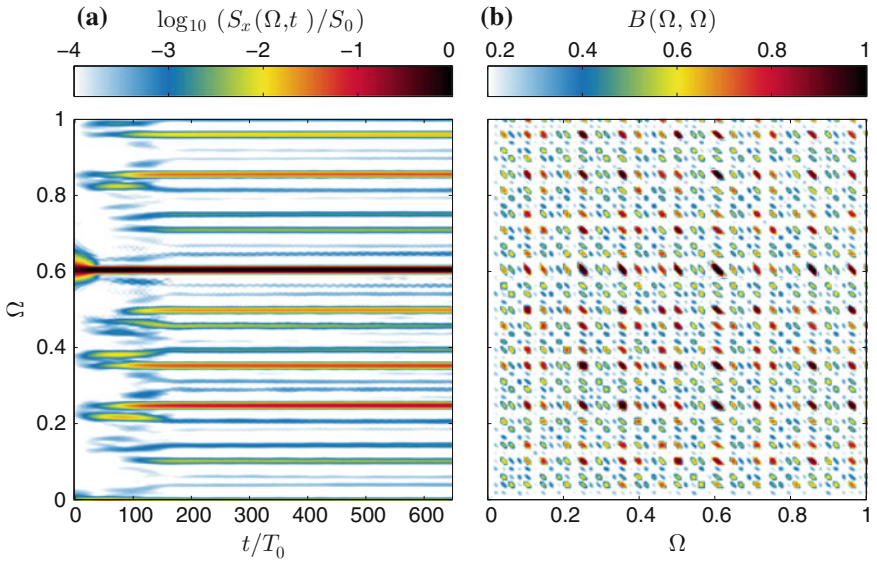


Fig. 8 Time-frequency diagram (**a**) and its associated bicoherence (**b**) of the PIV signal measured in a 5×5 cm^2 square centered around the most energetic branch of the attractor, after the focusing reflection

To detect the frequency triplets, we use the bispectrum analysis which measures the extent of statistical dependence among three spectral components (Ω_k, Ω_i, Ω_j) satisfying the relationship $\Omega_k = \Omega_i + \Omega_j$, with the quantity $M(\Omega_i, \Omega_j) = F(\Omega_i)F(\Omega_j)F^*(\Omega_i + \Omega_j)$, where F is the Fourier transform and $*$ denotes the complex conjugate. In practice, the bispectrum is usually normalized and considered in form of bicoherence which is 0 for triplets with random phases and 1 for triplets with perfect phase coupling. The bicoherence is shown in Fig. 8b. In addition to the strong peak $(0.61, 0.61)$ corresponding to the forcing frequency (therefore to self-correlation), the possible triplets satisfying the definition of triadic resonance at $\Omega_k = \Omega_0$ can be found on the line with slope -1 connecting the points $(0, 0.61)$ and $(0.61, 0)$. This emphasizes that the mechanism at play is triadic. Other peaks are also visible corresponding to other choices of Ω_k revealing that the instability mechanism is repeated and leads to a cascade.

Thanks to this beautiful representation, it can therefore be attested that the energy transfer from global to small scales in attractors operates via a hierarchy of triadic interactions producing a complex internal wave field with a rich multi-peak discrete frequency spectrum embedded in a continuous spectrum of weaker magnitude.

The bicoherence demonstration of the cascade is closely related to the bispectrum shown in Fig. 8b. Both axis represent frequencies, and the color on the bispectrum diagram is proportional to the product of the amplitudes. From this picture it can be also be easily seen that the amplitude of the daughter waves tend to lie on the antidiagonals. On very long time intervals the picture shown in Fig. 8b can be changed significantly due to slow evolution of the nonlinear interactions to a completely new regime.

A Route Towards Wave Turbulence

It is important to emphasize that the final stage is non-trivial since these phenomena are beyond the domain of pure wave-wave interactions: it corresponds to a regime usually called wave turbulence [4]. A similar situation takes place for surface waves, where the flourishing literature gives a fully consistent description of energy cascades between components of wave spectra, only in the case of weakly nonlinear processes, while experimental reality deals with cascades significantly "contaminated" by effects of a finite size fluid domain, wave breaking, wave cusps, nonlinear dispersion, viscous damping of wave-field components, etc. The very specific dispersion relation for internal waves introduces additional complications. For instance, in rotating fluids, which have a dispersion relation analogous to stratified fluids, the usefulness of the formalism of wave turbulence as a basis for the studies in rotating turbulence has been reported for experiments only recently [21]. The three dimensional structure of wave attractor and transition to wave turbulence in a rotating annular frustum was recently described in [22]. For internal waves, the question is still fully open, from both experimental and numerical points of view.

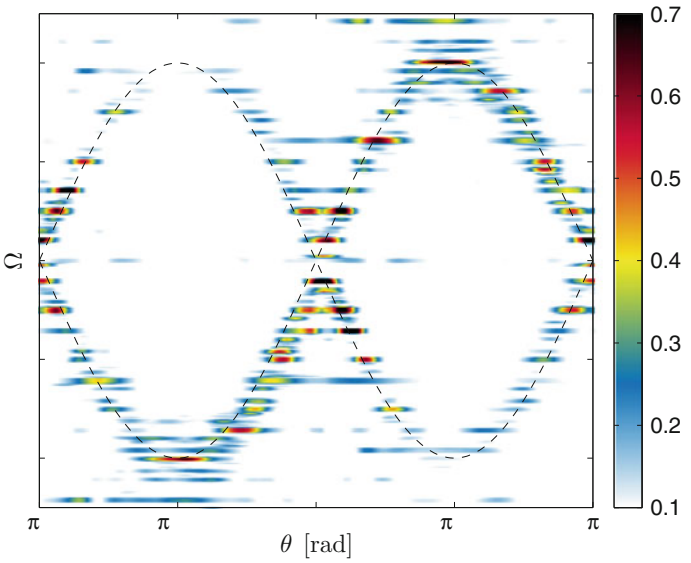


Fig. 9 Energy spectra presented as a function of the non-dimensional frequency Ω and of the slope of the wave beam θ. The dashed lines correspond to the dispersion relation $\Omega = \pm \sin\theta = \pm k_x / \sqrt{k_x^2 + k_z^2}$. Integration across different wavenumbers ranges from 0.22 to 1 rad·cm^{-1}, i.e. wave lengths 28.5–6.3 cm

The presence of wave turbulence-like phenomena is illustrated in Fig. 9 using the energy spectra experimentally obtained for large scales as a diagnostic tool [21]. Horizontal and vertical velocity fields $u(x, z, t)$ and $w(x, z, t)$ are obtained with 2D PIV measurements in the entire trapezoidal domain. A two dimensional Fourier transform for space and a one dimensional Fourier transform for time of these fields leads to $\hat{u}(k_x, k_z, \Omega)$ and $\hat{w}(k_x, k_z, \Omega)$. One can thus define the 2D energy spectrum by

$$E(k_x, k_z, \Omega) = \frac{|\hat{u}(k_x, k_z, \Omega)|^2 + |\hat{w}(k_x, k_z, \Omega)|^2}{2ST}, \tag{8}$$

where S is the area of the PIV measurement and T its duration.

In the dispersion relation for internal waves, $\Omega = \pm \sin\theta$, the wave vector $\mathbf{k}$ and its components do not appear directly but they are linked with the angle θ by $\sin\theta = \pm k_x / \sqrt{k_x^2 + k_z^2}$. To compute the energy spectrum as a function of variable θ, one can interpolate the energy spectrum $E(k_x, k_z, \Omega)$ to get $E(k, \theta, \Omega)$, where k is the norm of the wave vector. Then, one can integrate over the entire range of wave vectors $[k_{min}, k_{max}]$ as follows

$$E(\theta, \Omega) = \int_{k_{min}}^{k_{max}} E(k, \theta, \Omega)\, k \, \mathrm{d}k, \tag{9}$$

or on any range of wave vectors between k_{min} and k_{max}. Note that the accessible range of wave vectors $[k_{\mathrm{min}}, k_{\mathrm{max}}]$ is set by the experimental PIV mesh. The integration in Eq. (9) has been done in Fig. 9 with an integration range which represents 84% of the energy in the entire range $[k_{min}, k_{max}]$. The linear dispersion relation is seen to attract the maxima of the energy spectra. Above results are convincing signatures of a discrete wave turbulence framework for internal waves in this intermediate forcing amplitude regime.

If we repeat the same experiment with a larger amplitude, we have indications that the system is beyond the wave turbulence-like regime and has reached a mixing regime. Indeed, short-scale perturbations in particular clearly escape any relation to linear wave dynamics. This is expected to be due to overturnings, natural precursors to mixing.

Mixing Inferred from Vorticity Distribution

An important issue is whether or not sufficiently energetic internal wave motion can produce an irreversible energy contribution to mixing. Figure 10a presents the comparison between density profiles measured before and after experiments: while no modification of the density (within experimental error) can be observed for the intermediate amplitude forcing that leads to wave turbulence regime described in the previous section, one gets a clear evidence of mixing in case of a larger forcing amplitude.

Further, differences between the regimes corresponding to low and high mixing are clearly seen in statistics of extreme events. This statistics is obtained by the calculation of probability density functions (PDF). Since we are interested in small-scale events destabilizing the stratification, we take the horizontal y-component of vorticity $\xi(x, z, t) = \partial u/\partial z - \partial w/\partial x$ measured in the vertical midplane of the test tank as a relevant quantity and consider the PDF of the dimensionless quantity ξ/N. In Fig. 10b, we present the vorticity PDFs corresponding to different wave regimes in the attractor. In a stable attractor (see dash-dotted curve), extreme events are completely absent and the wave motion is concentrated within the relatively narrow branches of the attractor while the rest of the fluid is quiescent. Accordingly, the PDF has a sharp peak at zero vorticity and is fully localized between well-defined maximum and minimum values of vorticity. For larger forcing amplitudes (dashed and solid curves), the development of TRI increases the probability of extreme events due to summation of primary and secondary wave components.

The occurrence of local overturning events can be viewed as a competition between stratification and vorticity. In a two-dimensional flow, a relevant stability parameter is a version of the Richardson number, which can be introduced as $\mathrm{Ri}_{\xi} = N^2/\xi^2$. For a horizontal stratified shear flow this parameter reduces to the conventional gradient Richardson number $\mathrm{Ri} = N^2/(\mathrm{d}u/\mathrm{d}z)^2$, where $\mathrm{d}u/\mathrm{d}z$ is the velocity shear. Flows with large Ri are generally stable, and the turbulence is suppressed by the stratification. The classic Miles-Howard necessary condition for instability

requires that $\mathrm{Ri} < 1/4$ somewhere in the flow. If this condition is satisfied, the destabilizing effect of shear overcomes the effect of stratification, and some mixing occurs as a result of overturning. The threshold value $|\xi/N| = 2$ is marked on the plot of vorticity PDFs. It can be seen that data corresponding to large forcing amplitudes have "tails" extending into the domains $|\xi/N| > 2$. The area under the tails represents the probability of event of strength $|\xi/N| > 2$. In the larger forcing case (solid curve), this probability is an order of magnitude greater than in intermediate one (dashed curve), in qualitative agreement with the much higher mixing that has been reported.

Conclusion

The paramount importance of a cascade of mechanical energy in the dynamics of the ocean comes from the fact that the ocean is not a classic heat engine: mechanical forcing is needed to drive the meridional heat flux and deep-water renewal. The closure of the meridional circulation critically depends on the cascade of mechanical energy in the abyss and its contribution to mixing.

In the present work, tackling this question with a physicist approach,

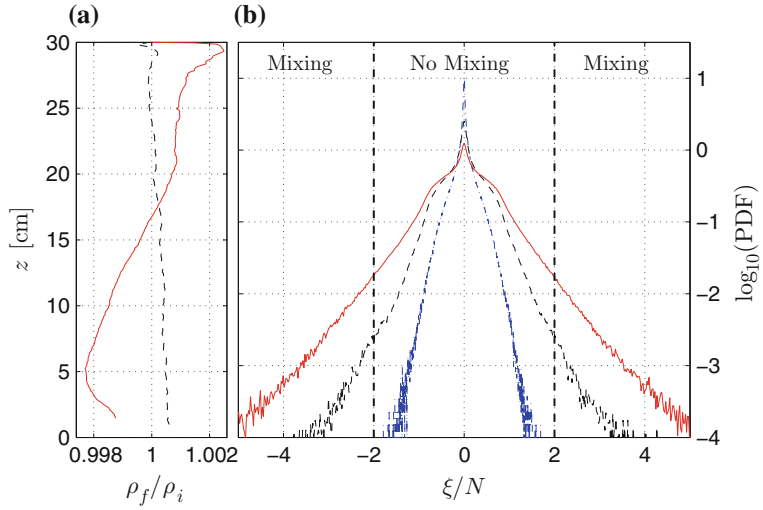


Fig. 10 **a** Ratio between the density profiles measured after and before the experiments for cases with intermediate (dashed) and large (solid) forcing amplitudes. **b** Experimental probability density functions of the vorticity, calculated on the grid from experimental images for low (dotted), intermediate (dashed) and large (solid) forcing amplitudes. Figures 8 and 9 correspond to the intermediate forcing amplitude

- We propose a unique self-consistent experimental (and also numerical) set-up that models a cascade of triadic interactions transferring energy from large-scale monochromatic input to multi-scale internal wave motion.
- We provide explicit evidences of a wave turbulence framework for internal waves, with a clear transition to a cascade of small-scale overturning events.
- We show how beyond the wave turbulence, this original set-up can induce mixing that can be inferred from the calculation of the potential energy or directly by measuring the stratification.

We take advantage of elaborate and recent signal processing tools (Hilbert transforms, time-frequency analysis, bicoherence, ...) to analyze experimental and numerical data. The observed model cascade employs wave attractors, whose significance is realized for stratified and/or rotating fluids (i.e. for a very broad class of celestial bodies) and for magnetized materials, attesting cross-disciplinary importance of the present study for a broad scientific community.

Confinement of the fluid domain and focusing of wave energy at an attractor play an important role in the cascade. However, these conditions are not very restrictive. Under natural conditions internal waves can travel thousands of kilometers which means that quite large bodies of water (for instance, seas) can be considered as confined domains. Also, since attractors can occur in laterally open domains [23], the mechanism of the triadic wave cascade and the bulk mixing described in the present paper is likely to occur in domains with multi-ridge topography as described in [3].

Physical systems supporting wave attractors are strong sources of natural wave turbulence and provide an "internal wave mixing box" that can give useful insights, in the laboratory, on abyssal mixing.

Acknowledgements This work was supported by the LABEX iMUST (ANR-10-LABX-0064) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). This work has achieved thanks to the resources of PSMN from ENS de Lyon. E.V.E. gratefully acknowledges his appointment as a Marie Curie incoming fellow at Laboratoire de physique ENS de Lyon. INS is gratefull for support Russian Foundation for Basic Research 15-01-06363 and Russian Science Foundation 14-50-00095. Direct numerical simulations were performed on supercomputer Lomonosov of Moscow State University. We thank L. Maas, G. Pillet and H. Scolan for helpful discussions and D. Le Tourneau and M. Moulin for technical assistance.

References

1. Morozov, E. G. (1995). Semidiurnial internal wave global field. *Deep-Sea Research*, *I*(42), 135–148.
2. Egbert, G. D., & Ray, R. D. (2000). Significant dissipation of tidal energy in the deep ocean inferred from satellite altimeter data. *Nature*, *495*, 775–778.
3. Polzin, K. L., Toole, J. M., Ledwell, J. R., & Schmitt, R. W. (1997). Spatial variability of turbulent mixing in the Abyssal Ocean. *Science*, *276*(5309), 93–96.
4. Nazarenko, S. V. (2011). *Wave turbulence*. Lecture Notes in Physics, Berlin: Springer.
5. Maas, L. R. M., Benielli, D., Sommeria, J., & Lam, F. P. A. (1997). Observations of an internal wave attractor in a confined stably stratified fluid. *Nature*, *388*, 557–561.

6. Brouzet, C., Sibgatullin, I.N., Ermanyuk, E.V., Joubaud, S., & Dauxois, T. (2017). Scale effects in internal wave attractors. *Physical Review Fluids*, *2*, 114803.
7. Scolan, H., Ermanyuk, E. V., & Dauxois, T. (2013). Nonlinear fate of internal waves attractors. *Physical Review Letters*, *110*, 234501.
8. Bourget, B., Scolan, H., Dauxois, T., Le Bars, M., Odier, P., & Joubaud, S. (2014). Finite-size effects in parametric subharmonic instability. *Journal of Fluid Mechanics*, *759*, 739–750.
9. Dauxois, T., Joubaud, S., Odier, P., & Venaille, A. (2018). Instabilities of internal wave beams. *Annual Review of Fluid Mechanics*, *50*, 131–156.
10. Phillips, O. M. (1966). *The dynamics of the upper ocean*. New York: Cambridge University Press.
11. Dauxois, T., & Young, W. R. (1999). Near-critical refection of internal waves. *Journal of Fluid Mechanics*, *390*, 271–295.
12. Maas, L. R. M., & Lam, F. P. A. (1995). Geometric focusing of internal waves. *Journal of Fluid Mechanics*, *300*, 1–41.
13. Berry, M. V. (1981). Regularity and chaos in classical mechanics, illustrated by three deformations of a circular billiard. *European Journal of Physics*, *2*, 91–102.
14. Fischer, P. F. (1997). An overlapping Schwarz method for spectral element solution of the incompressible Navier Stokes equations. *Journal of Computational Physics*, *133*, 84–101.
15. Fischer, P. F, Mullen, J. S. (2001). Filter-based stabilization of spectral element methods. *Comptes rendus de l'Académie des Sciences Paris Series I–Analyse Number*, *332*, 265–270.
16. Brouzet, C., Sibgatullin, I., Scolan, H., Ermanyuk, E. V., & Dauxois, T. (2016). Internal wave attractors examined using laboratory experiments and 3D numerical simulations. *Journal of Fluid Mechanics*, *793*, 109–131.
17. Brouzet, C. (2016). *Internal wave attractors: from geometrical focusing to non-linear energy cascade and mixing*. Ph.D. dissertation, ENS de Lyon. https://tel.archives-ouvertes.fr/tel-01361201/en.
18. Gostiaux, L., Didelle, H., Mercier, S., & Dauxois, T. (2007). A novel internal waves generator. *Experiments in Fluids*, *42*(1), 123–130.
19. Mercier, M. J., Martinand, D., Mathur, M., Gostiaux, L., Peacock, T., & Dauxois, T. (2010). New wave generation. *Journal of Fluid Mechanics*, *657*, 308–334.
20. Brouzet, C., Ermanyuk, E. V., Joubaud, S., Sibgatullin, I. N., & Dauxois, T. (2016). Energy cascade in internal wave attractors. *Europhysics Letters*, *113*, 44001.
21. Yarom, E., & Sharon, E. (2014). Experimental observation of steady inertial wave turbulence in deep rotating flows. *Nature Physics*, *10*, 510–514.
22. Sibgatullin, E., Ermanyuk, L., Maas, X., Xu & Dauxois, T. (2017). *Directnumerical simulation of three-dimensional inertial wave attractors*. Ivannikov ISPRAS open conference 2017 (ISPRAS) (pp. 137-143). Moscow, Russia.
23. Echeverri, P., Yokossi, T., Balmforth, N. J., & Peacock, T. (2011). Tidally generated internal-wave attractors between double ridges. *Journal of Fluid Mechanics*, *669*, 354–374.

Rogue Waves in the Ocean, the Role of Modulational Instability, and Abrupt Changes of Environmental Conditions that Can Provoke Non Equilibrium Wave Dynamics

Karsten Trulsen

Dedicated to Eugene G. Morozov on his 70th Birthday

Introduction

Rogue or freak waves are waves that are unexpectedly large in comparison with ambient waves. The term "freak wave" was apparently introduced by Draper [16] who discussed ship accidents that could have been provoked by such waves. Mallory [32] discussed serious consequences of extreme waves on ship traffic in the Agulhas current off South Africa. With the advent of offshore oil and gas exploration, well documented observations of freak waves accumulated (e.g. Sand et al. [48]). The "New Year" wave at Draupner in the North Sea 1/1–1995 is one of the best documented observations of a rogue wave [8, 25]. Later the even more extreme Andrea wave at Ekofisk in the North Sea 9/11–2007 was also well documented [7, 15, 31].

This is a good opportunity to give praise to the policy of openness demonstrated by the Norwegian oil company Statoil giving academic researchers unrestricted access to the Draupner "New Year" wave dataset. Few measured time series have been published more often, or inspired more wave research, than the one shown in Fig. 1.

There is currently no consensus on how to define rogue or freak waves. Common criteria are $H/H_s > 2$ or $\eta_c/H_s > 1.25$ where H is the zero-crossing wave height, η_c is the crest height, and H_s is the significant wave height defined as four times the standard deviation of the surface elevation typically calculated from a 20 min time series [17, 24].

K. Trulsen (✉)
Department of Mathematics, University of Oslo, Oslo, Norway
e-mail: karstent@math.uio.no

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_17

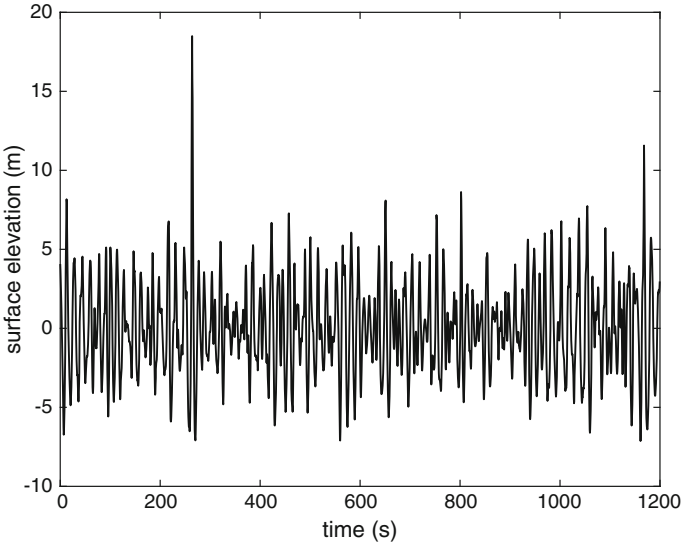


Fig. 1 Twenty minute wave elevation time series measured by a downward pointing radar at 16/11–E in the Norwegian sector of the North Sea on 1/1–1995. First axis is time in seconds, second axis is elevation in meters. Data courtesy of J. I. Dalane and O. T. Gudmestad of Statoil

The existence of rogue waves in the ocean is nowadays well accepted and they are recognized as a threat and a challenge for human activity offshore [6, 17, 27]. Indeed, recently a large wave killed one person and injured several others in the COSL Innovator accident at Troll in the Norwegian Sea 30/12–2015 [61].

Most attention to rogue wave generation has so far been given to random events within Gaussian seas, including linear refraction due to currents or bathymetries, random events within slightly non-Gaussian seas due to static nonlinearities or dynamic nonlinear evolution in equilibrium wave fields, and modulational instabilities which are nonlinear instabilities of perturbations around steady states. While such mechanisms can indeed produce freak waves, and such criteria are nowadays used as warning criteria in operational forecasting, validation performed by e.g. the Norwegian Meteorological Institute has shown that some improvements are still necessary for the warning criteria to be fully satisfactory [6].

The purpose of the present paper is to point out a possible future direction for rogue wave research that might provide some of the desired improvements. There is indeed a different path to rogue wave formation, one that has received little attention so far and is not accounted for in the above practices and criteria, namely the dynamic evolution of non-equilibrium wave fields, where the lack of equilibrium is not due to a small perturbation away from a steady state.

Recent laboratory experiments [47, 58] and numerical computations [22, 60] have shown that non-equilibrium evolution of wave fields can produce surprisingly rough wave conditions. We anticipate that wave fields that are brought out of

equilibrium due to rapidly varying meteorological conditions, or significantly non-uniform environments such as currents or bathymetries [36], or possibly the sudden appearance of a ship in a wave field [34, 35], can be a significant source of rogue waves.

The paper is organized as follows: section "Common Theories for Rogue Waves" gives a summary of common theories on which current rogue wave warning criteria are typically based, section "Case Study on Rogue Waves Through Common Theories: The Prestige Accident" presents a case study for the involvement of rogue waves in the Prestige oil tanker accident based on the common theories, section "Rogue Waves in Non Equilibrium Wave Fields" points out an alternative type of mechanism for rogue wave generation typically not included in todays warning criteria, and "Conclusions" provides a conclusion.

Common Theories for Rogue Waves

Within linear wave theory (LWT) there are several mechanisms that can provoke large waves, e.g. spatio-temporal focusing of waves, refraction over uneven depth and refraction over non-uniform currents. Within LWT, employing the principle of superposition and the Central Limit Theorem from probability theory, it is anticipated that the resulting distribution of surface elevation is Gaussian (e.g. Pierson, [46]).

Rogue waves are known to occur more often than anticipated from LWT, this enhanced occurrence is generally accepted to be due to nonlinearity. There are several known nonlinear mechanisms that can be responsible for this.

Second-order corrections to deep-water waves, static or bound wave nonlinear corrections in general, are known to provoke small deviation from Gaussian statistics. These corrections were derived for uniform waves by Stokes [50], for deep-water irregular gravity waves by Tick [53] and Longuet-Higgins [29] and further by Masuda et al. [33]. Static nonlinear corrections to linear wave theory form the basis for Tayfun-distributions [52].

Starting with the observation that steady uniform waves are unstable to small perturbations, the Benjamin–Feir instability [3, 4], or more general modulational instability (MI) was soon recognized as a mechanism that could initiate the generation of extreme waves. It was soon recognized that the cubic nonlinear Schrödinger (NLS) equation [5, 13, 23, 64] is the simplest nonlinear model that accounts for this instability mechanism. The MI occurs if the ratio between the steepness of the uniform wave and the spectral bandwidth of the perturbation is above a threshold.

Soon after the first well-documented observations of rogue waves in the ocean, e.g. the Draupner wave, it was suggested that the generation of such waves in the ocean could be explained by weakly nonlinear and narrow-banded models, in particular by the nonlinear Schrödinger (NLS) equation. Trulsen and Dysthe [56] argued that it would be an advantage to use a broader-bandwidth modification of the

modified nonlinear Schrödinger (MNLS) equation of Dysthe [18] for better representation of realistic bandwidths [55, 57].

Breather solutions [19, 28, 30, 44, 45] are long-time and long-distance solutions of the NLS equation, starting from an infinitesimal perturbation of a uniform wave train, resulting in extreme waves localized in space and time, occurring once or repeatedly in space and/or in time. Breather solutions have recently been studied in laboratory experiments [9–11].

Related to the discovery of modulational instability of uniform wave trains, it was also recognized that steady homogeneous wave fields are unstable to small inhomogeneous perturbations [1, 2, 12]. This instability occurs if the ratio between the steepness and the spectral bandwidth of the wave field is above a threshold.

The Benjamin–Feir index (BFI), coined by Janssen [26], previously suggested by Onoratoet al. [37, 39] under a different name, is precisely the ratio between the steepness and the spectral bandwidth of the wave field. It has been suggested as an indicator to predict increased probability of rogue waves. The BFI has been shown to be useful as an indicator for extreme waves in unidirectional seas [40, 41] easily reproduced in long and narrow laboratory tanks. However, the BFI has been found to be less useful in directional seas in simulations [21, 38] and in laboratory tests [42, 43] as well as an operational forecasting criterion [6].

A different path from swell and wind-sea interaction to freak wave generation was suggested by Tamura et al. [51] who speculated that the nonlinear coupling between swell and wind-sea could generate a narrow spectrum. Some ship accidents indeed seem to have occurred in conditions of narrowing wave spectra [51, 62, 63]. It has indeed been observed that the directional spread sometimes has been reduced ahead of increased rogue wave occurrences at sea [54].

Recently, researchers have started suspecting that MI is not the correct or the only path to explain real world ocean rogue waves, the waves being in general too broad-banded and short-crested for the BFI to be useful (e.g. Fedele et al. [20]).

Case Study on Rogue Waves Through Common Theories: The Prestige Accident

Freak waves are sometimes the subject of great controversies. There is for example still no consensus on the cause of the Prestige oil tanker initial accident on 13/11–2002, which subsequently led to a major environmental disaster after the sinking of the ship on 19/11–2002. The magnitude of the environmental disaster provoked heated debate in mass media and in court regarding the likelihood that the initial accident could have been caused by a rogue wave (see Trulsen [59]).

In the recent study of Trulsen et al. [59] newly computed hindcast spectra for every hour during the day of the accident were used as input data for four different nonlinear models capable of computing the phase-resolved sea surface, allowing to estimate statistical parameters that characterize the conditions for rogue waves. All

four models coincided that the wave conditions encountered by the tanker Prestige at the moment of the accident were slightly more extreme than those of a Gaussian sea state, and slightly less extreme than those of a Tayfun sea state. This study strongly suggests that the probability of a rogue wave hitting the oil tanker was neither greater than nor smaller than usual.

Rogue Waves in Non Equilibrium Wave Fields

Practical experience with the common theories for rogue waves, e.g. application of the BFI as a warning criterion, suggests that some improvements may still be necessary for the warning criteria to be fully satisfactory [6]. We here point out that there is indeed another path to the formation of rogue waves.

It is often observed that numerical simulation of nonlinear wave evolution, initialized with artificial initial conditions, can need some evolution time or distance before the wave fields become well-behaved. During the initial transient evolution extreme events are often seen to occur. Some effort has been made to suppress this behavior, e.g. by Dommermuth [14], although the "problem" is typically dealt with simply letting the numerical simulations run over sufficient time or distance before the results are used. On the other hand, a very interesting situation would arise if this "initial" strange behavior was the result of a sudden change of physical environment rather than artificial initialization of a numerical integration.

Recently it has been observed that irregular wave fields that propagate from deeper waters into shallower waters can have significant amplification of kurtosis and freak wave statistics some distance inside the transition to the shallower depth. This behavior was first discovered in an experimental dataset from MARIN in The Netherlands by Trulsen et al. [58], subsequently is was studied numerically by Sergeeva et al. [49], Zeng and Trulsen [65], Gramstad et al. [22] and Viotti and Dias [60]. This is a nonlinear effect that is neither explained by MI nor by linear refraction.

Recently Raustøl [47] carried out fine-resolution experiments and measured that the kurtosis could be amplified to a value of 6, occurring at a location approximately one wavelength on the inside of the depth transition to shallower water. She also identified thresholds for water depths when this amplifying behavior took place. It is interesting to note that this extreme amplification of kurtosis took place precisely in a wave field that was not modulationally unstable.

It is common to treat a sea state as being statistically stationary when in fact it varies. Meteorological forecasting services typically give forecasts for every three hours. In Trulsen et al. [59] the Prestige accident was studied with hindcasts every hour, making the assumption that the sea state was constant during each of the one-hour intervals. In the case that the sea state varied dramatically within the one-hour intervals, the nonlinear phase-resolving simulations of Trulsen et al. [59] could be rendered invalid. An insufficient amount of work has been done to identify what happens if the meteorological conditions and sea state change sufficiently fast that the wave field is not in an equilibrium state. Indeed, Tamura et al. [51] suggested that

sea states can vary in time in such a way that the occurrence of rogue waves may be affected.

The presence of a ship in a wave field is known to locally affect the wave field near the ship. If the ship appears suddenly into an already established wave field, it may also represent a perturbation that brings the wave field temporarily out of equilibrium [34, 35].

Indeed, in the recent review of Onorato and Suret [36] it is speculated that a change of ambient conditions can bring a wave field out of equilibrium, thus provoking amplification of kurtosis before the wave field is brought back to equilibrium.

Conclusions

Rogue waves are known to occur more frequently than expected from linear wave theory, more frequently than expected within Gaussian seas. The common nonlinear theories for rogue waves explain such deviation by weakly nonlinear effects on top of equilibrium linear sea states, or as the effect of modulational instability due to unstable perturbations of steady states. The degree of modulational instability of a steady sea state is sometimes assessed by the so-called Benjamin–Feir Index (BFI). There is however a different mechanism for rogue wave generation, viz. nonlinear dynamics of wave fields that are not in an equilibrium state. This mechanism is not indicated by the value of BFI, since the modulational instability is not relevant in the absence of a steady state. We have recently performed experiments at the Department of Mathematics at the University of Oslo revealing that a substantial amplification of kurtosis can occur in non-equilibrium wave fields that are not modulationally unstable.

Acknowledgements This work has been supported by the Research Council of Norway through the project “EXtreme wave WArning criteria for MARine structures” (ExWaMar) RCN 256466.

References

1. Alber, I. E. (1978). The effects of randomness on the stability of two-dimensional surface wavetrains. *Proceedings of the Royal Society of London A*, *363*, 525–546.
2. Alber, I. E, Saffman, P. (1978). Stability of random nonlinear deep water waves with finite bandwidth spectra. *Technical Report*, 31326–6035–RU–TRW Defense and Space System Group
3. Benjamin, T. B. (1967). Instability of periodic wavetrains in nonlinear dispersive systems. *Proceedings of the Royal Society of London A*, *299*, 59–75.
4. Benjamin, T. B., & Feir, J. E. (1967). The disintegration of wave trains on deep water. *Journal of Fluid Mechanics*, *27*, 417–430.
5. Benney, D. J., & Roskes, G. J. (1969). Wave instabilities. *Studies in Applied Mathematics*, *48*, 377–385.
6. Bitner-Gregersen, E. M, Gramstad, O. (2016). Rogue waves—Impact on ships and offshore structures. *Technical Report*, 05–2015, DNV-GL

7. Bitner-Gregersen, E. M., Fernandez, L., Lefèvre, J. M., Monbaliu, J., & Toffoli, A. (2014). The North Sea Andrea storm and numerical simulations. *Natural Hazards and Earth System Science*, *14*, 1407–1415.
8. Cavaleri, L., Barbariol, F., Benetazzo, A., Bertotti, L., Bidlot, J. R., Janssen, P., et al. (2016). The Draupner wave: A fresh look and the emerging view. *Journal of Geophysical Research: Oceans*, *121*, 6061–6075.
9. Chabchoub, A., Hoffmann, N. P., & Akhmediev, N. (2011). Rogue wave observation in a water wave tank. *Physical Review Letters*, *106*, 204,502.
10. Chabchoub, A., Akhmediev, N., & Hoffmann, N. P. (2012). Experimental study of spatiotemporally localized surface gravity water waves. *Physical Review Letters*, *86*, 016,311.
11. Chabchoub, A., Kibler, B., Dudley, J. M., & Akhmediev, N. (2014). Hydrodynamics of periodic breathers. *Philosophical Transactions of the Royal Society of London A*, *372*, 20140,005.
12. Crawford, D. R., Saffman, P. G., & Yuen, H. C. (1980). Evolution of a random inhomogeneous field of nonlinear deep-water gravity waves. *Wave Motion*, *2*, 1–16.
13. Davey, A. (1972). The propagation of a weak nonlinear wave. *Journal of Fluid Mechanics*, *53*, 769–781.
14. Dommermuth, D. (2000). The initialization of nonlinear waves using an adjustment scheme. *Wave Motion*, *32*, 307–317.
15. Donelan, M. A., & Magnusson, A. K. (2017). The making of the Andrea wave and other rogues. *Scientific Reports*, *7*, 44,124.
16. Draper, L. (1964). 'Freak' ocean waves. *Oceanus*, *10*, 13–15.
17. Dysthe, K., Krogstad, H. E., & Müller, P. (2008). Oceanic rogue waves. *Annual Review of Fluid Mechanics*, *40*, 287–310.
18. Dysthe, K. B. (1979). Note on a modification to the nonlinear Schrödinger equation for application to deep water waves. *Proceedings of the Royal Society of London A*, *369*, 105–114.
19. Dysthe, K. B., & Trulsen, K. (1999). Note on breather type solutions of the NLS as models for freak-waves. *Physica Scripta*, *T82*, 48–52.
20. Fedele, F., Brennan, J., Ponce de León, S., Dudley, J., & Dias, F. (2016). Real world ocean rogue waves explained without the modulational instability. *Scientific Reports*, *6*, 27,715.
21. Gramstad, O., & Trulsen, K. (2007). Influence of crest and group length on the occurrence of freak waves. *Journal of Fluid Mechanics*, *582*, 463–472.
22. Gramstad, O., Zeng, H., Trulsen, K., & Pedersen, G. K. (2013). Freak waves in weakly nonlinear unidirectional wave trains over a sloping bottom in shallow water. *Physics of Fluid*, *25*, 122103.
23. Hasimoto, H., & Ono, H. (1972). Nonlinear modulation of gravity waves. *Journal of the Physical Society of Japan*, *33*, 805–811.
24. Haver, S. (2000). Evidences of the existence of freak waves. In: 2000 *Rogue Waves* (pp. 129–140). Ifremer.
25. Haver, S. (2004). A possible freak wave event measured at the Draupner jacket Januar 1 1995. *Rogue Waves*, *2004*, 1–8.
26. Janssen, P. A. E. M. (2003). Nonlinear four-wave interactions and freak waves. *Journal of Physical Oceanography*, *33*, 863–884.
27. Kharif, C., Pelinovsky, E., Slunyaev, A. (2009). *Rogue waves in the ocean*. Springer
28. Kuznetsov, E. A. (1977). Solitons in a parametrically unstable plasma. *Soviet physics, Doklady*, *22*, 507–508.
29. Longuet-Higgins, M. S. (1963). The effect of non-linearities on statistical distributions in the theory of sea waves. *Journal of Fluid Mechanics*, *17*, 459–480.
30. Ma, Y. C. (1979). The perturbed plane-wave solutions of the cubic Schrödinger equation. *Studies in Applied Mathematics*, *60*, 43–58.
31. Magnusson, A. K., & Donelan, M. A. (2013). The Andrea wave characteristics of a measured North Sea rogue wave. *Journal of Offshore Mechanics and Arctic Engineering*, *135*, 031,108.
32. Mallory, J. K. (1974). Abnormal waves on the south east coast of South Africa. *The International Hydrographic Review*, *51*, 99–129.

33. Masuda, A., Kuo, Y. Y., & Mitsuyasu, H. (1979). On the dispersion relation of random gravity waves. Part 1. Theoretical framework. *Journal of Fluid Mechanics*, *92*, 717–730.
34. Molin, B., Remy, F., Kimmoun, O., & Jamois, E. (2005). The role of tertiary wave interactions in wave-body problems. *Journal of Fluid Mechanics*, *528*, 323–354.
35. Molin, B., Kimmoun, O., Remy, F., & Chatjigeorgiou, I. K. (2014). Third-order effects in wave-body interaction. *European Journal of Mechanics-B/Fluids*, *47*, 132–144.
36. Onorato, M., & Suret, P. (2016). Twenty years of progresses in oceanic rogue waves: the role played by weakly nonlinear models. *Natural Hazards*, *84*, 541–548.
37. Onorato, M., Osborne, A. R., Serio, M., & Bertone, S. (2001). Freak waves in random oceanic sea states. *Physical Review Letters*, *86*, 5831–5834.
38. Onorato, M., Osborne, A. R., & Serio, M. (2002a). Extreme wave events in directional, random oceanic sea states. *Physics of Fluids*, *14*, L25–L28.
39. Onorato, M., Osborne, A. R., Serio, M., Resio, D., Pushkarev, A., Zakharov, V. E., et al. (2002b). Freely decaying weak turbulence for sea surface gravity waves. *Physical Review Letters*, *89*(14), 144,501.
40. Onorato, M., Osborne, A. R., Serio, M., Cavaleri, L., Brandini, C., & Stansberg, C. T. (2004). Observation of strongly non-Gaussian statistics for random sea surface gravity waves in wave flume experiments. *Physical Review Letters*, *70*, 1–4.
41. Onorato, M., Osborne, A. R., Serio, M., Cavaleri, L., Brandini, C., & Stansberg, C. T. (2006). Extreme waves, modulational instability and second order theory: wave flume experiments on irregular waves. *European Journal of Mechanics - B/Fluids*, *25*, 586–601.
42. Onorato, M., Cavaleri, L., Fouques, S., Gramstad, O., Janssen, P. A. E. M., Monbaliu, J., et al. (2009a). a) Statistical properties of mechanically generated surface gravity waves: A laboratory experiment in a three dimensional wave basin. *Journal of Fluid Mechanics*, *627*, 235–257.
43. Onorato, M., Waseda, T., Toffoli, A., Cavaleri, L., Gramstad, O., Janssen, P. A. E. M., et al. (2009b). (b) Statistical properties of directional ocean waves: The role of the modulational instability in the formation of extreme events. *Physical Review Letters*, *102*, 114,502.
44. Osborne, A. R., Onorato, M., & Serio, M. (2000). The nonlinear dynamics of rogue waves and holes in deep-water gravity wave trains. *Physics Letters A*, *275*, 386–393.
45. Peregrine, D. H. (1983). Water-waves, non-linear Schrödinger-equations and their solutions. *Journal of the Australian Mathematical Society B*, *25*, 16–43.
46. Pierson, W. J. (1955). Wind generated gravity waves. *Advances in Geophysics*, *2*, 93–178.
47. Raustøl, A. (2014). Freake bølger over variabelt dyp. Master's thesis, University of Oslo
48. Sand, S. E., Ottesen Hansen, N. E., Klinting, P., Gudmestad, O. T., & Sterndorff, M. J. (1990). Freak wave kinematics. In O. T. Gudmestad, A. Tørum (Ed.), *Water Wave Kinematics* (pp. 535–549). Kluwer.
49. Sergeeva, A., Pelinovsky, E., & Talipova, T. (2011). Nonlinear random wave field in shallow water: variable Korteweg-de Vries framework. *Natural Hazards and Earth System Sciences*, *11*, 323–330.
50. Stokes, G. G. (1847). On the theory of oscillatory waves. *Transactions of the Cambridge Philosophical Society*, *8*, 441–455.
51. Tamura, H., Waseda, T., & Miyazawa, Y. (2009). Freakish sea state and swell-windsea coupling: Numerical study of the Suwa-Maru incident. *Geophysical Research Letters*, *36*, L01,607.
52. Tayfun, M. A. (1980). Narrow-band nonlinear sea waves. *Journal of Geophysical Research*, *85*, 1548–1552.
53. Tick, L. J. (1959). A non-linear random model of gravity waves I. *Journal of Mathematics and Mechanics*, *8*, 643–651.
54. Toffoli, A., & Bitner-Gregersen, E. M. (2011). Extreme and rogue waves in directional wave fields. *Open Ocean Engineering Journal*, *4*, 24–33.
55. Trulsen, K., & Dysthe, K. B. (1996). A modified nonlinear Schrödinger equation for broader bandwidth gravity waves on deep water. *Wave Motion*, *24*, 281–289.
56. Trulsen, K., & Dysthe, K. B. (1997). Freak waves—a three-dimensional wave simulation. In: *Proceedings of the 21st Symposium on Naval Hydrodynamics, National Academy Press* (pp 550–560).

57. Trulsen, K., Kliakhandler, I., Dysthe, K. B., & Velarde, M. G. (2000). On weakly nonlinear modulation of waves on deep water. *Physics of Fluids*, *12*, 2432–2437.
58. Trulsen, K., Zeng, H., & Gramstad, O. (2012). Laboratory evidence of freak waves provoked by non-uniform bathymetry. *Physics of Fluids*, *24*, 097,101.
59. Trulsen, K., Nieto Borge, J. C., Gramstad, O., Aouf, L., & Lefèvre, J. M. (2015). Crossing sea state and rogue wave probability during the Prestige accident. *Journal of Geophysical Research: Oceans*, *120*, 7113–7136.
60. Viotti, C., & Dias, F. (2014). Extreme waves induced by strong depth transitions: Fully nonlinear results. *Physics of Fluids*, *26*, 051,705.
61. Viste-Ollestad I, Andersen TL, Oma N, Zachariassen S (2016) Granskingsrapport etter hendelse med fatalt utfall på COSL Innovator 30. desember 2015. Tech. rep., Petroleumstilsynet
62. Waseda, T., Tamura, H., & Kinoshita, T. (2012). Freakish sea index and sea states during ship accidents. *Journal of Marine Science and Technology*, *17*, 305–314.
63. Waseda, T., In, K., Kiyomatsu, K., Tamura, H., Miyazawa, Y., & Iyama, K. (2014). Predicting freakish sea state with an operational third-generation wave model. *Natural Hazards and Earth System Science*, *14*, 945–957.
64. Zakharov, V. E. (1968). Stability of periodic waves of finite amplitude on the surface of a deep fluid. *Journal of Applied Mechanics and Technical Physics*, *9*, 190–194.
65. Zeng, H., & Trulsen, K. (2012). Evolution of skewness and kurtosis of weakly nonlinear unidirectional waves over a sloping bottom. *Natural Hazards and Earth System Science*, *12*, 631–638.

Simulation of Standing and Propagating Sea Waves with Three-Dimensional ARMA Model

Ivan Gankevich and Alexander Degtyarev

Introduction

Studying behaviour of a ship at sea is often based on some model of external excitations—any disturbance that displaces the vessel from equilibrium—major component of which is wind waves. Currently, the most popular sea wave simulation models are based on the linear expansion of a stochastic moving surface as a system of independent random variables. Such models were studied by St. Denis and Pearson [1], Rosenblatt [2], Sveshnikov [3], and Longuet-Higgins [4]. The most popular model is that of Longuet-Higgins (LH), which approximates propagating sea waves as a superposition of elementary harmonic waves with random phases ε_n and random amplitudes c_n:

$$\zeta(x, y, t) = \sum_n c_n \cos(u_n x + v_n y - \omega_n t + \varepsilon_n), \tag{1}$$

where the wave number (u_n, v_n) is continuously distributed on the (u, v) plane, i.e. the unit area $du \times dv$ contains an infinite number of wave numbers. The frequency ω_n associated with wave numbers (u_n, v_n) is given by a dispersion relation

$$\omega_n = \omega(u_n, v_n).$$

The phase ε_n are jointly independent random variables uniformly distributed in the interval $[0, 2\pi]$.

I. Gankevich (✉) · A. Degtyarev
Saint Petersburg State University, Universitetskii prospekt 35, Petergof,
Saint Petersburg 198504, Russia
e-mail: i.gankevich@spbu.ru

A. Degtyarev
e-mail: a.degtyarev@spbu.ru

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_18

Longuet-Higgins showed that under the above conditions, the function $\zeta(x, y, t)$ is a three-dimensional steady-state homogeneous ergodic Gaussian field, defined by

$$2E_\zeta(u, v)dudv = \sum_n c_n^2,$$

where $E_\zeta(u, v)$ is two-dimensional spectral density of wave energy.

Formula (1) is derived from equation of continuity and equation of motion for incompressible inviscid fluid. For ocean waves incompressibility and isotropy of a fluid is assumed; since the motion of ocean waves is due to gravitational forces, irrotational motion of the fluid is assumed which let us introduce the velocity potential ϕ. Under these assumptions the equation of continuity reduces to Laplace equation:

$$\Delta\phi = \frac{\partial^2\phi_x}{\partial x^2} + \frac{\partial^2\phi_y}{\partial y^2} + \frac{\partial^2\phi_z}{\partial z^2} = 0.$$

The Laplace equation is linear and its solution can be found using Fourier transforms. Thus, for plane waves a well-known solution is given in the form of a definite integral [5]:

$$\phi(x, z, t) = \int_0^\infty e^{kz} \left[A(k, t)\cos kx + B(k, t)\sin kx\right] dk.$$

A similar, but slightly more complicated solution is obtained for the three-dimensional case. The constants A and B are determined from the boundary conditions on the surface. In the linear formulation the equation of the wave profile (which is derived from linearised kinematic boundary condition and equation of motion, see section "Determining Wave Pressures for Discretely Given Wavy Surface") is

$$\begin{aligned}\zeta(x, t) &= -\frac{1}{g}\frac{\partial\phi(x, 0, t)}{\partial t} \\ &= \int_0^\infty \left[\frac{\partial A(k, t)}{\partial t}\cos kx + \frac{\partial B(k, t)}{\partial t}\sin kx\right] dk \\ &= \int_0^\infty C_t(k, t)\cos\left(kx + \varepsilon(k, t)\right).\end{aligned} \tag{2}$$

If we set $c_n = C_t(k_n, t)dk$, then wave model (1) may be associated with an approximation of integral (2).

Although, LH model is based on simple linear wave theory and has straightforward computational algorithm, it has some serious shortcomings.

- LH model is designed to represent a stationary Gaussian field. Normal distribution of the simulated process (1) is a consequence of the central limit theorem: its application to the analysis of storm or shallow water waves represents a significant challenge.
- LH model is periodic and need a large set of frequencies to perform long-term simulation.
- In the numerical implementation of the LH model, it appears that convergence rate of (1) is slow. This leads to a skewed simulated wave energy spectrum and skewed cumulative distribution functions of various wave parameters (heights, lengths, etc.). This problem becomes especially significant when simulating complex sea waves that have a wide spectrum with multiple peaks.

The latter point becomes particularly critical in long-term numerical simulation. In a time domain computation of the responses of a vessel in a random seaway, the repeated evaluation of the apparently simple Eq. (1) at hundreds of points on the hull for thousands of time steps becomes a major factor determining the execution speed of the code [6]. So, finding a less computationally intensive method for modelling ocean waves has the potential to increase performance of long-term simulation.

Related Work

Ocean Wave Modelling

Another approach to simulating sea waves involves representing stochastic moving surface as a linear transformation of white noise with memory, which allows to model stationary ergodic Gaussian random process with given correlation characteristics [7]. The first attempts to model two-dimensional disturbances were undertaken in [8], which resulted in the development of the resonance theory of wind waves, and the formal mathematical framework was developed in [9, 10]—the authors built a one-dimensional model of ocean waves based on autoregressive-moving average (ARMA) model.

One-dimensional ARMA model does not have some of the LH model deficiencies: it is both computationally efficient and requires less number of coefficients to converge. In [11] ARMA model is used to generate time series spectrum of which is compatible with Pierson–Moskowitz (PM) approximation of ocean wave spectrum. The authors carry out experiments for one-dimensional AR, MA and ARMA models. They mention excellent agreement between target and initial spectra and higher performance of ARMA model compared to models based on summing large number of harmonic components with random phases. They also mention that in order to reach agreement between target and initial spectrum MA model require lesser number of coefficients than AR model. In [12] the authors generalise ARMA model coefficients determination formulae for multi-variate case.

AR model was successfully applied to predict evolution of propagating wave profiles based on instantaneous wave recordings. In [13] AR model is used to predict swell waves to control wave-energy converters (WEC) in real-time. In order to make WEC more efficient its internal oscillator frequency should match the one of ocean waves. The authors treat wave elevation as time series and compare performance of AR model, neural networks and cyclical models in forecasting time series future values. AR model gives the most accurate prediction of low-frequency swell waves for up to two typical wave periods. It is an example of successful application of AR process to ocean wave modelling.

The feature that distinguishes present work with respect to afore-mentioned ones is the study of three-dimensional (2D in space and 1D in time) ARMA model, which is mostly a different problem.

1. Yule–Walker system of equations, which are used to determine AR coefficients, has complex block-block structure.
2. Optimal model order (in a sense that target spectrum agrees with initial) is determined manually.
3. Instead of PM spectrum, analytic formulae for standing and propagating waves ACF are used as the model input.
4. Three-dimensional wavy surface should be compatible with real ocean surface not only in terms of spectral characteristics, but also in the shape of wave profiles. So, model verification includes distributions of various parameters of generated waves (lengths, heights, periods etc.).

Multi-dimensionality of investigated model not only complexifies the task, but also allows to carry out visual validation of generated wavy surface. It is the opportunity to visualise output of the programme that allowed to ensure that generated surface is compatible with real ocean surface, and is not abstract multi-dimensional stochastic process that is real only statistically.

Pressure Field Determination Formulae

Small Amplitude Waves Theory

In [14–16] the authors propose a solution for inverse problem of hydrodynamics of potential flow within the framework of small-amplitude wave theory (under assumption that wave length is much larger than height: $\lambda \gg h$). In that case inverse problem is linear and reduces to Laplace equation with mixed boundary conditions, and equation of motion is solely used to determine pressures for calculated velocity potential derivatives. The assumption of small amplitudes means the slow decay of wind wave coherence function, i.e. small change of local wave number in time and space compared to the wavy surface elevation (z coordinate). This assumption allows to calculate elevation z derivative as $\zeta_z = k\zeta$, where k is wave number. In two-dimensional case the solution is written explicitly as

$$\left.\frac{\partial\phi}{\partial x}\right|_{x,t} = -\frac{1}{\sqrt{1+\alpha^2}} e^{-I(x)} \int_0^x \frac{\partial\dot{\zeta}/\partial z + \alpha\dot{\alpha}}{\sqrt{1+\alpha^2}} e^{I(x)} dx, \tag{3}$$

$$I(x) = \int_0^x \frac{\partial\alpha/\partial z}{1+\alpha^2} dx,$$

where α is wave slope. In three-dimensional case solution is written in the form of elliptic partial differential equation (PDE):

$$\frac{\partial^2\phi}{\partial x^2}\left(1+\alpha_x^2\right) + \frac{\partial^2\phi}{\partial y^2}\left(1+\alpha_y^2\right) + 2\alpha_x\alpha_y\frac{\partial^2\phi}{\partial x\partial y} +$$
$$\left(\frac{\partial\alpha_x}{\partial z} + \alpha_x\frac{\partial\alpha_x}{\partial x} + \alpha_y\frac{\partial\alpha_x}{\partial y}\right)\frac{\partial\phi}{\partial x} +$$
$$\left(\frac{\partial\alpha_y}{\partial z} + \alpha_x\frac{\partial\alpha_y}{\partial x} + \alpha_y\frac{\partial\alpha_y}{\partial y}\right)\frac{\partial\phi}{\partial y} +$$
$$\frac{\partial\dot{\zeta}}{\partial z} + \alpha_x\dot{\alpha}_x + \alpha_y\dot{\alpha}_y = 0.$$

The authors suggest transforming this equation to finite differences and solve it numerically.

As will be shown in section "Evaluation and Discussion" that (3) diverges when attempted to calculate velocity field for large amplitude waves, and this is the reason that it can not be used together with ARMA model, that generates arbitrary amplitude waves.

Linearisation of Boundary Condition

LH model allows to derive an explicit formula for velocity field by linearising kinematic boundary condition. Velocity potential formula is written as

$$\phi(x,y,z,t) = \sum_n \frac{c_n g}{\omega_n} e^{\sqrt{u_n^2+v_n^2}z} \sin(u_n x + v_n y - \omega_n t + \varepsilon_n).$$

This formula is differentiated to obtain velocity potential derivatives, which are plugged to dynamic boundary condition to obtain pressures.

Three-Dimensional ARMA Process as a Sea Wave Simulation Model

ARMA ocean simulation model defines wavy surface as three-dimensional (two dimensions in space and one in time) autoregressive moving average process: every

surface point is represented as a weighted sum of previous in time and space points plus weighted sum of previous in time and space normally distributed random impulses. The governing equation for 3-D ARMA process is

$$\zeta_{\boldsymbol{i}} = \sum_{\boldsymbol{j}=\boldsymbol{0}}^{\boldsymbol{N}} \boldsymbol{\Phi}_{\boldsymbol{j}} \zeta_{\boldsymbol{i}-\boldsymbol{j}} + \sum_{\boldsymbol{j}=\boldsymbol{0}}^{\boldsymbol{M}} \boldsymbol{\Theta}_{\boldsymbol{j}} \varepsilon_{\boldsymbol{i}-\boldsymbol{j}}, \tag{4}$$

where ζ—wave elevation, $\boldsymbol{\Phi}$—AR process coefficients, $\boldsymbol{\Theta}$—MA process coefficients, ε—white noise with Gaussian distribution, $\boldsymbol{N}$—AR process order, $\boldsymbol{M}$—MA process order, and $\boldsymbol{\Phi}_{\boldsymbol{0}} \equiv 0$, $\boldsymbol{\Theta}_{\boldsymbol{0}} \equiv 0$. Here arrows denote multi-component indices with a component for each dimension. In general, any scalar quantity can be a component (temperature, salinity, concentration of some substance in water etc.). Equation parameters are AR and MA process coefficients and order.

Any ARMA process can be uniquely represented as either MA or AR process of infinite order [10], and the parameters of the spectral representation are defined by the rule of division of power series (in a rational factorized form [9]):

$$S(\omega) = \frac{\Delta\sigma^2}{\pi} \frac{\prod_m (1 - z_m e^{-im\omega\Delta})(1 - z_m e^{im\omega\Delta})}{\prod_n (1 - p_n e^{-in\omega\Delta})(1 - p_n e^{in\omega\Delta})},$$

where z_m and p_n are the zeros of numerator (MA), and denominator (AR), respectively, which form a pair of mutually conjugate numbers. If some of the zeros are located near the unit circle, then the spectral density will have pronounced dips.

Autoregressive (AR) Process

AR process is ARMA process with only one random impulse instead of their weighted sum:

$$\zeta_{\boldsymbol{i}} = \sum_{\boldsymbol{j}=\boldsymbol{0}}^{\boldsymbol{N}} \boldsymbol{\Phi}_{\boldsymbol{j}} \zeta_{\boldsymbol{i}-\boldsymbol{j}} + \varepsilon_{i,j,k}. \tag{5}$$

The coefficients $\boldsymbol{\Phi}$ are calculated from auto-covariate function (ACF) via three-dimensional Yule–Walker (YW) equations, which are obtained after multiplying both parts of the previous equation by $\zeta_{\boldsymbol{i}-\boldsymbol{k}}$ and computing the expected value. Generic form of YW equations is

$$\gamma_{\boldsymbol{k}} = \sum_{\boldsymbol{j}=\boldsymbol{0}}^{\boldsymbol{N}} \boldsymbol{\Phi}_{\boldsymbol{j}}\, \gamma_{\boldsymbol{k}-\boldsymbol{j}} + \sigma_\varepsilon^2 \delta_{\boldsymbol{k}}, \qquad \delta_{\boldsymbol{k}} = \begin{cases} 1, & \text{if } \boldsymbol{k} = 0 \\ 0, & \text{if } \boldsymbol{k} \neq 0, \end{cases} \tag{6}$$

where γ—ACF of process ζ, σ_ε^2—white noise variance. Matrix form of three-dimensional YW equations, which is used in the present work, is

$$\Gamma \begin{bmatrix} \boldsymbol{\Phi_0} \\ \boldsymbol{\Phi}_{0,0,1} \\ \vdots \\ \boldsymbol{\Phi_N} \end{bmatrix} = \begin{bmatrix} \gamma_{0,0,0} - \sigma_\varepsilon^2 \\ \gamma_{0,0,1} \\ \vdots \\ \gamma_{\boldsymbol{N}} \end{bmatrix}, \qquad \Gamma = \begin{bmatrix} \Gamma_0 & \Gamma_1 & \cdots & \Gamma_{N_1} \\ \Gamma_1 & \Gamma_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \Gamma_1 \\ \Gamma_{N_1} & \cdots & \Gamma_1 & \Gamma_0 \end{bmatrix},$$

where $\boldsymbol{N} = \left(p_1, p_2, p_3\right)$ and

$$\Gamma_i = \begin{bmatrix} \Gamma_i^0 & \Gamma_i^1 & \cdots & \Gamma_i^{N_2} \\ \Gamma_i^1 & \Gamma_i^0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \Gamma_i^1 \\ \Gamma_i^{N_2} & \cdots & \Gamma_i^1 & \Gamma_i^0 \end{bmatrix} \qquad \Gamma_i^j = \begin{bmatrix} \gamma_{i,j,0} & \gamma_{i,j,1} & \cdots & \gamma_{i,j,N_3} \\ \gamma_{i,j,1} & \gamma_{i,j,0} & \ddots & x \quad \vdots \\ \vdots & \ddots & \ddots & \gamma_{i,j,1} \\ \gamma_{i,j,N_3} & \cdots & \gamma_{i,j,1} & \gamma_{i,j,0} \end{bmatrix},$$

Since $\boldsymbol{\Phi_0} \equiv 0$, the first row and column of Γ can be eliminated. Matrix Γ is block-toeplitz, positive definite and symmetric, hence the system is efficiently solved by Cholesky decomposition, which is particularly suitable for these types of matrices.

After solving this system of equations white noise variance is estimated from (6) by plugging $\boldsymbol{k} = \boldsymbol{0}$:

$$\sigma_\varepsilon^2 = \sigma_\zeta^2 - \sum_{\boldsymbol{j=0}}^{\boldsymbol{N}} \boldsymbol{\Phi_j}\, \gamma_{\boldsymbol{j}}.$$

Moving Average (MA) Process

MA process is ARMA process with $\boldsymbol{\Phi} \equiv 0$:

$$\zeta_{\boldsymbol{i}} = \sum_{\boldsymbol{j=0}}^{\boldsymbol{M}} \boldsymbol{\Theta_j} \varepsilon_{\boldsymbol{i-j}}. \tag{7}$$

MA coefficients $\boldsymbol{\Theta}$ are defined implicitly via the following non-linear system of equations:

$$\gamma_{\boldsymbol{i}} = \left[\sum_{\boldsymbol{j=i}}^{\boldsymbol{M}} \boldsymbol{\Theta_j \Theta_{j-i}} \right] \sigma_\varepsilon^2.$$

The system is solved numerically by fixed-point iteration method via the following formulae

$$\boldsymbol{\Theta_i} = -\frac{\gamma_{\boldsymbol{0}}}{\sigma_\varepsilon^2} + \sum_{\boldsymbol{j=i}}^{\boldsymbol{M}} \boldsymbol{\Theta_j \Theta_{j-i}}.$$

Here coefficients Θ are calculated from back to front: from $i = M$ to $i = 0$. White noise variance is estimated by

$$\sigma_\varepsilon^2 = \frac{\gamma_0}{1 + \sum_{j=0}^{M} \Theta_j^2}.$$

Authors of [7] suggest using Newton–Raphson method to solve this equation with higher precision, however, this method does not work in three dimensions. Using slower method does not have dramatic effect on the overall programme performance, because the number of coefficients is small and most of the time is spent generating wavy surface.

Mixed Autoregressive Moving Average (ARMA) Process

Generally speaking, ARMA process is obtained by plugging MA generated wavy surface as random impulse to AR process, however, in order to get the process with desired ACF one should re-compute AR coefficients before plugging. There are several approaches to "mix" AR and MA processes.

- The approach proposed in [7] which involves dividing ACF into MA and AR part along each dimension is not applicable here, because in three dimensions such division is not possible: there always be parts of the ACF that are not taken into account by AR and MA process.
- The alternative approach is to use the same (undivided) ACF for both AR and MA processes but use different process order, however, then realisation characteristics (mean, variance etc.) become skewed: these are characteristics of the two overlapped processes.

For the first approach there is a formula to re-compute ACF for AR process, but there is no such formula for the second approach. So, the best solution for now is to simply use AR and MA process exclusively for different types of waves.

Process Selection Criteria for Different Wave Profiles

One problem of ARMA model application to ocean wave generation is that for different types of wave profiles different processes *must* be used: standing waves are modelled by AR process, and propagating waves by MA process. This statement comes from practice: if one tries to use the processes the other way round, the resulting realisation either diverges or does not correspond to real ocean waves. So, the best way to apply ARMA model to ocean wave generation is to use AR process for standing waves and MA process for progressive waves.

The other problem is inability to automatically determine optimal number of coefficients for three-dimensional AR and MA processes. For one-dimensional processes this can be achieved via iterative methods [7], but they diverge in three-dimensional case.

The final problem, which is discussed in section "Mixed Autoregressive Moving Average (ARMA) Process", is inability to "mix" AR and MA process in three dimensions.

In practice some statements made for AR and MA processes in [7] should be flipped for three-dimensional case. For example, the authors say that ACF of MA process cuts at q and ACF of AR process decays to nought infinitely, but in practice making ACF of 3-dimensional MA process not decay results in it being non-invertible and producing realisation that does not look like real ocean waves, whereas doing the same for ACF of AR process results in stationary process and adequate realisation. Also, the authors say that one should allocate the first q points of ACF to MA process (as it often needed to describe the peaks in ACF) and leave the rest points to AR process, but in practice in case of ACF of a propagating wave AR process is stationary only for the first time slice of the ACF, and the rest is left to MA process.

To summarise, the only established scenario of applying ARMA model to ocean wave generation is to use AR process for standing waves and MA process for propagating waves. With a new formulae for 3 dimensions a single mixed ARMA process might increase model precision, which is one of the objectives of the future research.

The Shape of ACF for Different Types of Waves

Analytic Method of Finding the ACF

The straightforward way to find ACF for a given ocean wave profile is to apply Wiener–Khinchin theorem. According to this theorem the autocorrelation K of a function ζ is given by the Fourier transform of the absolute square of the function:

$$K(t) = \mathscr{F}\left\{|\zeta(t)|^2\right\}. \tag{8}$$

When ζ is replaced with actual wave profile, this formula gives you analytic formula for the corresponding ACF.

For three-dimensional wave profile (2D in space and 1D in time) analytic formula is a polynomial of high order and is best obtained via symbolic computation programme. Then for practical usage it can be approximated by superposition of exponentially decaying cosines (which is how ACF of a stationary ARMA process looks like [7]).

Empirical Method of Finding the ACF

However, for three-dimensional case there exists simpler empirical method which does not require sophisticated software to determine shape of the ACF. It is known

that ACF represented by exponentially decaying cosines satisfies first order Stokes' equations for gravity waves [17]. So, if the shape of the wave profile is the only concern in the simulation, then one can simply multiply it by a decaying exponent to get appropriate ACF. This ACF does not reflect other wave profile parameters, such as wave height and period, but opens possibility to simulate waves of a particular non-analytic shape by "drawing" their profile, then multiplying it by an exponent and using the resulting function as ACF. So, this empirical method is imprecise but offers simpler alternative to Wiener–Khinchin theorem approach; it is mainly useful to test ARMA model.

Standing Wave ACF

For three-dimensional plain standing wave the profile is given by

$$\zeta(t,x,y) = A\sin(k_x x + k_y y)\sin(\sigma t). \tag{9}$$

Find ACF via analytic method. Multiplying the formula by a decaying exponent (because Fourier transform is defined for a function f that $f \underset{x\to\pm\infty}{\longrightarrow} 0$) yields

$$\zeta(t,x,y) = A\exp\left[-\alpha(|t| + |x| + |y|)\right]\sin(k_x x + k_y y)\sin(\sigma t). \tag{10}$$

Then, apply 3D Fourier transform to the both sides of the equation via symbolic computation programme, fit the resulting polynomial to the following approximation:

$$K(t,x,y) = \gamma\exp\left[-\alpha(|t| + |x| + |y|)\right]\cos\beta t\cos\left[\beta x + \beta y\right]. \tag{11}$$

So, after applying Wiener–Khinchin theorem we get initial formula but with cosines instead of sines. This difference is important because the value of ACF at $(0,0,0)$ equals to the ARMA process variance, and if one used sines the value would be wrong.

If one tries to replicate the same formula via empirical method, the usual way is to adapt (10) to match (11). This can be done either by changing the phase of the sine, or by substituting sine with cosine to move the maximum of the function to the origin of coordinates.

Propagating Wave ACF

Three-dimensional profile of plain propagating wave is given by

$$\zeta(t,x,y) = A\cos(\sigma t + k_x x + k_y y). \tag{12}$$

For the analytic method repeating steps from the previous two paragraphs yields

$$K(t,x,y) = \gamma\exp\left[-\alpha(|t| + |x| + |y|)\right]\cos\left[\beta(t + x + y)\right]. \tag{13}$$

For the empirical method the wave profile is simply multiplied by a decaying exponent without need to adapt the maximum value of ACF (as it is required for standing wave).

Comparison of Studied Methods

To summarise, the analytic method of finding ocean wave's ACF reduces to the following steps.

- Make wave profile decay when approaching $\pm\infty$ by multiplying it by a decaying exponent.
- Apply Fourier transform to the absolute square of the resulting equation using symbolic computation programme.
- Fit the resulting polynomial to the appropriate ACF approximation.

Two examples in this section showed that in case of standing and propagating waves their decaying profiles resemble the corresponding ACFs with the exception that the ACF's maximum should be moved to the origin to preserve simulated process variance. Empirical method of finding ACF reduces to the following steps.

- Make wave profile decay when approaching $\pm\infty$ by multiplying it by a decaying exponent.
- Move maximum value of the resulting function to the origin by using trigonometric identities to shift the phase.

Evaluation and Discussion

In [18–20] for AR model the following items were verified experimentally:

- probability distributions of different wave characteristics (wave heights, lengths, crests, periods, slopes, three-dimensionality),
- dispersion relation,
- retention of integral characteristics for mixed wave sea state.

In this work we repeat probability distribution tests for three-dimensional AR and MA model.

In [9] the authors show that several ocean wave characteristics (listed in Table 1) have Weibull distribution, and wavy surface elevation has Gaussian distribution. In order to verify that distributions corresponding to generated realisation are correct, quantile-quantile plots are used (plots where analytic quantile values are used for *OX* axis and estimated quantile values for *OY* axis). If the estimated distribution matches analytic then the graph has the form of the straight line. Tails of the graph may diverge from the straight line, because they can not be reliably estimated from the finite-size realisation. Different methods of extracting waves from realisation produce variations in quantile function tails, it is probably impractical to extract every possible wave from realisation since they may (and often) overlap.

Table 1 Values of Weibull shape parameter for different wave characteristics

Characteristic	Weibull shape (k)
Wave height	2
Wave length	2.3
Crest length	2.3
Wave period	3
Wave slope	2.5
Three-dimensionality	2.5

Verification was performed for standing and propagating waves. The corresponding ACFs and quantile-quantile plots of wave characteristics distributions are shown in Figs. 1, 2 and 3.

Graph tails in Fig. 1 deviate from original distribution for individual wave characteristics, because every wave have to be extracted from the resulting wavy surface to measure its length, period and height. There is no algorithm that guarantees correct extraction of all waves, because they may overlap each other. Weibull distribution right tail represents infrequently occurring waves, so it deviates more than left tail.

Degree of correspondence for standing waves (Fig. 2) is lower for height and length, is roughly the same for surface elevation and is higher for wave period distribution tails. Lower correspondence degree for length and height may be attributed to the fact that Weibull distributions were obtained empirically for ocean waves which are typically propagating, and distributions may be different for standings waves. Higher correspondence degree for wave periods is attributed to the fact that wave

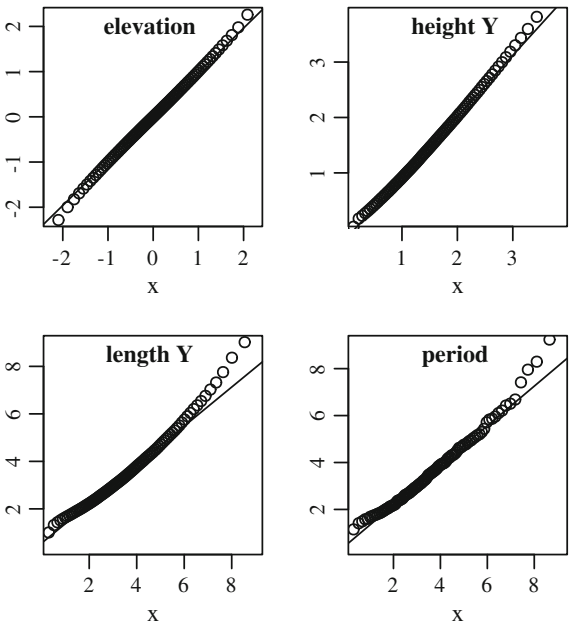


Fig. 1 Quantile-quantile plots for propagating waves

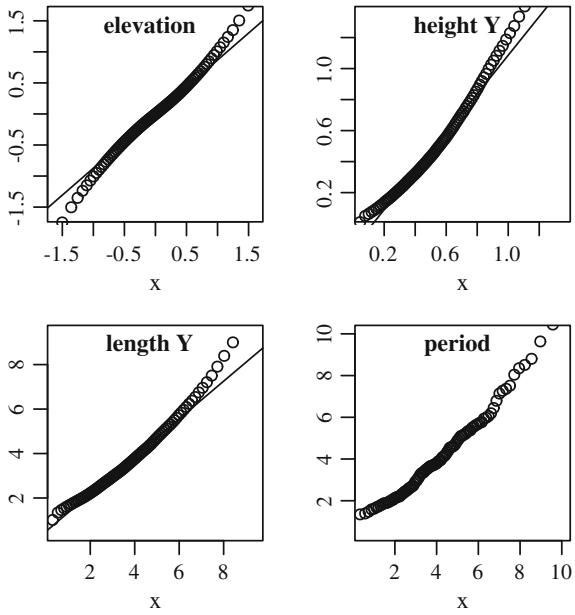


Fig. 2 Quantile-quantile plots for standing waves

periods of standing waves are extracted more precisely as the waves do not move outside simulated wavy surface region. The same correspondence degree for wave elevation is obtained, because this is the characteristic of the wavy surface (and corresponding AR or MA process) and is not affected by the type of waves.

ARMA model, owing to its non-physical nature, does not have the notion of ocean wave; it simulates wavy surface as a whole instead. Motions of individual waves and their shape are often rough, and the total number of waves can not be determined precisely. However, integral characteristics of wavy surface match the ones of real ocean waves.

Theoretically, ocean waves themselves can be chosen as ACFs, the only preprocessing step is to make them decay exponentially. This may allow to generate waves of arbitrary profiles, and is one of the directions of future work.

Determining Wave Pressures for Discretely Given Wavy Surface

Analytic solutions to boundary problems in classical equations are often used to study different properties of the solution, and for that purpose general solution formula is too difficult to study, as it contains integrals of unknown functions. Fourier method is one of the methods to find analytic solutions to a PDE. It is based on application of Fourier transform to each part of PDE, which reduces the equation to

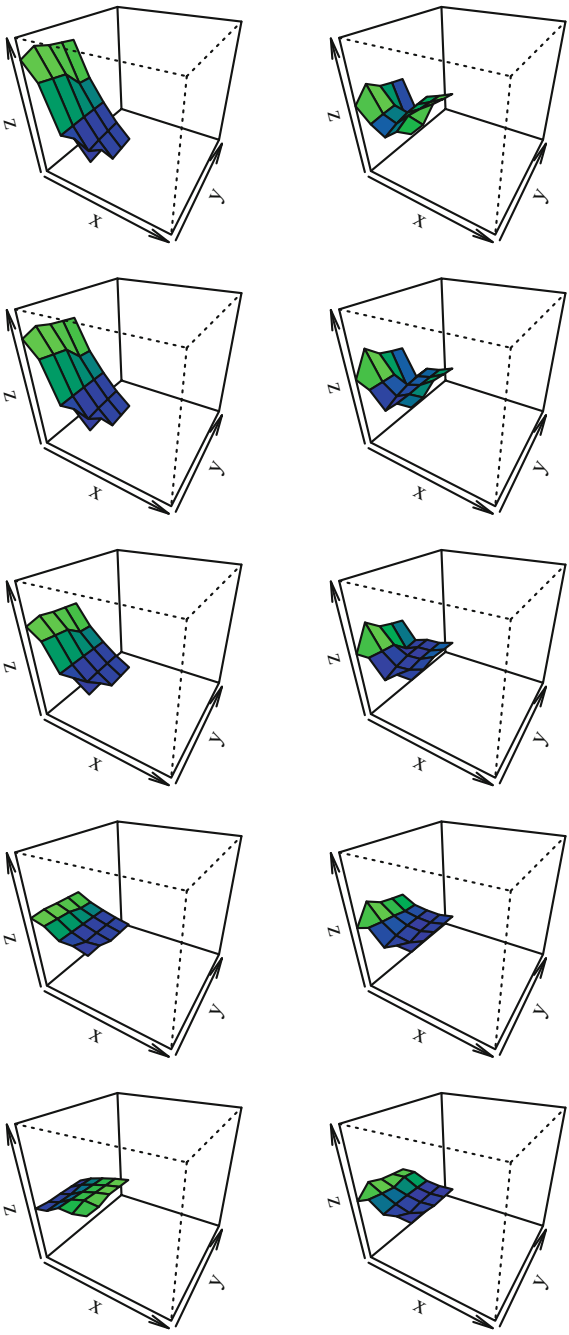


Fig. 3 Time slices of ACF for standing (left column) and propagating waves (right column)

algebraic, and the solution is written as inverse Fourier transform of some function (which may contain Fourier transforms of other functions). Since, it is not possible to write analytic forms of these Fourier transforms in all cases, unique solutions are found and their behaviour is studied in different domains instead. At the same time, computing discrete Fourier transforms on the computer is possible for any discretely defined function and efficient when using FFT algorithms. These algorithms use symmetry of complex exponentials to decrease asymptotic complexity from $\mathcal{O}(n^2)$ to $\mathcal{O}(n \log_2 n)$. So, even if general solution contains Fourier transforms of unknown functions, they still can be computed numerically, and FFT family of algorithms makes this approach efficient.

Alternative approach to solve a PDE is to reduce it to difference equations, which are solved by constructing various numerical schemes. This approach leads to approximate solution, and asymptotic complexity of corresponding algorithms is comparable to that of FFT. For example, stationary elliptic PDE transforms to implicit numerical scheme which is solved by iterative method on each step of which a tridiagonal or five-diagonal system of algebraic equations is solved via Thomas algorithm. Asymptotic complexity of this approach is $\mathcal{O}(nm)$, where n—number of wavy surface grid points, m—number of iterations. Despite their wide spread, iterative algorithms are inefficient on parallel computer architectures; in particular, their mapping to co-processors may involve copying data in and out of the co-processor in each iteration, which negatively affects their performance. At the same time, high number of Fourier transforms in the solution is an advantage, rather than a disadvantage. First, solutions obtained by Fourier method are explicit, hence their implementations scales with the large number of parallel computer cores. Second, there are implementations of FFT optimised for different processor architectures as well as co-processors (GPU, MIC) which makes it easy to get high performance on any computing platform. These advantages substantiate the choice of Fourier method to obtain explicit analytic solution to the problem of determining pressures under wavy ocean surface.

The problem of finding pressure field under wavy sea surface represents inverse problem of hydrodynamics for incompressible inviscid fluid. System of equations for it in general case is written as [5]

$$\begin{aligned} &\nabla^2 \phi = 0, \\ &\phi_t + \frac{1}{2}|\boldsymbol{v}|^2 + g\zeta = -\frac{p}{\rho}, && \text{at } z = \zeta(x, y, t), \\ &D\zeta = \nabla\phi \cdot \boldsymbol{n}, && \text{at } z = \zeta(x, y, t), \end{aligned} \tag{14}$$

where ϕ—velocity potential, ζ—elevation (z coordinate) of wavy surface, p—wave pressure, ρ—fluid density, $\boldsymbol{v} = (\phi_x, \phi_y, \phi_z)$—velocity vector, g—acceleration of gravity, and D—substantial (Lagrange) derivative. The first equation is called continuity (Laplace) equation, the second one is the conservation of momentum law (the so called dynamic boundary condition); the third one is kinematic boundary condition for free wavy surface, which states that rate of change of wavy surface elevation ($D\zeta$)

equals to the change of velocity potential derivative along the wavy surface normal ($\nabla\phi \cdot \boldsymbol{n}$).

Inverse problem of hydrodynamics consists in solving this system of equations for ϕ. In this formulation dynamic boundary condition becomes an explicit formula to determine pressure field using velocity potential derivatives obtained from the remaining equations. So, from mathematical point of view inverse problem of hydrodynamics reduces to Laplace equation with mixed boundary condition—Robin problem.

Two-Dimensional Case

Formula for Infinite Depth Fluid

Two-dimensional Laplace equation with Robin boundary condition is written as

$$\begin{aligned} &\phi_{xx} + \phi_{zz} = 0, \\ &\zeta_t + \zeta_x \phi_x = \frac{\zeta_x}{\sqrt{1+\zeta_x^2}}\phi_x - \frac{1}{\sqrt{1+\zeta_x^2}}\phi_z, \qquad \text{at } z = \zeta(x,t). \end{aligned} \tag{15}$$

Use Fourier method to solve this problem. Applying Fourier transform to both sides of the equation yields

$$-4\pi^2\left(u^2+v^2\right)\mathscr{F}_{u,v}\{\phi(x,z)\} = 0,$$

hence $v = \pm iu$. Hereinafter we use the following symmetric form of Fourier transform:

$$\mathscr{F}_{u,v}\{f(x,y)\} = \iint_{-\infty}^{\infty} f(x,y)e^{-2\pi i(xu+yv)}dxdy.$$

We seek solution in the form of inverse Fourier transform $\phi(x,z) = \mathscr{F}^{-1}_{x,z}\{E(u,v)\}$. Plugging[1] $v = iu$ into the formula yields

$$\phi(x,z) = \mathscr{F}^{-1}_{x}\left\{e^{2\pi uz}E(u)\right\}. \tag{16}$$

In order to make substitution $z = \zeta(x,t)$ not interfere with Fourier transforms, we rewrite (16) as a convolution:

$$\phi(x,z) = \mathscr{D}_1(x,z) * \mathscr{F}^{-1}_{x}\{E(u)\},$$

where $\mathscr{D}_1(x,z)$—a function, form of which is defined in section "Velocity Potential Computation" and which satisfies equation $\mathscr{F}_u\left\{\mathscr{D}_1(x,z)\right\} = e^{2\pi uz}$. Plugging formula

[1] $v = -iu$ is not applicable because velocity potential must go to nought when depth goes to infinity.

ϕ into the boundary condition yields

$$\zeta_t = (if(x) - 1)\left[\mathscr{D}_1(x,z) * \mathscr{F}_x^{-1}\{2\pi u E(u)\}\right],$$

where $f(x) = \zeta_x/\sqrt{1+\zeta_x^2} - \zeta_x$. Applying Fourier transform to both sides of this equation yields formula for coefficients E:

$$E(u) = \frac{1}{2\pi u}\frac{\mathscr{F}_u\left\{\zeta_t/\left(if(x) - 1/\sqrt{1+\zeta_x^2}\right)\right\}}{\mathscr{F}_u\left\{\mathscr{D}_1(x,z)\right\}}$$

Finally, substituting z for $\zeta(x,t)$ and plugging resulting equation into (16) yields formula for $\phi(x,z)$:

$$\boxed{\phi(x,z) = \mathscr{F}_x^{-1}\left\{\frac{e^{2\pi u z}}{2\pi u}\frac{\mathscr{F}_u\left\{\zeta_t/\left(if(x) - 1/\sqrt{1+\zeta_x^2}\right)\right\}}{\mathscr{F}_u\left\{\mathscr{D}_1(x,\zeta(x,t))\right\}}\right\}.} \tag{17}$$

Multiplier $e^{2\pi u z}/(2\pi u)$ makes a graph of a function to which Fourier transform is applied asymmetric with respect to OY axis. This makes it difficult to apply FFT which expects periodic function with nought on both ends of the interval. Using numerical integration instead of FFT is not faster than solving the initial system of equations with numerical schemes. This problem is alleviated by using formula (19) for finite depth fluid with wittingly large depth h. This formula is derived in the following section.

Formula for Finite Depth Fluid

On the sea bottom vertical fluid velocity component equals nought: $\phi_z = 0$ on $z = -h$, where h—water depth. In this case equation $v = -iu$, which came from Laplace equation, can not be neglected, hence the solution is sought in the following form:

$$\phi(x,z) = \mathscr{F}_x^{-1}\left\{\left(C_1 e^{2\pi u z} + C_2 e^{-2\pi u z}\right)E(u)\right\}. \tag{18}$$

Plugging ϕ into the boundary condition on the sea bottom yields

$$C_1 e^{-2\pi u h} - C_2 e^{2\pi u h} = 0,$$

hence $C_1 = \frac{1}{2}Ce^{2\pi u h}$ and $C_2 = -\frac{1}{2}Ce^{-2\pi u h}$. Constant C may take arbitrary value here, because after plugging it becomes part of unknown coefficients $E(u)$. Plugging formulae for C_1 and C_2 into (18) yields

$$\phi(x,z) = \mathscr{F}_x^{-1}\{\cosh(2\pi u(z+h))E(u)\}.$$

Plugging ϕ into the boundary condition on the free surface yields

$$\zeta_t = f(x)\mathscr{F}_x^{-1}\{2\pi i u \cosh(2\pi u(z+h))\, E(u)\} - \frac{1}{\sqrt{1+\zeta_x^2}}\mathscr{F}_x^{-1}\{2\pi u \sinh(2\pi u(z+h))\, E(u)\}.$$

Here sinh and cosh give similar results near free surface, and since this is the main area of interest in practical applications, we assume that $\cosh(2\pi u(z+h)) \approx \sinh(2\pi u(z+h))$. Performing analogous to the previous section transformations yields final formula for $\phi(x,z)$:

$$\phi(x,z,t) = \mathscr{F}_x^{-1}\left\{\frac{\cosh(2\pi u(z+h))}{2\pi u}\frac{\mathscr{F}_u\left\{\zeta_t / \left(if(x) - 1/\sqrt{1+\zeta_x^2}\right)\right\}}{\mathscr{F}_u\left\{\mathscr{D}_2(x,\zeta(x,t))\right\}}\right\}, \tag{19}$$

where $\mathscr{D}_2(x,z)$—a function, form of which is defined in section "Velocity Potential Computation" and which satisfies equation $\mathscr{F}_u\left\{\mathscr{D}_2(x,z)\right\} = \cosh(2\pi u z)$.

Reducing to the Formulae from Linear Wave Theory

Check the validity of derived formulae by substituting $\zeta(x,t)$ with known analytic formula for plain waves. Symbolic computation of Fourier transforms in this section were performed in Mathematica [21]. In the framework of linear wave theory assume that waves have small amplitude compared to their lengths, which allows us to simplify initial system of Eq. (15) to

$$\begin{aligned} &\phi_{xx} + \phi_{zz} = 0, \\ &\zeta_t = -\phi_z \qquad\qquad \text{at } z = \zeta(x,t), \end{aligned}$$

solution to which is written as

$$\phi(x,z,t) = -\mathscr{F}_x^{-1}\left\{\frac{e^{2\pi u z}}{2\pi u}\mathscr{F}_u\{\zeta_t\}\right\}.$$

Propagating wave profile is defined as $\zeta(x,t) = A\cos(2\pi(kx-t))$. Plugging this formula into (17) yields $\phi(x,z,t) = -\frac{A}{k}\sin(2\pi(kx-t))\cosh(2\pi kz)$. In order to reduce it to the formula from linear wave theory, rewrite hyperbolic sine in exponential form, discard the term containing $e^{-2\pi kz}$ as contradicting condition $\phi \underset{z\to-\infty}{\longrightarrow} 0$. Taking real part of the resulting formula yields $\phi(x,z,t) = \frac{A}{k}e^{2\pi kz}\sin(2\pi(kx-t))$, which corresponds to the known formula from linear wave theory. Similarly, under small-amplitude waves assumption the formula for finite depth fluid (19) is reduced to

$$\phi(x,z,t) = -\mathscr{F}_x^{-1}\left\{\frac{\cosh(2\pi u(z+h))}{2\pi u\cosh(2\pi uh)}\mathscr{F}_u\{\zeta_t\}\right\}.$$

Substituting $\zeta(x,t)$ with propagating plain wave profile formula yields

$$\phi(x,z,t) = \frac{A}{k}\frac{\cosh(2\pi k(z+h))}{\cosh(2\pi kh)}\sin(2\pi(kx-t)), \tag{20}$$

which corresponds to the formula from linear wave theory for finite depth fluid.

Different forms of Laplace equation solutions, in which decaying exponent is written with either "+" or "−" signs, may cause incompatibilities between formulae from linear wave theory and formulae derived in this work, where sinh is used instead of cosh. Equality $\frac{\cosh(2\pi k(z+h))}{\cosh(2\pi kh)} \approx \frac{\sinh(2\pi k(z+h))}{\sinh(2\pi kh)}$ becomes strict on the free surface, and difference between left-hand and right-hand sides increases when approaching sea bottom (for sufficiently large depth difference near free surface is negligible). So, for sufficiently large depth any function (cosh or sinh) may be used for velocity potential computation near free surface.

Reducing (17) and (19) to the known formulae from linear wave theory shows, that formula for infinite depth (17) is not suitable to compute velocity potentials with Fourier method, because it does not have symmetry, which is required for Fourier transform. However, formula for finite depth can be used instead by setting h to some characteristic water depth. For standing wave reducing to linear wave theory formulae is made under the same assumptions.

Three-Dimensional Case

Three-dimensional version of (14) is written as

$$\begin{aligned} &\phi_{xx} + \phi_{yy} + \phi_{zz} = 0, \\ &\zeta_t + \zeta_x\phi_x + \zeta_y\phi_y = \frac{\zeta_x}{\sqrt{1+\zeta_x^2+\zeta_y^2}}\phi_x + \frac{\zeta_y}{\sqrt{1+\zeta_x^2+\zeta_y^2}}\phi_y - \frac{1}{\sqrt{1+\zeta_x^2+\zeta_y^2}}\phi_z, \quad \text{at } z = \zeta(x,y,t). \end{aligned} \tag{21}$$

Again, use Fourier method to solve it. Applying Fourier transform to both sides of Laplace equation yields

$$-4\pi^2\left(u^2+v^2+w^2\right)\mathscr{F}_{u,v,w}\{\phi(x,y,z)\} = 0,$$

hence $w = \pm i\sqrt{u^2+v^2}$. We seek solution in the form of inverse Fourier transform $\phi(x,y,z) = \mathscr{F}^{-1}_{x,y,z}\{E(u,v,w)\}$. Plugging $w = i\sqrt{u^2+v^2} = i|\boldsymbol{k}|$ into the formula yields

$$\phi(x,y,z) = \mathscr{F}^{-1}_{x,y}\left\{\left(C_1 e^{2\pi|\boldsymbol{k}|z} - C_2 e^{-2\pi|\boldsymbol{k}|z}\right)E(u,v)\right\}.$$

Plugging ϕ into the boundary condition on the sea bottom (analogous to two-dimensional case) yields

$$\phi(x,y,z) = \mathscr{F}^{-1}_{x,y}\{\cosh(2\pi|\boldsymbol{k}|(z+h))\,E(u,v)\}. \tag{22}$$

Plugging ϕ into the boundary condition on the free surface yields

$$\begin{aligned}\zeta_t = {} & if_1(x,y)\mathscr{F}^{-1}_{x,y}\{2\pi u \cosh(2\pi|\boldsymbol{k}|(z+h))\,E(u,v)\} \\ & + if_2(x,y)\mathscr{F}^{-1}_{x,y}\{2\pi v \cosh(2\pi|\boldsymbol{k}|(z+h))\,E(u,v)\} \\ & - \frac{1}{\sqrt{1+\zeta_x^2+\zeta_y^2}}\mathscr{F}^{-1}_{x,y}\{2\pi|\boldsymbol{k}|\sinh(2\pi|\boldsymbol{k}|(z+h))\,E(u,v)\}\end{aligned}$$

where $f_1(x,y) = \zeta_x/\sqrt{1+\zeta_x^2+\zeta_y^2} - \zeta_x$ and $f_2(x,y) = \zeta_y/\sqrt{1+\zeta_x^2+\zeta_y^2} - \zeta_y$.

Like in section "Two-Dimensional Case" we assume that $\cosh(2\pi u(z+h)) \approx \sinh(2\pi u(z+h))$ near free surface, but in three-dimensional case this is not enough to solve the problem. In order to get analytic formula for coefficients E we need to assume, that all Fourier transforms in the equation have radially symmetric kernels, i.e. replace u and v with $|\boldsymbol{k}|$. There are two points supporting this assumption. First, in numerical implementation integration is done over positive wave numbers, so the sign of u and v does not affect the solution. Second, the rate growth of cosh term of the integral kernel is much higher than the one of u or $|\boldsymbol{k}|$, so the substitution has small effect on the magnitude of the solution. Despite these two points, a use of more mathematically rigorous approach would be preferable.

Making the replacement, applying Fourier transform to both sides of the equation and plugging the result into (22) yields formula for ϕ:

$$\begin{aligned}&\phi(x,y,z,t) \\ &= \mathscr{F}^{-1}_{x,y}\left\{\frac{\cosh(2\pi|\boldsymbol{k}|(z+h))}{2\pi|\boldsymbol{k}|}\frac{\mathscr{F}_{u,v}\left\{\zeta_t/\left(if_1(x,y)+if_2(x,y)-1/\sqrt{1+\zeta_x^2+\zeta_y^2}\right)\right\}}{\mathscr{F}_{u,v}\left\{\mathscr{D}_3(x,y,\zeta(x,y))\right\}}\right\},\end{aligned} \tag{23}$$

where $\mathscr{F}_{u,v}\left\{\mathscr{D}_3(x,y,z)\right\} = \cosh(2\pi|\boldsymbol{k}|z)$.

Evaluation and Discussion

Comparing obtained generic formulae (17) and (19) to the known formulae from linear wave theory allows to see the difference between velocity fields for both large and small amplitude waves. In general analytic formula for velocity potential in not known, even for plain waves, so comparison is done numerically. Taking into account conclusions of section "Two-Dimensional Case", only finite depth formulae are compared.

The Difference with Linear Wave Theory Formulae

In order to obtain velocity potential fields, ocean wavy surface was generated by AR model with varying wave amplitude. In numerical implementation wave num-

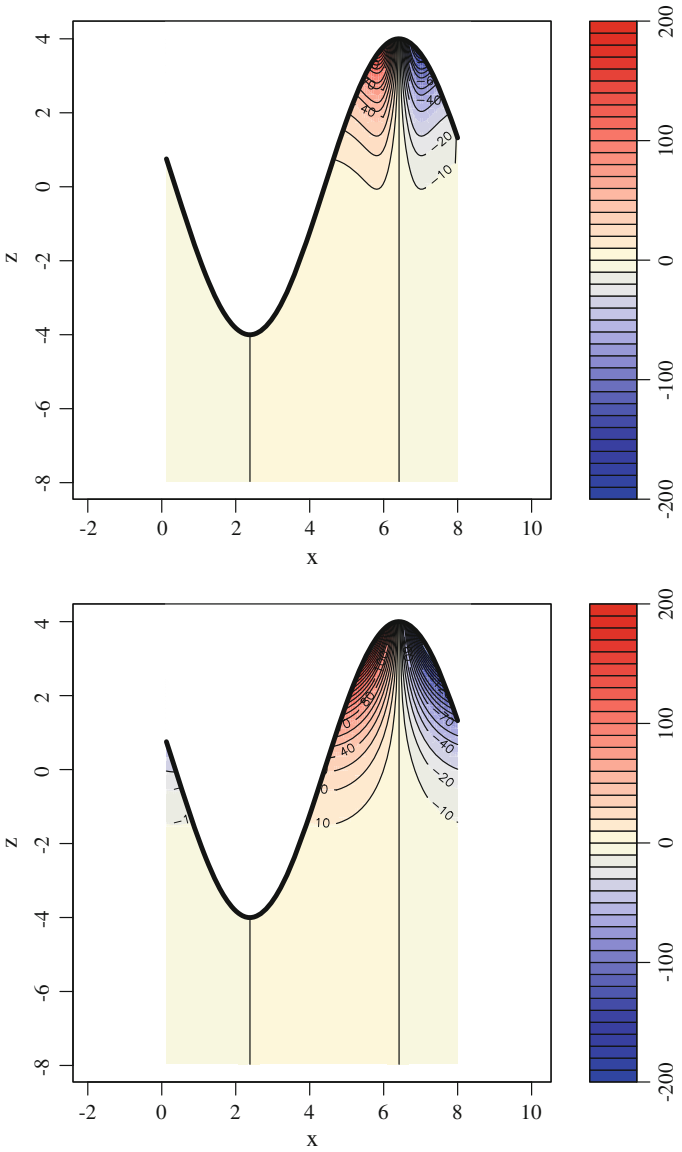


Fig. 4 Velocity potential field of propagating wave $\zeta(x, y, t) = \cos(2\pi x - t/2)$. Field produced by formula (19) (top) and linear wave theory formula (bottom)

bers in Fourier transforms were chosen on the interval from 0 to the maximal wave number determined numerically from the obtained wavy surface. Experiments were conducted for waves of both small and large amplitudes.

The experiment showed that velocity potential fields produced by formula (19) for finite depth fluid and formula (20) from linear wave theory are qualitatively different (Fig. 4). First, velocity potential contours have sinusoidal shape, which is different from oval shape described by linear wave theory. Second, velocity potential decays more rapidly than in linear wave theory as getting closer to the bottom, and the region where the majority of wave energy is concentrated is closer to the wave crest. Similar numerical experiment, in which all terms of (19) that are neglected in the framework of linear wave theory are eliminated, shows no difference (as much as machine precision allows) in resulting velocity potential fields.

The Difference with Small-Amplitude Wave Theory

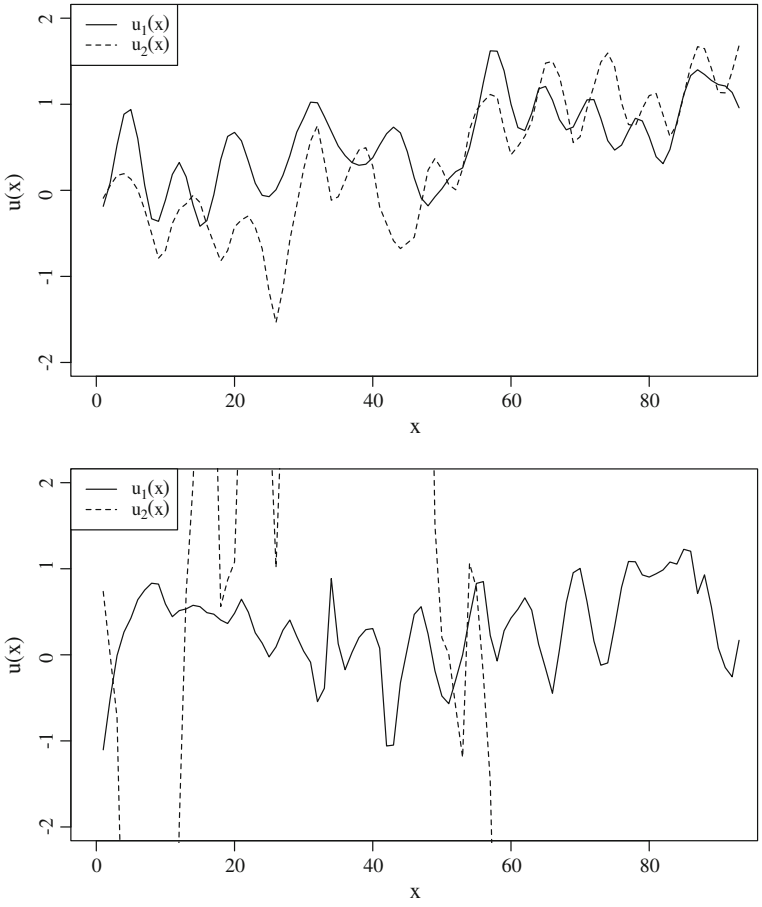


Fig. 5 Comparison of velocity field on the ocean wavy surface obtained by generic formula (u_1) and formula for small-amplitude waves (u_2). Velocity field for realisations containing small-amplitude (top) and large-amplitude (bottom) waves

The experiment, in which velocity fields produced numerically by different formulae were compared, shows that velocity fields produced by formulae (19) and (3) correspond to each other for small-amplitude waves. Two ocean wavy surface realisations were made by AR model: one containing small-amplitude waves, other containing large-amplitude waves. Integration in formula (19) was done over wave numbers range extracted from the generated wavy surface. For small-amplitude waves both formulae showed comparable results (the difference in the velocity is attributed to the stochastic nature of AR model), whereas for large-amplitude waves stable velocity field was produced only by formula (19) (Fig. 5). So, generic formula (19) gives satisfactory results without restriction on wave amplitudes.

High-Performance Software Implementation for Heterogeneous Platforms

White Noise Generation

In order to eliminate periodicity from generated wavy surface, it is imperative to use PRNG with sufficiently large period to generate white noise. Parallel Mersenne Twister [22] with a period of $2^{19937} - 1$ is used as a generator in this work. It allows to produce aperiodic ocean wavy surface realisations in any practical usage scenarios.

There is no guarantee that multiple Mersenne Twisters executed in parallel threads with distinct initial states produce uncorrelated pseudo-random number sequences, however, algorithm of dynamic creation of Mersenne Twisters [23] may be used to provide such a guarantee. The essence of the algorithm is to find matrices of initial generator states, that give maximally uncorrelated pseudo-random number sequences when Mersenne Twisters are executed in parallel with these initial states. Since finding such initial states consumes considerable amount of processor time, vector of initial states is created preliminary with knowingly larger number of parallel threads and saved to a file, which is then read before starting white noise generation.

Wavy Surface Generation

In ARMA model value of wavy surface elevation at a particular point depends on previous in space and time points, as a result the so called *ramp-up interval* (see Fig. 6), in which realisation does not correspond to specified ACF, forms at the beginning of the realisation. There are several solutions to this problem which depend on the simulation context.

If the realisation is used in the context of ship stability simulation without manoeuvring, ramp-up interval will not affect results of the simulation, because it is located

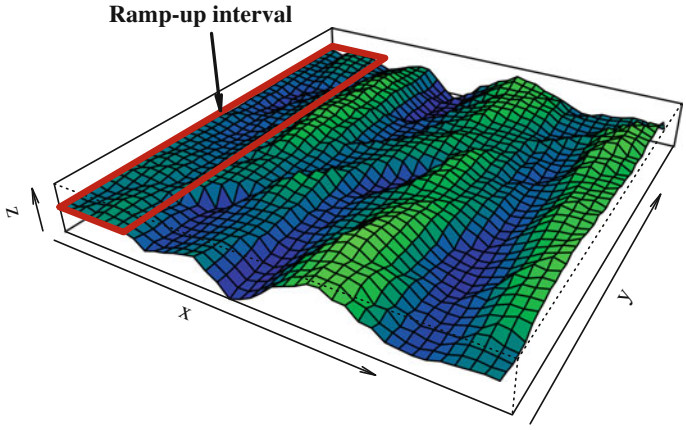


Fig. 6 Ramp-up interval at the beginning of the *OX* axis of the realisation

on the border (too far away from the studied marine object). If ship stability with manoeuvring is studied, then the interval may be simply discarded from the realisation (the size of the interval approximately equals the number of AR coefficients in each dimension). However, this may lead to a loss of a very large number of points, because discarding is done for each dimension. Alternative approach is to generate ocean wavy surface on ramp-up interval with LH model and generate the rest of the realisation with ARMA model.

Algorithm of wavy surface generation is data-parallel: the realisation is divided into equal parts along the time axis each of which is generated independently, however, in the beginning of each realisation there is ramp-up interval. To eliminate it for MA process, *overlap-add* method [24–26] (a popular method in signal processing) is used. The essence of the method is to add another interval, size of which is equal to the ramp-up interval size, to the end of each part. Then wavy surface is generated in each point of each part (including points from the added interval), the interval at the end of part N is superimposed on the ramp-up interval at the beginning of the part $N + 1$, and values in corresponding points are added. To eliminate the ramp-up interval for AR process, the realisation is divided into part along each dimension, and each part is computed only when all dependent parts are ready. For that purpose, an array of current part states is maintained in the programme, and all the parts are put into a queue. A parallel thread acquires a shared lock, finds the first part in the queue, for which all dependent parts have "completed" state, removes this part for the queue, frees the lock and generates the part. After that a thread updates the state of the part, and repeats the same steps until the queue becomes empty. This algorithm eliminates all ramp-up intervals except the one at the beginning of the realisation, and the size of the parts should be sufficiently small to balance the load on all processor cores.

Table 2 Formulae for computing $\mathscr{D}_1(x,z)$ and $\mathscr{D}_2(x,z)$ from section "Two-Dimensional Case", that use normalisation to eliminate uncertainty from definition of Dirac delta function of complex argument

Function	Without normalisation	Normalised
$\mathscr{D}_1(x,z)$	$\delta(x+iz)$	$\frac{1}{2h}\operatorname{sech}\left(\frac{\pi(x-i(h+z))}{2h}\right)$
$\mathscr{D}_2(x,z)$	$\frac{1}{2}\left[\delta(x-iz)+\delta(x+iz)\right]$	$\frac{1}{4h}\left[\operatorname{sech}\left(\frac{\pi(x-i(h+z))}{2h}\right)\right.$ $\left.+\operatorname{sech}\left(\frac{\pi(x+i(h+z))}{2h}\right)\right]$

Velocity Potential Computation

In solutions (17) and (19) to two-dimensional problem there are functions $\mathscr{D}_1(x,z) = \mathscr{F}_x^{-1}\left\{e^{2\pi uz}\right\}$ and $\mathscr{D}_2(x,z) = \mathscr{F}_x^{-1}\{\cosh(2\pi uz)\}$ which has multiple analytic representations and are difficult to compute. Each function is a Fourier transform of linear combination of exponents which reduces to poorly defined Dirac delta function of a complex argument (see Table 2). The usual way of handling this type of functions is to write them as multiplication of Dirac delta functions of real and imaginary part, however, this approach does not work here, because applying inverse Fourier transform to this representation does not produce exponent, which severely warp resulting velocity field. In order to get unique analytic definition, normalisation factor $1/\cosh(2\pi uh)$ (which is also included in the formula for $E(u)$) may be used. Despite the fact that normalisation allows to obtain adequate velocity potential field, numerical experiments show that there is little difference between this field and the one produced by formulae, in which terms with ζ are omitted. As a result, we do not use normalisation factors in the formula.

Evaluation

ARMA model does not require highly optimised software implementation to be efficient, its performance is high even without use of co-processors; there are two main causes of that. First, ARMA model itself does not use transcendental functions (sines, cosines and exponents) as opposed to LH model. All calculations, except model coefficients, are done via polynomials, which can be efficiently computed on modern processors using a series of fused multiply-add (FMA) instructions. Second, pressure computation is done via explicit analytic formula using nested FFTs. Since two-dimensional FFT of the same size is repeatedly applied to every time slice, its coefficients (complex exponents) are pre-computed for all slices, and computations are performed with only a few transcendental functions. In case of MA model, performance is also increased by doing convolution with FFT. So, high performance of ARMA model is due to scarce use of transcendental functions and heavy use of FFT,

Table 3 A list of mathematical libraries used in ARMA model implementation

Library	What it is used for
DCMT [23]	Parallel PRNG
Blitz [27, 28]	Multidimensional arrays
GSL [29]	PDF, CDF, FFT computation checking process stationarity
LAPACK, GotoBLAS [30, 31]	Finding AR coefficients
GL, GLUT [32]	Three-dimensional visualisation
CGAL [33]	Wave numbers triangulation

not to mention that high convergence rate and non-existence of periodicity allows to use far fewer coefficients compared to LH model.

ARMA implementation uses several libraries of reusable mathematical functions and numerical algorithms (listed in Table 3), and was implemented using OpenMP and OpenCL parallel programming technologies, that allow to use the most efficient implementation for a particular algorithm.

For the purpose of evaluation we use simplified version of (23):

$$\begin{aligned}\phi(x,y,z,t) &= \mathscr{F}^{-1}_{x,y}\left\{\frac{\cosh(2\pi|\boldsymbol{k}|(z+h))}{2\pi|\boldsymbol{k}|\cosh(2\pi|\boldsymbol{k}|h)}\mathscr{F}_{u,v}\{\zeta_t\}\right\}\\ &= \mathscr{F}^{-1}_{x,y}\{g_1(u,v)\mathscr{F}_{u,v}\{g_2(x,y)\}\}\,. \end{aligned} \tag{24}$$

This formula is particularly suitable for computation on GPUs:

- it contains transcendental mathematical functions (hyperbolic cosines and complex exponents);
- it is computed over large four-dimensional (t, x, y, z) region;
- it is analytic with no information dependencies between individual data points in t and z dimensions.

Since standing sea wave generator does not allow efficient GPU implementation due to autoregressive dependencies between wavy surface points, only velocity potential solver was rewritten in OpenCL and its performance was compared to existing OpenMP implementation.

For each implementation the overall performance of the solver for a particular time instant was measured. Velocity field was computed for one t point, for 128 z points below wavy surface and for each x and y point of four-dimensional (t, x, y, z) grid. The only parameter that was varied between subsequent programme runs is the size of the grid along x dimension. A total of 10 runs were performed and an average time of each stage was computed.

A different FFT library was used for each version of the solver. For OpenMP version FFT routines from GNU Scientific Library (GSL) [29] were used, and for OpenCL version clFFT library [34] was used instead. There are two major differences in the routines from these libraries.

- The order of frequencies in Fourier transforms is different and clFFT library requires reordering the result of (24) whereas GSL does not.
- Discontinuity at $(x, y) = (0, 0)$ of velocity potential field grid is handled automatically by clFFT library, whereas GSL library produce skewed values at this point.

For GSL library an additional interpolation from neighbouring points was used to smooth velocity potential field at these points. We have not spotted other differences in FFT implementations that have impact on the overall performance.

In the course of the numerical experiments we have measured how much time each solver's implementation spends in each computation stage to explain how efficient data copying between host and device is in OpenCL implementation, and how one implementation corresponds to the other in terms of performance.

Results

The experiments showed that GPU implementation outperforms CPU implementation by a factor of 10–15 (Fig. 7), however, distribution of time between computation stages is different for each implementation (Fig. 8). The major time consumer in CPU implementation is computation of g_1, whereas in GPU implementation its running time is comparable to computation of g_2. GPU computes g_1 much faster than CPU due to a large amount of modules for transcendental mathematical function computation. In both implementations g_2 is computed on CPU, but for GPU implementation the result is duplicated for each z grid point in order to perform multiplication of all XYZ planes along z dimension in single OpenCL kernel, and, subsequently copied to GPU memory which severely hinders overall stage performance. Copying the resulting velocity potential field between CPU and GPU consumes ≈20% of velocity potential solver execution time.

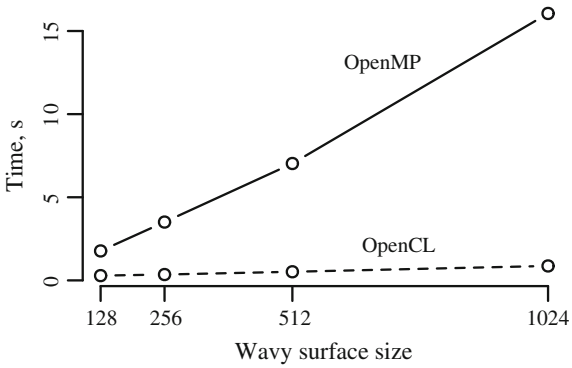


Fig. 7 Performance comparison of CPU (OpenMP) and GPU (OpenCL) versions of velocity potential solver

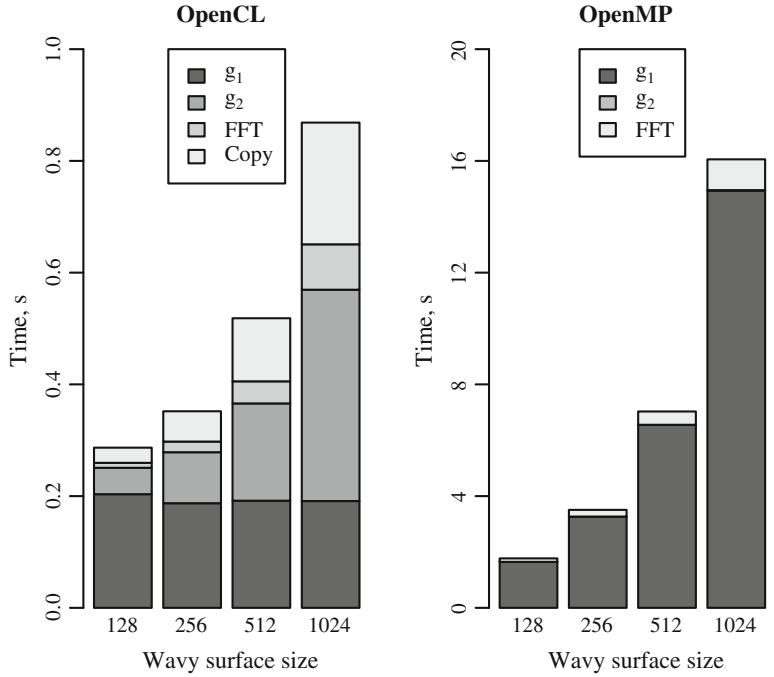


Fig. 8 Performance breakdown for GPU (OpenCL) and CPU (OpenMP) versions of velocity potential solver

Conclusion

Three-dimensional ARMA ocean simulation model coupled with analytic formula for determining pressures under wavy sea surface is computationally efficient of performing long-term ship behaviour simulations on the computer. Possible applications of the approach include studying ship behaviour in storm and shallow water waves. Its validity was visually and statistically verified in a number of experiments: distribution of characteristics of waves, produced by ARMA model, match the ones of real ocean waves, and velocity potential field, produced by the analytic formula correspond to the one produced by the formula for small-amplitude waves, and the formula itself reduces to the known one from linear wave theory.

Numerical experiments showed that wavy surface generation is efficient on CPU as it involves no transcendental mathematical functions, and velocity potential field computation is efficient on GPU due to heavy use of Fourier transforms. The use of dynamically generated Mersenne Twister PRNGs allows to produce uncorrelated sequences of pseudo-random numbers with no practical limitation on the realisation period, which in turn allows to perform long simulation sessions on parallel machines.

The future work is to make ARMA mathematical apparatus and its numerical implementation a base of virtual testbed for marine objects dynamics studies.

References

1. St. Denis, M., & Pierson, W. J., Jr. (1953). *On the motions of ships in confused seas*. Technical report, New York University, Bronx School of Engineering and Science
2. Rosenblatt, M. (1956). *A random model of the sea surface generated by a hurricane*. Technical report, DTIC Document
3. Sveshnikov, A. A. (1959). *Mathematics Akademii Mechanics and Engineering*, *3*, 32.
4. Longuet-Higgins, M. S. (1957). *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *249*(966), 321.
5. Kochin, N., Kibel, I., & Roze, N. (1966). *Theoretical hydrodynamics* (in Russian). FizMatLit.
6. Beck, R. F., Reed, A. M., Sclavounos, P. D., & Hutchison, B. L. (2001). *Transactions-Society of Naval Architects and Marine Engineers*, *109*, 1.
7. Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (revised ed.). Holden-Day.
8. Kostecki, M. (1972). *Stochastic model of sea waves*. Ph.D. thesis, CTO, Gdansk.
9. Rozhkov, V. A., & Trapeznikov, Y. A. (1990). *Probabilistic models of oceanographic processes*. Leningrad: Gidrometeoizdat.
10. Gurgenidze, A. T., & Trapeznikov, Y. A. (1988). *Probabilistic model of wind waves* (pp. 8–23). Leningrad: Gidrometeoizdat.
11. Spanos, P. D. (1982). *ARMA algorithms for ocean spectral analysis*. University of Texas at Austin, Engineering Mechanics Research Laboratory.
12. Spanos, P. D., & Zeldin, B. (1996). *Earthquake Engineering and Structural Dynamics*, *25*(5), 497.
13. Fusco, F., & Ringwood, J. V. (2010). *IEEE Transactions on Sustainable Energy*, *1*(2), 99.
14. Degtyarev, A., & Gankevich, I. (2012). *Proceedings of 11th International Conference on Stability of Ships and Ocean Vehicles, Athens* (pp. 841–852)
15. Degtyarev, A. B., & Podoliakin, A. B. (1998). *Proceedings of II International Conferences on Shipbuilding (ISC'98), Saint-Petersburg* (Vol. V, pp. 416–423).
16. Degtyarev, A., & Boukhanovsky, A. (1997). *Analysis of peculiarities of ship-environmental interaction*. Technical report, 09-97-1AB-1VA, Strathclyde University, Ship Stability Research Center, Glasgow.
17. Boccotti, P. (1983). *Meccanica*, *18*(4), 205.
18. Degtyarev, A. B., & Reed, A. M. (2011). *Proceedings of the 12th International Ship Stability Work-shop*.
19. Degtyarev, A. B., & Reed, A. M. (2013). *International Shipbuilding Progress*, *60*(1–4), 523.
20. Boukhanovsky, A. V. (1997). *Probabilistic modeling of wind wave fields taking into account their heterogeneity and nonstationarity*. Ph.D. thesis, Saint Petersburg State University.
21. Wolfram Research Inc. (2016). *Mathematica*. Champaign, Illinois.
22. Matsumoto, M., & Nishimura, T. (1998). *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, *8*(1), 3.
23. Matsumoto, M., & Nishimura, T. (1998). *Monte Carlo and Quasi-Monte Carlo Methods*, *2000*, 56.
24. Oppenheim, A. V., Schafer, R. W., Buck, J. R., et al. (1989). *Discrete-time signal processing* (Vol. 2). Englewood Cliffs, NJ: Prentice Hall.
25. Svoboda, D. (2011). *Image Analysis and Processing–ICIAP 2011* (pp. 453–462). Springer.
26. Pavel, K., & David, S. (2013). *Algorithms for efficient computation of convolution*. INTECH Open Access Publisher.

27. Veldhuizen, T. L., & Jernigan, M. E. (1997). *International Conference on Computing in Object-Oriented Parallel Environments* (pp. 49–56). Springer.
28. Veldhuizen, T. (2000). *Computer Science Technical Report*, *542*, 60.
29. Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Alken, P., et al. (2009). *GNU scientific library reference manual*. In B. Gough (Ed.) (3rd ed.). Network Theory Ltd.
30. Goto, K., & Van De Geijn, R. (2008). *ACM Transactions on Mathematical Software (TOMS)*, *35*(1), 4.
31. Goto, K., & Geijn, R. A. (2008). *ACM Transactions on Mathematical Software (TOMS)*, *34*(3), 12.
32. Kilgard, M. J. (1996). *The OpenGL Utility Toolkit (GLUT) Programming Interface*. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.34.8270
33. Fabri, A., & Pion, S. (2009). *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 538–539). ACM.
34. clFFT developers. clFFT: OpenCL Fast Fourier Transforms (FFTs). https://clmathlibraries.github.io/clFFT/.

Perturbation Theory for the Compound Soliton of the Gardner's Equation; Their Interaction and Evolution in a Media with Variable Parameters

Irina A. Soustova, Konstantin A. Gorshkov, Alexey V. Ermoshkin, Lev A. Ostrovsky and Yuliya I. Troitskaya

Introduction

The results of studying the interaction and propagation of compound solitons in the media with variable parameters are reported. Such solitons are encountered in the systems, in which the waves in the form of kinks (field jumps) may exist in addition to the solitary waves. These solitons are referred to as compound as they may be represented as an ensemble of interacting kinks which makes possible reduced description of their evolution to the evolution of kinks. This approach, in contrast to the traditional one [3], when the solitons are considered as an integral formation, permits studying non-quasi-stationary processes related to the interaction of compound solitons of strongly differing sizes and velocities, as well as to their propagation in the media with variable parameters. Some of these processes will be considered below within the framework of the Gardner equation with variable coefficients

$$\Phi_t + \Phi(\alpha(t) - \mu(t)\Phi)\Phi_x + \beta\Phi_{xxx} = 0 \tag{1}$$

This equation is frequently used for describing internal waves in the shelf region of the oceans and seas. Hence, the presented examples are of practical interest [4, 14, 16, 15, 17–19].

This contribution is a brief review of the results of an approximate description of the evolution and interaction of composite solitons obtained by the authors in 2001–2016. As one of the applications of the theory, the features of the evolution of intense internal waves in the shelf zone of the ocean are analyzed.

I. A. Soustova (✉) · K. A. Gorshkov · A. V. Ermoshkin · L. A. Ostrovsky
Y. I. Troitskaya
Institute of Applied Physics, Russian Academy of Sciences, Nizhny Novgorod, Russia
e-mail: soustova@hydro.appl.sci-nnov.ru

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_19

Basic Points of the Approximate Approach

Equation (1) with constant coefficients α, β, and $\mu > 0$, has a family of soliton solutions:

$$\Phi_s(x,t) = \overline{\Phi} + \frac{D}{2}[th\,\lambda(x - vt + \Delta) - th\,\lambda(x - vt - \Delta)], \tag{2}$$

which depend on an arbitrary pedestal $\overline{\Phi}(x,t)$ and on one more parameter, namely, dimensional variable $\lambda\Delta = r$; $D = \Phi_m\, th\, 2r$, $\lambda = \frac{1}{2}\sqrt{\frac{v_m - v_0}{\beta}}\, th\, 2r$, $\Delta = r\sqrt{\frac{\beta}{v_m - v_0}}cth2r$, $\frac{(v - v_0)}{(v_m - v_0)} = th^2 2r$, $\Phi_m = \frac{\alpha}{\mu} - 2\overline{\Phi}$, $v_m = \frac{\alpha^2}{6\mu} + \frac{\overline{\Phi}}{3}\left(\alpha - \mu\overline{\Phi}\right)$, and $v_0 = \overline{\Phi}\left(\alpha - \mu\overline{\Phi}\right)$. Soliton amplitude $A = \max\left(\Phi_s - \overline{\Phi}\right)$ and width L at HWFM ($\frac{1}{2}A$ level) is also expressed through parameter $r = \lambda\Delta$: $A = \Phi_m(1 - ch^{-1}2r)$, $ch\lambda L = 2 + ch2r$.

Besides soliton solutions, Eq. (1) has a one-parameter family of kink solutions

$$\Phi_k^+(x,t) = \overline{\Phi} + \frac{\Phi_m}{2}[1 \pm th\lambda_m(x - v_m t)], \tag{3}$$

where $\lambda_m = \frac{1}{2}\Phi_m\sqrt{\frac{\mu}{6\beta}}$.

Soliton solution (2) may be written in the exact form as a combination of kink solutions with re-normalized (due to kink interaction) parameters: $\Phi_m \to D, \lambda_m \to \lambda, v_m \to v$:

$$\Phi_s(x,t) = \Phi_k^+ + \Phi_k^- - (\overline{\Phi} + D) \tag{4}$$

Parameter r is equal to the ratio of the distance between the kinks (2Δ) and the characteristic scale of field jumps in the kinks $\left(\lambda_m^{-1}\right)$. At $r \ll 1$, the soliton amplitudes are small due to the strong overlapping of the kinks. Their velocity is close to minimal v_0 and the size does not depend on 2Δ; it is determined only by the λ_m^{-1} scale. In general, such solitons are close to the KdV solitons and have no obvious signs of a compound structure. At $r \gg 1$, the soliton amplitude and velocity are close to the maximum possible values of Φ_m, v_m. The soliton shape tends to a rectangle and size L, which almost coincides with the distance between the kinks 2Δ, hence the composite structure of the solitary wave becomes apparent. The structure of expression (4) is typical for the solutions obtained by the method of matched asymptotic expansions. The general solution consists of the sum of solutions in internal regions with fast field variation $\Phi_k^+ + \Phi_k^-$ (soliton field jumps $\approx[th\lambda_m(x - v_m t)]$) after deduction of their total asymptotic behavior that is a solution in the external regions with slow field variation (in this case, they are constant values $(\overline{\Phi} + D)$). Expression (4) gives a correct description of the general structure of the field. This representation will be used below for describing quasi-nonstationary processes. Construction of the solutions by the method of

matched asymptotic expansions implies finding local corrections near each kink and their subsequent matching in the regions between the kinks.

Compound Soliton Interaction

Let us consider the interaction of N compound solitons of the Gardner equation (parameters of the equation are assumed to be constant) with arbitrary ratio of their sizes. The basic expression for describing their evolution is a superposition of 2N kinks of alternating polarity (3) after deduction of their total asymptotic behavior:

$$\Phi_{Ns}^{(0)}(x,t) = \frac{1}{2}\sum_{i=1}^{2N}(-1)^{i+1}\tanh\lambda_m[x - V_m t - S_i(\varepsilon t, \varepsilon x)] \tag{5}$$

where S_i are kink center coordinates. The procedure of matching the first corrections yields equations for $S_i(x,t) = S_i(\eta), \eta = x - 3t$:

$$\left(\frac{dS_i}{dt}\right) = -4\left[e^{-(S_{i+1}-S_i)} - e^{(S_{i-1}-S_i)}\right] \tag{6}$$

and corrects the general solution for field distribution (5) [5]:

$$\Phi_{N_i}^{(0)} + \Phi_{N_i}^{(1)}(x,t) = \frac{1}{4}\sum_{i=1}^{2N}(-1)^{i+1}\left(1 - \frac{\partial S_i}{\partial x}\right)th\frac{1}{2}(\xi - S_i) \tag{7}$$

Exactly integrable system (6), the Langmuir chain equation, describes the elastic collision of N solitons without changing their number and parameters. Only phase shifts of the solitary waves appear. The compound character of the Gardner equation solitons manifests itself in the asynchronous motion of fronts and decays of each solitary wave at the stage of their closest convergence. When the kinks belonging to two different solitons are temporally combined as a result of soliton collision (Fig. 1), a narrow soliton transforms to an identical soliton of the opposite polarity moving on the crest of a wide soliton from its front to decay.

The collision process finishes by the inverse transformation of the narrow soliton to the soliton of initial polarity close to the decay of the wide solitary wave. Approximate description (5–7) is quite close to the exact one in a wide range of relative soliton velocities (up to $\dot{S}_i - S_{i\pm}/\dot{V}_m \sim 1/2$). It is interesting that the exact N-soliton solution, similarly to the approximate solution, may be represented as a superposition of quasi-kinks presented by form (7) [5, 6].

It is important that approximate description of N-soliton interactions may also be constructed for strongly nonlinear models used for describing intense internal waves (IIW), for example, the Miyta model [7, 8] for a two-layer liquid:

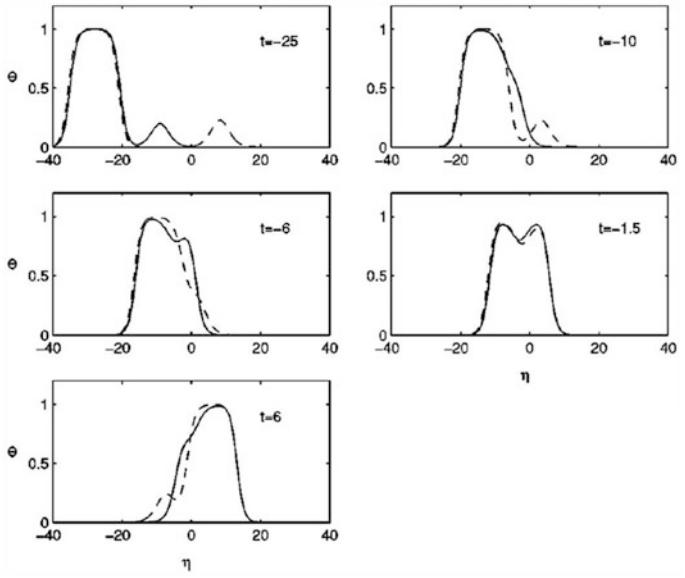


Fig. 1 Interaction of two solitons. Solid line shows exact solution. Dashed line shows an approximate solution

$$\eta_{1,2_t} + \left(\eta_{1,2} u_{1,2}\right)_x = 0;$$
$$\rho_{1,2}\left(u_{1,2_t} + u_{1,2} u_{1,2_x} + g\xi_x\right) = -p_x + \frac{1}{3\eta_{1,2}}\left(\eta_{1,2}^3 \left(\frac{\partial}{\partial t} + u_{1,2}\frac{\partial}{\partial x}\right)^2 \xi\right)_x, \tag{8}$$

where $\eta_{1,2} = h_{1,2} \mp \xi$, $h_{1,2}$ are the perturbed and unperturbed thicknesses of the upper and lower liquid layers, ξ is the vertical shift of the interface, $u_{1,2}$ are the average (over the vertical coordinate) values of the horizontal velocities of the liquid in the layers, $\rho_{1,2}$ are the liquid densities, g is acceleration of gravity, and p is pressure. The family of stationary solutions of system (8) is qualitatively similar to the family of solitary waves of the Gardner equation (Fig. 2).

However, the asymptotic forms of the kinks of model (8) have different speed of field decay (exponents) at $x \to \pm\infty$, whereas the asymptotic forms of kinks of the Gardner equation are identical. Nonsymmetry of soliton field jumps affects significantly the process of soliton collisions and changes the form of the equations for kink coordinates:

$$\frac{dS_i}{dt} = I(\tau,\rho) - 2\left\{\begin{array}{l} M_0 e^{-\Lambda_0(S_i - S_{i-1})} + M_m e^{-\Lambda_m(S_{i+1} - S_i)},\ i - \ odd \\ M_m e^{-\Lambda_m(S_i - S_{i-1})} + M_0 e^{-\Lambda_0(S_{i+1} - S_i)},\ i - \ even \end{array}\right\} \tag{9}$$

Equations (9) may be reduced to the integrable Langmuir chain system mentioned above, so that the interaction of the solitons of Eq. (8) persists to be elastic

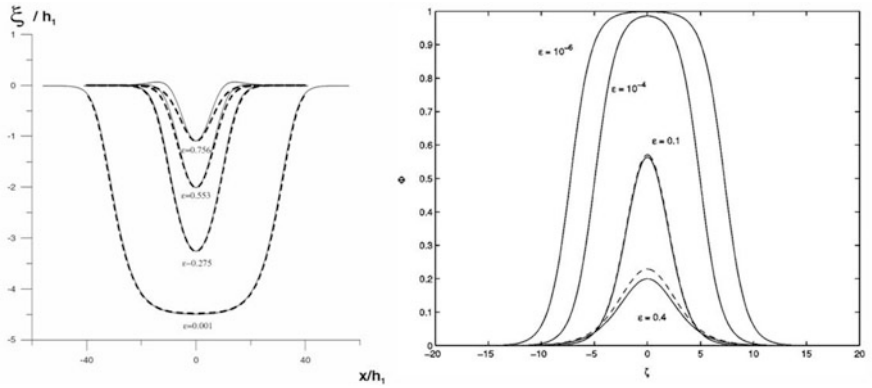


Fig. 2 The family of stationary solutions of system (8) (left panel) and Gardner equation (right panel)

and the solitary wave number and parameters remain unchanged after the interaction. However, anomalies appear in the dependence of the soliton phase shifts on their parameters that may be easily understood using an example of two-soliton interaction (Fig. 3) [7].

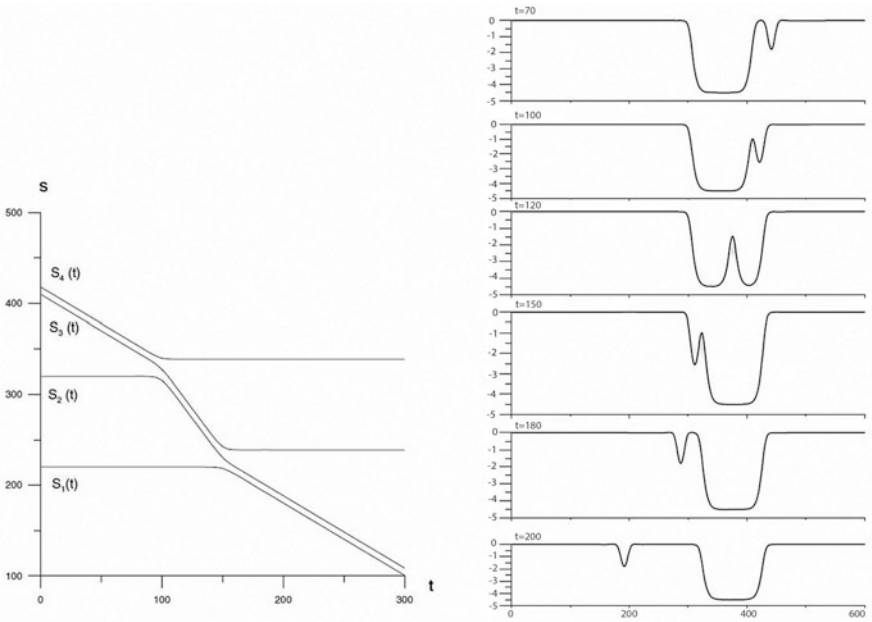


Fig. 3 Concurrent interaction of two solitons, one of which is close to the limiting soliton (CC-model)

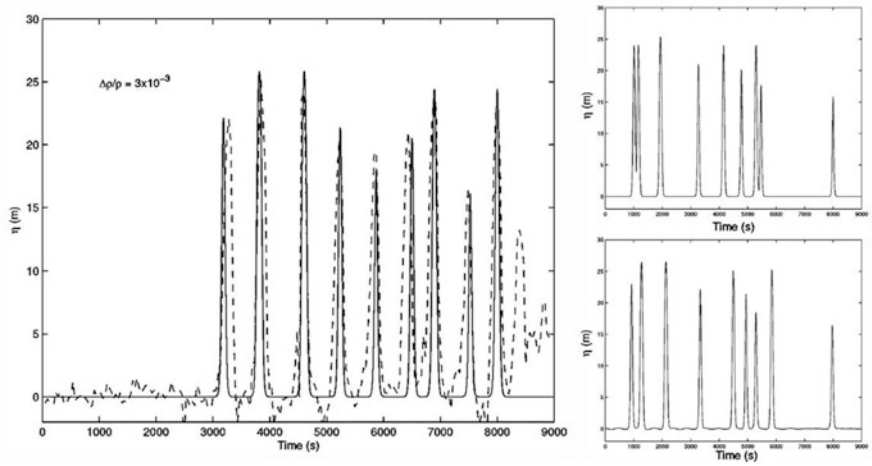


Fig. 4 Soliton distribution at the initial point (8 km offshore (left panel) and at the end point (20 km offshore (right panel). Dashed are related to the approximate model, and solid curve shows numerical computations [14]

One can see that the narrow soliton of reverse polarity arising behind the front of the wide soliton has a different amplitude and a different velocity (this is important!). During the motion on the crest of the wide soliton it acquires an additional phase shift which increases as the size of the wide soliton increases as well as the difference in velocities increases. Clearly, such singularity does not appear when the velocities of solitons have opposite polarities and coincide for slow solitons. This case is realized for the solitons of the Gardner equation. Thus, the proposed approach permits the distributed problem to be reduced to a system of ordinary differential equations and even the wave nature of the interacting structures (finite velocity of perturbation propagation, deformation of the shape, phase shifts, etc.) can be taken into account. The developed theory was used for describing the propagation of a train of solitons of strongly nonlinear internal waves (IIW) observed during the COPE experiment [14] (Fig. 4).

Compound Soliton Evolution in the Media with Variable Parameters

The proposed approach is especially efficient for investigating the evolution of compound solitons in the situations critical for quasi-stationary description, when the predicted increase of the solitary wave scales becomes unrestricted over finite space-time intervals. Such situations are revealed from the analysis of the quasi-stationary Gardner equation when coefficients $\alpha(t), \mu(t) > 0, \beta(t)$ vary slowly in time compared to the soliton duration (L/v). In this case, the basic equation

defining the solitary wave evolution follows from the law of conservation of the total soliton wave pulse $\int_{-\infty}^{+\infty} \Phi_s^2(x,t)dx = P_0 = const$; for $\overline{\Phi} = 0$ it has the following form:

$$2r - th2r = P_0 \left|\alpha^{-2}\mu^3\beta^{-1}\right|^{\frac{1}{2}} \tag{10}$$

The left-hand side of (10) as a function of r is increasing monotonically throughout the $0 < r < +\infty$ range, therefore the soliton scale increases without restriction when the combination of the coefficients in the right-hand side of (10) tends either to zero or to infinity. As the developed approach greatly relies on the composite character of solitons, we will further address the $r \to \infty, (\left|\alpha^{-2}\mu^3\beta^{-1}\right|^{\frac{1}{2}}) \to \infty$ limit. In a typical situation corresponding to this limit in the case of IWs, the α(t) function has a point $\alpha(t_{cr}) = 0$ which in real conditions corresponds to symmetric position of the thermocline. Such a problem was solved numerically in [9] using linear function $\alpha(t) = 1 + \varepsilon t$. In this work, this problem is solved analytically in a more general formulation.

An approximate description is constructed in conformity with the compound structure of the solitons of Eq. (1). Solutions are sought independently in relatively narrow regions of kinks and in more extended regions between them, and then they are matched. The scales of medium parameters variation $(\approx\Delta)$ are assumed to be much greater than the scales of the soliton field jumps $\left(\approx\lambda_m^{-1}\right)$, but remain small or comparable with the distances and intervals between them. The small parameter of the problem is of the order of $(\lambda_m\Delta)^{-1}$ (Fig. 5).

The above scale ratio permits us to assume that the evolution of field jumps is quasi-stationary and the fields outside the kinks vary slowly; they are described by the basic Eq. (1) at $\mu = 1$ in the dispersion-free approximation (simple wave equation):

$$\Phi_t + \Phi(\alpha(t) - \Phi)\Phi_x = 0 \tag{11}$$

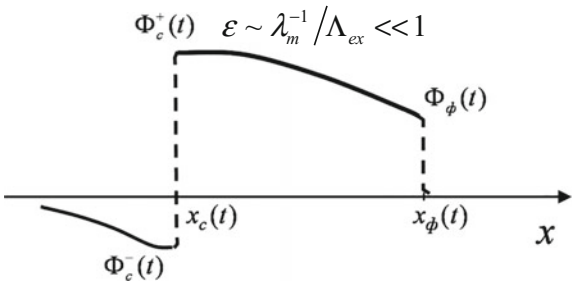


Fig. 5 Structure of a compound soliton of the Gardner equation

The quasi-stationary character of kinks allows us to connect the slowly varying fields $\Phi^{\pm}(x,t)$ (Fig. 5) from the regions adjoining the kink trajectory $x_k(t)$ (from $x > x_k(t)$ and $x < x_k(t)$)

$$\Phi^{+}(\mathrm{x_k},\mathrm{t}) = \alpha(\mathrm{t}) - \Phi^{-}(\mathrm{x_k},\mathrm{t}) \tag{12}$$

Kink trajectory $x_k(t)$ is defined by the stationary equation for its velocity:

$$\frac{dx_k}{dt} = \frac{\alpha^2}{6} + \frac{\Phi}{3}(\alpha(t) - \Phi) \tag{13}$$

As was stated in [9], the characteristic velocities of slowly varying perturbations from the regions $x > x_k(t)$ and $x < x_k(t)$ near the kink are identical and are always smaller than the kink velocity. It follows from this that kink trajectory $x_k(t)$ from (8) with the values of the field Φ_k^+ on this trajectory is the initial data line for finding field $\Phi(x,t)$ in the region $x < x_k(t)$ from Eq. (6). The quality of the characteristic velocities allows us to use the field value Φ_k^+ in (13) as well as Φ_k^-. As a result, Eq. (11) with conditions (12, 13) form a closed system for successive finding of slowly varying fields in all regions between the kinks, starting from the region before the kink corresponding to the quasi-soliton front. We will illustrate this algorithm for the case, when there are no perturbations before the soliton and the values of field are $\Phi_f(x_f(t),t) = \Phi_k(x_k(t) = x_f(t),t)$ and front coordinate $x_f(t)$ are found immediately from (12)–(14).

$$\Phi_{\Phi}(x_{\Phi},t) = \alpha(t),\ x_f(t) = \int_0^t v_m(t')dt' + x_f(0) = 1/6\int_0^t \alpha^2(t')dt' + x_f(0). \tag{14}$$

Initial data line (14) permits obtaining a solution describing field distribution in the implicit form at $x < x_k(t)$, i.e., on the quasi-soliton crest:

$$x - x_f(t_f(\Phi) = \Phi\big(A(t) - A(t_f(\Phi))\big) - ((\Phi))^2\big(t - t_f(\Phi)\big), \tag{15}$$

where $A(t)$ is the antiderivative of the function $\alpha(t)$, and $t_{\Phi}(\Phi)$ is the function inverse to $\alpha(t)$. Since the field near the quasi-soliton decay from the crest is described by the family of characteristics (13), then assuming $\mathrm{x} = \mathrm{x_c}(\mathrm{t}), \Phi(\mathrm{x_c}(\mathrm{t}),\mathrm{t}) = \Phi_{\mathrm{c}}^{+}(\mathrm{t})$ in (13) and using (15) we obtain a system of equations for determining the magnitudes of these quantities. Functions $\mathrm{x_c(t)}$, $\Phi_{\mathrm{c}}^{+}(\mathrm{t})$ obtained from this system form the initial data line for finding field $\Phi(\mathrm{x},\mathrm{t})$ for $\mathrm{x} < \mathrm{x_c(t)}$ in form (15) with the substitution $x_f \to x_c(t(\Phi_c),\ t_f \to t_c(\Phi)$. Solutions may be obtained in the analytical form for a rather wide class of power functions $\alpha(t) = (1 - \varepsilon t)^{\delta}$, however, the general pattern of the evolution is rather illustrative and reduces to the following. The perturbations arising near the front with velocities smaller than the velocity of the front move along the quasi-soliton crest in the

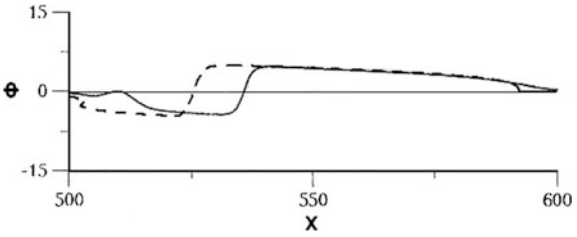


Fig. 6 Structure of quasi-soliton field at moment t_{cr}. Dashed line is the approximate method, solid line depicts numerical simulation of the Gardner equation [11]

direction of its decay without changing their magnitude: the value of field $\Phi_c^+(t)$ is equal to the value of the field near the front at a certain preceding moment of time $(t-\tau)$. For a monotonically decreasing function $\alpha(t)$, the field along the quasi-soliton crest decreases monotonically from its decay to the front. Such a character of the field persists up to the moment $t=t_{cr}$, when $\alpha(t_{cr})=0$; hence, $\Phi_f(t_{cr})=0$. The field near the decay $\Phi_c^+(t_{cr})$ remains finite, and the field jump becomes symmetric $(\Phi_c^+(t_{cr}) = -\Phi_c^-(t_{cr}))$ and remains finite. while the size of the quasi-soliton is $L(t)=x_f(t)-x_c(t)$ due to the finiteness of the front and decay speeds at all stages of the process. For decreasing function $\alpha(t)$, the decay velocity is always $\Phi_c^+\left(\Phi_c^+-\alpha\right)/3$ smaller than the front velocity, so that $L(t)=L(0)+1/3\int_0^t \Phi_c^+(\Phi_c^+-\alpha)dt$ increases monotonically in the $0<t<t_{cr}$ interval, and the decay velocity in the neighborhood of $\alpha=0$ is always negative $(\dot{x}_c(t_{cr}) = -(\Phi_c^+(t_{cr}))^2/3)$.

The regions near the quasi-soliton front and behind its decay remain problematic for approximate description. As $\alpha\to 0$, the kink evolution at the front becomes non-quasi-stationary due to its growing scale $\left(\lambda_m^{-1}\approx\alpha^{-1}\right)$. However, in view of the relative smallness of the field values and of the size of the front-line region practically all the quasi-soliton crest is described correctly within the framework of the proposed approach. In the region behind the decay the approximate approach gives correct values of field Φ_c^-, but it does not adequately describe the field evolution as a whole in this region for all δ. Already at $\delta = 1$ intense steeping and pronounced field profile ambiguities appear (Fig. 6).

Modeling of IIW Evolution in the Vicinity of the Critical Point Based on the Results of the Experiment; Radar Portraits of IIWs in the Shelf Zone

The example considered above is relevant to for studying the IIW evolution in the shelf zone of the oceans, where the Gardner Eq. (1) also holds true for the vertical shift function $A\sim F$ related to of the interface between two liquids. In the approximation of a two-layer liquid immobile, the nonlinearity and dispersion coefficients of the Gardner equation have the form [4]:

$$c = \sqrt{\frac{g\Delta\rho}{\rho}\frac{h_1 h_2}{h_1 + h_2}},\ \beta = \frac{h_1 h_2}{c^2 6},\ \alpha = \frac{3}{2}\frac{h_1 - h_2}{\sqrt{2\Delta\rho g c} h_1 h_2},$$

$\mu = -\frac{3}{16 g\Delta\rho c h_1^2 h_2^2}(h_1^2 + h_2^2 + 6h_1 h_2)$, $\Phi = \sqrt{2\Delta\rho g c}A$, $\Delta\rho$ is the density jump between the upper layer having thickness h_1 and is lower layer of thickness h_2. In the case typical of the shelf zone, the change in coefficients is due to the change in depth and coefficients of the Gardner equation are functions of the variable $\tau = \int^x \frac{dx}{c}$ that is a "ray" coordinate.

According to [18], a solitary wave was observed on the shelf of the Kamchatka Peninsula with the following parameters: propagation velocity 0.51 m/s, soliton amplitude 10 m (leading height 14 m), soliton width 500 m (17 min), linear propagation velocity c = (0.24 – 0.35) m/s, thermocline position h1 = 14.5 m; the depth of the ocean varied linearly with coefficient 0.017. Below, we will present the calculated coefficients of the Gardner equation and values of the limiting soliton amplitude for the given hydrological parameters (Fig. 7). One can see that the coefficient of quadratic nonlinearity changes its sign at a distance of 1 km from the shore; the coefficient of cubic nonlinearity is significantly non-zero. Comparison of the results of observations [18] with the results of calculations using the approximate model [10] shows complete qualitative coincidence of the basic feature of transformation of the leading and rear fronts of the solitary wave.

Remote sensing of intense internal waves in the shelf zone, as shown in [14], can be carried out using a radar installed on the shore. Unfortunately, in [18], the remote sensing data are not available, but it is possible to numerically simulate the manifestation of an evolving soliton of intense internal waves approaching the critical point at the sea surface and in the radar images.

Information about the inhomogeneous flow at the sea surface induced by intense internal waves is needed to calculate surface manifestations of IIW solitons. In the approximation of two-layer stratification, under the conditions of the "rigid lid" at the surface and the shallow water approximation we can find from the mass

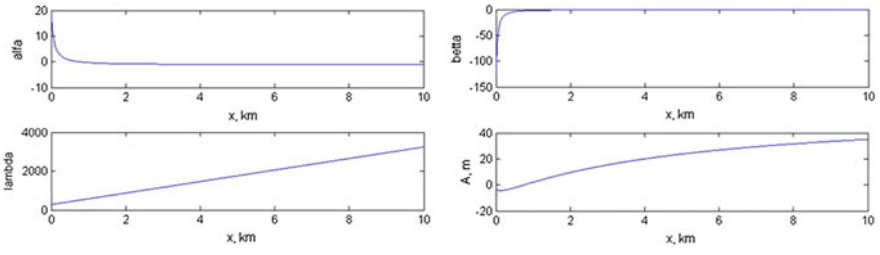


Fig. 7 Coefficients of the dimensionless Gardner equation and quasi-stationary amplitude of limiting soliton front

conservation law $\partial_t(h_1+\eta)+\partial_x[(h_1+\eta)U]=0$, the relationship between the IIW soliton amplitude $A(x,t)$ and the velocity field at the surface $U(x,t)$ [12]:

$$U(x,t)=c\frac{A(x,t)/h_1}{1+A(x,t)/h_1} \tag{16}$$

The surface flow velocity was calculated for the parameters of the experiment described above. The results of the calculations are presented in Fig. 8.

The transformation of surface waves in the field of an inhomogeneous flow induced by the IIW soliton propagating on the shelf was modeled on the basis of the kinetic equation for spectral density of the wave action of surface waves in the relaxation approximation:

$$\frac{\partial N}{\partial t}+\dot{\vec{r}}\frac{\partial N}{\partial \vec{r}}+\dot{\vec{k}}\frac{\partial N}{\partial \vec{k}}=\alpha N-\frac{\alpha N^2}{N_0}, \tag{17}$$

where $N\left(\vec{r},\vec{k},t\right)=W\left(\vec{r},\vec{k},t\right)/\sqrt{g\left|\vec{k}\right|}$ is the spectral density of the wave action of surface waves represented through the spatial spectrum of elevations $W\left(\vec{r},\vec{k},t\right)$ for which the JONSWAP expression was used [1]. The Hughes [13] increment was used for the wind wave growth rate α in Eq. (17).

The spatial distribution of flow velocity at the surface is shown in Fig. 9. The wind speed in the numerical experiment was 5 m/s, with an angle of 30° to the direction of IIW soliton propagation. It is the case of concurrent wave propagation relative to the flow moving with the speed of the soliton. In this case a definite range of wind wave spectrum falls into synchronism with the IW, i.e., it remains in the field of inhomogeneous flow for a long time, during which the effect is accumulated. The group velocity of wind waves should coincide with the velocity of IW soliton to meet the condition of synchronism. The variability of the spectral density

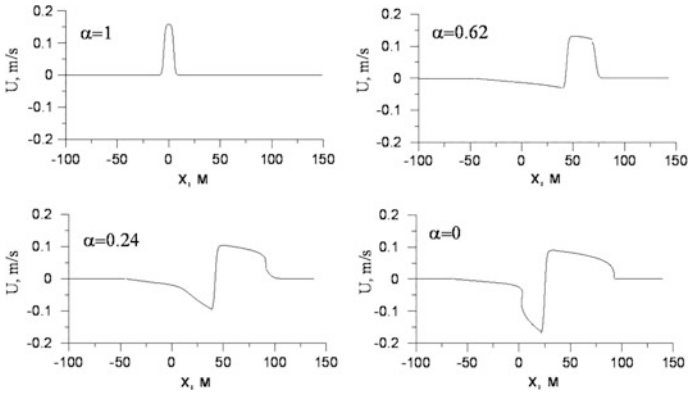


Fig. 8 Calculated surface flow velocities in the IIW field at different stages of the evolution

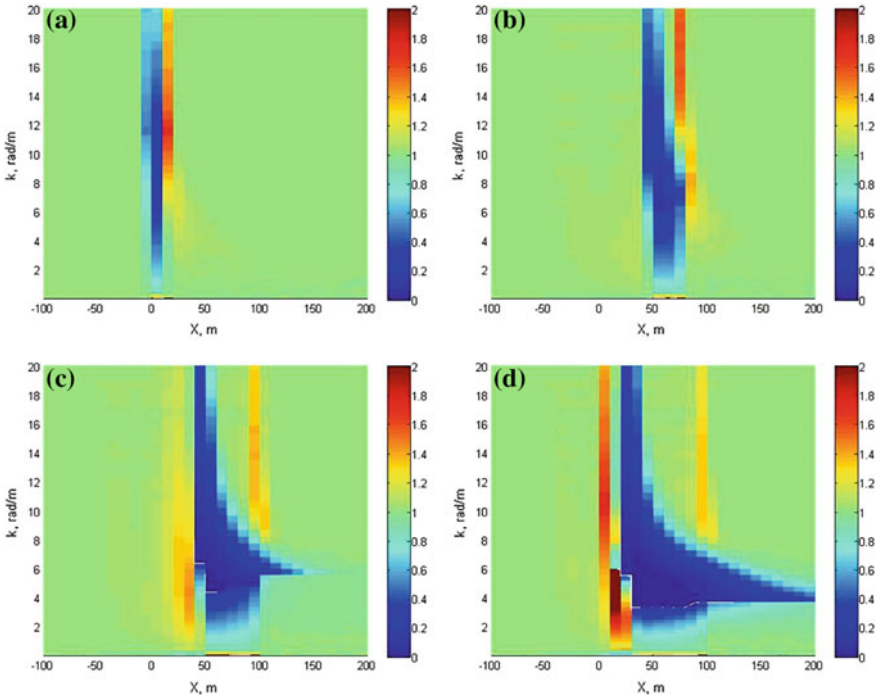


Fig. 9 Transformation of wind wave spectrum above the region of IIW soliton during its propagation on the shelf: $\alpha = 1$ (**a**), $\alpha = 0.62$ (**b**), $\alpha = 0.24$ (**c**), and $\alpha = 0$ (**d**). The color palette illustrates spectral density variability

of wind wave spatial spectrum above the region of IIW soliton propagating on the shelf is shown in Fig. 9 for different moments of time corresponding to Fig. 8. It is clearly seen that the region of wave numbers close to $k = 10$ rad/m is subjected to the maximum transformations by the field of inhomogeneous soliton flows.

The results of modeling the wind wave spectra transformation in the field of inhomogeneous flows of IIW soliton and the empirico-theoretical model of radio wave scattering by the rough water surface at sliding test angles described in [2] were used to obtain a radar panorama of the sea surface during sounding the field with horizontal polarization imaging at an angle of incidence of 88°, which is typical for coastal radars. The theoretical radar panoramas of the sea surface above the IIW region at different stages of soliton evolution are presented in Fig. 10. The images correspond to the following values of nonlinearity coefficient: $\alpha = 1$, $\alpha = 0.62$, $\alpha = 0.24$, and $\alpha = 0$, the latter being the critical point for the IIW soliton. Analysis of radar images leads to the conclusion that the waves and radar contrasts become weaker over the concurrent flow that is observed at the beginning of IIW soliton propagation, but when a counter flow appears a negative "tail" appears in the region of wave and radar contrast enhancement (it is shown in Fig. 8 for $\alpha = 0.24$ and $\alpha = 0$).

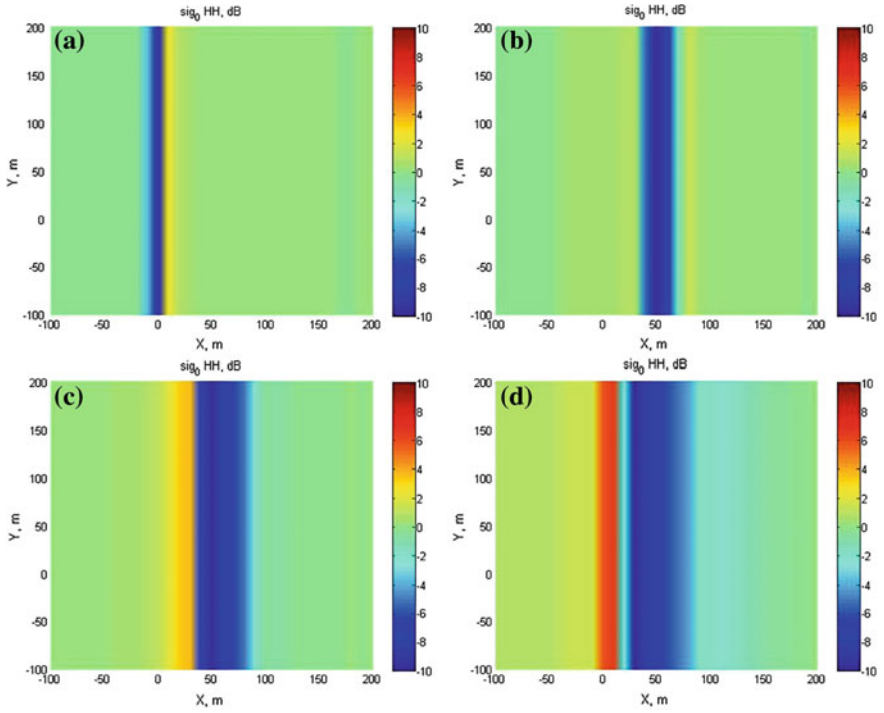


Fig. 10 Theoretical radar panoramas of the sea surface above the IIW soliton region at different stages: $\alpha = 1$ (**a**), $\alpha = 0.62$ (**b**), $\alpha = 0.24$ (**c**), $\alpha = 0$ (**d**). The color palette shows radar contrast in dB

Our calculations demonstrate high radar contrasts of the considered phenomenon: manifestations of IIW soliton at the sea surface during its propagation on the shelf. The quantitative calculations were made under the assumption characterizing the cascade mechanism: the energy transfer from the longwave region of the wind wave spectrum to short waves. It should be noted that the radar panoramas may be used for reconstructing the basic IIW soliton parameters: propagation velocity, distribution of the flow sign over the radar contrast, spatial size of the IIW soliton. These parameters together make radar imaging an efficient tool for IIW diagnostics on the shelf.

Conclusion

In this contribution we tried to briefly outline the main features and possibilities of an approximate analytical description of the non-quasistationary evolution of solitons. As an example of the application of the method, the features of the evolution of IIW in the vicinity of the critical points of the shelf zone of the ocean

are studied, where the depths of the upper and lower layers of the fluid with different densities are compared. The qualitative and quantitative coincidence with the results of the real IIW observation experiment on the shelf of the Kamchatka Peninsula is demonstrated. It is important that the analytical approach developed here allows rapid assessment of the amplitude and structure of the IIW at any arbitrary time with known hydrology and topography parameters of the surveyed region. The possibility of detecting the features of an IIW at critical points by radar images (recorded when sounding by the ship or shore radar) is demonstrated.

Acknowledgements This study was supported by the Russian Foundation for Basic Research (projects 16-55-52022, 18-05-00292 and 17-05-41117 RGS). Numerical code development and numerical modeling were supported by the Russian Science Foundation (grant 15-17-20009).

References

1. Battjes, J. A., Zitman, T. J., & Holtheijsen, L. H. (1987). A re-analysis of the spectra observed in JONSWAP. *Journal of Physical Oceanography, 17,* 1288–1295.
2. Ermoshkin, A. V., Bakhanov, V. V., & Bogatov, N. A. (2015). Development of an empirical model for Radar backscattering cross section of the ocean surface at grazing angles. *Sovremennye problemi distantsionnogo zondirovaniya Zemli iz kosmosa, 12*(4), 51–59.
3. Gorshkov, K. A., & Ostrovsky, L. A. (1981). Interaction of solitons in nonintegrable systems. *Physica D: Nonlinear Phenomena, 3,* 428–438.
4. Grimshaw, R., Pelinovsky, E., & Talipova, T. (2007). Modeling internal solitary waves in the coastal ocean. *Surveys In Geophysics, 28,* 273–287.
5. Gorshkov, K. A., & Soustova, I. A. (2001). Interaction of solitons as compound structures in the Gardner model. *Radiophysics and Quantum Electronics, 44*(5–6), 502–512.
6. Gorshkov, K. A., Ostrovsky, L. A., Soustova, I. A., & Irisov, V. G. (2004). Perturbation theory for kinks and application for multisoliton interactions in hydrodynamics. *Physical Review E, 69,* 1–10.
7. Gorshkov, K. A., Ostrovsky, I. A., & Soustova, I. A. (2011). Dynamics of strongly nonlinear solitons in the two-layer fluid. *Studies in Applied Mathematics, 126*(1), 49–73.
8. Gorshkov, K. A., Ostrovsky, L. A., Soustova, I. A., & Shevz, L. M. (2011). The interaction of intense internal waves within the framework of Choi-Camassa equation. *Izvestiya RAN, Atmosphere and Ocean Physics., 47*(3), 339–347.
9. Gorshkov, K. A., Soustova, I. A., Ermoshkin, A. V., & Zaitseva, N. V. (2012). Evolution of the composite soliton of the Gardner equation in media with variable parameter. *Radiophysics and Quantum Electronics, 55*(5), 380–392.
10. Gorshkov, K. A., Soustova, I. A., & Ermoshkin, A. V. (2016). Field structure of a quasisoliton approaching the critical point. *Radiophysics Quantum Electronics, 58*(10), 738–744.
11. Grimshaw, R., Pelinovsky, E., & Talipova, T. (1999). Solitary wave transformation in a medium with sing-variable quadratic nonlinearity and cubic nonlinearity. *Physica D: Nonlinear Phenomena, 132,* 40–62.
12. Gorshkov, K.A., Dolina, I. S., Soustova, I. A., & Troitskaya, Y. I. (2003). Transformation of short waves in a nonuniform flow field on the ocean surface. The effect of wind growth rate modulation. Radiophysics Quantum Electronics *46*(7), 464–485.
13. Hughes, B. (1978). The effect of internal waves on surface wind waves. 2. Theoretical analysis. *Journal of Geophysical Research, 83*(1), 455–465.

14. Kropfli, R. A., Ostrovski, L. A., Stanton, T. P., Skirta, E. A., Keane, A. N., & Irisov, V. (1999). Relationships between strong internal waves in the coastal zone and their radar and radiometric signatures. *Journal of Geophysical Research, 104,* 3133–3148.
15. Lamb, J. L. (1983). *Introduction to the theory of solitons* (p. 295).
16. Morozov, E. G., Paka, V. T., & Bakhanov, V. V. (2008). Strong internal tides in the Kara Gates Strait. *Geophysical Research Letters, 35,* L166603.
17. Nakoulima, O., Zahybo, N., Pelynovsky, E., Talipova, T., Slunyaev, A., & Kurkin, A. (2004). Analytical and numerical studies of the variable-coefficient Gardner equation. *Applied Mathematics and Computation, 152,* 449–471.
18. Serebryany, A. N., & Pao, H. P. (2008). Transition of a nonlinear internal wave through an overturning point on a shelf. *Doklady Earth Science, 420*(4), 714–719.
19. Trevorrow, M. (1998). Observations of internal solitary waves near the Oregon coast with an inverted echo sounder. *Journal of Geophysical Research, 103,* 7671–7680.

Part III
Scientific Contributions: Ocean Circulation

Geostrophic Adjustment Beyond the Traditional Approximation

Gregory M. Reznik

Introduction

In this contribution we discuss geostrophic adjustment in rotating fluid when the angular speed of rotation $\mathbf{\Omega}$ does not coincide in direction with the acceleration due to gravity; the traditional and hydrostatic approximations are not used. Gyroscopic waves (GW) are most susceptible to the "non-traditional" terms in equations of motion and play an important role in our consideration. The GWs exist owing to rotation (e.g., [18]); no stratification or gravity are necessary, although both of these factors strongly affect the structure and properties of these waves. In a "pure" form the GWs occur in a barotropic fluid layer of constant depth bounded by two rigid lids and rotating as a whole at a constant angular speed whose direction can be different from the gravity (see Fig. 1).

Under the traditional approximation (TA) when the horizontal component of the angular speed $\mathbf{\Omega}$ is neglected, the GWs in the barotropic layer are sub-inertial, i.e. their frequencies σ do not exceed the vertical component of twice the angular speed of rotation $f = 2\Omega \sin\varphi_0$ (see Fig. 1) i.e. $\sigma \leq f$; without the TA both sub-inertial and super-inertial GWs with $\sigma \geq f$ are possible [4, 13]. In stably stratified fluid under the TA the sub-inertial GWs exist together with the super-inertial internal waves only if the minimal buoyancy frequency is $N_{\min} < f$ [11]. In strongly stratified fluid, i.e. at $N_{\min} > f$, only super-inertial internal waves are possible. However, without the TA sub-inertial waves (so called internal inertio-gravity waves) occur even in the strongly stratified fluid [6, 7, 13]. Like the GWs these waves cannot exist without rotation.

The paper is organized as follows. In section "Barotropic Model" we examine the geostrophic adjustment of a barotropic fluid layer where GWs are the only

G. M. Reznik (✉)
Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
e-mail: greznikmd@yahoo.com

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_20

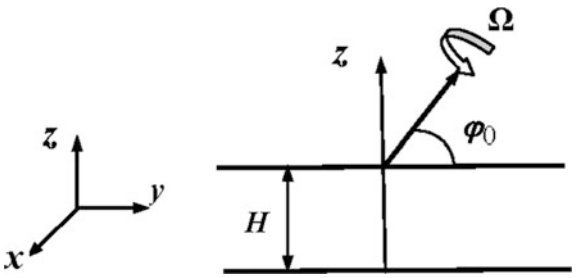


Fig. 1 Schematic representation of the barotropic fluid layer of constant depth H rotating at the angular speed $\boldsymbol{\Omega}$

possible wave motion. The adjustment of stratified fluid where GWs co-exist with internal waves is considered in section "Stably-Neutrally Stratified Fluid" using the model of stably-neutrally stratified (SNS) fluid. The fluid consists of a stratified upper layer with $N > f$ and a homogeneous lower layer, the density and other fields are continuous at the interface between the layers. The configuration is of practical interest since recent observations indicate that, at least in some parts of the World Ocean, there exist practically homogeneous or very weakly stratified (i.e. at $N \leq f$) near bottom layers several hundred meters thick [26, 27].

Both sections "Barotropic Model" and "Stably-Neutrally Stratified Fluid" are organized similarly. In subsections "Governing Equations" the governing equations with boundary and initial conditions are presented. In subsections "Linear Gyroscopic Waves" and "Linear Wave Modes" linear waves are discussed. The linear invariants, geostrophic modes and linear adjustment are examined in subsections "Geostrophic Mode and Linear Adjustment" and "Invariants of Motion and Geostrophic Mode". In subsections "Non-dimensional Equations and Asymptotic Solution" and "Non-dimensional Equations and the Lowest-Order Solution" non-dimensional equations, an asymptotic procedure for finding the solution, and the lowest-order solution are given. The slowly evolving components of motion (quasi-geostrophic (QG) flow and inertial oscillations) are described in subsections "Slow Evolution" and "Slow Evolution of the QG Component and Inertial Oscillations". Final section "Summary and Discussion" contains a discussion and conclusions.

Barotropic Model

Governing Equations

The equations of motion for the barotropic fluid layer represented in Fig. 1 can be written in the form:

$$\mathbf{u}_t + (\mathbf{u} \cdot \nabla)\mathbf{u} + 2\boldsymbol{\Omega} \times \mathbf{u} = -\nabla p/\rho_0, \quad \nabla \cdot \mathbf{u} = 0. \tag{1a, b}$$

Here $\mathbf{u} = (u, v, w)$ where u, v, w are the velocity components associated with the x, y, z axes, respectively, the z-axis is directed upward parallel to gravity (origin at the upper surface); ρ_0 is the fluid density; p the deviation of pressure from the hydrostatic one; $2\boldsymbol{\Omega} = \mathbf{e}_x f_s + \mathbf{e}_z f$ where $\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$ are the unit vectors along corresponding axes and $f = 2\Omega \sin \varphi_0$, $f_s = 2\Omega \cos \varphi_0$.

The velocity field obeys the no-flux conditions at the surface and bottom:

$$w|_{z=0,\,-H} = 0, \tag{2}$$

and the initial conditions

$$(u, v, w)_{t=0} = (u_I, v_I, w_I)(x, y, z); \quad w_I = -\int_{-H}^{z} (\partial_x u_I + \partial_y v_I) dz. \tag{3a, b}$$

Equations (1a, b) represent the so-called non-traditional f-plane approximation; in this case φ_0 is the reference latitude around which the west-east, south-north, and vertical Cartesian coordinates x, y, z are introduced (see Fig. 1). The ratio $q = f_s/f = \cot \varphi_0$ is assumed to be of the order of unity,

$$q = O(1), \tag{4}$$

which corresponds to the mid-latitudes.

Linear Gyroscopic Waves

In the absence of the free surface, beta-effect, and stratification the only wave-generating mechanism is rotation, i.e. only the gyroscopic waves are possible here. Linearized Eq. (1a, b) can be reduced to one equation for the vertical velocity Miropol'sky [19]:

$$(\partial_{tt} + f^2) w_{zz} + \nabla_h^2 w_{tt} + 2 f f_s w_{yz} + f_s^2 w_{yy} = 0, \; w|_{z=0,\,-H} = 0 \tag{5a, b}$$

where $\nabla_h^2 = \partial_x^2 + \partial_y^2$. Any solution to problem (5a, b) can be represented as a superposition of the wave solutions

$$w_n = W_n(z) \exp[\mathrm{i}(kx + ly - \sigma_n t)], \tag{6}$$

where k, l are the horizontal wavenumbers $\left(\kappa = \sqrt{k^2 + l^2}\right)$ and amplitude W_n has the form (see [21] for details):

$$W_n = e^{ia_n z} \sin b_n z,\ n = 1, 2, \ldots. \tag{7}$$

Here n is the number of vertical mode equal to the number of zeros of $W_n(z)$, and

$$a_n = -\frac{ff_s l}{f^2 - \sigma_n^2}, \quad b_n = \frac{\sigma_n \kappa}{|f^2 - \sigma_n^2|}\sqrt{f^2 - \sigma_n^2 + \bar{f}_s^2} = \frac{n\pi}{H}, \quad \bar{f}_s = f_s \frac{|l|}{\kappa}. \tag{8a, b, c}$$

The dispersion relation $\sigma_n = \sigma_n(k, l)$ consists of the sub-inertial branch σ_n^{sub}:

$$\sigma_n^{\text{sub}} = \left\{ f^2 - \frac{1}{2(1+\bar{b}_n^2)}\left[(f^2 - \bar{f}_s^2) + \sqrt{\sigma_0^4 + 4\bar{b}_n^2 f^2 \bar{f}_s^2}\right] \right\}^{1/2}, \tag{9a}$$

and the super-inertial branch σ_n^{sup}:

$$\sigma_n^{\text{sup}} = \left\{ f^2 + \frac{1}{2(1+\bar{b}_n^2)}\left[-(f^2 - \bar{f}_s^2) + \sqrt{\sigma_0^4 + 4\bar{b}_n^2 f^2 \bar{f}_s^2}\right] \right\}^{1/2}, \tag{9b}$$

where $\sigma_0 = \sqrt{f^2 + \bar{f}_s^2}$. The branches are presented in Fig. 2.

In the long-wave approximation $\kappa H \ll 1$ both the sub- and super-inertial frequencies (9a, 9b) are close to the inertial frequency f; in this case

$$\sigma_n^{\text{sub}} = f - \frac{f_s H}{2n\pi}|l| + fO(\bar{b}_n^{-2}), \quad \sigma_n^{\text{sup}} = f + \frac{f_s H}{2n\pi}|l| + fO(\bar{b}_n^{-2}). \tag{10a, b}$$

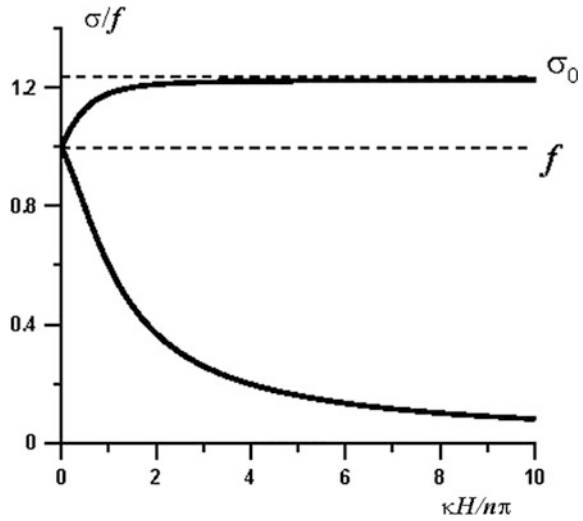


Fig. 2 Dispersion relation for the barotropic gyroscopic waves

The long-wave asymptotics (10a, b) are universal in the sense that they also remain valid in the stratified fluid [7]. We emphasize that the gyroscopic waves are close to the inertial oscillations if $L \gg H$; it is not the case for the surface and internal gravity waves, which are nearly inertial if $L \gg L_R$ where L, L_R are the horizontal and Rossby scales, respectively. Reznik [20] showed that this property of the GWs is also valid in the stratified fluid. Usually $L_R \gg H$, therefore the presence of the GWs results in the existence of inertial oscillations with shorter horizontal scales $L \leq L_R$.

Using the scales L, H, and f^{-1} as the time scale, we write (5a, b) in the non-dimensional form:

$$(\partial_{tt} + 1)w_{zz} + 2\delta q w_{yz} + \delta^2(\nabla_h^2 w_{tt} + q^2 w_{yy}) = 0,\ w|_{z=0,-1} = 0, \tag{11a, b}$$

where $\delta = H/L \ll 1$. The smallness of δ allows a solution to (11a, b) to be sought in the following asymptotic form:

$$w = w_0(x, y, z, t, T_1, \ldots) + \delta w_1(x, y, z, t, T_1, \ldots) + \ldots, \tag{12}$$

where $T_n = \delta^n t, \quad n = 1, 2, \ldots$ are the slow times.

Substitution of (12) into (11a, b) gives in the lowest order:

$$(\partial_{tt} + 1)w_{0zz} = 0,\ w|_{z=0,-1} = 0, \tag{13a, b}$$

Whence we have:

$$w_0 = W_0(x, y, z, T_1, \ldots)\mathrm{e}^{-\mathrm{i}t} + \text{c.c.}; \tag{14}$$

c.c. denotes complex-conjugate value. Thus, the lowest order solution is inertial oscillations modulated by the arbitrary amplitude $W_0(x, y, z, T_1, \ldots)$, which depends on the coordinates and slow time.

The dependence is determined from the first-order equation:

$$(\partial_{tt} + 1)w_{1zz} = -2\partial_{tT_1} w_{0zz} - 2q w_{0yz}. \tag{15}$$

The correction w_1 is bounded in the "fast" time t if the right-hand side of (15) is zero whence one obtains:

$$\partial_{T_1} W_{0zz} + \mathrm{i}q W_{0yz} = 0. \tag{16}$$

Equation (16) should be solved under the boundary and initial conditions

$$W_0|_{z=0,-1} = 0,\ W_0|_{T_1=0} = W_I = \frac{1}{2} w_I(x, y, z). \tag{17a, b}$$

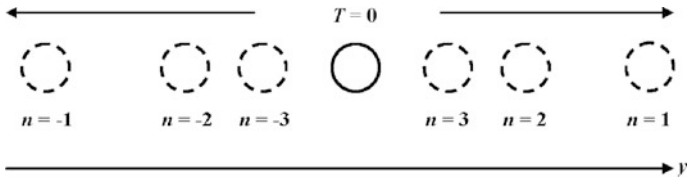


Fig. 3 Schematic representation of dispersion spreading of horizontally localized initial field (solid circle); n denotes the number of corresponding vertical mode (dashed circles)

Representing the derivative W_{0z} as the Fourier series

$$W_{0z} = \sum_{n=-\infty}^{n=\infty} \hat{W}_n(x, y, T_1) \mathrm{e}^{\mathrm{i}2n\pi z}, \tag{18}$$

and substituting (18) into (16) one obtains:

$$\hat{W}_{nT_1} + \frac{q}{2n\pi}\hat{W}_{ny} = 0,\ \hat{W}_n = \hat{W}_n\left(x, y - \frac{q}{2n\pi}T_1\right). \tag{19a, b}$$

Thus, amplitude W_0 is given by the formula:

$$W_0 = \frac{\mathrm{i}}{2\pi}\sum_{n=-\infty}^{n=\infty}\frac{1}{n}\hat{W}_n\left(x, y - \frac{q}{2n\pi}T_1\right)\left(1 - \mathrm{e}^{\mathrm{i}2n\pi z}\right). \tag{20}$$

Solution (20) describes an along-meridional (along the y-axis) dispersive spreading of the perturbation: each vertical mode with number n travels along the y-axis at the group velocity $q/2n\pi$ that, obviously, agrees with the asymptotics (10a, b). The group velocity does not depend on the horizontal wavenumbers k, l; therefore, the modes $\hat{W}_n(1 - \mathrm{e}^{\mathrm{i}2n\pi z})/n$ in the series (20) uniformly translate one after another conserving their shapes (see Fig. 3).

With increasing n the group velocity decreases, i.e. at a fixed point x, y the velocity field has a tendency to become more and more small-scale in the vertical direction. Under the TA ($q = 0$) the meridional dispersion at time $T_1 \sim 1$ disappears, in this case the inertial oscillations disperse in all allowable directions on the longer time $T_2 \sim 1$.

Geostrophic Mode and Linear Adjustment

Geostrophic adjustment is a particular case of the more general wave adjustment [23] which takes place in a physical system possessing in the linear approximation linear invariants and linear wave solutions harmonically depending on time. In the linear approximation Eq. (1a, b) take the form:

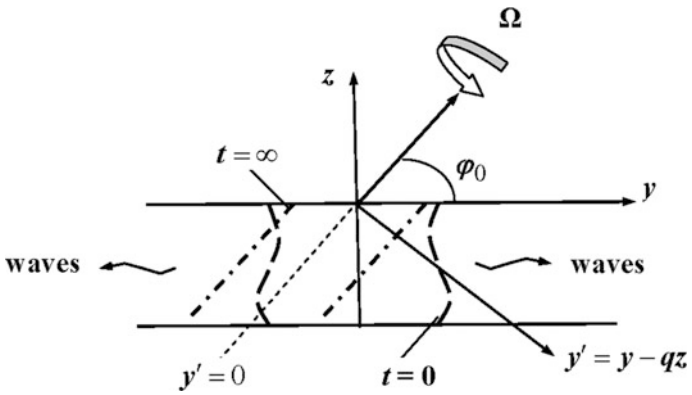


Fig. 4 Coordinates (22) and schematic representation of linear geostrophic adjustment of an initial perturbation (thick long-dashed lines) to a z'-independent vortex state (thick dot-dashed lines) oriented along $\mathbf{\Omega}$

$$u_t - fv + f_s w = -p_x/\rho_0,\ v_t + fu = -p_y/\rho_0, \tag{21a, b}$$

$$w_t - f_s u = -p_z/\rho_0,\ u_x + v_y + w_z = 0. \tag{21c, d}$$

To derive the linear invariant we introduce the new variables:

$$x' = x, \quad y' = y - qz, \quad z' = z. \tag{22}$$

Coordinates (22) are not orthogonal; the planes $y' = y - qz = const$ are parallel to the angular speed $\mathbf{\Omega}$ (see Fig. 4). We emphasize that only the coordinates are transformed, the velocity components are determined by the geometry of the boundaries as before. In the coordinates (22), Eq. (21a, b, 21c, d) are written as

$$u_t - fv + f_s w = -p_{x'}/\rho_0,\ v_t + fu = -p_{y'}/\rho_0, \tag{23a, b}$$

$$w_t - f_s u = -(p_{z'} - qp_{y'})/\rho_0,\ u_{x'} + v_{y'} + w_{z'} - qw_{y'} = 0. \tag{23c, d}$$

Excluding p from (23a, b) and using the continuity Eq. (23d) one obtains the following equation for the vertical vorticity:

$$(v_{x'} - u_{y'})_t = fw_{z'}. \tag{24}$$

Transformation (22) does not change the no-flux condition (2), i.e.

$$w|_{z'=0,\,-H} = 0, \tag{25}$$

therefore, the following conservation integral is obtained from (24):

$$\int_{-H}^{0} (v_{x'} - u_{y'}) dz' = \bar{\Omega}_I^{(z')}(x', y'). \tag{26}$$

The r.h.s. part of (26) is determined by the initial conditions (3a, b):

$$\bar{\Omega}_I^{(z')}(x', y') = \int_{-H}^{0} \Omega_I^{(z)}(x', y' + qz', z') dz', \tag{27}$$

where $\Omega_I^{(z)}$ is the initial vertical vorticity:

$$\Omega_I^{(z)} = \Omega_I^{(z)}(x, y, z) = \partial_x v_I - \partial_y u_I. \tag{28}$$

The gyroscopic wave (6) is a solution to the system (23a, b, 23c, d) with the boundary condition (25) therefore the invariant (26) also exists for the wave. One can readily see, however, that for the wave solution harmonically depending on time with frequency $\sigma > 0$, the invariant is zero (see [23] for details), i.e.

$$\int_{-H}^{0} (\tilde{v}_{x'} - \tilde{u}_{y'}) dz' = 0, \tag{29}$$

where the tilde denotes the wave solution.

This property of waves allows the solution to linear problem (21a, b, 21c, d), (22), (3a, b) to be represented as a sum of a stationary component $\bar{\mathbf{u}}(\mathbf{r}), \bar{p}(\mathbf{r})$ with nonzero conservation integral (26) and a wave component $\tilde{\mathbf{u}}(\mathbf{r}, t), \tilde{p}(\mathbf{r}, t)$ with the zero invariant:

$$(\mathbf{u}, p) = (\bar{\mathbf{u}}, \bar{p}) + (\tilde{\mathbf{u}}, \tilde{p}). \tag{30}$$

The stationary component obeys the equations:

$$-f\bar{v} + f_s\bar{w} = -\bar{p}_{x'}/\rho_0,\ f\bar{u} = -\bar{p}_{y'}/\rho_0, \tag{31a, b}$$

$$f_s\bar{u} = (\bar{p}_{z'} - q\bar{p}_{y'})/\rho_0,\ \bar{u}_{x'} + \bar{v}_{y'} + \bar{w}_{z'} - q\bar{w}_{y'} = 0; \tag{31c, d}$$

and the no-flux conditions (25). The component is a geostrophic mode [8] which does not depend on the depth z on the planes parallel to the angular speed $\mathbf{\Omega}$:

$$\bar{u} = -\frac{1}{f\rho_0}\bar{p}_{y'} \quad \bar{v} = \frac{1}{f\rho_0}\bar{p}_{x'}, \quad \bar{w} = 0, \quad \bar{p}_{z'} = 0. \tag{32}$$

The geostrophic mode is characterized by a columnar motion; the column axes are directed along the rotation speed $\mathbf{\Omega}$ so that the motion is parallel to the rigid boundaries and the vertical velocity is zero (see Fig. 4). Geostrophic pressure $\bar{p}$ is found from (26), (32):

$$\nabla_h^2 \bar{p} = \frac{f\rho_0}{H}\bar{\Omega}_I^{(z')}(x', y'). \tag{33}$$

The wave component of solution obeys Eq. (21a, b, 21c, d):

$$\tilde{u}_t - f\tilde{v} + f_s\tilde{w} = -\tilde{p}_x/\rho_0, \ \tilde{v}_t + f\tilde{u} = -\tilde{p}_y/\rho_0, \tag{34a, b}$$

$$\tilde{w}_t - f_s\tilde{u} = -\tilde{p}_z/\rho_0, \ \tilde{u}_x + \tilde{v}_y + \tilde{w}_z = 0, \tag{34c, d}$$

with boundary conditions (2) and initial conditions

$$(\tilde{u}_I, \tilde{v}_I) = (u_I - \bar{u}, v_I - \bar{v}). \tag{35}$$

In addition, the conservation integral (26) for the wave component is zero, i.e. (29) is valid. Solution to the problem (34a, b, 34c, d) to (35), (29), (2) is a superposition of the gyroscopic waves considered in subsection "Linear Gyroscopic Waves". The waves are dispersive, therefore for localized initial conditions (when $u_I, v_I \to 0$ as $r = \sqrt{x^2 + y^2} \to \infty$) the wave solution $\tilde{\mathbf{u}}, \tilde{p}$ decays with increasing time at a fixed point of space and the full solution (30) tends to the geostrophic mode (32), (33). In other words, any localized initial state tends with time to a geostrophically balanced localized vortex with axis parallel to $\mathbf{\Omega}$ (see Fig. 4). This tendency to columnar motion seems to be very persistent and is observed, for example, in laboratory experiments with turbulence in rotating tanks (see, e.g. [5, 25], and references therein).

Typical time T_w of the wave adjustment can be defined as $T_w = L/c_g$ where L is the typical horizontal scale of initial perturbation and c_g is the typical group velocity of radiated waves. It readily follows from the dispersion relations (9a, 9b), (10a, b) that for the large and moderate scales $L \geq H$ the group velocity $c_g = O(fH)$ and for the small scales $L \ll H$ in the super-inertial (sub-inertial) range $c_g = O(fL^3/H^2)$ ($c_g = O(fL^2/H)$), i.e.

$$T_w = \frac{L}{H}f^{-1} \quad \text{for} \quad \delta = \frac{H}{L} \leq 1, \tag{36a}$$

$$T_w \geq \frac{H}{L}f^{-1} \quad \text{for} \quad \delta = \frac{H}{L} \gg 1. \tag{36b}$$

Thus, the typical time of the wave adjustment is of the order of the inertial time f^{-1} for perturbations of moderate scales with $L \sim H$ and greatly exceeds this time in the large-scale ($L \gg H$) and short-scale ($L \ll H$) domains.

Nonlinear adjustment at small Rossby number $Ro = U/fL \ll 1$ (U is the horizontal velocity scale) results in a slow (as compared to the inertial time f^{-1}) evolution of the geostrophic component on the advective time $T_a = O(1/Rof)$. Scenario of the adjustment depends on the relationship between the typical flow velocity U and group velocity c_g of fast waves (see Reznik [23] for more detail). In the case $T_w \ll T_a$, group velocity c_g greatly exceeds the flow velocity U, i.e. the waves rapidly run away from the initial perturbation and do not interact effectively with the geostrophic mode. The residual flow left behind, after all the waves have been propagated away, slowly changes on the advective time and is close to the geostrophic balance. This scenario is realized for perturbations with moderate scale $L \sim H$ since in this case, $T_w = O(f^{-1}) \ll T_a = O(Ro^{-1} f^{-1})$. For large- and small-scale perturbations, time $T_w \gg f^{-1}$, therefore in these scale domains the waves can effectively interact with the geostrophic mode if $T_w \geq T_a$ and, therefore, $c_g \leq U$. In the rest of paper we examine nonlinear evolution of large-scale perturbations with $L \gg H$ assuming that the advective time T_a and the wave time T_w from (36a) are of the same order. The assumption means that group velocity c_g is of the order of the flow velocity U:

$$c_g = O(fH) \sim U. \tag{37}$$

Non-dimensional Equations and Asymptotic Solution

We now write the system (1a, b) in coordinates (22) and then in non-dimensional form using the scales L, H, f^{-1}, U and the scales of vertical velocity $W = (H/L)U$ and of pressure $P = \rho_0 fUL$ (the primes are omitted):

$$u_t + Ro(uu_x + vu_y + wu_z - \delta qwu_y) - v + \delta qw = -p_x, \tag{38a}$$

$$v_t + Ro(uv_x + vv_y + wv_z - \delta qwv_y) + u = -p_y, \tag{38b}$$

$$\delta^2 w_t + \delta^2 Ro(uw_x + vw_y + ww_z - \delta qww_y) - \delta qu = -p_z + \delta qp_y, \tag{38c}$$

$$u_x + v_y + w_z - \delta qw_y = 0; \tag{38d}$$

in the boundary and initial conditions (2), (3a, b) the depth H is replaced by 1 and x, y, z in (2), (3a, b)—by the variables (22). In terms of the small Rossby number $Ro = U/fL$ and parameter $\delta = H/L$ condition (37) means that

$$Ro = \delta \ll 1. \tag{39}$$

Solution to the problem (38a, 38b, 38c, 38d), (2), (3a, b) is represented in the form of multiple timescale asymptotic expansions analogous to (12):

$$(u, v, w, p) = (u_0, v_0, w_0, p_0)(x, y, z, t, T_1, \ldots) + \delta(u_1, v_1, w_1, p_1) + \ldots. \tag{40}$$

Details of the calculation of successive terms in (40) can be found in Reznik [21]; here we present only the results. On times $T_1 \sim 1$ within to small terms the motion is split in a unique way into slow and fast components:

$$(u, v, w, p) = (\bar{u}, \bar{v}, 0, \bar{p}) + (\tilde{u}, \tilde{v}, \tilde{w}, 0); \quad (\bar{u}, \bar{v}) = \int_{-1}^{0} (u, v) dz, \quad \int_{-1}^{0} (\tilde{u}, \tilde{v}) dz = 0 \tag{41a, b, c}$$

The slow component $(\bar{u}, \bar{v}, 0, \bar{p})$ is not influenced by the fast one $(\tilde{u}, \tilde{v}, \tilde{w}, 0)$ and does not depend on the fluid depth; it is quasigeostrophic (QG) and obeys the QG potential vorticity equation coinciding here with 2D fluid dynamics equation:

$$\bar{u} = \bar{u}(x, y, T_1) = -\psi_y, \quad \bar{v} = \bar{v}(x, y, T_1) = \psi_x, \quad \frac{\partial \nabla_h^2 \Delta \psi}{\partial T_1} + J(\psi, \nabla_h^2 \psi) = 0; \tag{42a, b, c}$$

$\psi = \psi(x, y, T_1)$ is the QG streamfunction.

The fast component $(\tilde{u}, \tilde{v}, \tilde{w}, 0)$ consists of long gyroscopic waves; it is a packet of inertial oscillations modulated by amplitude depending on coordinates and the slow time:

$$\tilde{u} + i\tilde{v} = A(x, y, z, T_1) e^{-it}, \; \tilde{w} = -\frac{1}{2} \mathrm{e}^{-\mathrm{i}t} \int_{-1}^{z} \mathrm{s}(A) dz + \mathrm{c.c.}, \; \mathrm{s} = \partial_x - \mathrm{i}\partial_y. \tag{43a, b, c}$$

Amplitude A obeys the equation with coefficients depending on ψ, i.e. the fast component is coupled to the slow one:

$$\frac{\partial A}{\partial T_1} + J(\psi, A) + \frac{i}{2} \nabla_h^2 \psi A + iq \left(\int_{-1}^{z} A dz + \int_{-1}^{0} z A dz \right)_y = 0, \quad \int_{-1}^{0} A dz = 0. \tag{44a, b}$$

The term proportional q in (44a) is due to the "non-traditional" terms in the equations of motion related to the horizontal component of $\boldsymbol{\Omega}$.

Slow Evolution

Under the TA, i.e. for $q=0$, Eq. (44a) is substantially simplified, especially if the QG streamfunction ψ is axisymmetric and, therefore, it does not depend on time as follows from (42c), i.e. $\psi=\psi(r)$. In this case, the solution to (44a, b) has the form:

$$A=\exp\left(-\frac{\mathrm{i}}{2}\nabla_h^2\psi T_1\right)A_I\left(r,\theta-\frac{\psi'}{r}T_1\right), \tag{45}$$

where the prime means differentiation with respect to r and $A_I(r,\theta)$ is the initial amplitude in (43a) written in polar coordinates. The exponential factor in (45) shifts the inertial frequency f to the so-called effective inertial frequency $f+\nabla_h^2\psi/2$ [17]. Factor $A_I(\ldots)$ describes advection of the inertial oscillations by the QG flow, the radial gradients of the amplitude becoming sharp due to the differential rotation. In accordance with (45) the inertial oscillations are trapped by the QG vortex, (45) displaying no asymmetry between cyclonic and anticyclonic vortices in their "trapping ability" (cf. [17]). This lack of asymmetry is related to the lack of dispersion of the long gyroscopic waves here: under the TA the dispersion becomes significant on longer times $T_2\sim 1$. It is seen from (45) that the magnitude $|A|$ behaves exactly as a passive scalar in the steady QG flow. For $q=0$ the same is valid for any ψ since by virtue of (44a) we have:

$$|A|_{T_1}+\mathrm{J}(\psi,|A|)=0, \tag{46}$$

Equation (46) means that under the TA the inertial oscillations are trapped by the QG velocity field $\bar{u},\bar{v}$.

At $q\neq 0$ the "non-traditional" term in (44a) changes the situation. In the absence of the slow component, i.e. for $\psi=0$, (44a, b) is similar to (16) in section "Linear Gyroscopic Waves" (one can readily see this by differentiating (44a) with respect to z, setting $\psi=0$ and applying operator (43c) to the resulting equation). Therefore, the non-traditional term in (44a) produces a tendency for the meridional (along the y-axis) propagation of the inertial oscillations. To analyze the general case $\psi\neq 0$, $q\neq 0$ we represent the solution to (44a, b) in the form analogous to (18):

$$A=\sum_{n=-\infty}^{n=\infty}\hat{A}_n(x,y,T_1,\ldots)\mathrm{e}^{\mathrm{i}2n\pi z}. \tag{47}$$

The equation for the Fourier amplitude $\hat{A}_n$ is written as

$$\hat{A}_{nT_1}+\mathrm{J}(\psi,\hat{A}_n)+\frac{\mathrm{i}}{2}\nabla_h^2\psi\hat{A}_n+\frac{q}{2n\pi}\hat{A}_{ny}=0. \tag{48}$$

The equation for the module $|\hat{A}_n|$ analogous to (46) simply follows from (48):

$$|\hat{A}_n|_{T_1} + \mathrm{J}(\bar{\psi}_n, |\hat{A}_n|) = 0. \tag{49}$$

Here $\bar{\psi}_n$ is the sum of ψ and a superimposed constant meridional flow $\frac{q}{2n\pi}x$:

$$\bar{\psi}_n = \psi + \frac{q}{2n\pi}x. \tag{50}$$

Thus, the field $|\hat{A}_n|$ behaves as a passive scalar in the velocity field $\bar{u}_n = -\bar{\psi}_{ny}$, $\bar{v}_n = \bar{\psi}_{nx}$. Let the QG component ψ contain intense vortices with closed streamlines, in this case the streamline field (50) consists of the closed streamlines related to the vortices, and unclosed ones, each of the unclosed streamlines tending to the straight line $\bar{\psi}_n = \frac{q}{2n\pi}x = const$ as $y \to \pm\infty$. If the QG flow is time-independent then the module $|\hat{A}_n|$ is trapped in the domains with the closed streamlines and travels away from the initial disturbance location along the unclosed streamlines. The "propagation ability" depends on the mutual strength of the field ψ and the superimposed flow $\frac{q}{2n\pi}x$ and decreases with increasing n. In the case of time-dependent QG flow the situation is more complicated since Lagrangian trajectories do not coincide with the streamlines. However, one can assume that the time-dependent ψ, at least, does not reduce the "propagation ability" of the inertial oscillations since in this case the Lagrangian trajectory can escape from the closed streamlines (e.g., [1]).

An analog of Eq. (44a) in stratified fluid was derived by Young and Ben Jelloul [28], and analyzed by Balmforth et al. [2], Balmforth and Young [3], Klein and Llewellyn-Smith [14], Klein et al. [15]; the QG flow in these works was assumed to be prescribed. In the context of geostrophic adjustment an analog of (44a, b) was derived by Reznik et al. [24] (for barotropic shallow water with free surface) and by Zeitlin et al. [29] (for stratified fluid). In all these works, the TA was used and the inertial oscillations were the long gravity (surface or internal) waves with horizontal scales greatly exceeding the corresponding Rossby scales.

It readily follows from (42c) and (44a, b) that

$$\frac{\partial}{\partial T_1}\int dxdydz|A|^2 = 0, \quad \frac{\partial}{\partial T_1}\int dxdy(\nabla\psi)^2 = 0, \tag{51a, b}$$

i.e. on times $t \sim 1/\delta$ both the fast and the slow components conserve their total energies. However, energy transfer between the inertial oscillations and QG flow is possible on longer times $t \sim 1/\delta^2$. This follows from the "refined" QG equation valid on times of the order of $1/\delta^2$ [21]:

$$\nabla_h^2\psi_{T_1} + \mathrm{J}(\psi, \nabla_h^2\psi) + \delta[\mathrm{G}(\psi, A) + q\mathrm{H}(A)] = 0, \tag{52}$$

where

$$\mathrm{G}(\psi, A) = \frac{1}{2}\left(\mathrm{M}(\psi)\overline{|A|^2}\right)_{xy} - \frac{1}{2}\mathrm{M}\left(\psi_{xy}\overline{|A|^2}\right), \tag{53a}$$

$$\mathrm{H}(A) = -\frac{\mathrm{i}}{4}\mathrm{s}\left[\int_{-1}^{0} dz A \int_{-1}^{z} \mathrm{s}^*(A^*)dz\right]_y + \text{c.c.}, \tag{53b}$$

operator $\mathrm{M} = \partial_{xx} - \partial_{yy}$, $\bar{a} = \int_{-1}^{0} a dz$, and the asterisk denotes the complex conjugation.

If all fields decay at infinity, then we have

$$\int \psi \mathrm{G}(\psi, A) dxdy = 0, \tag{54}$$

therefore, energy $\bar{E}$ of the QG component changes in time as

$$\partial_T \bar{E} = \delta q \int \psi \mathrm{H}(A) dxdy, \quad \bar{E} = \frac{1}{2}\int (\nabla_h \psi)^2 dxdy. \tag{55a, b}$$

For $q \neq 0$ the right-hand side of (55a), is, generally, non-zero whence the important conclusion follows that without the TA a transfer of energy between the QG component and inertial oscillations can exist.

Stably-Neutrally Stratified Fluid

Governing Equations

We consider a stably-neutrally stratified fluid of constant depth H, bounded by two rigid lids and rotating as a whole at a constant angular speed $\mathbf{\Omega}$, which, generally, is not parallel to gravity (directed along the z-axis in Fig. 5). The fluid density ρ, being continuous, depends on z in the upper layer of depth $h_1 - \eta$ and is constant in the lower layer of depth $h_2 + \eta$ where $h_1, h_2 = H - h_1$ are constant mean depths of the

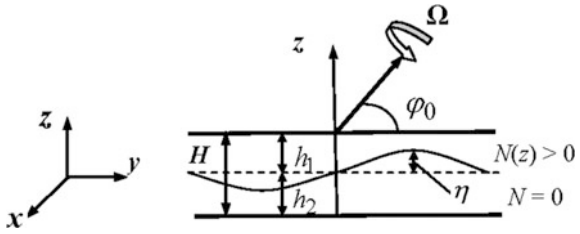


Fig. 5 Schematic representation of rotating stably-neutrally stratified fluid

layers, and $\eta = \eta(x, y, t)$ is the perturbation of interface between the layers (see Fig. 5). Obviously, the condition $N_{\min} < f$ of coexistence of internal and gyroscopic waves is satisfied here.

Density ρ and pressure p in the layers are given by the formulae:

$$\rho = \begin{cases} \rho_s(z) + \rho', & 0 \geq z \geq -h_1 + \eta \\ \rho_0 = const, & -h_1 + \eta \geq z \geq -H \end{cases}; \tag{56a}$$

$$p = \begin{cases} g \int\limits_z^0 \rho_s dz + p', & 0 \geq z \geq -h_1 + \eta \\ -g\rho_0(z + h_1) + g \int\limits_{h_1}^0 \rho_s dz + p', & -h_1 + \eta \geq z \geq -H \end{cases}, \tag{56b}$$

where $\rho_s = \rho_s(z)$ is the upper layer density at the rest state, the equilibrium density is continuous, i.e. $\rho_s(-h_1) = \rho_0$; ρ', p' are the variations of density and pressure from their hydrostatic profiles.

Motion of the fluid obeys the following equations:

$$\mathbf{u}_t + \mathbf{u} \cdot \nabla \mathbf{u} + 2\boldsymbol{\Omega} \times \mathbf{u} + \mathbf{e}_z g\rho/\rho_0 = -\nabla p'/\rho_0, \tag{57a}$$

$$\rho_t + \mathbf{u} \cdot \nabla \rho - \rho_0 N^2 w/g = 0, \ \nabla \cdot \mathbf{u} = 0 \tag{57b, c}$$

in the domain $0 \geq z \geq -h_1 + \eta$, and

$$\mathbf{u}_t + \mathbf{u} \cdot \nabla \mathbf{u} + 2\boldsymbol{\Omega} \times \mathbf{u} = -\nabla p'/\rho_0, \quad \nabla \cdot \mathbf{u} = 0 \tag{58a, b}$$

in the domain $-h_1 + \eta \geq z \geq -H$. Here, g is the acceleration due to gravity and $N^2 = -g\partial_z \rho_s/\rho_0$. The prime in density variations is omitted and (assuming the variations are small) the density is replaced by the constant value ρ_0 where it is not differentiated.

The no-flux conditions at the rigid surface and bottom and the initial conditions coincide with (2) and (3a, b). Conditions at the interface $z = -h_1 + \eta$ are discussed in detail in Reznik [22]. We are interested in a regime when the density and velocity fields are continuous at the interface at the initial moment and remain continuous all the time. The continuity of fields prevents possible Kelvin-Helmholtz instability and makes it possible to study the "buffer" zone between the stably stratified and homogeneous domains, in which a non-stationary boundary layer can arise [20]. The density is conserved in the fluid elements so they cannot intersect the interface, which, therefore, is a material surface, at which the conditions

$$(\rho_s + \rho)_{z = -h_1 + \eta} = \rho_0, \ w|_{z = -h_1 + \eta} = \eta_t + u\eta_x + v\eta_y \tag{59a, b}$$

should be satisfied. The continuity of all fields means that in addition to (59a, b) the horizontal velocity and pressure are also continuous at $z = -h_1 + \eta$:

$$[u]_{z=-h_1+\eta}=0, \quad [v]_{z=-h_1+\eta}=0, \quad [p]_{z=-h_1+\eta}=0; \tag{59c, d, e}$$

here and below $[a]_{z=z_0}=a|_{z=z_0+0}-a|_{z=z_0-0}$.

If the buoyancy frequency $N(z)$ is not zero at $z=-h_1$, i.e.

$$N(-h_1)\neq 0, \tag{60}$$

then the interface is a weak discontinuity [16], since the physical fields here are continuous but their gradients are discontinuous. In some calculations below we will use the simplest configuration with weak discontinuity when the background upper layer density profile $\rho_s(z)$ linearly depends on z, and N is a constant, i.e.

$$\rho_s=-\frac{\rho_0}{g}N^2(z+h_1)+\rho_0. \tag{61}$$

Linear Wave Modes

The linearized version of (56a, 56b), (57a, 57b, c) can be reduced to two equations for the vertical velocity w (e.g., Miropol'sky [19]; cf. (5a)):

$$(\partial_{tt}+f^2)w^+_{zz}+\nabla_h^2 w^+_{tt}+2ff_s w^+_{yz}+f_s^2 w^+_{yy}+N^2\nabla_h^2 w^+=0, \tag{62a}$$

$$(\partial_{tt}+f^2)w^-_{zz}+\nabla_h^2 w^-_{tt}+2ff_s w^-_{yz}+f_s^2 w^-_{yy}=0. \tag{62b}$$

The boundary conditions for (62a, 62b) follow from (59a, 59c, d, e) and the continuity Eq. (57c), (58b):

$$w^+|_{z=0}=w^-|_{z=-H}=0, \quad [w]=[w_z]=0. \tag{63a, b}$$

Here and below, superscripts $+$ and $-$ denote quantities in the upper and lower layers, respectively; $[a]=a^+|_{z=-h_1}-a^-|_{z=-h_1}$.

The wave solutions

$$w^{\pm}=W^{\pm}(z)\exp[\mathrm{i}(kx+ly-\sigma t)]+c.c,\ \sigma>0; \tag{64}$$

to (62a, 62b), (63a, b) were examined in detail in Reznik [20]. The wave spectrum consists of gyroscopic, internal, and internal inertio-gravity waves with non-zero frequency. Here we are interested in the physically important case when the stratification is strong, $f/N_0\ll 1$ (N_0 is the characteristic buoyancy frequency), and the waves are long,

$$H \ll L \leq L_R. \tag{65}$$

Here $L_R = HN_0/f$ is the Rossby scale. Neglecting small terms in (62a, 62b) gives the following approximate equations:

$$(\partial_{tt} + f^2)w^+_{zz} + N^2\nabla_h^2 w^+ = 0, \quad (\partial_{tt} + f^2)w^-_{zz} = 0. \tag{66a, b}$$

The wave spectrum of system (66a, b), (63a, b) consists of super-inertial internal waves (64) with $\sigma > f$ and inertial oscillations with $\sigma = f$. Amplitudes $W^{\pm}(z)$ of the internal wave obey the equations

$$W^+_{zz} + b^2 W^+ = 0, \quad W^-_{zz} = 0, \; b^2 = \frac{\kappa^2 N^2}{\sigma^2 - f^2}, \tag{67a, b, c}$$

where $\kappa = \sqrt{k^2 + l^2}$. In the case (61) of constant N, the solution to (67a, b) satisfying (63a,b) is readily found:

$$W^+_n = -\sin b_n z / \sin b_n h_1, \quad W^-_n = (z + H)/h_2; \; n = 1, 2, \ldots \tag{68}$$

Here $b_n = s_n/h_1$, where s_n is the n-th root of the equation

$$s \cot s = -h_1/h_2. \tag{69}$$

The corresponding dispersion relation has the form:

$$\sigma_n = \sigma_n^{iw} = \sqrt{f^2 + \kappa^2 h_1^2 N^2/s_n^2}. \tag{70}$$

In accordance with (68) the vertical mode amplitude oscillates in the upper stratified layer and depends linearly on depth in the lower one.

For the inertial oscillations with $\sigma = f$ it follows from (66a, b) that

$$w^+ = 0, \; w^- = A_s(x, y, z)\sin ft + A_c(x, y, z)\cos ft; \tag{71}$$

here the amplitudes $A_{s,c}$ are arbitrary functions obeying the conditions:

$$A_{s,c}|_{z=-h_1} = \partial_z A_{s,c}|_{z=-h_1} = A_{s,c}|_{z=-H} = 0. \tag{72}$$

Thus, the inertial oscillations here are confined to the homogeneous lower layer and do not penetrate into the upper one.

The non-hydrostatic and "non-traditional" terms neglected in (62a, 62b) to derive (66a, b) are of no importance for the internal waves but they affect the inertial oscillations (71) (see [20] for more details). The terms transform the oscillations into sub-inertial long gyroscopic waves of "standard" form (64) with the approximate dispersion relation:

$$\sigma_n = \sigma_n^{gw} = f - \frac{f_s h_2}{2n\pi}|l|,\ n = 1, 2, \ldots \tag{73}$$

coinciding with the dispersion relation of sub-inertial long gyroscopic waves in barotropic fluid of depth h_2 (cf. (10a)). In these waves the vertical mode amplitudes W_n oscillate in the lower layer and decay exponentially in the upper one with increasing distance from the interface. As follows from (70) and (73), in the scale range (65) the internal waves are strongly dispersive, whereas the gyroscopic ones are characterized by weak dispersion.

Invariants of Motion and Geostrophic Mode

We now write the linearized problem (57a, 57b, c), (58a, b) in terms of the variables (22) (cf. (23a, b, 23c, d)):

$$u_t^{\pm} - fv^{\pm} + f_s w^{\pm} = -p_{x'}^{\pm},\ v_t^{\pm} + fu^{\pm} = -p_{y'}^{\pm}, \tag{74a, b}$$

$$w_t^{\pm} - f_s u^{\pm} + g\rho^{\pm}/\rho_0 = -p_{z'}^{\pm} + qp_{y'}^{\pm},\ \rho_t^{+} - (\rho_0 N^2/g)w^{+} = 0, \tag{74c, d}$$

$$u_{x'}^{\pm} + v_{y'}^{\pm} + w_{z'}^{\pm} - qw_{y'}^{\pm} = 0. \tag{74e}$$

Here and below, $\rho^{-} = 0$, $p^{\pm} = p'^{\pm}/\rho_0$.

Elimination of the pressure $p^{\pm}$ from (74a, b) gives the vorticity equation:

$$\left(v_{x'}^{\pm} - u_{y'}^{\pm}\right)_t - fw_{z'}^{\pm} = 0. \tag{75}$$

It follows from (75) and (74d) that in the stratified layer the potential vorticity (PV) Π^{+} conserves:

$$\Pi^{+} = v_{x'}^{+} - u_{y'}^{+} - (fg/\rho_0)(\rho^{+}/N^2)_{z'} = \Pi_I^{+}(x', y', z'). \tag{76}$$

To find an analogous invariant in the homogeneous layer we use the boundary condition at interface (59b), which in the linear approximation can be written as

$$w|_{z'=-h_1} = \eta_t. \tag{77}$$

Integration of (75) over z' from $-H$ to $-h_1$ taking into account (77) and (2) gives the PV in the homogeneous layer (cf. (26)):

$$\Pi^- = \frac{1}{h_2}\int_{-H}^{-h_1} (v_{x'}^- - u_{y'}^-)dz' - \frac{f}{h_2}\eta = \Pi_I^-(x', y'). \tag{78}$$

The values of $\Pi_I^{\pm}$ of the invariants are determined by the initial fields (3a, b).

The third invariant is the surface density $\rho^+|_{z'=0}$; in view of (74d) and the no-flux condition (2) we have:

$$\rho^+|_{z'=0} = \rho_I^+|_{z'=0}. \tag{79}$$

The solution to the problem (74a, b, 74c, d, 74e) is represented as a sum of a stationary geostrophic mode $(\mathbf{u}_g, \rho_g, p_g)$ with non-zero invariants $\Pi^{\pm}$, $\rho^+|_{z'=0}$ and an ageostrophic wave part $(\mathbf{u}_a, \rho_a, p_a)$ consisting of the harmonic waves (64) with non-zero frequencies and the zero invariants. One can readily show, using the bottom no-flux condition from (2), that in the stationary solution the vertical velocity is zero and the lower layer motion does not depend on z':

$$u_g^{\pm} = -(1/f)\partial_{y'}p_g^{\pm}, \quad v_g^{\pm} = (1/f)\partial_{x'}p_g^{\pm}, \quad w_g^{\pm} = 0, \tag{80a, b, c}$$

$$\rho_g^{\pm} = -(\rho_0/g)\partial_{z'}p_g^{\pm}, \quad \partial_{z'}p_g^- = 0. \tag{80d, e}$$

The geostrophic pressure $p_g^{\pm}$ is determined using (76), (78):

$$\nabla_h'^2 p_g^+ + f^2(\partial_{z'}p_g^+/N^2)_{z'} = f\Pi_I^+(x', y', z'), \tag{81a}$$

$$\nabla_h'^2 p_g^- - (f^2/h_2)\eta_g = f\Pi_I^-(x', y'), \tag{81b}$$

where $\nabla_h'^2 = \partial_{x'x'} + \partial_{y'y'}$. The boundary condition at the surface $z' = 0$ follows from (79) and (80d):

$$\partial_{z'}p_g^+\Big|_{z'=0} = -(g/\rho_0)\rho_I^+|_{z'=0}. \tag{82a}$$

The boundary condition for p_g^+ at the interface is discussed in Reznik [22] and can be written in the form suitable both for the continuous $(N^2(-h_1) = 0)$ and discontinuous $(N^2(-h_1) \neq 0)$ buoyancy frequency profiles:

$$\lim_{z' \to -h_1}\left[\left(\partial_{z'}p_g^+/N^2\right)_{z'} - \partial_{z'}p_g^+/(h_2N^2)\right] = \frac{1}{f}(\Pi_I^+ - \Pi_I^-)_{z'=-h_1}. \tag{82b}$$

Knowing p_g^+ one can determine p_g^- by continuity of pressure, $p_g^- = p_g^+\Big|_{z=-h_1}$, and then η_g from (81b).

It is seen from (80a, b, c, 80d, e) that the motion in the geostrophic mode occurs in the planes parallel to the rigid boundaries; the motion in the homogeneous layer is columnar with columns elongated parallel to the angular rotation speed $\mathbf{\Omega}$. In the upper stratified layer the motion is more complicated; its structure is not columnar and depends on the initial horizontal velocity and density.

Under the TA (when $q=0$) the geostrophic mode is described by (80a, b, c, 80d, e), (81a, 81b) with x, y, z instead of x', y', z'. This means that the non-traditional terms are of importance in QG dynamics if the dominating horizontal scale L of initial perturbation is smaller or of the order of the fluid depth:

$$L \leq H. \tag{83}$$

For long-wave perturbation with

$$L \gg H \tag{84}$$

the contribution of the term qz in (22) is small, i.e. the non-traditional terms have a weak effect on the long-wave geostrophic mode.

The ageostrophic wave component $(\mathbf{u}_a, \rho_a, p_a)$ obeys the same Eq. (74a, b, 74c, d, 74e) but the invariants (76), (78) and the surface density $\rho_a(x', y', 0)$ are zero. The wave component is a superposition of harmonic waves considered in Reznik [20]. The waves are dispersive, therefore for localized initial conditions (when $u_I, v_I \to 0$ as $r = \sqrt{x^2 + y^2} \to \infty$) the wave part decays with increasing time at a fixed point of space and the full solution tends to the above stationary geostrophic mode. Thus, in the lower barotropic layer any localized initial state tends with time to a geostrophically balanced vortex state with axis parallel to $\mathbf{\Omega}$, exactly as in the purely barotropic case (see subsection "Geostrophic Mode and Linear Adjustment").

Similar to the barotropic case, the nonlinear adjustment at small Rossby number Ro results in a slow evolution of the geostrophic component on the advective time $T_a = O(1/Rof)$. In the rest of the paper, we examine the nonlinear evolution of large-scale perturbations with $H \ll L \leq L_R$. As was shown in subsection "Linear Wave Modes", in this range the wave spectrum consists of internal waves and the gyroscopic ones, which are close to inertial oscillations. In view of (70) and (73) the corresponding group velocities c_g^{iw} and c_g^{gw} of the internal and gyroscopic waves are $O(fL_R)$ and $O(fH)$, respectively. For L in the range (65) and small Rossby number this means that the group velocity of internal waves greatly exceeds both the flow velocity U and the group velocity of gyroscopic waves, and the internal waves only weakly interact with the slow geostrophic component. At the same time, the interaction between the gyroscopic waves and geostrophic flow is much more effective, especially if $c_g^{gw} \leq U$. In what follows we assume that the group velocity c_g^{gw} is of the order of the flow velocity U (cf. (37)):

$$c_g^{gw} = O(fH) \sim U. \tag{85}$$

Non-dimensional Equations and the Lowest-Order Solution

Non-dimensional Equations

To write system (57a, 57b, c), (58a, b), (59a, b, 59c, d, e) in non-dimensional form we use the scales from subsection "Non-dimensional Equations and Asymptotic Solution" with $L = L_R$, and the scales of density variations and interface perturbations $R = \rho_0 Ro(N_0^2 H/g)$ and $Z = RoH$. In the vector form the non-dimensional equations are written as:

$$\hat{\mathbf{u}}_t + Ro\hat{\mathbf{u}} \cdot \hat{\nabla}\hat{\mathbf{u}} + 2\hat{\Omega} \times \hat{\mathbf{u}} + \mathbf{e}_z \rho/\delta = -\hat{\nabla} p, \tag{86a}$$

$$\rho_t + Ro\hat{\mathbf{u}} \cdot \hat{\nabla}\rho - N^2 w = 0, \ \hat{\nabla} \cdot \hat{\mathbf{u}} = 0 \tag{86b, c}$$

in the domain $0 \geq z \geq -h_1 + Ro\eta$; and

$$\hat{\mathbf{u}}_t + Ro\hat{\mathbf{u}} \cdot \hat{\nabla}\hat{\mathbf{u}} + 2\hat{\Omega} \times \hat{\mathbf{u}} = -\hat{\nabla} p, \quad \hat{\nabla} \cdot \hat{\mathbf{u}} = 0 \tag{87a, b}$$

in the domain $-h_1 + Ro\eta \geq z \geq -1$. Here $\hat{\mathbf{u}} = (u, v, \delta w)$, $\hat{\nabla} = (\partial_x, \partial_y, \partial_z/\delta)$, $2\hat{\Omega} = \mathbf{e}_y q + \mathbf{e}_z$, $\delta = H/L = f/N_0$; here and below the notations for non-dimensional h_1, h_2 and N remain unchanged. In the boundary and initial conditions (2), (59a, b, 59c, d, e) and (3a, b), written in non-dimensional form the depth H is replaced by 1 and the interface surface becomes $z = -h_1 + Ro\eta$.

We are interested in the case when the motion is large-scale and the rotation is fast i.e. both the aspect ratio δ and the Rossby number Ro are small parameters. Condition (85) means that (cf. (39))

$$\delta = Ro \ll 1. \tag{88}$$

Similar to section "Barotropic Model" the solution is represented in the asymptotic form (cf. (40)):

$$(u, v, w, p, \rho) = (u_0, v_0, w_0, p_0, \rho_0)(x, y, z, t, T_1, \ldots) + \delta(u_1, v_1, w_1, p_1, \rho_1) + \cdots \tag{89}$$

Substitution of (89) into (86a, 86b, c), (87a, b) gives in the lowest order:

$$\partial_t u_0^{\pm} - v_0^{\pm} = -\partial_x p_0^{\pm}, \ \partial_t v_0^{\pm} + u_0^{\pm} = -\partial_y p_0^{\pm}, \ \rho_0^{\pm} = -\partial_z p_0^{\pm} \tag{90a, b, c}$$

$$\partial_t \rho_0^{\pm} - N_{\pm}^2 w_0^{\pm} = 0, \ \partial_x u_0^{\pm} + \partial_y v_0^{\pm} + \partial_z w_0^{\pm} = 0; \tag{90d, e}$$

$$w_0^+\big|_{z=0} = w_0^-\big|_{z=-1} = 0, \tag{91a}$$

$$w_0^\pm\big|_{z=-h_1} = \partial_t \eta_0, \quad [u_0, v_0, p_0] = 0, \tag{91b, c}$$

$$(u_0^\pm, v_0^\pm, \rho_0^\pm)_{t=0} = (u_I^\pm, v_I^\pm, \rho_I^\pm)(x, y, z). \tag{91d}$$

Geostrophic Component

System (90a, b, c, 90d, e), up to constant coefficients, is a simplification of system (74a, b, 74c, d, 74e) at $f_s = q = 0$ and the analysis in subsection "Invariants of Motion and Geostrophic Mode" is directly applied to (90a, b, c, 90d, e). The invariants (76), (78), and (79) are now non-dimensional and take the form:

$$\Pi^+ = \partial_x v_0^+ - \partial_y u_0^+ - (\rho_0^+ / N^2)_z = \Pi_I^+(x, y, z), \tag{92a}$$

$$\Pi^- = \frac{1}{h_2}\int_{-1}^{-h_1} (\partial_x v_0^- - \partial_y u_0^-)dz - \frac{\eta_0}{h_2} = \Pi_I^-(x, y), \tag{92b}$$

$$\rho_0^+\big|_{z=0} = \rho_I^+(x, y, 0). \tag{92c}$$

Exactly as in subsection "Invariants of Motion and Geostrophic Mode" the lowest-order solution is represented as the sum of a geostrophic part (coinciding with (80a, b, c, 80d, e) *mutatis mutandis*) and an ageostrophic component with the zero invariants (92a, 92b, 92c). For simplicity of notations the geostrophic and ageostrophic components will be denoted here by the subscripts "g" and "a" (without the subscript "0"). The quasi-geostrophic PV (81a, 81b) take the form:

$$\Pi^+ = \nabla_h^2 p_g^+ + (\partial_z p_g^+ / N^2)_z = \Pi_I^+(x, y, z), \tag{93a}$$

$$\Pi^- = \nabla_h^2 p_g^- - \eta_g / h_2 = \Pi_I^-(x, y). \tag{93b}$$

Equations (93a, 93b) should be solved with the non-dimensional boundary conditions

$$\partial_z p_g^+\Big|_{z=0} = -\rho_g^+\Big|_{z=0} = -\rho_I^+(x, y, 0), \tag{94a}$$

$$\lim_{z \to -h_1}\left[\left(\partial_z p_g^+ / N^2\right)_z - \partial_z p_g^+ / (h_2 N^2)\right] = \Pi^+\big|_{z=-h_1} - \Pi^-, \tag{94b}$$

which follow from (82a, 82b).

Ageostrophic Component

The ageostrophic components in the layers obey Eq. (90a, b, c, 90d, e) i.e.:

$$\partial_t u_a^{\pm} - v_a^{\pm} = -\partial_x p_a^{\pm},\ \partial_t v_a^{\pm} + u_a^{\pm} = -\partial_y p_a^{\pm},\ \rho_a^{\pm} = -\partial_z p_a^{\pm}, \tag{95a, b, c}$$

$$\partial_t \rho_a^{\pm} - N_{\pm}^2 w_a^{\pm} = 0,\ \partial_x u_a^{\pm} + \partial_y v_a^{\pm} + \partial_z w_a^{\pm} = 0,\ \rho_a^{-} = 0. \tag{95d, e, f}$$

The boundary conditions for the ageostrophic quantities are the same as (91a, 91b, c) and the initial state is determined after calculating the geostrophic fields (see subsection "Geostrophic Component"):

$$(u_a, v_a, \rho_a)_{t=0} = (u_{aI}, v_{aI}, \rho_{aI}) = (u_I - u_g, v_I - v_g, \rho_I - \rho_g). \tag{96}$$

In addition, the ageostrophic fields are imposed by the restrictions that the PV in the layers are zero:

$$\partial_x v_a^{+} - \partial_y u_a^{+} - (\rho_a^{+}/N^2)_z = 0,\ \frac{1}{h_2}\int_{-1}^{-h_1} (\partial_x v_a^{-} - \partial_y u_a^{-})dz - \frac{\eta_a}{h_2} = 0, \tag{97a, b}$$

and the agestrophic density at $z = 0$ is zero in view of (94a):

$$\rho_a^{+}\big|_{z=0} = 0. \tag{98}$$

Some Properties of the Ageostrophic Solution

We now discuss some general properties of the ageostrophic component. Using (95a, b, c, 95d, e, f), (97a, b), and (98) one can show that the vertically integrated ageostrophic horizontal velocities and pressure are zero:

$$\int_{-1}^{0} (u_a, v_a, p_a)dz = 0. \tag{99}$$

The lower layer pressure p_a^{-} does not depend on z, therefore:

$$p_a^{-} = -\frac{1}{h_2}\int_{-h_1}^{0} p_a^{+}\, dz. \tag{100}$$

Continuity of the pressure at the interface $z=-h_1$ gives the relation:

$$p_a^+\big|_{z=-h_1}+\frac{1}{h_2}\int\limits_{-h_1}^{0}p_a^+\,dz=0. \tag{101}$$

It readily follows from (95a, b) and (101) that

$$\partial_t\hat{u}_a^+-\hat{v}_a^+=0,\ \partial_t\hat{v}_a^++\hat{u}_a^+=0, \tag{102}$$

where

$$(\hat{u}_a^+,\hat{v}_a^+)=(u_a^+,v_a^+)\big|_{z=-h_1}+\frac{1}{h_2}\int\limits_{-h_1}^{0}(u_a^+,v_a^+)dz. \tag{103}$$

Solution to (102) is readily written:

$$\hat{u}_a^++\mathrm{i}\hat{v}_a^+=(\hat{u}_{aI}^++\mathrm{i}\hat{v}_{aI}^+)e^{-\mathrm{i}t}. \tag{104}$$

An important issue is that the upper stratified layer does not contain the inertial oscillations $\sim\sin t,\cos t$ i.e.

$$u_{as}^+=v_{as}^+=u_{ac}^+=v_{ac}^+=0; \tag{105}$$

and similarly for other fields [22]. Here,

$$g_{s,c}=\lim_{T\to\infty}\frac{2}{T}\int\limits_0^T(\sin\,t,\cos\,t)g(t)dt. \tag{106}$$

In view of (105) the integral in (103) also does not contain the inertial oscillations; therefore, as readily follows from (103), (104), the horizontal velocities at the interface $(u_a^+,v_a^+)\big|_{z=-h_1}$ *contain* the inertial oscillations (104), i.e. (105) is valid at $z>-h_1$ and is *not valid* at $z=-h_1$. Such a solution structure is typical for a boundary layer, see also Reznik [20] when at large times the velocity u_a^+ (for example) is represented in the form

$$u_a^+=C_s[x,y,(z+h_1)t]\sin\,t+C_c[x,y,(z+h_1)t]\cos\,t. \tag{107}$$

in a close vicinity of the interface at $z=-h_1$.

If functions $C_{s,c}$ tend to zero as $t\to\infty$ at any fixed $z>-h_1$, but, for example, $C_s(x,y,0)\neq0$, then $u_{as}^+=0$ at $z>-h_1$ and $u_{as}^+=C_s(x,y,0)\neq0$ at $z=-h_1$. We note that in the particular case

$$\hat{u}_{aI}^{+} = \hat{v}_{aI}^{+} = 0, \tag{108}$$

relations (105) are valid everywhere in the upper layer $z \geq -h_1$ and the boundary layer in the vicinity of interface does not arise. We emphasize that details of the buoyancy frequency profile $N(z)$ are unimportant in the above consideration therefore one can expect the boundary layer to exist for any upper layer stratification, i.e. for both smooth and discontinuous profiles of $N(z)$.

Motion in the Stratified Upper Layer

In the upper layer, Eq. (95a, b, c, 95d, e, f) can be reduced to one equation for the vertical velocity (e.g., Miropol'sky, [19]:

$$(\partial_{tt} + 1)\partial_{zz} w_a^{+} + N^2 \nabla_h^2 w_a^{+} = 0, \tag{109}$$

which should be solved under the initial conditions:

$$(w_a^{+}, \partial_t w_a^{+})_{t=0} = (w_I^{+}, \dot{w}_I^{+})(x, y, z), \tag{110a}$$

the no-flux boundary condition:

$$w_a^{+}\big|_{z=0} = 0, \tag{110b}$$

and the boundary condition at $z = -h_1$, which simply follows from (103), (95e):

$$\mathrm{m}(w_a^{+})\big|_{z=-h_1} = -(\partial_x \hat{u}_a^{+} + \partial_y \hat{v}_a^{+}), \tag{110c}$$

where $\mathrm{m} = \partial_z - 1/h_2$. The initial fields w_I^{+} and $\dot{w}_I^{+}$ can be expressed in terms of the initial fields $u_{aI}^{+}, v_{aI}^{+}, \rho_{aI}^{+}$ [20]. Using (104), (102) one can represent (110c) in the form:

$$\mathrm{m}(w_a^{+})\big|_{z=-h_1} = [\mathrm{m}(w_I^{+})\cos t + \mathrm{m}(\dot{w}_I^{+})\sin t]_{z=-h_1}. \tag{111}$$

It is convenient to represent all variables in (109) to (111) in the form of Fourier-integrals, for example

$$w_a^{+} = \frac{1}{2\pi} \int \tilde{w}_a^{+}(k, l, z, t) e^{\mathrm{i}(kx+ly)} dk dl; \tag{112}$$

and similarly for the other values. Here and below, the tilde denotes the Fourier-amplitude of the corresponding variable. The amplitude $\tilde{w}_a^{+} = \tilde{w}_a^{+}(k, l, z, t)$

is sought as an expansion in the vertical modes, which here are the eigenfunctions of the problem:

$$W_{zz} + b^2 N^2 W = 0, \quad W|_{z=0} = 0, \quad \mathrm{m}(W)|_{z=-h_1} = 0. \tag{113}$$

Eigenfunctions W_n, eigenvalues b_n, and the corresponding wave frequencies σ_n are readily found in the case $N = \mathrm{const}$ (cf. (68)–(70)):

$$W_n = \sin b_n z, \quad b_n = s_n/h_1, \; \sigma_n = \sqrt{1 + \kappa^2/b_n^2}, \; n = 1, 2, \ldots \tag{114}$$

here s_n is the n-th root of (69).

Amplitude $\tilde{w}_a^+$ can be written as (see Reznik [22] for details):

$$\tilde{w}_a^+ = \sum_{n=1}^{\infty} [\tilde{w}_{In}^+ \cos \sigma_n t + (\tilde{\dot{w}}_{In}^+ / \sigma_n) \sin \sigma_n t] W_n(z). \tag{115}$$

We see that the upper layer ageostrophic component is the superposition of long super-inertial internal waves $\sim \exp[\mathrm{i}(kx + ly - \sigma_n t)]$ considered in subsection "Linear Wave Modes". The function w_a^+ does not contain the inertial oscillations (cf. subsection "Some Properties of the Ageostrophic Solution") and at the interface any finite partial sum of the series (115) satisfies boundary conditions (111) with zero r.h.s. This means that in the very close vicinity of the interface $z = -h_1$ the wave modes with very large numbers n play an important role. Since $b_n \to \infty$, $\sigma_n \to 1$ for $n \to \infty$, vertical scales and frequencies of these modes are close to zero and to the inertial frequency, respectively. Joint effect of these modes forms the near interface boundary layer considered below in subsection "Boundary Layer".

Motion in the Homogeneous Lower Layer

It readily follows from (95a, b, c, 95f) that in the lower layer:

$$\partial_z(\partial_t u_a^- - v_a^-) = 0, \; \partial_z(\partial_t v_a^- + u_a^-) = 0, \tag{116a, b}$$

whence

$$U_a^- = U_{aI}^-(x, y, z) e^{-\mathrm{i}t} + \overline{U}(x, y, t). \tag{117}$$

Here, $\overline{U} = \bar{u} + i\bar{v}$ is still unknown depth-independent complex velocity and

$$U_a^- = u_a^- + \mathrm{i}v_a^-, \; U_{aI}^- = u_{aI}^- + \mathrm{i}v_{aI}^-. \tag{118a, b}$$

Integrating (117) over z from –1 to $-h_1$ and using (99), (105) one finds:

$$\overline{U} = -\frac{e^{-it}}{h_2}\int_{-1}^{-h_1} U_{aI}^{-}\,dz + \overline{\overline{U}},\ \overline{\overline{U}} = -\frac{1}{h_2}\int_{-h_1}^{0} U_a^{+}\,dz,\quad \overline{\overline{U}}_{s,c} = 0. \qquad (119a, b, c)$$

The "non-inertial" depth-independent velocity $\overline{\overline{U}}(x, y, t)$ is induced by the upper-layer internal waves and can be calculated from known vertical velocity w_a^+, which is given by (112), (115).

In view of (117), (119a, b, c) the horizontal velocity U_a^- can be written as:

$$U_a^- = A(x, y, z)e^{-it} + \overline{\overline{U}},\ A = A_I = U_{aI}^- - \frac{1}{h_2}\int_{-1}^{-h_1} U_{aI}^-\,dz. \qquad (120a, b)$$

The velocity components are given by the formulae:

$$(u_a^-, v_a^-) = \left(\frac{1}{2}, \frac{-i}{2}\right)A(x, y, z)e^{-it} + c.c. + (\overline{\overline{u}}, \overline{\overline{v}})(x, y, t), \qquad (121a)$$

$$w_a^- = -\frac{1}{2}e^{-it}\int_{-1}^{z} s(A)dz + c.c. - (z+1)(\overline{\overline{u}}_x + \overline{\overline{v}}_y),\quad s = \partial_x - i\partial_y. \qquad (121b)$$

It follows from (120b) that

$$\int_{-1}^{-h_1} A\,dz = 0. \qquad (122)$$

Thus, in the homogeneous layer the ageostrophic motion is the sum of inertial oscillations and a field induced by super-inertial internal waves. Condition (122) provides non-penetration of the inertial signal into the stratified layer in the vertical velocity field. If the amplitude of inertial oscillations A_I in (120a, b) is zero at the interface $z = -h_1$:

$$A_I(x, y, -h_1) = 0, \qquad (123)$$

then, non-penetration is also provided for the horizontal velocity. One can readily show that (123) is equivalent to the condition (108) of absence of the boundary layer. Obviously, (123) and, therefore, (108) are satisfied only for particular initial velocity fields; if this is not the case a non-stationary boundary layer develops near the interface.

Boundary Layer

We now consider the case of general initial conditions when (123) is not valid and the inertial signal in the horizontal velocity is non-zero at the interface $z=-h_1$. As discussed above, in this case the solution in the domain $z\geq -h_1$ has a boundary layer structure in the vicinity of the interface at large times. Representation (115), (112) of the solution is poorly suitable for description of such regimes since any finite partial sum of the series (115) obeys boundary condition (111) with the zero r.h.s.

To describe the boundary layer dynamics we introduce two new variables:

$$\bar{w}_s=\frac{1}{t}\int_0^t w_a^+\,\sin tdt,\quad \bar{w}_c=\frac{1}{t}\int_0^t w_a^+\,\cos tdt \tag{124}$$

(cf. [9, 10, 12]). The meaning of the variables is that the impact of not near-inertial harmonics to $\bar{w}_{s,c}$ becomes negligible at large times $t\gg 1$ as seen from (115).

From (109), (110b), and (111) we find:

$$(t\bar{w})_{zztt}+2\mathrm{i}(t\bar{w})_{zzt}+t\nabla_h^2\bar{w}=\partial_{zz}(-w_I^+ +\mathrm{i}\dot{w}_I^+),\quad \bar{w}=\bar{w}_s+\mathrm{i}\bar{w}_c; \tag{125a, b}$$

$$\bar{w}|_{z=0}=0,\ \mathrm{m}(\bar{w})|_{z=-h_1}=0.5\mathrm{m}(\dot{w}_I^+ +\mathrm{i}w_I^+)_{z=-h_1}. \tag{126a, b}$$

Boundary condition (126b) is written up to small terms of the order of $1/t$. When writing (125a, b) we assume for simplicity N to be constant i.e. $N=1$.

Outside the boundary layers $\bar{w}_z\sim 1$, therefore an approximate solution $\bar{w}_0$ satisfying (125a) in this domain at large t is determined as

$$\bar{w}_0=S(x,y,z)/t,\ \nabla_h^2 S=\partial_{zz}(-w_I^+ +\mathrm{i}\dot{w}_I^+)=R \tag{127a, b}$$

Obviously, $\bar{w}_0$ satisfies neither (126a) (since, generally, $R|_{z=0}\neq 0$) nor (126b), therefore in the vicinities of the boundaries $z=0,\ -h_1$, narrow boundary layers arise, in which $\partial_z\gg 1$. In the boundary layer near the interface the leading order solution is sought in the form $\bar{w}=\bar{w}_b=t^\alpha\hat{w}_b(x,y,\xi)$ where $\xi=(z+h_1)t^\beta$ is the boundary layer stretched coordinate. Parameters α and β are determined as follows. First, in virtue of (126b), $\partial_z\bar{w}_b=t^{\alpha+\beta}\partial_\xi\hat{w}_b\sim 1$, and, second, maxima of the terms with derivatives with respect to z that were neglected outside the boundary layer, and the third term on the l. h.s. of (125a) should be of the same order, i.e. $t^{2\beta-1}\sim 1$. As a result, we have $\beta=-\alpha=1/2$ i.e. in the boundary layer

$$\bar{w}=\bar{w}_b=\frac{1}{\sqrt{t}}\hat{w}_b(x,y,\xi),\quad \xi=(z+h_1)\sqrt{t}. \tag{128}$$

Substituting (128) into (125a) and neglecting small terms we have:

$$\xi\partial_{\xi\xi\xi}\hat{w}_b + 3\partial_{\xi\xi}\hat{w}_b - \mathrm{i}\nabla_h^2\hat{w}_b = 0. \tag{129}$$

In terms of the Fourier amplitude $\tilde{\hat{w}}_b$ (see (112)) (129) is rewritten as:

$$\xi\partial_{\xi\xi\xi}\tilde{\hat{w}}_b + 3\partial_{\xi\xi}\tilde{\hat{w}}_b + \mathrm{i}\kappa^2\tilde{\hat{w}}_b = 0. \tag{130}$$

For the boundary layer to exist the solution $\tilde{\hat{w}}_0$ to (130) satisfying the conditions

$$\tilde{\hat{w}}_0 \to 0 \text{ as } \xi \to \infty; \quad \partial_\xi\tilde{\hat{w}}_0\big|_{\xi=0} = 1, \tag{131a, b}$$

must exist. Analysis in Reznik [22] confirms possibility of such a solution. The corresponding Fourier amplitude $\tilde{\tilde{w}}$ is given by:

$$\tilde{\tilde{w}}_b = C(k,l)\tilde{\hat{w}}_0(k,l,\xi)/\sqrt{t},\ C = 0.5\mathrm{m}(\dot{w}_I^+ + \mathrm{i}w_I^+)_{z=-h_1}. \tag{132a, b}$$

The boundary layer near the surface $z=0$ is similar to that near the interface [22].

Slow Evolution of the QG Component and Inertial Oscillations

The lowest-order solution constructed in subsection "Non-dimensional Equations and the Lowest-Order Solution" is the sum of a time-independent geostrophic component, lower-layer inertial oscillations, and dispersive internal waves; the non-stationary boundary layers are the result of the joint impact of the internal waves with very short vertical lengths. To derive the solution of the lowest-order system (90a, b, c, 90d, e), (91a, 91b, c, 91d) depending on slow times one should "allow" parameters related to the initial fields to depend on the slow times. The geostrophic part of the solution is determined by the PV $\Pi^\pm$ in the layers (see (93a, 93b)) directly related to the initial fields u_I, v_I, ρ_I. In what follows we assume that:

$$\Pi^\pm = \Pi^\pm(x,y,z,T_1,T_2,\ldots), \quad \Pi^\pm(x,y,z,0,0,\ldots) = \Pi_I^\pm. \tag{133a, b}$$

Similarly, for the internal waves the initial fields $(w_I^+, \dot{w}_I^+)(x,y,z)$ are replaced by the functions $(w_s^+, \dot{w}_s^+)(x,y,z,T_1,T_2,\ldots)$ with

$$(w_s^+, \dot{w}_s^+)(x,y,z,0,0,\ldots) = (w_I^+, \dot{w}_I^+)(x,y,z). \tag{134}$$

Finally, in the formulae (121a, 121b), (122) describing the lower layer inertial oscillations we put:

$$A = A(x, y, z, T_1, T_2, \ldots), \quad A(x, y, z, 0, 0, \ldots) = A_I(x, y, z). \tag{135a, b}$$

Slow evolution of the fields (133a, b) to (135a, b) is determined from condition of boundedness of higher approximations. In the rest of the paper we discuss dependence on the slow time T_1, which is obtained from analyses of the first approximation. The calculations are rather cumbersome and details can be found in Reznik [22]; here we represent only the results.

Assuming all fields to decay at infinity as $r = \sqrt{x^2 + y^2} \to \infty$ one can show that contribution of the dispersive internal waves tends to zero with increasing time and the resulting flow is a sum of a slowly changing QG component and inertial oscillations confined to the homogeneous layer. The QG part is governed by the pair of coupled equations of conservation PV in the layers:

$$\Pi^{\pm}_{T_1} + J(p_g^{\pm}, \Pi^{\pm}) = 0; \tag{136}$$

$$\Pi^{+} = \nabla_h^2 p_g^{+} + (\partial_z p_g^{+}/N^2)_z, \; \Pi^{-} = \nabla_h^2 p_g^{-} - \eta_g/h_2. \tag{137a, b}$$

The QG lower layer pressure p_g^- is related to p_g^+ by the continuity at the interface:

$$p_g^- = p_g^+\Big|_{z=-h_1}. \tag{138}$$

Boundary condition (94a) for p_g^+ at $z = 0$ can be used only to calculate the initial geostrophic pressure; on times $t \sim 1/\delta$ one should take into account the slow evolution of density, which is unknown in advance. The correct condition is

$$\partial_{zT_1} p_g^{+} + J(p_g^{+}, \partial_z p_g^{+}) = 0 \text{ at } z = 0 \tag{139}$$

(see [22].)

The complete set of equations describing the slow evolution of the QG component on times $t \sim 1/\delta$ includes PV-Eq. (136), (137a, b), the boundary conditions (139), (94b), (138), and the initial conditions (133b).

The equation governing slow evolution of the inertial oscillation amplitude is very similar to the corresponding Eq. (44a) of the barotropic model:

$$A_{T_1} + J(p_g^-, A) + \frac{i}{2}\nabla_h^2 p_g^- A + iq\left(\int_{-1}^{z} A dz + \frac{1}{h_2}\int_{-1}^{-h_1} zA dz\right)_y = 0. \tag{140}$$

In addition to (140), amplitude A should satisfy condition (122), which prevents the inertial signal from penetration into the stratified upper layer.

Discussion of the Slow Evolution

The algorithm for calculation of the QG component is as follows. Knowing $p_g^+(T_1)$ one determines from (137a, b) the PVs $\Pi^+(T_1)$, $\Pi^-(T_1)$, then from (136), (139) the fields $\Pi^+(T_1+\Delta T)$, $\Pi^-(T_1+\Delta T)$, $\partial_z p_g^+(T_1+\Delta T)\Big|_{z=0}$. Knowing Π^+, Π^- at the step $T_1+\Delta T$ one can calculate the right-hand side in the boundary condition (94b), and determine the field $p_g^+(T_1+\Delta T)$ from (137a) and the "new" boundary conditions for p_g^+ at $z=0, -h_1$.

The QG component conserves its energy; it can be shown [22] that:

$$E = E^+ + E^- = const, \tag{141a}$$

$$E^+ = \frac{1}{2}\int dxdy \int_{-h_1}^{0} dz\left[(\nabla_h p_g^+)^2 + \frac{(\partial_z p_g^+)^2}{N^2}\right],\ E^- = \frac{h_2}{2}\int dxdy(\nabla_h p_g^-)^2. \tag{141b, c}$$

Here, E is the full QG energy, E^+ and E^- are the energies of the upper and lower layers, respectively.

Let the QG motion in the lower layer be absent at some time $T_1 = T_I$, i.e.

$$p_g^- = p_g^+\Big|_{z=-h_1} = 0. \tag{142}$$

Generally, for times $T_1 > T_I$ the QG pressure p_g^- becomes non-zero i.e. the QG energy transfers into the lower layer. To show this we assume that (142) is satisfied at all times. In this case the quantities $\Pi^+|_{z=-h_1}$, Π^- do not depend on time by virtue of (136), and we have from (93a) an additional boundary condition for p_g^+ at $z=-h_1$:

$$\lim_{z\to -h_1}(\partial_z p_g^+/N^2)_z = \Pi_I^+\Big|_{z=-h_1}. \tag{143}$$

Obviously, the problem (136) for Π^+, (142), (143), (94b) and (139) is overdetermined and hence (142) cannot be valid at all times.

Equation (140) for the amplitude of inertial oscillations almost exactly coincides with the corresponding equation for the barotropic case (44a). The last term in (140) arises due to the non-zero horizontal component of the Earth's rotation. Under the TA $q=0$ and the inertial oscillations are trapped by the QG component as in the barotropic case (see subsection "Slow Evolution"). The "non-traditional" term results in a meridional dispersion of the inertial oscillations and in doing so it provides an effective energy radiation from the initial perturbation domain (see section "Barotropic Model").

Similarly to the barotropic case the slow QG component does not depend on the fast ageostrophic waves on times $\sim 1/\delta$, whereas evolution of the inertial oscillations depends on the geostrophic streamfunction p_g^-. At the same time, it follows from (140) and (122) that

$$\frac{\partial}{\partial T_1}\int dxdy \int_{-1}^{-h_1} dz|A|^2 = 0, \tag{144}$$

i.e., the total energy of inertial oscillations is conserved along with that of the QG component. We note, however, that the conservation takes place on times $\sim 1/\delta$, on longer times the "non-traditional" terms in the equations of motion can, in principle, give rise to energy exchange between the components as it takes place in the barotropic fluid (see section "Barotropic Model"). If this is the case, the energy of QG motion in the lower layer can be transferred to the inertial oscillations. Of course, the mechanism of dissipation of QG energy in a homogeneous layer is highly speculative and it would be useful to verify it numerically using a non-hydrostatic model without the TA.

It was shown in subsection "Non-dimensional Equations and the Lowest-Order Solution" that condition (122) forbids the inertial signal in the vertical velocity field to penetrate into the stratified layer. Equation (140) "supports" limitation (122): it is readily to show integrating (140) over z from –1 to $-h_1$ that if (122) is satisfied at some moment T_0 then it is valid for all times $T > T_0$. If the initial conditions do not satisfy (108) and, therefore, (123), then the analogous screening in horizontal velocity is provided by a non-stationary boundary layer developing in the upper layer near the interface. Generally, condition (123) is not supported by (140) because of the "non-traditional" term in (140), i.e., the quantity $A|_{z=-h_1}$ ceases to be zero even if (123) is satisfied at some moment. This means that without the TA the near interface boundary layer develops at any initial conditions.

Summary and Discussion

We have examined geostrophic adjustment in a rotating fluid of constant depth confined between two rigid lids. The angular speed of rotation $\mathbf{\Omega}$ does not coincide in direction with the gravity; the traditional and hydrostatic approximations are not used. Two models were considered: the barotropic one and the SNS fluid consisting of a stratified upper layer with $N \gg f$ and a homogeneous lower layer, the density and other fields being continuous at the interface between the layers. In both the cases, the wave spectra contain the gyroscopic waves which exist due to rotation and are susceptible to the non-traditional terms in equations of motion.

Geostrophic adjustment is a particular case of more general wave adjustment [23], which takes place in a physical system possessing in the linear approximation linear invariants and linear wave solutions harmonically depending on time. In our

models, the linear invariants are potential vorticities written in special non-orthogonal coordinates related to the vector $\mathbf{\Omega}$; the invariants are determined by the initial conditions in a unique way. The steady state related to the invariants is a geostrophic mode, which is a geostrophically balanced motion parallel to the layer boundaries. In the barotropic fluid, the motion is columnar, the columns are parallel to $\mathbf{\Omega}$. In the SNS fluid, the same is valid in the homogeneous layer; in the stratified layer, in the general case, the geostrophic mode is not columnar. The non-traditional terms are of importance in the QG dynamics if the dominating horizontal scale L of the initial perturbation is smaller or of the order of the fluid depth, $L \leq H$, for the long-wave perturbation with $L \gg H$, the contribution of the terms is small.

In the barotropic model the gyroscopic waves are the only possible wave motions; in the SNS fluid the wave spectrum consists of internal and gyroscopic waves (see [20] for details). All the waves are dispersive; therefore, in the process of linear geostrophic adjustment the motion tends to the geostrophic mode with increasing time.

Using multiple-time-scale perturbation theory we studied the non-linear adjustment in the long-wave approximation $H \ll L \leq L_R$ and small Rossby numbers Ro. In the barotropic case, $L_R = \infty$; in the SNS fluid, $L_R = HN_0/f$. In this scale range, the gyroscopic waves are close to weakly dispersive inertial oscillations (confined to the lower layer in the SNS fluid). The internal waves in the SNS fluid are strongly dispersive and penetrate into the homogeneous lower layer down to the bottom.

The general scenario of the adjustment is similar to the one with gravity waves (cf. [24, 29]): an arbitrary perturbation is split in a unique way into slow and fast components evolving with characteristic time scales $(Rof)^{-1}$ and f^{-1}, respectively. In both cases the slow component is close to the geostrophic balance and is not influenced by the fast one on times $t \sim (fRo)^{-1}$. In the barotropic model the slow component does not depend on depth and is described by the 2D fluid dynamics equation for the geostrophic streamfunction. In the SNS fluid, the slow component is governed by two coupled nonlinear equations of conservation of QG potential vorticity in the upper and lower layers. The upper and lower layer QG flows are not independent: if at some moment the QG motion in the lower homogeneous layer vanishes, then at subsequent times the QG energy is transferred from the upper layer to the lower one.

The fast component consists of the inertial oscillations modulated by the amplitude depending on coordinates and the slow time, and (in the SNS fluid) the internal waves. The inertial oscillations are long gyroscopic waves; the depth-integrated horizontal flow induced by the oscillations is zero. Typical group speed of the internal waves greatly exceeds the typical horizontal slow velocity U; at the same time, the group speed of the inertial oscillation is of the order of U. As a result, in the course of non-linear geostrophic adjustment the internal waves decay because of dispersion, and the residual flow consists of the QG slow component and inertial oscillations.

At times $t \sim (fRo)^{-1}$ the energy of inertial oscillations is conserved but they are coupled to the slow component: their amplitude obeys an equation with coefficients depending on the geostrophic streamfunction. Under the TA the inertial oscillations

are trapped by the QG component; dispersion of the inertial oscillations packet occurs on times $t \sim (fRo^2)^{-1}$. Without the TA the "non-traditional" terms in the amplitude equation result in a meridional dispersion of the inertial oscillations on much shorter times $t \sim (fRo)^{-1}$ and in doing so the terms provide an effective energy radiation from the initial perturbation domain.

Another feasible effect of the "non-traditional" terms is energy exchange between the slow QG component and inertial oscillations in the homogeneous layer at times $O(f^{-1}Ro^{-2})$. The possibility of such exchange was demonstrated in the barotropic case. In the case of SNS fluid, one can speculate that the QG energy is transferred from the stratified layer to the homogeneous one and then to the fast inertial oscillations. It would be useful to elucidate the existence and efficiency of this mechanism using a non-hydrostatic numerical model without the TA.

Inertial oscillations with horizontal scale $L \leq L_R$ cannot exist in the stratified upper layer. To prevent their penetration from the lower layer to the upper one, a non-stationary boundary layer develops in the upper layer near the interface at large times. The boundary layer is a result of joint impact of internal modes with large vertical wavenumbers whose frequencies are close to the inertial frequency *f*. Thus the near interface domain is characterized by large vertical gradients of the horizontal velocities that can result in strong mixing and instability here.

In geostrophic adjustment with gravity waves (surface or internal) the inertial oscillations arise only if the dominating scale L of the initial perturbation exceeds the corresponding Rossby scale L_R, i.e., $L \gg L_R$ (cf. [24, 29]). In the presence of gyroscopic waves the "shorter" inertial oscillations with the scales $H \ll L \leq L_R$ are possible. The significant vertical velocities of the near-inertial oscillations observed by van Haren and Millot [27] in the practically barotropic deep Western Mediterranean Sea can be related to this property of the gyroscopic waves. We note that some other regions of the deep ocean are also characterized by very weak stratification, as for example, the Canada Basin in the Arctic Ocean [26], or the Pacific Ocean near 179° E [6].

Acknowledgements This work was supported by the Russian Science Foundation grant no. 14-50-00095 (section "Barotropic Model"), the Ministry of Education and Science of Russian Federation grant no. 14.W03.31.0006 (section "Stably-Neutrally Stratified Fluid"), and the Russian Foundation for Basic Research grant no. 17-05-00094 (analysis of slow evolution).

References

1. Aref, H. (1984). Stirring by chaotic advection. *Journal of Fluid Mechanics, 143,* 1–21.
2. Balmforth, N. J., Llewellyn Smith, S. G., & Young, W. R. (1998). Enhanced dispersion of near-inertial waves in an idealized geostrophic flow. *Journal of Marine Research, 56,* 1–40.
3. Balmforth, N. J., & Young, W. R. (1999). Radiative damping of near-inertial oscillations in the mixed layer. *Journal of Marine Research, 57,* 561–584.
4. Brekhovskikh, L. M., & Goncharov, V. (1994). *Mechanics of continua and wave dynamics*. Berlin: Springer.

5. Davidson, P. A., Staplehurst, P. J., & Dalziel, S. B. (2006). On the evolution of eddies in a rapidly rotating system. *Journal of Fluid Mechanics, 557,* 135–144.
6. Gerkema, T., Zimmerman, J. T. F., Maas, L. R. M., & van Haren, H. (2008). Geophysical and astrophysical fluid dynamics beyond the traditional approximation. *Reviews of Geophysics*, *46*, RG2004.
7. Gerkema, T., & Shrira, V. I. (2005). Near-inertial waves in the ocean: Beyond the traditional approximation. *Journal of Fluid Mechanics, 529,* 195–219.
8. Greenspan, H. P. (1968). *Theory of rotating fluids*. Cambridge University Press.
9. Il'in, A. M. (1970). Asymptotic properties of a solution of a boundary-value problem. *Mathematical Notes*, *8*, 625–632.
10. Il'in, A. M. (1972). On the behaviour of the solution of a boundary-value problem when $t \to \infty$. *Mathematics of the USSR-Sbornik*, *16*, 545–572.
11. Kamenkovich, V. M. (1977). *Fundamentals of ocean dynamics*. Amsterdam: Elsevier.
12. Kamenkovich, V. M., & Kamenkovich, I. V. (1993). On the evolution of Rossby waves, generated by wind stress in a closed basin, incorporating total mass conservation. *Dynamics of Atmospheres and Oceans, 18,* 67–103.
13. Kasahara, A. (2003). The roles of the horizontal component of the Earth's angular velocity in nonhydrostatic linear models. *Journal of the Atmospheric Sciences, 60,* 1085–1095.
14. Klein, P., & Llewellyn-Smith, S. (2001). Horizontal dispersion of near-inertial oscillations in a turbulent mesoscale eddy field. *Journal of Marine Research, 59,* 697–723.
15. Klein, P., Llewellyn-Smith, S., & Lapeyre, G. (2004). Organization of near-inertial energy by an eddy field. *Quarterly Journal of the Royal Meteorological Society, 130,* 1153–1166.
16. Kochin, N. E., Kibel, L. A., & Rose, N. V. (1964). *Theoretical hydromechanics*. New York: Wiley-Interscience.
17. Kunze, E. (1985). Near-inertial wave propagation in geostrophic shear. *Journal of Physical Oceanography, 15,* 544–565.
18. LeBlond, P. H., & Mysak, L. A. (1978). *Waves in the ocean*. Amsterdam: Elsevier.
19. Miropol'sky, Y. Z. (2001). *Dynamics of internal gravity waves in the ocean*. Dordrecht: Kluwer.
20. Reznik, G. M. (2013). Linear dynamics of a stably-neutrally stratified ocean. *Journal of Marine Research, 71,* 253–288.
21. Reznik, G. M. (2014). Geostrophic adjustment with gyroscopic waves: Barotropic fluid without traditional approximation. *Journal of Fluid Mechanics, 743,* 585–605.
22. Reznik, G. M. (2014). Geostrophic adjustment with gyroscopic waves: stably-neutrally stratified fluid without the traditional approximation. *Journal of Fluid Mechanics, 747,* 605–634.
23. Reznik, G. M. (2015). Wave adjustment: general concept and examples. *Journal of Fluid Mechanics, 779,* 514–543.
24. Reznik, G. M., Zeitlin, V., & Ben Jelloul, M. (2001). Nonlinear theory of geostrophic adjustment. Part 1 rotating shallow-water model. *Journal of Fluid Mechanics, 445,* 93–120.
25. Staplehurst, P. J., Davidson, P. A., & Dalziel, S. B. (2008). Structure formation in homogeneous freely decaying turbulence. *Journal of Fluid Mechanics, 598,* 81–105.
26. Timmermans, M. L., Melling, H., & Rainville, L. (2007). Dynamics in the Deep Canada Basin, Arctic Ocean, inferred by thermistor chain time series. *Journal of Physical Oceanography, 37,* 1066–1076.
27. van Haren, H., & Millot, C. (2005). Gyroscopic waves in the Mediterranean Sea. *Geophysical Research Letters, 32,* L24614.
28. Young, W. R., & Ben Jelloul, M. (1997). Propagation of near-inertial oscillations through a geostrophic flow. *Journal of Marine Research, 55,* 735–766.
29. Zeitlin, V., Reznik, G. M., & Ben Jelloul, M. (2003). Nonlinear theory of geostrophic adjustment. Part 2. Two-layer and continuously stratified primitive equations. *Journal of Fluid Mechanics, 491,* 207–228.

Evolution of an Intrathermocline Lens over the Lofoten Basin

Boris N. Filyushkin, Mikhail A. Sokolovskiy and Konstantin V. Lebedev

Introduction

Systematic oceanographic surveys conducted by the Arctic and Antarctic Research Institute during the period 1970–1990 in the region of the Lofoten Basin (LB) of the Norwegian Sea proved the existence of a quasistationary anticyclonic vortex (AV) [2]. Six consecutive oceanographic surveys occupied in 1987–1988 have shown the existence of a synoptic-scale (diameter was about 60 km) vortex (a lens) and instrumentally confirmed its anti-cyclonic character. At the same time, during a year, the AV have been drifting almost along a closed cyclonic orbit (with a horizontal latitude scale about 130 km) while its core was still located in the deepest part of the LB (the isobath 3200 m) [6, 7] and Fig. 1 of this contribution.

The development of the Argo project in the first years of the 21st century in addition to the neutral buoyancy floats observations started in the beginning of the 1990s, provides opportunities for studying the structure and circulation of water in the LB. New observations confirmed the existence of a quasi-permanent AV and the fact that the cyclonic circulation dominates in the depth range 600–1000 m within the deepest part of the LB [5, 12, 14, 22]. Detailed measurements performed by the RAFOS floats and Doppler current velocity meters revealed a complex pattern of the formation and existence of a large number of mesoscale vortices that differ in their lifetime and scale. At the same time, the number of vortices in the LB fluctuated throughout the year [19]. Similar results were also obtained from satellite observations of variations in the sea surface height (SSH) and the sea surface temperature [8, 15, 23]. In these publications, satellite images based on weekly

B. N. Filyushkin (✉) · M. A. Sokolovskiy · K. V. Lebedev
Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
e-mail: borisfil@yandex.ru

M. A. Sokolovskiy
Water Problems Institute, Russian Academy of Sciences, Moscow, Russia

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_21

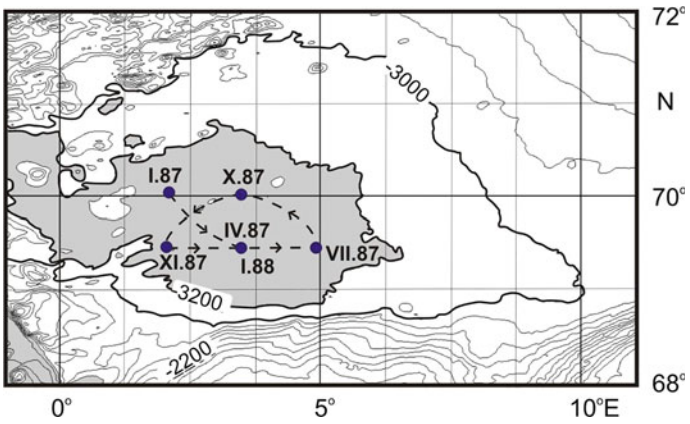


Fig. 1 Map of the LB bottom topography. The dashed line shows the trajectory of the anticyclonic vortex (lens) and the dots represent the lens location at the times of measurements [7]. The area with the depths deeper than 3200 is shown in gray color

altimeter data have shown the location of mesoscale long-lived vortices of different rotation signs that moved along the cyclonic orbit. At specific stages of the motion, they merged with an anticyclonic vortex located at the center of the deepest part of the LB (3200 m) at 69° 30′ N, 3° E [15]. Currently, the daily data show that the previously recorded long-living vortices represented a cluster of separate short-living eddies, and they better demonstrate the mean eddy drift [16].

The Lofoten Basin is the main reservoir of heat in the Polar seas (PS) [1, 2, 15, 23]. However, the heat content of the central AV is less than the total heat content of the entire LB waters. This means that warm Atlantic waters should submerge [2, 8] in the upper layers in the center of the deepest part of the LB, and at the same time, a deep topographic cyclonic water rotation should exist [5, 14, 15, 22]. In addition, the deepest part of the LB is located between two frontal zones: in the northwest, the front is located along the Mona Ridge, and in the southeast, the front is represented by the Norwegian Current. Physical regularities of water exchange between the North Atlantic and the Arctic Ocean in 1958–2009 were analyzed on the basis of numerical experiments with eddy-permitting model of the ocean circulation with a resolution of 0.25° both along the longitude and latitude; they have shown that transversal oscillations of the Norwegian Current front were responsible for the formation of intermediate dense waters [11]. In this region, the zone of the vortex formation moving inside the LB has been noted [14, 15, 17].

A detailed review of the theoretical models by different authors describing the merging, regeneration, and stability of vortices is given in Bashmachnikov et al. [4].

In the introductory part of this contribution, the annual mean values of temperature, salinity, density, and ocean currents at the sea surface and depths up to 1500 m have been calculated for the whole region of the PS using the Argo model utilizing the Argo floats observations over the period 2005–2014. These observations showed the existence of the AV over the entire region of the LB with the

prevailing surface currents from 7–10 cm/s at the outer boundaries and up to 1–2 cm/s at the center of the LB; the current velocities decrease significantly with depth. At the same time, observations confirm the existence of a mesoscale lens at medium depths of 250–700 m over the deepest part of the LB. The evolution of the intratermocline lens, originally located in the central part of the LB depending on different types of velocity fields and the topographic effect are investigated using the contour dynamics method within the frames of the three-layer model.

Observation Materials and Methods

Temperature and salinity data measured by the Argo floats covering the 10-year period from 2005 to 2014 [3] were used to study the hydrological structure and water dynamics in the LB. The studied area includes the North Atlantic (NA) and the Polar Seas (PS) of the Arctic Ocean (AO) from 55° to 80° N and from 30° W to 15° E. In this area about 17,600 temperature and salinity profiles from 125 Argo floats have been collected during 10 years of observations. In the LB area (68°–73° N and 5° W–12° E) there are 3273 profiles measured by 73 Argo floats. Seasonal mean data in addition to the 10-year annual mean fields were calculated in this region using 835 Argo profiles in winter, 679 in spring, 891 in summer, and 868 in autumn.

We used the AMIGO method ([10], see also http://argo.ocean.ru) for the data processing. Spatial resolution of data in the calculated database is one degree by longitude and latitude, the time resolution is one month. During the processing of the Argo profiles, we used the method of the variational interpolation of measurements onto a regular grid, followed by a model hydrodynamic adjustment of the fields. Such a procedure minimizes the errors when transferring irregularly located measurements into the regular grid; hence, the solution passes as close to the data as possible. At the final stage, using the ocean general circulation model in the diagnostic and hydrodynamic adjustment mode, we calculated the balanced monthly means and climatological temperature, salinity, density, and velocity fields based on the Argo data [10].

Taking into account that the one-degree calculation grid is rough for detecting and tracking the motion of mesoscale vortices with an average radius of about 37 km [15], we made an attempt to describe the structure and dynamics of the PS waters on the basis of the materials described above. The annual mean temperature, salinity, density, and flow velocity fields in this region were constructed at the depths from the sea-surface (associated with the 30 m level) to 1500 m depth. The maps and vertical sections of these characteristics in the central part of the LB, limited on the map by the isobath 3200 m, have clearly shown the existence of the AV at depths of 200–800 m. The vortex center is located at the point with coordinates 69.5° N, 3.5° E, which agrees with the results by Alekseev et al. [1], Ivanov and Korablev [6, 7], Köhl et al. [8], and Raj et al. [15].

Figure 2a shows a map of the annual mean velocity vectors in 1° squares at a depth of 30 m in the entire PS region. According to these data, two jets of the main transit currents are distinguished, transporting the warm and salty waters of the Atlantic Ocean to the AO. The eastern jet NwAC (Norwegian Atlantic Current) transports Atlantic waters from the area of the Faroe-Shetland Strait and follows along the Norway shore, round the Voring Plato from the east. Further, it flows to the north towards the Fram Strait, forming the Spitsbergen Current and a branch of the warm current in the Barents Sea. The western branch of NwAC enters the Norwegian Sea east of Iceland and flowing over the Norwegian Basin from the southeast, passes along the LB from the south and northwest, and then leaves the LB towards the Fram Strait. The Norwegian Coastal Current (NCC) is also clearly pronounced. This scheme of the location of the main jets completely coincides with the studies by Orvik and Niiler [13], Köhl et al. [8], Raj et al. [15], Rossby et al. [18], and Volkov et al. [24].

The maps in Fig. 2 show the current behavior in each layer of the model considered below: surface layer 0–255 m, intermediate layer 255–650 m, and deep layer up to 3000 m. The current velocity in the deepest layer is shown in the two lower panels.

Let us consider in detail the nature of the velocity field in the region adjacent to the LB. A large-scale anticyclonic vortex is observed in the water column from the surface and up to 1500 m above the entire area of the basin, shown on the map

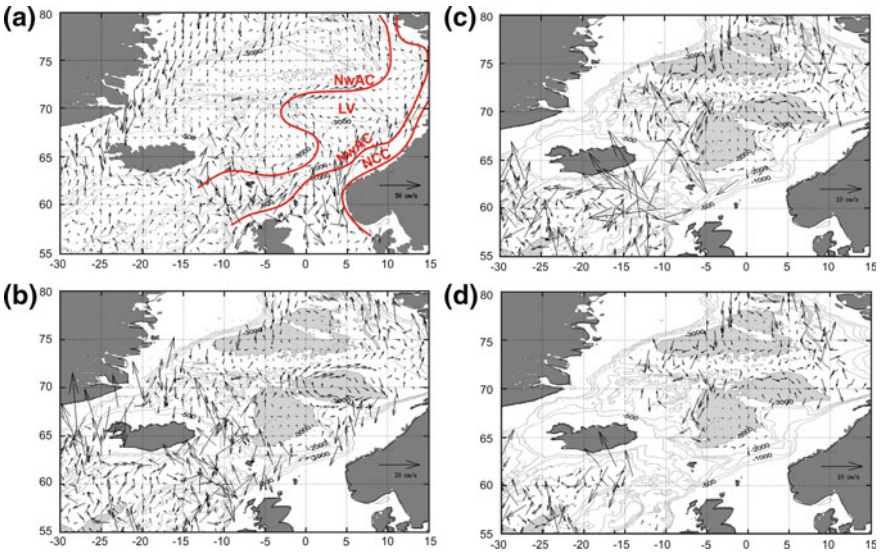


Fig. 2 Annual mean velocity vectors in 1° squares of the AMIGO model at depths: **a** 30 m, **b** 400 m, **c** 1000 m, and **d** 1500 m. On panel (**a**), the jets of the main transit currents are marked by red lines, namely: eastern and western jets of the NwAC (Norwegian Atlantic Current) and the jet of NCC (Norwegian Coastal Current). The scales of the velocity vectors are different on different maps

within the isobath of 3000 m. Note that the western and northern parts of this vortex are formed by the western NwAC jet. In the northern part, there are two currents: a jet of the Atlantic waters directed northward and confined to the Mona Ridge, as well as a jet along the edge of the LB, which weakens when moving eastward. The structure of the AV in the LB is asymmetric: the velocities at its northern periphery at the 30 m level are 6–12 cm/s, 4–7 cm/s at 400 m, and 1–3 cm/s at 1000 m; the velocities at its southern periphery are 3–6 cm/s, 3–5 cm/s, and 0–1 cm/s, respectively, and in the center they range from 1–2 cm/s to zero. This means that a circular current exists in the 0–1500 m layer along the perimeter of the isobath 3000 m, which is better pronounced in its northern part. The velocities are weak (0–1 cm/s) in its central part. This current is located within the boundaries of the isobath 3200 m. Therefore, after analyzing these materials, it is possible to assume that an intrathermocline lens exists at the depths of 250–700 m within the LB with an average long-term position of its center at 69.5° N, 3.5° E. The lens forced by the topographic beta-effect drifts along a cyclonic trajectory. A similar result, according to the hydrological observations, was obtained by Alekseev et al. [1], in which the authors proposed to consider the scales of 60 and 130 km for describing the AV and LB.

Although the Argo observations give us a clear pattern of currents in the LB as part of the general circulation in the PS, they cannot serve as a basis for explaining the mechanisms of formation and displacement of vortices. In this relation, it is extremely important to conduct model studies within the framework of an adequate mathematical model, using estimates of the natural space and time scales. Below, we investigate the evolution of an intrathermocline lens in the central part of LB for different types of current fields using the three-layer version of the Contour Dynamics Method.

Numerical Model and Simulation

Let us consider a model basin to explain the mechanism of the impact of the underwater depression on the behavior of the anticyclonic intrathermocline lens. We model the motion in the form of a set of two circular cylinders with displaced centers and with vertical walls which coincide with the 3000 and 3200 m isobaths.

Let the outer circle radius be equal to 132 km and the internal to 60 km; the height difference on each circle is 200 m, and their centers are in coordinates $(X_1;\ Y_1)$ and $(X_2;\ Y_2)$, respectively. We will use a quasi-geostrophic three-layer ocean model [20] with a piecewise-constant densities of ρ_1, ρ_2, ρ_3 between 0–250 m, 250–655 m, and 655–3000 m, respectively, and $\Delta\rho_1 = \rho_2 - \rho_1 = 0.00018\ \mathrm{g\,cm^{-3}}$, $\Delta\rho_2 = \rho_3 - \rho_2 = 0.00015\ \mathrm{g\,cm^{-3}}$. These values are typical for this water basin [4]. We assume that the vertical and horizontal scales are H = 3 km and L = 40 km, respectively; for the dimensionless layer thicknesses and radii of cylinders we get $h_1 = 0.0833$, $h_2 = 0.1383$, $h_3 = 0.7784$, and $R_1 = 3.3$, $R_2 = 1.5$.

We will also assume that the unperturbed external field consists of a combination of a homogeneous flow with horizontal velocities U, V, and a large-scale quasistationary anticyclonic vortex with characteristic azimuthal velocity A.

Taking into account the above assumptions, one can write the analytical expressions for stream functions ψ_i of the horizontal motion in the layers [20, 21]:

$$\psi_i(x, y) = -Uy + Vx + A\left(x^2 + y^2\right) - \sum_{j=1}^{2} \sigma_j T_{ji}, \quad i = 1, 2, 3; \tag{1}$$

where

$$T_{ji} = \begin{cases} \frac{r_j}{4} - \frac{q_{i2}s_{23}}{s_{13}\gamma_1^2}\left[1 - \gamma_1 \mathrm{K}_1\left(\gamma_1 R_j\right)\mathrm{I}_0\left(\gamma_1 r_j\right)\right] - \frac{q_{i3}s_{33}}{s_{13}\gamma_2^2}\left[1 - \gamma_2 \mathrm{K}_1\left(\gamma_2 R_j\right)\mathrm{I}_0\left(\gamma_2 r_j\right)\right], & r_j \le R_j, \\ \frac{1 + \ln(r_j)^2}{4} - \frac{q_{i2}s_{23}}{s_{13}\gamma_1}\mathrm{I}_1\left(\gamma_1 R_j\right)\mathrm{K}_0\left(\gamma_1 r_j\right) - \frac{q_{i2}s_{23}}{s_{13}\gamma_2}\mathrm{I}_1\left(\gamma_2 R_j\right)\mathrm{K}_0\left(\gamma_2 r_j\right), & r_j \ge R_j, \end{cases}$$
$$j = 1, 2; \quad i = 1, 2, 3.$$

Here, index i is the number of the layers (from top to bottom), and index j is the cylinder number; σ_j is an oriented height (negative for the depression) normalized by the horizontal section square of each of the cylinders, $r_j = \sqrt{\left(x - X_j\right)^2 + \left(y - Y_j\right)^2}$, where x and y are the coordinates of the observation point along the axes of the Cartesian coordinate system directed to the east and north respectively; I_n, K_n $(n = 0, 1)$ are modified Bessel functions of the n order; q_{ji} and s_{ji} are the matrix elements.

$$Q = \begin{pmatrix} 1 & \frac{h_3\lambda_2}{\lambda_2 - \lambda_1} & -\frac{F_1}{h_1\lambda_2} \\ 1 & \frac{1}{\lambda_2 - \lambda_1}\left(h_2\lambda_2 + \frac{F_2}{h_2}\right) & -\left(\frac{F_1}{h_1\lambda_2} + 1\right) \\ 1 & \frac{1}{\lambda_2 - \lambda_1}\left[h_2\lambda_2 + \frac{F_2}{h_2} + \lambda_1 + \frac{F_1(h_1 + h_2)}{h_1 h_2}\right] & -\left[\frac{F_1}{h_1\lambda_2} + 1 + \frac{h_2}{F_2}\left(\lambda_2 + \frac{F_1(h_1 + h_2)}{h_1 h_2}\right)\right] \end{pmatrix}$$

and

$$S = Q^{-1} = \begin{pmatrix} h_1 & h_2 & h_3 \\ -\frac{h_2}{F_2}\left[\lambda_2 + \frac{F_1(h_1 + h_2)}{h_1 h_2}\right] & \frac{h_2}{F_2}\left[\lambda_2 + \frac{F_1(h_1 + h_2) + F_2 h_1}{h_1 h_2}\right] & -1 \\ -\frac{1}{\lambda_2 - \lambda_1}\left[\lambda_1 + \frac{F_1(h_1 + h_2)}{h_1 h_2}\right] & \frac{1}{\lambda_2 - \lambda_1}\left[\lambda_1 + \frac{F_1(h_1 + h_2) + F_2 h_1}{h_1 h_2}\right] & -\frac{1}{\lambda_2 - \lambda_1}\frac{F_2}{h_2} \end{pmatrix}.$$

Here, $\lambda_{2,1} = -\frac{1}{2}\left[\frac{F_1}{h_1} + \frac{F_1 + F_2}{h_2} + \frac{F_2}{h_3} \pm \sqrt{\left(\frac{F_1}{h_1} + \frac{F_1 + F_2}{h_2} + \frac{F_2}{h_3}\right)^2 - 4\frac{F_1 F_2}{h_1 h_2 h_3}}\right]$; $F_1 = (fL)^2/g_1' H$ and $F_2 = (fL)^2/g_2' H$ are Froude numbers $(g_1' = g\Delta\rho_1/\rho_0$, $g_2' = g\Delta\rho_2/\rho_0)$; f is the constant Coriolis parameter, g is the acceleration of gravity, ρ_0 is the mean density value.

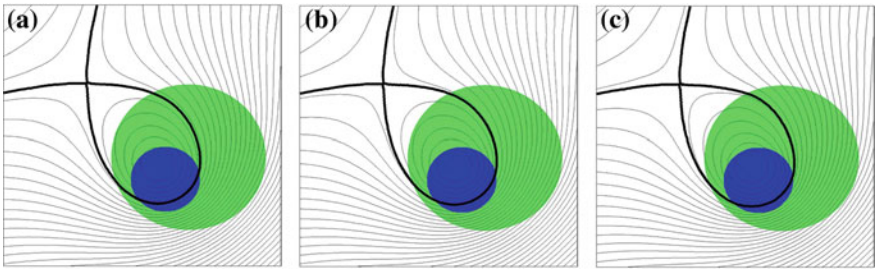


Fig. 3 Phase portraits are shown as thin black lines for the barotropic external flow [1] over a cylindrical stepwise depression when $U = V = 10$ cm/s and $A = 0$. The section of the external cylinder is green, and the internal cylinder section is blue. Thick lines show separatices, which divide the areas of capture (with closed trajectories) and the areas with running flows of two types on the opposite sides of separatrices. Panels **a**, **b**, **c** are related to the upper, middle and lower layers, respectively

Figure 3 shows examples of fields ψ_i when $(X_1;\ Y_1) = (0;\ 0)$ and $(X_2;\ Y_2) = (-1.5;\ -1.5)$. Panels "**a**", "**b**", "**c**" represent phase portraits in the upper, middle and lower layers, respectively, in the particular case $A = 0$ when the depression is surrounded by the barotropic northeast stream. Cyclonic relative vorticity is generated owing to the conservation of the potential vortex in the vicinity of the depression [9]. Even at a considerable distance from the depression the flow remarkably twists counterclockwise, and a quasi-barotropic topographic cyclone with a small predominance of capture in the lower layer is formed in the western part of the depression.

Figure 4 depicts stream function ψ_2 in the middle layer, in the case $A < 0$. Panels **a–d** demonstrate the effect of an external anticyclonic vortex on the field function behavior of horizontal velocities. Here, four phase portraits are given for the case of the current type [1] and $A = -0.005,\ -0.01,\ -0.02\ -0.04$, which correspond to the azimuthal velocity values $v = -0.9,\ -1.8,\ -3.6,\ 7.2\,\mathrm{cm\,s^{-1}}$ on the circular isobath 3000 m. It can be seen that the presence of an external large-scale anticyclone, contributes to the reduction of the area of the cyclonic Taylor column until it disappears completely (panel "**d**"), when it fills the space above the depression almost completely suppressing the topographic effect. At moderate values of the

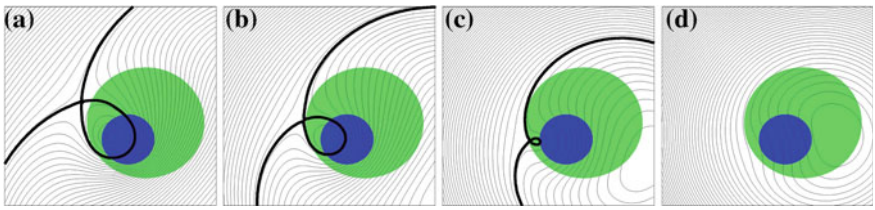


Fig. 4 The same as in Fig. 3, but for the middle layer: **a** $A = -0.005$, **b** $A = -0.01$, **c** $A = -0.02$, **d** $A = -0.04$

intensity of the anticyclone (panels "**a**" and "**b**"), one can speak about the qualitative agreement with Fig. 2, identifying the bundle of isolines, located to the right from the separatrix loop, with the eastern stream NwAC, and the one to the left with the west side. The same can be said about the phase portraits in Fig. 3.

The resulting velocity distributions will now be used as "external fields" for the anticyclonic lens, localized in the pycnocline [6, 7, 23]. Hereinafter, we will consider processes that take place only in the corresponding middle layer.

Without resting on the problem of the anticyclonic lens in the vicinity of a topographic vortex (these questions are discussed in detail in Ivanov and Korablev [6, 7]), we consider several versions of the different behavior of the lens depending on its initial location.

Let us consider separately cases (I) without taking into account the vortex ($A = 0$), and (II) taking into account the effect of anticyclonic rotation on the external flow ($A < 0$).

Case (I). At the initial time moment, we assume that the lens has a circular shape with dimensionless radius $r_{lens} = 0.6$, which corresponds to 60 km, and represents a vortex patch belonging to the middle layer, which is a region with a constant value of potential vorticity (PV). In this case, the initial PV value determines the time scale. Assuming the azimuthal velocity at the outer edge of the lens to be 20 cm/s, we will choose PV such that the rotation period of its liquid particles is equal to 20 days. This period will correspond to a unit of dimensionless time in numerical experiments.

In the presence of a lens, the phase portrait is constructed taking into account its effect on the current function field using the Contour Dynamics Method developed for the three-layer rotating liquid [20].

In the first experiment, shown in Fig. 5, the initial position of the lens center coincides with the stationary elliptic point inside the cyclonic topographic vortex, which is located inside the separatrix loop of the current function for the middle

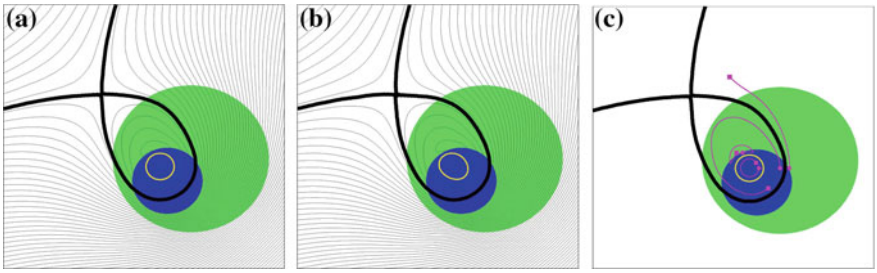


Fig. 5 Stream function field of the middle layer ψ_2 in the vicinity of the submerged depression in the case of the lens presence. The center of the lens is located at a "motionless" point of the separatrix loop in "**a**" and "**b**" panels (a 20-day interval corresponds to the unit of dimensionless time). Panel "**c**" shows the initial parts of the trajectory of fluid particles, placed both inside the lens and at its vicinity. In all panels, a thick yellow line represents the lens contour. Trajectories of fluid particles and markers (round as the initial position, and square as the last calculated position) are shown in by magenta color in panel "**c**"

layer. Assuming the selected external parameters, this point has coordinates (−1.3098; −0.3511). Panel "**a**" shows that, in this case, the lens does not virtually affect the kinematic characteristics of the flow outside the topographic vortex. Changes are observed only in the inner part of the separatrix loop. Panel "**b**" shows that the position of the lens can be considered almost quasistationary. Indeed, when $t = 10$, i.e., more than 6 months after the beginning of motion, the lens center remains in its initial position. Of course, the lens does not retain a circular shape, since the external field is not axisymmetric, but the position of its center stably holds almost the same place. Panel "**c**" shows the behavior of several liquid particles inside and outside the lens. Three of the four selected material points are initially located at the latitude of the lens center and move as follows: (1) the particle inside the lens rotates along a closed circular path in the anticyclonic direction, (2) the particle between the lens and the boundary of the separatrix loop rotates along the cyclonic path, remaining inside the topographic cyclone, (3) the particle located outside the separatrix loop is carried away by the current beyond the depression. Finally, the particle, initially located north of the lens boundary, moves along a hole-shaped path reminding the configuration of the corresponding isoline of the current function.

This experiment suggests that the region of the capture of an anticyclonic lens by a cyclonic topographic vortex cannot only be confined to a motionless elliptic point but may have finite dimensions.

The next series of calculations presented in Fig. 6, confirms and concretizes the latter assumption. In this figure, the lenses themselves are not shown; it gives only the trajectories of their centers. In panel "**a**", at the initial time, this center is placed at the hyperbolic point of the separatrix of the current function field with coordinates (−4.4001, 3.3899). Theoretically, this singular point is the place of attraction of the trajectories, and the calculation does, in fact, show that the center of the lens remains in place for quite a long time. But since the coordinates of the hyperbolic point are calculated approximately, then, because of the instability of the equilibrium position, the lens begins to move along a trajectory passing in a very close neighborhood of the separatrix. After a nearly complete tour along the loop, the lens is again slowing down in the vicinity of the hyperbolic point, but then it breaks off

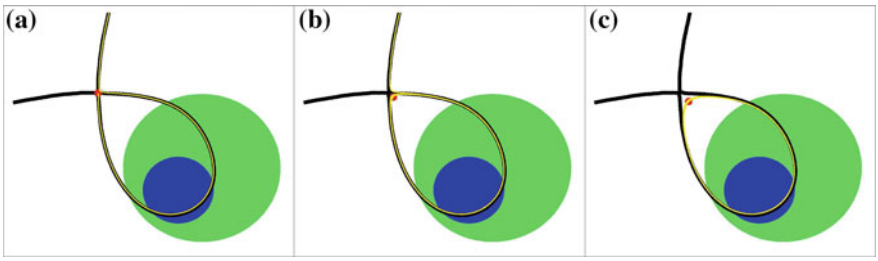


Fig. 6 Trajectories (yellow lines) of the lenses' centers are initially localized in the hyperbolic point of the unperturbed separatrix and in its vicinity. Red markers depict initial positions of the lenses' centers: (−4.4001; 3.3899) in panel "**a**"; (−4.2; 3.2) in panel "**b**"; (−4.0; 3.0) in panel "**c**"

from the closed part of the trajectory and is carried away by a stream along one of the separatrix "whiskers".

A small displacement of the initial position of the lens center inside the separatrix loop (Fig. 6b, where the initial coordinates of the center are indicated in the caption) leads to the fact that now the lens makes three revolutions inside the loop along an insignificantly untwisting spiral, and then, like the previous case, is carried away by flow beyond the depression. Finally, if the displacement of the initial position of the lens center toward the center of the topographic vortex becomes larger (Fig. 6c), the lens is completely captured by the topography. At the end of the calculation time interval, the lens performed 10 complete cycles. This series of calculations shows that the region of the lens capture by the depression occupies almost the entire separatrix loop of the current function except for a small neighborhood near the hyperbolic singular point. It is important to note that the capture can take place even in the case when the lens is originally located outside the depression (the figure in panel "**c**" shows this clearly). Thus, the trajectory of the lens center in panel "**c**" can be approximately taken as the outer contour of the cross section (in the middle layer) of the Taylor column [7, 9, 25], which limits the region of the initial positions of the lens centers that do not leave the vicinity of the depression.

The unperturbed separatrix of the phase portrait of the middle layer is shown in all three panels of Fig. 6. However, it is obvious that the presence of a lens (if it is not initially placed at a stationary elliptic point, as in Fig. 5) substantially alters the entire structure of the flow and the configuration of the separatrices in particular.

Figure 7 shows how a lens affected by an "external field" changes the field of the current function over the depression and at its vicinity. Here, we show a sequence of instantaneous lens configurations in the case presented in Fig. 6a, when, at the initial moment, the lens center was located at the hyperbolic point of the unperturbed separatrix. In all panels, the stationary unperturbed separatrix is depicted by a grey solid line. It is not related to the real non-stationary velocity field and is only a marker curve that tracks the motion of the lens center. The first panel shows that already at $t = 0$, two separatrices begin their formation (red and black solid lines) in the current function field. Note that, theoretically, at this stationary point they should merge, but as mentioned above, the coordinates of the hyperbolic point are calculated with an error, and therefore two separatrices have a small relative displacement (the black line has a self-intersection point located to the southwest of the lens center, and red one to the northeast). This experiment also gives an opportunity to demonstrate the uneven motion of the lens along its trajectory. Indeed, since the starting position is connected with a stationary point of the phase portrait, at the initial stage of the movement, the lens moves very slowly (the second panel corresponds to time $t = 8$). Then the motion accelerates, and the maximum velocity of the lens along the trajectory occurs at the maximum distance from the initial position. In the course of the lens re-approach to the hyperbolic point of the unperturbed separatrix, its motion slows down again (see indicated times at the panels). Note that a similar effect appeared in the calculations by Köhl et al. [8] (see Fig. 11).

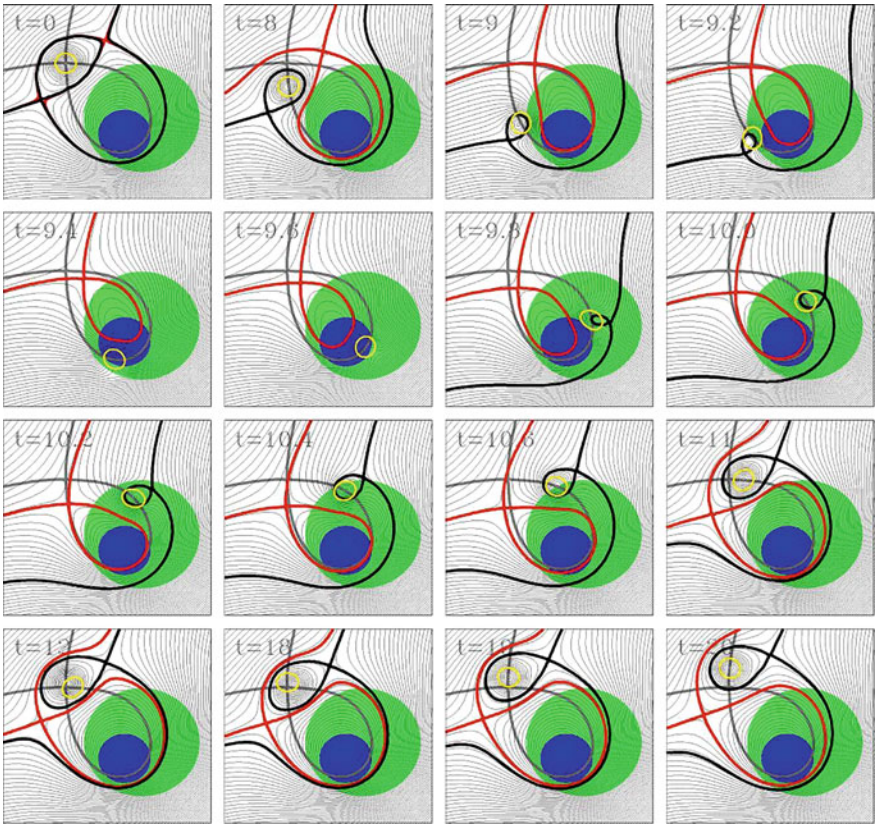


Fig. 7 Time scans of the sequence of the lens instantaneous positions at its initial location at the hyperbolic point of the unperturbed separatrix (gray solid lines) at the indicated moments of the dimensionless time. Two separatrices are forming in the field of the current function represented by black and red solid lines

It is easy to see that at each time moment the black separatrix loop is associated with the local position of the lens, and the red separatrix loop is associated with the bottom relief. It should be noted that the first of them decreases (and even disappears in the interval 9.2–9.4) in the vicinity of the maximum velocities of the lens motion. This can be explained by the weakening of the local anticyclonic swirl of the flow due to the increased influence of the topographic cyclone. The demonstration of the controlling role of the unperturbed separatrix during all stages of the lens motion, despite the complex nature of the evolving field of horizontal velocities is an important result of this and previous experiments.

Case (II). Let now the external flow, in addition, have a general anticyclonic rotation ($A < 0$).

Figure 8 shows the behavior of the captured lens against the background of instant phase portraits at $A = -0.01$. In this case, the hyperbolic self-intersection

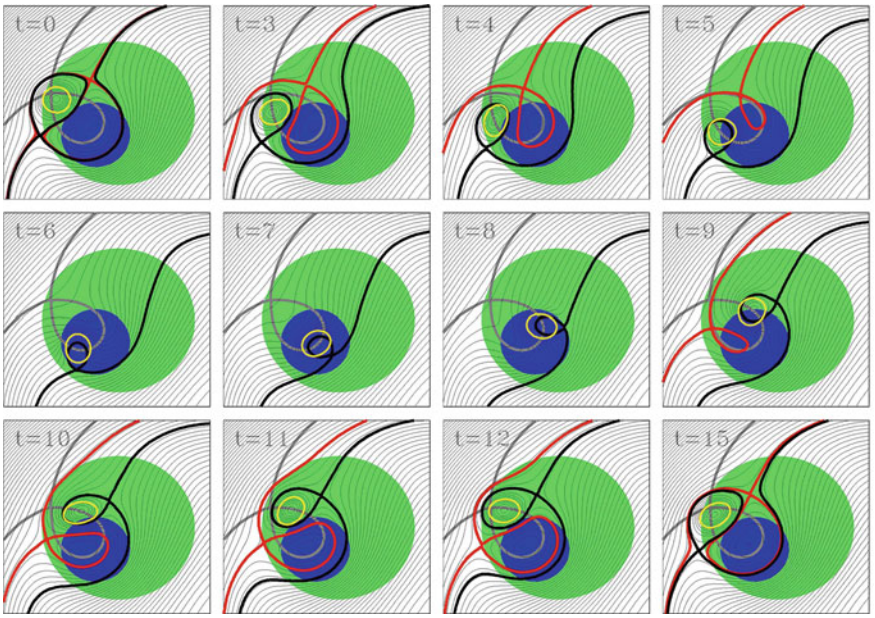


Fig. 8 The same as in Fig. 7, but taking into account a large-scale anticyclonic vortex in the local area of the depression at $v = -1.8\,\mathrm{cm\,s^{-1}}$, in the case when the initial position of the lens is located inside the loop of the unperturbed separatrix in the vicinity of the hyperbolic point

point of the separatrix has coordinates (−2.899, 0.905), and the initial position of the lens center is inside the loop at the point (−2.70, 0.63). It is seen that under such conditions the lens remains trapped, and its center describes a closed trajectory inside the separatrix loop. This figure shows the calculations results over an interval, slightly longer than one period. As in the experiment shown in Fig. 7, the rotation is not uniform: it slows down along the trajectory near the hyperbolic singular point and accelerates in the course of motion away from this point along the trajectory. The most significant changes in the volume of the loop are observed on the red separatrix associated with the bottom relief. It decreases, and can even completely disappear (moments 60, 70, and 80), when the anticyclonic lens passes over the deepest part of the basin and, thus, partially neutralizes the cyclonic twist induced by the depression that prevents the formation of a region of closed trajectories.

We note an interesting side effect: a periodic occurrence of a meandering jet at time intervals when the black and red separatrices are separated by a specific distance (intervals 3–5 and 9–12). For example, at $t = 12$, fluid particles in this stream, located between the "whiskers" of the black and red separatrices, approach the depression from the southwest; over the depression, they round one of the loops, pass between two loops, round the second loop, and finally leave the vicinity of the depression in the northeastern direction. When the separatrices are close to each

other ($t = 0$ and $t = 15$), the jet vanishes, and the flow rounds the region of the hugged loops, like one solid body.

Figures 4 and 8 show that, despite the external anticyclonic rotation, a cyclonic Taylor column can be formed over the depression during a specific interval of the anticyclone intensity. This column is capable to capture and twist counter-clockwise passive liquid particles as well as vortices inside it. This gives an explanation to the seemingly strange cyclonic direction of the observed lens motion in the works by Ivanov and Korablev [6, 7]. In the case where the initial location of the lens center is outside the Taylor column, the lens leaves the vicinity of the LB.

Conclusions

The formation, drift, and destruction of mesoscale vortices within the LB occur against a background of the large-scale processes in the study site. Hence, the PS is a buffer zone of the water exchange between the North Atlantic and the Arctic Ocean [11]. The complex character of their interaction determines the seasonal and multi-year variability of the hydrological regime and dynamics of the PS and, in particular, the LB. Raj et al. [15] note the inter-annual and seasonal variability of the number of observed vortices in the investigated basin and their relationship to these external processes. It should be kept in mind that it is almost impossible to recreate in the model experiment the complete similarity with the real ocean conditions. However, such modeling, in spite of some assumptions, allows us to trace all the stages of the influence of the external parameters on the behavior of the intrathermocline lens. In this contribution based on the results of various experimental observations, the existence of a quasi-permanent intratermocline lens in the vicinity of the deep part of the LB is postulated as a fact. Within the framework of the numerical model, we model various scenarios of the lens motion, which can take place in the ocean. In particular, it is clearly shown that (1) the lens not captured by the quasi-stationary Taylor column leaves the LB region, (2) the lens located in the vicinity of the hyperbolic point can remain for a long time in the quasistatic state, (3) the lens belonging to the inner part of the Taylor column, which is associated with the separatrix loop of the stream function, performs closed periodic rotations in the cyclonic direction. The latter can occur even in presence of a large-scale anticyclonic vortex of moderate intensity. The model conclusions about the behavior of the quasi-constant anticyclonic vortex in the LB coincide with the main observed results. They also give an idea of the physical processes that determine the various conditions of its cyclonic drift depending on the variability of the velocity field characteristics and the topographic effect.

Acknowledgements The work was supported by Russian Science Foundation (grant 14-50-00095) (analysis of the ocean data) and Ministry of Education and Science of the Russian Federation (grant 14.W03.31.0006, (numerical simulation), and Russian Foundation of Basic Research (grant 16-55-150001) (vortex dynamics).

References

1. Alekseev, G. V., Bagryantsev, M. V., Bogorodskiy, P. V., Vasin, V. V., & Shirokov, P. E. (1991). Structure and circulation of water in the area of anticyclonic eddy in the northeastern Norwegian Sea. *Problems of the Arctic and Antarctic, 65,* 14–23 (in Russian).
2. Alekseev, G. V., Nikolaev, Yu V, Romanov, A. A., Romantsev, V. A., & Sarukhanyan, E. I. (1986). Results of natural investigations in the Norwegian energy active zone. *Itogi Nauki i Tekhniki, Atmosphere, Ocean, Space Program RAZREZY, 7,* 46–72 (in Russian).
3. Argo. (2000). Argo float data and metadata from Global Data Assembly Center (Argo GDAC). SEANOE. http://doi.org/10.17882/42182.
4. Bashmachnikov, I. L., Sokolovskiy, M. A., Belonenko, T. V., Volkov, D. L., Isachsen, P. E., & Carton, X. (2017). On the vertical structure and stability of the Lofoten vortex in the Norwegian Sea. *Deep-Sea Research Part I* (in press).
5. Gascard, J.-C., & Mork, K. A. (2008). Climatic importance of large-scale and mesoscale circulation in the Lofoten Basin deduced from Lagrangian observations. In R. R. Dickson, J. Meincke, & P. Rhines (Eds.), *Chapter 6: Arctic-Subarctic Ocean fluxes. Defining the role of the Northern Seas in climate* (pp. 131–143). Dordrecht: Springer.
6. Ivanov, V. V., & Korablev, A. A. (1995). Formation and regeneration of the pycnocline lens in the Norwegian Sea. *Russian Meteorology and Hydrology, 9,* 62–69.
7. Ivanov, V. V., & Korablev, A. A. (1995). Interpycnocline lens dynamics in the Norwegian Sea. *Russian Meteorology and Hydrology, 10,* 32–37.
8. Köhl, A. (2007). Generation and stability of a quasi-permanent vortex in the Lofoten Basin. *Journal of Physical Oceanography, 37,* 2637–2651.
9. Kozlov, V. F. (1984). *Models of the topographic vortices in ocean* (p. 200). Moscow: Nauka.
10. Lebedev, K. V. (2016). An argo-based model for investigation of the Global Ocean (AMIGO). *Oceanology, 56,* 172–181.
11. Moshonkin, S. N., Bagno, A. V., Gusev, A. V., Filyushkin, B. N., & Zalesny, V. B. (2017). Physical properties of the Atlantic-Arctic water exchange formation. *Izvestiya Atmospheric and Oceanic Physics, 53,* 213–223.
12. Orvik, K. A. (2004). The deepening of the Atlantic water in the Lofoten Basin of the Norwegian Sea, demonstrated by using an active reduced gravity model. *Geophysical Research Letters, 31,* L01306. https://doi.org/10.1029/2003GL018687.
13. Orvik, K. A., & Niiler, P. (2002). Major pathways of Atlantic water in the northern North Atlantic and Nordic Seas toward Arctic. *Geophysical Research Letters, 29*. https://doi.org/10.1029/2002GL015002.
14. Poulain, P.-M., Warn-Varnas, A., & Niiler, P. P. (1996). Near-surface circulation of the Nordic Seas as measured by Lagrangian drifters. *Journal Geophysical Research, 101,* 18237–18258.
15. Raj, R. P., Chafik, L., Nilsen, J. E. Ø., Eldevik, T., & Halo, I. (2015). The Lofoten Vortex of the Nordic Seas. *Deep-Sea Research Part I, 96,* 1–14.
16. Raj, R. P., & Halo, I. (2016). Monitoring the mesoscale eddies of the Lofoten Basin: Importance, progress, and challenges. *International Journal of Remote Sensing, 37,* 3712–3728.
17. Rodionov, V. B., & Kostianoy, A. G. (1998). *Oceanic fronts of the North-European basin seas* (293 pp.). Moscow: GEOS (in Russian).
18. Rossby, T., Ozhigin, V., Ivshin, V., & Bacon, Sh. (2009). An isopycnal view of the Nordic Seas hydrography with focus on properties of the Lofoten Basin. *Deep-Sea Research Part I, 56,* 1955–1971.
19. Søiland, H., & Rossby, T. (2013). On the structure of the Lofoten Basin Eddy. *Journal of Geophysical Research: Oceans, 118,* 4201–4212.
20. Sokolovskiy, M. A. (1991). Modeling triple-layer vortical motions in the ocean by the Contour Dynamics Method. *Izvestiya Atmospheric and Oceanic Physics, 27,* 380–388.

21. Sokolovskiy, M. A., & Verron, J. (2014). Dynamics of vortex structures in a stratified rotating fluid. In *Series Atmospheric and oceanographic sciences library* (Vol. 47, p. 382). Switzerland: Springer International Publishing.
22. Voet, G., Quadfasel, D., Mork, K. A., & Søiland, H. (2010). The mid-depth circulation of the Nordic Seas derived from profiling float observations. *Tellus, 62A,* 516–529.
23. Volkov, D. L., Belonenko, T. V., & Foux, V. R. (2013). Puzzling over the dynamics of the Lofoten Basin—A sub-Arctic hot spot of ocean variavility. *Geophysical Reseach Letters, 40,* 738–743.
24. Volkov, D. L., Kubryakov, A. A., & Lumpkin, R. (2015). Formation and variability of the Lofoten Basin vortex in a high-resolution ocean model. *Deep-Sea Research Part I, 105,* 142–157.
25. Zyryanov, V. N. (1995). *Topographic eddies in sea currents dynamics* (p. 240). Moscow: Water Problems Institute of RAS (in Russian).

The Global Atmosphere Oscillations in the Context of the Recent Climate Change

Victor G. Neiman, Vladimir I. Byshev, Yury A. Romanov and Ilya V. Serykh

Introduction

It is commonly accepted [1–4] that variability of the Earth's climate is characterized by a wide range of oscillations. Despite their origin, all those variations may be classified into two types. The first of them is presented by resonance perturbations at the frequencies corresponding to the external forcing. The second one combines local fluctuations of both inertial and auto-oscillation origin.

An assessment of the regional current climate change is usually based on the indices of various oscillations, such as the North Atlantic, North Pacific, Arctic, Southern, etc. [5]. These oscillations are indicative of some features of the internal dynamics of the climate system. Physical background of frequency-amplitude characteristics of the oscillations may signify the existence of the global scale atmospheric variations, whose dynamic structure includes all known regional indices.

Therefore the problem arises of acquiring a better understanding of the global atmospheric oscillations related to the structure of the current climate variability. The goal of our study is to elaborate appropriate indices suitable for prompt and adequate detecting of this phenomenon. An attempt to find an answer to some of these questions is undertaken in this contribution.

V. G. Neiman (✉) · V. I. Byshev · Y. A. Romanov · I. V. Serykh
Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
e-mail: vneiman2007@yandex.ru

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_22

The Data and Methods

The analysis of the spatial structure of the anomalies of the hydro-meteorological characteristics appearing under the influence of the global atmospheric oscillations was carried out on the basis of the average monthly fields of the atmospheric pressure P at the sea level (HadSLP2 database) and the surface temperature T (CRUTEM4 database) both prepared with a resolution of 5° × 5° by the British Met Office Hadley Center for Climate Science for the period of 1900–2012 [6, 7]. As it was shown in [8], the deformation fields of P and T reflecting the draw effect of the interannual GAO involve the time period of the corresponding impacts of El Niño—La Niña as well. Taking this fact into account, the assessment of P and T anomalies caused by the interannual GAO has been performed by formula (1). It is important, however, to mention that large-scale anomalies within T and P global fields caused by the GAO were found to appear well before the ENSO events onset [8]. The latter (El Niño and La Niña) coincide in time with the GAO extreme phase states. Therefore, our calculation of the GAO amplitude values was performed using the sets of the global fields of T and P coinciding in time with the culminating periods of El Niño and La Niña. In turn, these periods were well marked by the commonly accepted appropriate indices [9].

$$X_H(\varphi, \lambda, z_0) = \frac{1}{N_{EL}} \sum_{i=1}^{N_{EL}} \delta X_{EL}(\varphi, \lambda, z_0, t_i) - \frac{1}{N_{LA}} \sum_{j=1}^{N_{LA}} \delta X_{LA}(\varphi, \lambda, z_0, t_j) \qquad (1)$$

With the goal in mind to perform independent control of the reliability of our method we calculated the difference between the mean fields separately for each half-century: 1900–1949 and 1950–1999. The first period included nine El Niño and six La Niña events, the second included 10 and 5 events, respectively. We summarized the results of these calculations for the second half of the 20th century in Fig. 1, which was found similar to the one in the first period.

The GAO manifestations in the hydro-meteorological fields on inter-decadal time scale (Fig. 2) was estimated by formula (2) as the difference between the mean fields of two consecutive climate phases: 1975–1999 and 1950–1974 [10, 11].

$$X_L(\varphi, \lambda, z_0) = \frac{1}{N_1} \sum_{i=1}^{N_1} X(\varphi, \lambda, z_0, t_i) - \frac{1}{N_2} \sum_{j=1}^{N_2} X(\varphi, \lambda, z_0, t_j) \qquad (2)$$

Notations in formulas (1, 2) are: X (φ, λ, z_0, t) is the field of sea level pressure P (φ, λ, z_0, t) or the surface temperature T (φ, λ, z_0, t); N_{EL}, N_{LA}, and N_1, N_2 are time samples (average monthly number of X, or anomalies δX) of global fields, compiled in conjunction with the temporary episodes of El Niño and La Niña in the first case, and with the phases of climate, respectively; φ, λ, z_0, t are the latitude, longitude, level, and time, respectively; δX are the anomalies of the atmospheric pressure and surface temperature after removing its seasonal variation. Subscripts $_H$ and $_L$ in the characteristics of X are attributed to the high frequency (inter-annual) and low frequency (inter-decadal) GAO.

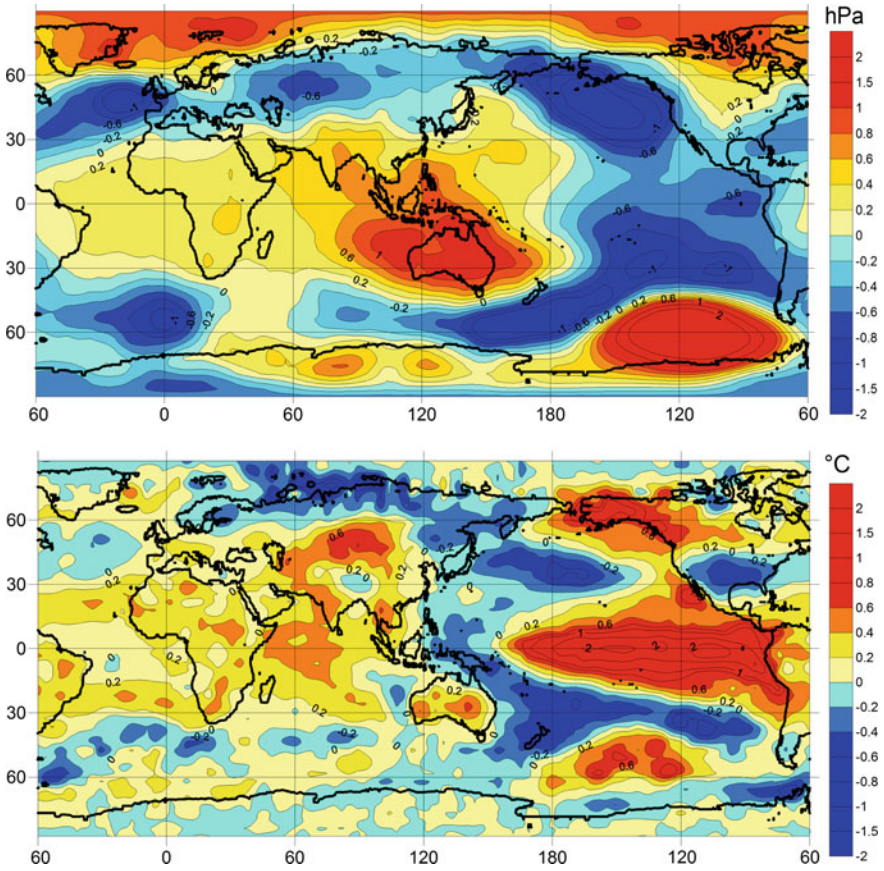


Fig. 1 Average thermo-baric structure of the interannual GAO in 1950–1999. The GAO amplitude in the fields of atmospheric pressure anomalies at the sea level (top panel) and near-surface air temperature (bottom panel)

Discussion

The initial discussion on the interannual GAO [8] showed this phenomenon in the form of an anomalous structure of the planetary pressure field, as presented in Fig. 1 (top panel), which contains the corresponding modes of all known regional atmospheric pressure oscillations. The key element of this structure was a large-scale positive anomaly in the equatorial-tropical region surrounded by a low pressure belt along its outer boundary. Figure 1 shows that this anomaly spreads from the West Atlantic to the Date Line and extends in the latitudinal direction between 30° S and 30° N. The width of the low pressure belts just over the northern and southern boundaries of the positive anomaly is as high as 2000–3000 km. In

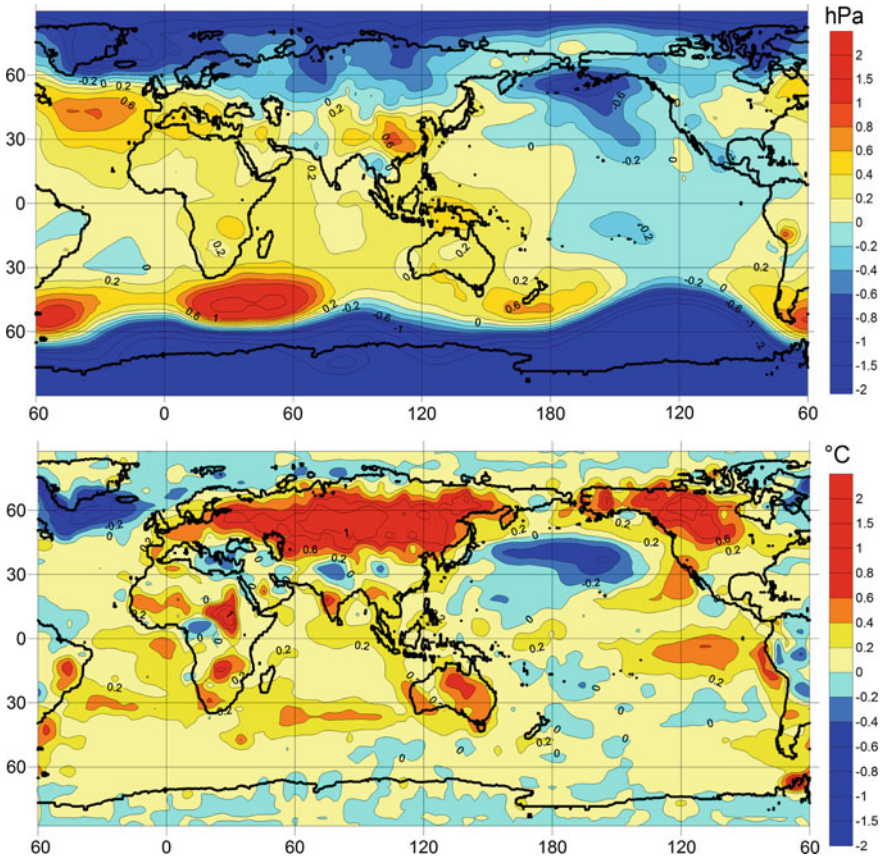


Fig. 2 The amplitude of the global atmospheric oscillation on inter-decadal time scale in the fields of the atmospheric pressure anomalies at the sea level (top panel) and surface air temperature (bottom panel)

the northern polar region and in the major part of the southern polar zone, the positive pressure anomalies dominate again.

Comparison of Figs. 1 and 2 shows the similarity and fundamental differences of the GAO structural features on the two time scales. In both cases, the GAO is characterized by positive atmospheric pressure anomalies in the equatorial-tropical region with its core in the Indian Ocean.

Though the inter-decadal GAO appears less intense in the terms of its anomalies, it occupies a greater latitudinal belt. Moreover, the interannual GAO is characterized by a positive anomaly along its entire outer boundary, which in its turn is adjacent to the region of low atmospheric pressure. Zones of high pressure are sometimes observed at higher latitudes of both hemispheres. Otherwise in the second case, low atmospheric pressure characterizes the whole area beyond the central positive anomaly.

Figure 2 (top panel), as already mentioned, shows a map of the mean difference of the surface atmospheric pressure between two consequent climate scenarios, namely the relatively warmer (1975–1999) and the cooler (1950–1974) periods. As one can see in this case, the pressure distribution in the tropical zone of the Pacific Ocean is similar to the distribution characteristic of the El Niño events (Fig. 1). In both cases, the region of the positive anomalies of the atmospheric pressure is surrounded by the zones of negative anomalies. However, in the case of the interannual GAO at the polar latitudes, the negative anomalies are replaced with the positive ones, and in the case of the inter-decadal GAO they remain negative up to the Pole.

The difference in the temperature distribution in the tropical Pacific (Fig. 2, bottom panel) is similar to the distribution shown in Fig. 1, i.e., the presence of positive T anomalies in the eastern zone. Therefore, in many studies [3, 12–15] these changes in the pressure and temperature are called the inter-decadal oscillations like El Niño. However, Fig. 2 (top panel) shows that changes in the pressure considered here are global but not limited to the tropical Pacific. Therefore, it is more correct to call them inter-decadal global oscillations, what we do and offer.

If we consider the properties of the respective thermal perturbations, the most significant positive temperature anomaly (Fig. 1) occurs, as expected, in the eastern tropical Pacific under the influence of the El Niño events induced by the GAO. A smaller positive anomaly is observed in the northern part of the Pacific, including the territory of Alaska and the Chukchi Peninsula.

An extensive negative temperature anomaly covers the mid-latitudes of the Pacific, Northeast Asia, the southern and northeastern parts of the North American continent. The main distinguishing feature of the spatial distribution of temperature anomalies generated by the inter-decadal GAO (Fig. 2) is presented by a four-core structure of the positive anomaly in the northwest of North America and in the northeast of the Eurasian continent, as well as by the negative anomalies in the North Pacific and Atlantic oceans. There are grounds to believe that this pattern reflects qualitatively the phase redistribution of heat forced by the GAO in the ocean-atmosphere-continent system.

The interannual GAO is obviously related to the seasonal cycle of the atmospheric processes, and its appearance coincides with the periods of the equilibrium state of the climate system [8]. This condition occurs whenever the Sun is above the Earth's equator (radiation equilibrium), and the system of the global monsoon circulation becomes unstable after its change to the interphase mode. The initial moment of the GAO as shown in [16] is linked directly to the atmospheric pressure increase in the equatorial-tropical region of the Atlantic and Indian oceans (thermal tide) where negative sea surface temperature anomalies were reported to exist in this period.

We suggest that the inter-decadal GAO (Fig. 2) appears as a shift in the dynamics of the climate system, which is followed by a change of the climate scenario either on the regional or even the global scale. According to our estimates [17], this process is related immediately to the intensification of the heat flux from the oceans to the atmosphere, which is succeeded by the climate warming on land.

Since the ocean accumulates heat, the continents should experience extremes of the climate severity.

Taking the previous discussion into account, we propose to consider an idea of the inter-decadal GAO as an inhibitor of the global climate phase variability, which manifests itself in the form of successive turnover of the climatic events or even in the change of the climate scenario [11]. This is a mechanism how climatic effect of each of the two GAO modes may be revealed according to their specific time scale. It is appropriate to make a nontrivial conclusion [8] that the high-frequency (interannual) mode of the GAO plays a role of the trigger mechanism for the Pacific ENSO events. Independent verification of this important result has fully proved its correctness. This can be shown by the example of the GAO generated field of the near-surface air temperature anomalies (Fig. 1), which appears explicitly under the El Niño forcing. This assumption is also supported by the comparison of time functions of the GAO and ENSO indices (Fig. 3), which are generally similar. We note that the El Niño—Southern Oscillation (ENSO) index was calculated here as the monthly average anomaly of the sea surface temperature (°C) in the area [18] of Niño-3 (5° S–5° N, 160° W–90° W).

The pattern of the sea surface pressure anomaly field formed under the influence of the global atmospheric oscillations is the goal of this analysis. An actual insight into the phenomenon can help us to plot the phase pattern of the GAO evolution in greater detail. But now we would pay attention only to the fact that the field of

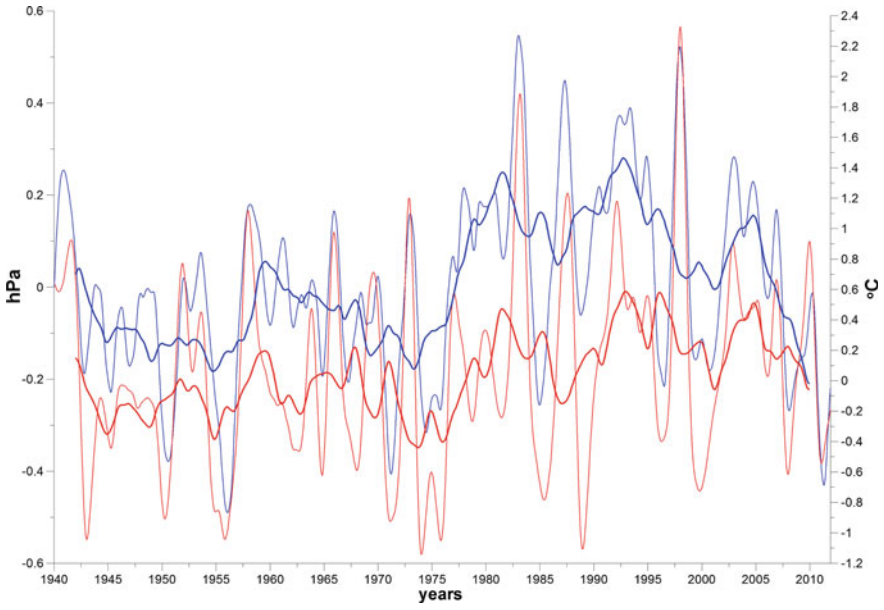


Fig. 3 Time evolution of the GAO index (blue line is related to the left scale) and ENSO index (red curve, right scale). Thin lines show annual smoothing; bold curves show the data after 7-year running smoothing

pressure anomalies corresponding to the inter-decadal GAO is apparently quite stable, perhaps on the scale of a few decades.

As a preliminary study, the main elements of the interannual GAO may be considered timely in conjunction with the well defined phases of the ENSO variability, for which the GAO is the background or the primary process, as it has been already mentioned. In other words, in such cases it is appropriate to use the common ENSO (El Niño and La Niña) indices as the markers of the GAO. Further, by the analogy with the various regional oscillations, which are identified by their conventional indices, an appropriate index for the GAO that assesses the characteristic deformation of the planetary atmospheric pressure fields in the equatorial-tropical region (30° S–30° N, 60° W–180°) could be described by the formula:

$$I_1(t) \sim P'(z_0, t) = \int_{\varphi_1}^{\varphi_2} \int_{\lambda_1}^{\lambda_1} \delta P^*(\varphi, \lambda, z_0, t)\, d\varphi\, d\lambda \tag{3}$$

The evolution of two climatic indices shown in Fig. 3 makes it possible to analyze the intensity of the ENSO events on the interannual time scale in the period under study. In particular, on the basis of these backgrounds one can easily distinguish 15 events of El Niño and the same number of La Niña events. This follows directly from the main positive and negative extremes ($T > +0.5$ °C; $T < -0.5$ °C) on the average annual temperature curve (thin red line). The less significant positive extremes on that line are likely related to the weaker events named El Niño-Modoki [19]. Taking these extremes into account we find that the average occurrence interval of the El Niño was about 3.7 years, which is consistent with the existing concept about the temporal characteristic of the phenomenon [20]. This is also related to the La Niña events with the minus sign responding to the negative temperature anomaly drawn by the phenomenon.

The clearly seen peaks of intense ENSO index indicate two outstanding El Niño events of 1982–1983 and 1997–1998, which are remarkable due to their absolute intensity over the past century. In addition, it should be mentioned that this graph also shows the period from the mid-1970s until the beginning of the 20th century marked by the more frequent and intense El Niño events.

Let us focus attention on the comparable change of the interannual GAO index (thin blue line). The time evolution of this index corresponds to the temperature curve, which is reflecting the ENSO signal. Such a correspondence indicates clearly the existence of the abovementioned correlation between the GAO and El Niño processes.

Both our estimates [16] and the independent study results [12] show that a sequence of the considered events with a system of their causal relationship is the following. The global hydrometeorological anomalies, or in our terminology the GAO signals, appear long prior to the onset of the ENSO events in the Pacific region. A closer look at Fig. 3 may reveal the presence of the positive intermediate

phase shift between some extremes of the ENSO and GAO indices. This evidence strongly supports our hypothesis that the interannual GAO anticipates the occurrence of the El Niño acting as its trigger mechanism.

Figure 3 shows that along with the general similarity of the two indices behavior they have, at least, one significant difference, namely, the number of extreme points of the interannual GAO index exceeds the corresponding marks of the ENSO. This could generally mean that the frequency parameters of the GAO and ENSO may vary. In such case, the situation is not excluded when the GAO occurs more often than the ENSO events. However, in any case, as it is clearly seen in Fig. 3, the signals of all actual ENSO events, distinguished by its index, certainly correspond to the specific phase of the interannual GAO. Therefore, all events like El Niño and La Niña should not be considered separately just as anomalous regional features in the dynamics of the regional Pacific climate system only, although, this approach is sometimes applied to these phenomena. Nevertheless, according to the previous discussion, it is not correct to consider these effects beyond the concept of the GAO. The fact based on the GAO analysis indicates that the signals of the ENSO and other regional climatic disturbances should not be interpreted otherwise than the factors of the lower, secondary level.

The bold lines in Fig. 3 provide essential information on the properties and characteristics of the inter-decadal GAO. First of all, smooth transformation of the data on the ENSO and GAO indexes, which were converted into the curves of almost the same form, is worth noting. On this occasion we would assume that such a result is an actual verification of the adequacy of our understanding of the climate variability based on the correlations between the signals in the fields of completely different hydrometeorological characteristics, such as temperature and atmospheric pressure. These fields indirectly represented here by the corresponding indices of the GAO and ENSO are distinguished from the quasi-cyclical fluctuations with an average estimated period of 16 years.

The traces of incompletely filtered interannual GAO (blue curves) and ENSO (red lines) signals are revealed over the background of the lower frequency fluctuations. The fact that the smoothed ENSO index is similar to the inter-decadal GAO signal provides evidence that these events are caused by correlated physical processes. Thus, as it is shown in Fig. 3, the smoothed curves represent, according to our definition, nothing but the actual inter-decadal GAO. It is clear that the oscillation frequency characteristics require a further clarification due to the lack of statistical reliability of its current assessment. Incidentally, we note that even lower-frequency harmonics of such type of oscillations in the range of periods about 35–50 years were revealed in our study of the climate change in the North Atlantic region [11].

Conclusions

It is well known that during El Niño notable changes in the atmospheric and oceanic processes appear in different parts of the Earth. In some places they occur synchronously with the El Niño phenomenon, in the other regions they appear with a certain time shift. In general, these phenomena are defined as a remote response to the El Niño events.

We note that, in addition to El Niño, researchers use a very large number of different regional atmospheric oscillations such as the North Atlantic, Arctic, Antarctic, East Pacific and many others. Many of these oscillations contain the modes, which are synchronized with the events of El Niño [21]. The causes of the possible changes in the weather regime in different parts of the Earth during the events of El Niño can be understood in general even from the analysis of the peculiarities of appropriate global atmospheric pressure field, like the one shown in Fig. 1. Let us consider here just one example. From this figure it may be clarified, in particular, that cooling of the Barents Sea region during El Niño occurs due to the development of anomalously cold easterly winds that bring chill and prevent the spread of the warm currents from the west. Many scientists studying the remote atmosphere regime changes during El Niño use to apply the paradigm of the El Niño remote response. They believe that the key cause of these changes is the event of El Niño whose impact is more or less manifested very far from the source of the signal.

Anticipated physical mechanisms of the El Niño signal transmission over long distances were assumed in many papers [13, 22–24]. In almost all of the cases the opinion of the authors is based on the concept of the atmospheric or oceanic "bridges" suggested for the fast propagating Rossby and Kelvin waves. Actually, some changes associated with El Niño in the atmosphere and ocean, but mainly within the Pacific could be explained on the basis of these waves [25]. We are not familiar with the works, in which transmission mechanisms of El Niño signal to more remote areas were considered. Neither we know the publications, which explain how under the influence of the regional El Niño signal in the field of the atmospheric pressure anomalies the clearly manifested structures covering the entire globe, i.e. of the scale significantly larger than the scale of the El Niño are immediately formed.

Thus, it is reasonable to propose a change in the paradigm of the Global Response of El Niño on the paradigm of El Niño occurrence as one of the manifestations of the real Global Oscillation. We understand the latter as the global structure in the field of the atmospheric pressure anomalies shown, in particular, in Fig. 1 (top panel). The indications of the atmospheric process anomalies shown in Fig. 1 should be reversed in the opposite phase of this Global process revealed in this paper. The event of La Niña usually appears in the Pacific region at this phase of the GAO.

The true causes of the recent global fluctuations in the dynamics of the climate system are not yet clear. We believe that it has, relatively speaking, a tidal character.

We consider that the Chandler wobble [26] with a period of 14 months and its overtones can trigger this disturbance. However, the final determination of the nature of this phenomenon is a task for the future research.

We adhere to the concept that the simplest explanation is the most correct. The main advantage in adopting the new paradigm is a significant simplification of the answers to many questions related to the problem of long-term effects. In particular, there is no need to invent anything regarding telecommunication mechanisms in the atmosphere. Under the new approach, all the features of the atmospheric circulation that occur on the Earth during El Niño, including the origin of the El Niño phenomenon, are explained by the existence of the global atmospheric oscillations. This approach is inter alia the most essential for the already mentioned synchronization of many regional oscillations with the events of El Niño.

It was shown in [27] that the global occurrences of the sea surface atmospheric pressure anomaly differences along with the events of El Niño and La Niña, as shown in Fig. 1 (top panel) are statistically significant. The significance of the major part of these differences is at the 90% significance level. This testifies objectively that the Global Oscillation shown in Fig. 1 can be considered statistically a very real phenomenon. Some papers devoted to the study of the El Niño event, have similar results [20, 28]. The main difference here is in the setting of priorities. Most of the researchers preferred to assume the El Niño event as the primary; they interpreted the characteristic changes that occur simultaneously in the remote areas of the Earth as the global response to these events. We believe that the interannual Global Atmosphere Oscillation is the primary one, whereas the El Niño event is a part of the GAO and one of its most notable consequences. We want to emphasize that one of the main arguments among the others in favor of our hypothesis is its simplicity and actuality in explaining many environmental changes remotely accompanying the El Niño events.

Thus, we attempted to suggest a conceptual description of the multi-scale global atmospheric oscillations, some distinct signs of which were implicitly detected in a number of previous studies. The principal result of this work is the finding that interannual and inter-decadal global atmospheric oscillations (GAO) are the main governing causes of the actual short-period variability in the current climate. First of all, the interannual GAO acts as a trigger mechanism for the El Niño events, anticipating their occurrence of the anomalous temperature, atmospheric pressure and wind conditions in the Indo-Pacific region. The crucial role of the inter-decadal GAO is related to the fact that this shift in the climate system dynamics may lead to a climate scenario transition both on the regional and, perhaps, on the global scale.

Acknowledgements This work was supported by the Russian Science Foundation, grant 14-50-00095.

References

1. Gribbin, D. (1980). *Climate change* (p. 360). Leningrad: Gidrometeoizdat.
2. Huybers, P., & Curry, W. (2006). Links between annual, Milankovitch, and continuum temperature variability. *Nature, 441,* 329–332.
3. Kondrat'ev, K. J. (1992). *Global climate* (p. 359). Saint-Petersburg: Nauka.
4. Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., et al. (Eds.). (2013). *Climate change 2013; IPCC, The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (p. 1535). Cambridge: Cambridge University Press.
5. Bronnimann, S., Stickler, A., Griesser, T., Fisher, A. M., Grant, A., Ewen, T., et al. (2009). Variability of large-scale atmospheric circulation indices for the Northern Hemisphere during the past 100 years. *Meteorologische Zeitschrift, 18*(4), 379–396.
6. Allan, R. J., & Ansell, T. J. (2006). A new globally-complete monthly historical gridded mean sea level pressure data set (HadSLP2): 1850–2004. *Journal of Climate, 19,* 5816–5846.
7. Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., & Morice, C. P. (2012). Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research, 117,* D05127. https://doi.org/10.1029/2011JD017139.
8. Byshev, V. I., Neiman, V. G., Romanov, Y. A., & Serykh, I. V. (2012). On El Niño as a consequence of the global oscillation in the dynamics of the Earth's climatic system. *Doklady Earth Sciences, 446*(1), 1089–1094.
9. Neelin, J. D., Battisti, D. S., Hirst, A. C., Jin, F.-F., Wakata, Y., Yamagata, T., et al. (1998). ENSO theory. *Journal of Geophysical Research, 103*(C7), 14261–14290.
10. Byshev, V. I., Neiman, V. G., Romanov, Y. A., & Serykh, I. V. (2009). On the spatial nonuniformity of some parameters of the modern climate. *Doklady Earth Sciences, 426*(1), 705–709.
11. Byshev, V. I., Neiman, V. G., Romanov, Y. A., & Serykh, I. V. (2011). Phase variability of some characteristics of the present-day climate in the northern Atlantic region. *Doklady Earth Sciences, 438*(2), 887–892.
12. Kozlenko, S. S., Mokhov, I. I., & Smirnov, D. A. (2009). Analysis of the cause and effect relationships between El Niño in the Pacific and its analog in the equatorial Atlantic. *Izvestiya, Atmospheric and Oceanic Physics, 45*(6), 704–713.
13. Steinman, B. A., Mann, M. E., & Miller, S. K. (2015). Atlantic and Pacific multidecadal oscillations and Northern Hemisphere temperatures. *Science, 347*(6225), 988–991.
14. Wyatt, M. G., Kravtsov, S., & Tsonis, A. A. (2012). Atlantic multidecadal oscillation and Northern Hemisphere's climate variability. *Climate Dynamics, 38,* 929–949. https://doi.org/10.1007/s00382-011-1071-8.
15. Zhang, L., Wang, C., & Wu, L. (2012). Low-frequency modulation of the Atlantic warm pool by the Atlantic multidecadal oscillation. *Climate Dynamics, 39,* 1661–1671. https://doi.org/10.1007/s00382-011-1257-0.
16. Byshev, V. I., Neiman, V. G., Romanov, Y. A., & Serykh, I. V. (2011). On global scale of El Niño events within the Earth climatic system. *Recent problems of the Earth observation from space*. *8*(4), 200–208 (in Russian).
17. Anisimov, M. V., Byshev, V. I., Zalesny, V. B., Moshonkin, S. N., Neiman, V. G., Romanov, Y. A., et al. (2012). On inter-decadal variability of ocean and atmosphere characteristics in the Northern Atlantic. *Recent problems of the Earth observation from space*. *9*(2), 304–311 (in Russian).
18. Petrosjants, M. A., Semenov, E. K., & Gushchina, D. J. (2005). *Atmospheric circulation in tropics: Climate and its variability* (p. 640). Moscow: Maks Press.
19. Kao, H.-Y., & Yu, J.-Y. (2009). Contrasting Eastern-Pacific and Central-Pacific types of ENSO. *Journal of Climate, 22,* 615–632. https://doi.org/10.1175/2008JCLI2309.1.

20. Sidorenkov, N. S. (2002). *Physics of the Earth's rotation instabilities* (p. 384). Moscow: Nauka, Fizmatlit.
21. Liu, Z., & Alexander, M. (2007). Atmospheric bridge, oceanic tunnel, and global climate teleconnections. *Reviews of Geophysics, 45,* R62005.
22. Bronnimann, S. (2007). Impact of El Niño-Southern Oscillation on European climate. *Reviews of Geophysics, 45*, RG3003. https://doi.org/10.1029/2006RG000199.
23. Li, S., & Luo, F.-F. (2013). Lead-lag connection of the Atlantic multidecadal oscillation (AMO) with East Asian surface air temperatures instrumental records. *Atmospheric and Oceanic Science Letters, 6*(3), 138–143.
24. Varotsos, C. A. (2013). The global signature of the ENSO and SST-like fields. *Theoretical and Applied Climatology, 113,* 197–204. https://doi.org/10.1007/s00704-012-0773-0.
25. Liu, P., & Sui, C.-H. (2014). An observational analysis of the oceanic and atmospheric structure of global-scale multi-decadal variability. *Advances in Atmospheric Science, 31,* 316–330.
26. Serykh, I. V., & Sonechkin, D. M. (2016). On the influence of the polar tide on El Niño. Recent problems of the Earth observation from Space. *13*(2), 44–52 (in Russian).
27. Byshev, V. I., Neiman, V. G., Romanov, Y. A., Serykh, I. V., & Sonechkin, D. M. (2016). Statistical significance and climatic role of the global atmospheric oscillation. *Oceanology, 56*(2), 165–171.
28. de Viron, O., Dickey, J. O., & Ghil, M. (2013). Global modes of climate variability. *Geophysical Research Letters, 40,* 1832–1837. https://doi.org/10.1002/grl.50386.

Influence of the Current Field Non-stationarity and the Non-simultaneity of Hydrographic Measurements on ADCP-based Transport Estimates

R. Yu. Tarakanov

Introduction

The horizontal absolute geostrophic velocities in the entire water column are determined by the following equations:

$$u(z) = -\frac{g}{f}\frac{d\zeta}{dy} + \frac{g}{f\rho_0}\int_z^0 \frac{d\rho(z)}{dy}dz, \tag{1a}$$

$$v(z) = \frac{g}{f}\frac{d\zeta}{dx} - \frac{g}{f\rho_0}\int_z^0 \frac{d\rho(z)}{dx}dz; \tag{1b}$$

here u, v are the zonal and meridional components of the current velocity, ζ is the elevation of the ocean surface relative to the geopotential surface, ρ is the density in situ, ρ_0 is the mean density of the ocean, g is the acceleration due to gravity, f is the Coriolis parameter, the z-axis is directed upwards. The first term in (1a, 1b) is the absolute geostrophic current velocity at the ocean surface. Hereinafter, we shall consider this variable as the barotropic component of the absolute geostrophic current. The second term in (1a, 1b) is depth-dependent (baroclinic) geostrophic velocity profile with zero value at the ocean surface. The dependence of this profile on the horizontal density gradient determines the simplicity of calculating the baroclinic current component from the data of temperature, salinity (conductivity) and pressure, and consequently water density at the stations of transoceanic hydrographic sections or surveys. The barotropic components of the geostrophic

R. Yu.Tarakanov (✉)
Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
e-mail: rtarakanov@gmail.com

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_23

currents in the practice of research are estimated on the basis of the satellite altimetry data [9] using the data of current velocity at stations measured using Lowered Acoustic Doppler Profiler (LADCP) [1, 2, 5, 6, 10, 11, 13] or by Shipboard Acoustic Doppler Current Profiler (SADCP) [12].

The estimation of the correctness of the absolute geostrophic current calculation if the barotropic component determined on the basis of the ADCP data can be divided into two tasks:

1. Estimation of the accuracy of the absolute geostrophic current approximation at the segments of the section (survey) based on the measured profiles at the stations corresponding to these segments.
2. Estimation of the distortion of the field of geostrophic currents associated with the combination of non-stationarity of this field and non-simultaneity of the measurements over a section (survey). This task can, in turn, be divided into two more tasks:
   2a. Estimation of the distortion of the baroclinic current component;
   2b. Estimation of the distortion of the barotropic current component.

In this contribution, we will consider the latter of the above mentioned tasks applied to particular hydrographic sections in relation to the barotropic component of the current and the barotropic transport by the current. Here, we consider that the measured current profiles accurately approximate the real geostrophic current with respect to the determination of its barotropic component. We assume that the current (its barotropic and baroclinic components) is considered as the normal component of the actual current to the sections. In such a formulation of the task, its solution does not require the involvement of field data in the entire ocean column, but can be reduced to the investigation of the Sea Level Anomaly (SLA) at the ocean surface, which is measured by modern satellite altimeters. Note that the isolines of the level surface (absolute dynamic topography) determined by the first term in Eq. (1a, 1b) represent the streamlines of the absolute geostrophic currents at the ocean surface. The SLA at a particular location in the ocean is the instantaneous deviation of the level from the mean value over the time period of observations.

The structure description of applied data is essential for the theoretical justification of this study. Hence, first, we provide the information about these data (section "Data"), and then describe the method of the data analysis (section "Method of Processing"). Section "Error Estimates" describes the error sources and gives quantitative estimates of the errors of calculations of the barotropic component of the current and barotropic transport. In section "Discussion", we analyze the results of calculations performed according to the method described in section "Method of Processing". Section "Conclusions" summarizes the conclusions.

Data

This contribution is mainly based on the daily data (digital maps) of the SLA (DT-Global-MSLA-Upd product) with a grid of $1°/4 \times 1°/4$, which is produced by the French CLS (Collected Localization Satellites) agency using satellite altimetry observations. The data are available in the Internet (DT Global-MADT-Upd product, http://www.aviso.altimetry.fr). Each daily map corresponds to 00:00 GMT. In addition to these data, the data on the Formal Mapping Error (FME) and Absolute Dynamic Topography (ADT) (DT-Global-MADT-Upd product) published by the CLS agency with the same resolution were also used. The FME characterizes the accuracy of satellite measurements and the mathematical self-evaluation of the procedure for interpolating data from satellite tracks onto the ocean surface in space and time. The ADT data are the sum of the SLA and the mean ADT version of CNES-CLS13 produced by the CLS Space Oceanography Division and distributed by AVISO, with support from CNES (http://www.aviso.altimetry.fr/) [15]. The data on ADT, SLA, and FME are given in centimeters.

The analysis was carried out for the tracks of a number of hydrographic sections occupied from one continental slope to the other in the era of the altimetric observations of the ocean surface. These are quasi-zonal sections AR05 and A05 in the North Atlantic, the tracks of which pass nominally along 60° N and 25° N, sections P02 along 30° N in the Pacific, as well as sections carried out in the Southern Hemisphere, nominally along 30° S: A10 in the Atlantic, I05 in the Indian, and P06 in the Pacific. In the Southern Ocean the following tracks of quasi-meridional sections in different sectors of the ocean were analyzed: SR01, A21 and a series of the Russian sections in the Drake Passage, SR02 and I06 in the region south of Africa, I09, SR03, P11A south of Australia, and P19 in the southeastern sector of the Pacific Ocean. More detailed information about these sections and their locations are given in Table 1 and Fig. 1. The WOCE and CLIVAR data are available at https://cchdo.ucsd.edu.

The bottom relief along the hydrographic section lines was determined from the data of the digital array of the bottom topography [17], V.18.1 (http://topex.ucsd.edu/cgi-bin/get_data.cgi). The resolution of the array is $1/\cos\varphi$ of angular minutes by latitude and 1 minute by longitude; φ is the latitude.

Method of Processing

The nonstationarity of the current field in combination with the non-simultaneity of measurements over the section leads to a distorted pattern of the field of currents recovered from these data. The ADCP-measurements of instantaneous velocity profiles at the ocean surface at specific points are actually equivalent to the measuring of instantaneous sea level gradients at these points. The level difference thus accumulated along the line of the section differs from the real one because of the

Table 1 Information about the sections in the World Ocean (see Fig. 1) studied in this contribution

Section	R/V	Dates	Measurement durations (days)	Number of stations	Typical distance between stations in the open sea (miles)
Drake_2003	RV "Akademik Sergey Vavilov"	11–15.12.2003	5	25	20
Drake_2007	RV "Akademik Ioffe"	13–19.11.2007	9	53	10
Drake_2010 Jan	RV "Akademik Ioffe"	02–10.01.2010	9	54	10
Drake_2010 Oct	RV "Akademik Sergey Vavilov"	05–12.11.2010	8	47	10
Drake_2011	RV "Akademik Ioffe"	28.10–04.11.2011	8	45	10
A21 1999	RRS "James Clark Ross"	18–26.03.1999	9	43	15
SR01 1993	RRS "James Clark Ross"	21–26.11.1993	5	30	20
SR01 1994	RRS "James Clark Ross"	16–21.11.1994	6	27	20
SR01 1996	RRS "James Clark Ross"	15–20.11.1996	6	29	20
SR01 1998	RRS "James Clark Ross"	30.12.1997–07.01.1998	9	50	10
SR01 2013	RRS "James Clark Ross"	19–27.03.2013	8	30	20
SR01 2015	RRS "James Clark Ross"	13–18.03.2015	5	30	20
SR01 2016	RRS "James Clark Ross"	06–11.01.2016	6	28	20
SR02 2001	RV "PolarStern"	20.12.2000–11.01.2001	22	64	30–60
I06 2008	RV "Roger Revelle"	05.02–08.03.2008	32	92	30
I09 1995	RV "Knorr"	01–18.01.1995	17	61	30
I09 2004	RV "Aurora Australis"	24.12.2004–15.01.2005	22	68	30
I09 2012	RV "Aurora Australis"	20.01–10.02.2012	21	70	30

(continued)

Table 1 (continued)

Section	R/V	Dates	Measurement durations (days)	Number of stations	Typical distance between stations in the open sea (miles)
SR03 1994	RV "Aurora Australis"	02–16.01.1994	14	57	30
SR03 1995	RV "Aurora Australis"	18.07–01.08.1995	15	54	30
SR03 1996	RV "Aurora Australis"	30.08–21.09.1996	22	57	30
SR03 2001	RV "Aurora Australis"	29.10–28.11.2001	30	61	30
SR03 2008	RV "Aurora Australis"	28.03–15.04.2008	18	59	30
SR03 2011	RV "Aurora Australis"	04–18.01.2011	14	54	30
P11A 1993	RV "Aurora Australis"	04–30.04.1993	26	62	30
P19 1993	RV "Knorr"	10.01–01.03.1993	61	51	30
AR05 1997	RV "Professor Shtokman"	14–23.10.1997	9	27	60
AR05 2002	RV "Akademik Mstislav Keldysh"	17–27.08.2002	10	39	30
AR05 2003	RV "Akademik Ioffe"	27.06–07.07.2003	10	47	15–30
AR05 2004	RV "Akademik Ioffe"	20–28.06.2004	8	38	30
AR05 2005	RV "Akademik Ioffe"	23–30.06.2005	9	38	30
AR05 2006	RV "Akademik Ioffe"	03–12.07.2006	10	53	20
AR05 2007	RV "Akademik Ioffe"	30.06–12.07.2007	12	80	10
A05 1998	RV "Ronald H. Brown"	24.01–23.02.1998	31	130	30–45
A05 2004	RV "Discovery"	05.04–09.05.2004	35	123	40–45
A05 2010	RV "Discovery"	07.01–15.02.2010	40	134	30–45

(continued)

Table 1 (continued)

Section	R/V	Dates	Measurement durations (days)	Number of stations	Typical distance between stations in the open sea (miles)
A05 2011	RV "Sarmiento de Gamboa"	28.01–14.03.2011	45	167	30
A05 2015	RV "Discovery"	10.12.2015–20.01.2016	42	134	40
A10 1993	RV "Meteor"	30.12.1992–28.01.1993	30	111	27–45
A10 2003	RV "Mirai"	07.11–02.12.2003	25	111	30–45
A10 2011	RV "Ronald H. Brown"	28.09–29.10.2011	32	120	25–40
I05 2002	RV "Charles Darwin"	04.03–14.04.2002	41	133	20–50
I05 2009	RV "Roger Revelle"	24.03–12.05.2009	49	189	30
P02 1994	RV "Shoyo"	08.01–10.02.1994	33	63	60–140
P02 2004	RV "Melville"	16.06–27.08.2004	72	187	30–35
P02 2013	RV "Melville"	22.03–01.06.2013	71	159	30–60
P06 2003	RV "Mirai"	03.08–12.10.2003	70	235	30–40
P06 2009	RV "Melville"	22.11.2009–09.02.2010	80	250	30–40

current field variation during this time. The accumulation of distortion of the barotropic component of the current and the barotropic transport corresponding to it can be, in principle, estimated from the satellite altimetry. A distortion constituent of the barotropic transport, which is not taken into account using these data, also exists. This constituent is related to the displacement of isohypses (streamlines of the geostrophic current at the ocean surface) relative to the isobaths. Due to this displacement during the execution of the section, the instantaneous barotropic transport varies within certain limits. It is clear that, depending on the restriction on the total transport across the section, this variability should be completely or partially compensated by a change in the instantaneous baroclinic transport. The latter cannot be accounted directly from the satellite altimetry data. Note that the

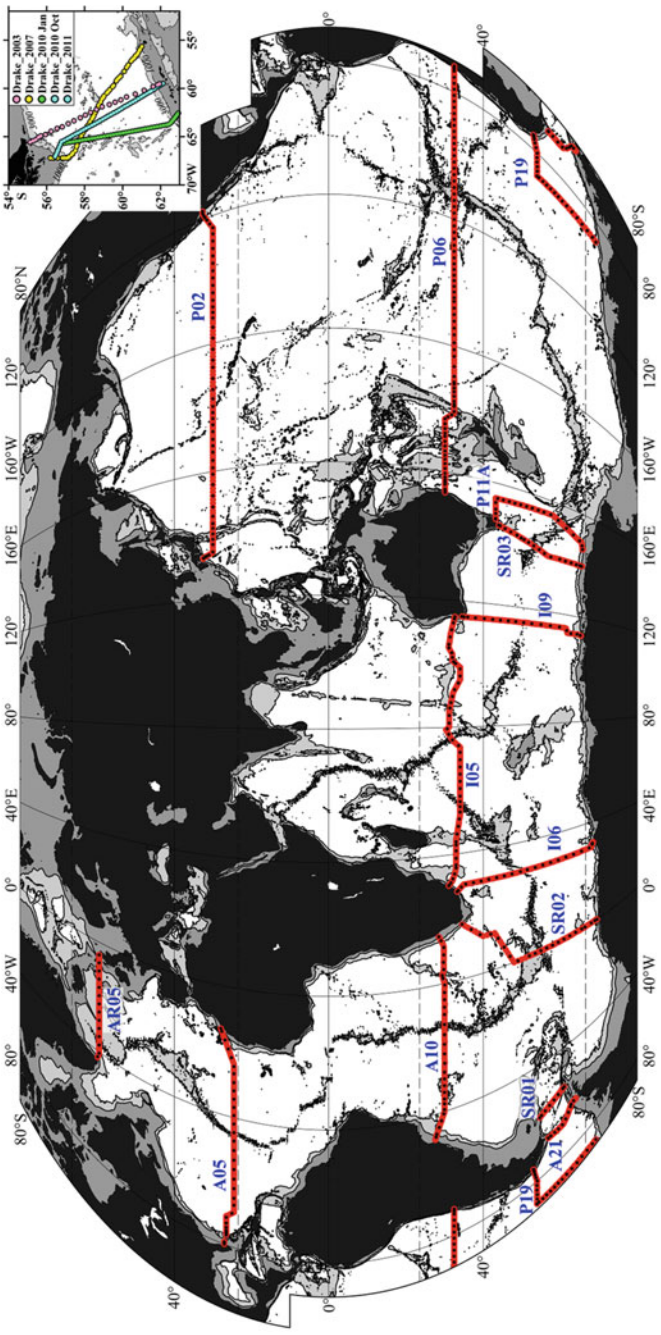


Fig. 1 Locations of the section lines in the World Ocean (shown in red). Information on the sections is given in Table 1. The areas of the bottom with depths shallower than 3000 and 1000 m are shaded with light and dark gray tones, respectively. Dots with different colors in the inset show the lines of the Russian sections in 2003–2011

barotropic transport over the section may appear beyond the variability range of its instantaneous values. This fact indicates that the total transport derived from the barotropic transport deviation differs from the real values by unknown quantity, which requires an assessment of the baroclinic transport variability.

In order to estimate quantitatively the distortions of the barotropic component of the current, let us consider the level of the ocean surface ζ along the section line as a function of two variables: time t and distance along section x; the positive direction along the x axis corresponds to the direction of the section:

$$\zeta = \zeta(x, t). \tag{2}$$

The total differential of this function, $d\zeta$ is written as:

$$d\zeta = \left(\frac{\partial \zeta}{\partial x}\right)_t dx + \left(\frac{\partial \zeta}{\partial t}\right)_x dt. \tag{3}$$

We perform a transition from continuous function ζ to its approximation over the section by the finite differences. Then, the total differential can be written as:

$$\Delta\zeta = \zeta(x_{i+1}, t_{i+1}) - \zeta(x_i, t_i), \tag{4}$$

where the subscript corresponds to the station number. In this approach, each station is assigned a single point in time and a single point in the ocean. However, in the practice, the measurements at each station take a certain time interval and, in addition, the drift of the vessel takes place during the station. The resulting variability of the level can be interpreted as a variation (determination error) of the mentioned parameter at the station. This problem will be considered separately in section "Error Estimates".

In the representation of the level difference (and hence of the barotropic component of the current) between the stations according to formula (4), there is an exact physical correspondence in time between the ocean level and the CTD profiles measured at these points. In this case, from the point of view of the contribution of the baroclinic and barotropic components of the current to the total transport for a given pair of stations, we also obtain an exact physical relation between these components. We note that because of the nonstationarity of the current field and the fact that the measurements at different stations are made at different moments of time, the calculated profiles of absolute geostrophic velocities do not exist in reality, i.e. they are virtual. A similar method for calculating absolute geostrophic currents was used in [9].

We add and subtract $\zeta(x_{i+0.5}, t_i)$ and $\zeta(x_{i+0.5}, t_{i+1})$ on the right side of formula (4), which are the values of the level in the middle of the space between the i-th and $i + 1$-st stations at the i-th and $i + 1$-st time moments, respectively. We obtain

$$\zeta(x_{i+1}, t_{i+1}) - \zeta(x_i, t_i) \equiv$$
$$\equiv (\zeta(x_{i+1}, t_{i+1}) - \zeta(x_{i+0.5}, t_{i+1})) + (\zeta(x_{i+0.5}, t_{i+1}) - \zeta(x_{i+0.5}, t_i)) + (\zeta(x_{i+0.5}, t_i) - \zeta(x_i, t_i)). \tag{5}$$

Summing over all the spaces between the stations (from the 1-st to the n-th), we get:

$$\zeta(x_n, t_n) - \zeta(x_1, t_1) \equiv \sum_{i=1,n-1} (\zeta(x_{i+1}, t_{i+1}) - \zeta(x_i, t_i)) \equiv \sum \Delta\zeta = \sum \Delta\zeta_x + \sum \Delta\zeta_t, \tag{6}$$

where

$$\sum \Delta\zeta_x = (\zeta(x_{1.5}, t_1) - \zeta(x_1, t_1)) + \sum_{i=2,n-1} (\zeta(x_{i+0.5}, t_i) - \zeta(x_{i-0.5}, t_i)) + (\zeta(x_n, t_n) - \zeta(x_{n-0.5}, t_n)), \tag{6a}$$

$$\sum \Delta\zeta_t = \sum_{i=1,n-1} (\zeta(x_{i+0.5}, t_{i+1}) - \zeta(x_{i+0.5}, t_i)). \tag{6b}$$

Here, $\sum \Delta\zeta_x$ is the sum of the level differences at the intervals between adjacent interstation midpoints at fixed time moments, $\sum \Delta\zeta_t$ is the sum of the level deviations in time in the middle of the spaces between stations, and n is the number of stations. In the sketch shown in Fig. 2 explaining formula (6), the bold line shows the ocean level curve consisting of $\Delta\zeta_x$ and $\Delta\zeta_t$, the dashed line is the curve formed of $\Delta\zeta$. In the sum over the entire section, these two curves give the same level difference.

The $\Delta\zeta_x$ value at the i-th interval of the section approximates the barotropic component of the geostrophic current at fixed time corresponding to the moment of the i-th station. In the physical sense, this component is absolutely similar to the barotropic component of the geostrophic current measured using ADCP. Thus, the level deviation in time in the middle of the spaces between stations determines the physical difference between the calculations of the barotropic component of the absolute geostrophic current from the satellite altimetry and ADCP data.

In practice, it is not possible to determine the barotropic component of the current directly from the ADCP velocity measurements at one station, since the measured baroclinic profile contains unknown non-geostrophic components (e.g., purely drift current, inertial oscillations). In addition, the measurements in the upper several tens of meters are usually obtained with a very large error or these data are not available in the case of shipboard ADCP. Therefore, the barotropic component based on the ADCP measurements is determined for the baroclinic geostrophic profile calculated from the CTD data from two neighboring adjacent stations, which usually consists of the following procedure [1, 6, 11, 13]. Initially, the measured velocity profiles of two neighboring stations are averaged at each level up to the greatest common depth of the CTD and ADCP measurements at these stations. Then, the depth averaged values of the velocity for this averaged profile $\langle u_{adcp} \rangle$ and

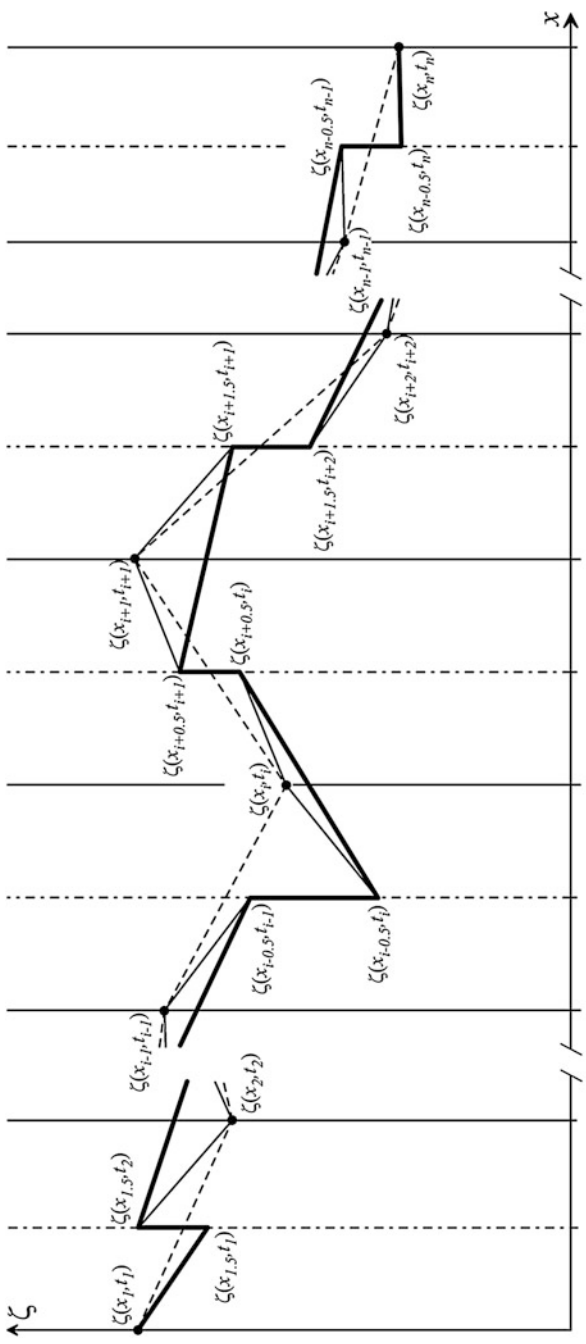


Fig. 2 Representation of the ocean level along the section by two finite-difference schemes. The dashed line shows the level of the ocean at the stations, the bold line is the level restored in the middle of the spaces between stations. Vertical bold lines correspond to the stations; dashed lines correspond to the midpoints of spaces between stations

of the baroclinic geostrophic profile $\langle u_{ctd} \rangle$ are calculated, and the difference between these average values $\langle u_{adcp} \rangle - \langle u_{ctd} \rangle$ is calculated. Finally, the absolute geostrophic velocity at each depth (including the barotropic component, as the value at the ocean surface) is defined as the sum of the baroclinic geostrophic velocity and the mentioned difference. With regard to the described procedure, it is necessary to note the following:

1. The procedure is based on the assumption of an averaged geostrophic character of the ADCP measurements, i.e. it is assumed that non-geostrophic perturbations in the velocity profile are reduced to zero when averaging over the entire profile.
2. The calculation of the average arithmetic profile from the ADCP measurements at two stations does not change the essence of the physical difference between the determination of the barotropic component from the altimetry and the ADCP data and does not change the quantitative differences between these definitions in the entire section.
3. Similarly to the case of the determination of the barotropic component from the altimetry data the resulting profile of the absolute geostrophic velocity does not actually exist because of the non-simultaneity of measurements at stations.

Let us consider from the point of view of estimating the total transport through a hydrographic section the two methods described above for calculating the barotropic component of the absolute geostrophic current, i.e. by means of the total differential $\Delta\zeta$ (calculation from the altimetry data) and using $\Delta\zeta_x$ (analog of the ADCP measurements). Suppose that such a section is occupied from shore to shore in the ocean with a flat bottom and vertical walls. To simplify the discussion, we fix level ζ in time and the vertical density profiles at the ends of the section, and assume that g and f are constant over the entire section. Then, due to the linearity of the geostrophic relationships (1a, 1b), the total transport not only through the section line, but through any trajectory connecting the points of the beginning and the end of the section in time and space of the ocean, will be the same, i.e. it will not depend on the intermediate vertical density distributions along such a trajectory, and on the configuration of level ζ and its time variability. Besides, the barotropic and baroclinic transports will also be constant separately. The condition of such constancy is satisfied by the method of calculating the level over a particular section using total differential $\Delta\zeta$. The calculation using $\Delta\zeta_x$ will differ from it by $\sum \Delta\zeta_t$ by virtue of (6), which depends, in principle, on time and trajectory. Temporal variability of the level at the ends of the section, i.e. the refusal to fix ζ at these points, as well as the dependence of g and f on latitude, does not change the essence of the described difference. Thus, to estimate the transport across the section the recovery of the ocean level using total differential $\Delta\zeta$ is physically correct, since in this case, exact correspondence in time is gained between the ocean level and the CTD profile at the stations. In this sense, the value of $\sum \Delta\zeta_t$ is a constituent of the distortion of the barotropic component of the current, which can be taken into account and estimated using the satellite altimetry data.

The real ocean is not flat and the walls are not vertical. In this case, the instantaneous barotropic transport across the section line changes with time due to the displacement of the level isolines (stream lines of the absolute geostrophic velocity at the ocean surface) relative to the isobaths. When calculating in finite differences, the transport also depends on the particular scheme of the section splitting into segments. Instantaneous barotropic transport calculated from the level values in the middle of the spaces between stations, i.e. the transport approximating ADCP measurements, can be written as:

$$Q_{adcp}(t) = \sum_{i=1,n} q_{adcp}(x_i,t) \equiv (\zeta(x_n,t) - \zeta(x_{n-0.5},t))\frac{g_n\bar{h}_n}{f_n} + (\zeta(x_{1.5},t) - \zeta(x_1,t))\frac{g_1\bar{h}_1}{f_1} + \\ + \sum_{i=2,n-1} (\zeta(x_{i+0.5},t) - \zeta(x_{i-0.5},t))\frac{g_i\bar{h}_i}{f_i} \quad , \tag{7}$$

where g_i, f_i are the acceleration due to gravity and the Coriolis parameter at the point of the i-th station, $\bar{h}_i$ is the depth averaged over the interval of the section corresponding to the i-th station (over the i-th interval between the interstation midpoints of the spaces), $q_{adcp}(x_i,t)$ is the barotropic transport in this interval of the section. The sign of the transport summation by stations is determined in (7) conventionally. For example, for quasi-zonal sections, it is convenient to choose the positive direction to the left of the direction of the section if it runs from west to east. Then, the positive transport values correspond to the flow to the north, the negative transport values correspond to the flow to the south. For example, in the case of quasi-meridional sections it is convenient to select the positive direction to the left of the ship motion along the section from west to east. Then, the positive values correspond to the easterly flow, the negative values correspond to the westerly flow.

The barotropic transport across a particular section calculated according to the same scheme consists of elements $q_{adcp}(x_i,t_i)$ corresponding to the moments of the execution time of the stations:

$$Q_{adcp} = \sum_{i=1,n} q_{adcp}(x_i,t_i) \equiv (\zeta(x_n,t_n) - \zeta(x_{n-0.5},t_n))\frac{g_n\bar{h}_n}{f_n} + (\zeta(x_{1.5},t_1) - \zeta(x_1,t_1))\frac{g_1\bar{h}_1}{f_1} + \\ + \sum_{i=2,n-1} (\zeta(x_{i+0.5},t_i) - \zeta(x_{i-0.5},t_i))\frac{g_i\bar{h}_i}{f_i} \quad . \tag{8}$$

The total barotropic transport (i.e. determined by the total differential of the ocean level) is represented in the following form:

$$Q_{alt} = Q_{adcp} + Q_t, \tag{9}$$

where

$$Q_t = \sum_{i=1,n-1} (\zeta(x_{i+0.5}, t_{i+1}) - \zeta(x_{i+0.5}, t_i)) \frac{g_{i+0.5}\bar{h}_{i+0.5}}{f_{i+0.5}}; \quad (10)$$

here g_i, f_i are the acceleration due to gravity and the Coriolis parameter in the middle of the space between the i-th and $i + 1$-st stations, $\bar{h}_{i+0.5}$ is the depth averaged over the mentioned space. Variable Q_t represents the total deviation of the transport in time along the section in the middle of the spaces between the stations and it is the component of the distortion of the barotropic flow that can be taken into account and estimated using the altimetry data.

If we subtract the transport across the section from instantaneous barotropic transport $Q_{adcp}(t)$, we obtain an anomaly of the barotropic transport approximating the ADCP measurements at the stations:

$$\Delta Q_{adcp}(t) = \sum_{i=1,n} \Delta q_{adcp}(x_i, t) \equiv (\Delta\zeta(x_n, t) - \Delta\zeta(x_{n-0.5}, t)) \frac{g_n\bar{h}_n}{f_n} + \\ + (\Delta\zeta(x_{1.5}, t) - \Delta\zeta(x_1, t)) \frac{g_1\bar{h}_1}{f_1} + \sum_{i=2,n-1} (\Delta\zeta(x_{i+0.5}, t) - \Delta\zeta(x_{i-0.5}, t)) \frac{g_i\bar{h}_i}{f_i}, \quad (11)$$

where $\Delta\zeta(x_{i\pm0.5}, t) = \zeta(x_{i\pm0.5}, t) - \zeta(x_{i\pm0.5}, t_i)$. It is important that the value of $\Delta Q_{adcp}(t)$ does not depend on the mean dynamic topography of the ocean and like the deviation in time Q_t over spaces between stations can be calculated on the basis of the SLA satellite data. To proceed to the discrete distribution of the barotropic transport in time it is necessary to replace t by t_j in formula (11).

Two-dimensional distribution $\Delta q(x_i, t_j)$ per unit length (top panel) and distribution $\Delta Q_{adcp}(t_j)$ (bottom panel) for the section carried out in the Drake Passage in October 2010 are shown in Fig. 3 as an example; time counts t_j correspond to the moments of the stations. The points on curve $\Delta Q_{adcp}(t_j)$ are the sum of elements $\Delta q(x_i, t_j)$ for fixed t_j. Diagonal elements $\Delta q(x_i, t_i)$ (i.e., when $i = j$) are zero. The range of values $\Delta Q_{adcp}(t_j)$ determines the variability of the barotropic transport approximating the ADCP measurements during the time of the section. An example in Fig. 3 also shows that the barotropic transport over the section can appear beyond the variability of the instantaneous barotropic transport during the time of the section.

Instantaneous barotropic currents calculated from the level values at stations can be written as:

$$Q'_{alt}(t) = \sum_{i=1,n-1} q_{alt}(x_{i+0.5}, t) \equiv \sum_{i=1,n-1} (\zeta(x_{i+1}, t) - \zeta(x_i, t)) \frac{g_{i+0.5}\bar{h}_{i+0.5}}{f_{i+0.5}}. \quad (12)$$

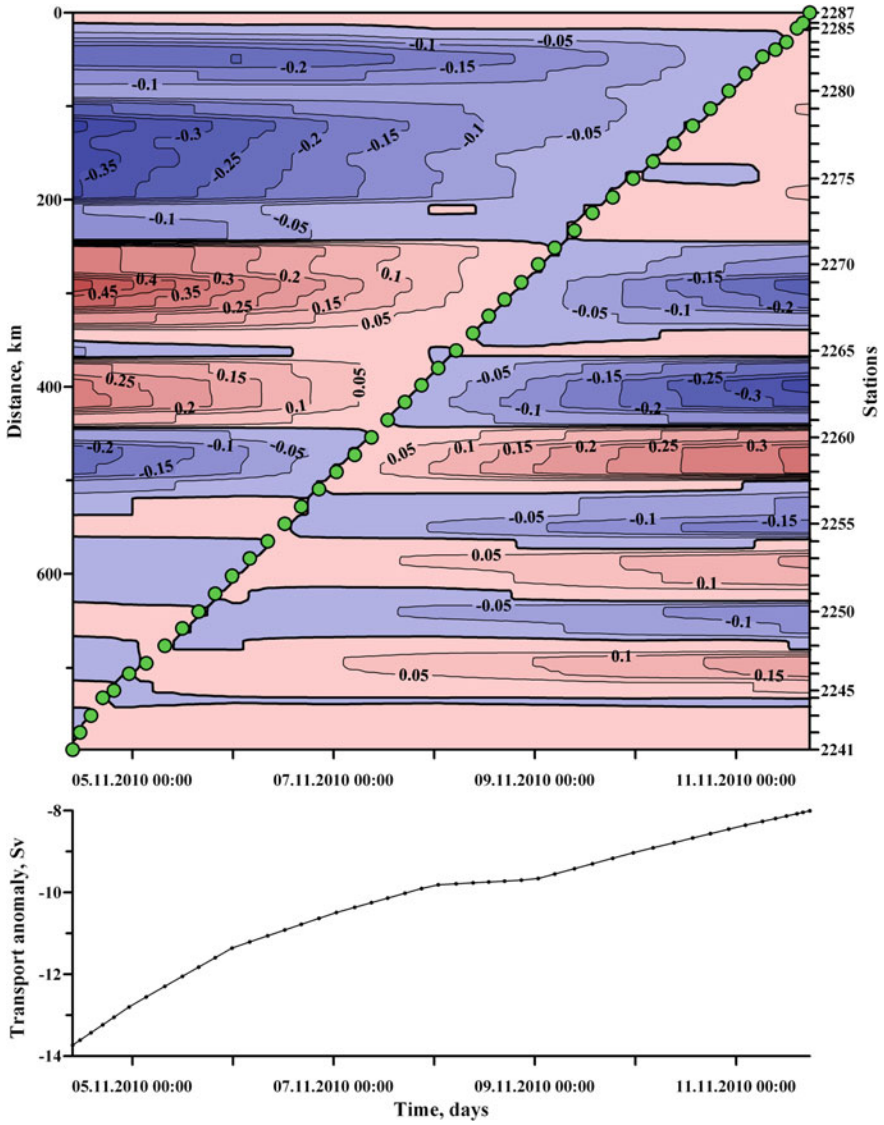


Fig. 3 Two-dimensional distribution $\Delta q(x_i, t_j)$ (see formula (11)) in Sv (1 Sv = 10^6 $m^3 s^{-1}$) per unit length (top panel) and anomaly distribution of the barotropic transport $\Delta Q_{adcp}(t_j)$ approximating the ADCP measurements at the stations (bottom panel) for the section in the Drake Passage in October 2010

In this scheme, the total barotropic transport over the section is written as:

$$Q'_{alt} = \sum_{i=1,n-1} q_{alt}(x_{i+0.5}, t_{i+0.5}) \equiv \sum_{i=1,n-1} (\zeta(x_{i+1}, t_{i+1}) - \zeta(x_i, t_i)) \frac{g_{i+0.5}\bar{h}_{i+0.5}}{f_{i+0.5}}; \tag{13}$$

the barotropic current, approximating ADCP measurements:

$$Q'_{adcp} = Q'_{alt} - Q_t, \tag{14}$$

here Q_t, as in the first scheme, is calculated from formula (10); the anomaly of the total barotropic transport is:

$$\Delta Q'_{alt}(t) = \sum_{i=1,n-1} \Delta q_{alt}(x_{i+0.5}, t) \equiv \sum_{i=1,n-1} (\Delta\zeta(x_{i+1}, t) - \Delta\zeta(x_i, t)) \frac{g_{i+0.5}\bar{h}_{i+0.5}}{f_{i+0.5}}, \tag{15}$$

where $\Delta\zeta(x_i, t) = \zeta(x_i, t) - \zeta(x_i, t_i)$. The value of $\Delta Q'_{alt}(t)$ does not depend on the mean dynamic topography of the ocean. The range of values $\Delta Q'_{alt}(t)$ determines the variability of the total barotropic transport during the time of the section.

The anomalies of the total barotropic $\Delta Q'_{alt}(t)$, and the barotropic transports $\Delta Q_{adcp}(t)$ approximating the ADCP-measurements are related as:

$$\Delta Q_{adcp}(t) = \Delta Q'_{alt}(t) + Q_t + \Delta Q - \Delta Q(t), \tag{16}$$

where variables

$$\Delta Q = Q'_{alt}(t) - Q_{adcp} - Q_t, \tag{17}$$

and

$$\Delta Q(t) = Q'_{alt}(t) - Q_{adcp}(t) \tag{18}$$

determine the quantitative difference between two finite difference schemes of the representation of the section track in the course of measurements over the section and for the instantaneous values of the transport in time, respectively. In this sense, they are important characteristics in estimating the accuracy of the calculation. Separately, these variables depend on the mean dynamic topography.

According to the previous theoretical justification, the calculations of the barotropic component of the current, barotropic transport, their temporal variability and values characterizing the accuracy of calculations over individual sections presented in Table 1 were performed. The data on SLA and ADT were interpolated linearly in space along the tracks of the sections according to two schemes: to the points of the stations and to the midpoints of the spaces between stations, and also in time for the

moments of the position of the measuring equipment at the bottom. The validity of exactly this choice of points and moments of stations is considered in section "Error Estimates" to estimate the errors. The calculation of the transport deviation Q_t due to the current non-stationarity and the non-simultaneity of hydrographic measurements was performed at the midpoints of the spaces between the stations. The bottom relief along the tracks of the sections necessary for estimating the transports was calculated by linear interpolation from the nodes of the digital data array to the section line over steps of approximately 0.1 km, so that the whole number of spaces between the station points would be located between the station points. The results of all the calculations are shown in Table 2 and in Figs. 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 and 15.

Error Estimates

Before proceeding to the analysis of the results of calculations it is necessary to estimate their accuracy. It has two constituents: the accuracy of the data and the accuracy of the calculation procedure.

The basis for estimating the calculation error based on the digital level maps of the ocean surface is the **FME (Formal Mapping Error)**, $\delta\zeta$. The distributions of this error over the sections are shown in Figs. 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 and 15. If we interpret it as the root-mean-square deviation of the level at a point, then the error in calculating the level difference between two points is defined as $\sqrt{\delta\zeta_1^2 + \delta\zeta_2^2}$. In the ocean with a flat bottom and depth h, the error of the barotropic transport along any trajectory connecting these points (provided that g and f are constants) is defined as:

$$\delta Q = \frac{gh}{f}\sqrt{\delta\zeta_1^2 + \delta\zeta_2^2}; \tag{19}$$

if we constrict the trajectory to a point, we obtain formally the variation of the transport at this point:

$$\delta Q = \sqrt{2}\frac{gh}{f}\delta\zeta. \tag{20}$$

The error of the barotropic transport over the section in the real ocean in finite differences can be represented as follows:

$$\langle \delta Q \rangle = \sqrt{\frac{\sum_{j=1,m} 2\left(\frac{g_j h_j}{f_j}\delta\zeta_j\right)^2 w_j}{\sum_{j=1,m} w_j}}; \tag{21}$$

Table 2 Estimates of the level deviation, transport, and errors in calculations over the investigated sections (see Table 1)

Section	FBTE $\langle \delta Q \rangle$ (Sv), see (21)	Deviation $\sum \Delta\zeta_t$ (cm), see (6b)	Deviation Q_t (Sv), see (10)[b]	Range $\Delta Q_{adcp}(t)$ (Sv), see (11)[c]	Range $\Delta Q'_{alt}(t)$ (Sv), see (15)[c]	Discrepancy ΔQ (Sv), see (17); SLA(ADT)	Range $\Delta Q - \Delta Q(t)$ (Sv), see (18)	Range $Q_{adcp} - \bar{Q}_{adcp}$ (Sv), see (25)
Drake_2003	10.0	0.94	2.4	−2.6 to 1.1	**−4.8 to −1.5**	−1.9(−2.3)	−0.2 to 0.2	0.3
Drake_2007	9.0	−1.39	−3.3	**−3.3 to −1.5**	−0.1 to 1.8	−0.1(−0.1)	0.1 to 0.2	−0.3
Drake_2010 Jan	11.0	3.92	**11.8**	−0.4 to 5.4	**−12.2 to −6.4**	0.0(0.0)	0.0 to 0.0	0.5
Drake_2010 Oct	12.4	−3.34	−10.4	**−13.7 to −8.0**	−3.3 to 2.6	−0.3(−0.2)	−0.3 to −0.1	0.0
Drake_2011	11.5	1.32	4.2	−7.9 to 1.2	**−11.9 to −2.9**	−0.1(−0.1)	−0.3 to −0.1	−0.6
A21 1999	9.1	−0.42	−2.1	**−11.2 to −4.5**	**−9.0 to −2.6**	2.1(0.0)	0.0 to 0.2	−0.1
SR01 1993	23.2	2.69	7.1	**0.5 to 18.1**	−6.7 to 12.5	0.0(0.1)	−1.5 to 0.1	0.6
SR01 1994	19.2	0.8	2.0	**−5.2 to −1.0**	**−7.3 to −3.2**	0.0(0.0)	−0.1 to 0.5	−0.5
SR01 1996	19.9	−3.91	−11.3	−11.9 to 5.8	−1.5 to 18.8	0.3(0.4)	−1.7 to 0.9	−1.7
SR01 1998	14.4	−1.25	−3.2	**1.1 to 6.4**	**4.4 to 9.6**	0.9(1.1)	0.0 to 0.1	−2.3
SR01 2013	14.3	1.26	3.1	**2.7 to 8.3**	**−0.9 to 4.1**	3.8(3.8)	0.3 to 1.2	−2.2
SR01 2015	14.9	−1.08	−3.4	**−8.6** to **−6.8**	**−6.1 to −2.8**	0.3(0.3)	−0.6 to 0.8	0.7
SR01 2016	13.0	0.79	2.3	−2.9 to 12.9	−5.1 to 12.5	−3.7(−2.3)	−1.9 to 0.1	1.6
SR02 2001	18.9	−7.95	**−34.6**	−36.1 to 37.0	−2.6 to 45.6	19.2(15.1)	−1.8 to 26.2	−3.1
I06 2008	26.6	−6.41	−24.0	−23.7 to 21.1	**2.1 to 40.6**	2.9(3.9)	−3.0 to 4.6	2.8
I09 1995	16.6	0.33	1.9	−19.6 to 4.3	−25.0 to 2.5	0.4(0.0)	−0.6 to 1.2	−0.3
I09 2004	13.0	−0.01	−1.4	−10.4 to 15.1	−10.6 to 16.3	−0.4(−0.1)	−1.2 to 1.5	−0.2
I09 2012	12.8	5.25	**17.8**	−3.8 to 25.3	−21.0 to 4.7	1.6(1.6)	−1.0 to 2.9	−2.2
SR03 1994	23.0	−0.01	0.1	**2.9** to **15.4**	**3.0 to 17.6**	2.0(1.5)	−2.3 to 0.0	−0.8

(continued)

Table 2 (continued)

Section	FBTE $\langle \delta Q \rangle$ (Sv), see (21)	Deviation $\sum \Delta\zeta_t$ (cm), see (6b)	Deviation Q_t (Sv), see (10)[b]	Range $\Delta Q_{adcp}(t)$ (Sv), see (11)[c]	Range $\Delta Q'_{alt}(t)$ (Sv), see (15)[c]	Discrepancy ΔQ (Sv), see (17); SLA(ADT)	Range $\Delta Q - \Delta Q(t)$ (Sv), see (18)	Range $Q_{adcp} - \bar{Q}_{adcp}$ (Sv), see (25)
SR03 1995[a]	17.5	−0.86	−2.8	−5.7 to 1.4	−3.7 to 4.8	0.5(0.7)	−0.2 to 0.7	2.7
SR03 1996[a]	16.4	−2.38	−8.6	−4.9 to 11.3	**5.6 to 16.9**	−5.5(−4.9)	−2.9 to 3.0	3.5
SR03 2001	11.6	0.22	1.3	−15.3 to 5.2	−5.2 to 11.0	0.6(0.6)	−1.3 to 3.0	−1.8
SR03 2008	11.9	3.15	10.3	−0.5 to 16.1	−8.5 to 3.5	4.5(4.1)	−2.3 to 2.3	−3.3
SR03 2011	12.5	−3.17	−10.0	**−14.4 to −5.8**	−4.5 to 5.2	−1.7(−2.1)	−1.2 to 0.8	0.3
P11A 1993	16.3	−13.53	**−47.3**	**−59.2 to −27.2**	−9.0 to 21.2	2.6(2.4)	−2.8 to 1.8	−2.6
P19 1993	12.5	−3.97	**−17.1**	−12.8 to 1.3	**1.2 to 18.0**	1.7(2.0)	−0.4 to 3.0	−1.3
AR05 1997	4.3	−0.08	0.2	−4.0 to 3.5	−3.0 to 0.7	2.2(3.0)	−1.2 to 2.5	−0.1
AR05 2002	3.4	−0.99	−1.6	−4.0 to 0.3	−2.8 to 3.2	−0.5(−1.6)	−1.3 to 0.4	0.7
AR05 2003	2.9	−0.41	−0.9	−4.3 to 2.1	−3.4 to 2.8	−0.7(−1.4)	−0.1 to 0.2	2.4
AR05 2004	2.9	0.25	0.5	**0.2 to 3.1**	−0.6 to 3.3	−1.7(−2.1)	−0.8 to 0.4	3.6
AR05 2005	3.0	−0.21	−1.1	−6.6 to 3.2	**−6.9 to −0.3**	2.6(0.9)	0.2 to 4.6	1.2
AR05 2006	3.5	−0.56	−1.2	−9.7 to 2.3	−8.5 to 3.9	−1.9(−2.3)	−0.6 to 0.1	0.3
AR05 2007	3.6	0.87	1.3	−7.3 to 5.6	−8.5 to 4.6	0.0(−0.2)	−0.3 to 0.2	0.2
A05 1998	13.2	2.32	**21.8**	**5.5 to 20.0**	−14.7 to 1.7	−5.1(−4.2)	−5.0 to 1.3	3.7
A05 2004	8.9	0.54	5.6	**10.0 to 28.5**	**3.3 to 21.9**	−1.5(−0.5)	−1.8 to 3.3	2.0
A05 2010	10.0	−6.65	**−47.2**	−31.8 to 8.8	**14.1 to 60.5**	−2.3(−1.9)	−3.5 to 2.1	1.3
A05 2011	9.6	1.53	**12.3**	−18.2 to 18.9	−31.4 to 5.4	−0.1(0.6)	−0.9 to 2.2	−0.3
A05 2015	9.3	−5.13	**−39.3**	−6.0 to 36.8	**23.1 to 64.9**	9.1(9.5)	9.6 to 15.0	2.9
A10 1993	14.9	0.68	4.5	−30.8 to 11.8	−22.5 to 25.1	2.3(1.9)	0.3 to 6.3	−2.4
A10 2003	9.1	−1.46	−8.7	−21.3 to 34.0	−13.2 to 43.3	3.6(3.2)	−1.6 to 0.6	1.4

(continued)

Table 2 (continued)

Section	FBTE $\langle \delta Q \rangle$ (Sv), see (21)	Deviation $\sum \Delta\zeta_t$ (cm), see (6b)	Deviation Q_t (Sv), see (10)[b]	Range $\Delta Q_{adcp}(t)$ (Sv), see (11)[c]	Range $\Delta Q'_{alt}(t)$ (Sv), see (15)[c]	Discrepancy ΔQ (Sv), see (17); SLA(ADT)	Range $\Delta Q - \Delta Q(t)$ (Sv), see (18)	Range $Q_{adcp} - \bar{Q}_{adcp}$ (Sv), see (25)
A10 2011	10.0	2.86	**16.6**	−14.9 to 43.2	−30.2 to 29.2	−2.0(−2.6)	−3.3 to 0.8	−0.3
I05 2002	11.7	1.42	5.7	−67.0 to 85.2	−60.2 to 89.4	0.1(−1.9)	−15.3 to 18.1	−9.5
I05 2009	10.9	3.55	**15.6**	−184.2 to 166.7	−201.3 to 150.4	−0.5(0.4)	12.8 to 20.5	−3.4
P02 1994	33.7	−4.79	−32.6	−45.1 to 57.7	−44.1 to 89.1	8.6(−21.4)	−33.0 to 31.8	−28.1
P02 2004	14.5	15.82	**88.0**	−331.1 to 153.2	−416.4 to 64.0	−0.5(−1.9)	−3.6 to 14.8	−5.3
P02 2013	16.9	−7.37	**−43.2**	**−249.5 to −57.9**	**−191.2 to −0.1**	−13.8(−14.5)	−23.8 to −3.3	−2.0
P06 2003	8.2	3.69	**13.6**	**11.4 to 293.7**	−7.8 to 276.4	1.8(2.4)	−1.6 to 11.1	−5.7
P06 2009	9.4	−5.32	**−37.1**	−128.4 to 234.4	−85.4 to 282.0	−3.2(−2.9)	−12.6 to 4.1	−4.6

[a]The estimates for the SR02 sections in 1995 and 1996 are given for the entire section, except for a small southern segment, enclosed in the data by an ice mask
[b]Bold face in this column indicates the transport deviations greater than FBTE
[c]Bold face in this column corresponds to the ranges of the transport variability that do not cross the zero line

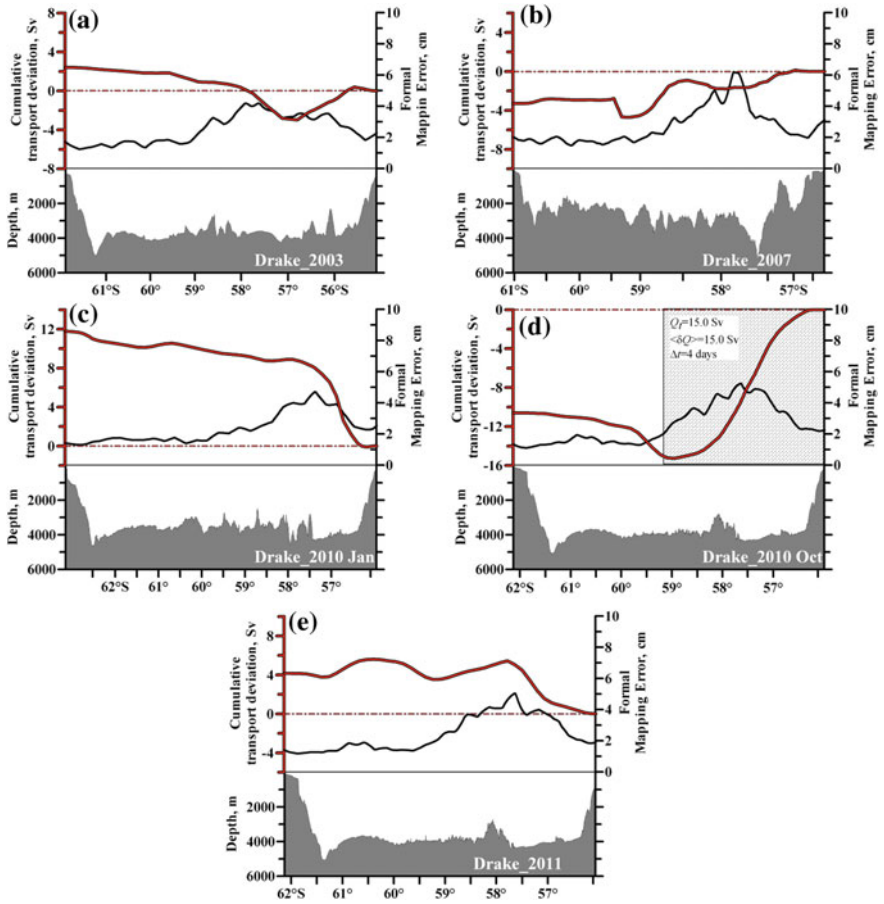


Fig. 4 Depth distributions (lower parts of the figures), Formal Mapping Errors (heavy black curves), and the cumulative deviations of the barotropic transports at the midpoints between stations Q_t (red curves) over the Russian sections in the Drake Passage: **a** 2003, **b** 2007, **c** January 2010, **d** October–November 2010, **e** 2011. Transport cumulation is from south to north. In Fig. 4d, the segment of the section with the magnitude of the absolute value of Q_t greater the Formal Error of the Barotropic Transport $\langle \delta Q \rangle$ calculated for this segment is shaded; the duration of measurements in this section is also indicated. The position of the sections is shown in the inset in Fig. 1, information about them is given in Table 1

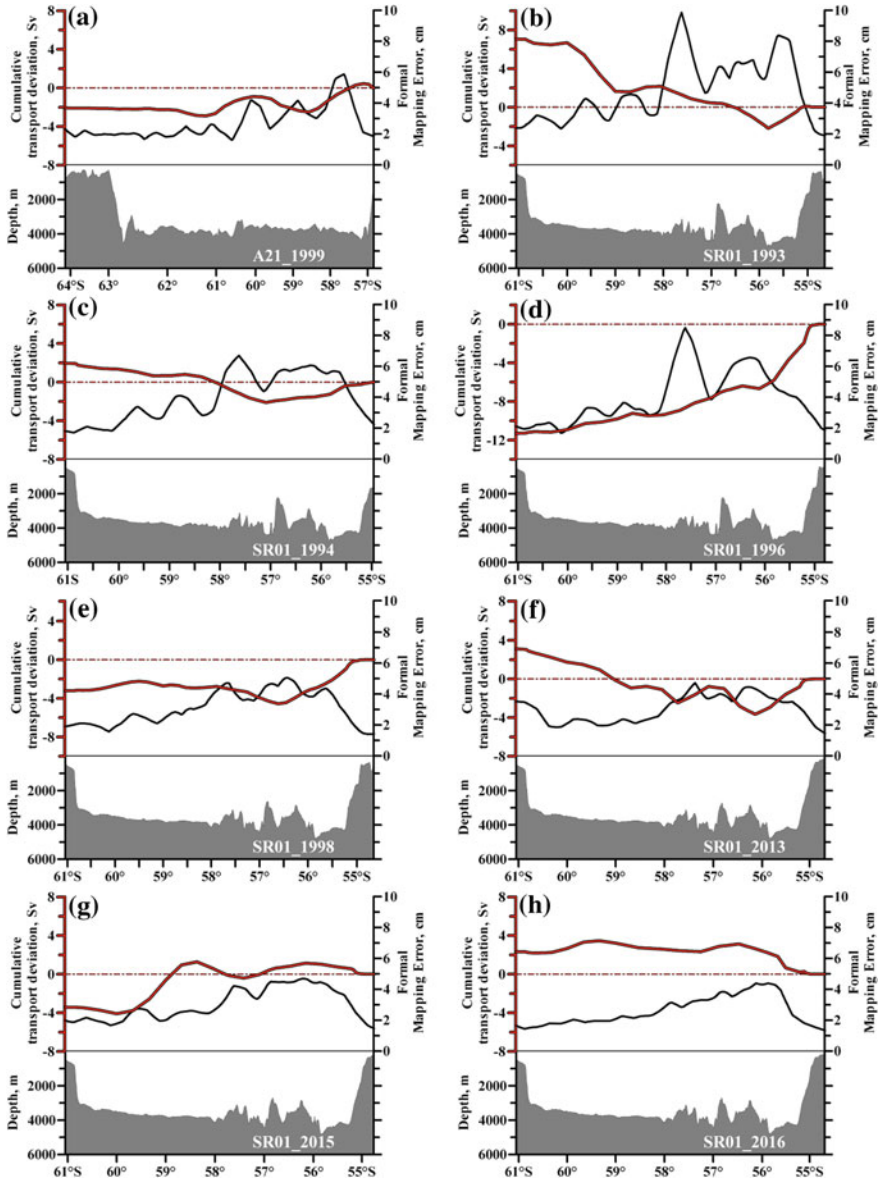


Fig. 5 The same as in Fig. 4 for quasi-meridian sections in the Drake Passage: **a** A21 in 1999, **b** SR01 in 1993, **c** SR01 in 1994, **d** SR01 in 1996, **e** SR01 in 1998, **f** SR01 in 2013, **g** SR01 in 2015, **h** SR01 in 2016. The locations of the sections are shown in Fig. 1; information about them is given in Table 1

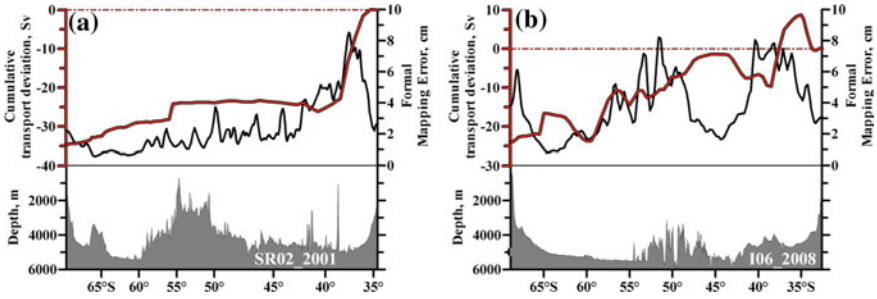


Fig. 6 The same as in Fig. 5, but for sections south of Africa: **a** SR02 in 2001, **b** I06 in 2008

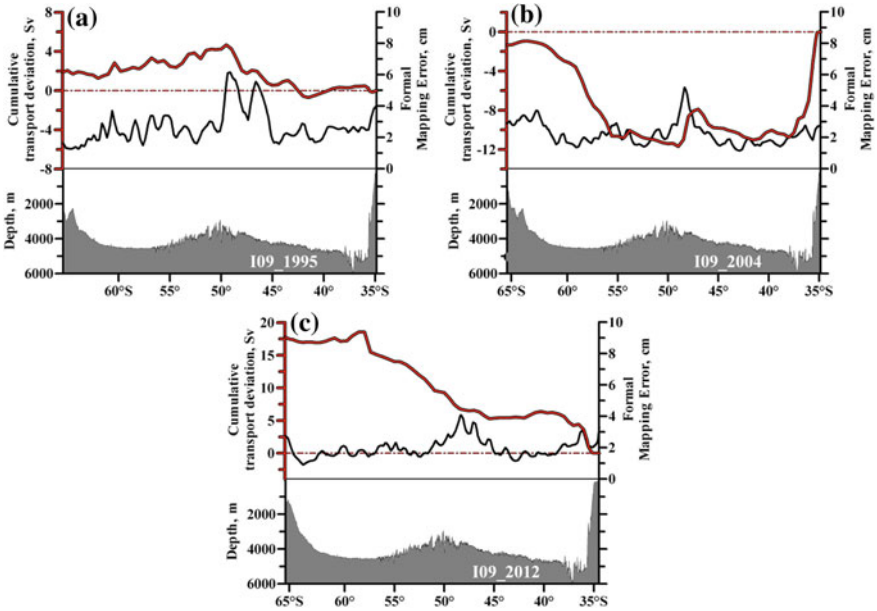


Fig. 7 The same as in Fig. 5, but for sections I09 south of the southwestern tip of Australia: **a** 1995, **b** 2004, **c** 2012

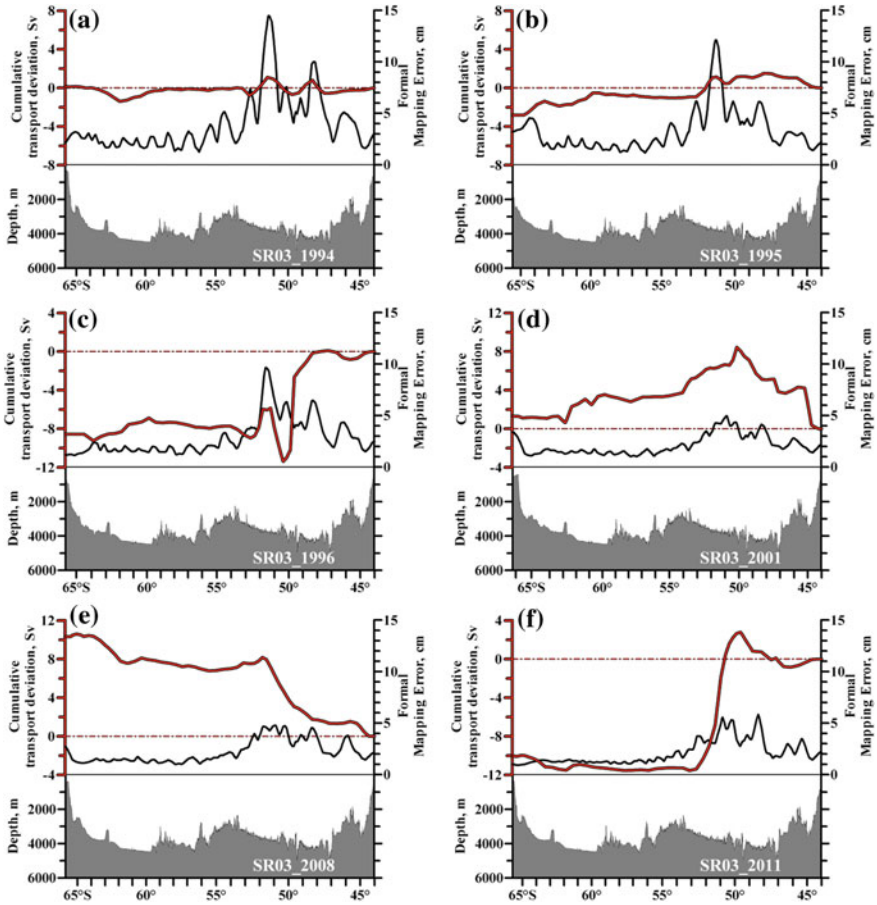


Fig. 8 The same as in Fig. 5, but for SR03 sections from Tasmania to Antarctica: **a** 1994, **b** 1995, **c** 1996, **d** 2001, **e** 2008, **f** 2011

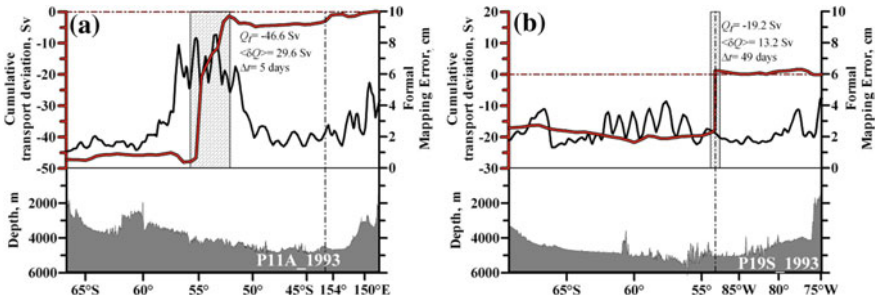


Fig. 9 The same as in Fig. 5, but for sections in the South Pacific: **a** P11A in 1993, **b** P19 in 1993

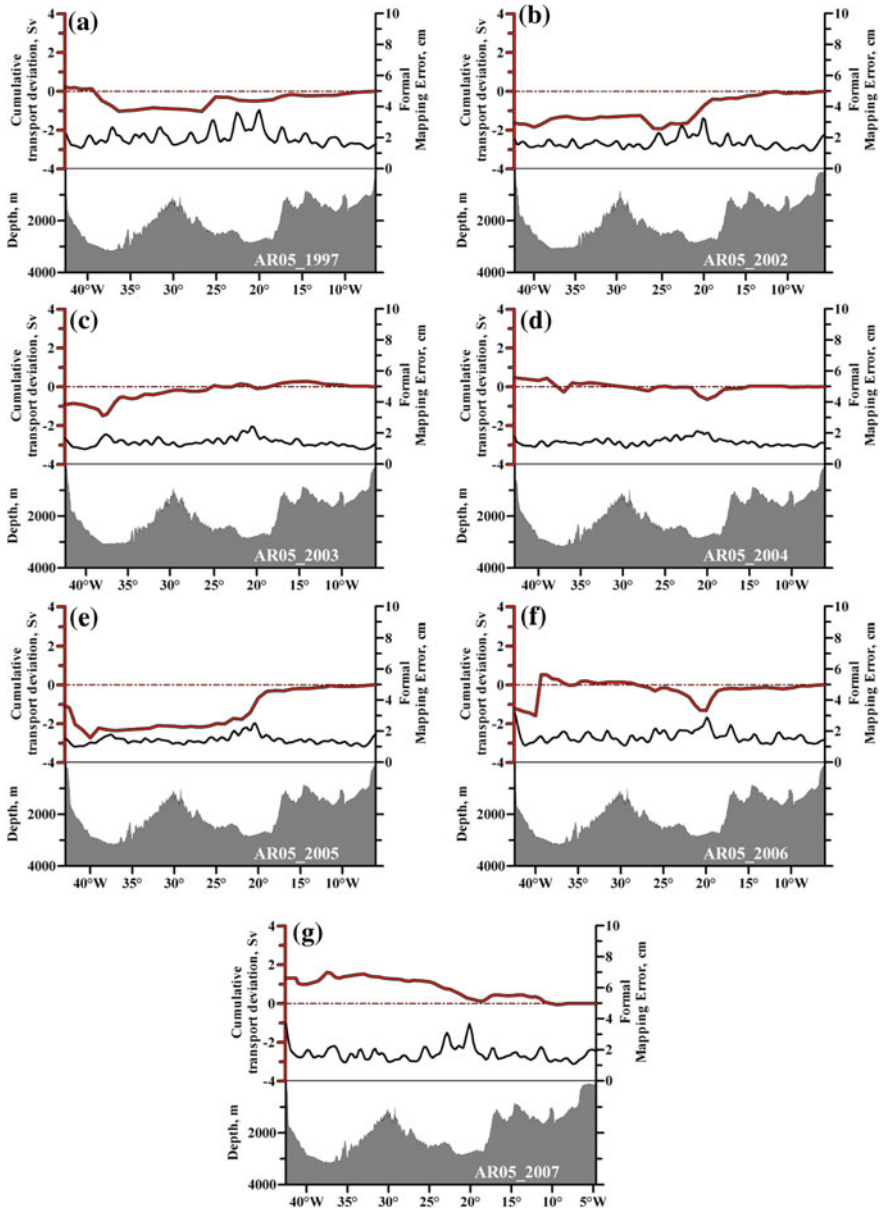


Fig. 10 The same as in Fig. 5, but for quasi-zonal AR05 sections in the subarctic area of the Atlantic: **a** 1997, **b** 2002, **c** 2003, **d** 2004, **e** 2005, **f** 2006, **g** 2007

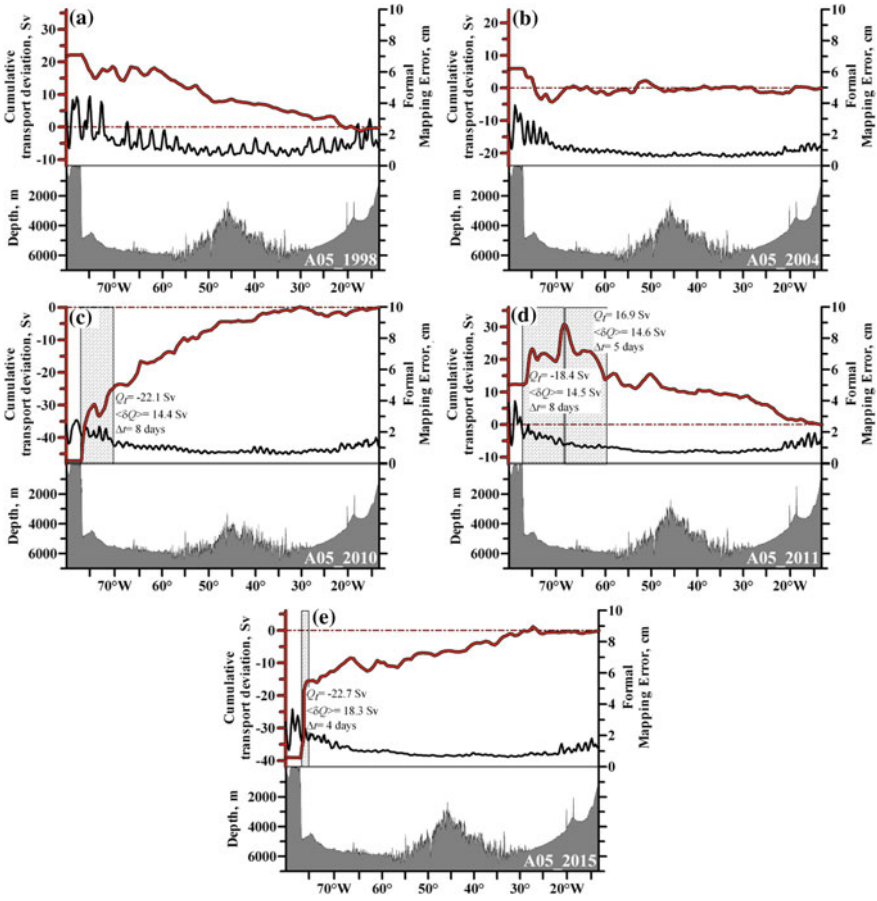


Fig. 11 The same as in Fig. 10, but for sections A05 in the North Atlantic: **a** 1998, **b** 2004, **c** 2010, **d** 2011, **e** 2015

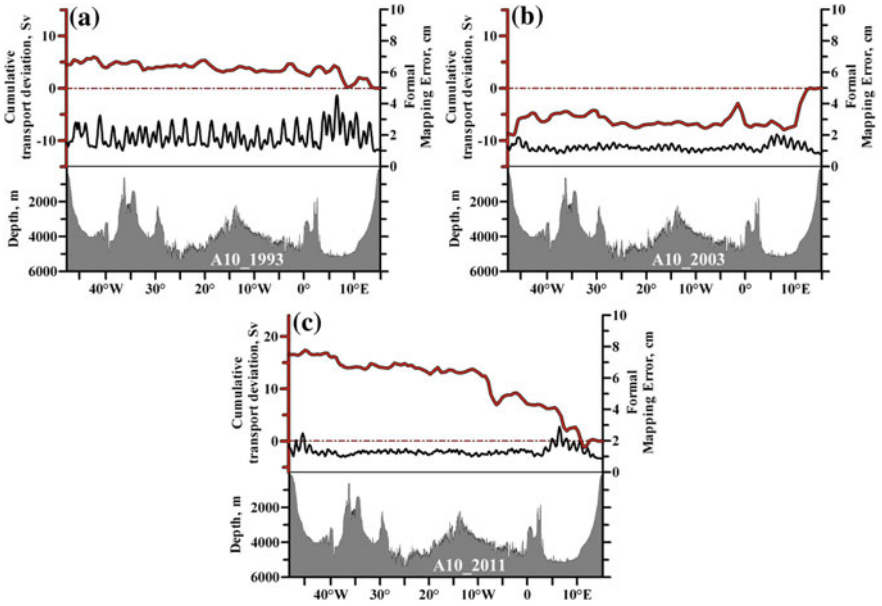


Fig. 12 The same as in Fig. 10, but for sections A10 in the South Atlantic: **a** 1993, **b** 2003, **c** 2011

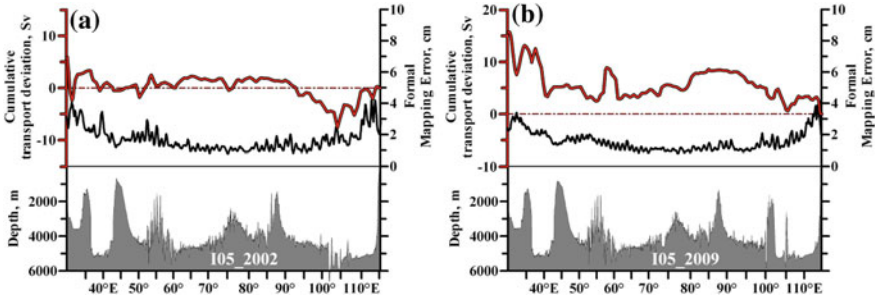


Fig. 13 The same as in Fig. 10, but for sections I05 in the southern Indian Ocean: **a** 2002, **b** 2009

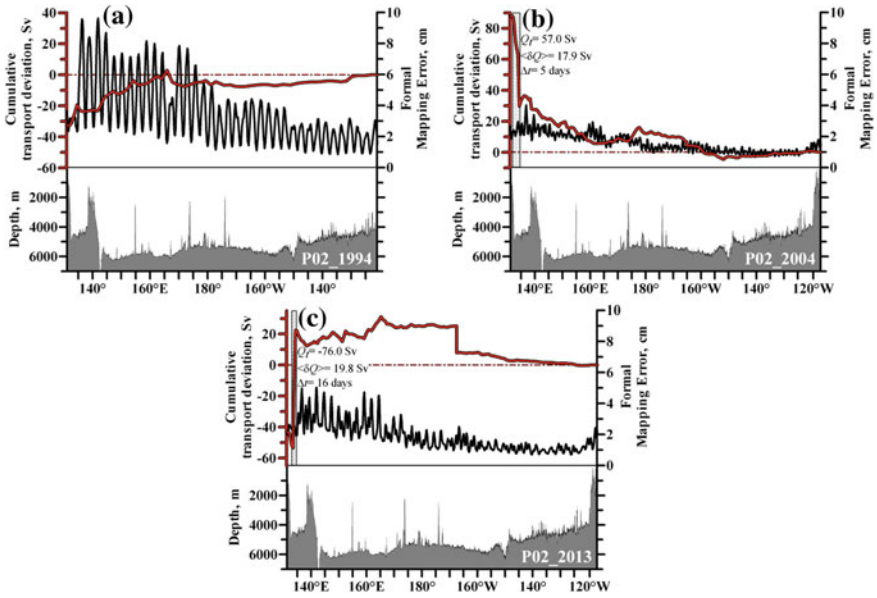


Fig. 14 The same as in Fig. 10, but for sections P02 in the North Pacific: **a** 1994, **b** 2004, **c** 2013

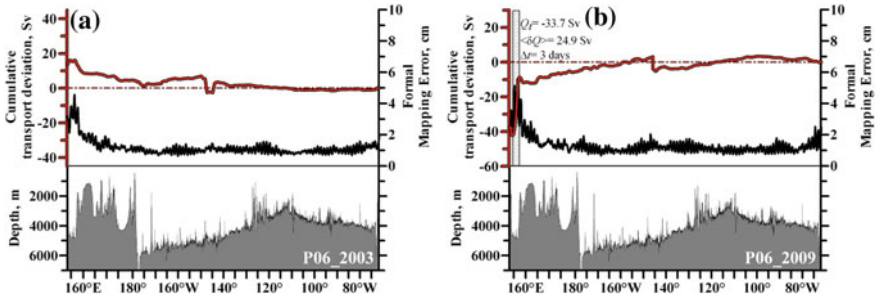


Fig. 15 The same as in Fig. 10, but for sections P06 in the South Pacific: **a** 2003, **b** 2009

the weight coefficients w_j must satisfy the condition for the dependence of the transport error only on the values of $\delta\zeta$ at the ends of the section in the ocean with the bottom slope compensating for the latitudinal variation of the Coriolis parameter. Therefore, they must be proportional to the depth at the ends of the section, while at the remaining points they must be proportional to the sum of the variation of the depth and the mentioned change in the section segments corresponding to these points. The account for the latitudinal variation of the Coriolis parameter can be important for the tracks of quasi-meridional sections at low latitudes. Since the step of the section partitioning we used in calculating the bottom relief along the section is much smaller than the resolution of the initial array of the bottom

topography and, even more, it is smaller than the resolution of the satellite altimetry data used in this version of the calculation scheme we accept the following schemes of calculation for this version:

$$\begin{aligned} w_1 &= |g_1 h_1 / f_1|, \\ w_m &= |g_m h_m / f_m|, \\ w_j &= \left| \frac{g_j}{f_j} \left(\frac{h_{j+1} - h_{j-1}}{2} - h_j \frac{f_{j+1} - f_{j-1}}{f_j} \right) \right| \text{ for } j = 2, m-1; \end{aligned} \tag{22}$$

here, square brackets denote the absolute value of the magnitude.

Estimates of the error of the barotropic transport calculated on the basis of the data from the formal mapping error over all the sections studied are given in Table 2. Hereinafter, for simplicity, this error will be called **the Formal Error of the Barotropic Transport (FBTE)**.

The inaccuracy in the determination of the ocean level, due to the duration of the measurements and the drift of the vessel at the station, should also be considered as an error of the data. It can be interpreted, as well as the FME, as the root-mean-square deviation with respect to some average level value during the station execution time. Then, their total error at the station is represented as:

$$\delta\zeta' = \sqrt{\frac{1}{l} \sum_{k=1,l} \left((\langle\zeta\rangle - \zeta_k)^2 + (\delta\zeta_k)^2 \right)}, \tag{23}$$

where $\langle\zeta\rangle = \frac{1}{l} \sum_{k=1,l} \zeta_k$ is the ocean surface level averaged over the station time, the summation is carried out over all time points from the beginning to the end of the station. Considering that the time of the station is usually much shorter than a day, i.e. than the time step of the digital maps formula (23) can be simplified to evaluate $\delta\zeta'$:

$$\delta\zeta' = \sqrt{\left(\frac{\zeta_a - \zeta_b}{2} \right)^2 + \frac{(\delta\zeta_a)^2 + (\delta\zeta_b)^2}{2}}, \tag{24}$$

where ζ_a, $\delta\zeta_a$, and ζ_b, $\delta\zeta_b$ are the values of the level and the FME, respectively, at the beginning and at the end of the station. Note that the value of $\delta\zeta'$ when taking into account the vessel drift depends on the mean dynamic topography and it does not depend on the topography if we do not take the ship's drift into account. It is the total error that can be used instead of the FBTE $\langle\delta Q\rangle$.

The example in Fig. 16 shows the distributions of the FME $\delta\zeta$ and the absolute value of the level change $|\zeta_a - \zeta_b|$ at the stations with and without taking into account the vessel's drift over the section across the Drake Passage in October 2010. The calculation was made both from the SLA and ADT data. As can be seen from Fig. 16, the level change without taking into account the vessel's drift is more than one order of magnitude smaller than the FME. It is expected that the largest changes in the level appeared during the calculation based on the ADT data taking

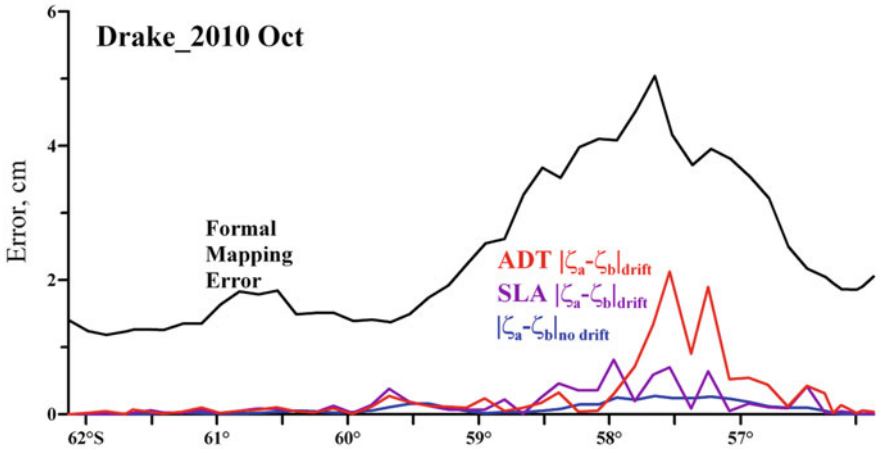


Fig. 16 Distribution of errors over the Russian section in the Drake Passage in October–November 2010. Black color shows the Formal Mapping Error. The color curves correspond to the errors related to the duration of measurements and drift of the vessel at the stations: red curve is calculated from the data of the Absolute Dynamic Topography taking into account the drift of the vessel, purple is the same, but according to the Sea Level Anomaly, blue is the same, but without taking into account the drift of the vessel

into account the vessel drift. They are as large as 2 cm, which corresponds to a 2-mile drift of the vessel at the station. However, even these errors appeared two and a half times smaller than the FME at these points, which corresponds to a contribution to the total error given by formula (22) of only about 2%.

The previous example of the section across the Drake Passage shows that the existence of the time period of measurements and the drift of the ship at the stations do not lead to any significant additive to the FME over the entire section, although the local effect may be significant. Another conclusion follows from this fact: the choice of the station point in time (the beginning, the position of the measuring equipment at the bottom, and the end of the station) is not important for estimating the barotropic transport over the entire section based on the satellite altimetry data.

One more error in the data that affects the accuracy of estimating the transport in the sections is **the error of the digital array of the bottom topography**. Since the magnitude of the barotropic transport error depends on the depth over the section, the error of the digital bottom topography array, which is a percentage of the depth and obviously does not make a significant additional contribution to their total error.

As already noted in section "Method of Processing", **the values of ΔQ, and $\Delta Q - \Delta Q(t)$ determined by formulas (17) and (18), characterize the quantitative difference between the estimates of barotropic transports in two finite-difference schemes for representing the section**: with the calculation of the sea level at the points of the stations and the midpoints of the spaces between them. We note that when the distance between the stations tends to zero, discrepancy ΔQ

corresponding to the calculation over the section length and time-dependent discrepancy $\Delta Q - \Delta Q(t)$ also tends to zero.

Over most of the sections studied, with a typical distance between stations less than 30 miles, the absolute value of discrepancy ΔQ was several times smaller than the FBTE $\langle \delta Q \rangle$ (Table 2). We noted above that the specified distance, depending on the latitude range of a particular section from Table 1, is 2–3 steps of a quarter-degree grid of altimetry data. However, there are examples when the values of ΔQ and $\langle \delta Q \rangle$ are comparable (sections SR02 in 2001, A05 in 2015, P02 in 1994, and P02 in 2013), which implies insufficient resolution of these sections with the occupied stations. The characteristic distance between stations over section SR02 was 60 miles in its northern part, while the distance between stations over the P02 (1994) section was 90 miles. These are the maximum values among all the sections. The characteristic distance between stations in the open ocean over the sections A05 (2015) and P02 (2013) was 40 miles. A wide range of values $\Delta Q - \Delta Q(t)$ also indicates lack of resolution of the sections by the stations. This criterion adds sections I05 in 2002 (up to 50 miles between stations), P02 in 2004 (30–35 miles), P06 in 2003 (30–35 miles) and P06 in 2009 (30–35 miles). We note that almost zero errors were estimated over the lines of the Russian sections in the Drake Passage with a resolution of 10 miles.

Calculation of transport deviation Q_t due to the non-stationarity of the current field and the non-simultaneity of hydrographic measurements does not depend on the type of the finite-difference scheme of the partition of the section (see formula (10)). Therefore, discrepancy ΔQ has nothing to do with the estimates of the accuracy of this value. In order to obtain such an estimate, it is possible to virtually increase the number of stations over a section, gradually reducing the time intervals and the distance between stations to a value smaller than the grid step of the altimetric data. Then, the values of Q_t are calculated for these versions of the partition of the section by formula (10). Their discrepancy with the estimate of Q_t at the original resolution characterizes the unknown **error due to the difference between the horizontal resolution of the maps and distance between the stations over the section**. Figure 17 shows the examples of the cumulative (from the northern end for the meridional, and from the eastern end for the zonal sections) values of Q_t discrepancy over several sections for two-, four-, and eightfold subdivision of sections by stations. These examples show that the calculation error of Q_t over the lines of the real sections is much smaller than the FBTE. Even over the P02 section (1994), with the worst station resolution, their ratio was one to four (8 Sv vs. 34 Sv; 1 Sv = 10^6 m^3 s^{-1}). In most cases, the calculation error of Q_t did not exceed several tenths of a Sverdrup.

As already indicated in section "Method of Processing", the averaging of the ADCP data to the center of the space between the stations (when estimating the barotropic component of the current) does not involve quantitative differences in the calculation of the level surface. However, when estimating the transport across a section, this procedure leads to the following discrepancy:

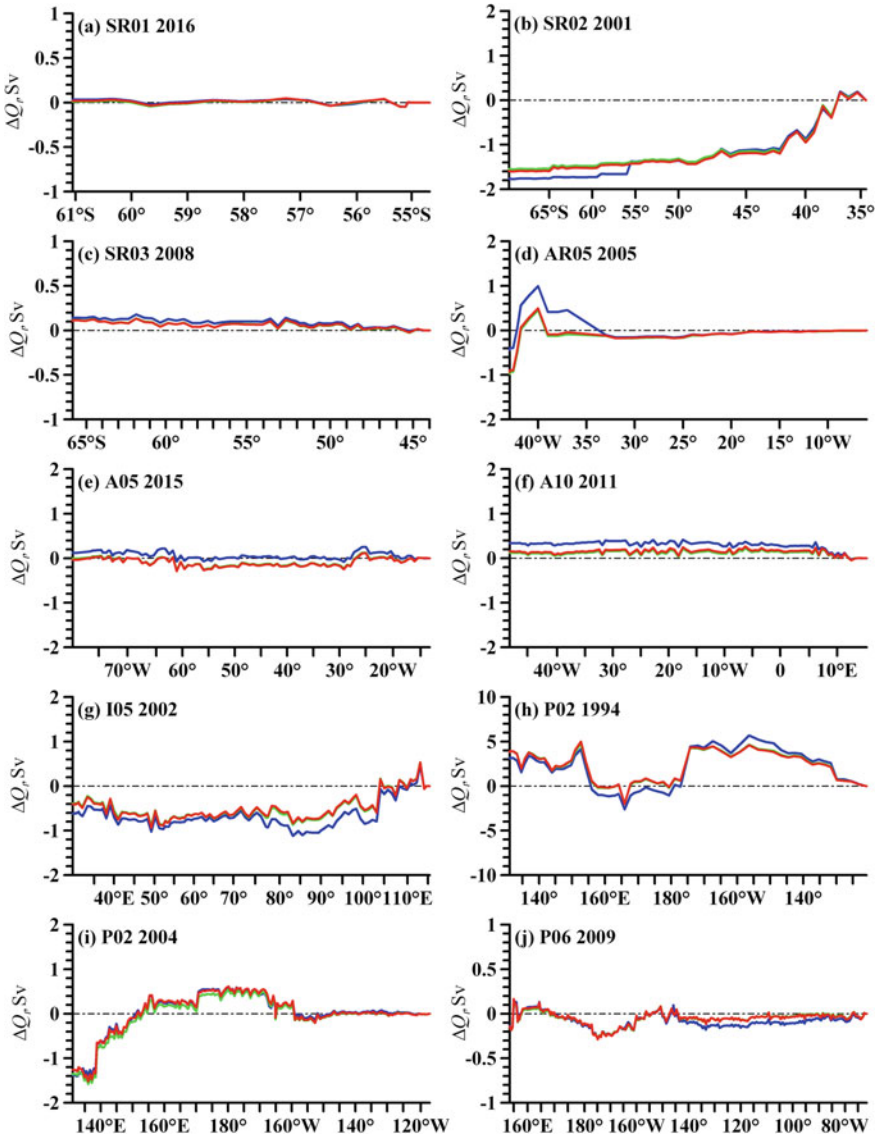


Fig. 17 Cumulative distributions of the ΔQ_t difference between the initial and increased resolution of the sections by the stations: **a** SR01 in 2016, **b** SR02 in 2001, **c** SR03 in 2008, **d** AR05 in 2005, **e** A05 in 2015, **f** A10 in 2011, **g** I05 in 2002, **h** P02 in 1994, **i** P02 in 2004, **j** P06 in 2009. Blue, green, and red curves are for double, fourfold, and eightfold increase in resolution respectively

$$Q_{adcp} - \bar{Q}_{adcp} = \frac{1}{2} \sum_{i=1,n-1} (\bar{h}_{i+0.75}(x_{i+1} - x_{i+0.5}) - \bar{h}_{i+0.25}(x_{i+0.5} - x_i))(u_{i+1} - u_i); \quad (25)$$

here, $\bar{Q}_{adcp}$ is the transport based on the averaged data, x_i and u_i are the distances from the beginning of the section and the barotropic component of the measured velocity at the point of the i-th station, respectively, $x_{i+0.5}$ is the distance from the beginning of the section to the midpoint point between the i-th and $i + 1$-st stations, $\bar{h}_{i+0.25}$ is the average depth between the i-th and $i + 0.5$-th points of the section. As the distance between the stations tends to zero, this discrepancy also tends to zero. Although it characterizes the accuracy of the calculation procedure from the ADCP data, it can be easily simulated by the altimetry data. Velocity values can be either interpolated over the section line from the digital maps of velocity at the ocean surface or calculated using the ADT data over the section line, for example, at the midpoints of the spaces between stations. Table 2 shows the estimates of the discrepancy for the second method of calculating the velocity. The values of the discrepancy over sections with a spatial resolution between stations less than 30 miles are usually substantially smaller than the FBTE. These values are comparable on the sections with a coarser resolution. The exception is only valid for a few AR05 sections where the formal error is very small.

Summarizing the discussion in this section, it can be argued that the FBTE is the main component of the error in estimating the barotropic transports over hydrographic sections from satellite altimetry. The remaining components in most cases are smaller. An exception may be the sections, where the distance between stations is much larger than the linear quarter-degree step of the grid of the altimetric data.

Discussion

As already noted above, the non-stationarity of the current field and the non-simultaneity of hydrographic measurements in calculating the barotropic transport across the hydrographic section from the ADCP measurements are manifested in two phenomena: in the transport deviation Q_t based on the satellite altimetry data, which can be taken into account using the altimetry data, and in the unaccountable temporal variability of the barotropic transport due to the displacement of the streamlines of the geostrophic currents with respect to isobaths.

Total transport deviation Q_t over 12 of the 49 sections studied in the World Ocean (Table 1), was greater than the estimate of the FBTE $\langle \delta Q \rangle$ (Table 2). Among them are the following quasi-meridional sections: January 2010 in the Drake Passage, SR02 south of Africa, I09 in 2012 from the southwestern tip of Australia to Antarctica, P11A south of Australia, P19 in 1993 in the Southeast Pacific; quasi-zonal sections: A05 in 1998, 2010, 2011, and 2015 in the North Atlantic, A10 in 2011 in the South Atlantic, I05 in 2009 in the Southern Indian Ocean, P02 in

2004 and 2011 in the North Pacific, P06 in 2003 and 2009 in the South Pacific. It is expected that in most of these cases there was a long duration of measurements over the sections (more than 20 days). However, there is an example of faster measurements (9 days) in January 2010 in the Drake Passage. There are also several examples of "rapid" sections, when deviation Q_t was smaller, but comparable to the error $\langle \delta Q \rangle$: the October 2010 section in the Drake Passage (8 days), SR03 in 2008 (18 days) and 2011 (14 days) the south of Australia (Table 2).

Figures 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 and 15 show the distributions of cumulative deviations Q_t over the 49 sections studied here (Table 1). The accumulation of the deviation occurs from north to south for quasi-meridional, and from east to west for quasi-zonal sections. Configurations of level deviations $\sum \Delta\zeta_t$ are almost similar to these distributions. The coefficient of similarity between them can be estimated from Table 2. In addition, the intervals of the sections with sharp local deviation Q_t, which are not less than the FBTE $\langle \delta Q \rangle$, are highlighted in some of Figs. 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 and 15. These sharp deviations can be associated both with a significant time pause in the measurements and a rapid change in the current field in short time intervals.

Segments with sharp deviation Q_t are found in the western parts of the oceans over quasi-zonal sections: A05 in 2010 and 2015 close to the continental slope of North America in the North Atlantic (Fig. 11c, e); P02 in 2004 and 2013 crossing the Kuroshio zone and the Kuroshio Recirculation in the North Pacific (Fig. 14b, c), P06 in 2009 crossing the retroflection of the East Australian Current in the South Pacific (Fig. 15b). In the latter case, significant deviation Q_t compared with $\langle \delta Q \rangle$ was accumulated just in 3 days. It should be noted that in the P02 section in 2004 a deviation of 57 Sv, which occurred only in 5 days, is comparable to the total transport of the Kuroshio estimated at 52–57 Sv [7, 14]. Two segments with a significant sharp deviation are located in the western part of section A10 in 2011 (Fig. 11d).

In the Southern Ocean, the segments of sections Drake_2010 in October 2010 (Fig. 4d), and P11A in 1993 (Fig. 9a) with sharp deviations Q_t corresponded to the northern periphery of the Antarctic Circumpolar Current (ACC). Of particular note is the P11A section south of Australia, where about 47 Sv were accumulated just in 5 days, which is more than a quarter of the average total transport over this section in the Southern Ocean, consisting of the transport of the ACC (173 Sv according to [3], transport compensating the flow of water from the Pacific into the Indian Ocean through the Indonesian Straits (14–18 Sv) (e.g., [4, 18]), and transport compensating the flow from the Pacific to the Arctic Ocean through the Bering Strait (0.8 Sv) [16]. A sharp jump of Q_t in the northern part of section P19 (Fig. 14c) is associated with a long time break in the measurements.

The above examples show that the non-stationarity of the current field in combination with non-simultaneous measurements at stations of a hydrographic section can lead to significant distortions in the transport estimates over the sections based on the ADCP measurements, i.e. "instantaneous" measurements of the velocity of currents at stations. Often, the accumulation of these distortions along

the line of the section takes place in small time intervals of several days and it is localized in relatively small parts of this section, where the largest non-stationarity has been noted. Thus, in the analysis of the real field data in the ocean, the estimates of such distortions are absolutely necessary both with respect to the calculation of the total transport across the section, and with respect to the investigation of the structure of this transport. This conclusion concerns even such "fast" measurements as sections in the Drake Passage, the duration of which is less than 10 days.

As can be seen from Table 2, the range of variability of the instantaneous barotropic transports (the range of barotropic transport anomaly values in Table 2) over most sections is significant compared with the FBTE $\langle \delta Q \rangle$; the exception is related to almost all "quick" sections in the Drake Passage (except for SR01 in 2016) and four SR03 sections south of Australia (1994, 1995, 1996, and 2011). These facts of such significance indicate that the estimation of the variability of the barotropic transport caused by the displacement of streamlines at the ocean surface relative to the isobaths is essential in the analysis of the field data, even over "fast" sections. For example, the duration of the above mentioned SR01 section in 2016 was only 6 days.

The range of instantaneous barotropic transport variability should be interpreted depending on the restrictions on the total transport across the section. Under the conditions of constant total transport across the section line, the oscillations of its baroclinic and barotropic components must completely compensate each other. It is exactly the case that is observed over the lines of the transoceanic quasi-zonal sections crossing the Atlantic, Indian, and Pacific oceans. In the Atlantic Ocean, instantaneous total transport corresponds to the inflow of the Pacific waters through the Bering Strait; in the Indian Ocean this is the inflow through the Indonesian straits (over the sections from Africa to Australia); in the Pacific Ocean this is a combination of the outflow through the Indonesian straits depending on the location of the section. Thus, the range of the barotropic transport anomaly values $\Delta Q'_{alt}(t)$ in Table 2 for the corresponding sections also gives an indirect estimate of the oscillations of baroclinic transports.

The time variability of the total transport over the quasi-meridian sections crossing the Southern Ocean exists, due to the variability of the transport of the ACC, as a result of which exact compensation of the oscillations of the instantaneous barotropic transport by the baroclinic apparently does not occur. At the same time, the measurements in the Drake Passage [8] show a high degree of anticorrelation between the barotropic and baroclinic transports in the Southern Ocean. Under these conditions, the range of the barotropic transport anomaly values $\Delta Q'_{alt}(t)$ in Table 2 represents an indirect combined estimate of fluctuations in the baroclinic transport and the variability of the total transport. We also note that no exact compensation occurs over the lines of the sections not ending at the continental slope.

The account for Q_t allows us to complete the total transport, determined from the data of the instant ADCP measurements over a section, to such a total transport that has exact physical correspondence between its baroclinic and barotropic components. This is caused by the gained exact time correspondence between the ocean

level and the CTD profile at the stations. However, the total transport obtained using the suggested method does not have to tend to a strictly defined value (in the case of quasi-zonal sections across the Atlantic, Indian, and Pacific oceans) or to some range of temporal variability (in the case of quasi-meridional sections in the Southern Ocean). The same is related to the barotropic and baroclinic components separately. An illustration of this statement is that quite often the total barotropic transport over the section exceeds the variability of its instantaneous values. In Table 2, this corresponds to the range of values $\Delta Q'_{alt}(t)$ that do not cross the zero line. The total transport obtained in this manner is only an approximation to the real values. The same conclusion is related to the total transport recovered from the ADT data at stations. To obtain a final estimate of the total transport across the section or the range of its variability, depending on the restrictions on the total transport, it is necessary to study the variability of the baroclinic transport component during the section time period based on analysis of the measurements in the ocean depth.

Conclusions

1. The main error component in the estimates of the barotropic transports over oceanic sections from the satellite altimetry data is caused by the Formal Mapping Error.
2. The non-stationarity of the current field in combination with non-simultaneous measurements at stations of a hydrographic section can lead to significant (i.e., exceeding the error in calculating the barotropic transport, based on the Formal Mapping Error) distortions of the transport estimations over the sections based on the ADCP measurements of velocity. Thus, the estimates of these non-stationarity and non-simultaneity are absolutely necessary in the analysis of the field data in the ocean, both with respect to the calculation of the total transport across the section, and with respect to the investigation of the structure of this transport.
3. The account for the deviation of the barotropic transport caused by the non-stationarity of the current field and non-simultaneity of hydrographic measurements allows us to add an increment to the total transport, determined from the instant ADCP measurements over a section, only up to some approximation of the real total transport. To obtain a final estimate of the total transport across the hydrographic section, it is necessary to study the variability of the baroclinic current component on the basis of the analysis of the measurements in the ocean column.

Acknowledgements The work was supported by the Russian Science Foundation, grant 16-17-10149. The author is grateful to E. G. Morozov, and K. V. Lebedev for useful comments when preparing the text of this contribution.

References

1. Arhan, M., Naveira Garabato, A. C., Heywood, K. J., & Stevens, D. P. (2002). The Antarctic Circumpolar Current between the Falkland Islands and South Georgia. *Journal of Physical Oceanography, 32*(6), 1914–1931.
2. Cunningham, S. A., Alderson, S. G., King, B. A., & Brandon, M. A. (2003). Transport and variability of the Antarctic Circumpolar Current in Drake Passage. *Journal of Geophysical Research, 108*(C5), 8084. https://doi.org/10.1029/2001JC001147.
3. Donohue, K. A., Tracey, K. L., Watts, D. R., Chidichimo, M. P., & Chereskin, T. K. (2016). Mean Antarctic Circumpolar Current transport measured in Drake Passage. *Geophysical Research Letters, 43,* 11760–11767. https://doi.org/10.1002/2016GL070319.
4. Dye, S., Hansen, B., Østerhus, S., Quadfasel, D., & Rudels, B. (2007). The overflow of dense water across the Greenland-Scotland Ridge. *Exchanges, 40,* 20–22.
5. Gladyshev, S., Arhan, M., Sokov, A., & Speich, S. (2008). A hydrographic section from South Africa to the southern limit of the Antarctic Circumpolar Current at the Greenwich meridian. *Deep-Sea Research Part I, 55*(10), 1284–1303.
6. Gladyshev, S. V., Koshlyakov, M. N., & Tarakanov, R. Y. (2008). Currents in the Drake Passage based on observations in 2007. *Oceanology, 48*(6), 759–770.
7. Imawaki, S., Uchida, H., Ichikawa, H., Fukasawa, M., Umatani, S., & ASUKA Group. (2001). Satellite altimeter monitoring the Kuroshio Transport South of Japan. *Geophysical Research Letters, 28,* 17–20.
8. Koenig, Z., Provost, C., Park, Y. H., Ferrari, R., & Sennéchael, N. (2016). Anatomy of the Antarctic Circumpolar Current volume transports through Drake Passage. *Journal of Geophysical Research: Oceans, 121,* 2572–2595. https://doi.org/10.1002/2015JC011436.
9. Koshlyakov, M. N., Lisina, I. I., Morozov, E. G., & Tarakanov, R. Y. (2007). Absolute Geostrophic Currents in the Drake Passage Based on observations in 2003 and 2005. *Oceanology, 47*(4), 451–463.
10. Koshlyakov, M. N., Gladyshev, S. V., Tarakanov, R. Y., & Ryzhikov, N. I. (2010). Deep Currents in the Central Part of the Drake Passage based on the data of the 2008 hydrographic survey. *Oceanology, 50*(6), 821–828.
11. Koshlyakov, M. N., Gladyshev, S. V., Tarakanov, R. Y., & Fedorov, D. A. (2011). Currents in the Western Drake Passage according to the observations in January of 2010. *Oceanology, 51*(2), 187–198.
12. Koshlyakov, M. N., Gladyshev, S. V., Tarakanov, R. Y., & Fedorov, D. A. (2012). Currents in the Drake Passage based on the observations in November of 2010. *Oceanology, 52*(3), 299–308.
13. Koshlyakov, M. N., Gladyshev, S. V., Tarakanov, R. Y., & Fedorov, D. A. (2013). Currents in the Drake Passage by the observations in October–November of 2011. *Oceanology, 53*(1), 1–12.
14. Qiu, B., & Joyce, T. M. (1992). Interannual variability in the mid- and low latitude western North Pacific. *Journal of Physical Oceanography, 22,* 1062–1079.
15. Rio, M. H., Mulet, S., & Picot. N. (2013). New global mean dynamic topography from a GOCE geoid model, altimeter measurements and oceanographic in-situ data. In *Proceedings of the ESA Living Planet Symposium, Edinburg, September 2013*.
16. Roach, A. T., Aagaard, K., Pease, C. H., Salo, S. A., Weingartner, T., Pavlov, V., et al. (1995). Direct measurements of transport and water properties through the Bering Strait. *Journal of Geophysical Research, 100*(C9), 18443–18457.
17. Smith, W. H. F., & Sandwell, D. T. (1997). Global seafloor topography from satellite altimetry and ship depth soundings. *Science, 277,* 1957–1962.
18. Talley, L. D., Reid, J. L., & Robbins, P. E. (2003). Data-based meridional Overturning stream functions for the global ocean. *Journal of Climate, 16*(10), 3213–3226.

Satellite Remote Sensing of Submesoscale Eddies in the Russian Seas

Andrey G. Kostianoy, Anna I. Ginzburg, Olga Yu. Lavrova and Marina I. Mityagina

Introduction

Numerous field measurements (especially large in number since the 1970s when the Soviet experiment POLYGON-70, the American experiment MODE, and the Soviet-American experiment POLYMODE were performed) and satellite imagery showed that the oceans are full of eddies of different formation mechanisms and different spatial and temporal characteristics. Among the inhabitants of the eddy zoo [32]: frontal eddies of large-scale currents, open sea eddies, topographic eddies, eddies generated under atmospheric influence, intrathermocline lenses, dipoles, tripoles, etc. (see, for example, [4, 12, 17, 26, 27]). Their diameters vary in the wide range from a few kilometers to over 200 km and lifetimes from a few days to tens of months and even years (in case of intrathermocline lenses). Eddies with a horizontal scale D of order of the Rossby internal radius of deformation R_d relate to mesoscale (synoptic in Russian literature) ones. These geostrophic or quasi-geostrophic eddies with anticyclonic or cyclonic rotation (clockwise or counter-clockwise in the Northern Hemisphere, respectively) are the most studied to date both with in-situ measurements and with satellite observations. They are a powerful mechanism of horizontal (due to the involvement of the surrounding

A. G. Kostianoy (✉) · A. I. Ginzburg
Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
e-mail: kostianoy@gmail.com

A. G. Kostianoy
S.Yu. Witte Moscow University, Moscow, Russia

A. G. Kostianoy
Interfacultary Center for Marine Research (MARE) and Modelling for Aquatic Systems (MAST), University of Liège, Liège, Belgium

O. Yu. Lavrova · M. I. Mityagina
Space Research Institute, Russian Academy of Sciences, Moscow, Russia

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_24

waters on the eddy periphery) and vertical (due to upwelling/downwelling motions of the water at the eddy center) mixing in the ocean.

Observations (photographs in sunglitter areas on the sea surface) of American space oceanographers from the spacecrafts, especially during shuttle mission 41-G, US (1984), have shown that submesoscale (D less than R_d) eddies exist throughout the ocean [23, 29–31]. They have a spiral form, rotate in the cyclonic direction and are observed in the ocean beyond six degrees north and south latitude [31]. Such eddies are also found on the radar images in different regions of the ocean/seas (e.g., [2, 10, 20]).

It is assumed that submesoscale flows are characterized by a larger Rossby number (Ro of O(1)) and higher (on order of magnitude) vertical velocity than the mesoscale, which determines their important role in the vertical exchange between surface and deeper layers of the ocean/sea [35]. However, to date, information on the mechanisms of their generation, life history and vertical structure is clearly insufficient. In-situ measurements in the eddies' areas are practically absent because of their small time of life and small spatial sizes. Therefore, the accumulation of satellite observations of submesoscale vortices in various oceans/seas together with related information on hydrophysical and meteorological conditions is extremely important.

In this paper, observations of submesoscale eddies in the Russian seas (the Black, Caspian, Baltic, and White) are discussed and assumptions are made about possible mechanisms of their formation.

Manifestation of Eddies on the Sea Surface and Data Used

Submesoscale eddies rarely appear in satellite visible band and infrared (IR) images acquired by sensors MODIS Aqua/Terra and AVHRR because of the weak thermal or optical contrasts of the structures relative the surrounding water and insufficient spatial resolution of the sensors (1 km). However, sometimes they were detected in satellite optical images with high spatial resolution (meters–tens of meters): for example, in optical images of the *QuickBird* sensor (resolution of 0.6–2.44 m) [18] or the Landsat-5 Thematic Mapper (TM) sensor (resolution of 30 m) [20].

The main source of information on submesoscale eddies are satellite radar images [10, 11, 13–15, 20, 21]. The advantage of the latter images is the all-weather survey (independence from time of day, and cloudiness) and high spatial resolution (10 to 100 m) in a swath from 100 to 500 km. The best conditions for eddies to be manifested in radar images (RI), wind speed range 2–5 m/s and the presence of surface films of surface-active substances (oil films, films of biogenic origin during the phytoplankton bloom), which smooth the sea surface due to the suppression of small-scale component of sea waves. These films can be found virtually everywhere on the vast sea surface, mostly in warm seasons, they are very sensitive to surface currents and typically form the shape of a local circulation pattern. A manifestation of eddies on the sea surface is due to the involvement of the slick strips in the eddy

movement. At high wind speeds, only contours of the eddy or its center can be drawn on RI due to shear-wave mechanism (the interaction of waves and currents); in this case it is often not possible to determine eddy size and direction of rotation. The manifestation of eddies in radar imagery can be also due to the young, not solid ice.

For the purposes of illustration we shall use in this paper a few visible band images (from the *QuickBird* and the Landsat-5 TM sensors) as well as satellite radar images obtained from the ASAR Envisat, SAR ERS-2, and Sentinel-1.

Submesoscale Eddies in the Russian Seas

The Black Sea

The semi-enclosed Black Sea is characterized by intense mesoscale dynamics. Typical elements of the circulation are mesoscale (~20–100 km) eddies, vortex dipoles, jets, the lifetime of which reaches several months (R_d in the open Black Sea is equal to 15–20 km, in the shelf zone near Gelendzhik it is 8 km [37]). The main results of the study of mesoscale dynamics of the Black Sea and its connection with the large-scale water circulation (the Rim Current propagating cyclonically), dependence on the wind forcing and topography of the continental slope, horizontal and vertical exchange in the active layer of the Black Sea are presented in numerous works (e.g., [6, 7, 20, 24, 33, 36]).

Satellite high-resolution images reveal submesoscale eddies both in the coastal zone and in the deep area of the Black Sea. A set of visible images of the *QuickBird* satellite, which was exhibited in the programme "Google Earth" in 2009 [18], demonstrate the variety of eddies with sizes of 1–10 km in the coastal zone, which appear due to the presence of water with high content of suspended matter and surfactants (see Figs. 1, 2, 3, 4, 5 and 6), the dates of these images were not specified).

Anticyclonic eddy with a diameter of about 7 km is clearly manifested in the region of the villages of Khosta and Kudepsta in Fig. 1. Source of suspended matter is the outflow of the rivers Mzymta and Psou located to the southeast of this point. The formation of this eddy may be caused by the instability of the coastal north-westward current associated with local features of the shoreline.

A cyclonic eddy with a diameter of 4 km is seen in Fig. 2, in the area of the Bakalskaya Spit (north-western coast of Crimea). This eddy is manifested in the field of suspended matter, the source of which may be only local resuspension of waters near the sandy spit under the action of wind waves. It could be formed by rounding the spit with a coastal current.

A cyclonic eddy with a diameter of 2 km is observed at a distance of 5 km from the coast in the area of a northern suburb of Sochi (village of Mamay, Fig. 3). In this case, a spiral vortex is manifested due to the presence on the sea surface of a large amount of surfactant. This cyclone could be a result of short-term wind

Fig. 1 Anticyclonic eddy (AC) offshore of Khosta and Kudepsta (the northeastern Black Sea)

Fig. 2 Cyclonic eddy (C) westward of the Bakalskaya Spit (Crimea)

Fig. 3 Cyclonic eddy (C) offshore of Sochi City (the northeastern Black Sea)

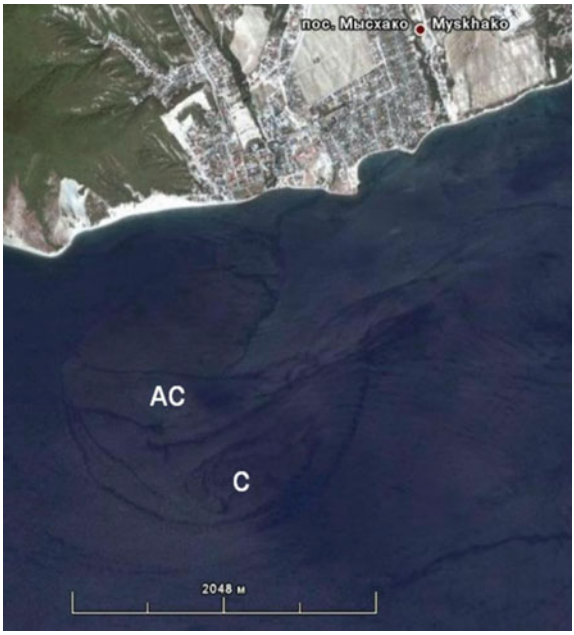


Fig. 4 Vortex dipole offshore of Myskhako (the north-eastern Black Sea)

forcing at the sea surface directed to the south-southeastward, i.e. against the main northward stream of the Rim Current.

A vortex dipole with a horizontal size of about 2 km (the width of the "cap") has been observed in the region of the village of Myskhako in the field of surfactants (Fig. 4). The dipole is oriented from the northeast to southwest; it is probably also caused by the corresponding local wind burst. A source of a number of surfactants

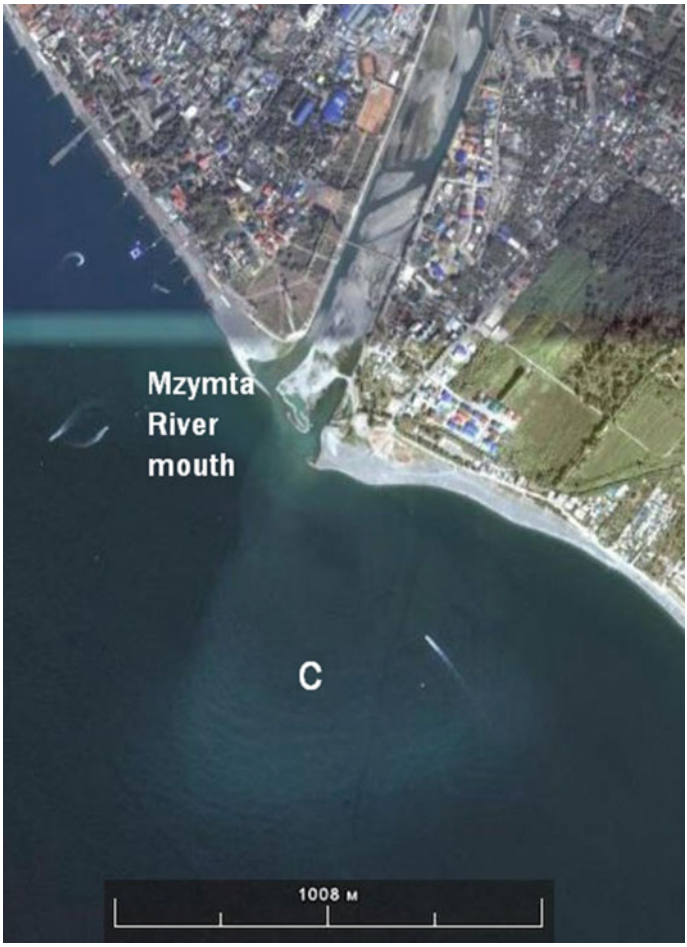


Fig. 5 Cyclonic eddy (C) offshore of the Mzymta River mouth (the northeastern Black Sea)

in this area are the waste water of the village of Myskhako, as well as contaminated surface waters that flowed from Tsemess Bay and the port of Novorossiysk.

In Fig. 5 a small volume of river water of the river Mzymta with high content of suspended matter propagates perpendicular to the shore at a distance of 1 km from the mouth and terminates with a cyclonic eddy in a south-easterly direction, against the usual direction (northwestward) of the coastal water flow. It should be noted that the river plume can form a dipole or a cyclone/anticyclone at its terminal depending on the direction of wind and/or coastal current. This process has been observed in the coastal zone of Adler village. Formation of such eddies is conditioned directly by the river flow transferring the local momentum to the surface layer of the sea.

Two anticyclonic eddies with diameters of 2 and 1 km are seen in Fig. 6 off the coast in the region of the village of Shakhe in the field of suspended matter. The

Fig. 6 Two anticyclonic (A) eddies offshore of Shakhe (the northeastern Black Sea)

larger eddy is located opposite to the mouth of the Shakhe River and may be caused by the river outflow. The smaller eddy can be formed be shear instability of the coastal current.

Submesoscale eddies with a diameter of 2–8 km, both cyclones and anticyclones, are often observed on a narrow (2–10 km) Caucasian shelf in the area of Gelendzhik, north of Cape Idokopas. The generation of anticyclonic eddies in this region is conditioned by two mechanisms: (i) shear (barotropic) instability of the Rim Current propagating over the upper part of the continental slope, and (ii) the separation of the water flow from the shore due to the extension of the shelf more than 2 times west of Cape Idokopas [37]. Observation of such anticyclone in September 2008 (IR data from the NOAA satellites with 1 km spatial resolution, hydrological measurement, and ADCP-survey) during a few days showed that this eddy formed first as submesoscale one with a radius of 3 km and then turned into a mesoscale eddy with a radius of 10 km when moving northwestward. Estimates of the Rossby number resulted in the values of 1.3 and 0.4, respectively [37]. Periodic formation of anticyclonic eddies behind Cape Idokopas was repeatedly observed on MERIS-Envisat satellite images (resolution of 250 m, see [3]).

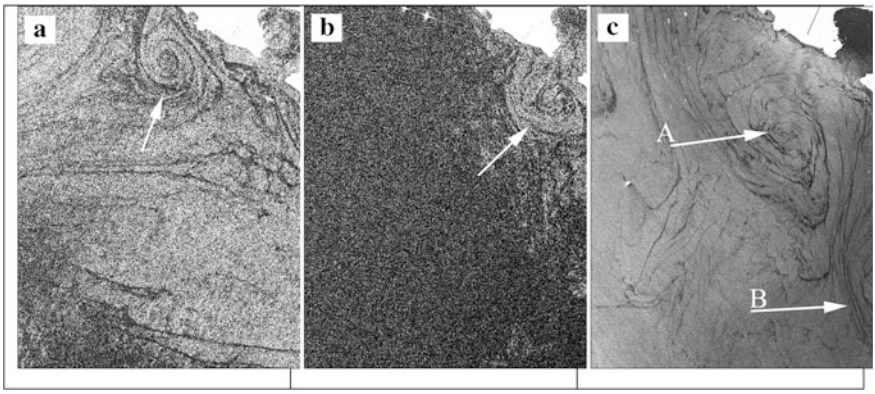


Fig. 7 Examples of submesoscale cyclonic eddies detected in SAR imagery of the northeastern Black Sea offshore of Gelendzhik: **a** part of an Envisat ASAR image of August 3, 2006, 19:13 UTC, showing a cyclonic eddy with a diameter of 3.5 km; **b** part of an Envisat ASAR image of June 26, 2006, 07:32 UTC, showing a cyclonic eddy with a diameter of 5.3 km; **c** part of a TerraSAR-X image of September 13, 2012, 15:15 UTC, showing two cyclonic eddy with diameters of 5 km (A) and 3.5 km (B)

Generation of submesoscale cyclonic eddies occurs, in particular, in the cases of southeastward currents on the shelf associated with the passage of coastal mesoscale anticyclonic eddies that results in the cyclonic shear between the eddy and the shore [20, 37]. Examples of manifestation of submesoscale cyclonic eddies in the region of Gelendzhik Bay are presented in Fig. 7. Survey of velocity field with ADCP towed behind the vessel in the area of a submesoscale cyclone near Gelendzhik in September 2012 showed that the velocity of the orbital motion in the eddy core reached 50 cm/s; active dynamics was observed only in the upper 15–20 m layer, i.e. above the thermocline. The eddy core was characterized by increased concentration of phytoplankton due to the upward motions (upwelling) in its center [38].

Submesoscale eddies in the coastal zone (cyclones, anticyclones and vortex dipoles) can also be associated with coastal upwelling [7, 20]. They are manifested well in the regions of river outflows with increased concentration of surface-active substances. Dense packing of interconnected vortex structures in the Danube Delta region is shown in Fig. 8 (see also Fig. 5.23a in [20]).

Submesoscale eddies are formed in the open part of the sea as well. One of the obvious mechanisms for the occurrence of such short-living entities is the shear instability on the periphery (front) of mesoscale anticyclones (see Fig. 2a in [36]; Fig. 6.13 in [20]). It is not excluded that similar situations follow intense wind forcing. An example of radar image with solitary cyclonic eddy about 8 km in diameter north of Cape Sinop is given in Fig. 2.4 in [14]. Dense packing of multiple spiral vortices 2.5–10 km in diameter in the deep eastern part of the Black Sea in the absence of any front (velocity shear) is presented in Fig. 9 (Fig. 2 in [21]). Another example of four spiral cyclonic vortices with diameters of about 15 km,

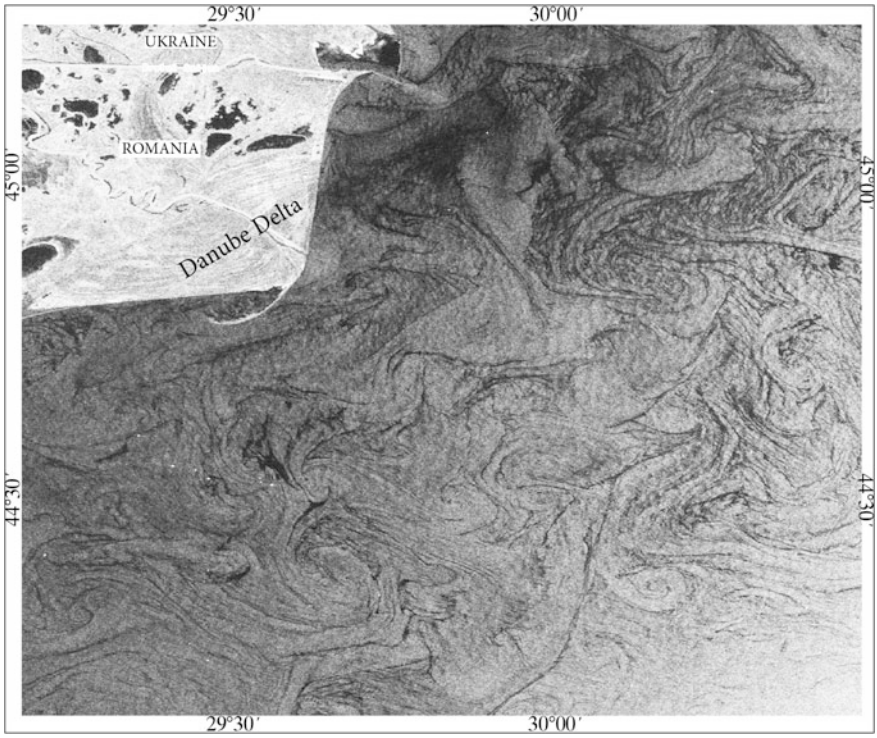


Fig. 8 Dense packing of eddies offshore of the Danube Delta (the northwestern Black Sea). Part of an Envisat ASAR image of August 29, 2009, 08:14 UTC

one interconnected pair of which was located within about 20 km of the shoreline and another pair at a distance about 50 km from the shore northeast of the Bosporus Strait, was given in a number of publications (see [31]; Fig. 1 in [23]; Fig. 6.15 in [20]). The comment to this picture (picture in the sunglitter on September 18, 1992) by Robert Stevenson [31] is interesting: "This spiral-eddy field was in that location throughout the mission and was caught in several photos over three days, although individual eddies could not be identified one day to the next. There is no explanation for the absence of eddies in the waters east and northeast from the Black Sea mouth of the Bosporous". It follows from this comment that the eddy pattern was local, it appeared spontaneously, its time scale was a few days, and the lifetime of individual vortices did not exceed a day. Mechanism or mechanisms of generation of eddy fields in such cases are unknown. Heterogeneity of atmospheric flow (wind) or convection in the surface layer of the ocean were mentioned as the hypothetical mechanisms of formation of solitary spiral eddies, group/fields of such eddies in the open sea [5, 9, 20].

Fig. 9 Manifestation of submesoscale eddies in a radar image of the sea surface during autumn phytoplankton bloom in the eastern part of the Black Sea. A part of a Sentinel-1A image of October 26, 2014, 03:24 UTC

The Caspian Sea

Large-scale circulation of the enclosed Caspian Sea is well studied on the basis of hydrological measurements and modeling, whereas the representation of the mesoscale dynamics of the sea in the warm season (period of less cloud cover) is provided mainly by satellite images in the visible and IR bands [20]. Mesoscale eddies in the Middle and Southern Caspian have a wide range of spatial scales from about 10 km to about 80 km (R_d in the open Caspian Sea is equal to 17–22 km, in the shelf zone it is 3–8 km [1]). The greatest number of submesoscale eddies are found along the western coast of the Middle Caspian and to the south of the Absheron Peninsula, which is connected, apparently, with a high concentration of the surface films due to river runoff and petroleum products [14]. Examples of submesoscale eddies manifestation south of the Absheron Peninsula and in the souteastern part of the sea are given in Figs. 10 and 11, respectively. Vortex dipole in the Northern Caspian (south of the Volga River Delta) manifested on the ERS-2 SAR image is shown in Fig. 13 in [10].

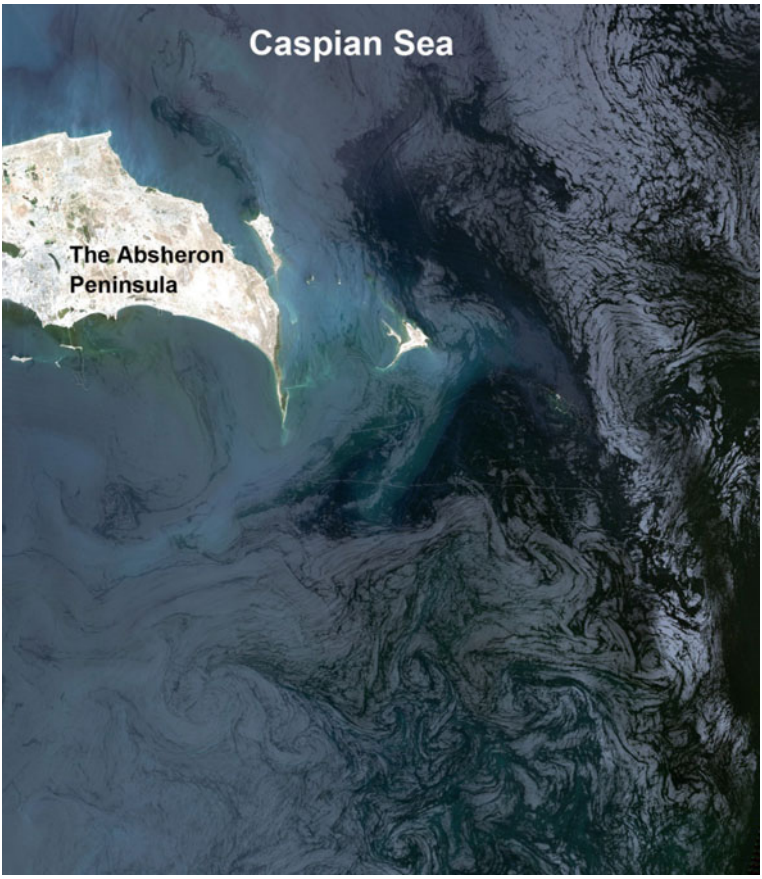


Fig. 10 Submesoscale eddies south of Absheron peninsula (the Southwestern Caspian Sea, Landsat-5, 02.06.2010)

The Baltic Sea

In the semi-enclosed and shallow Baltic Sea there are no intense permanent currents. Drift currents caused by wind are variable in direction and velocity, which usually do not exceed 10–15 cm/s. This sea is characterized by intense mesoscale and submesoscale dynamics [8, 14–16, 19–21, 34]. The typical size of submesoscale eddies in the Baltic is 2.5 km [14] (R_d is 7–10 km in the open sea and 1–6 km in the shelf zone [25]). Examples of manifestation of submesoscale eddies in the Baltic can be found in [8, 16, 20]. Some mechanisms of the formation of submesoscale vortices in the sea are obvious: they are river runoff, frontal instability (vortices at the boundary of mesoscale eddies), vortices at the terminals of filaments in the upwelling zones. However, the mechanism of generation of closely packed eddy fields of the type shown in Figs. 12 and 13 is still unclear.

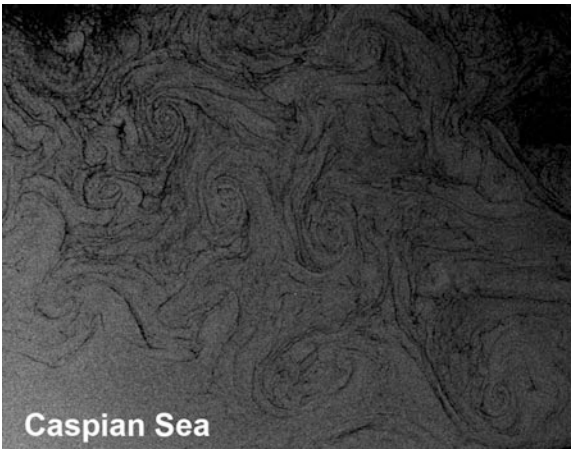


Fig. 11 Submesoscale eddies in the Southeastern Caspian Sea (ASAR Envisat, 19.10.2010)

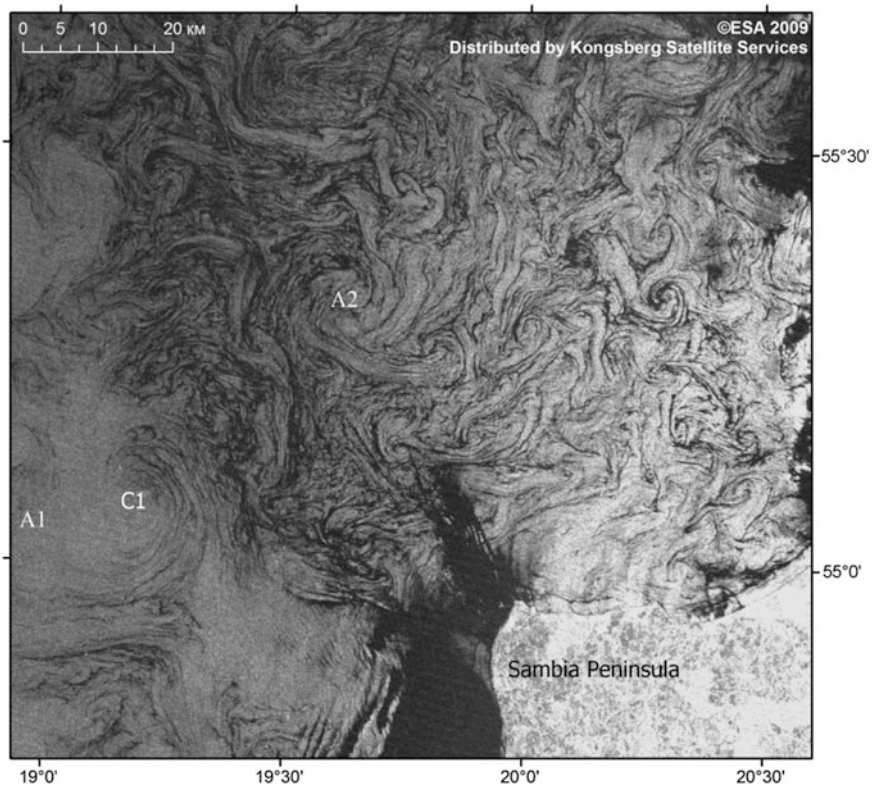


Fig. 12 Eddy field to the west and north of the Sambia Peninsula (ASAR Envisat, April 25, 2009, 09:11 UTC)

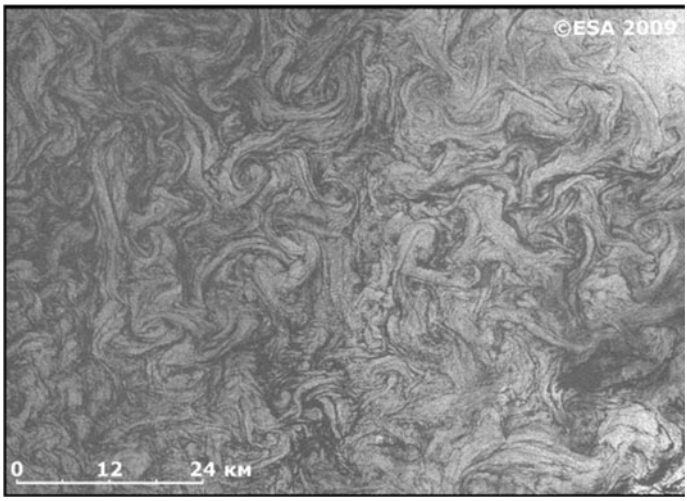


Fig. 13 Submesoscale eddy field west of Latvian coast (ASAR Envisat, April 25, 2009, 09:09 UTC)

Figure 12 demonstrates an example of co-existing of submesoscale and mesoscale eddies. The mesoscale dipole A1–C1 with an eddy diameter of about 10–12 km is seen in the lower left corner of the image. Anticyclone A2 with diameter of about 10 km forms a package of three dipoles with cyclones of a lesser size (3–5 km). The diameters of counterparts of numerous submesoscale dipoles in Figs. 12 and 13 are 2–5 km.

Note that both images were acquired practically at the same time but in different areas: to the west and north of the Sambia Peninsula (Fig. 12) and in the area limited to 55°45′ −56°40′N, 18°30′ −20°55′E (Fig. 13). Obviously, the cause of these eddy fields was the same. It is likely that they were caused by heterogeneities of the atmospheric effects.

Another type of submesoscale eddies in the Baltic Sea is vortex streets of the Karman's street type (see, for example, Fig. 14). Such a vortex street was recorded in the Bering Strait behind Fairway Rock on the ERS-1 SAR image (see Fig. 5 in [10]). However, in the case of the vortex street in Fig. 14, no rocks or islands that might explain its formation exist. Various orientations of streets observed by the authors in this area (north of the Gulf of Gdansk, near 55°30′N) allows us to suggest that their appearance is related to different wind directions in different situations.

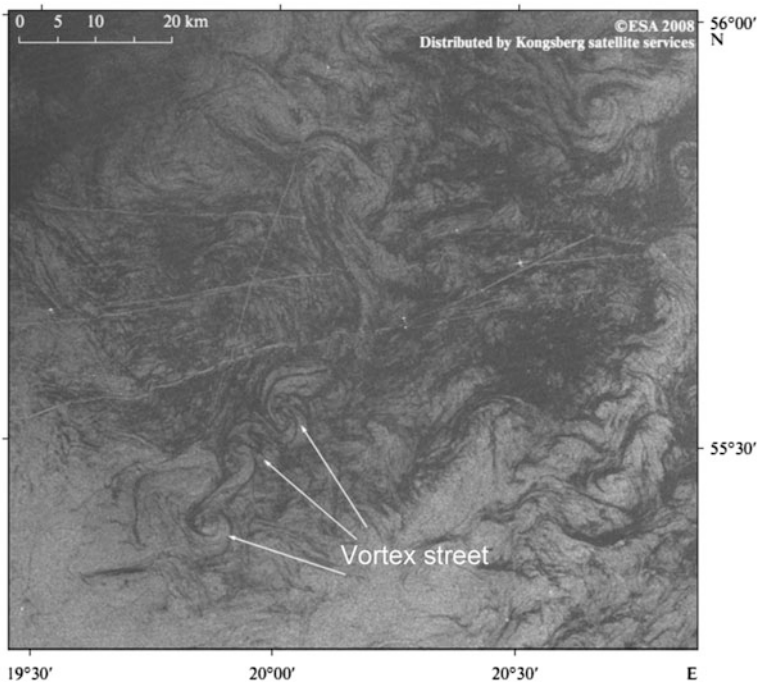


Fig. 14 Vortex street in the Southeastern Baltic Sea (ASAR Envisat, May 17, 2008, 20:16 UTC)

The White Sea

The White Sea is one of the smallest inner seas of Russia with complex topography and intense tidal currents. Radar images and field measurements have shown that submesoscale eddies were observed throughout the waters of the sea, the greatest number of them were recorded near the fronts and in the areas of sharp increase in depth [28, 39]. Characteristic diameters of these eddies are 2–8 km (R_d varied in the range of 1–8 km with the lowest values in the shelf areas with intense tidal mixing and the maximum values in the deep ones [39]. Preferential direction of rotation of the vortices is cyclonic. Short-lived intrapycnocline eddies are also observed in the same phase of tide over the irregularities of bottom relief.

Conclusions

Satellite remote sensing has shown that submesoscale eddies are a common element of the dynamics of the inner Russian seas. They have short lifetime (a day or a few days) and small dimensions (a few kilometers); hence, the main tool for their

observation is satellite information with high spatial resolution. Such eddies as well as mesoscale ones provide to a great extent the horizontal mixing in the open sea and shelf/open sea water exchange, the latter is particularly important for inland seas with a high level of contamination by river discharges, city and tourist resort wastes, etc. Besides, submesoscale eddies can play an important role in the vertical exchange between the surface and deeper layers of the sea/ocean [35]. Measurements with ADCP in a cyclonic submesoscale eddy area in the Black Sea showed that active eddy motions were observed only in the upper 15–20-m layer [38]. Thus, we can put forward an assumption that stratification may be important for their formation.

The mechanisms of generation of submesoscale eddies in the coastal zones of seas are: wind momentum; shear instability of coastal currents; river runoff; interaction of vortices and dissipation; the interaction of flows with small-scale features of the shoreline (headlands, spits) or irregularities of the bottom topography. In the deep sea, the possible generation mechanisms are: frontal instability, including instability at the boundaries of mesoscale structures (eddies, filaments of upwelling). Less obvious mechanisms of formation of submesoscale eddies or fields of such eddies are in the regions where fronts or velocity shear are absent (examples of radar images of such eddy fields in the Black, Caspian, and Baltic seas are presented in Figs. 8, 9, 10, 11, 12, 13). It was suggested in literature that their formation may be a result of the atmospheric effects [5, 20] or convection [5, 9]. We hope that further observations of remote sensing data jointly with the hydrometeorological data will help to clarify this problem.

Acknowledgements Analysis of high resolution optical satellite imagery made by A. G. Kostianoy and A. I. Ginzburg was supported by the Russian Science Foundation Grant 14-50-00095. O. Yu. Lavrova and M. I. Mityagina analyzed SAR imagery and studied submesoscale eddies in the framework and with a support of the Russian Science Foundation Grant 14-17-00555.

References

1. Arkhipkin, V. S., Bondarenko, A. L., Vedev, D. L., & Kosarev, A. N. (1992). Peculiarities of water circulation at eastern coast of the Middle Caspian Sea. *Vodnye Resursy, 6,* 36–43. (in Russian).
2. Eldevik, T., & Dysthe, K. B. (2002). Spiral eddies. *Journal of Physical Oceanography, 32*(3), 851–869.
3. Elkin, D. N., & Zatsepin, A. G. (2013). Laboratory investigation of a mechanism of periodic eddy formation behined capes in a Coastal Sea. *Oceanology, 53*(1), 29–41. (in Russian).
4. Fedorov, K. N., & Ginsburg, A. I. (1986). "Mushroom-like" currents (vortex dipoles) in the ocean and in a laboratory tank. *Annales Geophysicae*, *4*(B, 5), 507–516.
5. Ginzburg, A. I. (1992). Nonstationary eddy motions in the ocean. *Oceanology, 32*(6), 689–694.
6. Ginzburg, A. I. (1994). Horizontal exchange processes in the near-surface layer of the Black Sea. *Earth Observation and Remote Sensing, 12,* 190–202.
7. Ginzburg, A. I., Kostianoy, A. G., Soloviev, D. M., & Stanichny, S. V. (2000). Remotely sensed coastal/deep-basin water exchange processes in the Black Sea surface layer. In D. Halpern (Ed.) *Satellites, oceanography and society* (pp. 273–287). Amsterdam: Elsevier.

8. Ginzburg, A. I., Bulycheva, E. V., Kostianoy, A. G., & Soloviev, D. M. (2015). Vortex dynamics in the southeastern Baltic Sea from satellite radar data. *Oceanology, 55*(6), 805–813.
9. Golitsyn, G. S. (2012). On the nature of spiral eddies on the surface of seas and oceans. *Izvestiya, Atmospheric and Oceanic Physics, 48*(3), 350–354.
10. Ivanov, A. Yu., & Ginzburg, A. I. (2002). Oceanic eddies in synthetic aperture radar images. *Proceedings of Indian Academy of Sciences. Earth and Planetary Sciences*, *111*(3):281–295.
11. Johannessen, J. A., Kudryavtsev, V., Akimov, D., Eldevik, T., Winther, N., & Chapron, B. (2005). On radar imaging of current features: 2. Mesoscale eddy and current front detection. *Journal of Geophysical Research, 110* (C07017). https://doi.org/10.1029/2004JC002802.
12. Kamenkovich, V. M., Koshlyakov, M. N., & Monin, A. S. (1987). *Synoptic eddies in the ocean*. Hydrometeoizdat, Leningrad (in Russian), 511 p.
13. Karimova, S. S. (2010). On manifestation of vortical structures in satellite radar images. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa, 7*(3), 152–160. (in Russian).
14. Karimova, S. S. (2012). *Investigation of submesoscale vortices of the Baltic, Black and Caspian Seas based on satellite radar data*. PhD Thesis, Moscow, 187 p.
15. Karimova, S. S. (2012). Statistical Analysis of submesoscale eddies in the Baltic, Black and Caspian Seas using satellite SAR images. *Issledovanie Zemli iz kosmosa*, *3*, 31–47.
16. Karimova, S. S., Lavrova, O. Yu., & Soloviev, D. M. (2012). Observation of eddy structures in the Baltic Sea with the use of radiolocation and radiometric satellite data. *Izvestiya, Atmospheric and Oceanic Physics*, *48*(9), 1006–1013.
17. Kostianoy, A. G., & Belkin, I. M. (1989). A survey of observations on intrathermocline eddies in the World Ocean. In J. C. J. Nihoul, B. M. Jamart (Eds.), *Mesoscale/synoptic coherent structures in geophysical turbulence. Proceedings of 20th International Liege Colloq. Ocean Hydrodynamics* (pp. 821–841). Amsterdam: Elsevier.
18. Kostianoy, A. G., Ginzburg, A. I., Sheremet, N. A., Lavrova, O. Yu., & Mityagina, M. I. (2010). Small-scale eddies in the Black Sea. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa, 7*(1), 248–259. (in Russian).
19. Lavrova, O. Yu. (2005). Slicks as indicators of eddy activity in the coastal area. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa, 2,* 118–123. (in Russian).
20. Lavrova, O. Yu., Kostianoy, A. G., Lebedev, S. A., Mityagina, M. I., Ginzburg, A. I., & Sheremet, N. A. (2011). *Complex satellite monitoring of the Russian seas*. IKI RAS: Moscow, 424 p. (in Russian).
21. Lavrova, O. Yu., & Mityagina, M. I. (2016). Manifestation specifics of hydrodynamic processes in satellite images of intense phytoplankton bloom areas. *Izvestiya, Atmospheric and Oceanic Physics, 52*(9), 974–987.
22. Mityagina, M. I., Lavrova, O. Yu., & Karimova, S. S. (2010). Multi-sensor survey of seasonal variability in coastal eddy and internal wave signatures in the north-eastern Black Sea. *International Journal of Remote Sensing, 31*(17), 4779–4790.
23. Munk, W., Armi, L., Fischer, K., & Zachariasen, F. (2000). Spirals on the sea. *Procedings of the. Royal Society of London A, 456,* 1217–1280.
24. Oguz, T., Latun, V. S., Latif, M. A., Vladimirov, V. V., Sur, H. I., Markov, A. A., et al. (1993). Circulation in the surface and intermediate layers of the Black sea. *Deep-Sea Research, 40,* 1597–1612.
25. Osinski, R., Pak, D., Walczowski, W., & Piechura, J. (2010). Baroclinic radius of deformation in the southern Baltic Sea. *Oceanologia, 52*(3), 417–429.
26. Richardson, P. L., Bower, A. S., & Zenk, W. (2000). A census of Meddies tracked by floats. *Progress in Oceanography, 45,* 209–250.
27. Robinson, A. R. (Ed.) (1983). *Eddies in marine science*. Springer, 609 p.
28. Rodionov, A. A., Zimin, A. V., Kozlov, I. E., & Shapron, B. (2014). Submesoscale structures and dynamics in the White Sea. State of art and directions of research. *Fundamentalnaya i prikladnaya gidrofizika, 7*(3), 29–41. (in Russian).

29. Scully-Power, P. (1986). Navy Oceanographer Shuttle Observations, STS 41-G, Mission Report. In *Naval underwater systems center technical report NUSC TD* 7611. 71 p.
30. Stevenson, R. E. (1989). *Oceanography from the Space Shuttle. Office of Naval Research*. The University Corporation for Atmospheric Research. 200 p.
31. Stevenson, R. E. (1998). Spiral eddies: The discovery that changed the face of the oceans. *21st Century Science and Technology, 11*, 58–71.
32. Stommel, H., Meinke, J., & Zenk, W. (1977). New animals for the eddy zoo. *Polymode News* (Unpublished Newsletter), *22*(1).
33. Sur, H. I., & Ilyin, Yu. P. (1997). Evolution of satellite derived mesoscale thermal patterns in the Black Sea. *Progress in Oceanography, 39,* 109–151.
34. Tavri, A., Singha, S., Lehner, S., & Topouzelis, K. (2016). Observation of sub-mesoscale eddies over Baltic Sea using TERRASAR-X and oceanographic data. In *Proceedings of Living Planet Symposium 2016*, Prague, Czech Republic, 9–13 May 2016 (ESA SP-740, August 2016).
35. Thomas, L. N., Tandon, A., & Mahadevan, A. (2008). Submesoscale processes and dynamics. *Ocean Modeling in an Eddying Regime. Geophysical Monograph Series* Vol. 177:17–37. The American Geophysical Union, Washington, DC. https://doi.org/10.1029/177gm04.
36. Zatsepin, A. G., Ginzburg, A. I., Kostianoy, A. G., Kremenetsky, V. V., Krivosheya, V. G., Stanichny, S. V., et al. (2003). Observation of Black Sea mesoscale eddies and associated horizontal mixing. *Journal of Geophysical Research, 108*(C8), 3246. https://doi.org/10.1029/2002JC001390.
37. Zatsepin, A. G., Kondrashov, A. A., Korzh, A. O., Kremenetskiy, V. V., Ostrovskii, A. G., & Soloviev, D. M. (2011). Submesoscale eddies at the Caucasus Black Sea shelf and the mechanisms of their generation. *Oceanology, 51*(4), 554–567.
38. Zatsepin, A. G., Ostrovskii, A. G., Kremenetskiy, V. V., Piotoukh, V. B., Kuklev, S. B., Moskalenko, L. V., Podymov, O. I., et al. (2013). On the nature of short-period oscillations of the main Black Sea pycnocline, submesoscale eddies, and response of the marine environment to the catastrophic showers of 2012. *Izvestija RAN. Fizika Atmosphery i Okeana, 49*(6), 717–732. (in Russian).
39. Zimin, A. V., Atadzhanova, O. A., Romanenkov, D. A., Kozlov, I. E., & Chapron, B. (2016). Submesoscale eddies in the White Sea based on satellite SAR data. *Issledovanie Zemli iz kosmosa, 1–2,* 129–135. (in Russian).

Ship-Based Monitoring of the Northern North Atlantic Ocean by the Shirshov Institute of Oceanology. The Main Results

Artem Sarafanov, Anastasia Falina, Alexey Sokov, Vyacheslav Zapotylko and Sergey Gladyshev

Introduction

The Meridional overturning circulation (MOC) in the North Atlantic is one of the main drivers of the widely known global oceanic "conveyor belt", an important element of the Earth's climate system (e.g., [1]). Warm upper-ocean waters transported northward by the North Atlantic Current release heat to the atmosphere, gain density due to cooling, and eventually sink in the subpolar North Atlantic and adjacent Arctic seas thereby generating the return southward flow of colder waters at depths (Fig. 1) [2, 3]. Temporal variability of the large-scale circulation and associated heat transport in the subpolar North Atlantic is one of the principal factors behind the high-latitude climate anomalies in the Northern Hemisphere.

Progress in understanding the causes of the ongoing climate change and forecasting climate variability in the Arctic and over the European part of Russia for the next decades require reliable observation-based estimates of the variability of the North Atlantic circulation and the Atlantic–Arctic heat and freshwater fluxes, as well as elucidation of the underlying mechanisms. In a number of recent studies, radical changes in the thermohaline regime and large-scale circulation in the Atlantic Ocean have been suggested to occur under global warming. For example, the long-term freshening of the subpolar North Atlantic deep waters since the mid-1960s [9] has been (cautiously) attributed to climate change-related factors [10, 11]. Hypothetically, under global warming, an increased evaporation in the tropics and increased precipitation at high latitudes, coupled with an intensified melting of Arctic ice, lead to the upper-ocean freshening in the regions of deep water formation and, hence, to the deep water freshening in the Atlantic Ocean. At the same time, milder winters along with the upper-ocean freshening lead to a decrease in the

A. Sarafanov (✉) · A. Falina · A. Sokov · V. Zapotylko · S. Gladyshev
Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
e-mail: sarafanov@mail.ru

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_25

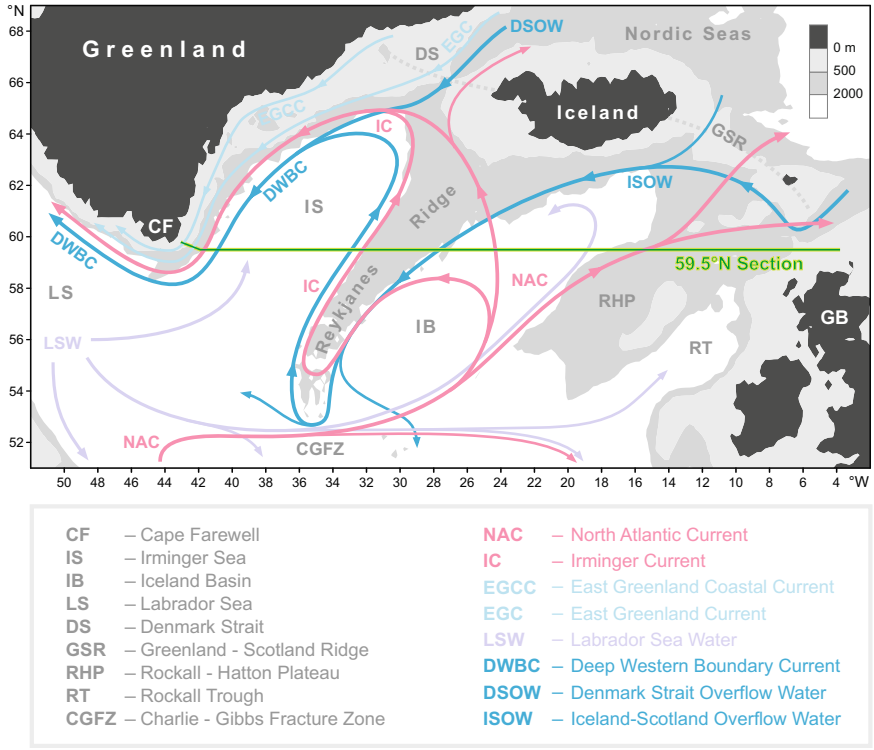


Fig. 1 Schematic diagram of the large-scale circulation in the northern North Atlantic compiled from [4–7]. Abbreviations for the main topographic features, currents, and water masses are explained in the legend. The nominal location of the 59.5° N hydrographic section (1997—present) is shown with the solid green line. Adapted from [8]

deep water production rates, which results in slowing of the Atlantic MOC (e.g., [11, 12]

To better understand the past and present changes in the ocean-atmosphere dynamical system, as well as their causes and consequences, data on the full-depth oceanic variability are needed. An indispensable effective tool for assessing the large-scale circulation and thermohaline changes in the deep ocean and investigating mechanisms governing these changes are repeated full-depth transoceanic observations.

Since 1997, the Shirshov Institute of Oceanology has carried out the long-term monitoring of the North Atlantic circulation and water mass properties along the 59.5° N hydrographic section between Cape Farewell (Greenland) and Scotland (Fig. 1). Since 2002, the section has been repeated yearly on board the Russian research vessels, providing high precision data on temperature, salinity, oxygen and nutrients concentrations, and current velocities in the entire water column—"from

shore to shore", from the sea surface to the bottom. Based on the unique data set thus collected, a number of fundamental findings have been achieved. Below, we briefly summarize the main subjects and results of our research.

The 59.5° N transatlantic section (Fig. 1) was designed for monitoring the large-scale circulation and thermohaline/chemical properties of oceanic waters at the northern periphery of the North Atlantic, the region where the warm upper-ocean waters are transformed by deep convection and mixing into the colder intermediate and deep waters, the Labrador Sea Water (LSW), Iceland Scotland Overflow Water (ISOW) and Denmark Strait Overflow Water (DSOW) (Fig. 1) transported southward in the lower limb of the Atlantic MOC. Hydrographic data collected at 59.5° N along with those obtained within the framework of the kindred projects, primarily the French OVIDE (http://www.ifremer.fr/lpo/ovide), and historical data sets have been used for studying the dense water production [13, 14], decadal temperature and salinity changes in the intermediate–deep water column [15–17], causes of these changes [17, 18], the mean state [8] and long-term variability of the large-scale circulation in the region [19–21].

Deep Convection in the Irminger Sea

The oxygen data collected in 1997 in the northern North Atlantic in several sections ending nearby the southern tip of Greenland provided the observation-based support for the hypothesis [22] that winter convection in the Irminger Sea may penetrate deep into the LSW layer (1000–2000 m), thus causing local renewal of this water mass. A separate lateral maximum of oxygen concentrations in the deep LSW layer was detected east of Cape Farewell (59.5° N, 36–40° W): the concentrations increased (by ~0.1 ml/l) from the Labrador Sea eastern edge toward the Irminger Sea (Fig. 2) rather than the reverse, as would be expected if LSW observed in the Irminger Sea interior in 1997 were solely of advective origin [13].

Reversal of the Deep-Water Freshening

The LSW and Nordic Seas overflow-derived deep waters, ISOW and DSOW, freshened in the northern North Atlantic during the last three–four decades of the 20th century [9]. Between the 1960s and 1990s, the water column in the region freshened on average by about 0.03 [10].

The long-term freshening reversed in the mid-1990s [15–17]. The salinification (and warming) of the intermediate and deep waters since the mid-1990s (Fig. 3) was much more intense than the preceding freshening. Over nearly a decade (1997–2006), temperature/salinity in the intermediate–deep water column ($\sigma_0 \geq 27.45$, depths > 500–1000 m) at 59.5° N increased by ~0.3 °C/0.03–0.04 [16].

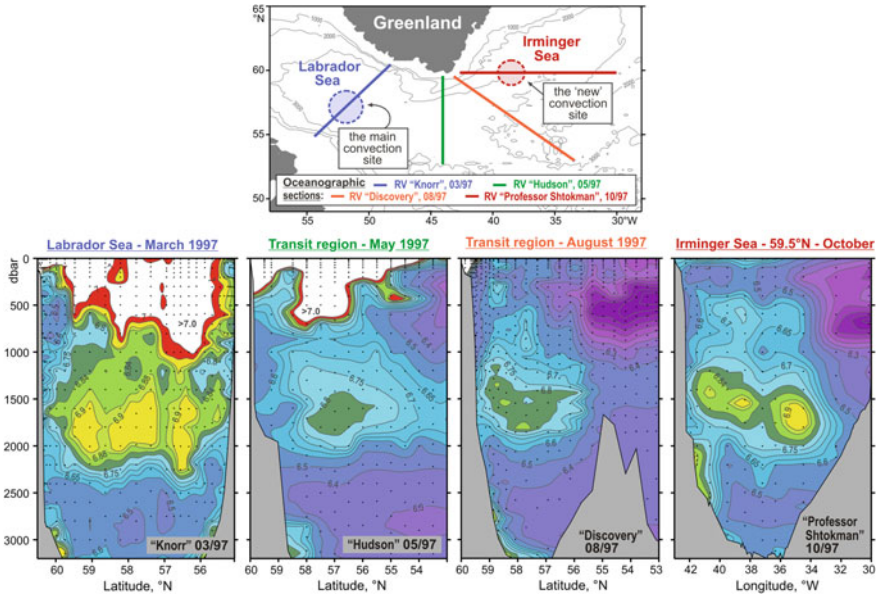


Fig. 2 Oxygen concentrations (ml/l) in the water column (lower panel) as observed in March–October 1997 in four hydrographic sections (upper panel) ending nearby the southern tip of Greenland. A separate oxygen maximum in the (LSW) layer (1000–2000 m) in the Irminger Sea at 59.5° N strongly implies local convective renewal of LSW before 1997.
Adapted from [13]

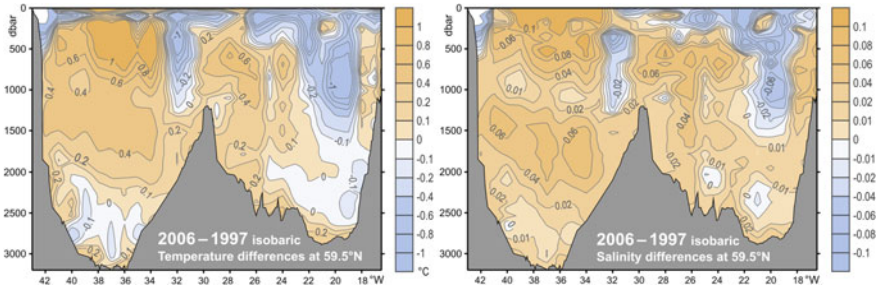


Fig. 3 Warming and salinification in the northern North Atlantic between the mid-1990s and mid-2000s, as observed at 59.5° N. The figure shows the 2006–1997 temperature (°C, left) and salinity (right) differences on isobaric surfaces in the Irminger Sea and Iceland Basin.
Adapted from [15]

In the Irminger Sea, the long-term freshening in the deep water column ($\sigma_0 > 27.80$, depths $>\sim 2000$ m) reversed in the early 2000s [17]. The observed freshening reversal was a lagged consequence of the persistent ISOW salinification

that occurred upstream, in the Iceland Basin, after 1996 due to salinification of the Northeast Atlantic waters entrained into the overflow.

It was demonstrated [17] that the entrainment salinity increase was associated with the North Atlantic Oscillation (NAO)-induced weakening and contraction of the Subpolar Gyre and corresponding northwestward advance of subtropical waters that followed the NAO decline in the mid-1990s and continued through the mid-2000s. Remarkably, the deep water freshening reversal was not related to changes in the overflow water salinity.

Deep-Ocean Salinity Changes and the NAO

Close relationship between the thermohaline properties of the northern North Atlantic intermediate and deep waters and the winter NAO index on a decadal time scale ($r^2 \approx 0.65$, 1950s–2000s, Fig. 4b, c) was revealed [18] from the observation-based salinity time series for LSW in the Labrador Sea [23] and ISOW in the Iceland Basin [15, 24]. Persistent NAO decline (amplification) leads to warming and salinification (cooling and freshening) in the intermediate–deep water column.

An explanation for the close link between the NAO and the coherent decadal changes in the intermediate and deep water properties in the region was proposed [18]. The two factors dominate this link (Fig. 4d): (i) intensity of convection in the Labrador Sea controlling injection of relatively cold fresh waters into the intermediate layer and (ii) zonal extent of the Subpolar Gyre that regulates the relative contributions of cold fresh subpolar waters and warm saline subtropical waters to the entrainment into the Norwegian Sea overflow south of the Iceland–Scotland Ridge and to the Atlantic inflow to the Nordic Seas. These factors act in phase leading to the observed coherent thermohaline changes in the intermediate–deep water column.

Due to weakening of the surface forcing associated with the NAO transition into neutral to low phase (1950s to mid-1960s, mid-1990s to mid-2000s), convection in the Labrador Sea weakens diminishing cold fresh water penetration into the intermediate layer. This results in warming and salinification at the intermediate depths in the Subpolar Gyre. Concurrently, the Subpolar Gyre contracts allowing northward advance of warm saline upper-ocean and intermediate subtropical waters in the northeastern North Atlantic. Northward progression of subtropical waters increases temperature and salinity at the upper intermediate levels and, correspondingly, increases temperature and salinity of the Northeast Atlantic waters entrained into the Iceland–Scotland overflow along its pathway to the deep Iceland Basin. As a result, temperature and salinity at the deep levels increase. The contrary changes: intensification of deep convection in the Labrador Sea and expansion of the Subpolar Gyre, caused by amplifying surface forcing (mid-1960s to mid-1990s) lead to cooling and freshening at the intermediate–deep levels. Additionally, under high-NAO conditions, deep convection may occur in the Irminger Sea potentially

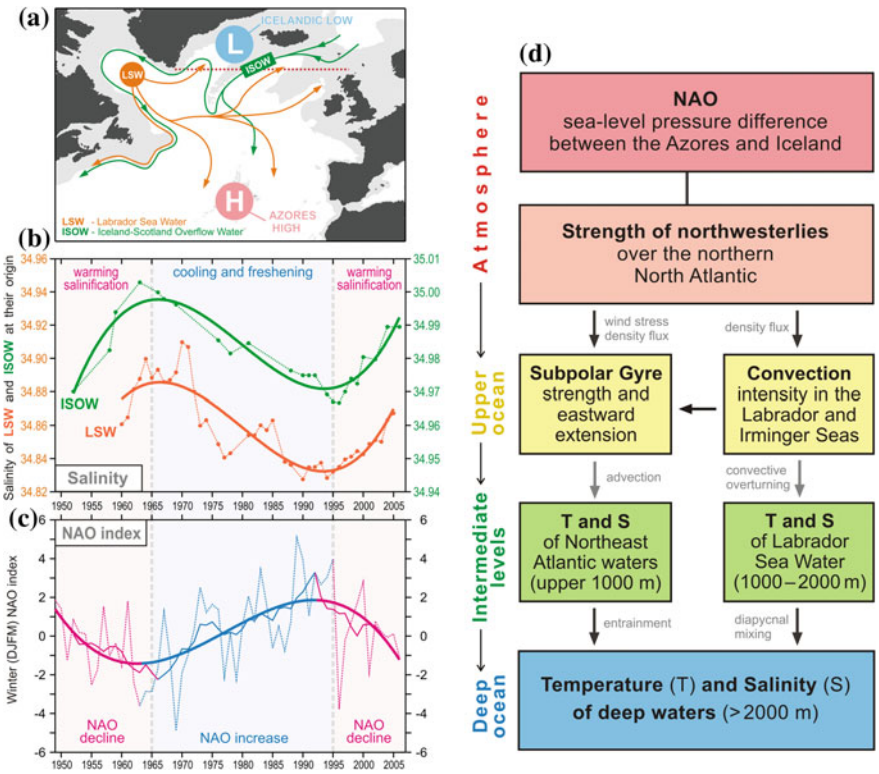


Fig. 4 Coherence of the decadal salinity changes (1950s–2000s) of the intermediate (LSW) and deep (ISOW) waters in the northern North Atlantic and their link to the North Atlantic oscillation (NAO) index. **a** Schematic representation of the LSW and ISOW pathways and locations of the Icelandic Low (L) and Azores High (H) centers constituting the NAO dipole pattern. The red dotted line indicates the 59.5° N transatlantic section. **b** Salinity time series for LSW in the Labrador Sea [23] and ISOW in the Iceland Basin [15, 24] overlaid by the third order polynomial fits. **c** Time series of the winter NAO index, after [25], overlaid by 7-year running mean and third order polynomial fit. **d** Mechanism of the NAO effect on the decadal changes in temperature (T) and salinity (S) of the northern North Atlantic intermediate and deep waters. Positive/negative links shown with the dark/light grey arrows mean that changes in 'causative' and 'consequential' characteristics have the same/opposite sign(s). The overall effect of the NAO on T and S of the water column is negative: persistent NAO decline leads to warming and salinification of the water masses and vice versa, as shown in (**b**) and (**c**).
Adapted from [18]

contributing to cooling and freshening at the intermediate (LSW) levels. The two regimes of convection and large-scale circulation corresponding to stronger (early 1990s) and weaker (mid-1960s, mid-2000s) NAO-related atmospheric forcing are schematically visualized in Fig. 5.

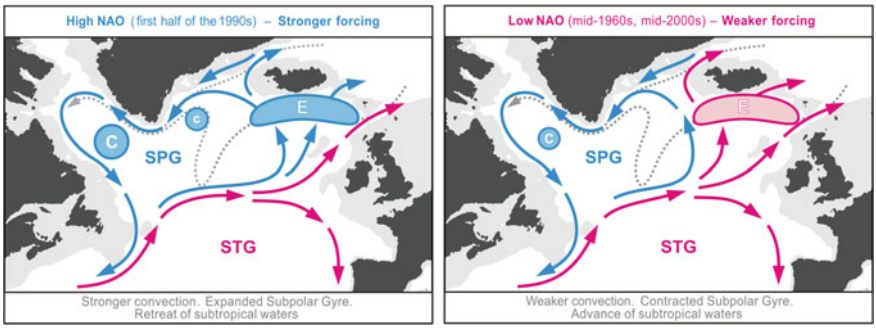


Fig. 5 Schematic representation of the upper-ocean circulation and convection intensity in the northern North Atlantic under high (left) and low (right) NAO conditions. Blue (magenta) solid arrows indicate the upper-ocean flows with higher fraction of colder fresher subpolar (warmer saltier subtropical) waters. The main pathways of the Nordic overflow-derived deep waters are shown with the dotted curves. "C" and "E" symbols are used to denote, respectively, the deep convection sites and the domain, where the Atlantic waters are entrained into ISOW. Larger (smaller) circles indicate stronger (weaker) convection. SPG and STG are the subpolar and subtropical gyres, respectively.
Adapted from [18]

Deep-Ocean Salinity Changes and Climate Change

There are increasing concerns that in the warmer climate, the MOC may substantially decline due to a decrease in the convective activity in the northern North Atlantic and Nordic Seas (e.g., [26]). The long-term freshening in the Nordic Seas and freshening of the northern North Atlantic deep waters in the 1960s–1990s have been considered as a likely indicator or precursor of the dramatic change in the MOC (e.g., [11]). The freshening has been attributed to a combination of factors potentially associated with the global warming: the increasing ice melt and net precipitation at high latitudes (e.g., [10]). A probable causality between the climate change and the decreasing North Atlantic deep water salinity has supported the concerns and unfavorable predictions, thus 'warming up' the reasonable scientific debate on climate change and overblown speculations in media.

Despite the long-term increase in freshwater input to the Arctic, freshening in the northern North Atlantic had reversed in the mid-1990s, as we demonstrated above. This reversal forces us to revise the hypotheses on the mechanisms behind the deep-water thermohaline anomalies. It seems doubtful that the persistent global temperature growth may lead to the opposite decadal trends (positive-then-negative-then-positive, Fig. 4) in the deep water salinity.

Our results [16–18] suggest that natural atmospheric variability over the North Atlantic plays the major role in the deep-water thermohaline variability on a decadal time scale. There are no reasons to associate the deep-water freshening in the 1960s–1990s with climate change, unless the 3-decade-long surface forcing amplification is evidently shown to be a consequence of the latter. Having said that,

the net 1950s–2000s trends in the water mass salinities are negative implying that the global factors (e.g., probable intensification of hydrological cycle [10]) may act on longer time scales.

Decadal Variability of the Deep Western Boundary Current at Cape Farewell

Recent decadal changes in the Deep Western Boundary Current (DWBC) transport southeast of Cape Farewell were assessed from the hydrographic data (1991–2007, Fig. 6a), direct velocity measurements (2002–2006) and satellite altimetry (1992–2007). Following the approach used in earlier studies (e.g., [27]), we first determined that the DWBC ($\sigma_0 > 27.80$) baroclinic transport (T_{BC}) referenced to 1000 m depth increased by ~2 Sv between the mid-1990s (1994–1997) and 2000s (2000–2007) (Fig. 6b) [19]. In the next step, we quantified velocity changes at the reference level (1000 m) by combining estimates of the hydrography-derived velocity changes in the water column and the altimetry-derived velocity changes at the sea surface [20]. The inferred increase in the southward velocity at 1000 m above the DWBC in 1994–2007 indicates that the increase in the DWBC absolute transport was larger but very close to the 2 Sv increase in the DWBC T_{BC}. This result along with the observed coherence of the DWBC absolute and baroclinic transport changes between individual observations [20] imply that the DWBC absolute transport variability in the region is well represented by its baroclinic component on the decadal and shorter time scales.

The historical record of the DWBC T_{BC} (1955–2007, Fig. 6c) updated after Bacon [27] shows distinct decadal variability (±2–2.5 Sv) with the transport minima in the 1950s and mid-1990s, maximum in the early 1980s and moderate-to-high transport in the 2000s. The DWBC T_{BC} decadal variability is consistent with the general pattern of the recent decadal hydrographic and circulation changes in the northern North Atlantic. The DWBC T_{BC} anomalies negatively correlate ($R = -0.80$, 1955–2007) with the thickness anomalies of LSW at its origin implying a close link between the DWBC transport southeast of Cape Farewell and LSW production in the Labrador Sea (Fig. 6d). During the recent three decades (late 1970s–late 2000s), the DWBC T_{BC} changes were also in-phase with the changes in the strength and zonal extent of the Subpolar Gyre [20]. In particular, the Gyre weakening at shallow levels in the mid-1990s–mid-2000s was accompanied by the DWBC strengthening in the Irminger Sea [19–21]. The results imply that the decadal changes in (i) LSW production, (ii) SPG strength, and (iii) DWBC transport in the Irminger Sea are linked, representing a complex coherent oceanic response to the decadal variability of the surface forcing.

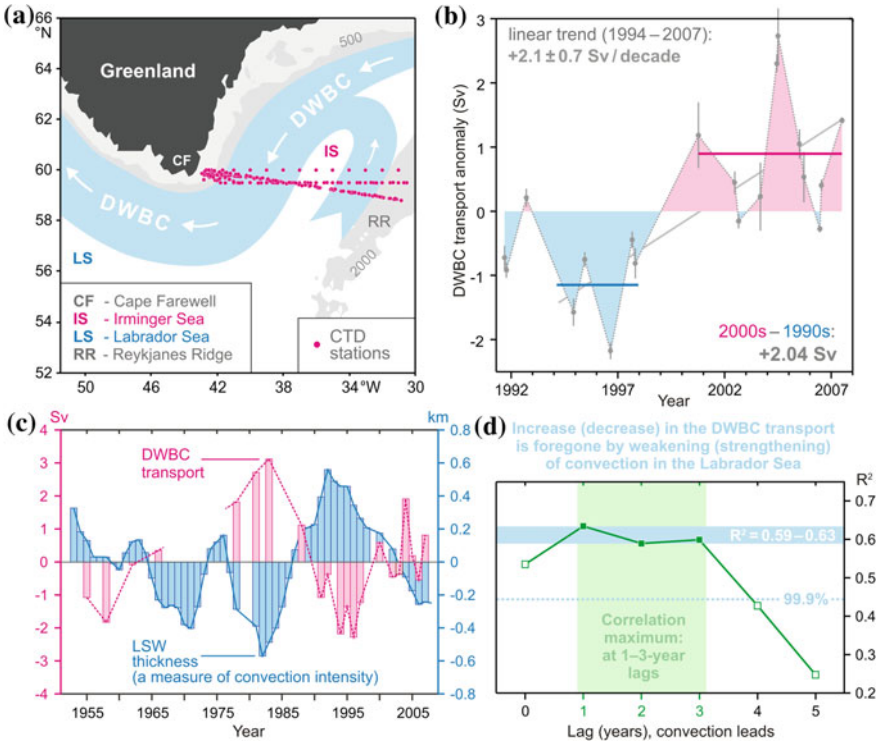


Fig. 6 The deep western boundary current (DWBC) transport variability and its link to the convection intensity in the Labrador Sea. **a** Locations of the hydrographic sections (1991–2007) and schematic of the deep water circulation in the Irminger Sea. **b** The DWBC transport anomalies at Cape Farewell in 1991–2007, 1 Sv = 10^6 m^3 s^{-1}. The 1994–1997 and 2000–2007 mean anomalies and the 1994–2007 linear trend are shown. **c** Anomalies of the DWBC transport at Cape Farewell and LSW thickness in the Labrador Sea in the 1950s–2000s. **d** Correlation coefficient (R^2) for the two time series shown in (**c**) at the 0–5-year lag, the LSW thickness leads. The correlation maximum is achieved at the 1–3-year lag. The DWBC transport anomalies in the southern Irminger Sea are foregone by the convection intensity anomalies in the Labrador Sea. Adapted from [19]

Mean State of the Full-Depth Circulation in the 2000s

The mean state of the full-depth summer circulation in the Atlantic Ocean in the region in between Cape Farewell (Greenland), Scotland and the Greenland-Scotland Ridge (Fig. 1) was assessed by combining 2002–2008 yearly hydrographic measurements at 59.5° N, mean dynamic topography, satellite altimetry data, and available estimates of the Atlantic–Nordic Seas exchange [8]. The mean absolute transports by the upper-ocean, mid-depth and deep currents and the MOC (MOCσ = 16.5 ± 2.2 Sv, at σ_0 = 27.55) at 59.5° N were quantified in the density

space. Inter-basin and diapycnal volume fluxes in between the 59.5° N section and the Greenland-Scotland Ridge were then estimated from a box model.

The estimated meridional and diapycnal volume fluxes contributing to the MOC are schematically visualized in Fig. 7. The dominant components of the meridional exchange across 59.5° N are the North Atlantic Current (NAC, 15.5 ± 0.8 Sv, $\sigma_0 < 27.55$) east of the Reykjanes Ridge, the northward Irminger Current (IC, 12.0 ± 3.0 Sv) and southward Western Boundary Current (WBC, 32.1 ± 5.9 Sv) in the Irminger Sea and the deep water export from the northern Iceland Basin (3.7 ± 0.8 Sv, $\sigma_0 > 27.80$). About 60% (12.7 ± 1.4 Sv) of waters carried in the MOCσ upper limb ($\sigma_0 < 27.55$) by the NAC/IC across 59.5° N (21.1 ± 1.0 Sv) recirculates westwards south of the Greenland-Scotland Ridge and feeds the WBC; 80% (10.2 ± 1.7 Sv) of the recirculating NAC/IC-derived upper-ocean waters gains density of $\sigma_0 > 27.55$ and contributes to the MOCσ lower limb. Accordingly, the contribution of light-to-dense water conversion south of the Greenland-Scotland Ridge (~10 Sv) to the MOCσ lower limb at 59.5° N is one and a half times larger than the contribution of dense water production in the Nordic Seas (~6 Sv).

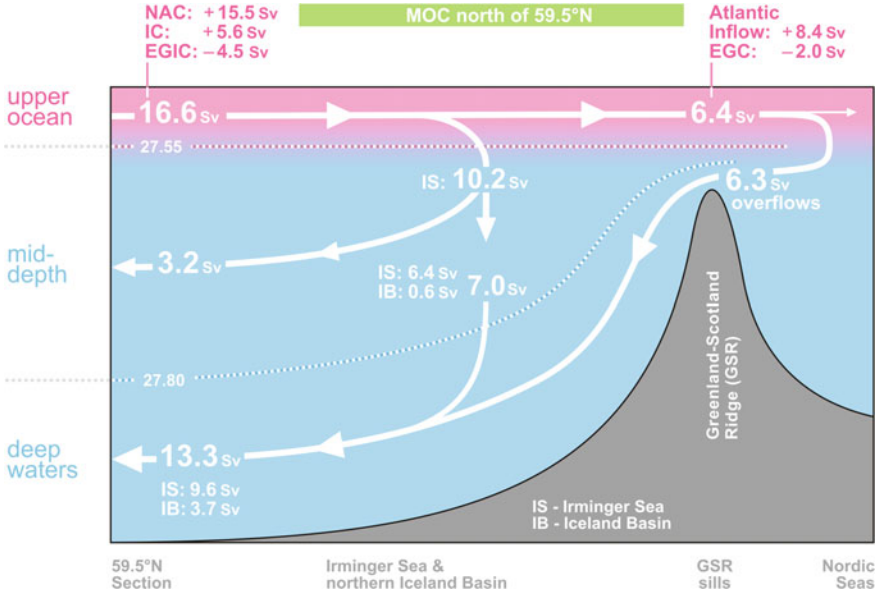


Fig. 7 Schematic diagram of the meridional overturning circulation (MOC) at the northern periphery of the Atlantic Ocean, northeast of Cape Farewell. The dotted lines refer to the σ_0 isopycnals 27.55 and 27.80. The arrows denote the integral meridional and diapycnal volume fluxes. Where the signs are specified, the positive (negative) transports are northward (southward). The NAC and EGIC transports in the upper layer ($\sigma_0 < 27.55$) at 59.5° N are the throughputs accounting for the recirculations. The East Greenland/Irminger Current (EGIC) refers to the upper part of the western boundary current. Other abbreviations are explained in the legend to Fig. 1. Adapted from [8]

Cascading of Dense Shelf Waters in the Irminger Sea

Based on the hydrographic data collected at 59.5° N, 64.3° N and 65–66° N in the western Irminger Sea in the 1990s–2000s, an observational evidence for the deep-reaching cascading of dense shelf waters south of the Denmark Strait was found [14]. The data collected in the northwestern Irminger Sea (65–66° N) indicate that the East Greenland Current ~200 km south of the Denmark Strait occasionally carries shelf waters as dense as the overflow-derived deep waters transported by the DWBC ($\sigma_0 > 27.80$). Hydrographic traces of cascading of dense shelf waters down the East Greenland slope were found from repeat measurements at 64.3° N, where the densest fresh plumes were observed within the DWBC ($\sigma_0 > 27.80$) (Fig. 8). Using the data collected at 59.5° N, we showed that the fresh 'signals' originating from the shelf can be traced in the DWBC as far downstream as the latitude of Cape Farewell, where the anomalously fresh oxygenated plumes are repeatedly observed in the ISOW and DSOW density classes.

The results of our analysis along with the results from earlier studies (e.g., [28, 29]) indicate that shelf water cascading in the northern Irminger Sea is an intermittent process occurring in all seasons of the year. This implies that, despite the apparent short duration of a particular cascading event, the cumulative contribution of such events to the thermohaline variability and southward export of the

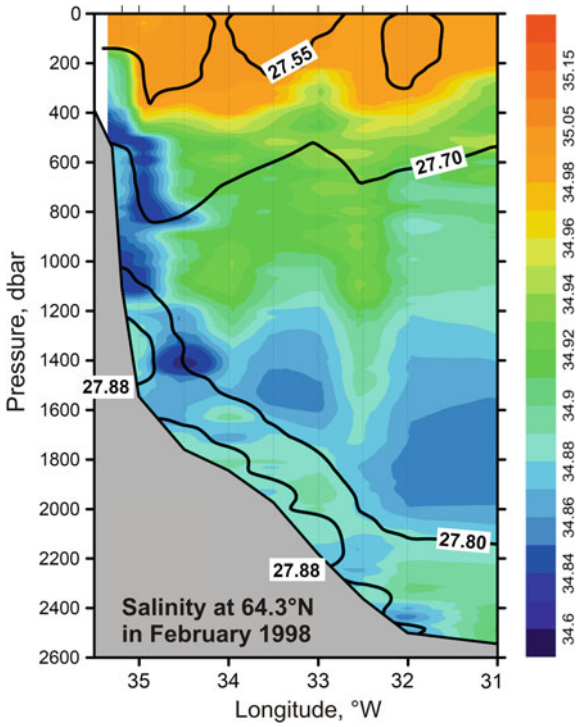


Fig. 8 Salinity observed in the northwestern Irminger Sea at 64.3° N in February 1998. The σ_0 isopycnals 27.55, 27.70, 27.80 and 27.88 are plotted as the thick black lines; the station locations are marked with the ticks on the top axis. The plot shows fresh dense waters descending (cascading) down the continental slope of Greenland down to the LSW layer ($27.70 < \sigma_0 < 27.80$) and the layer of the Nordic Seas overflow-derived deep waters ($\sigma_0 > 27.80$). Adapted from [14]

deep waters in the WBC can be considerable. Our tentative estimate based on the data from two synoptic surveys at ~59.5° N suggests that the transient contribution of a cascading event in the northern Irminger Sea to the DWBC transport at Cape Farewell can be as large as ~25%.

Acknowledgements This work was performed in the framework of the state assignment of FASO Russia (theme No. 0149-2018-0003), and supported in part by the Russian Science Foundation, grant 14-50-00095.

References

1. Van Aken, H. M. (2007). *The oceanic thermohaline circulation: An introduction* (p. 326). New York:Springer. ISBN 978-0-387-36637-1.
2. Dickson, R. R., & Brown, J. (1994). The production of North Atlantic deep water: Sources, rates and pathways. *Journal Geophysical Research, 99*(C6), 12319–12341.
3. Koltermann, K. P., Sokov, A., Tereschenkov, V., Dobroliubov, S., Lorbacher, K., & Sy, A. (1999). Decadal changes in the thermohaline circulation of the North Atlantic. *Deep Sea Research Part II: Topical Studies in Oceanography, 46,* 109–138. https://doi.org/10.1016/S0967-0645(98)00115-5.
4. Lherminier, P., Mercier, H., Huck, T., Gourcuff, C., Perez, F. F., Morin, P., et al. (2010). The Atlantic Meridional Overturning Circulation and the subpolar gyre observed at the A25–Ovide section in June 2002 and 2004. *Deep Sea Research Part I, 57,* 1374–1391. https://doi.org/10.1016/j.dsr.2010.07.009.
5. Schmitz, W. J., Jr., & McCartney, M. S. (1993). On the North Atlantic Circulation. *Reviews of Geophysics, 31,* 29–49.
6. Schott, F. A., & Brandt, P. (2007). Circulation and deep water export of the subpolar North Atlantic during the 1990s. In A. Schmittner, J. Chiang, & S. Hemmings (Eds.), *Ocean circulation: Mechanisms and impacts* (Vol. 173, pp. 91–118). Geophysical Monograph Series. https://doi.org/10.1029/173gm08.
7. Sutherland, D. A., & Pickart, R. S. (2008). The East Greenland Coastal Current: Structure, variability, and forcing. *Progress in Oceanography, 78,* 58–77. https://doi.org/10.1016/j.pocean.2007.09.006.
8. Sarafanov, A., Falina, A., Mercier, H., Sokov, A., Lherminier, P., Gourcuff, C., Gladyshev, S., Gaillard, F., & Daniault N. (2012). Mean full-depth summer circulation and transports at the northern periphery of the Atlantic Ocean in the 2000s. *Journal of Geophysical Research*, *117*(C01014). https://doi.org/10.1029/2011JC007572.
9. Dickson, R., Yashayaev, I., Meincke, J., Turrell, B., Dye, S., & Holfort, J. (2002). Rapid freshening of the deep North Atlantic Ocean over the past four decades. *Nature, 416,* 832–837.
10. Curry, R., Dickson, R., & Yashayaev, I. (2003). A change in the freshwater balance of the Atlantic Ocean over the past four decades. *Nature, 426,* 826–829.
11. Hansen, B., Osterhus, S., Quadfasel, D., & Turrell, W. (2004). Already the day after tomorrow? *Science, 305,* 953–954.
12. Bryden, H. L., Longworth, H. R., & Cunningham, S. A. (2005). Slowing of the Atlantic meridional overturning circulation at 25°N. *Nature*, *438*, 655–657. https://doi.org/10.1038/nature04385.
13. Falina, A., Sarafanov, A., & Sokov, A. (2007). Variability and renewal of Labrador Sea Water in the Irminger Basin in 1991–2004. *Journal Geophysical Research, 112,* C01006. https://doi.org/10.1029/2005JC003348.

14. Falina, A., Sarafanov, A., Mercier, H., Lherminier, P., Sokov, A., & Daniault, N. (2012). On the cascading of dense shelf waters in the Irminger Sea. *Journal of Physical Oceanography, 42*, 2254–2267. http://dx.doi.org/10.1175/JPO-D-12-012.1.
15. Sarafanov, A., Sokov, A., Demidov, A., & Falina, A. (2007). Warming and salinification of intermediate and deep waters in the Irminger Sea and Iceland Basin in 1997–2006. *Geophysical Reseach Letters, 34,* L23609. https://doi.org/10.1029/2007GL031074.
16. Sarafanov, A., Falina, A., Sokov, A., & Demidov, A. (2008). Intense warming and salinification of intermediate waters of southern origin in the eastern subpolar North Atlantic in the 1990s to mid-2000s. *Journal Geophysical Research, 113,* C12022. https://doi.org/10.1029/2008JC004975.
17. Sarafanov, A., Mercier, H., Falina, A., Sokov, A., & Lherminier, P. (2010). Cessation and partial reversal of deep water freshening in the northern North Atlantic: Observation-based estimates and attribution. *Tellus, 62A,* 80–90. https://doi.org/10.1111/j.1600-0870.2009.00418.x.
18. Sarafanov, A. (2009). On the effect of the North Atlantic Oscillation on temperature and salinity of the subpolar North Atlantic intermediate and deep waters. *ICES Journal of Marine Science, 66*(7), 1448–1454. https://doi.org/10.1093/icesjms/fsp094.
19. Sarafanov, A., Falina, A., Mercier, H., Lherminier, P., & Sokov, A. (2009). Recent changes in the Greenland-Scotland overflow-derived water transport inferred from hydrographic observations in the southern Irminger Sea. *Geophysical Reseach Letters, 36,* L13606. https://doi.org/10.1029/2009GL038385.
20. Sarafanov, A., Falina, A., Lherminier, P., Mercier, H., Sokov, A., & Gourcuff, C. (2010). Assessing decadal changes in the deep western boundary current absolute transport southeast of Cape Farewell (Greenland) from hydrography and altimetry. *Journal Geophysical Research, 115,* C11003. https://doi.org/10.1029/2009JC005811.
21. Våge, K., Pickart, R., Sarafanov, A., Knutsen, Ø., Mercier, H., Lherminier, P., et al. (2011). The Irminger Gyre: Circulation, convection, and interannual variability. *Deep Sea Research Part I, 58,* 590–614. https://doi.org/10.1016/j.dsr.2011.03.001.
22. Pickart, R. S., Spall, M., Ribergaard, M. H., Moore, G. W. K., & Milliff, R. (2003). Deep convection in the Irminger Sea forced by the Greenland tip jet. *Nature, 424,* 152–156.
23. Yashayaev, I. (2007). Hydrographic changes in the Labrador Sea, 1960–2005. *Progress in Oceanography, 73,* 242–276.
24. Boessenkool, K. P., Hall, I. R., Elderfield, H., & Yashayaev, I. (2007). North Atlantic climate and deep-ocean flow speed changes during the last 230 years. *Geophysical Research Letters, 34,* L13614. https://doi.org/10.1029/2007GL030285.
25. Hurrell, J. W. (1995). Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science, 269,* 676–679.
26. Meehl, G. A. (2007). Global climate projections. In S. Solomon et al. (Eds.), *Climate change 2007: The physical science basis* (pp. 747–847). Cambridge University Press.
27. Bacon, S. (1998). Decadal variability in the outflow from the Nordic seas to the deep Atlantic Ocean. *Nature, 394,* 871–874.
28. Rudels, B., Eriksson, P., Grönvall, H., Hietala, R., & Launiainen, J. (1999). Hydrographic observations in Denmark Strait in fall 1997, and their implications for the entrainment into the overflow plume. *Geophysical Reseach Letters, 26,* 1325–1328.
29. Rudels, B., Fahrbach, E., Meincke, J., Budeus, G., & Eriksson, P. (2002). The East Greeenland Current and its contribution to the Denmark Strait overflow. *ICES Journal of Marine Science, 59,* 1133–1154.

Thermohaline Structure and Salt Fingering in the Lomonosov Equatorial Undercurrent as Observed in April 2017

Tatiana A. Demidova

Introduction

Current velocity measurements using the OS-75 shipboard ADCP operating at a frequency of 76.8 kHz manufactured by the RDI-Teledyne Co. (sADCP) have been carried out on the route of the R/V "Akademik Sergey Vavilov" since 2014 in the equatorial zone of the central and western Atlantic over the sections across the Lomonosov Equatorial Undercurrent (EUC). In the (boreal) spring, the ship crossed the equator from south to north, and in autumn it crossed the equator from north to south. In 2014–2017, we carried out six crossings and some of them were accompanied by CTD casts. The main task of this research was the study of the EUC velocity field and thermohaline subsurface structure. It included estimates of velocities and spatial location of the undercurrent and its fronts relative to the equator and depth, estimates of water transport, observations of thermohaline stratification and its stability, and estimates of vertical mixing. Results of studies in 2015–2016 are briefly discussed in [5, 6]. In this contribution, we consider the results of measurements in April 2017. We discuss the properties of the EUC velocity field and focus on the analysis of the thermohaline stratification structure in the subsurface layer down to the depths of 600–800 m from the data of CTD-casts at three stations with a goal of finding convectively unstable layers favorable for salt fingering vertical mixing.

Vertical mixing related to the equatorial undercurrents was considered and discussed since the discovery of the subsurface flows, for example in [16, 17, 22]. Many studies report about the role of vertical mixing owing to the turbulence and velocity shear in the undercurrents, for example in [12–14, 23, 24]. In 1973, [14] determined that while considering vertical mixing it is important to present the

T. A. Demidova (✉)
Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
e-mail: tademidova@ocean.ru

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_26

counter-current flow as a jet of salty water moving in the background environment. At the same time, the mixing occurring below its core should be considered as processes developing more or less continuously due to the turbulence and other mechanisms.

In a flow similar to the EUC with vertical density gradients and mean vertical shears, stability may be expressed as the ratio of the buoyancy effects of the density change and the vertical shear arising due to the inertial effects caused by the motion of this flow [14]. This ratio is usually expressed as the Richardson number. The effects of the contribution of the buoyancy are given in terms of the local Brunt-Väisälä frequency, which characterizes the *static* (thermohaline) stability. The *dynamic* stability is defined by the mean vertical shear. This approach to determination of stability and the mixing ability is widely used in the publications on vertical mixing. At the same time, gravitationally stable salinity-temperature stratification of the ocean water column may actually be unstable due to the double diffusion convection that occurs owing to the fact that the molecular diffusivity of heat is much greater than the diffusivity of salt [10, 11, 29, 31].

We keep the latter in mind and emphasize that the vertical thermohaline field across the Lomonosov undercurrent is characterized by extremely high gradients of salinity. They arise owing to the existence of the high-gradient upper halocline with salinity in the core over 36.7 psu, while under the halocline (and above it) there is water of relatively low salinity: 34.5–35.5 psu. The occurrence of inversion and steps of different scales confirms the assumption of the unstable character of water under the upper sharp halocline-thermocline. Thus, owing to continuous delivery of high salinity water by the flow of EUC, the especially favorable conditions for development of the convective instability and processes of convective mixing in the form of double diffusion may exist.

As the convective mixing is caused by the local decrease in density with depth, it propagates vertically and develops regardless the layers move or not. Therefore, realizing inevitable influence of the turbulence and shear effects on the vertical mixing in the case of a rapid flow, as well as the influence of internal waves, we allocate and address a question of the convective instability leading to the vertical convective diffusion in the form of salt fingers.

Some aspects of convective mixing in the equatorial undercurrents were addressed in some early publications (for example, [24]). However, to the author's knowledge, the first comprehensive study of the equatorial convective mixing in the Atlantic EUC started in [4]. The authors used the data from the western equatorial Atlantic in 2010 and considered the mixing process along the countercurrent flow axis from its sources at 43° W in the west up to 32° W in the east. They have shown evidence that along this distance the double-diffusive mixing in the form of salt fingers took place; and its intensity varied with longitude. They emphasize the importance of "the combination of horizontal advection and vertical instability" in intense "sequence of downward-convective events" probably occurring there.

In this research we follow [4] and contribute to the deep studying of fine thermohaline structures and vertical convective mixing processes in the thermocline equatorial undercurrent area on the basis of our shipboard sADCP and CTD-cast

measurements across the EUC. We focus on the detailed discussion of vertical thermohaline stratification of the section and on detecting favorable conditions for the development of double diffusion and salt fingering. In order to reveal such conditions we calculated the thermohaline stability parameters of the subsurface water. The results provided us with a reliable pattern of spatial distribution of convectively unstable intervals that were closely related in depth to thin high-gradient parts of the stepwise structures of the fields. Such unstable layers that are capable to maintain and promote vertical mixing, redistributing salt and heat downward from the EUC core, were located within the thermocline layer, which resists mixing.

The contribution is organized as follows. In section "Data and Methods", we discuss the data used for the estimates and analysis. Then, in section "Results and Discussion", we consider the results of sADCP describing the circulation features of the EUC and the results of CTD casts discussing the horizontal and vertical structure of the thermohaline field across the EUC as well as relation between thermohaline and velocity fields. In section "Conclusions", we consider briefly the criteria of vertical stability stratification types and then we discuss the computed profiles of the density ratio R_ρ and vertical stability parameter $E_{\theta,S}$, defined from the CTD casts. In the end, some important conclusions of the study are given.

Data and Methods

In this study, we analyze the direct high-resolution velocity current measurements and CTD casts on April 14, 2017 in the region of the Lomonosov EUC. The CTD stations were located in the core of the Lomonosov EUC along 33° W at latitudes 0° 30′ S (station 2698), 0° (equatorial station 2699), and 0° 30′ N (station 2700).

The data of the shipboard sADCP were processed using the software package of the manufacturing company. The velocities that did not satisfy the reliability criteria were discarded. The declared accuracy of velocity measurements in the high-resolution mode was 2–3 cm/s. The measurements were carried out at a speed of the ship of 6–10 knots. The depth range of profiling was from 22 m to 600–800 m with a depth resolution of 10 m. We focused special attention on the accuracy of the data and methods of data acquisition. In this study, 15 min averaged velocities were used.

The sADCP measurements were combined with the CTD/LADCP casts to a depth of 1500 m using the SBE 19 plus profiler and lowered Sentinel 300 kHz LADCP. The accuracy of the temperature sensor was 0.002 °C while the salinity was measured with an accuracy to 0.003 psu. Unfortunately, due to the conditions of measurements and data processing, the upper boundary of the reliable LADCP data was at a depth of ∼100 m. This restriction is unacceptable for the analysis of the velocity measurements. Therefore, in this work we used only the ship-mounted sADCP to plot the profiles and velocity sections up to the depths of 600–800 m.

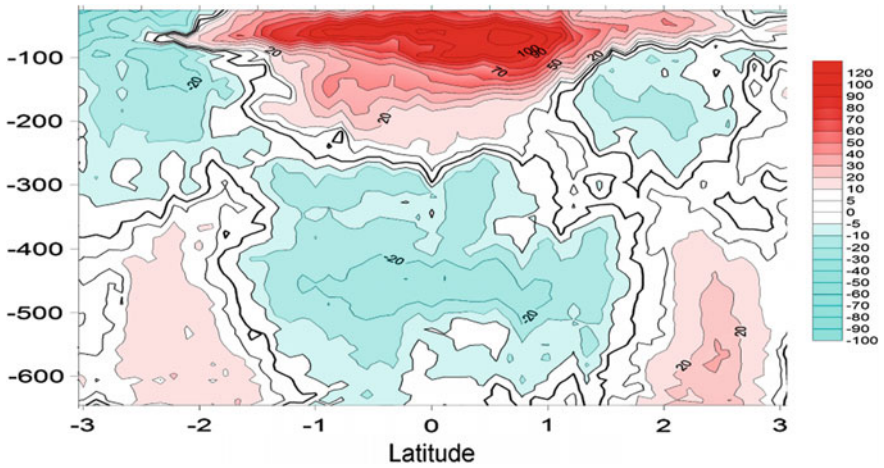


Fig. 1 The Lomonosov EUC on April 14, 2017. Section of zonal velocity across the equator at 33° W based on the shipboard sADCP data. The pink tones are related to the eastward velocities and the blue tones are related to the westward velocities

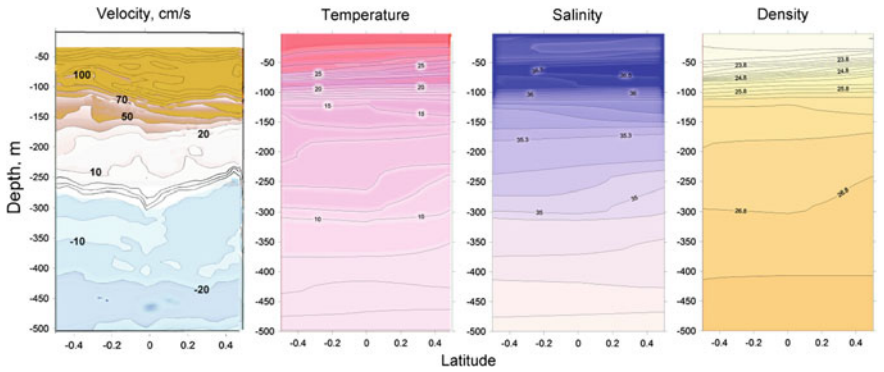


Fig. 2 Sections of the zonal velocity, temperature, salinity, and potential density anomaly down to a depth of 500 m over the equatorial section on April 14, 2017. The zonal velocity was averaged over the entire run time of the casts. These sections illustrate vertical inhomogeneity of the thermohaline structure and velocity field

Results and Discussion

The Equatorial Subsurface Circulation

On April 14, 2017, the EUC was located between 2.2° S and 2.4° N. It is likely that at the equator, the subsurface current rose to the surface in a band limited at both sides by 1.5° latitudes (Fig. 1). High eastward velocities reaching 80 cm/s at a level

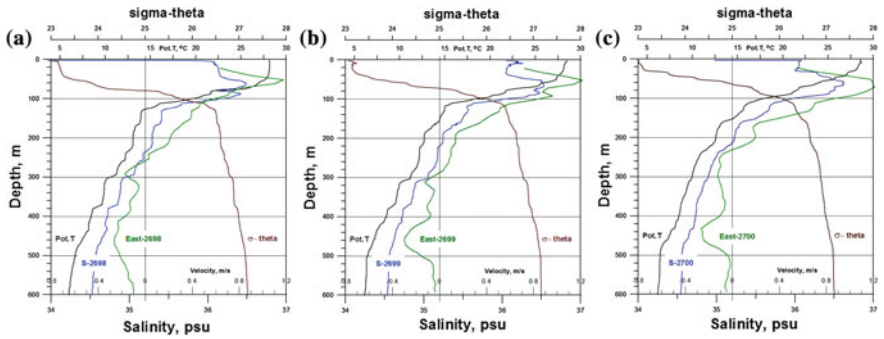


Fig. 3 Vertical thermohaline structure. Salinity, potential temperature, potential density anomaly, and mean zonal velocity component (sADCP) on April 14, 2017. Station 2698 (**a**), 0.5° S, Station 2699, 0° (**b**), Station 2700, 0.5° N (**c**)

of 22 m indicate that this current occupied the entire upper layer up to the surface. The maximum eastward speeds exceeding 120 cm/s were found at depths of 62–82 m at a distance of 30–40 miles north of the equator between latitudes 0.51–0.86° N (the maximum velocity equal to 127 cm/s was measured at a depth of 62 m). The velocities over 50–80 cm/s within boundaries of the core were found to a depth of 150 m, while less intense eastward velocities reached a depth of 250 m. Below this depth, the eastward flow changed to the westward one, which is the Equatorial Intermediate Current (EIC).

High velocities in the core of the current are characteristic of the EUC in the spring period due to its seasonal variations [2, 21]; however, the velocities recorded in our expedition are among the highest reported in the publications. Such velocities were sometimes reported based on the data of moorings, for example, by Kolesnikov et al. [18] as well as by Katz [15] and [30] or during the PIRATA experiment by [1]. Short time averaging (15 min) used in our study compared with the most publications known to the author could cause a slight increase in the velocity values.

Our measurements indicate that the meridional transport was poorly developed. The general background on the section (not shown here) was formed by the northward flows, while the southward flows were seen as small patches. The velocities of the meridional flows of both directions seldom exceeded 20 cm/s. The weak influence of the meridional component on the resulting direction of the current in the EUC core of 150 m thick is reflected in almost strictly eastward direction of the EUC.

The westward EIC is developed immediately under the Lomonosov Current below the depths of 250–300 m. This current is limited from below by the maximum depths of our observations (650–700 m). It divides the eastward Northern and Southern Intermediate Counter-Currents (NICC and SICC) located almost symmetrically relative to the equator at latitudes of 2–3° N and 1.5–2.5° S, respectively, with the cores at a depth of about 500 m. The northern branch, the NICC, is more

intense: the velocities in the core were up to 40 cm/s, while the velocities in the Southern branch of the intermediate counter-current (SICC) did not reach 20 cm/s.

Analysis of CTD-Casts and sADCP Profiles

Vertical distribution of thermohaline parameters is based on the data of CTD-casts at points 0° 30′ S, 0°, and 0° 30′ N (stations 2698–2700). The velocity profiles were measured by the sADCP profiling. Sections of temperature, salinity, density, and zonal velocity are shown in Fig. 2 for better visualization and interpretation of these fields. Profiles of the thermohaline parameters and profiles of the zonal velocity averaged over the entire run time of the casts are shown in Fig. 3.

The sections in Fig. 2 show condensation of contour lines of potential temperature and potential density anomaly (hereinafter, density for short) in the upper part of the section (below the uppermost mixed layer) down to a depth of 120–140 m and their sparseness below that level. Condensation of the isohalines is observed at depths from 40 to 60 m and from 90 to 130 m. A local sparseness between the two intervals is related to the salinity (and velocity) core. The condensation of isotherms and isopycnals above 120–140 m reflects the development of a sharp, high-gradient upper part of the thermo-pycnocline at all stations. The sharp upper halocline occurs lower than this layer below the salinity maximum at a depth of 60–70 m.

Sparseness of all isolines is especially clearly pronounced in the central part of the main thermocline at depths of 200–500 m. It demonstrates the existence of moderate gradients of temperature, salinity, and density in this range of depths. Over some intervals, the gradients are weaker than the background gradients. One can see an example of almost homogeneous layers of temperature, salinity, and potential density in the interval 310–365 m at station 2698 south of the equator (Fig. 3a). Similar decrease in the density gradient weaker than the background values may reflect the development of intense processes of vertical transport.

One can also see in Fig. 2 the formation of elevations and depressions in the isopycnal structure. The most pronounced features are isopycnal depressions in the northern part of the section at depths above 150 m in the north and at a depth of ∼300 m at the equator. The upper depression is associated with the corresponding decrease in depths of isotherms and isohalines and also with the lower boundary of the undercurrent core. The isopycnal depression at a depth of ∼300 m corresponds to the depressions and elevations in the structure of isotherms and isohalines. The existence of the depressions of isotherms and isohalines between 150 and 300 m leads to the local warming and salinity increase in the range of these depths as well as below them. Note that the boundary between the eastward EUC flow and the westward SIC flow is exactly in the depth interval of the lower isopycnal depression. The maximum isopycnal depression at the equator corresponds to the similar but sharper depression of the isotachs (contour lines of equal velocity).

Strong Lomonosov undercurrent transports water of high salinity in the subsurface layers of the equatorial zone. It forms a subsurface maximum of salinity

whose depth position coincides, in general, with the maximum velocities in the core of the current. This is well seen over the sections and on the profiles in Fig. 3. Profiles of velocity and salinity are relatively well correlated.

Intervals of high gradient velocity below the velocity maximum are located in the upper part of the main thermocline. The structure of the velocity profiles here is similar to the corresponding upper sharp thermo-halocline.

According to the CTD data, high salinity of surface water exceeding 36.0–36.2 psu was observed at the equator (Fig. 3). These high salinities in the surface layer were caused by the clearly seen shallow location of the EUC core and its exposure to the surface. A high temperature mixed layer (~28 °C) was observed near the surface within the first 40 m at all CTD stations. It was almost homogeneous south of the equator and less mixed at the equator and north of it.

In the same depth interval, a quasi-homogeneous layer with minimum salinity was formed by a surface intrusion of fresher water, probably of the Amazonian origin [28]. At the equatorial station 2699, this surface intrusion is characterized by the salinity minimum at a depth of 40 m within the homogeneous mixed layer. It is accompanied only with weak distortions of the temperature layer structure. Its traces were seen at stations 2698 and 2700.

The upper high-gradient thermocline is located immediately below the small transition zone under the quasi-homogeneous temperature layer (starting from 70 m at station 2698, from 60 m at station 2699, and from 40 m at station 2700). It is extended down to a depth of 140 m south of the equator and to a depth of 130 m at the equator and north of it.

As seen in Fig. 3, the beginning of a high-gradient thermocline approximately corresponds to the maxima of salinity and velocity at the equator. The beginning of the sharp thermocline is displaced downwards approximately by 20 m from the maximum of speed and salinity depth at the southern station and upwards from these maxima by 20 m at the northern station.

Salinity increases down from a depth of 40 m to the maximum values at a depth of 63–66 m that is approximately the same as the depth of the velocity maximum. At the southern station, the depth interval of high salinity values just below its maximum is complicated by a sharp local minimum 10–15 m thick at a depth of ~76 m (Fig. 3). The traces of it are seen as intrusions at the equator and at the northern station. This local decrease in salinity is probably caused by the local reduction of speed in the current core. The temperature field is weakly influenced by these structural inhomogeneities of the velocity and salinity fields: only small steps in the sharp-gradient thermocline are noted at each of the three stations.

Owing to this local complication, the upper high-gradient halocline at the three stations begins deeper than the levels of the salinity and speed maxima and much deeper than the upper level of the high-gradient thermocline: approximately at a depth of 95 m.

Vertical Inhomogeneities of Profiles

Several inversions of temperature and salinity and intrusion inhomogeneities (due to the isopycnal advection) were recorded in the upper part of the section. Stepwise structures (due to the vertical structure forming processes) of various scales from a few to several tens of meters are the characteristic features of the vertical structure of the thermohaline field below the upper high-gradient thermo-halocline and down to approximately 500 m, which is the transition region to the intermediate salinity minimum.

The steps consist of thin high-gradient layers ("sheets") separated by thick low-gradient layers. The vertical gradients in high-gradient layers exceed 10–100 times and greater the corresponding mean gradients. Such fine structures are clearly pronounced on the vertical profiles at the southern station 2698. These steps are clearly seen both in the temperature and salinity fields, but owing to the fact that they compensate each other they are less manifested in the density field (Fig. 3). From south to north, the number of steps decreases and the vertical size (the thickness of well-mixed layers) increases while the salinity (and temperature) gradients decrease.

Similar structures are also seen on the vertical profiles of velocity but here they are smoother. The latter can be related to a lower depth resolution of the velocity profile compared with the temperature and salinity profiles (the vertical steps of the profiles are 10 and 2 m, respectively). Well-correlated vertical structure of both thermohaline parameters and velocity is typical for the sections across the EUC, which was already noticed in the early publications.

The existence of intense stepwise structures on the profiles provides the first preliminary evidence about intense processes of mixing in the stratified water column. Below we will address a question of the reality of the significant contribution of convective mixing below the EUC core.

Thermohaline Stratification and Vertical Stability

In this paragraph, we consider the criteria of the types of stratification and vertical stability and then present and discuss the computed profiles of the density ratio R_ρ and vertical stability parameter $E_{\theta,S}$, computed from the CTD casts. We focus here on the joint analysis of the thermohaline field structure and the computed parameters with the purpose of detecting conditions, favorable for the development of double diffusion and salt fingering.

Depending on the background stratification, the effects of double-diffusion instability are manifested in the modes of salt fingers and interleaving (or diffusion) convections. Salt fingers occur in steadily stratified layers when both temperature and salinity decrease with depth. When **T** and **S**, on the contrary, increase downwards, the convection in the mode of interleaving occurs [7, 8, 20].

Under the conditions when the stabilizing contribution to a density gradient is provided by the gradients of temperature and salinity, any disturbances decay over time owing to the stable state. The other situation is observed when one of the components (temperature or salinity) has the vertical gradient that destabilizes the density gradient. In such a case, the potential energy of the unstable component, which is released owing to the double diffusion, leads to continuing intensification of any (even initially small) disturbances of the thermohaline fields over time. Those disturbances manifest themselves in the form of *thin and fine structures* of the fields.

Intense development of stepwise structures in the thermocline below the core borders of the EUC, revealed from our measurements, gave us additional grounds for deeper consideration of the question about the reality of the double diffusion convective mixing during the time period of our experiments. The detection and identification of intervals of probable double diffusion in the form of salt fingers was provided with the use of diagnosis of the types of stratification and thermohaline instability. The vertical thermohaline stability parameter and the density ratio were computed from the CTD casts, and the relationship of both parameters were considered.

Double diffusion condition criteria. It was found from numerous experiments, that the intensity of formation of the elements of temperature and salinity fine structure is determined by the background of the vertical distributions of these parameters. This fact became the basis for parameterization of the fine structure of the thermohaline fields. Following [3, 9] we will consider, as the qualitative criteria, the interrelation between the density ratio R_ρ and vertical stability parameter $E_{\theta,S}$ taking into account the background gradients and their signs.

In the upper layer of the ocean (neglecting adiabatic effects), the density ratio following from the linearized equation of state is written as:

$$R_\rho = \alpha \frac{d\theta}{dz} / \beta \frac{dS}{dz},$$

where

$\frac{d\theta}{dz}$ and $\frac{dS}{dz}$ are the gradients of potential temperature and salinity, respectively, and $\alpha = -\frac{1}{\rho}\frac{d\rho}{dT}$ and $\beta = \frac{1}{\rho}\frac{d\rho}{dS}$ are thermal expansion and saline contraction coefficients, respectively, ρ is the density and z is the vertical coordinate (pressure or depth).

Thus, density ratio R_ρ characterizes the relative contribution of changes in temperature and salinity to the density gradient and to the vertical stratification stability. In other words, it defines the relation of the stabilizing and destabilizing factors of stratification and serves as a useful indicator of the relative intensity of double diffusion processes.

Under the same conditions of the upper layer, parameter $E_{\theta,S}$ is written as

$$E_{\theta,S} \equiv -\frac{1}{\rho}\frac{d\rho}{dz} \equiv \frac{1}{g} \quad \mathrm{N}^2 = \alpha\frac{dT}{dz} + \beta\frac{dS}{dz} = (-\eta + \varsigma),$$

where N is the Brunt-Väisälä frequency. It represents the vertical thermohaline stability for adiabatic conditions needed for the qualitative assessment of the stability of seawater stratification. Along with the buoyancy frequency, $E_{\theta,S}$ is a measure of the tendency of a water particle to move vertically relative to its surroundings. There is a correspondingly more complicated expression for the stability when the adiabatic effects are taken into account, which is usually necessary at great depths.

The first term of the stability parameter $E_T = -\eta = \alpha\frac{dT}{dz}$ expresses the temperature contribution to stability (temperature stability), and the other term $E_S = \xi = \beta\frac{dS}{dz}$ shows the salinity contribution (salinity stability). Such a division allows one to reveal the relative influence of salinity and temperature on the formation of the total thermohaline stability $E_{\theta,S}$. Here we can note that the density ratio may be considered as the ratio of the temperature component of vertical thermohaline stability to the salinity stability component.

According to the definitions given above, the type of stability of the background stratification in the real ocean environment is determined by the sign and value of parameter $E_{\theta,S}$:

$E_{\theta,S} = (-\eta + \xi) > \mathbf{0}$ is related to the positive, or hydrostatically stable stratification (stable equilibrium of water);

$E_{\theta,S} = (-\eta + \xi) < \mathbf{0}$ is related to the negative, or hydrostatically unstable stratification (unstable equilibrium);

$E_{\theta,S} = (-\eta + \xi) = \mathbf{0}$ is related to the neutral, adiabatic equilibrium stratification.

The range of variation of parameters of stability $E_{\theta,S}$ and density ratio R_ρ is generally very wide:

$$-\infty < E_{\theta,S}, E_T, E_S, R_\rho < \infty.$$

At the same time it is shown, for example by Fedorov [8] that formation of the fine structure due to the effects of double diffusion convection is possible only at positive background density ratio and stability, i.e. $R_\rho > \mathbf{0}$ and $E_{\theta,S} = (E_T + E_S) > \mathbf{0}$, although separately both E_T, and E_S may be negative. At $R_\rho < \mathbf{0}$ and/or $E_{\theta,S} < \mathbf{0}$ the influence of the effects of double diffusion convection is excluded.

The question of the values of the density ratio R_ρ necessary for the development of salt fingers was investigated in many laboratory and numerical experiments, for example, in already mentioned above [8, 19, 26]. There are some observations of the fingers at great numbers of R_ρ, up to $R_\rho = 10$ [26]. However the limitations $\mathbf{1} < R_\rho < \mathbf{2}$ are now accepted as the condition of their formation with an assumption that most efficiently the fingers develop and transfer mass against the density

gradient at $\mathbf{1} < R_\rho < \mathbf{1.5}$. The reason for the upper limit $(R_\rho < 2)$ follows from the conclusions of [27] that the finger growth rate depends on the local buoyancy period and should be shorter than this period, which happens only when R_ρ is smaller than 2. The intensity and speed of their formation increase as R_ρ decreases. Recently, it was found [25], that at $R_\rho \to 1$, the behavior of the salt finger system becomes transient and the growth rate of salt fingers becomes very high.

The criteria for evaluation of the type of the stratification as well as of the fine structure formation occurring in the stratified water medium are based on the interrelations of the vertical stability of stratification and the density ratio [3, 9]. We used these criteria, and added to them, when appropriate, the limiting restrictions on R_ρ for salt fingering development.

1. *Complete, or absolute, stability*: $\Delta\mathbf{T} < \mathbf{0}$, $\Delta\mathbf{S} > \mathbf{0}$, $E_{\theta,S} > 0$, $(E_T > \mathbf{0}, E_S > \mathbf{0}), R_\rho < \mathbf{0}$. Stratification is convectively stable. The effects of double-diffusion convection are excluded.
2. *Convectively unstable stratification (in the form of salt fingers)*: $\Delta\mathbf{T} < \mathbf{0}$, $\Delta\mathbf{S} < \mathbf{0}$, $E_{\theta,S} > 0$, $(E_T > \mathbf{0}, E_S < \mathbf{0}), R_\rho > \mathbf{1}$.
   At $R_\rho > \mathbf{1}$ stratification is convectively unstable due to double diffusion.
   At $\mathbf{0} < R_\rho < \mathbf{1}$ the stratification is hydrostatically unstable.
3. *Convectively unstable stratification (in the form of interleaving convection)*:
   $\Delta\mathbf{T} > \mathbf{0}$, $\Delta\mathbf{S} > \mathbf{0}$,
   $E_{\theta,S} > 0$, $(E_T < \mathbf{0}, E_S > \mathbf{0}), R_\rho > \mathbf{1}$.
   At $R_\rho > \mathbf{1}$ stratification is convectively unstable due to double diffusion.
   At $\mathbf{0} < R_\rho < \mathbf{1}$ the stratification is hydrostatically unstable.
4. *Absolute instability*: $\Delta\mathbf{T} > \mathbf{0}$, $\Delta\mathbf{S} < \mathbf{0}$, $E_{\theta,S} < 0$, $(E_T < \mathbf{0}, E_S < \mathbf{0}), R_\rho < \mathbf{0}$.

With the use of the limitative restrictions for salt fingering development, the conditions in the brackets are redundant. With our goal of showing the reality of salt fingering convection, evaluation of intensity and types of the vertical thermohaline structure included an assessment of the character of $\Delta\mathbf{T}$ and $\Delta\mathbf{S}$ background profiles and separate intervals, estimates of parameters of stability $E_{\theta,S}$ and R_ρ and their relationship. In the vertical interval satisfying the second criteria, the restrictive conditions for intense development of salt fingers $\mathbf{1} < R_\rho < \mathbf{2}$ were applied. Such approach was used for detecting and identifying vertical intervals of favorable conditions for salt fingering development. Computed profiles of the density ratio and of the vertical stability together with thermohaline profiles in the pressure field are considered below.

Vertical Distribution of the Stability Parameters

The stability parameters based on the CTD data were first computed using vertical averaging of 1, 3, 5, 10, and 20 dbar. It turned out that a step of 3 db provided the relatively high vertical resolution at satisfactory stability of the profiles. It provides

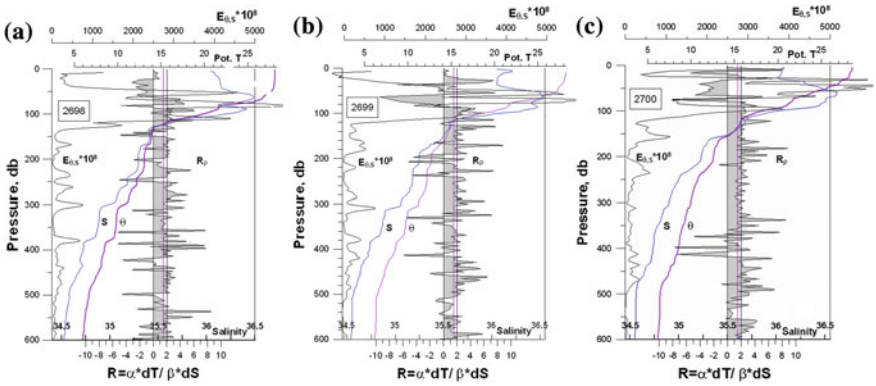


Fig. 4 Profiles of stability $E_{\theta,S}$, density ratio R_ρ, θ, and S at station 2698 (**a**), station 2699 (**b**), station 2700 (**c**). Vertical averaging is 3 m. The gray color highlights positive and negative values of R_ρ

a possibility to consider stepwise structures of the same vertical scale and larger, and the profiles with such averaging have been further analyzed.

Profiles of thermohaline stability $E_{\theta,S}$ and R_ρ together with the profiles of potential temperature θ and salinity **S** are shown in Fig. 4.

We can locate four distinct intervals at each station on the basis of the vertical profiles of the thermohaline stability parameter $E_{\theta,S}$, (Fig. 4).

(1) The uppermost interval 20–40 m thick includes the surface quasi-homogeneous mixed layer with local surface low-salinity intrusion. It corresponds to the minimal vertical stability whose values depend on the degree of mixing. In the mixed layer at stations 2698 and 2700, the minimum stability is almost the same: $\sim 270 \times 10^{-8}$ m^{-1}. At the equatorial station 2699, the thermohaline stability is negative: -325×10^{-8} m^{-1} at the vertical averaging equal to 5 db (and -590×10^{-8} m^{-1} if the step is 1 db). This corresponds to the upper mixed layer resulting from an intrusion of low saline water mentioned above.

(2) In the interval of depths of the upper pycnocline and in the thermocline that started from the transient zone below the mixed layer down to 100–110 m, high numbers of the stability are within the interval of the salinity and velocity core. The maximum values of stability reach 5000×10^{-8} m^{-1} and greater within the interval of positive salinity gradients. A complicated distribution of $E_{\theta,S}$ with the maxima alternating relative minima is bound to the intrusion complications of the salinity core.

(3) The third interval includes the main pycnocline below the EUC salinity and velocity core. The maxima and minima of stability alternate from low negative and first positive units (10^{-8} m^{-1}) to 500×10^{-8} m^{-1}. These "oscillations" of stability exactly correspond (Fig. 5) to the well developed stepwise fine structure of salinity and temperature profiles, which is also reflected on the

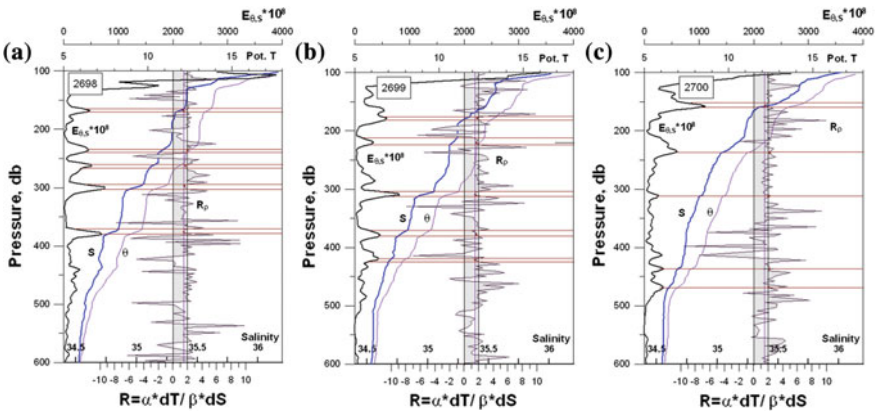


Fig. 5 Intervals of manifestation of possible double diffusion convection in the form of salt fingers on the profiles of R_ρ related to high-gradient steps on the T and S profiles and peaks of $E_{\theta,S}$. Intervals of possible fingering are limited with brown lines. Brown dots denote the minimum $R_\rho =$ within each selected interval. (The profiles are limited with depths below the salinity maximum. Vertical averaging is 3 db. **a** Station 2698, **b** station 2699, **c** station 2700

profile of density (Fig. 3). The minima are within the well-mixed portions of steps and maxima are related to the high-gradient portions.

(4) The fourth, lowest interval with the minimum mean values of stability is the transition region to the intermediate water salinity minimum. It is characterized by the lowest gradients of thermohaline parameters. The mean values of $E_{\theta,S}$ do not exceed the first tens of units, while in the range of depths of 550–700 m the minimum values equal to zero or even minimal negative values are observed. This provides evidence about the neutral and weakly negative stability in some depth intervals in the transition region to the intermediate salinity minimum.

Combining these observations with vertical distribution of the density ratio R_ρ, and taking into account the T and S gradients (Figs. 4 and 5), we can now define the types of stratifications and select the possible double diffusion convective intervals applying the criteria considered above. The main fact is that clear stepwise structures occur only in the main thermocline at the interval of "oscillating" $E_{\theta,S}$ where $R_\rho \leq \mathbf{2}$ or slightly higher than 2. This provides us with the initial information about the intervals, in which salt fingering is most possible. Therefore, we start with a brief discussion of some important phenomena within the first two intervals including the core of the EUC. We consider in more detail the depth interval of the main thermo-halocline below the undercurrent core borders (deeper than 120 m).

As seen from Fig. 4, in the uppermost interval, negative thermohaline stability at 9 m at the equatorial station 2699 related to the low negative density ratio ($E_{\theta,S} = -347\,\mathrm{m}^{-1}$ and $R_\rho = -1.35$), which appears due to the decrease in salinity, is characteristic of the absolute instability ($\Delta\mathbf{T} > \mathbf{0}, \Delta\mathbf{S} < \mathbf{0}, E_{\theta,S} < 0, R_\rho < \mathbf{0}$). Effects of double-diffusion convection are excluded.

In the two upper intervals down to the bottom boundary of the sharp upper thermo-halocline, there are two (stations 2698 and 2699) to three (station 2700) intervals with the negative density ratio related to the positive stability parameter as well as to positive ΔS and negative ΔT gradients. At the equator (station 2699) there is a distinct interval with $R_\rho < \mathbf{0}$ at the depths of the salinity maximum (60–85 m) and at the northern station 2700, at depths from ~40 to 65 m. The corresponding positive values of $E_{\theta,S}$ are extremely high here. Such combination of parameters $\mathbf{\Delta T < 0}$, $\mathbf{\Delta S > 0}$, $E_{\theta,S} > 0, R_\rho < \mathbf{0}$ determines absolute, convectively stable stratification. The effects of double-diffusion convection are excluded. The same conclusion is related to station 2698, where the effect is less pronounced.

There are several thin intervals, where R_ρ is positive and $E_{\theta,S} > 0$, and $\mathbf{\Delta T < 0}$, $\mathbf{\Delta S < 0}$ (60–80 m at station 2698; 30, 55, and 65 m at station 2699; and 15, 30, 60, and 80 m at station 2700). These conditions define convectively unstable stratification, but owing to the fact that R_ρ is too high (~4–8) the probability of the development of the salt fingering double diffusion is very low within the intervals.

Now we address the depths of the main thermo-halocline below its sharp upper part (deeper than 120 m) with the development of stepwise structures and oscillating thermohaline stability $E_{\theta,S}$ (Fig. 5).

As seen from Fig. 5, the peaks of $E_{\theta,S}$ "oscillations" that exactly correspond (Fig. 5) to the thin, well-developed high-gradient portions ("sheets") of the stepwise structures over the salinity, temperature, and density profiles are also associated with the low positive values of density ratio $1 < R_\rho < 2$, while some of them are within $1 < R_\rho < 1.5$.

Pronounced examples of such thin intervals with peaks of positive stability and low density ratio, related to the "sheets" of the steps can be found at the depths shown in Table 1.

We can conclude that these minimum positive values $1 < R_\rho < 2$ and <1.5 are found against the background of stable stratification $E_{\theta,S} > \mathbf{0}$, when both temperature and salinity decrease downwards with sharp negative gradients; hence, such intervals meet the strict conditions of the development of convective instability and double diffusion convection of the salt finger type. The intervals of possible fingering are from 4 to 14 m thick. The most intense manifestation of the process occurs at the southern station 2698 and at the equator (station 2699). At the northern station 2700, the manifestation of fingering is least expressed. We show in Table 1 the depths of possible fingering at this station but only the interval 4 m thick between 155 and 159 m with $R_\rho = 1.62$ may be regarded here as reliable.

It is important to note that the approximate coincidence of the most clearly pronounced fingering intervals at depths of 300 and 375 m at stations 2698 and 2699 located 30 miles apart may show the possible great horizontal scales of such processes.

In the rest part of the main thermocline within 100–600 m, as was shown above and seen from Fig. 5, there are thick intervals of well-mixed portions of steps related to low positive to low negative values of stability $E_{\theta,S}$ (intervals close to 212 and 219 m at station 2698 are examples of slightly negative $E_{\theta,S}$). This implies

Table 1 Depth intervals of possible salt finger convection, related values of $E_{\theta,S}$ at the peaks and minimum density ratio R_ρ

Station	Depth (m)	Vertical thermohaline stability $E_{\theta,S}$ (10^{-8} m^{-1})	Min density ratio R_ρ
2698	164–172	478	1.48
2698	234–238	446	2.0
2698	260–270	500	1.65
2698	294–303	742	1.55
2698	371–379	673	1.41
2699	175–181	572	1.75
2699	210–224	331	1.11/1.64
2699	303–312	812	1.46
2699	371–378	379	1.6
2699	518–424	408	1.66
2700	155–159	1096	1.62
2700	236	580	2.0
2700	311	246	2.14
2700	435	437	2.2
2700	469	331	2.2

intervals of the neutral stability to the slight instability. The density ratio at such intervals varies from high negative to high positive numbers. The variations in R_ρ at minimal to negative $E_{\theta,S}$ and minimal T and S gradients only indicate the small local changes in the conditions of well mixed layers, and some incidental intervals with $1 < R_\rho < 2$ are not related here to the double diffusion convection as the whole situation does not meet the second criteria. Neutral to unstable background stratification excludes the development of double diffusion convection in such well-mixed depth intervals in spite of some occurrence of low positive density ratio within them.

Thus, we found thin vertical intervals with conditions favorable for the development of double diffusive convection in the form of salt fingers below the EUC core boundary. These intervals 4–14 m thick are located within the high-gradient portions of the stepwise structures of the thermohaline profiles across the Lomonosov EUC. Occurrence of such intervals is the evidence of a high probability of the salt fingering processes that has to be an effective mechanism of the redistribution of salt (and heat) from the high saline core of the EUC down to a depth of 400–500 m and obviously into the deeper layers. The manifestation of favorable conditions for salt fingering processes is strongest at the southern station 2698 and decreases from south to north over our section. At the same time, as mentioned above, there are some indications of a large horizontal latitudinal scale of the detected "salt-fingering" intervals.

The results of the analysis are very important as we found spatial distribution of thin convectively unstable layers able to induce and maintain vertical mixing and

redistribution of salt and heat down from the EUC core within the thermo-halocline, which, on the contrary, resists any mixing.

The fact of the reality of salt-fingering convection as a consequence of the propagation of the high salinity flow of EUC, which occurred in the thermocline below its core along the section at 33° W on April, 2017 agrees with the results in [4] along the EUC axis based on the data in 2010.

Conclusions

Our study showed that in the region of the Lomonosov EUC that transports water of very high salinity, the vertical structure of the thermohaline field below the boundaries of the current core (approximately between the depths of 150 and 500 m) is characterized by vertical inhomogeneities of the stepwise type. The estimates of the vertical water stratification types there showed "oscillating" numbers of thermohaline stability $E_{\theta,s}$ with alternating thick intervals of the minimal positive to negative values (neutral to unstable stratification) and thin "peak" intervals of positive stability. The high-gradient portions ("sheets") of the steps with negative T and S gradients were related to these peaks of higher stable stratification and low values of density ratio $1 < R \leq 2$. Such intervals 3–4 to 10–15 m thick meet the strict criteria of the development of double diffusion convection of the salt finger type and provide evidence of the high probability of the development of vertical convective mixing. It implies that in the course of the propagation of the high-salinity current the fingering mechanism is responsible for the redistribution of salt and heat from the core of the EUC to the deeper layers of the water column. The manifestation of salt fingering effects is strongest at the southern station and decreases from south to north of our section. Some indications of large horizontal scale of the fingering processes are found.

Thus, as a result, we obtained clear evidence of numerous thin, convectively unstable layers able to induce and maintain vertical mixing and redistribution of salt and heat downward from the EUC core, within the main thermo-halocline that resists any mixing. Our assessment of the spatial distribution of such unstable intervals relative to depth and latitude with some evidence of their large horizontal scale is also very important.

Here, we need to emphasize that convective mixing of the salt fingering type is a part of the overall vertical mixing in the area, which also includes other convective forms as well as the processes of the dynamic mechanisms of mixing, related to turbulence, internal waves, and intense vertical shear. The importance of such types of mixing at the depth intervals above and below the boundaries of the EUC cores in the Atlantic and Pacific oceans is shown in [12–14, 20, 22–24, 27], etc. At the same time, it was also found that the existence of a shear current not only does not impede the existence of salt fingers, but it is even an additional mechanism of facilitating conditions for their formation [27]. Thus, the convective instability in the form of salt-fingering double diffusion in the region of the Lomonosov EUC

may develop together with the dynamic vertical shear mixing and other dynamic mechanisms. The problem of the influence and relative contribution of these factors to the overall vertical mixing in the EUC region from our data requires further detailed research.

Acknowledgements The field works were supported by the Program of the Presidium of the Russian Academy of Sciences (project I3П). Data analysis was supported by the Russian Science Foundation (project no. 14-50-00095).

References

1. Bourlès, B., et al. (2008). The PIRATA program: History, accomplishments, and future directions. *Bulletin of the American Meteorological Society, 89,* 1111–1125.
2. Bubnov, V. A., Moroshkin, K. V., Egorikhin, V. D., & Matveeva, Z. N. (1976). Variability of currents in the equatorial Atlantic. *Okeanologiya, 16,* 408–414.
3. Carlin, L. N., Klyuykov, E. Y., Kutko, V. P. (1988). *Small-scale structure of hydrophysical fields of the upper layer of the ocean* (p. 162) Moscow: Gidrometeoizdat. (in Russian).
4. Claret, M., Rodríguez, R., & Pelegrí, J. L. (2012). Salinity intrusion and convective mixing in the Atlantic equatorial undercurrent. *Scientia Marina, 76*(S1), 117–129. https://doi.org/10.3989/scimar.03611.19B.
5. Demidova, T. A. (2017). Distribution of thermohaline parameters of the Lomonosov undercurrent as an evidence of its static instability. In *The modern methods and means of oceanological research (MSOI-2017) Conference materials,* Moscow (pp. 148–149). (in Russian).
6. Demidova, T. A., Frey, D. I. (2017). On the velocity field of the Lomonosov current as observed on the ship route. In *The modern methods and means of oceanological research (MSOI-2017) Conference materials,* Moscow (pp. 150–151). (in Russian).
7. Fedorov, K. N. (1976). *Fine thermohaline structure of ocean water* (p. 184). Leningrad: Gidrometeoizdat. (in Russian).
8. Fedorov, K. N. (1984). Conditions of stratification and convection in the form of salt fingers in the ocean. *Doklady AN SSSR (Earth Science Sections), 275*(3), 749–753.
9. Fedorov, K. N., & Pereskokov, A. I. (1986). Typification of thermohaline conditions of stratification in the World Ocean. *Meteorology and Hydrology, 12,* 71–77. (in Russian).
10. Gershuni, G. Z., & Zhukhovitsky, E. M. (1963). About a convective instability of two-component mixture in the gravitational field. *Applied Mathematics and Mechanics, 27*(2), 301–308. (in Russian).
11. Gershuni, G. Z., & Zhukhovitsky, E. M. (1972). *Convective stability of incompressible liquid* (p. 392). Moscow, Nauka. (in Russian).
12. Gregg, M. C. (1976). Temperature and salinity microstructure of the Pacific Equatorial Undercurrent. *Journal of Geophysical Research, 81*, 1180–1196.
13. Gregg, M. C., Peters, H., Wesson, J. C., Oakey, N. S., & Shay, T. J. (1985). Intensive measurements of turbulence and shear in the equatorial undercurrent. *Nature, 318,* 140–144.
14. Jones, J. H. (1973). Vertical mixing in equatorial undercurrent. *Journal of Physical Oceanography, 3,* 286–296.
15. Katz, E. J., Bruce, J. G., & Petrie, B. D. (1979). Salt and mass flux in the Atlantic equatorial undercurrent. *Deep-Sea Research GATE Supplement II to V, 26,* 137–160.
16. Knauss, J. A. (1960). Measurements of the Cromwell Current. *Deep Sea Research, 6,* 265–286.
17. Knauss, J. A. (1960). Further measurements and observations on the Cromwell Current. *Journal of Marine Research, 24,* 204–240.

18. Kolesnikov, A. G., Boguslavsky, S. G., Grigoriev, G. N. (1968). Opening, the pilot study and development of the theory of the Lomonosov current. MGI, Academy of Sci of USSR, Sevastopol. (in Russian).
19. Kunze, E. (1990). The evolution of salt fingers in inertial wave shear. *Journal of Marine Research, 48,* 471–504.
20. Kuzmina, N., Lee, J. H., & Zhurbas, V. (2004). Effects of turbulent mixing and horizontal shear on double-diffusive interleaving in the central and western equatorial Pacific. *Journal of Physical Oceanography, 34*(1), 122–141.
21. Mercier, H., Arhan, M., & Lutjeharms, J. (2003). Upper-layer circulation in the eastern equatorial and South Atlantic Ocean in January–March 1995. *Deep-Sea Research I, 50,* 863–887.
22. Metcalf, W. G., Voorhis, A. D., & Stalcup, M. C. (1962). The Atlantic equatorial undercurrent. *Journal Geophysical Research, 67,* 2499–2508.
23. Moum, J. N., Osborn, T. R., & Paulson, C. A. (1989). Mixing in the equatorial surface layer. *Journal of Geophysical Research, 94,* 2005–2021. https://doi.org/10.1029/JC094iC02p02005.
24. Pacanowsky, R. C., & Philander, S. G. H. (1981). Parametrization of vertical mixing in numerical models of tropical oceans. *Journal of Physical Oceanography, 11,* 1443–1451.
25. Rehman, F., Dhiman, M., & Singh, O. P. (2016). Effect of eigenvalue solution on the characteristics of double diffusive salt fingers. *Journal of Mechanical Science and Technology, 30*(6), 2557–2563.
26. Schmitt, R. W. (1994). Double diffusion in oceanography. *Annual Review of Fluid Mechanics, 26,* 255–285.
27. Schmitt, R. W., & Evans, D. L. (1978). An estimate of the vertical mixing due to salt fingers based on observations in the North Atlantic central water. *Journal Geophysical Research, 83* (C6), 2913–2919.
28. Silva, A. C., Bourles, B., & Araujo, M. (2009). Circulation of the thermocline salinity maximum waters off the Northern Brazil as inferred from in situ measurements and numerical results. *Annales Geophysicae, 27,* 1861–1873.
29. Stern, M. E. (1960). The salt-fountain and thermohaline convection. *Tellus, 12,* 172–175.
30. Stramma, L., & Schott, F. (1999). The mean flow field of the tropical Atlantic Ocean. *Deep-Sea Research II, 46,* 279–303.
31. Turner, J. S. (1973). *Buoyancy effects in fluids*. Cambridge University Press:Cambridge. https://doi.org/10.1017/CBO9780511608827.

Numerical Realization of Hybrid Data Assimilation Algorithm in Ensemble Experiments with the MPIESM Coupled Model

Konstantin P. Belyaev, Ingo Kirchner, Andrey A. Kuleshov and Natalia P. Tuchkova

Introduction

The results of numerical experiments dedicated to the data assimilation technology and their applications to the computations of different parameters in the World Ocean are presented in this contribution. The last decade witnesses a huge interest to the climate research both in quality and quantity, which stems from the numerical model developments, methods of data storage, control, and climate predictions. This might be also explained as consequences of the increase in the computer power, data transforming, and processing, parallel technology developments, on-line result visualization, and many other achievements of science.

The new global project such as the Tropical Oceans Global Atmosphere (TOGA, URL: http://isccp.giss.nasa.gov/projects/togacoare.html), Pilot Research Array in the Tropical Atlantic (PIRATA, URL: http://www.pmel.noaa.gov), World Ocean Circulation Experiment (WOCE, URL: https://www.nodc.noaa.gov/woce), Argo (URL: http://www.argo.ucsd.edu), Archiving, Validating and Interpolating Satellite

K. P. Belyaev (✉)
Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
e-mail: kosbel55@gmail.com

I. Kirchner
Institute of Meteorology, Free University of Berlin, Berlin, Germany
e-mail: ingo.kirchner@met.fu-berlin.de

A. A. Kuleshov
Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Moscow, Russia
e-mail: andrew_kuleshov@mail.ru

K. P. Belyaev · N. P. Tuchkova
Federal Research Center "Computer Science and Control", Russian Academy of Sciences, Moscow, Russia
e-mail: natalia_tuchkova@mail.ru

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_27

Oceanography Data (AVISO, URL: http://www.aviso.oceanobs.com), and others are underway. The data from oceanic floats and moorings are gathered and collected permanently. However, the problem of data assimilation in the ocean and seas did not only lose its actuality and scientific significance, but became even more needed and practically actual. The new technology of oil and gas extraction on the sea shelf leads to the extreme importance of 3D oceanic parameter prediction in these zones. The large platform transportation on sea routes leads to the actuality of the wind and wave forecasts in the open ocean and oceanic shore zones. The recreation areas development and their protection from dangerous external forcing as well as the protection from anthropogenic influence bring new problems to be solved by the researchers in the pollution spreading, forecasting, and decrease of the environment contamination. All these problems cannot be solved without data assimilation methods and their application in the numerical state-of-the-art models.

The development of adequate mathematical models of dynamics and environment variability is a very important and actual problem, the solution of which will be even more needed in the near future. Numerical modeling is necessary not only as a mathematical object but its conjunction with the field data would help to correct and improve the model.

In the Earth modeling history it is possible to indicate the models by Bryan [4], two-layer model by Moiseev [11], which for first time could predict the phenomenon of the Nuclear Winter. The specific works have been also developed in the Siberian Branch of the Russian Academy of Science in Novosibirsk, under the leadership of Marchuk and Sarkisyan [10]. The last years of 20th century and the beginning of 21st century are known for an increased interest to the coupled ocean-atmosphere models and in general the Earth System Models, see, for example, the works by Kirtman and Shukla [9], in which a very complex Earth dynamics, including the ocean dynamics, atmosphere dynamics, vegetation and forest moisture exchange, river flows, and other process are mathematically described and solved numerically using the modern state-of-the-art clusters, supercomputers, and networks.

In general, in the beginning of the 1970s, the principal important step was made in the mathematical modeling of the ocean, atmosphere, and their interaction, as well as on the synoptic, seasonal, and long-term variability of the natural parameters. However, the modern state-of-the art models do not guarantee the best forecasting and understanding of the complex natural processes and phenomena. The problems remain, which are not yet able to be described or predicted adequately. For example, the quasi-biannual oscillation in the North Atlantic, the dipole in the Equatorial Atlantic, Antarctic circumpolar waves, and several other events are not modeled and correctly predicted numerically. The cause of the observed Global Warming is not also quite clear. The oscillation and migration of energetic-active domains, where the ocean-atmosphere interaction processes are several times more intense than in the other regions of the World Ocean do not have any adequate explanation.

These and other problems in modeling of the climate variability processes as well as the processes of smaller scales, for example, weather prediction, have

numerous causes. They may be related to the fact that the physical processes which lie in the basis of the models are poorly known. In addition, the processes of subgrid description are not resolved correctly. The lack of the input data or their poor quality may also cause wrong prediction. However, the main cause is not adequate knowledge of the physical processes in full details. In other words, the basic equations, which are applied for modeling are not completely correct and describe the reality only partially. Therefore, the further progress in the computer power and facilities as well as in the development of the mathematical algorithms, and increase in the model resolution would not necessary lead to the better prediction and analysis of the natural phenomena. In relation to the aforementioned ideas it is necessary to highlight that the development and further progress in the observational database will need improvements in testing and validation of the models and correction of the model outputs towards the observations.

Eventually, one may come to a conclusion that, in order to improve substantially the knowledge and understanding of the climate and other geophysical processes and their predictability it is necessary to combine the mathematical models and field data. This combination makes possible improving the modeling quality and enriching and restoring the existing database.

The problem of the optimal combination of the data and models is known in geophysics as the data assimilation (DA) problem. This is a non-trivial scientific problem because the extreme solutions such as mere replacement of all the model values with the observed data without changing the other values or simply ignoring the observations, while the model values remain unchanged are obviously not optimal. Indeed, if we only replace the model values by observations, the model output field will be totally unbalanced and on next time-step it will lead to large false fluxes, which may completely destroy the model. On the contrary, leaving the model values unchanged and ignoring the observations may result in a completely wrong output, which would appear far from reality.

In meteorology and oceanography, the DA-problems have been ongoing during several decades. Originally these problems came into consideration in the end of 1960s. Apparently, one of the first results was reported by Gandin [6]. He proposed the objective analysis method to correct the model output based on the linear interpolation ideas. This idea turned out very fruitful and this method became widely used [7]. Further, the DA-methods became an important part of the numerical weather prediction. However, there is a big gap between the theoretical sophisticated approaches and their practical realizations in the numerical algorithms. Even a very well-developed theoretical idea faces serious difficulties to be implemented in practice. Due to this and other causes the DA-problems remain actual.

The DA-methods are divided into two big groups; namely the variational and dynamical-stochastic methods. The first group is referred, for example, in [15], its modern realizations are known as a 4D-var scheme. The second group of methods is mostly represented in literature as the Kalman-filter approach, for example, in [5]. There are, also, several hybrid schemes, combining both approaches [1, 12].

In this study, the original hybrid scheme of the DA-method and several versions of its application are presented and discussed. The coupled Max Plank Institute Earth System model (MPIESM), developed at the Max Plank Institute for Meteorology in Hamburg [8, 13] is considered the main tool for the implementation of the DA-scheme.

Data Assimilation Method

The DA method presented earlier in [2, 3] is briefly described here. In a certain grid domain, we consider a given mathematical model of the ocean circulation that is integrated at a specific level [0, T]. We denote by X the state vector of the ocean, which includes its potential temperature (θ), salinity (S), and sea surface height (SSH). Further, we designate by Y the vector of the observed parameters. The theory of the data assimilation operates with two vectors of the ocean parameters, namely, the ones before and after data assimilation (the background and analysis states) X_b, X_a, respectively). They are interconnected by

$$X_a = X_b + K(Y - HX_b). \tag{1}$$

In formula (1), matrix K is the Kalman gain. Matrix H defines an operator projecting the model space onto the observation space. In fact, this operator nullifies the unobserved components of the vector and performs the linear interpolation of the observed components into the observation points.

In the classical theory [5], the Kalman gain matrix is determined by the formula

$$K = BH'(HBH' + R)^{-1}, \tag{2}$$

where B, R are given matrices, symbol $'$ stands for transposition of a vector or matrix.

Formula (1) can be expanded in the case when all parameters depend on time. Let the time-discretization $t_{n+1} = t_n + \Delta t$, $n = 0, 1, \ldots$ with time-step Δt be introduced where t_{n+1} is the moment of the computation and assimilation. Therefore, all parameters X_b, X_a, K etc. are supplied by subindex n, which corresponds to the assimilation moment.

A new method to determine the gain matrix is proposed, theoretically justified, and verified by the authors [2, 3]. When the results of calculations are corrected, the method makes possible a better account for the dynamics of the model using the classical Kalman filter. The optimal gain matrix K is found according to the formula

$$K = \sigma^{-1}(\Lambda - C)(H\Lambda)'Q^{-1}, \tag{3}$$

where $\sigma^{-1} = (H\Lambda)^{'} Q^{-1} H\Lambda$, $Q = E(Y - HX_b)(Y - HX_b)^{'} + R$ is the covariance matrix of the model errors plus the covariance matrix of the instrumental errors R, symbol E stands as usual for the mathematical expectation (average with respect to ensemble), vectors Λ and C at each step of the model computations are calculated by formulae

$$\Lambda_{n+1} = \frac{X_{b,n+1} - X_{a,n}}{\Delta t}, \quad C_{n+1} = \frac{Y_{n+1} - HX_{a,n}}{\Delta t}.$$

The physical sense of parameters Λ and C is the following: Λ means the model trend per time-unit, while C defines the observed trend. Scalar σ is the normalizing factor.

Selection of Observation Parameters

The studied assimilation scheme depends on two outside-of-the model determined parameters, namely vector C and covariance matrix Q. According to their definition vector C is set at each grid point and at each time step for all considered model variables. Its mathematical and physical sense is clear; this vector defines the observational trend, the value which becomes a benchmark under model trend approximation. In order to define this value one can use several schemes. For the directly observed variables it is possible to choose the time-difference of those variables per time interval previously interpolated onto the model grid. For unobserved model variables, such as model velocities, heat content, etc., which however, participate in the model assimilation procedure, it is possible to calculate these values through ensemble experiments, following the Monte Carlo scheme. The similar scheme is applied in the Kalman-filter algorithm to obtain background covariance matrix B in formula (2). Finally, vector C may be defined as a time-average of all model outputs over time-interval $[0, t_n]$ at any current time moment t_n. Certainly, this time moment should be far enough from the model start, albeit its precise value is not entirely clear. It is possible to prove that under some reasonable conditions the time-average will coincide with the ensemble average (weak ergodic theorem). However, the detailed consideration of the problem how to optimally define vector C is out of the scope of this study.

Different methods can be applied to obtain covariance matrix Q. As it follows from this relation, matrix Q is constructed from both observations and model output at moment t_n. This matrix is calculated in the observational phase-space, i.e. at observational points. For the given time-moment there may be insufficient amount of observations to define this matrix reliably. However, the existing databases, which collect data over many years, normally contain a sufficient amount of data to make all estimations with the necessary accuracy. Therefore, to determine this matrix it is sufficient to select the needed sample of observations, interpolate them at the considered points, and calculate the matrix according to its formula. Another

scheme to calculate this matrix is the Monte-Carlo method, which is widely used in the Kalman-filter theory [5]. The issue to calculate this matrix optimally is also beyond the consideration of the current work.

Numerical Experiments and their Results

The numerical experiments with the MPIESM model in conjunction with the DA scheme (1)–(3) have been performed.

The ocean block is the MPIOM model (version 1.2) with a resolution of 256 × 220 × 40 grid points, by longitude, latitude, and vertical direction, respectively. The resolution is not uniform, it is fine in the high latitude domain and in the upper ocean levels; it is coarse in the Southern Ocean and in the deep layers. In the Arctic zone, the horizontal grid mesh is about 30 km. In the ocean model block, ice dynamics and confluence module are developed specially. A full description is presented in the model manual, which can be found in [8, 13]. The atmosphere block represents the ECHAM6 model. Its resolution is T63L47 (the spectrum model with 63 basis eigenfunctions and 47 vertical levels), which regards to approximately 240 km of the spatial grid distance. The ocean-atmosphere interaction is realized through the OASIS block.

Initialization of the model has been forced from climate Atlas NCEP (URL: http://www.cpc.ncep.noaa.gov/products/precip/atlas_2/cont_data.html), and the spin up run has been performed to establish the constant energy regime of the currents in the ocean from 1850 to 2010. The date 01.01.2010 was set up as the initial date for the further DA experiments.

Several numerical experiments have been carried out with the DA, in which the model result has been corrected by formulae (1) and (3), while the observed data of the sea level were taken from AVISO data. Vector C for the sea level has been set up as interpolation onto the model grid points of the observed sea level minus model sea level at the previous time-step. For other model variables the values were taken as time average of the corresponding variables minus their model value on the previous time step. Covariance matrix Q has been defined from the ensemble experiments with $N_{ens} = 50$ ensemble terms and has been calculated as

$$Q = N_{ens}^{-1} \sum_{l=1}^{N_{ens}} (Y_l - HX)(Y_l - HX)' + R,$$

where Y_l is the model calculation of the sea level from the ensemble statistics with ensemble number l; X_b is the current model output at the given time-moment. Matrix R was chosen as the diagonal matrix with equal values on the diagonal 0.01 m, as it was recommended in [14].

The experiments have been executed as follows: starting from modeled 2010, the model has been integrated without assimilation (free run) until 2012, then during

3 years (2012–2014) the model has been integrated with the correction each 3 months according to formulae (1) and (3), then again it has been integrated free until 2016. The observations were included at the end of each 3-month period; previously they have been smoothed with 11-point filter along the track. In parallel, the reference experiment, i.e. integration without correction has been performed over the entire period 2010–2016.

Analysis of the computed geophysical fields is presented below. The sea level is the integral variable, which means that its variability stems from the integral variability of many other variables both in the atmosphere and in the ocean, such as wind stress, thermal expansion, precipitation, and many others. This leads to the fact that the assimilation of the sea level implies the changing of many other variables; hence, it will affect the entire system forming the sea level. These processes are very complex; however, one can see the model output of all characteristics after a specific period of model integration. Among the output parameters the sea level fields were the direct output, but we also analyzed the ice thickness and ice concentration. Special attention was focused on the Arctic zone of Russia.

Figure 1 demonstrates the sea level difference (analysis minus reference) after assimilation on 31.12.2011. The main perturbations were found in the domain where the main satellite tracks pass.

Figures 2 and 3 show the ice thickness characteristics. Figure 2 demonstrates the difference between the reference and analysis in the middle of the DA experiments, i.e. on 30.06.2013.

It is seen from Fig. 2 that the ice thickness after assimilation substantially decreases practically everywhere in the Arctic including the Arctic zone of Russia except for a relatively narrow strip north of Canada. These results are confirmed by

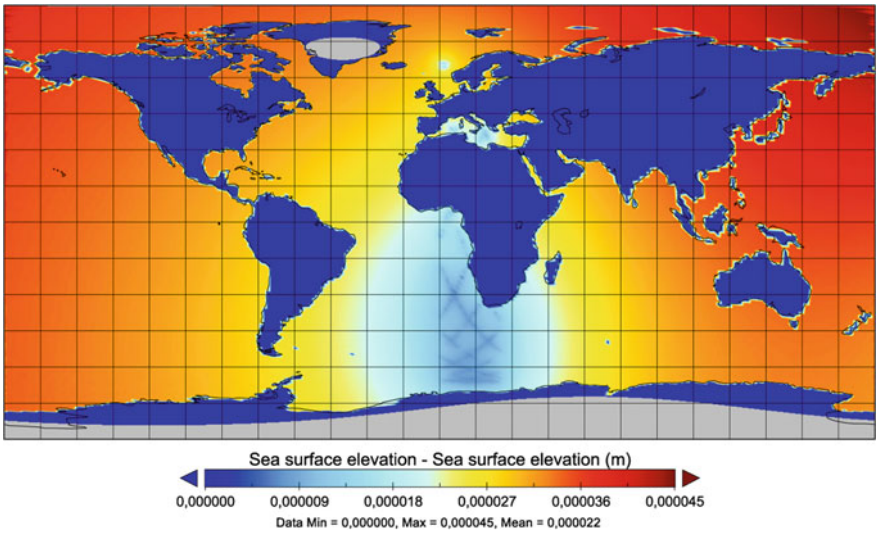


Fig. 1 The difference in the sea level after and before assimilation on 31.12.2011

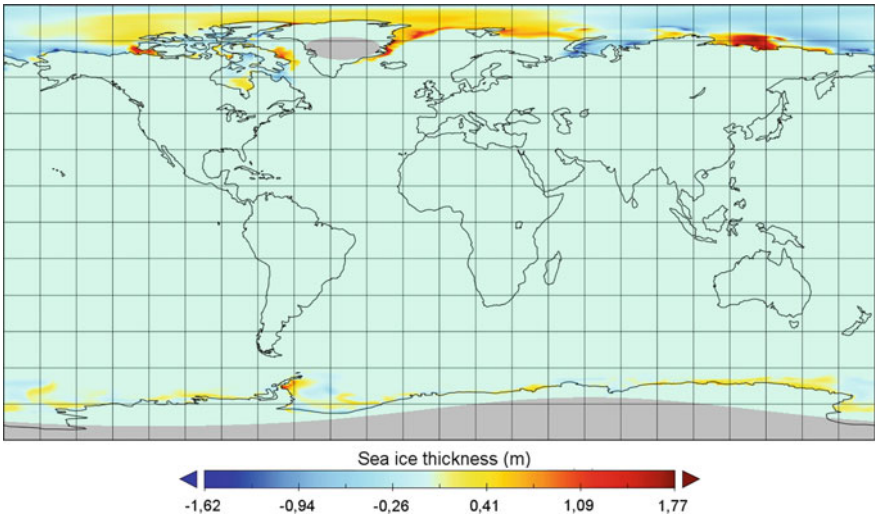


Fig. 2 The difference in the ice thickness (reference minus analysis on 30.06.2013)

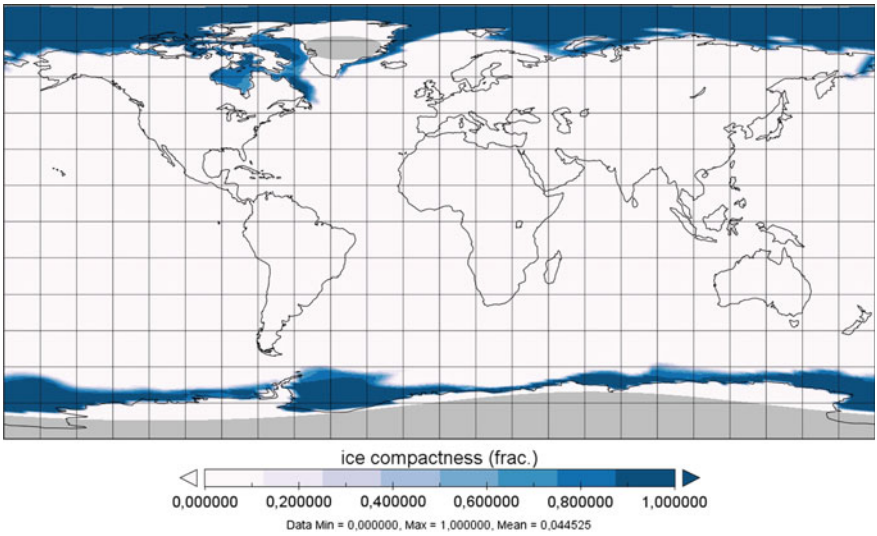


Fig. 3 Computed values of ice compactness (conventional units from 0 to 1) after 18 months of integration

independent observations. For example, in the Kara Sea, the observed values of ice concentration are approximately 72–75% while the model analysis of ice concentration yields about 70%, but the model reference experiment yields in more than 80%.

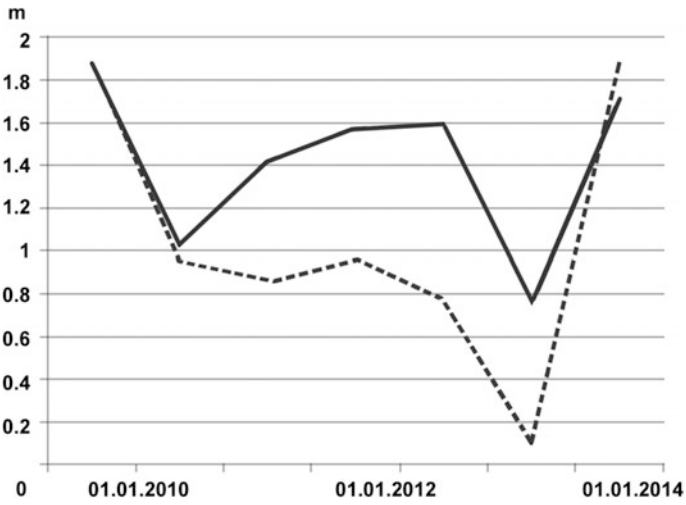


Fig. 4 Time-variability of the average ice thickness in the Arctic in 2010–2014. Reference experiment is shown with the solid line, analysis is shown with a dashed line

The results of calculating of the ice compactness presented in Fig. 3 with the data assimilation correspond well to the observed tendencies. The decrease in the ice thickness in the Arctic indicates the warming of sea temperature at the Polar latitudes and agrees with the observations.

Figure 4 shows the pronounced difference of time-variability of average ice thickness both reference (solid line) and analysis (dashed line). One can see that the reference results overestimates the forecast of ice thickness, but both curves maintain a general tendency, which agrees with the observations during the period 2010–2014. After halting the assimilation, the analysis curve returns to the model reference one.

The results of numerical experiments bring us to the conclusions that the model climate fields after assimilation agree with the natural tendencies. In the future experiments we plan to increase the number of observations and improve the comparison based on the assimilation methods and strategy.

The second series of experiments has been carried out with the temperature-salinity profiles downloaded from the WOCE data. This was the WOCE section A06, the chief scientist was Christian Colin from the Office de la Recherche Scientifique et Technique Outre-Mer (ORSTOM, France). The WOCE section was occupied along 8° N and continued along the 10° E meridian. The dates of the cruise on R/V "L'Atalante' were from January 2, 1993 to March 19, 1993.

The data from the Conductivity-Temperature-Depth (CTD) instrument have been collected from the sea surface to the bottom and include temperature and salinity profiles. Before assimilation all data passed the quality control, the badly recorded data have been discarded. A total of about 700 temperature and salinity records have been included into the assimilation procedure. It is reasonable to note

that only the DA validation and feasibility is presented using this set of observations, their full analysis is not conducted in this study.

The same assimilation technique as in the previous series of experiments has been used. The initial fields, spin up run, and all forces were the same. To create vector *C* the similar procedure has been applied, namely, for the observed values temperature and salinity at each level independently, *C* was calculated as the difference between the observed values at the observational time-step minus the model values at the previous time-step at each grid point. For non-observed values, vector *C* was set up as the model time-average values during the entire spin up run minus

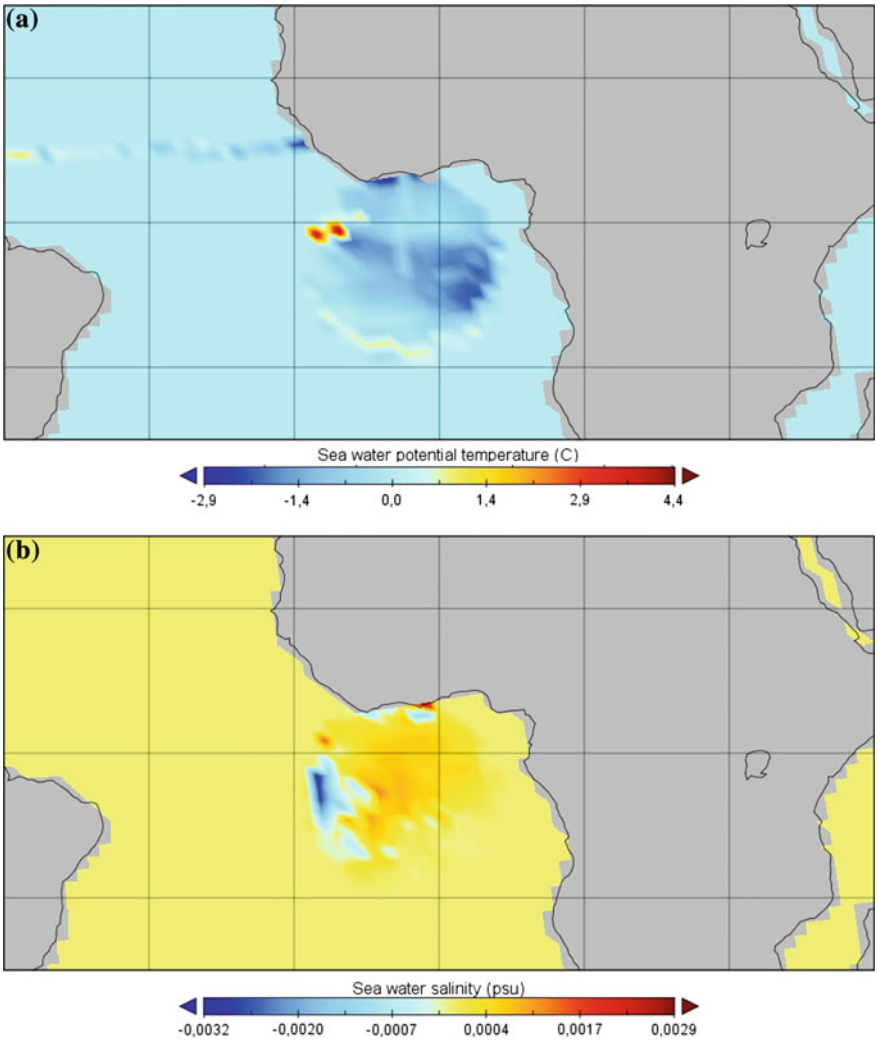


Fig. 5 The DA impact (assimilation minus reference) at a depth of 6 m. **a** Temperature, **b** salinity

the model values at the time step preceding to assimilation. The procedure of the definition of covariance matrix Q was the following: ensemble statistics from the spin up run was taken at the observational points and calculation was done according to its formula. Matrix R for temperature was set at 0.1 for salinity 0.01 on the diagonal and zero at the other positions.

Figure 5 demonstrates the difference between the model and reference experiments after and before assimilation at the sea surface for temperature (Fig. 5a) and salinity (Fig. 5b). It is seen that the main difference in both figures is concentrated in the Guinea Gulf, where the pattern has an ellipsoidal shape in the southern direction. The difference is mostly negative for temperature except two pronounced spots in the western part of the gulf, which appeared due to the occurrence of the local eddies with negative vorticity. This means that the model mostly overestimates the reality. It is reasonable that salinity pattern has an opposite structure, since the total density energy balance must be conserved. The difference along track in Fig. 5a is weakly manifested, and almost disappears in Fig. 5b. This can be explained due to the relatively weak observed difference at these points and due to the fact that the covariance mostly facilitates the propagation of the signal along the track because of the equatorial current.

Figure 6 presents the potential temperature field after assimilation near the sea bottom (3500 m depth). Since the potential temperature at these levels is substantially less then temperature in situ (up to 0.7 °C), it becomes smaller than 1 °C after assimilation (observed temperature is about 1.5 °C). The corrected zone propagates along the equator and enlarges south of the Guinea Gulf. The contrast between the assimilated and non-assimilated zone is very well pronounced, which may be not be physically reasonable. However, one may note that this effect occurs

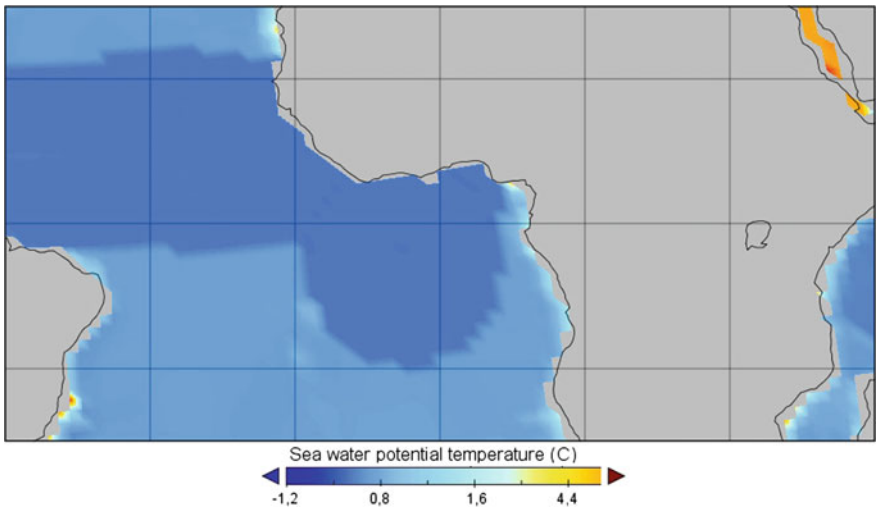


Fig. 6 Corrected potential temperature field near the sea bottom (3500 m)

mostly because of the significant difference between the potential and in situ temperature and due to the fact that the grid resolution at the sea bottom is large enough.

Conclusions

We emphasize the importance and significance of the DA strategy in the climate research. By changing the initial conditions and setting them up with the DA methods and different set of data it is possible to investigate the sensitivity of the model and also the influence of the different regions on the initial data perturbations. It allows us to make the long- and medium range forecasts of different parameters. It is important to note, that in these experiments it is necessary to keep the balance between the input variables and conserve the integral mass and energy. The data assimilation method is developed exactly for these purposes.

Acknowledgements This work was supported by the Russian Science Foundation: numerical calculations were supported by project 14-11-00434.

References

1. Belyaev, K., Meyers, S., & O'Brien, J. J. (2000). Application of the Fokker-Planck equation to data assimilation into hydrodynamical models. *Journal of Mathematical Sciences, 99*(4), 1393–1402.
2. Belyaev, K. P., Kuleshov, A. A., Tuchkova, N. P., & Tanajura, C. A. S. (2016). A correction method for dynamic model calculations using observational data and its application in oceanography. *Mathematical Models and Computer Simulations, 8*(4), 391–400.
3. Belyaev, K. P., Kuleshov, A. A., & Tanajura, C. A. S. (2016). An application of a data assimilation method based on the diffusion stochastic process theory using altimetry data in Atlantic. *Russian Journal of Numerical Analysis and Mathematical Modelling, 31*(3), 137–148.
4. Bryan, K. (1969). A numerical method for the study of the World Ocean. *Journal of Computational Physics, 3,* 347–376.
5. Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics, 53,* 343–367.
6. Gandin, L. S. (1963). *An objective analysis of hydrometeorologycal fields*. Leningrad: Gidrometizdat (in Russian).
7. Ghil, M., & Malnotte-Rizzoli, P. (1991). Data assimilation in meteorology and oceanography. *Advances in Geophysics, 33,* 141–266.
8. Haak, H. (2004). Simulation of low-frequency climate variability in the north atlantic ocean and the arctic, V.1. Max Planck Institute for Meteorology.
9. Kirtman, B., & Shukla, J. (2002). Interactive coupled ensemble: A new coupling strategy for CGCMs. *Geophysics Research Letters, 29*(10), 245–272.
10. Marchuk, G. I., & Sarkisayn, A. S. (1988). Mathematical modeling of ocean circulation. Moscow, Nauka (in Russian).

11. Moiseev, N. N., Alexandrov, V. V., & Tarko, A. M. (1985). A man and biosphere. In The system experiments with models. Moscow, Nauka (in Russian).
12. Tanajura, C. A. S., & Belyaev, K. (2009). A sequential data assimilation method based on the properties of diffusion-type process. *Applied Mathematical Modelling, 33,* 2165–2174.
13. Wetzel, P., Haak, H., Jungclaus, J., & Maier-Reimer, E. (2004). The Max-Planck-institute global ocean/sea-ice model. Model MPI-OM Technical report. http://www.mpimet.mpg.de/fileadmin/models/.
14. Xie, J., Counillon, F., Zhu, J., & Bertino, L. (2011). An eddy-resolving tidal -driven model in China south sea assimilation along track satellite data using EnOI. *Ocean Science, 7,* 609–627.
15. Zalesny, V. B., & Rusakov, A. S. (2007). Numerical algorithm of data assimilation based on splitting and adjoint equation methods. *Russian Journal of Numerical Analysis and Mathematical Modelling, 22*(2), 199–219.

Sea of Azov Waters in the Black Sea: Do They Enhance Wind-Driven Flows on the Shelf?

Peter O. Zavialov, Alexander S. Izhitskiy and Roman O. Sedakov

Introduction

The region of this study is shown in Fig. 1. The principal factor determining the regime of this part of the Crimean shelf of the Black Sea is its proximity to the Kerch Strait connecting the Black Sea and the Sea of Azov. Shallow and brackish, the latter is largely controlled by continental discharges from several rivers, of which the biggest are the Don and the Kuban. The transformed continental runoffs drain to the Black Sea through the Kerch Strait (or, more precisely, through the upper layer of the strait, while in the lower layer, typically, there exist less intense flow of the opposite sign). On long-term average, the annual discharge from the Kerch Strait is estimated at 75 km^3, and the return flow totals to about 50% of this volume [4, 5]. These numbers constitute significant components of the Black Sea water budget even at the entire basin scales, not to mention the shelf immediately adjacent to the strait.

It can be seen in many satellite scenes of the region how turbid sediment carrying waters of the Sea of Azov propagate mainly westward as a surface-trapped low salinity plume along the Crimean coast (see example in Fig. 2). This flow affects significantly the oceanographic regime and the ecological state of the area, as it transports nutrients but also pollutants [8, 9], given that the Kerch Strait and the Sea of Azov proper are subject to very strong anthropogenic loads. In many cases, the Sea of Azov plume can be traced for over 100 km in a relatively narrow (≈10 km) stripe stretching westward along the shore up to Feodosia Bay and even farther west. In contrast with the area of the Kerch Strait, the part of the Black Sea shelf west of the strait has been relatively poorly explored in the past, only a handful of research cruises took place there (e.g., [6, 7]).

P. O. Zavialov (✉) · A. S. Izhitskiy · R. O. Sedakov
Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
e-mail: peter@ocean.ru

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_28

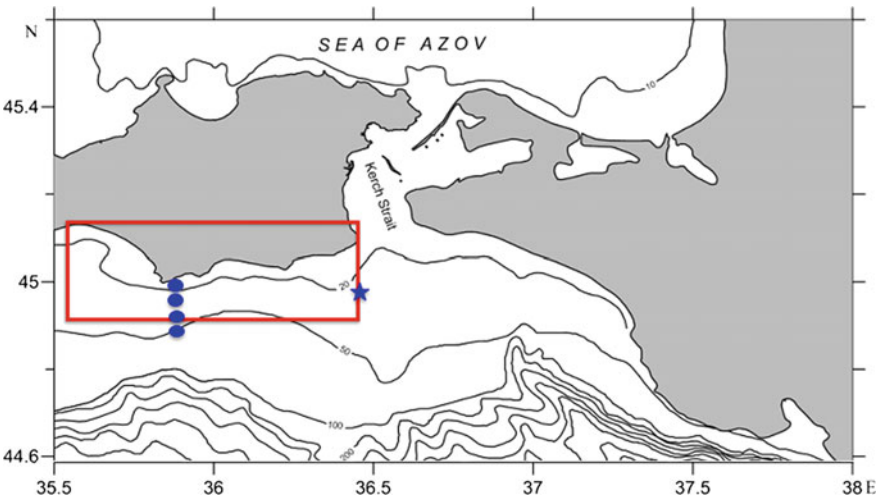


Fig. 1 Map of the study region and bathymetry. The blue star indicates the location of the mooring station, and the blue bullets show CTD stations referred to in the text

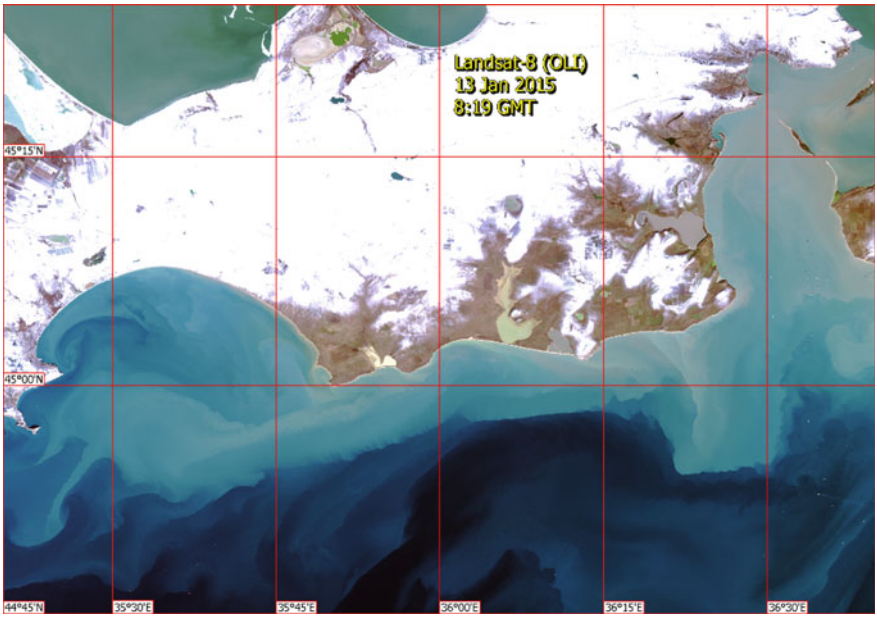


Fig. 2 LANDSAT visible band satellite image of the study region (January 13, 2015), cleary showing high turbidity, low salinity water originating from the Kerch Strait and propagating westward along the coast of Crimea

In May 2015, the Shirshov Institute of Oceanology conducted a field survey in the area. By means of direct measurements at a mooring station deployed at a depth of 22 m on the shelf just west of the strait outlet, it was observed that the cross-shore components of the current velocity in the surface and the bottom layers were strongly anti-correlated during the periods of stratifications, and completely uncorrelated during the periods when stratification relaxed. This suggests that the buoyant discharges from the Kerch Strait may enhance wind-driven currents on the eastern Crimean shelf. If so, this may represent an important mechanism responsible for cross-shelf exchanges in the region.

Browsing through the literature, we encountered some reports on the influence of stratification generated by continental discharges on the one hand and the wind-driven cross-shelf flows on the other. For example, as argued by [2], both observations and numerical modeling of the Pearl River plume in the South China Sea shelf showed that the enhancement of stratification by the plume thins the surface frictional layer and enhances the cross-shelf circulation in the upper water column so that the surface Ekman current and compensating flow beneath the plume are indeed amplified. Some related issues were discussed in early papers dedicated to the theory of upwellings (e.g., [1, 3]). However, to our knowledge, the ability of a buoyant Sea of Azov plume to enhance the upwelling-downwelling type wind-driven cross-shelf circulations in the Black Sea has not been investigated.

In what follows, we first describe the observational evidence that gave motivation to this paper. Then, we present two simple, semi-analytical 2D models of wind-driven flow on the Crimean shelf, aimed to simulate the effect of damping vertical mixing on the cross-shelf transport. One model represents the two-layered stratification and the other one the continuous stratification with eddy viscosity linearly decreasing downwards.

Observations

The map of field measurements on May 18–24, 2015, was shown above in Fig. 1. The mooring station indicated by the blue star was deployed at a depth of 22 m and functioned during 6 days. It was equipped with acoustic Nortek Aquadopp current meter at a depth of 5 m and mechanical Sea Horse tilt current meter at the bottom, and had also temperature sensors at the same depths. All data were recorded as 1 min averages.

The CTD profiling by SBE19plus instrument conducted on May 20, 2015, at four stations situated along a north-to-south transect from the Cape Chauda as indicated in Fig. 1, gives an idea of the vertical distributions of thermohaline fields (Fig. 3). It exhibits a pronounced two-layer stratification, with a warm (up to 16 °C) and relatively fresh (below 17.8 psu), 10–15 m thick upper layer formed by admixture of Sea of Azov water above the colder (below 13 °C) and saltier (up to 18.2 psu) Black Sea water. Between the two layers, there was a steep picnocline.

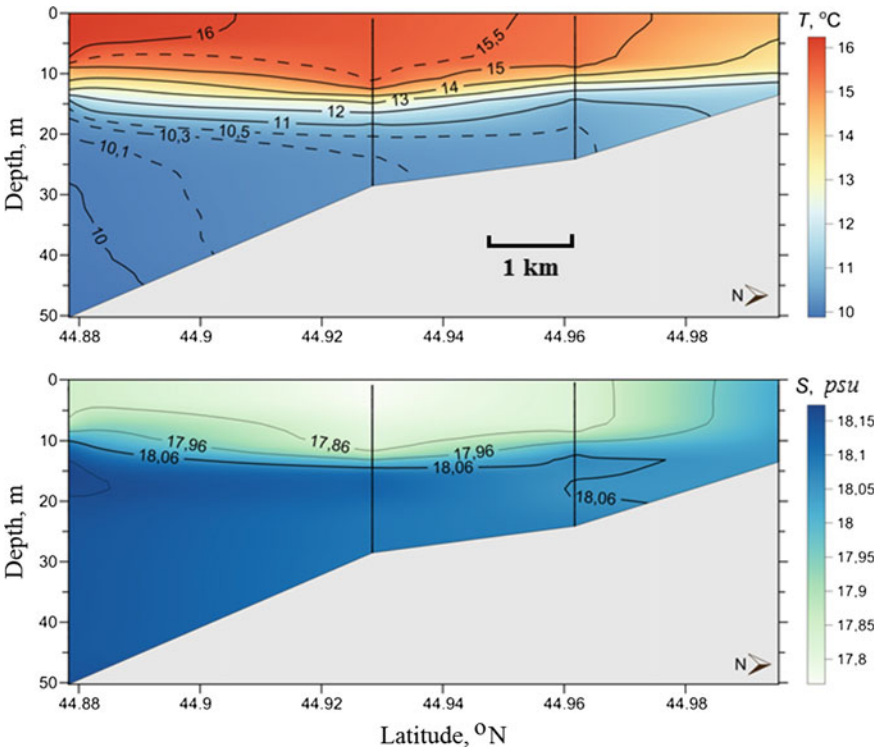


Fig. 3 Vertical distributions of temperature and salinity based on CTD profiling along the north-to-south transect from the Chauda Cape; May 21, 2015

Wind speed and direction were recorded for most of the observation period as 3 h averages at a meteorological station in the city of Kerch at the western bank of the strait (Fig. 4, top panel). The wind record shows moderate westerly winds on May 17–19. During this period, the cross-shore component of the wind exhibited significant diurnal variability, changing from onshore during the first half of the day to offshore afterwards. On May 20, the wind regime changed, and more constant northeasterly winds prevailed since then. We will see from what follows that this variability generated certain response in sea currents.

The data from the mooring station are presented in Fig. 4. In the central panel, the blue and red curves represent water temperature at 5 m and 21.5 m, respectively. The difference between them is a good proxy for stratification. It can be seen that during much of the observation period, the stratification was well manifested, with the temperature difference attaining values up to 4 °C. This stratification, however, relaxed fully on the last two days of the measurements. This relaxation occurred because the warm Sea of Azov plume occupied the entire water column, and the temperature at the bottom increased and matched the temperature at the surface.

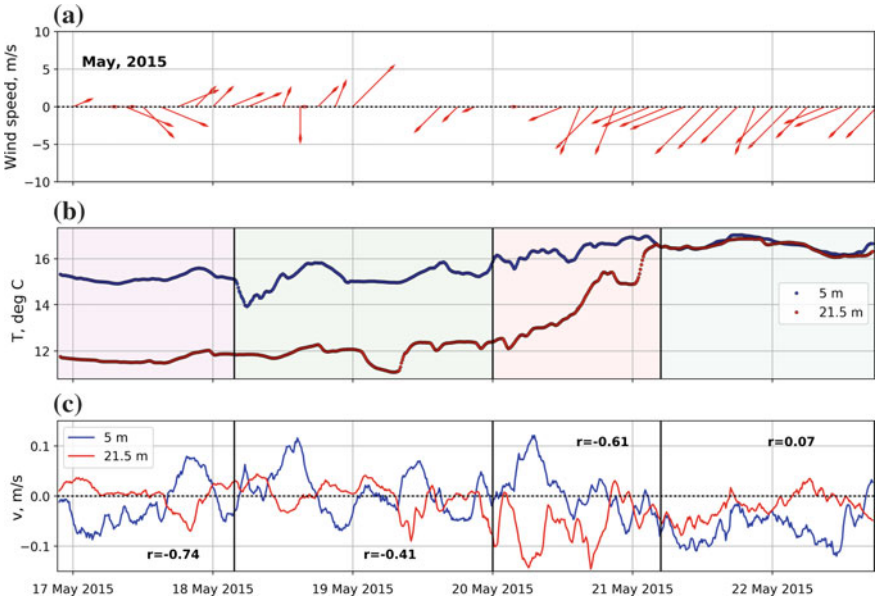


Fig. 4 Data from the thermistors and current meters at the mooring station. Top panel: temperature in the surface layer (5 m depth, blue curve) and the bottom layer (21.5 m depth, red curve). Central panel: temperature difference between the surface layer and the bottom layer (original in red and detrended in blue). Bottom panel: cross-shore velocity component at 5 m (red) and 21.5 m (blue). Numbers indicate correlation coefficient between the two series for selected time periods whose limits are shown by vertical lines

The lower panel in Fig. 4 exhibits the cross-shore component of the current velocity (which in this case coincides with the north-to-south component) at the depths of 5 and 21.5 m. It can be seen even visually that during the period of stratification, the cross-shelf currents in the two layers are anti-correlated and respond to wind variability in the upwelling/downwelling pattern: when the wind is onshore, the surface current is onshore and the bottom current offshore, and vice versa. The cross-shore currents respond to the cross-shore, and not the along-shore, wind stress component, which means that the mooring was located too close to the coast for the Coriolis force to interfere with the direct wind drag, and we deal with non-Ekman upwelling/downwelling circulation.

The correlation coefficient between the flows at the surface and at the bottom at different periods is shown as numbers in the bottom panel of Fig. 4, it reaches −0.74 when stratification is the highest. On the other hand, the correlation vanishes as soon as the stratification relaxes. We hypothesize, therefore, that stratification generated by discharge through the Kerch Strait may intensify wind-driven cross-shelf exchanges. In an attempt to verify and quantify this effect, in the next section we make use of idealized mathematical models.

Simple Mathematical Models

It is our objective to reproduce and quantify the effect of the flow intensification by means of the simplest models possible. We will consider the stationary state, 2-dimensional setting (i.e., conditions uniform in along-shore direction), and no Coriolis force as commented above. On the other hand, we retain the "real" bottom topography in the cross-shore direction.

Model 1: Density is uniform, eddy viscosity linearly decreases with depth

Schematic illustrating settings and notation for this model is shown in Fig. 5. Assuming the balance between pressure gradient and vertical friction, the momentum equation in this case reads:

$$g\frac{\partial h}{\partial x} = \frac{\partial}{\partial z}\mu(1+\kappa z)\frac{\partial u}{\partial z}$$

where g is gravity acceleration, h is the elevation of the sea surface with respect to the equilibrium level, x is the cross-shore coordinate, z is the vertical coordinate, μ is eddy viscosity at the surface, κ is the constant decrement describing linear decrease of eddy viscosity with depth, u is the horizontal (cross-shore) velocity.

This equation should be complemented with the continuity equation

$$\frac{\partial w}{\partial z} = -\frac{\partial u}{\partial x}$$

and the boundary conditions:

$$\rho\mu\frac{\partial u}{\partial z}\Big|_{z=0} = \tau,$$

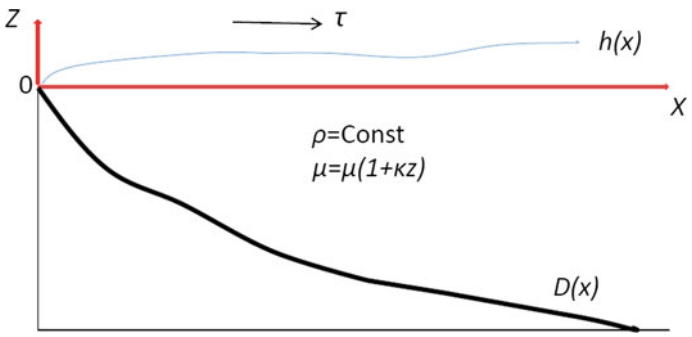


Fig. 5 Schematic illustrating settings and notation for Model 1

$$u(D) = 0,$$

$$w(D) = 0,$$

$$\frac{w(0)}{u(0)} = h'(x).$$

The boundary conditions, where, τ is the wind stress module and ρ is water density, represent the balance of frictional stresses at the surface, the no-motion conditions at the bottom, and the no normal flow condition at the surface. Hereinafter, the prime designates the derivative with respect to x.

The momentum equations can be integrated straightforwardly with respect to z even analytically, but all coefficients in the respective expressions are unknown functions of x. We use the boundary conditions to determine these functions and arrive at equations that, for a specific bathymetry D(x), can only be solved numerically. After simple but rather cumbersome calculations, we obtain the following solution:

$$u(x,z) = \frac{S(x)}{k^2}[1 + kz - \ln(1 + kz)] + \frac{A}{\mu k}\ln(1 + kz) + B(x),$$

$$w(x,z) = \frac{S'}{k^3}(1 + kz)ln(1 + kz) - \frac{S'}{2k}z^2 + \left[f(x) - \frac{2S'}{k^2}\right]z - \frac{S'}{k^3} + C(x),$$

$$h(x) = \frac{\mu}{g}\int_0^x Sdx,$$

where

$$A = \frac{\tau}{\rho},$$

$$B(x) = -\frac{\tau}{\rho\mu k}ln(1 + kD) - \frac{S}{k^2}\{1 + kD - ln(1 + kD)\},$$

$$C(x) = -\frac{S'}{k^3}(1 + kD)ln(1 + kD) + \frac{S'}{2k}D^2 - \left\{f(x) - \frac{2S'}{k^2}\right\}D + \frac{S'}{k^3},$$

$$f(x) = \frac{-S'}{k^2} + \frac{S'D + D'S}{k} - \frac{1}{k^2}\left\{S'ln(1 + kD) + \frac{SkD'}{1 + kD}\right\} + \frac{\tau}{\rho\mu}\frac{D'}{(1 + kD)},$$

and function S(x) is a solution of the following equation:

$$S' = -\frac{S^2\left\{\frac{k\mu}{g}(ln(1+kD)-kD)\right\} - S\left[k^2DD' + \frac{k^2DD'}{1+kD} - \frac{\tau k^2}{g\rho}ln(1+kD)\right] + \frac{k^3\tau DD'}{\rho\mu(1+kD)}}{(1+kD)ln(1+kD) + \frac{k^2D^2}{2} - kD(1+ln(1+kD))}.$$

We obtain S(x) by integrating the latter equation numerically, starting from the coast where we set S(0) = 0 and then obtain all other unknown functions, which fully determine the solutions for u(x, z), w(x, z), and h(x).

An example of the solution is graphically shown in Fig. 6. This example clearly showing downwelling-type circulation uses real bathymetry and corresponds to moderate onshore wind stress at 0.01 N/m^2. The coefficient μ is 10^{-4} m^2/s, and the decrement k is 10^{-3}.

We then used Model 1 to investigate the influence of stratification on cross-shelf transport determined as the vertical integral of u from the bottom upwards to the depth where the velocity changes sign (of course, the transport integrated over the entire water column equals to zero). In the model experiments, we varied decrement k and analyzed how the transport changed. The larger is k, the more stratified is the column and mixing in its lower portion supressed. The results are shown in Fig. 7. It can be seen that the enhancement of stratification with respect to μ can indeed significantly increase the transport, for example, the increase of the parameter k by 50% results in 10% increase of the transport at av wind speed of 3 m/s and 30% increase of the transport at wind speed 10 m/s. The relations shown in Fig. 7 seem

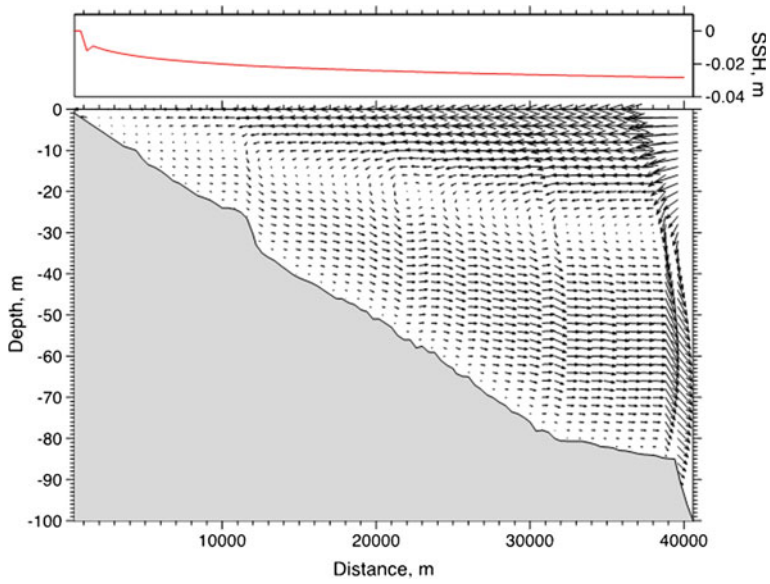


Fig. 6 Example of solution of Model 1 with real bathymetry and constant, moderate onshore (i.e., downwelling) wind stress at 0.01 N/m^2. The μ coefficient is 10^{-4} m^2/s, and the decrement k is 10^{-3}. The velocity scale is such that the average velocity module is about 0.06 m/s

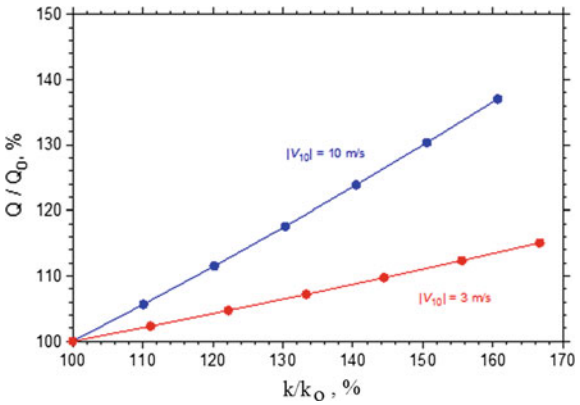


Fig. 7 Dependencies of the normalized cross-shore volume transport on the normalized decrement k for two different values of the wind speed. Q_0 is the volume transport at $k = k_0 = 10^{-3}$ and $\mu = \mu_0 = 10^{-4}$ m^2/s, in further experiments with increased k the values of μ were adjusted to keep the vertically averaged eddy viscosity constant

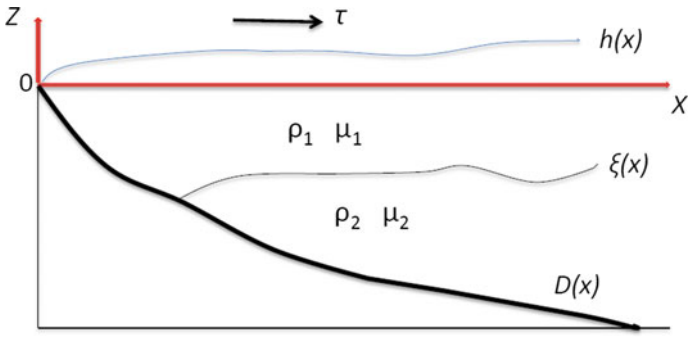


Fig. 8 Schematic illustrating the setup and notation for Model 2

linear, but in fact they are not, for example, in an experiment with extremely strong wind ($\tau = 0.8$ N/m^2, not shown here) the dependence was strongly nonlinear.

Model 2: Two layers of constant density and eddy viscosity, $\rho_1 < \rho_2$, $\mu_1 \geq \mu_2$

Schematic illustrating settings and notation for this model is shown in Fig. 8. Now we deal with two layers, each characterized by constant density (that of the upper layer lower) and eddy viscosity (that of the upper layer higher). The notation used is similar to that of Model 1, except that subscripts 1 and 2 refer to the variables pertaining to the upper and the lower levels, and the depth of the interface between them is $\zeta(x)$.

The momentum equations for the two layers now read

$$\mu_1 \frac{\partial^2 u_1}{\partial z^2} = g \frac{\partial h}{\partial x},$$

$$\rho_2 \mu_2 \frac{\partial^2 u_2}{\partial z^2} = g\rho_1 \frac{\partial h}{\partial x} - g(\rho_2 - \rho_1) \frac{\partial \xi}{\partial x}.$$

The continuity equations take the form

$$\frac{\partial w_1}{\partial z} = - \frac{\partial u_1}{\partial x},$$

$$\frac{\partial w_2}{\partial z} = - \frac{\partial u_2}{\partial x}.$$

In this case, there are also as many as 6 boundary conditions:

$$\rho_1 \mu_1 \frac{\partial u_1}{\partial z}\bigg|_{z=0} = \tau,$$

$$u_2(D) = 0,$$

$$w_2(D) = 0,$$

$$\frac{w_1(\xi)}{u_1(\xi)} = \xi'(x),$$

$$\frac{w_2(\xi)}{u_2(\xi)} = \xi'(x),$$

$$\rho_1 \mu_1 \frac{\partial u_1}{\partial z}\bigg|_{z=\xi} = \rho_2 \mu_2 \frac{\partial u_2}{\partial z}\bigg|_{z=\xi}.$$

The solution for Model 2 is given by the following formulas:

$$u_1(x,z) = \frac{S_1(x)}{2} z^2 + A_1 z + B_1(x),$$

$$u_2(x,z) = \frac{S_2(x)}{2} z^2 + A_2 z + B_2(x),$$

$$w_1(x,z) = - \frac{S_1'(x)}{6} z^3 - \frac{A_1'(x)}{2} z^2 - B_1'(x) z + C_1(x),$$

$$w_2(x,z) = -\frac{S_2'(x)}{6}z^3 - \frac{A_2'(x)}{2}z^2 - B_2'(x)z + C_2(x).$$

$$S_1(x) \equiv \frac{g}{\mu_1}\frac{\partial h}{\partial x},$$

$$S_2(x) \equiv \frac{\rho_1}{\rho_2}\frac{\mu_1}{\mu_2}S_1 - \frac{g}{\mu_2}\frac{\Delta\rho}{\rho_2}\frac{\partial \xi}{\partial x}, \Delta\rho = \rho_2 - \rho_{1,}$$

$$A_1 = \frac{\tau}{\rho_1\mu_1},$$

$$A_2(x) = \frac{\rho_1}{\rho_2}\frac{\mu_1}{\mu_2}[S_1\xi + A_1] - S_2\xi,$$

$$B_2(x) = -\frac{S_2}{2}D^2 - A_2D,$$

$$C_1(x) = \frac{\mu_1}{g}S_1B_1,$$

$$C_2(x) = \frac{S_2'}{6}D^3 + \frac{A_2'}{2}D^2 + B_2'D,$$

where functions $S_1(x)$ and $B_1(x)$ are solutions of the following equations:

$$S_1'\left[-\frac{rD^3}{3} + \frac{rD^2}{2}\xi - \frac{r\xi^3}{6}\right] + S_1[\xi + 2D^2D'r - \frac{\xi\xi^2}{2} + \frac{\xi'D^2}{2}r] + F = 0, \quad \text{where}$$

$$F = \frac{2}{3}\alpha\xi^3\xi'' + \alpha\left(D^2 - \frac{\xi^2}{2}\right)(\xi')^2 + D'r\frac{\tau}{\rho_1\mu_1}(D - \xi) - \alpha\xi''\xi^2D - \alpha\xi'\xi^2D' - \frac{\alpha}{2}\xi''\xi D^2$$
$$- 3DD'\alpha\xi'\xi - \frac{2}{3}\alpha D^3\xi'' - 2D^2\alpha D'^{\xi'} - \frac{\alpha}{2}(\xi')^2\xi^2 - \xi'\frac{r\tau}{\mu_1\rho_1}(D + \xi) - D\alpha(\xi')^2\left(\xi + \frac{D}{2}\right),$$
$$r = \frac{\rho_1}{\rho_2}\frac{\mu_1}{\mu_2},$$
$$\alpha = \frac{g}{\mu_2}\frac{\Delta\rho}{\rho_2},$$

and

$$\xi B_1' = \frac{\mu_1}{g}S_1B_1 - \frac{S_1'}{6}\xi^3 - \xi'\left\{\frac{1}{2}S_1\xi^2 + A_1\xi + B_1\right\}.$$

An example of the solution is shown in Fig. 9. This particular experiment used real bathymetry, constant onshore (i.e., downwelling) wind stress at 0.1 N/m^2, and

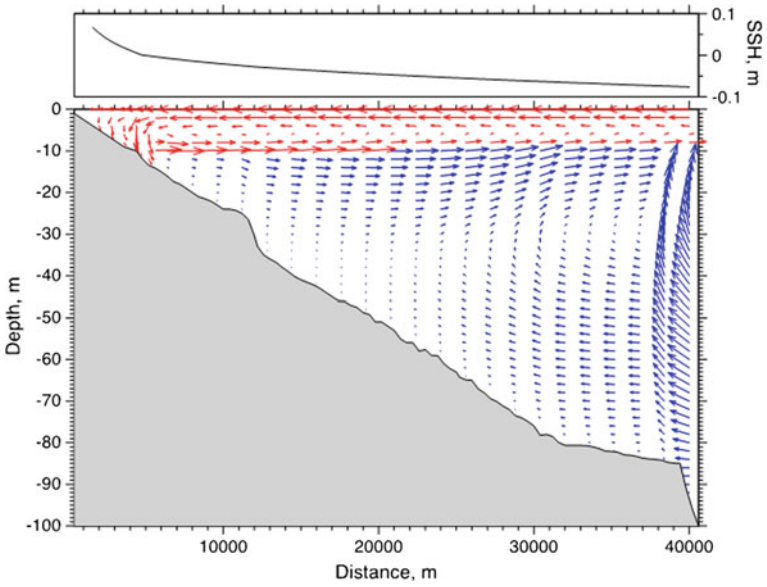


Fig. 9 Example of solution of Model 2 with real bathymetry and constant onshore (i.e., downwelling) wind stress at 0.1 N/m^2. Coefficients μ_1 and μ_2 are 3×10^{-3} m^2/s and 3×10^{-4} m^2/s, respectively. The velocity scale is such that the average velocity module for this image is about 0.21 m/s

prescribed gently shallowing seaward interface between the two layers at a depth of about 10 m. Coefficients μ_1 and μ_2 are 3×10^{-3} m^2/s and 3×10^{-4} m^2/s, respectively.

The velocity field demonstrates two vertical cells forming downwelling in the upper layer and upwelling in the lower level. In terms of momentum budget, this complex pattern can be explained as follows. Near the surface, the current follows the wind stress. The frictional drag weakens with the depth, and at some depth within the upper layer the balance between barotropic pressure gradient and friction implies the reversal of the flow. Beneath the interface, the current is initially driven by shear stress with the upper layer and baroropic pressure gradient, but, as the depth increases further, the baroclinic pressure gradient eventually prevails, and the balance implies another reversal of the current at some depth within the lower layer.

How does the total cross-shelf transport in such two-layered system depend on stratification? The outcome from model experiments is shown in Fig. 10 presenting the normalized total transport (defined as above for Model 1) as function of the ratio $\frac{\mu_1}{\mu_2}$. The higher is the ratio, the more pronounced is the stratification of the water column generated by the Sea of Azov "plume". It can be seen from Fig. 10 that an increase of the ratio leads to significant enhancement of the vertically integrated transport which is almost linear with respect to the ratio and also increases with μ_1.

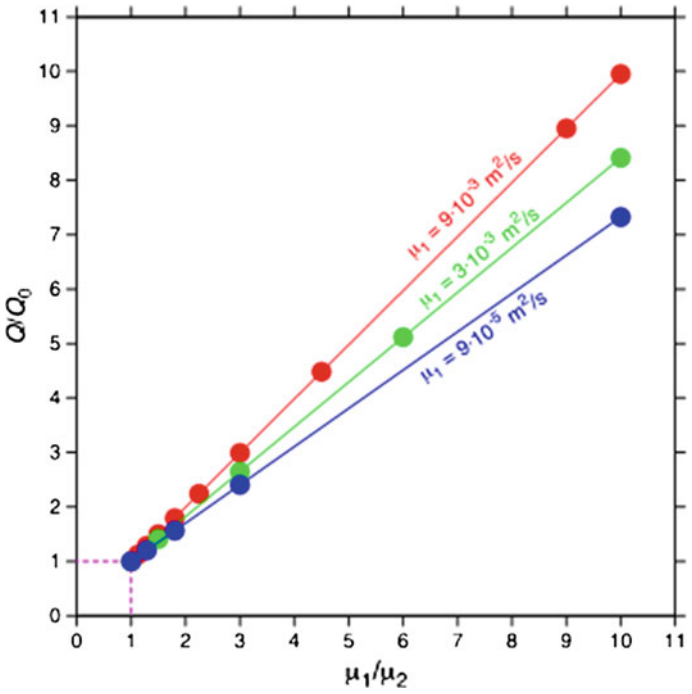


Fig. 10 Dependencies of the normalized cross-shore volume transport on the ratio $\frac{\mu_1}{\mu_2}$ for Model 2

Conclusions

In the field survey conducted in May 2015 on the eastern shelf of the Crimea Peninsula in the Black Sea, it was observed that stratification of the water column was largely controlled by wind-driven dynamics of buoyant surface plume originating from the discharges of the Sea of Azov water through the Kerch Strait. Further, it was observed that the cross-shore components of the current velocity in the surface and bottom layers were strongly anti-correlated during the periods of stratification, but completely uncorrelated during the periods when stratification relaxed. This suggests that the buoyant discharges from the Kerch Strait may enhance wind-driven currents responsible for cross-shelf exchanges. To verify this hypothesis, we developed two idealized semi-analytical 2D models of wind-driven flow on the Crimean shelf, aimed to simulate the effect of damping vertical mixing on the cross-shelf transport. One model represented the two-layered stratification, and the other one continuous stratification with turbulent viscosity linearly decreasing downwards. Both models demonstrated that, indeed, the plume-generated stratification may significantly enhance both the cross-shelf wind drift in the upper layer and the compensating flow in the bottom layer. This effect can contribute significantly to the efficiency of cross-shelf exchanges in the study

region. To accurately quantify this mechanism, further studies based on more extensive observations and perhaps less idealized modeling are needed.

Acknowledgements The field work was partly funded by the Russian Science Foundation (grant 14-50-00095), the modeling study was supported by the Russian Ministry of Science and Education (project 14.W03.31.0006).

References

1. Csanady, G. T. (1981). On the structure of transient upwelling events. *Journal of Physical Oceanography, 12,* 84–97.
2. Gan, J., Li, L., Wang, D., & Guo, X. (2009). Interaction of a river plume with coastal upwelling in the northeastern South China Sea. *Continental Shelf Research, 29,* 728–740. https://doi.org/10.1016/j.csr.2008.12.002.
3. Garvine, R. W. (1972). The effect of bathymetry on the coastal upwelling of homogeneous water. *Journal of Physical Oceanography, 3,* 47–56.
4. Ilyin, Y. P., Repetin, L. N., Belokopytov, V. N., Goryachkin, Y. N., Dyakov, N. N., Kubryakov, A. A. et al. (2012). Meteorological conditions of the seas of Ukraine. *Black Sea. ECOSY-Gidrofizika, 2*. Sevastopol (in Russian).
5. Ivanov, V. A., & Belokopytov, V. N. (2011). *Oceanography of the Black Sea*. NAS of Ukraine, Marine Hydrophysical Institute, Sevastopol.
6. Izhitskiy, A. S., & Zavialov, P. O. (2017). Hydrophysical state of the Gulf of Feodosia in May, 2015. *Oceanology* (in press).
7. Kondrat'ev, S. I. (2009). Dynamics of hydrochemical composition of waters in the Feodosia Bay as the result of water inflow from the sea of Azov during winter in 2006–2007. In *Ecological safety of coastal and shelf zones and complex exploration of shelf resources* (Vol. 18, pp. 30–38). Marine Hydrophysical Institute: Sevastopol.
8. Kotel'yanets, E. A., & Konovalov, S. K. (2008). Distribution of heavy metals in bottom sediments of the Feodosia Bay, ecological safety of coastal and shelf zones and complex exploration of shelf resources (Vol. 17, pp. 171–175). Marine Hydrophysical Institute, Sevastopol.
9. Lomakin, P. D., Chepyzhenko, A. I., & Chepyzhenko, A. A. (2006). Evaluation of concentration of dissolved petroleum products in the coastal waters of Crimea using the data of optical measurements. In *Ecological safety of coastal and shelf zones and complex exploration of shelf resources* (Vol. 14, pp. 245–258). Marine Hydrophysical Institute, Sevastopol.

Bottom Water Flows in the Vema Channel and over the Santos Plateau Based on the Field and Numerical Experiments

Dmitry I. Frey, Vladimir V. Fomin, Roman Yu. Tarakanov, Nikolay A. Diansky and Nikolay I. Makarenko

Introduction

Antarctic Bottom Water (AABW) occupies the bottom layer of oceanic basins of the whole Atlantic. According to the classical definition by Wüst, the upper boundary of this water mass is the isotherm of potential temperature $\theta = 2$ °C [21]. This water is formed at a few locations around the Antarctic slope in the autumn-winter period as a result of cooling and ice formation [17]. In the Atlantic part of the Southern Ocean, it is formed mainly in the Weddell Sea. When AABW reaches the ocean floor it propagates to the north flowing from one basin to another. The final points of its propagation are the East Azores Ridge (37° N) in the Northeast Atlantic and the deep basin east of the Newfoundland Bank in the Northwest Atlantic. The water exchange between these deep basins occurs through the abyssal channels in the bottom topography [15]. The mean velocities of the northward propagation of AABW are less than 1 cm/s; however, when AABW flows through the narrow abyssal channels the current strongly accelerates. Direct measurements of the velocities in these channels make possible calculations of the transport of the northward AABW spreading, which is important because the flows

D. I. Frey (✉) · R. Yu. Tarakanov
Russian Academy of Sciences, Shirshov Institute of Oceanology, Moscow, Russia
e-mail: dima.frey@gmail.com

V. V. Fomin · N. A. Diansky
Zubov State Oceanographic Institute, Moscow, Russia

N. A. Diansky
Lomonosov Moscow State University, Moscow, Russia

N. A. Diansky
Russian Academy of Sciences, Institute of Numerical Mathematics, Moscow, Russia

N. I. Makarenko
Lavrent'ev Institute of Hydrodynamics, Novosibirsk, Russia

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_29

of the coldest bottom waters influence the heat balance of the ocean and the Earth's climate as a whole.

The major part of the Antarctic waters from the Weddell Sea spreads to the north in the western part of the Atlantic Ocean. The further propagation to the East Atlantic occurs through numerous fracture zones of the Mid-Atlantic Ridge. Hence, almost all Antarctic waters propagate through the Argentine and Brazil basins in the southwestern part of the Atlantic. The zonally aligned Rio Grande Rise and Santos Plateau separate these deep basins from each other and prevent the free propagation of bottom water to the north. The key passages for AABW here are the Vema and Hunter channels and also the small channels on the Santos Plateau. The Vema Channel was formed due to the long-term erosion of the bottom by the abyssal currents [7]. Now it is the deepest pathway for Antarctic waters. The Hunter Channel and Santos Plateau are much shallower. Therefore, the coldest part of AABW propagates through the Vema Channel; however, the flow over the wide Santos Plateau makes a significant contribution to the total AABW transport [8, 18].

The most well studied part of the Vema Channel is the Vema Sill at 31° 12′–31° 14′ S. A bathymetric survey of the southern part of the channel was performed in 1993 with a multi-beam echo-sounder [22]. More than 20 CTD sections across the channel have been occupied in this region since the 1970s. Several hydrographic sections were occupied at the Vema Extension at 26° 40′ S [14, 16, 19]. The measurements of velocities in the channel were performed using the Lowered Acoustic Doppler Current Profiler (LADCP) and on a few moorings deployed in 1991–1992 and 2001–2003. The transport of Antarctic waters through the channel is estimated at 1.3–5.3 Sv [15]. In contrast to the Vema Channel, the bottom currents over the Santos Plateau are less studied. Velocities in this region were measured on several moorings deployed across the southern boundary of the Brazil Basin in 1991 [8] and in the northern part of the plateau in 2003 [15]. These measurements show strong variability of the bottom currents. Existing estimates of AABW transport over the Santos Plateau range from −1.5 to 2 Sv [13, 18] indicating that water transport here is possible in both directions.

As for numerical modeling, only one work to our knowledge was dedicated to the simulation of the bottom currents in the Vema Channel [9]. The authors used a two-dimensional version of the sigma-coordinate Princeton Ocean Model for simulating the cross-channel distribution of the thermohaline properties. It was shown that the bottom friction causes the cross-channel Ekman flux and modifies the lateral structure of the flow. That work does not describe the flow in the longitudinal direction. Existing three-dimensional global models of the ocean do not have sufficient lateral resolution for the simulation of the currents in narrow abyssal channels. In addition, the vertical resolution of these models in the bottom layers is not satisfactory for studying the lower part of the ocean circulation.

The goal of this paper is to study the variability of the bottom currents along abyssal channels in the Southwest Atlantic using regional three-dimensional model of the oceanic circulation with high resolution near the bottom. The data of direct measurements collected using the LADCP instrument were used for verification of the simulated velocity fields. These measurements were made at a few locations

along the Vema Channel (33° 34′, 32° 17′, 31° 12′, 28° 55′, 26° 40′ S) in 2009, 2010, 2014, and 2017. We also report about our new hydrographic measurements performed over the Santos Plateau in the spring and autumn seasons of 2016.

Data and Methods

We used both experimental and simulated velocity data to analyze the bottom currents in the region. The hydrographic data were measured at a few locations along the Vema Channel and over the Santos Plateau (Fig. 1). The historical CTD-stations are also shown in Fig. 1. The locations of our stations have been chosen on the basis of the existing bathymetry data. We performed echo-sounder surveys around the stations to obtain precise data of the local seabed topography. All experiments were made from the research vessel "Akademik Sergey Vavilov" of the Russian Academy of Sciences; the measurements were performed almost to the bottom (usually 5 m above the seabed).

Profiles of temperature, salinity, and current velocities were measured by the CTD (SBE 19 plus) and LADCP (RDI WorkHorse Sentinel 300 kHz) profilers. The CTD-data were processed using standard programming package (SBE Data Processing, version 7.23), the LADCP data were processed using programming package

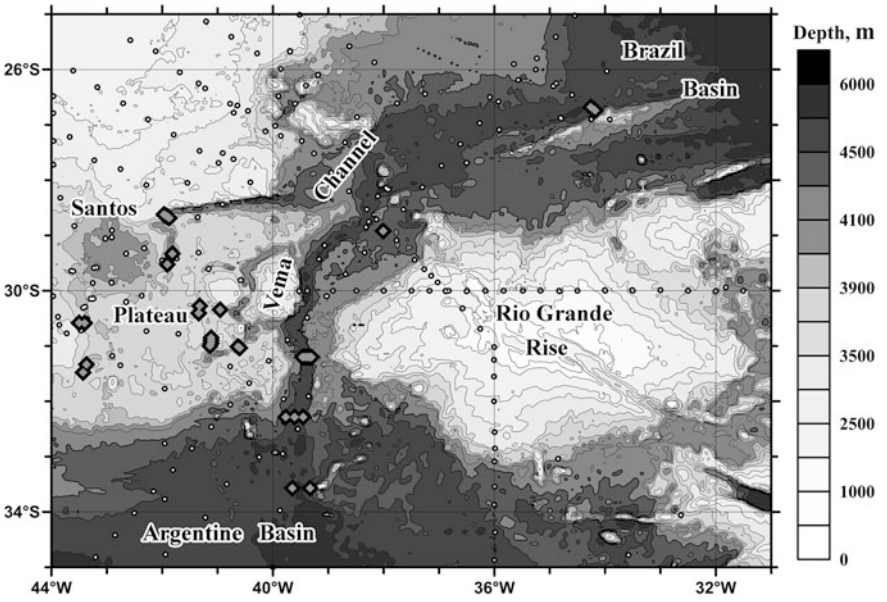


Fig. 1 The bottom topography of the study site based on the GEBCO 2014 data. White circles indicate historical stations; gray diamonds denote our stations with LADCP velocity measurements used in this work

LDEO Software (version IX.10), developed by M. Visbeck. This package is an implementation of the velocity inversion method [20]. The results of processing were corrected by subtracting the tidal velocities. The barotropic tide was calculated from the data of the satellite observations using the method described in [5]. Usually, the velocities of the bottom currents exceed the tidal velocities by one order of magnitude. However, this correction is not negligible for some slow currents over the Santos Plateau. The accuracy of velocity measurements is 2 cm/s, CTD temperature measurements is 0.003 °C, CTD salinity measurements is 0.005 PSU.

The simulation of the near-bottom dynamics was performed using the Institute of Numerical Mathematics Ocean Model (INMOM). The INMOM is a σ-coordinate ocean circulation model based on the primitive equations of ocean hydrothermodynamics with the Boussinesq and hydrostatic approximations [4]. Dimensionless variable σ is used in the model as the vertical coordinate:

$$\sigma = \frac{z - \zeta(x, y, t)}{(H(x, y) - \zeta(x, y, t)},$$

where z is the ordinary vertical coordinate in meters measured from the unperturbed sea surface towards the center of the Earth, H(x,y) is the depth at the given point, $\zeta(x, y, t)$ is the sea surface height (x,y) are the longitude and latitude, respectively. The use of a σ-coordinate model for studying bottom currents is important because the lower layers in the σ-coordinates exactly follow the bottom topography [6]. The INMOM has been well tested in the different regions of the World Ocean including the Atlantic in the frame of the Coordinated Ocean-ice Reference Experiments (CORE) international program [3] and in the eddy resolving regime [1].

In this work we used a regional version of the INMOM developed for the Southwest Atlantic. We selected a rectangular domain between 37.0° and 22.8° S and between 45.2° and 29.0° W, which completely covers the Santos Plateau, the Rio Grande Rise, and parts of the Argentine and Brazil basins. The grid domain contains 814 by 764 nodes in the horizontal plane, the spatial resolution is 0.02°. The vertical direction is divided into 33 σ-levels not uniformly distributed over the depth. The best resolution was at the bottom; at the ocean point with a depth of 4850 m, the 10 deepest levels were specified in the bottom layers from 4400 to 4850 m with a step of 50 m. The model time step was 3 min.

The bottom topography was constructed from the data of Smith and Sandwell digital atlas with an initial resolution of 30′. The original data were corrected using the multi-beam echo sounder survey [22], interpolated to the model domain and smoothed using the Tukey filter. The data from the climatological Levitus Atlas [2, 12] were used to construct the initial conditions for temperature and salinity. These monthly data include the mean climatic fields of temperature and salinity in the World Ocean with a spatial resolution of 0.25°. The Levitus data were also specified in a buffer zone with a width of 10 boundary points at the liquid boundaries [10]. The atmospheric forcing was taken into account by specifying the fluxes of heat, freshwater, and momentum at the boundary. The surface turbulent fluxes were calculated using the bulk formulas. The atmospheric surface characteristics

including air temperature, humidity, and wind velocity at a height of 10 m, precipitation, incident short- and long-wave radiation, and atmospheric pressure were taken from the CORE database [11]. At the ocean surface, the relaxation to the climate salinity and temperature data from the Levitus Atlas was performed with a time scale of about 1 month.

We calculated three-dimensional fields of the horizontal velocities, potential temperature, salinity at σ-levels, and sea surface height. In this work we analyze mainly the velocity fields. For the analysis, we interpolated the results from the σ-levels to the ocean depth. The total duration of the simulations was 45 days. During the first 30 days, the model was spun up in the diagnosis regime using "frozen" temperature and salinity fields. Then, the simulation was continued for 15 days in the adjustment regime. This method allows us to calculate the ocean dynamics on the basis of the known temperature and salinity data.

Vema Channel

The Vema Channel is the main pathway for the coldest bottom waters transported to the tropical Atlantic. The depths of the channel exceed 4600 m against the background depths of 4200 m. The shallowest point of the Vema Channel is the Vema Sill at 31° 12′ S. The maximum depth of this sill is approximately 4614 m, and the width of the channel here is 18 km. The Vema Channel is the best region for investigating the bottom gravitational currents in a deep narrow channel. The transport of AABW in this channel is the highest among all abyssal channels in the Atlantic. The length of the channel (700 km) and relatively constant width (15–20 km) in the southern part of the channel allows us to study AABW flow far away from its inflow from the Argentine Basin. The local bottom relief in the channel is not as rough as in the fracture zones of the Mid-Atlantic Ridge. The Vema Channel is located relatively far from the equator; therefore, it is possible to investigate the influence of the Coriolis force on the bottom current. Regular measurements here have been performed since 1970s, which is helpful for studying long-term variability of the bottom water properties.

Current velocities in the southern part of the Vema Channel calculated from our measurements and numerical simulations are shown in Fig. 2. Since the depths of a significant part of the channel are 4650–4700 m, the data were presented at a depth of 4600 m. We measured the velocities in this region at three zonal sections of the channel at 33° 34′, 32° 17′, and 31° 12′ S. Stations 1708 and 1709 (Fig. 2, left panel) were occupied far away from the inflow of the bottom water to the channel. The bottom velocities here were directed to the south and did not exceed 5–10 cm/s. The numerical simulation shows significant velocities in the channel north of 33° S. The bottom velocities based on experimental observations and numerical modeling were 10–15 cm/s at 32° 17′ S and 25–30 cm/s at 31° 12′ S. According to the model simulations, velocities up to 30 cm/s are observed in the Vema Channel over its entire length. The minimum potential temperature of the bottom waters varies from

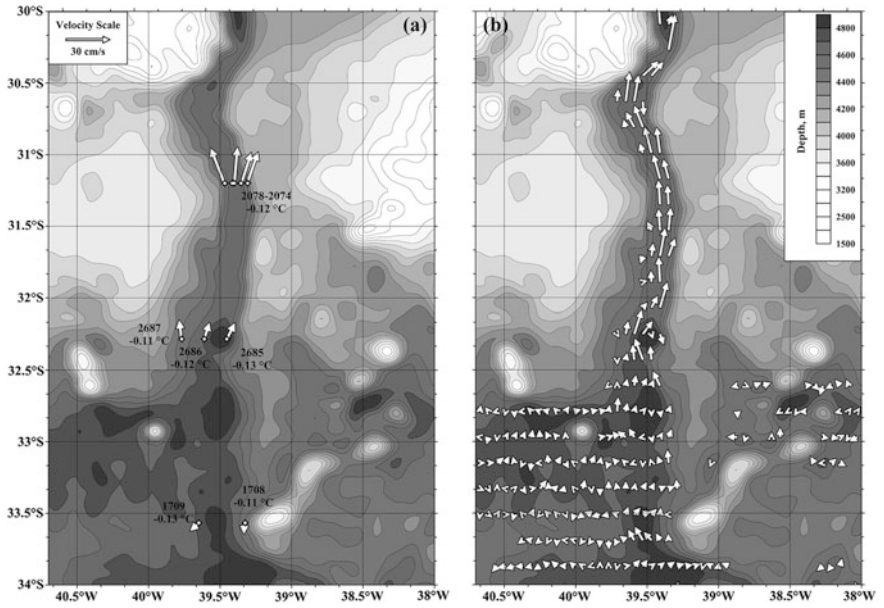


Fig. 2 Inflow of the bottom waters from the Argentine Basin to the Vema Channel. Measured (left panel) and simulated (right panel) velocities are shown at a depth of 4600 m. The numbers of stations and the lowest measured potential temperatures are shown in the left panel

−0.11 to −0.13 °C (Fig. 1a). The increase of the temperature in this part of the channel is approximately 0.01 °C/100 km. Due to the bottom friction, the core of cold water displaces to the east (in this channel) or to the right relative to the flow direction [9]. The observed difference in the bottom temperatures between the eastern and western walls of the channel is 0.02 °C both over the sections at 32° 17′ and 31° 12′ S.

Two branches of the current are observed in the middle part of the channel (Fig. 3a). The northern jet is located in the deepest part of the channel with depths of approximately 4700 m; it is the main passage for the coldest waters. The additional southern jet with velocities up to 20–25 cm/s is observed at depths of 4300–4400 on the southern slope of the channel. The lengths of the jets are 200 km and the maximum distance between them is 50 km. The existence of the southern jet was confirmed by the direct measurements at one station (2535) at 28° 55.3′ S, 38° 00.5′ W. The depth of the ocean at the place of measurements was 4384 m, the minimum potential temperature was −0.09 °C, the current with a velocity of 15 cm/s was directed to the southeast.

Calculated velocities near the exit from the Vema Channel to the Brazil Basin are shown in Fig. 3b. The mean velocities in the center of the flow are up to 30 cm/s, which is close to the velocities measured at stations 2436–2438 (25 cm/s directed to the east and northeast). After the outflow from the Vema Channel, the bottom current turns to the north and propagates along the western boundary of the Brazil Basin

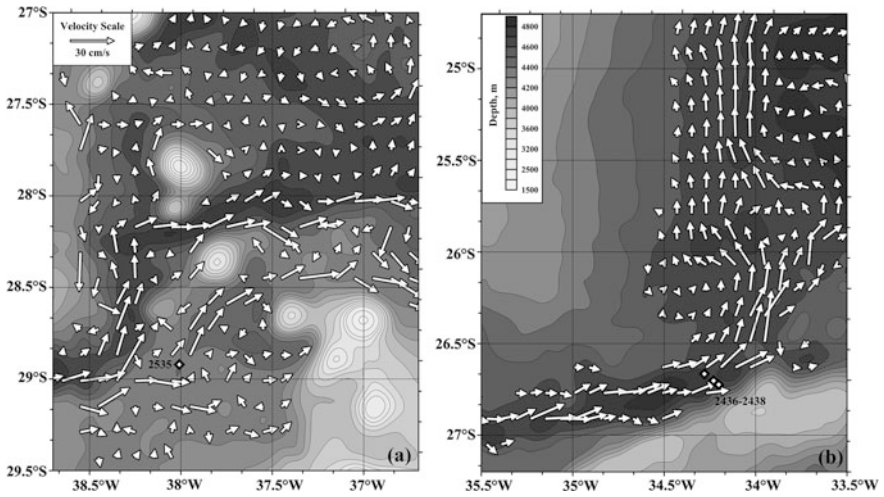


Fig. 3 Calculated velocities at a depth of 4300 m in the middle part of the channel (**a**) and in the region of its outflow to the Brazil Basin (**b**). White diamonds show our stations with current measurements. The station numbers are indicated

gradually slowing down. The minimum potential temperature observed at stations 2436–2438 was −0.08 °C. Hence, the difference between the minimum temperature at the beginning and at the end of the Vema Channel is 0.05 °C. In 2003, we performed CTD-measurements along the channel, and the corresponding difference was 0.04 °C. Therefore, the mean gradient of the bottom temperature in the channel is less than 0.01 °C/100 km.

Currents over the Santos Plateau

The Santos Plateau is located to the west of the Vema Channel at 28–32° S and 40–44° W. The depths of the major part of the plateau are approximately 4 km. According to the measurements, the potential temperature isotherm $\theta = 2$ °C is located at a depth of 3300–3400 m; hence, the Antarctic Bottom Water layer above the plateau is 600–700 m thick. Numerous small channels are located along the southern and eastern edges of the plateau. The mean depths of these channels relative to the surrounding plateau are 200–300 m.

In April–May and October–November 2016, we carried out two expeditions onboard the R/V "Akademik Sergey Vavilov" studying the bottom currents over the Santos Plateau. Twenty one stations with CTD and LADCP measurements were occupied in the small channels over the plateau. The locations of these stations, bottom potential temperatures, and maximum velocities in the bottom layer are shown in Fig. 4a. A section of seven stations was occupied across the zonally

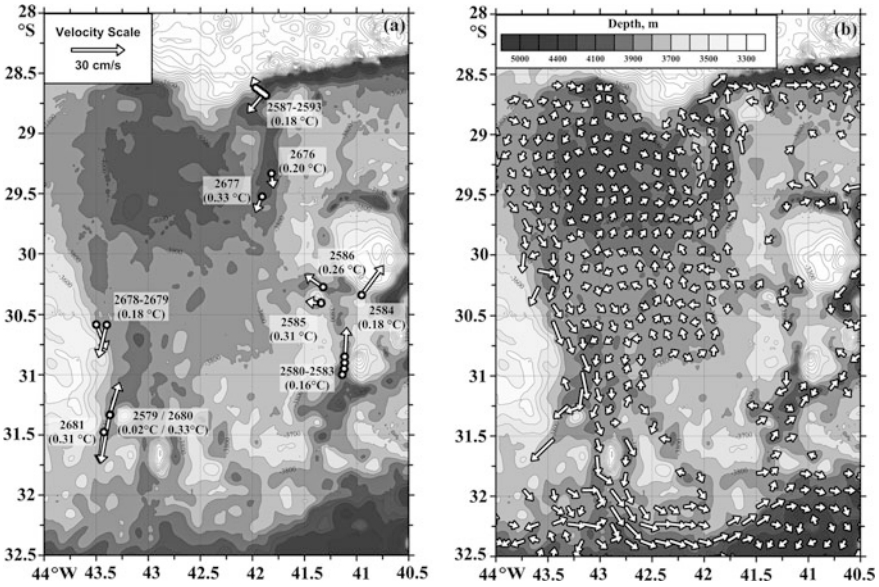


Fig. 4 Bottom currents over the Santos Plateau based on LADCP-observations (**a**) and numerical modeling (**b**). White circles denote our stations occupied in 2016. The numbers of stations and the lowest measured potential temperatures are shown in the left panel

oriented channel in the northwestern part of the plateau. The maximum depth of the channel is 4200 m, the potential temperature at this point was 0.18 °C. The main flow with velocities up to 20 cm/s was directed to the southwest, but a counter-current with easterly velocities of 10 cm/s was found near the northern slope of the channel. Potential temperatures at the bottom in the southeastern part of the plateau varied from 0.26 to 0.31 °C. An inflow with velocities of 5–10 cm/s through two small channels with the maximum depths of 3886 and 4026 m was found.

We simulated three-dimensional velocity fields using the INMOM model in the region from 37.0° to 22.8° S and from 45.2° to 29.0° W, which completely includes the Santos Plateau. Sigma-levels allow us to study bottom dynamics in arbitrary depth range. We specified 10 vertical levels in the bottom layer of the Vema Channel approximately from 4400 to 4850 m with a step of 50 m, while similar 10 levels over the plateau are at depths from 3630 to 4000 m with a step of 37 m (exact values depend on the ocean depth). The calculated velocity fields were interpolated to a depth of 3800 m for the analysis of the currents over the Santos Plateau (Fig. 4b). The velocities over the major part of the plateau are insignificantly low and do not exceed 5 cm/s. A strong southward current was found in the southwestern part of the plateau. The bottom waters propagate along the western slope of the plateau with velocities of 20 cm/s. We occupied four stations along this current in October 2016 at 30° 35′ S, 31° 20′ S, and 31° 29′ S. The presence of the southward current was confirmed by the direct LADCP measurements; the bottom

velocities at these stations varied from 15 to 30 cm/s. The bottom potential temperatures ranged from 0.18 to 0.33 °C. The current turns to the east at 32° 00′–32° 30′ S and flows along the northern boundary of the Argentine Basin. Part of these waters can flow back to the Vema Channel.

The measurements in the southwestern part of the plateau at 31° 20′ S, 43° 22′ W were performed twice in April and October 2016 (stations 2579 and 2680). Significant variations in the potential temperature, salinity, and velocities in the bottom layer were observed at this point (Fig. 5). The minimum potential temperature at the bottom in April was 0.02 °C; the velocities of the northerly current were up to 18 cm/s. In October, the current with the same velocities (15–20 cm/s) was directed to the south. The minimum recorded potential temperature was 0.33 °C, which was by 0.31 °C warmer than in April. The AABW upper boundary is located at a depth of 3320 m; the differences were observed only in the AABW layer (Fig. 5a). The properties of North Atlantic Deep Water layer above the AABW based on these two measurements were the same. The differences in temperature show that the inflow of cold Antarctic water occurs through the southwestern part of the plateau, where this water mixes with the relatively warm upper layers. When this water flows from the plateau to the south, its temperature is higher.

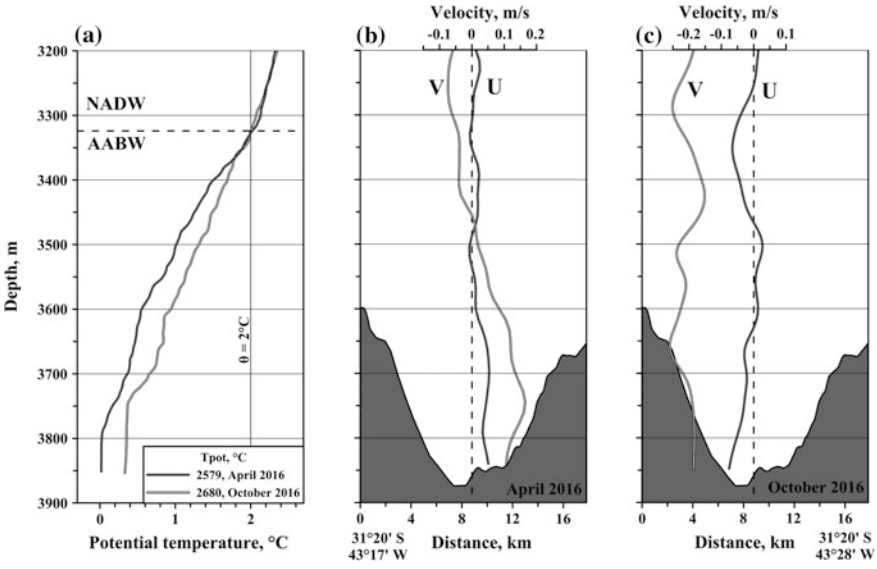


Fig. 5 Profiles of potential temperature and velocity at two stations at 31° 20′ S, 43° 22′ W measured in April (station 2579) and October (station 2680) 2016

Conclusions

We studied bottom water flows between the Argentine and Brazil basins on the basis of a three-dimensional regional model with high spatial resolution in the bottom layers and our field CTD and LADCP measurements in this region. From this point of view, we considered two main passages for AABW in this region: the deep and narrow Vema Channel and the relatively shallow and wide Santos Plateau. The main results of this work are as follows.

1. A steady northward bottom water flow is observed in the southern part of the Vema Channel north of 33° S. According to our measurements, the velocities were 10–15 cm/s at the beginning of the channel (32° 17′ S) and 25–30 cm/s in its middle part (31° 12′ S). The increase in the bottom temperature along the channel was of the order of 0.01°/100 km. The core of cold waters is displaced to the east at any section across the channel; the difference between the bottom temperatures near the eastern and western walls of the channel was 0.02 °C.
2. Numerical modeling revealed two branches of the current in the middle part of the channel. The major jet was located in the deepest part of the channel; the additional jet was found at depths of 4300–4400 on the southeastern slope. The lengths of the jets were 200 km and the maximum distance between them was 50 km. After the outflow from the Vema Channel, the bottom current turns to the north and propagates along the western boundary of the Brazil Basin.
3. A strong unsteady current was found in the southwestern part of the Santos Plateau. The bottom velocities here vary from 30 cm/s to the south up to 20 cm/s to the north. According to our CTD-measurements, the corresponding difference in the bottom temperatures was 0.31 °C. At 32°–32° 30′ S, the current turns to the east and propagates along the northern slope of the Argentine Basin.

Acknowledgments The work has been supported by the Russian Foundation for Basic Research grants 15-05-07539, 15-01-03942, 17-08-00085 (field studies); Russian Science Foundation project 16-17-10149 (data analysis).

References

1. Anisimov, M. V., & Diansky, N. A. (2008). Physical mechanism of the westward drift of the frontal current rings in the ocean. *Oceanology, 48*(3), 321–327.
2. Antonov, J. I., Seidov, D., Boyer, T. P., et al. (2010). World Ocean Atlas 2009 Vol. 2: Salinity. In S. Levitus (Ed.), *NOAA Atlas NESDIS 69* (p. 184). Washington, D.C.: U.S. Government Printing Office.
3. Danabasoglu, G., Yeager, S. G., Bailey, D., et al. (2014). North Atlantic simulations in coordinated ocean-ice reference experiments phase II (CORE-II) Part I: Mean states. *Ocean Modelling, 73,* 76–107.
4. Diansky, N. A., Bagno, A. V., & Zalesny, V. B. (2002). Sigma model of global ocean circulation and its sensitivity to variations in wind stress. *Izvestiya Atmospheric and Oceanic Physics, 38*(4), 537–556.

5. Egbert, G. D., & Erofeeva, S. (2002). Efficient inverse modeling of barotropic ocean tides. *Journal of Atmospheric and Oceanic Technology, 19,* 183–204.
6. Griffies, S., Gnanadesikan, A., Dixon, K. W., et al. (2005). Formulation of an ocean model for global climate simulations. *Ocean Science, 1,* 45–79.
7. Gamboa, L. A. P., Buffler, R. T., & Barker, P. F. (1983). Seismic stratigraphy and geologic history of the Rio Grande gap and Southern Brazil Basin. In *Initial Reports of the DSDP* (Vol. 72, pp. 481–498). US Government Printing Office.
8. Hogg, N., Siedler, G., & Zenk, W. (1999). Circulation and variability at the southern boundary of the Brazil Basin. *Journal of Physical Oceanography, 29,* 145–157.
9. Jungclaus, J., & Vanicek, M. (1999). Frictionally modified flow in a deep ocean channel: Application to the Vema Channel. *Journal Geophysical Research, 104*(C9), 21123–21136.
10. Klinck, J. M. (1995). Thermohaline structure of an eddy-resolving North-Atlantic model: The influence of boundary conditions. *Journal of Physical Oceanography, 25,* 1174–1196.
11. Large, W., & Yeager, S. (2009). The global climatology of an interannually varying air–sea flux data set. *Climate Dynamics, 33,* 341–364.
12. Locarnini, R. A., Mishonov, A. V., Antonov J. I., et al. (2010). World Ocean Atlas 2009. Vol. 1: Temperature. In S. Levitus (Ed.), *NOAA Atlas NESDIS 68*, (p. 182). Washington, D.C.: U.S. Government Printing Office.
13. McDonagh, E. L., Arhan, M., & Heywood, K. J. (2002). On the circulation of bottom water in the region of the Vema Channel. *Deep Sea Research, 49,* 1119–1139.
14. Morozov, E. G., Demidov, A. N., & Tarakanov, R Yu. (2008). Transport of Antarctic waters in the deep channels of the Atlantic Ocean. *Doklady Earth Sciences, 423*(8), 1286–1289.
15. Morozov, E. G., Demidov, A. N., Tarakanov, R. Y., & Zenk, W. (2010). *Abyssal channels in the Atlantic Ocean: Water structure and flows* (p. 266). Dordrecht: Springer.
16. Morozov, E. G., & Tarakanov, R. Y. (2014). The flow of Antarctic Bottom Water from the Vema Channel to the Brazil Basin. *Doklady Earth Sciences, 456*(1), 598–601.
17. Orsi, A. H., Johnsson, G. C., & Bullister, J. L. (1999). Circulation, mixing, and production of Antarctic Bottom Water. *Progress in Oceanography, 43,* 55–109.
18. Speer, K. G., & Zenk, W. (1993). The flow of Antarctic Bottom Water into the Brazil Basin. *Journal of Physical Oceanography, 23,* 2667–2682.
19. Tarakanov, R. Y., & Morozov, E. G. (2015). Flow of Antarctic Bottom Water at the output of the Vema Channel. *Oceanology, 55*(2), 153–161.
20. Visbeck, M. (2002). Deep velocity profiling using lowered acoustic doppler current profiler: bottom track and inverse solution. *Journal of Atmospheric and Oceanic Technology, 19,* 794–807.
21. Wüst, G. (1936). Schichtung und Zirkulation des Atlantischen Ozeans. In A. Defant (Ed.), Wissenschaftliche Ergebnisse, Deutsche Atlantische Expedition auf dem Forschungs und Vermessungsschiff "Meteor" 1925–1927 (Vol. 6, no. 1). Berlin: Walter de Gruyter & Co.
22. Zenk, W., Speer, K. G., & Hogg, N. G. (1993). Bathymetry at the Vema Sill. *Deep-Sea Research, 40,* 1925–1933.

Modeling Study of the Antarctic Circumpolar Current Variability Based on Argo Data

Konstantin V. Lebedev

Introduction

The Antarctic Circumpolar Current (ACC) is the largest ocean current with a mean transport reaching 173 Sv (1 Sv = 10^6 m^3 s^{-1}) according to recent research [5]. We also note the previous estimations of the mean ACC transport 134 ± 11 Sv in [24, 23]. There are also "instantaneous" estimates of the ACC transports in the Drake Passage based on the instrumental measurements over sections: 96, 152, and 184 Sv [3], 124 Sv [7], 147 Sv [10], 186 Sv [11], 127 Sv [12], and 156 Sv [19]. The ACC flows eastward around Antarctica connecting the southern parts of the Pacific, Atlantic, and Indian oceans. It is one of the major components of the global climate system. Accurate estimates of the ACC transport are very important to understand the Southern Ocean impact on the climate change. Due to the low amount of the oceanographic data and sustained observations in the Sothern Ocean, currently there is no consensus between the scientists about the amount of water transported by the current.

The development of the Argo project in the first years of the 21st century (the Argo array is part of the Global Climate Observing System/Global Ocean Observing System) designed to collect the data in the upper 2000 m of the ocean provides unique opportunities for continuous monitoring of the state of the Southern Ocean. Argo floats operate in the following manner. A typical float drifts for about 10 days at a given depth, then descends to a depth of 2000 m and rises to the surface measuring temperature and salinity. Battery power maintains the instrument for 3–4 years. In addition to temperature and salinity, such drifting floats also make it possible to obtain data on currents at the drift depth and at the ocean surface.

K. V. Lebedev (✉)
Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
e-mail: KLebedev@ocean.ru

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_30

In November 2007, the Argo program reached a coverage of 3000 simultaneously operating floats (one float in a 3° square) planned during the development of the program (measurements with the Argo floats cover the majority of the Global Ocean since 2005). Currently, 4000 Argo floats autonomously profile the upper 2000 m water column of the ocean from Antarctica to Spitsbergen with a 10-day interval. Constantly increasing number of measurements (World Ocean temperature and salinity database increases by 12,000 profiles every month) makes it possible to reconstruct and study the Southern Ocean dynamics and variability.

The Argo-Based Model for Investigation of the Global Ocean (AMIGO) was recently developed at the Shirshov Institute of Oceanology. It consists of a block for variational interpolation of Argo profiles to a regular grid and a block for model hydrodynamic adjustment of variationaly interpolated fields. It makes possible to obtain a full set of oceanographic characteristics: temperature, salinity, and current velocity from irregularly located Argo measurements [15]. The AMIGO technique was used for the calculation of the North Atlantic circulation [17] and estimates of the transports by oceanic currents [15, 21]. The analysis of model simulations and their comparison with the drifter data showed that irregularly distributed temperature and salinity data measured by the Argo floats variationaly interpolated on a regular grid restore realistic velocity fields [17, 21]; the AMIGO thermo-hydrodynamic information can be successfully used as the initial conditions in the ocean global circulation models [17].

In this paper we analyze the AMIGO simulations results in the Southern Ocean covering a 10-year period from 2005 to 2014. The spatial resolution of the data is 1° by latitude and longitude, and the temporal resolution is one month. The data are represented as monthly mean, seasonal, and annual means and climatological fields and are freely available at the Shirshov Institute web site: http://argo.ocean.ru/.

Data and Methods

The AMIGO dataset is used to study the variability of large-scale circulation in the Southern Ocean. The AMIGO is a new end-user oriented dataset compiled at the Shirshov Institute of Oceanology based on the Argo measurements [2] and satellite altimetry [6]. The most prominent features of this new method of Argo data processing are the application of variational interpolation of irregularly located Argo measurements to a regular grid [13, 16] followed by model hydrodynamic adjustment of the obtained fields with the concurrent use of satellite altimetry to verify solutions when selecting the model parameters [15].

The principle of the variational interpolation technique is to minimize the misfit between the interpolated fields defined on a regular grid and the irregularly distributed data, so that the optimal solution passes as close to the data as possible. The method of hydrodynamic adjustment [8, 9, 14] makes it possible to obtain hydrodynamically balanced oceanographic data using the ocean general circulation

model in the diagnostic and hydrodynamic adjustment modes. The new method of Argo data processing makes it possible to obtain a full set of oceanographic characteristics: temperature, salinity, and current velocity using irregularly located Argo measurements.

The simulations were performed for the entire globe limited in the north by 85.5° N using 1° grid spacing both by longitude and latitude. At the depths exceeding 2000 m, in which the Argo data are lacking, the temperature and salinity data were taken from the WOA09 database [1, 18]. The constant temperature and salinity values from the Argo data for the corresponding month (year, season) derived using the variational technique described above were specified as the boundary conditions at the ocean surface. The constant wind stress in the corresponding month (year, season) was specified from the ECMWF ERA-Interim reanalysis data [4].

Results

Maps of the mean salinity and currents of the Southern Ocean are shown in Figs. 1, 2 and 3 for the Atlantic (Fig. 1), Indian (Fig. 2), and Pacific (Fig. 3) sectors at a depth of 200 m. The ACC location can be traced along the salinity minimum. The Subantarctic Front (SAF), which is the northern border of the ACC, is clearly visible in the salinity and velocity fields. East of the Drake Passage, the front turns northward following the South American continental shelf, then crosses the Argentine Basin (60°–10° W) and the Mid-Atlantic Ridge (Fig. 1). Between 60° and 80° E, the SAF is constrained along the northern edge of the Kerguelen Plateau, then gradually turns southward along the Southeast Indian Ridge between 80° and 140° E (Fig. 2). Around 170° E (Campbell Plateau) the SAF turns sharply southward and northward again, follows the Pacific Antarctic Ridge (170°–230° E)

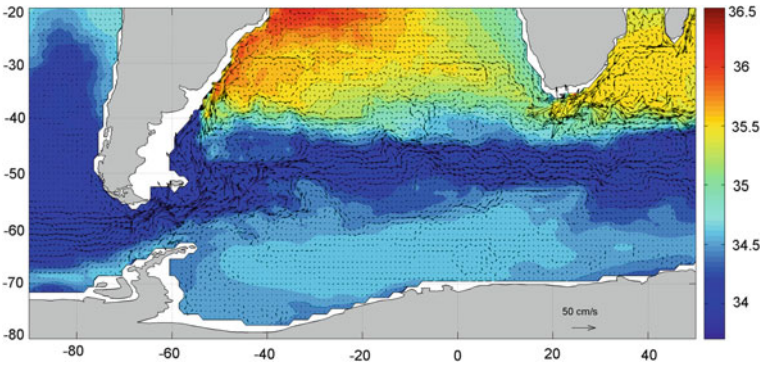


Fig. 1 Annual mean velocities and salinity field in the Atlantic sector of the Southern Ocean at 200 m. The salinity color scale is shown on the right

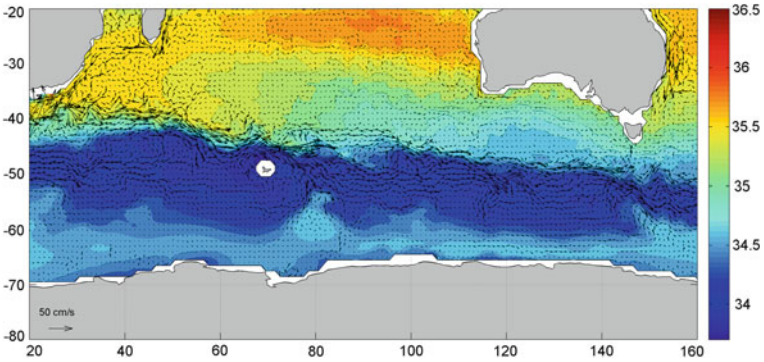


Fig. 2 Annual mean velocities and salinity field in the Indian sector of the Southern Ocean at 200 m. The salinity color scale is shown on the right

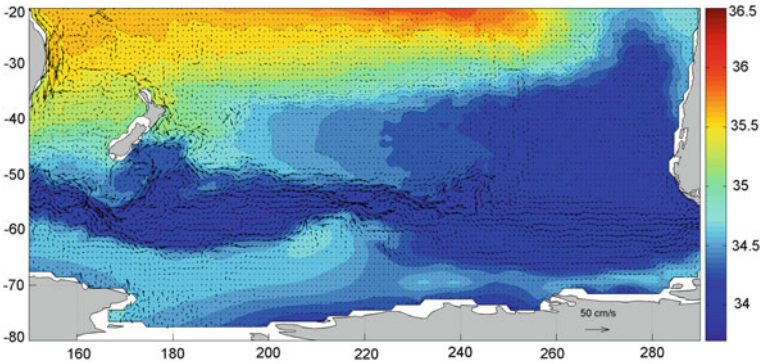


Fig. 3 Annual mean velocities and salinity field in the Pacific sector of the Southern Ocean at 200 m. The salinity color scale is shown on the right

and sharply turns northward at 240° E (Fig. 3). Between 240° and 300° E, the SAF crosses the Bellingshausen Basin and enters the Drake Passage.

In order to monitor the ACC variability we selected seven key sections, which separate the Southern Ocean into the Atlantic, Indian, and Pacific sectors: the Drake Passage at 66° W, African section at 21° E, West-Australian section at 117° E, East-Australian section at 149° E, and three zonal sections in the Atlantic, Indian, and Pacific oceans at 35° S.

The mean ACC transport over a period of 2005–2014 through the Drake Passage based on the AMIGO data is diagnosed as 162 ± 5 Sv. This value is between the recent estimates of the mean ACC transport in the Drake Passage (144 Sv according to Cunningham et al. [3]; 173 Sv according to Donohue et al. [5]). The mean heat transport calculated relative to 0 °C is 1.69 ± 0.07 PW with an amplitude of seasonal variations equal to 0.07 PW; the mean salt flux is 5.73 ± 0.17 Tg s^{-1}. The maximum value of mass transport was diagnosed in May 2006 as

173 Sv, the minimum value (149 Sv) was found in July 2007. The mean inflow of mass and salt through the Atlantic sector northern boundary are 0.6 ± 0.6 Sv and 0.02 ± 0.02 Tg s^{-1}, respectively. The mean heat flux through the Atlantic sector northern boundary (0.29 ± 0.19 PW) is directed to the Atlantic Ocean. The mean volume southward transport through the Atlantic Ocean is in a good agreement with the accepted climatology value of 0.83 Sv water transport from the Pacific to the Arctic Ocean based on the direct velocity measurements in the Bering Strait [20].

The mean mass, heat, and salt transports through the African section south of Cape Town were calculated as 162.6 ± 5 Sv, 1.19 ± 0.13 PW, and 5.75 ± 0.18 Tg s^{-1}, respectively. The mean inflow of mass, heat, and salt through the Indian sector northern boundary were 15.4 ± 3 Sv, 1.17 ± 0.28 PW, and 0.56 ± 0.11 Tg s^{-1}, respectively. The maximum value of mass transport was diagnosed in July 2010 as 24 Sv, the minimum value of 8 Sv was found in January 2014. The mean inflow from the Indian Ocean to the Southern Ocean is consistent with the multiyear mean value of 15 Sv transport through the Indonesian Throughflow (ITF) from the Pacific Ocean to the Indian Ocean determined from full-depth velocity measurements from January 2003 to December 2006 [22]. The mean transports through the West-Australian section were diagnosed as 178 ± 6 Sv, 2.04 ± 0.12 PW, and 6.30 ± 0.22 Tg s^{-1}, with well pronounced seasonal variations with the amplitudes of 5 Sv, 0.1 PW, and 0.2 Tg s^{-1}, respectively. The maximum value of mass transport was diagnosed in July 2010 as 194 Sv, the minimum value of 165 Sv was found in January 2010.

The mean mass, heat, and salt transports through the East-Australian section are diagnosed as 178 ± 6 Sv, 2.55 ± 0.13 PW, and 6.30 ± 0.22 Tg s^{-1}, respectively. The mean outflow through the Pacific sector northern boundary is 16.0 ± 3 Sv, with the maximum value of 24.8 Sv in July 2010 and the minimum value of 8.4 Sv in January 2014. The mean heat and salt fluxes through the northern boundary are 0.45 ± 0.39 PW and 0.56 ± 0.12 Tg s^{-1}, respectively. Seasonal variations are very strong with the boreal summer maxima and boreal winter minima. The flow through the Pacific sector northern boundary could be presented as the inflow between Australia and New Zealand (9.6 ± 5 Sv, 0.30 ± 0.22 PW, and 0.34 ± 0.18 Tg s^{-1}) and outflow between New Zealand and South America (25.6 ± 4 Sv, 0.75 ± 0.27 PW, and 0.90 ± 0.15 Tg s^{-1}). As a result, the mean transport through the section between New Zealand and Antarctica at 168° E increases by 9.6 Sv and is diagnosed as 187.6 ± 7 Sv, the values of heat and salt fluxes through this section are calculated as 2.55 ± 0.12 PW and 6.64 ± 0.24 Tg s^{-1}, respectively.

The Southern Ocean heat budget is negative except for the Australian region. The heat losses were diagnosed as 0.2 PW in the Atlantic sector, 0.3 PW in the Indian sector, and 0.4 PW in the Pacific sector. The net Southern Ocean heat loss is 0.4 PW due to 0.5 PW heat gain in the region south of Australia.

Conclusions

The mean transports estimated here based on the AMIGO dataset appeared close to those reported in [20, 22] from the computations based on direct velocity measurements. Our findings can be viewed as the further quantification of the ACC characteristics obtained using an alternative approach based on the newly developed method utilizing constantly increasing data flow from autonomous Argo profiling floats. The major results are as follows.

1. The mean ACC transport over a period of 2005–2014 through the Drake Passage based on the AMIGO data is diagnosed as 162 ± 5 Sv. The transport through the African section is calculated as 162.6 ± 5 Sv. The net mean flow of 178 ± 6 Sv is diagnosed between Australia and Antarctica; the mean transport through the section between New Zealand and Antarctica is found as 187.6 ± 7 Sv.
2. The maximum values of transports are diagnosed as 173 Sv in the Drake Passage (May 2006), 194 Sv through the section between Australia and Antarctica (July 2010), and 24.8 Sv from the Southern ocean to the Pacific Ocean (July 2010).
3. The minimum values of transports are found as 149 Sv in the Drake Passage (July 2007), 165 Sv through the section between Australia and Antarctica (January 2010), and 8.4 Sv from the Southern Ocean to the Pacific Ocean (January 2014).

Acknowledgements This work was supported by the Russian Science Foundation (project no. 16-17-10149).

References

1. Antonov, J. I., Seidov, D., Boyer, T. P., et al. (2010). World Ocean Atlas 2009, Vol. 2: Salinity. In S. Levitus (Ed.), *NOAA Atlas NESDIS 69 Ser.* Washington, D.C.: US Government Printing Office.
2. Argo. (2000). Argo float data and metadata from Global Data Assembly Center (Argo GDAC). SEANOE. http://doi.org/10.17882/42182.
3. Cunningham, S. A., Alderson, S. G., King, B. A., & Brandon, M. A. (2003). Transport and variability of the Antarctic Circumpolar Current in Drake Passage. *Journal Geophysical Research, 108*(C5), 8084. https://doi.org/10.1029/2001JC001147.
4. Dee, D. P., Uppala, S. M., Simmons, A. J., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society, 137,* 553–597.
5. Donohue, K. A., Tracey, K. L., Watts, D. R., et al. (2016). Mean Antarctic Circumpolar Current transport measured in Drake Passage. *Geophysical Reseach Letters, 43,* 11760–11767. https://doi.org/10.1002/2016GL070319.

6. Ducet, N., Le Traon, P. Y., & Reverdin, G. (2000). Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2. *Journal Geophysical Research, 105* (C8), 19477–19498.
7. Gladyshev, S. V., Koshlyakov, M. N., & Tarakanov, R. Yu. (2008). Currents in the Drake Passage based on observations in 2007. *Oceanology, 48*(6), 759–770.
8. Ivanov, Yu. A, Lebedev, K. V., & Sarkisyan, A. S. (1997). Generalized hydrodynamic adjustment method (GHDAM). *Izvestiya Atmospheric and Oceanic Physics, 33*(6), 752–757.
9. Ivanov, Yu. A, & Lebedev, K. V. (2000). Integral average monthly characteristics of the World Ocean climate. *Izvestiya Atmospheric and Oceanic Physics, 36*(2), 244–252.
10. Koshlyakov, M. N., Gladyshev, S. V., Tarakanov, R. Yu., & Fedorov, D. A. (2011). Currents in the Western Drake Passage according to the observations in January of 2010. *Oceanology, 51*(2), 187–198.
11. Koshlyakov, M. N., Gladyshev, S. V., Tarakanov, R. Yu., & Fedorov, D. A. (2012). Currents in the Drake Passage based on the observations in November of 2010. *Oceanology, 52*(3), 299–308.
12. Koshlyakov, M. N., Gladyshev, S. V., Tarakanov, R. Yu., & Fedorov, D. A. (2013). Currents in the Drake Passage by the observations in October–November of 2011. *Oceanology, 53*(1), 1–12.
13. Kurnosova, M. O., & Lebedev, K. V. (2014). Study of transport variations in the Kuroshio extension system at 35° N, 147° E based on the data of Argo floats and satellite altimetry. *Doklady Earth Science, 458*(1), 1154–1157.
14. Lebedev, K. V. (1999). Average annual climate of the ocean. Part 2: Integral characteristics of the World Ocean climate (mass, heat, and salt transports). *Izvestiya Atmospheric and Oceanic Physics, 35*(1), 87–96.
15. Lebedev, K. V. (2016). An argo-based model for investigation of the Global Ocean (AMIGO). *Oceanology, 56*(2), 172–181.
16. Lebedev, K. V., DeCarlo, S., Hacker, P. W., et al. (2010). Argo Products at the Asia-Pacific Data-Research Center. *Eos Trans. AGU*, *91*(26). Ocean Science Meeting Suppl., Abstract IT25A-01.
17. Lebedev, K. V., Sarkisyan, A. S., & Nikitin, O. P. (2016). Comparative analysis of the North Atlantic surface circulation reproduced by three different methods. *Izvestiya Atmospheric and Oceanic Physics, 52*(4), 410–417.
18. Locarnini, R. A., Mishonov, A. V., Antonov, J. I., et al. (2010). World Ocean Atlas 2009, Vol. 1: Temperature. In S. Levitus (Ed.), *NOAA Atlas NESDIS 68 Ser*. Washington, D.C.: US Government Printing Office.
19. Morozov, E. G., Tarakanov, R. Yu., Ansorge, I., & Swart, S. (2014). Jets and Transport of the Antarctic Circumpolar Current in the Drake Passage. *Fundamentalnaya i Prikladnaya Gidrofizika, 7*(3), 23–28.
20. Roach, A. T., Aagaard, K., Pease, C. H., et al. (1995). Direct measurements of transport and water properties through the Bering Strait. *Journal Geophysical Research, 100*(C9), 18443–18457.
21. Sarkisyan, A. S., Nikitin, O. P., & Lebedev, K. V. (2016). Physical characteristics of the Gulf Stream as an indicator of the quality of large-scale circulation modeling. *Doklady Earth Science, 471*(2), 1288–1291.
22. Sprintall, J., Wijffels, S. E., Molcard, R., & Jaya, I. (2009). Direct estimates of the Indonesian through flow entering the Indian Ocean: 2004–2006. *Journal Geophysical Research, 114,* C07001. https://doi.org/10.1029/2008JC005257.
23. Whitworth, T. (1983). Monitoring the Transport of the Antarctic Circumpolar Current at Drake Passage. *Journal of Physical Oceanography*, *13*(11), 2045–2057
24. Whitworth, T., & Peterson, R. G. (1985). Volume transport of the Antarctic circumpolar current from bottom pressure measurements. *Journal of Physical Oceanography, 15*(6), 810–816.

Tareev Equatorial Undercurrent in the Indian Ocean

Albert K. Ambrosimov, Dmitry I. Frey and Sergey M. Shapovalov

Introduction

In February 2017, in cruise 42 of the R/V "Akademik Boris Petrov", measurements of the currents at two meridional sections in the equatorial zone of the Indian Ocean were performed using a lowered ADCP instrument. The aim of the work was to study the distribution of velocities in the upper 700-m layer of the ocean, including the subsurface Tareev undercurrent, named after the Soviet oceanographer Boris Tareev [1]. For the first time this current was discovered in 1960 during the First International Indian Ocean Expedition during cruise 31 of the R/V "Vityaz". Here we cannot agree with Schott and McCreary [7], who noted that this undercurrent was first documented by Taft [9]. We would like to send the mentioned authors to the publication of Ovchinnikov [4], which describes the discovery of this equatorial undercurrent.

Subsequently, the study of the Tareev current in the system of the equatorial currents of the Indian Ocean was carried out in numerous expeditions. A rather complete summary of the results collected during 30 years is presented in the monograph by Neiman et al. [3]. We also note two later reviews [7, 8].

The previous studies show that one of the determining factors in the formation of the circulation in the upper layer in the equatorial zone of the Indian Ocean is the periodic change in the monsoon winds that affect the currents of this region. Relatively stable southeastern trade winds are observed south of 10° S; north of this latitude there is a strong variability of winds associated with the monsoon periodicity. As distinguished from the tropical parts of the Atlantic and Pacific Oceans, there is a deficiency of stable eastern winds along the equator during the year. Equatorial winds have an eastern component only at the end of winter/early spring.

A. K. Ambrosimov (✉) · D. I. Frey · S. M. Shapovalov
Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
e-mail: ambrosimov@ocean.ru

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_31

This is the cause of another phenomenon in the Indian Ocean: the eastern equatorial undercurrent (Tareev Current) exists only during a part of the year, usually in February–June, when the winds have an eastern component, and only when anomalous east winds arise this undercurrent may appear in the other seasons.

According to the observations made over the cross-equatorial sections in the region of 55° E in 1975 and 1976 [2], the existence of an equatorial counterflow (Tareev Current) was confirmed during the boreal spring. Significant meandering of the flow with a deviation of more than 100 km to the south from the equator has been recorded.

The measurements made with the ADCP over several equatorial sections in the 50–60° E longitude band [6] and moored ADCP in the equatorial zone [5] also revealed significant temporal and spatial variability of the equatorial undercurrent.

The Data and Measurements

In this contribution, we analyze the results of measurements over two meridional sections using two DVS-750 Doppler current meters. All measurements were carried out in the lowering and lifting mode from the ship maintaining stable position using the GPS control system. Taking into account the specific properties of these devices, the measurements were carried out according to the following procedure. Both meters were installed on the SBE-32 water sampling system, one with down-looking sonic transmitter and another with the up-looking one. The time sampling of the current speed measurement was 1 s, the depth of lowering was 700 m. The SBE-32 system was stopped for one minute at specific depths (700, 500, 300, 250, 200, 150, 100, and 50 m) to accumulate statistically valid signal. This allowed us to collect the data from each device with a high resolution in the range of up to 5 m. The measurements at the 68° E section were carried out on February 15–18, 2017, from 3° N to 3° S. Thirty three stations were occupied here. Thirty two additional stations more were occupied over the section along 65° E on February 18–23, 2017, from 3° S to 3.75° N. The distances between the stations were 10–15 miles. The stations were occupied under the northeastern wind with a wind speed not exceeding 12 m/s.

Thus, 65 vertical profiles of the current speed to a depth of 700 m were collected at the sections. The data of measurements for the analysis in this paper were used up to a depth of 300 m.

Analysis

According to the data of measurements, two sections of the distribution of the zonal component of velocity were plotted. They are shown in the Fig. 1.

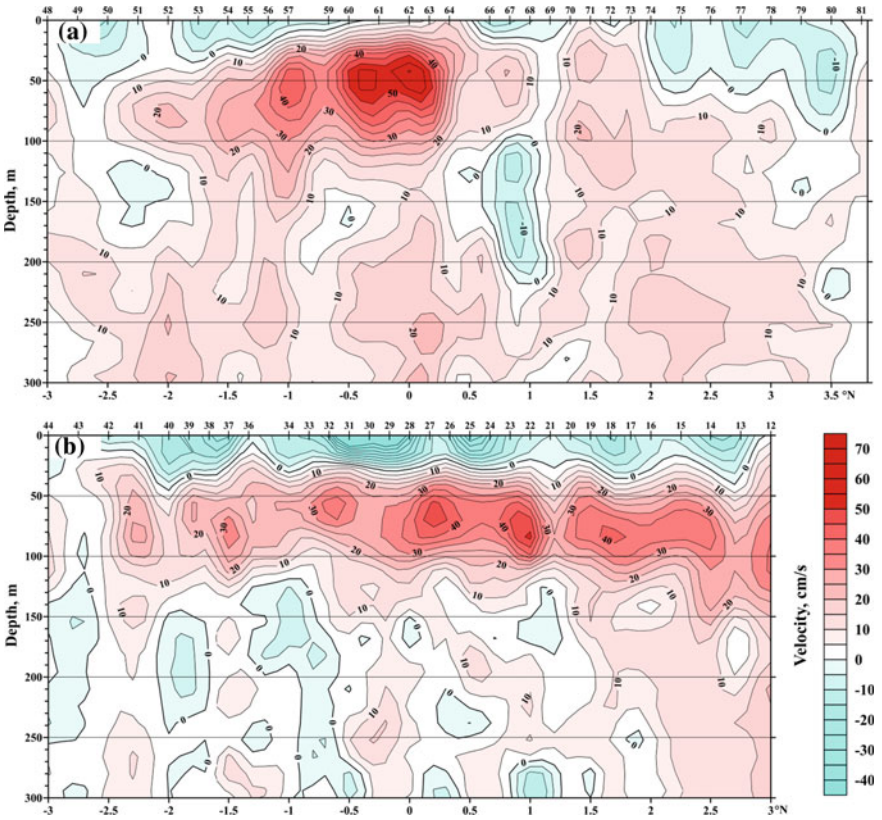


Fig. 1. Zonal current components of velocity in the upper 300 m over the cross equator sections at 65° E (**a**) and 68° E (**b**), measured by two up-looking and down-looking DVS-750 showing the eastward Equatorial Undercurrent from February 15 to 23, 2017. The red tones are related to the eastern direction

Figure 1 shows that the uppermost layer is occupied by a western direction stream, which is more developed at 68° E and noticeably weaker at 65° E. At the same time, in the northern and southern parts of the 65° E section this current penetrates to a depth of 30 m (in the south) and up to 80 m in the north and almost completely disappears at the equator. This current fairly stably occupies the entire upper 30-m layer along the 68° E section. The maximum speeds here exceed 40 cm/s.

Under the westward stream, the equatorial eastward countercurrent (Tareev Current) is clearly pronounced. Here we see the opposite situation. At 65° E, the core of the current in the form of a powerful jet of the eastern direction is located between 2.2° S and 1° N with the maximum speeds almost at the equator. In the region of the core, the current occupies the depths from the surface to 150 m. The velocities in the current exceed 60 cm/s at a depth of about 50 m. The width of the stream in the meridional direction is about 350 km. In the northern part of the

section, the velocities of the eastern jets are reduced to 10–20 cm/s. Note that the eastward flow extends over this section to a maximum observation depth of 700 m with velocities up to 30 cm/s.

At 68° E, the Tareev Current is slightly weaker and deeper. The core of the current disintegrates into separate jets, the depth of which ranges from 50 to 100 m, and the speed in the jets becomes lower than 50 cm/s. In general, there is a deviation of the core and individual stream jets to the north. In the southern part of the section, the Tareev Current begins to weaken south of 2° S, where the velocity falls to 20 cm/s. Nevertheless, the current can be seen up to the end of the section. At the northern stations of the section, the speed of the undercurrent is still quite high, thus, we can assume that it extends noticeably north of 3° N. The width of the current in the meridional direction over this section is more than 700 km. Below, the jets of the stream of the eastern direction are weaker and interspersed by the counter flows, the number of which considerably increases below 300 m (not shown in the figure).

Thus, we can formulate the following conclusions:

(1) A strong surface eastward current with velocities exceeding 40 cm/s was observed in the equatorial region of the Indian Ocean in the winter season of the Northern Hemisphere in February 2017. During the entire observation period, the northeastern winds with speeds of 5–12 m/s dominated.
(2) A thin surface westward current was observed over the meridional section along 68° E from 3° N to 3° S at all stations at depths from 0 to 50 m. Below them, the eastward equatorial undercurrent (Tareev Current) was recorded at depths of 50–150 m with velocities ranging from 20 to 50 cm/s. At the southern end of the section, the currents were much weaker.
(3) The Tareev Current is clearly observed over the 65° E meridional section from 2.5° S to 1° N. The undercurrent core is located in the equatorial region within one degree in the meridional span. At the ends of the section the flow weakens and it is masked by the vortex perturbations.
(4) The width of the Tareev equatorial current, according to our measurements, was more than 700 km over the section along 68° E and about 350 km over the 65° E section.

Thus, we can conclude that in February 2017 subsurface equatorial undercurrent exists during the time of the development of the western surface current generated by the northeastern summer monsoon in the equatorial zone of the Indian Ocean. The development of this flow spans from east to west and, ultimately, transforms to the flow of the eastern direction, existing during the winter monsoon as a subsurface undercurrent.

Acknowledgements The R/V "Akademik Boris Petrov" expedition was supported by the Federal Agency for Scientific Organizations of Russia and with the assistance of the Agency staff. The analysis was supported by the Russian Science Foundation, grant 14-50-00095. We thank all the participants of this expedition, including the scientific team and crew of the ship, who facilitated the scientific observations.

References

1. Koshlyakov, M. N., Morozov, E. G., & Neiman, V. G. (2016). Historical findings of the Russian physical oceanographers in the Indian Ocean. *Geoscience Letters, 3,* 19. https://doi.org/10.1186/s40562-016-0051-6.
2. Leetmaa, A., & Stommel, H. (1980). Equatorial current observations in the western Indian Ocean in 1975 and 1978. *Journal of Physical Oceanography, 10,* 258–269.
3. Neiman, V. G., Burkov, V. A., & Scherbinin, A. D. (1997). Dynamics of the Indian Ocean (232 pp.). Moscow: Scientific World (in Russian).
4. Ovchinnikov, I. M. (1961). Circulation in the northern part of the Indian Ocean in the period of the winter monsoon. In: *Oceanological Researches no. 4* (pp. 18–34). Leningrad, Hydrometeoizdat (in Russian).
5. Reppin, J., Schott, F. A., Fischer, J., & Quadfasel, D. (1999). Equatorial currents and transports in the upper central Indian Ocean: Annual cycle and interannual variability. *Journal Geophysical Research, 104,* 15495–15514.
6. Schott, F., Fischer, J., Garternicht, U., & Quadfasel, D. (1997). Summer monsoon response of the Northern Somali Current, 1995. *Geophysical Research Letters, 24,* 2565–2568.
7. Schott, F. A., & McCreary, J. P. (2001). The monsoon circulation of the Indian Ocean. *Progress in Oceanography, 51,* 1–123.
8. Schott, F. A., Xie, S. P., & McCreary, J. P. (2009). Indian Ocean circulation and climate variability. *Reviews of Geophysics*, *47*, RG1002. https://doi.org/10.1029/2007rg000245.
9. Taft, B. A. (1967). Equatorial undercurrent in the Indian Ocean, 1963. *Studies in Tropical Oceanography, 40,* 3–14.

The Bering Sea Regional Data Assimilation System: From Climate Variability to Short Term Hindcasting

Gleb G. Panteleev, Max Yaremchuk, Vladimir Luchin and Oceana Francis

Introduction

In the last decades, 4d variational (4dVar) data assimilation (DA) has become the most acknowledged tool for advanced hindcasting and forecasting of the ocean circulation. Starting from DA systems constrained by simple dynamics (e.g., [1, 16]), the 4dVar approach [9] has gradually evolved into advanced DA systems based on state-of-the-art ocean models (MIT, ROMS) and is now routinely applied toward reconstructing the circulation in the Arctic, Pacific, or World Ocean (e.g., [5, 15]). These state-of-the-art DA systems involve the assimilation of a variety of observations from satellites, surface and Argo drifters and climatological observations from oceanographic databases. Assimilation of these publicly accessible, near real-time observations is natural since the major goal of the "global" DA systems is to provide hindcasting of the global circulation and/or analysis of observed climate changes on a global scale.

Meanwhile, due to the low flexibility of the global DA systems in assimilating regional data on smaller scales, the regional DA systems have been under extensive development within the last decade (e.g. [6, 8]). These DA systems routinely assimilate alongtrack satellite altimetry (Jason-1, Jason-2, CryoSat), hourly SSTs from GOES, and surface currents from HFRs and provide regional updates and 3-day forecasts of the ocean state. It is necessary to note, that all the regional

G. G. Panteleev (✉) · M. Yaremchuk
Naval Research Laboratory, Stennis Space Center, Hancock, USA
e-mail: gleb.panteleev@nrlssc.navy.mil

V. Luchin
Pacific Oceanological Institute, Vladivostok, Russia
e-mail: vlad_luchin@mail.ru

O. Francis
University of Hawaii, Honolulu, USA
e-mail: oceanaf@hawaii.edu

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_32

systems were developed in the domains where TS observations were routinely collected year-round for a long period of time. This allows for reliable estimation of the background state, which is one of the key components of any DA system (e.g. [32]). There are, however, many regions where the temporal and spatial distribution of the observations is irregular and/or sparse in time and space. A prominent example is the Bering/Chukchi Sea basin which becomes increasingly important due to the rapid growth of the sea traffic and emerging prospects for development of oil and mining industries on the shelves.

In this contribution, we discuss the key components of the regional Bering Sea DA system developed by the authors in recent years [21, 22, 25] and present results of the analysis of the Bering Sea circulation on different scales.

The Regional 4dVar Data Assimilation System

The dynamic components, which synthesize diverse oceanographic data within a single framework are the Semi-Implicit Ocean Model (SIOM) and its adjoint configured to optimize the SIOM trajectory with respect to observations. The SIOM is a modification of the C-grid, z-coordinate Ocean Global Circulation Model (OGCM) developed at the Laboratoire d'Oceanographie Dynamique et de Climatologie [12]. The model was specifically designed for operating within the regional 4dVar framework controlled by fluxes at the open model boundaries and sea surface. The model is semi-implicit both in barotropic and baroclinic modes, permitting simulations with relatively large time steps of approximately 0.1 day [17, 20–22]. The tangent linear model was obtained by direct differentiation of the forward model code. The adjoint code of the model was built analytically by transposition of the operator of the tangent linear model, linearized in the vicinity of the given solution of the forward model [32]. The spatial structure of the background error covariances is modeled by the quadratic polynomials of the Laplacian operator [34, 35].

The SIOM 4dVar DA system supports assimilation of diverse oceanographic data such as: (a) In situ, *Temperature and salinity* observations from the database of the Far Eastern Branch of the Russian Academy of Science; (b) *Historical velocity observations* from moorings and sea surface drifters, and satellite-tracked drifter trajectories (Fig. 1b); (c) *Estimates of the mean transport through the Bering Strait.* Mean climatological estimates of 0.9 ± 0.2 Sv and corresponding seasonal transport are taken from [31]; (d) *Atmospheric forcing data* (momentum, heat and salt fluxes) for assimilation are taken from the National Center for Environmental Prediction (NCEP); (e) *The SSH anomalies* from the AVISO database at http://www.aviso.altimetry.fr. All observations are supplied with corresponding spatially varying covariances.

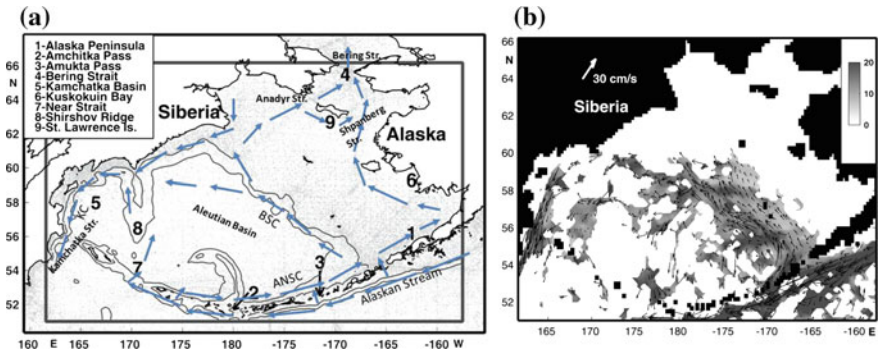


Fig. 1 **a** Spatial distribution of the historical temperature data in the Bering Sea. The model domain is shown by the rectangle. Contour lines mark the coastline and 1000 and 3000 m isobaths. Arrows show the schematic of the Bering Sea circulation according to Stabeno et al. [28] including the Bering Slope Current (BSC), the Aleutian North Slope Current (ANSC), and the Kamchatka Current (KC). **b** The multiyear mean surface drifter velocities averaged over the model grid cells. Gray shading shows the rms variance (cm/s) of the drifter velocity in each grid cell

Climatological Seasonal States of the Bering Sea

In oceanography, the volume of the in situ observations is usually insufficient to control all the model degrees of freedom, resulting in an ill-conditioned inverse problem. A reliable climatological evolution of the sea state is one of the best options for specifying the background state needed for regularization of the ill-conditioned inverse problem. Conventionally, the gridded climatological TS distributions are derived from historical observations using an optimal interpolation (OI) technique (e.g. [11]), which does not take into account the dynamical constraints and often leads to oversmoothing. On the contrary, the 4dVar DA obtains climatological states through the quasi-stationary variational DA approach [30]. Dynamical constraints of the model naturally introduce inhomogeneous advective smoothing in the regions of strong currents. This process allows for the derivation of a high resolution climatological state where all variables are dynamically balanced, which is not the case for "conventional" oceanographic Atlases derived through optimal interpolation algorithms.

This 4dVar approach was applied for the Bering Sea (BS) with a goal of obtaining the climatological annual mean and four seasonal states. The model was configured for the domain shown in Fig. 1a in a non-eddy-resolving mode on a relatively coarse (~18 km) regular z-coordinate grid with meridional and zonal resolutions of 0.16° and 0.3°, respectively, and a time step of 4 h. Vertically, the grid has 34 levels with unequal spacing ranging from 5 m near the surface to 500 m in the deeper layers. All data sets, outlined in section "The Regional 4dVar Data Assimilation System", were used to constrain the model solution. As a result, the annual mean climatological and four seasonal states were obtained. They are available at http://people.iarc.uaf.edu/~gleb/nprb_aleutian_passes/bering_sea_atlas_register.php.

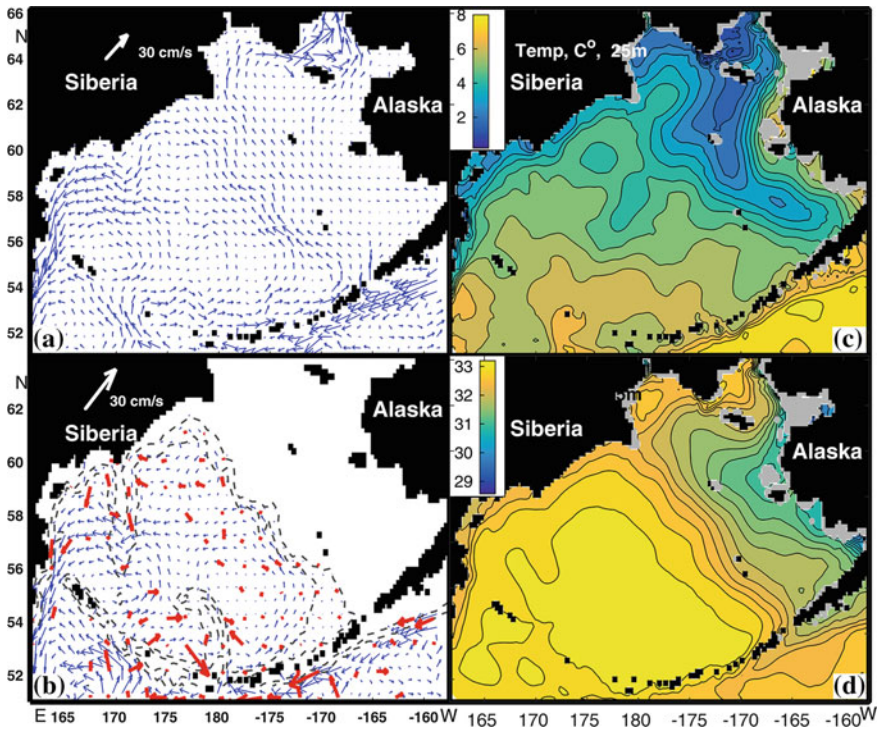


Fig. 2 Optimized maps of velocities at 25 m (**a**) and 1000 m (**b**) and temperature (**c**) and salinity (**d**) in summer. Red arrows denote mean velocities at 1000 m estimated from Argo floats parked at 1000 m [33]

Figure 2a–d presents the reconstructed mean summer state of the Bering Sea. The structure of the reconstructed circulation (Fig. 2a, b) is in good agreement with the schematic of the Bering Sea circulation provided by Stabeno et al. [28]. The circulation pattern reveals an intense (30–50 cm/s) Alaska Stream, a somewhat weaker (10–20 cm/s) Aleutian North Slope Current flowing along the southern and northern flanks of the Aleutian Arc and a cyclonic circulation in the deep part of the Bering Sea that includes a relatively weak (5–15 cm/s) Bering Slope Current and more pronounced (30–40 cm/s) Kamchatka Current along the eastern and western Bering Sea shelves, respectively. According to Fig. 2a, b, a significant portion of the inflow through the Near Strait forms a cyclonic gyre in the south-western part of the Bering Sea and then merges with the Kamchatka Current. The other portion of the Kamchatka Strait outflow comes from the Bering Slope Current that originates from multiple inflows through the eastern Aleutian Passes. The Alaskan Stream is the most intense current in the region. It flows along the southern flank of the Aleutian Arc and feeds the flows through the Aleutian Passes [29]. The Bering Slope Current splits into two branches near the point 57° N, 180° E. This is in a good qualitative agreement with the flow pattern described by Stabeno et al. [28],

but additionally, provides a statistically reliable quantification of the transport. The mean relative mismatch between the reconstructed surface velocities (Fig. 2a) and assimilated drifter velocities is 0.67.

Figure 2b provides a comparison of the optimized velocity field at 1000 m with independent (not assimilated) velocity data derived from the Argo floats. The average speed of the Argo floats is 4.6 cm/s, which is close to the mean optimized velocity amplitude of 3.7 cm/s. This 25% difference can be considered to be in reasonable agreement with the observations since a significant fraction of the ARGO drifters are involved in transient eddy motions resulting in higher Lagrangian mean velocities compared to the Eulerian mean speed estimates.

Figure 2c, d shows the summer TS distributions in the BS. In comparison with the optimal interpolation results (e.g., https://www.nodc.noaa.gov/OC5/PACIFIC2009/), the reconstructed fields are less smoothed, especially in the Northern Bering Sea (NBS) and along the Aleutian Arc. These are signatures of the strong topographically controlled flows through the Bering and Anadyr Straits and through the island system of the Aleutian Arc. The most important advantage of these TS distributions is that they are dynamically balanced with corresponding velocity fields (Fig. 2a, b), which allows for identifying finer structures of regional circulation while simultaneously providing estimates of climatological mass, heat and salt transports.

Except for the summer season, Pacific water masses entering the Bering Sea within the upper subsurface layer are characterized by higher temperature and salinity than the water mass residing in the Bering Sea at the same depth (Fig. 2c, d). The major cause of such temperature/salinity signal is due to the relatively high heat and salt content of the Pacific Water compared to the Bering Sea Water, as well as intense vertical and lateral mixing in the Aleutian Passes.

In the summer season, the warmest and least salty subsurface layer waters are formed in the nearshore regions of Bristol Bay, Norton Sound, and the Gulf of Karaginski. The presence of warm and relatively fresh salty waters in these areas is a consequence of strong stratification observed after the melting of the sea-ice cover and relatively weak winds, which are not able to mix the upper warm layer with the waters underneath. A somewhat colder and saltier upper subsurface layer water is present in the summer in several regions of the Bering Sea (e.g. in the central and eastern parts of the Aleutian Arc, and in the Chirikov Basin, Fig. 2c, d). Due to intense tidal and non-periodical currents in these regions, surface waters are mixed with the underlying cold waters formed in the winter and fall.

The optimized estimates of the mean volume transport through the major Aleutian Passes are found to be 2.5–7 times larger than those in the dynamical method [28]. This discrepancy is likely due to significant underestimation of the barotropic velocity component in the Aleutian Straits by the dynamical method. The latter suggestion is well supported by recent velocity observations in several Aleutian Passes, which reveal significant northward flow at a 100–200 m depth [29], indicating the importance of the regional barotropic flow in the overall Bering Sea volume balance.

Another important advantage of the 4dVar DA in the reconstructed climatological state of the Bering Sea is obtaining the dynamically balanced Sea Surface Height (SSH), which can formally be treated as MDOT, or the "*difference between a time-averaged SSH and geoid*". The importance of the MDOT availability is closely connected to the possibility of retrieving absolute SSH from the satellite altimetry anomalies observed by various satellite platforms (e.g. TOPEX/Poseidon, Jason-1/2, Earth Resources Satellite (ERS)-1/2, Envisat, Saral, etc.). Therefore, a number of research groups have developed methods to combine various data with altimetry to obtain more accurate estimates of the global MDOT [13, 18, 26]. For various reasons, all of these products have substantial deficiencies in the Bering Sea.

Figure 3 compares 4dVar MDOT and three other MDOT products in the region. Since the optimized surface velocities below the Ekman layer are in approximate geostrophic balance with the MDOT (optimized SSH field), the reconstructed MDOT contours (Fig. 3a) are conveniently interpreted as streamlines of the mean geostrophic currents at the surface. The circulation pattern reveals the following major structures: (a) an intense (30–40 cm/s) Alaskan Stream south of the Alaska Peninsula; (b) a somewhat weaker (10–20 cm/s) Aleutian North Slope Current embracing the southern and northern flanks of the Aleutian Arc; (c) the 30–40 cm/s

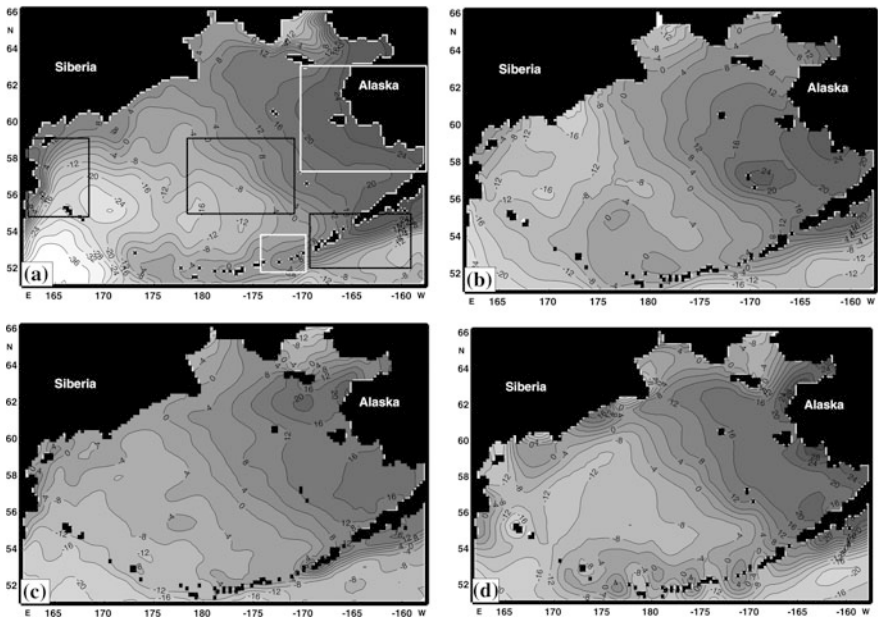


Fig. 3 The mean dynamic topographies of the Bering Sea **a** this study, **b** by Rio et al. [26], **c** by Rio et al. [27], and **d** by merging the EGM08 geoid model with altimeter data. Black rectangles in Fig. 3a show the drifter validation regions. White rectangles show the domain of high resolution SIOM DAS for the Amukta Pass and part of the Eastern Bering Sea shelf discussed below

strong Kamchatka Current to the west; and (d) a relatively weak (5–15 cm/s) cyclonic circulation occupying the deep part of the Bering Sea (Figs. 2a and 3a). According to Fig. 3a, a significant portion of this cyclonic gyre originates in the Near Strait with the rest coming from the inflow through other Aleutian passages. The circulation shown in Fig. 3a is in good qualitative agreement with the results of Stabeno et al. [29] who describes a gradual leakage of the Alaskan Stream though the passages in the Aleutian Arc.

The mean relative mismatch between the reconstructed surface velocities and assimilated drifter velocities (Fig. 1a) is 8.2 cm/s. It is unlikely that a better agreement could be obtained between the climatological velocity with mean amplitude of approximately 10 cm/s and the highly variable surface currents derived from the drifter trajectories affected by eddies and small-scale variations of the wind stress.

The derived MDOT were validated using the standard AVISO methodology: three regions well-covered by drifters were selected, gridded AVISO SSH anomalies from www.aviso.oceanobs.com were added to each MDOT, and the resulting geostrophic currents were compared with the currents deduced from drifter trajectories for the three regions shown by black rectangles in Fig. 3. The validation has shown that 4Dvar MDOT significantly (10–35%) outperforms all other MDOT products [22].

Obtaining a high resolution regional MDOT and its posterior errors is a prominent feature delivered by the 4dVar analysis of the climatological observations. In the next sections, we show how the derived MDOT benefits the analyses of decadal variability and hindcasts of the BS circulation at interannual time scales.

4dVar Analysis of the Circulation in the Aleutian Passes

This section provides another illustration that a realistic MDOT is a key component of a regional data assimilation system. In application to the regions with complicated topography, such as Aleutian Passes, MDOT inaccuracies may result in significant underestimation of the volume transports. To illustrate this, the 4dVar DA system was configured at a 6 km resolution with a domain around the Amukta Pass shown by the smaller white rectangle in Fig. 3a. A larger rectangle shows the domain of a similar experiment conducted in the central part of the Eastern Bering Sea shelf (see [23]).

Amukta Pass is one of the most important pathways for the inflow of the Pacific Water into the Bering Sea. Two assimilation experiments have been conducted in the domain: In the first experiment, the circulation for the period between January 10 and February 5, 2002 was reconstructed by assimilating AVISO sea level anomalies referenced to the 4dVar MDOT (Fig. 3a) in combination with the NCEP sea surface heat, salt and momentum fluxes. Climatological (winter) temperature/salinity in the Amukta Pass distributions were utilized as a first guess. The second

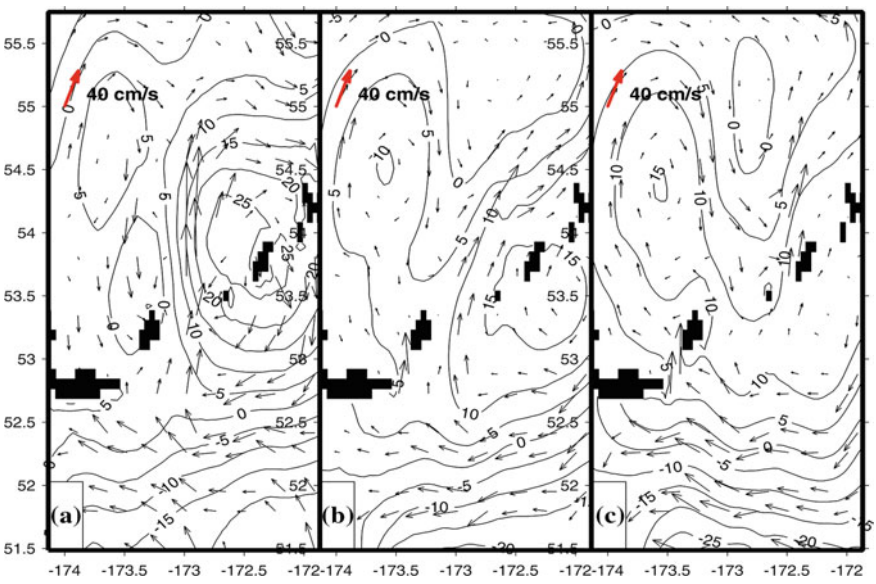


Fig. 4 Circulation in Amukta Pass on January 15 (**a**) and February 1 (**b**), 2002, derived by assimilation of the AVISO sea level anomaly data referenced to the MDOT shown in Fig. 3a. Panel (**c**) shows the circulation in the Amukta Pass on February 1, 2002, derived by assimilation of the AVISO sea level anomaly referenced to Fig. 3c (RIO05)

experiment had an identical configuration except that the 4Dvar MDOT was replaced by the RIO05 MDOT (Fig. 3c).

The mean circulation patterns on January 10–20 and January 25–February 5 derived in the first experiment are shown in Fig. 4a, b. The Amukta Pass northward transports in these two periods were 6.1 Sv and 3.5 Sv, respectively. A very intense northward current on January 10–20 and January 25-February 5 agrees well with the large positive anomalies of 7 Sv and 4 Sv in the Amukta transport observed by Stabeno et al. [29], who used records from four bottom-mounted ADCPs deployed in the region for 2 years [29]. Note that resolution of the gridded climatological MDOT (~18 km) and AVISO sea level anomaly (1/3° × 1/3°) are not sufficient enough to resolve the complicated geometry of the Amukta Pass. Therefore, better agreement with observation may be expected, if higher resolution climatological MDOT and/or AVISO products were available.

On the contrary, Fig. 4c demonstrates a much weaker circulation around the Amukta Pass which was derived by assimilating AVISO sea level anomalies referenced to the RIO05 MDOT shown in Fig. 3c. The respective transport averaged over the period of January 25-February 5, 2002 is close to 1.6 Sv, which is 2.5 times smaller than 4 Sv observed by Stabeno et al. [29]. These experiments indicate that RIO05 significantly underestimates the MDOT gradient in the region of the

Aleutian Passes. In the context of the Bering Sea large-scale dynamics, accurate MDOT appears to be especially important since the processes of the ventilation of the deep and shallow parts of the BS occurs through the few relatively narrow Aleutian Passes. These ventilation processes are one of the key mechanisms controlling the BS ecosystem and, therefore, should be reconstructed with increased accuracy.

In relation to ecosystem dynamics, the vast shelf of the Eastern Bering Sea (EBS) represents the major hub located on the crossroads of the pathways between the Pacific and Arctic Oceans. Extreme complexity of the advective and mixing processes in the shallow waters of EBS shape the seasonal ecosystem parameters and require accurate treatment in terms of observational and dynamical consistency.

Eastern Bering Sea: Regional Ocean in Motion

The Eastern Bering Sea (EBS) plays an important role in United States' fisheries, dynamics of the immense populations of marine birds and mammals, and subsistence activities of native communities. Therefore, this region has been permanently studied in the last decades. The most comprehensive studies were conducted in 2007–2010 in the framework of the Bering Sea Integrated Ecosystem Research Program (BSIERP). Typically, 2–3 hydrophysical (CTD) surveys per year were conducted and 3 mooring arrays were supported during 2007–2010 (Fig. 5). Note, however, that each CTD survey usually lasts for 2–3 weeks, providing only a "snapshot" of a limited area of the EBS, while during the rest of the year, there is no reliable information about the hydrophysical state. At the same time, moorings provide continuous time series in a few locations lacking adequate spatial coverage. Obviously, these data cannot adequately describe the circulation in the entire EBS, if considered, without additional dynamical constraints. The 4dVar DA provides a good guide on how to combine observations from different sources for reconstructing the non-stationary state of the ocean in motion.

Nested 4dVar system. The BSIERP observations and other sources of data provide an excellent opportunity to improve our understanding of changes in the physical forcing of the Bering ecosystem in its response to the climate change. But due to ice coverage in the winter, a relatively large regions of the EBS including the key passages through the Anadyr and Bering Straits were not covered by these studies. In view of the necessity to adequately resolve the Bering and Anadyr Straits, a nested SIOM 4dVar DA system was developed. The nested DA system utilized as a first guess, the momentum, heat and salt fluxes from Bering Ecosystem Study Ice Ocean Modeling and Assimilation System (BESTMAS) [10, 36].

The nested models were configured in the domains shown in Fig. 5 with a resolution of 7 km for the fine model grid, ω, embedded into a coarser (15–20 km) grid, Ω (Fig. 1). In the vertical, the coarse (fine) resolution model had 35(15) unevenly spaced levels with 5 (2.5) m near-surface spacing, which increased to 500(25) m at

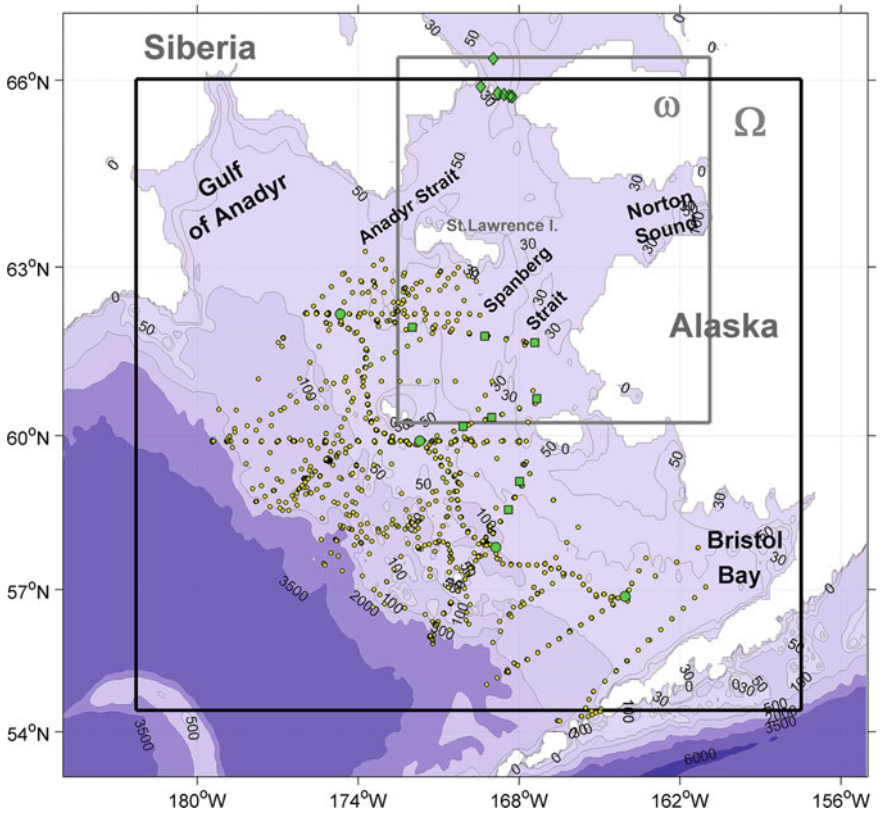


Fig. 5 Model domain and assimilated observations. The boundaries of the coarse (17 km) and fine (7 km) resolution grids are shown by the black and gray rectangles, respectively. Small circles show hydrographic stations, and mooring locations from different observational experiments are shown as white circles, squares, and diamonds

the bottom. Trajectories of both models were controlled by the initial/boundary conditions and surface fluxes. The details and validation of the novel two-way nested DA iterative algorithm were discussed in Panteleev et al. [25]. It was shown that the nested DA approach allows for a much better reconstruction of the flow through the Bering Strait and a significant reduction of the model-data observation misfits in comparison with the BESTMAS data assimilative solution. In the entire range of the velocity observations from moorings, 4dVar velocities demonstrate a much better linear fit to the data compared to the one characterizing the BESTMAS solution [25].

The two-way nested 4dVar scheme allowed us to achieve much better consistency between the optimal solutions within the nested domains. As a result, a significantly better fit was obtained to SSH data, constraining circulation in a coarse resolution domain, and to higher resolution velocity data controlling the Bering Strait throughflow.

The major sources of data needed for model forcing and assimilation are shown in Fig. 5 and include velocity observations from three mooring arrays; multiple CTD surveys. In addition, we assimilate the gridded SSH anomalies from the AVISO data set referenced to the regional MDOT and background information from the reconstructed seasonal climatologies of the Bering Sea (Section "The Regional 4dVar Data Assimilation System").

The optimized evolution of the EBS state was obtained by assimilating the above described data over 49 4-week time windows spanning the period of January 1, 2007 to October 3, 2010. On the average, $300 \times 4 = 1200$ iterations were required to achieve the nested 4dVar convergence of a single 4-week problem.

Mass, heat and freshwater transports. The 4dVar-optimized weekly averaged fluxes are shown in Fig. 6 in the Bering and Anadyr Straits. Because mass is conserved, and the surface heat/freshwater fluxes in the Northern Bering Sea can be neglected, the difference between the curves gives an insight on the partitioning of the respective Bering Strait transports between the Anadyr and Spanberg inflows. The heat and freshwater transports in Fig. 10 were obtained relative to −1.9 °C and 34.8 psu, respectively.

An interesting feature in Fig. 6 are the events when the Bering Strait fluxes were much larger than the advective heat/freshwater transports entering the Chirikov Basin through the Anadyr Strait (e.g., September, November 2007, August–September 2010). Near-surface circulation and wind stress patterns (Fig. 7a, c)

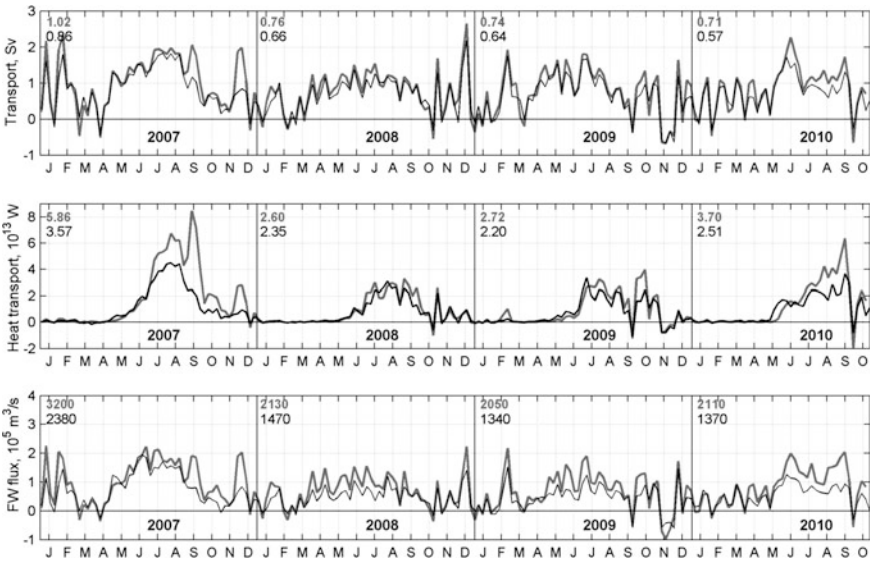


Fig. 6 Evolution of the weekly averaged mass, heat, and freshwater fluxes through the Bering (thick gray line) and Anadyr Straits (solid black line). Annual mean values of the fluxes are shown in the upper right corners of the respective boxes. The units are, respectively, 10^6 m^3/s, 10^{20} J, and 10^3 m^3/s

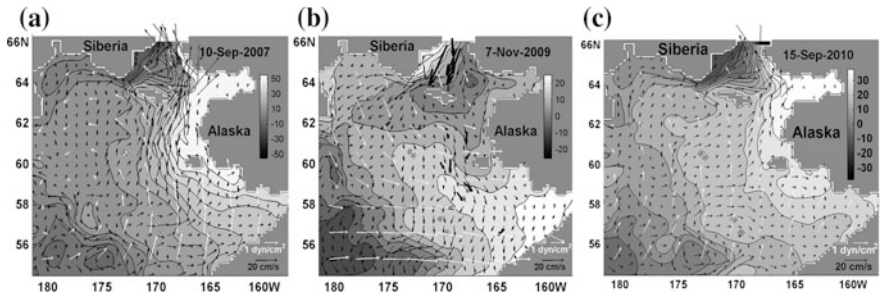


Fig. 7 Velocity at 25 m depth (black arrows), wind stress (white arrows), and SSH (shading, cm), at 0.00 UTC on September 10, 2007, November 7, 2009, and September 2, 2010. Thick arrows in the middle of the plot show velocities observed on the moorings

indicate that these anomalies were caused by strong southeasterly winds over the EBS shelf, which raise the sea level along the Alaskan coast and enhance transport of the warm and fresh waters through the Spanberg Strait.

Wind stress patterns in Fig. 7a, c indicate that intensification of the northward current along the Alaskan coast was forced by atmospheric lows over the central Bering Sea. Such cyclones are quite a common phenomena in late fall/winter and are known to cause intensification of the Kamchatka Current [7]. However, since EBS is mostly ice covered in winter, these winds have little effect on the heat transport, but may cause a noticeable increase in the freshwater transport (lower panels in Fig. 6). In summer, strong southeasterly winds associated with cyclones over the BS basin are less common, although a certain increase in the summer storm activity has been documented [14] over the last decade. These storms may contribute significantly to the observed decrease of the ice cover in the Chukchi Sea. Winter storms have a larger impact on the deep mixing in the central BS and freshwater flux through the Bering Strait.

Another interesting phenomena in Fig. 6 are the anomalous (southward) transport events through the Bering Strait. Most of them occur in fall-winter and have a typical duration of 1–2 weeks. The exception is a 28-day southward transport observed in November 2009. Comparison with the wind forcing has shown that all the negative transport anomalies in September–November 2008–2010 are associated with the southward winds. This phenomenon has been documented by Coachman [4] and appears to be a typical feature for the freezing and (to a lesser extent) melting seasons. It is also remarkable that in 2007–2010, the reversing of the Bering Strait throughflow mostly occurred synchronously with the Anadyr Current, when the contribution of the Spanberg Strait volume transport becomes negligible. Analysis of the optimized circulation shows that forcing by the northeasterly winds completely blocks the Spanberg Strait throughflow causing dramatic structural changes in the SSH and circulation patterns. Northwesterly wind surge tends to change the sign of the large-scale SSH gradient in the southern EBS (cf. Fig. 7a–c) and completely reverse the flow field in the Bering Strait along the Alaskan and Siberian coastlines. Prolonged flow reversals, such as the one observed

in November 2009, may have a significant impact on the EBS ecosystem by affecting the residence times and displacing the cold pool (e.g. [3]) boundary which is advected by the general southward flow clearly seen in Fig. 7b on a larger scale.

Water pathways and residence times. Analysis of trajectories of the Lagrangian particles is a convenient approach to study the regional ecosystems (e.g. [2, 19]). It could be also useful in the analysis of water mass pathways and their transformation rates (e.g. [22]). In the analysis described below, we focus on two important sub-regions: the Unimak Strait (the easternmost passage from the open Pacific into the EBS) and the BSC source region properly located north of Unalaska Island. Lagrangian particles were released in each region once a week during 3.5 years. Trajectories of the fastest and slowest particles from these releases are shown in Fig. 8a. It is noteworthy that the largest residence times τ (up to 40 months) were observed exclusively for the particles released in the Unimak Pass while the smallest residence times (8–10 months) were documented for the particles released north of Unalaska Island, where the North Alaskan Slope Current meets the EBS continental slope. The residence time distributions in both regions (Fig. 8b) are quite different in shape. The Unalaska release region is characterized by a double peak structure. A more detailed analysis shows that the fastest particles, corresponding to the first peak at $\tau = 10$ months were mostly released in the fall season (October-November), when the BSC is spun up by the storm activity [24]. The majority of Unalaska released particles reach the Arctic Ocean in 11–15 months as they travel with the BSC of normal strength. The near-exponential tail of the distribution ($\tau \sim$ 16–35 months) is formed by the particles carried by mesoscale eddies across the continental slope into the EBS shelf.

In contrast to the Unalaska particles, particles released at Unimak Pass are characterized by much more homogeneous distribution over the residence times (thin line in Fig. 8b). This could be explained by the low probability of their capture by the core of the BSC. As a consequence, most of these particles start their travel

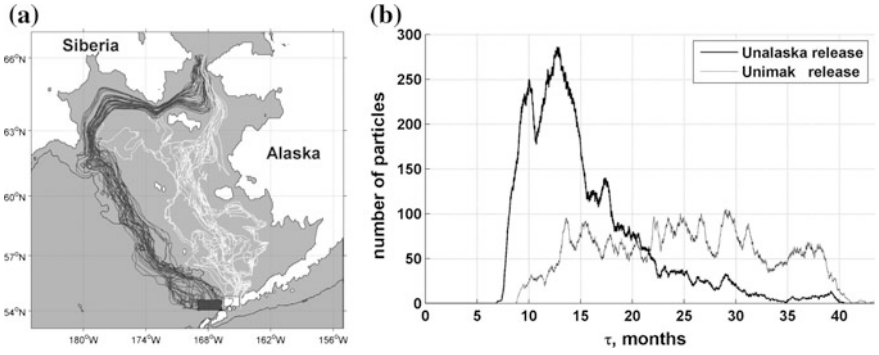


Fig. 8 **a** Trajectories of the particles reaching the Bering Strait and released in the Unimak Pass (white rectangle) and north of the Unalaska Island (dark gray). **b** EBS residence time distribution of the particles released in the Unimak Pass (thin line) and north of Unalaska Island

history in the EBS shelf environment which is characterized by much weaker and less organized currents. Trajectories of the Unimak Pass particles with the largest residence times ($\tau \sim 35$–40 months) shown in Fig. 8b first wander around the southern EBS shelf west of the Pribilof Islands, and then slowly proceed north following the Spanberg Strait pathway to the Arctic Ocean. Only a small fraction of these particles join the Anadyr Current far downstream, and they are never engaged in the eddy-induced exchange with the BSC. The same property characterizes the fastest particles populating the left peak ($\tau < 11$ months) of the Unalaska distribution in Fig. 8b. These particles strictly follow the main stream of the BSC and Anadyr Current.

Summary

We presented a systematic approach for developing a regional DAS as applied to the Bering Sea. The method is based on applying the 4dVar technique to both climatological and instantaneous data with the ultimate goal of reconstructing dynamically and statistically consistent states at various temporal and spatial scales: from large-scale interannual circulation to eddy resolving transient current systems. The backbone of the system are dynamically balanced seasonal climatological states and the corresponding MDOT of the Bering Sea, which distinguishes the developed Bering Sea DAS from other regional systems developed in the last decade, such as for the Oregon Coast by Kurapov et al. [8], and for the Hawaiian Islands by Janekovich et al. [6].

The developed seasonal climatologies of the Bering Sea include dynamically consistent temperature, salinity, velocity, and sea surface height fields for each season (winter, fall, spring and summer) and the corresponding fields for the climatological mean state of the Bering Sea. These climatologies have sufficiently fine resolution to allow accurate specification of the dynamically balanced background state for any subregion of the Bering Sea. An important by-product is the high-resolution regional MDOT, whose accuracy and quality significantly outperforms the other MDOTs derived from various global reanalyses.

Combining the developed Bering Sea MDOT with instantaneous satellite altimetry data provided an important observational constraint on the regional dynamics in the entire Bering Sea. The 4dVar assimilation of satellite altimetry anomalies referenced to the 4Dvar MDOT provides much better estimates of the transport through the Aleutian Passes. These flows control the supply of nutrients to the Bering Sea, affect the stability of the Aleutian North Slope Current and the process of eddy formation along the Aleutian Arc, which is an important component of local ecosystem dynamics.

The seasonal climatologies were also used to constrain a more detailed and comprehensive study of the EBS circulation in 2007–2010 by means of the nested 4dVar DAS developed by Panteleev et al. [25]. Comparison with the BESTMASS DA system of Lindsay and Zhang [10] has shown that the new DAS

provides a more realistic EBS circulation and Bering Strait throughflow both in terms of model-data misfits and general dynamical consistency. This allows for a more accurate analysis of the property exchange processes between the Pacific Ocean and the Bering Sea, including residence times, pathways and the transformation rates of various water masses on their way to the Arctic Ocean.

The wealth of new information on the Bering Sea dynamics delivered by the developed DAS illustrates the importance of regional DA systems for further improvement in understanding and forecasting environmental conditions in the marginal seas and nearshore areas. Our experience shows that 4dVar reconstruction of the dynamically consistent regional climatologies is of particular importance for the reanalysis of the local data sets. In turn, the improved accuracy of the hindcasts would inevitably result in better forecasting of the regional environmental conditions that are extremely important in many applications.

Acknowledgements This study was supported by the University of Hawaii, the International Arctic Research Center, NSF grants 1107925, 1203740 and ARC-1107327. G. Panteleev and M. Yaremchuk were also supported by the ONR core project "Arctic data assimilation" and program element 0602435N as part of the project "Adjoint-free 4dVar for Navy ocean models". The authors are indebted to P. Stabeno for providing drifter and current data.

References

1. Brasseur, P. (1991). A variational method for the reconstruction of general circulation fields in the Northern Bering Sea. *Journal Geophysical Research, 96,* 4891–4907.
2. Carlotti, F., & Wolf, K.-U. (1998). A Lagrangian ensemble model of Calanus finmarchicus coupled with a 1-D ecosystem model. *Fisheries Oceanography, 7*(3/4), 191–204.
3. Cianelli, L., & Bailey, K. M. (2005). Landscape dynamics and resulting species interactions: The cod-capelin system in the southeastern Bering sea. *Marine Ecology Progress Series, 291,* 227–236.
4. Coachman, A. K. (1993). On the flow field in the Chirikov Basin. *Continental Shelf Research, 13,* 481–492.
5. Forget, G., Campin, J.-M., Heimbach, P., Hill, C. N., Ponte, R. M., & Wunsch, C. (2015). ECCO version 4: An integrated framework for non-linear inverse modeling and global ocean state estimation. *Geoscientific Model Development Discussions, 8,* 3653–3743. https://doi.org/10.5194/gmdd-8-3653-2015.
6. Janekovich, I., Powell, B. S., Matthews, D., McManus, M. A., & Sevadjian, J. (2013). 4D-Var data assimilation in a nested, coastal ocean model: A Hawaiian case study. *Journal Geophysical Research, 118,* 1–14.
7. Hughes, F. W., Coachman, L. K., 7 Aagard, K. (1974). Circulation transport and water exchange in the western Bering Sea. In D. W. Hood and E. J. Kelly (Eds.), *Oceanography of the Bering Sea with emphasis on the renewable resources* (pp. 59–98). Institute of Marine Science, University of Alaska, Fairbanks.
8. Kurapov, A. L., Foley, D., Strub, P. T., Egbert, G. D., & Allen, J. S. (2011). Variational assimilation of satellite observations in a coastal ocean model off Oregon. *Journal Geophysical Research, 116,* C05006. https://doi.org/10.1029/2010JC006909.
9. Le Dimet, F. X., & Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus Series A, 38,* 97–100.

10. Lindsay, R. W., & Zhang, J. (2006). Assimilation of ice concentration in an ice-ocean model. *Journal of Atmospheric and Oceanic Technology, 23,* 742–749.
11. Luchin, V., Kruts, A., Sokolov, O., Rostov, V., Rudykh, N., Perunova, T., et al. (2009) Climatic Atlas of the North Pacific Seas 2009: Bering Sea, Sea of Okhotsk, and Sea of Japan. In V. Akulichev, Yu. Volkov, V. Sapozhnikov and S Levitus (Eds.), *NOAA Atlas NESDIS 67* (380 pp.). U.S. Government Printing Office, Washington, D.C., CD Disc.
12. Madec, G., Delecluse, P., Imbard, M., et al. (1999) OPA 8.1 Ocean General Circulation Model. Reference Manual, Note du Pole de Modelisation (Institut Pierre-Simon Laplace (IPSL), France.
13. Maximenko, N., Niiler, P., Rio, M.-H., Melnichenko, O., Centurioni, L., Chambers, D., et al. (2009). Mean dynamic topography of the ocean derived from satellite and drifting buoy data by three different techniques. *Journal of Atmospheric and Oceanic Technology, 26*(9), 1910–1919.
14. Mesquita, M. S., Atkinson, D. E., & Hodges, K. I. (2010). Characteristics and variability of the storm tracks in the North Pacific, Bering Sea and Alaska. *Journal of Climate, 23,* 294–311. https://doi.org/10.1175/2009/JCLI3019.1.
15. Moore, A. M., Arango, H. G., Broquet, G., Edwards, C., Veneziani, M., Powell, B., et al. (2011). The regional ocean modeling system (ROMS) 4-dimensional variational data assimilation systems Part II. Performance and application to the california current system. *Progress in Oceanography, 91*(1), 50–73.
16. Nechaev, D. A., & Yaremchuk, M. I. (1994). Conductivity-temperature-depth data assimilation into a three-dimensional quasigeostrophic open ocean model. *Dynamics of Atmospheres and Oceans, 21,* 137–165.
17. Nechaev, D., Panteleev, G., & Yaremchuk, M. (2005). Reconstruction of the circulation in the limited region with open boundaries: Circulation in the Tsushima Strait. *Okeanologiya, 45*(6), 805–828.
18. Niiler, P., Maximenko, N., & McWilliams, J. (2003). Dynamically balanced absolute sea level of the global ocean from near-surface velocity observations. *Geophysical Research Letters, 30* (22), 2164. https://doi.org/10.1029/2003/GL018628.
19. Panteleev, G. G., deYoung, B., Reiss, C., & Taggart, C. (2004). Passive tracer reconstruction as a least squares problem with a semi-lagrangian constraint: An application to fish eggs and larvae. *Journal of Marine Research, 62,* 787–878.
20. Panteleev, G., Nechaev, D., & Ikeda, M. (2006). Reconstruction of summer Barents Sea circulation from climatological data. *Atmosphere-Ocean, 44*(2), 111–132.
21. Panteleev, G. G., Proshutinsky, A., Kulakov, M., Nechaev, D. A., & Maslowski, W. (2007). Investigation of summer Kara Sea circulation employing a variational data assimilation technique. *Journal of Geophysical Research*, *112*, C04S15. https://doi.org/10.1029/2006jc003728.
22. Panteleev, G., Proshutinsky, A., Nechaev, D., Woodgate, R., & Zhang, J. (2010). Reconstruction and analysis of the Chukchi Sea circulation in 1990–1991. *Journal of Geophysical Research, 115,* C08023. https://doi.org/10.1029/2009JC005453.
23. Panteleev, G., Yaremchuk, M., Stabeno, P., Luchin, V., Nechaev, D., & Kukuchi, T. (2011). Dynamic topography of the Bering Sea. *Journal Geophysical Research, 116,* C05017. https://doi.org/10.1029/2010JC006354.
24. Panteleev, G., Yaremchuk, M., Nechaev, D., & Kikuchi, T. (2012). Variability of the Bering Sea circulation in 1992–2010. *Journal of Oceanography, 68,* 485–496. https://doi.org/10.1007/s10872-012-0113-0.
25. Panteleev, G., Yaremchuk, M., Francis, O., Stabeno, P. J., Weingartner, T., & Zjang, J. (2016). An inverse modeling study of circulation in the Eastern Bering Sea during 2007–2010. *Journal of Geophysical Research*, *121*, 3970–3989. https://doi.org/10.1002/2015jc011287.
26. Rio, M.-H., Schaeffer, P., Lemoine, J.-M., & Hernandez, F. (2005). Estimation of the ocean mean dynamic topography through the combination of altimetric data, in-situ measurements

and GRACE geoid: From global to regional studies. In *Proceedings of the GOCINA International Workshop*.
27. Rio, M.-H., Schaeffer, P., Moreaux, G., Bourgogne, S., Lemoine, J.-M., & Bronner, E. (2009). A new mean dynamic topography computed over the global ocean from GRACE data altimetry and in-situ measurements. In *Proceedings of OceanObs'09 Conference*, Venice, Italy.
28. Stabeno, P. J., Schumacher, J. D., & Ohtani, K. (1999). The physical oceanography of the Bering Sea. In *Dynamic of the Bering Sea* (838 pp.). Fairbanks: Alaska Sea Grant College Program.
29. Stabeno, P. J., Kachel, D. G., & Sullivan, M. E. (2005). Observation from moorings in the Aleutian Passes: Temperature, salinity and transport. *Fisheries Oceanography, 14,* 39–54.
30. Tziperman, E., & Thacker, W. C. (1989). An optimal-control/adjoint equation approach to studying the oceanic general circulation. *Journal of Physical Oceanography, 19,* 1471–1485.
31. Woodgate, R. A., Aagaard, K., & Weingartner, T. (2005). Monthly temperature, salinity, and transport variability of the Bering Strait through flow. *Geophysical Research Letters, 32,* L04601. https://doi.org/10.1029/2004GL021880.
32. Wunsch, C. (1996). *The ocean circulation inverse problem* (p. 442). Cambridge: Cambridge University Press.
33. Yoshinari, H., Maximenko, N. A., & Hacker, P. W. (2006). YoMaHa'05: Velocity data assessed from trajectories of Argo floats at parking level and at the sea surface. IPRC Technical Note No. 4, 20 pp.
34. Yaremchuk, M., & Sentchev, A. (2012). Multi-scale correlation functions associated with polynomials of the diffusion operator. *Quarterly Journal of the Royal Meteorological Society, 138,* 1948–1953.
35. Yaremchuk, M., Carrier, M., Smith, S., & Jacobs, G. (2013). Background error correlation modeling with diffusion operators, In S. K. Park and L. Xu (Eds.), *Data assimilation for atmospheric, oceanic and hydrologic applications* (Vol. II, pp. 177–203). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-35088-78.
36. Zhang, J., & Rothrock, D. A. (2005). The effect of sea ice rheology in numerical investigations of climate. *Journal Geophysical Research, 110,* C08014. https://doi.org/10.1029/2004JC002599.

Monitoring Strong Tidal Currents in Straits and Nearshore Regions

Alexei Sentchev, Max Yaremchuk and Maxime Thiébaut

Introduction

Open-ocean currents flow in complex patterns affected by wind, water density gradients, topography of the ocean floor, and the Earth's rotation. These large scale and relatively slow evolving forcing terms create water motions that are relatively persistent in magnitude and direction. In contrast, near-shore currents exhibit much stronger magnitude and variability both in space and time. Near the boundaries, in shallow water regions, and in narrow straits, flow dynamics is increasingly complex: tides and tidal currents are amplified, stratification effects are more pronounced, wind driven flows can contribute to extremely large variation of the sea level, while frictional effects in strong currents result in complex turbulent motions.

The strongest currents that can be experienced in coastal regions have tidal origin. Near the coast, tidal dynamics can be considerably modified by the shallow water and the coastline, and tidal current speed attains values of several meters per second that open ocean currents never reach. Regions of such high-tidal current speeds are sparse, and are typically the result of topographic flow amplification. A fast moving tide passing through a constriction can result in a tidal race (extremely strong current). It can also give rise to the formation of waves, eddies, and hazardous currents. In extreme cases, such as in Skookumchuck Narrows in British Columbia, through which tides can travel at more than 17 knots, very large whirlpools develop, which can be extremely hazardous to navigation. In other

A. Sentchev (✉) · M. Thiébaut
Lab. Oceanography and Geosciences, UMR 8187, LOG,
Université du Littoral - Côte d'Opale, CNRS, Université de Lille, Wimereux, France
e-mail: alexei.sentchev@univ-littoral.fr

M. Yaremchuk
Naval Research Laboratory, Stennis Space Center, Hancock, USA
e-mail: max.yaremchuk@nrlssc.navy.mil

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_33

coastal areas, located in their majority along the UK and French coast, not so large but still impressive current speeds (more than 10 knots) are encountered.

Coastal ocean hydrodynamic models, routinely used for SSH and currents forecasting, are not always capable of resolving local bathymetric features and complex physics governing a fast moving tide. In such situations, monitoring of coastal currents by a bottom-based or towed current recording system remains a real but not unique opportunity. In turn, in situ observations by bottom-mounted Acoustic Doppler Current Profiler (ADCP) in strong tidal flows are very difficult to perform because they require highly qualified staff, boats, specific equipment, and if the survey is successfully conducted, the acquired data do not allow to characterize the spatial variability of the flow neither its long-term variation.

Underway velocity measurements by towed or vessel-mounted ADCP is another efficient tool for tidal flow characterization in nearshore areas. In the early studies [8, 32], velocity profiles were sampled with sufficient frequency while the vessel steamed around a circuit allowing to resolve the vertical structure of the tidal current and its spatial irregularities. Using broadband ADCP surveys Goddijn-Murphy et al. [9], Polagye and Thomson [18] documented fine scale velocity variations caused by tidal flow interaction with land and islands. However, the rapid change of the tidal flow during surveying period tends to induce errors in the velocity map interpretation. Goddijn-Murphy et al. [9] showed that the accuracy in reconstruction of the full 4-dimensional tidal flow can be significantly increased by merging observed velocities with the dynamical constraints provided by numerical models. More recently, Sentchev and Yaremchuk [25] used the optimal 4-dimensional interpolation technique for reconstructing space-time evolution of the velocity field derived from towed ADCP surveys in the Boulogne harbour (English Channel). This approach, which combines underway velocity measurements and modelling, offers a real opportunity for short-term monitoring of the strong nearshore circulation.

Another technique of coastal flow monitoring involves remote sensing of surface currents by High Frequency Radars (HFR). Although application of the HF radars in oceanography has more than 30-year history [17, 19, 26], its efficiency for monitoring powerful coastal flows, jets and tidal generated transient eddies was demonstrated only recently (c.f. [22]). A new interest for tidal current monitoring in straits and narrows is motivated by the growing development of the tidal energy conversion by in-stream devices. The UK was a pioneer in this field by installing the first commercial 1.2 MW tidal turbine in the Strangford Loch, a narrow strait connecting the inland tidal basin to the Irish Sea. In France, the Sabella Co. installed its first of three 0.5 MW tidal turbines in the Fromveur Strait (W. Brittany). Other projects aim to install tidal stream conversion systems at various sites in France (Alderney Race) or in Scotland (Pentland Firth). In this context, long term monitoring of strong tidal flows at these sites is of primary importance for engineering applications and assessing the tidal stream potential.

In this chapter we give an overview of the two prospective techniques for monitoring strong currents in the nearshore regions with high tidal stream potential: the remote sensing by HFRs, and underway measurements by a towed ADCP. In

the next section, we first consider the basic principles of HFR-based observations of surface currents and then describe application of the method for assessment of spatial and temporal variability of the tidal flows at two sites where powerful tidal currents dominate local circulation: off the western coast of France and in the Strait of Dover. In section "Measurements of Tidal Currents with a Towed ADCP System", we briefly describe a compact ADCP towing system and demonstrate its application in two regions (the Solent and Dover Strait) in the English Channel. In section "Summary", we summarize the results and discuss the prospects of using the techniques, in particular, for applications related to tidal energy conversion.

Monitoring Surface Currents with HF Radars

Basic principle High Frequency radars operate in a frequency band 5–45 MHz. A part of the electromagnetic energy transmitted at these frequencies is trapped in the "surface-wave" mode and remains bound to the sea surface, following the Earth curvature. Moreover, in this frequency band, the electromagnetic waves have wavelengths that are commensurate with wind-driven gravity waves on the ocean surface. The ocean waves whose wavelengths are exactly half as long as those of the broadcast radio waves are responsible for the resonant backscatter, an example of a phenomenon known as Bragg scattering. Crombie (1955) first observed that the peak energy of HF radio waves reflected from the sea surface occupies a narrow frequency range near the frequencies $\pm f_b$, called the Bragg frequency; f_b represents the Doppler shift of the backscattered signal above and below the HFR frequency, it is defined as $f_b = g/\pi\lambda$, where λ is the radar wavelength, g is the acceleration of gravity. These two Bragg peaks in the power spectrum are produced by the coherent scattering of the HFR waves by the Bragg sea surface waves travelling radially toward and away from the radar.

In an important early study, Stewart and Joy [26] observed that there are often small frequency differences between the observed Doppler shift of the backscattered HFR waves and the Doppler shift predicted from the surface gravity waves moving over still water. Using radar-tracked surface drogues to directly measure ocean currents, they showed that the observed differences resulted from surface ocean currents and that the current speed toward or away from the radar could be estimated. In particular, radial velocity v_r of the surface current (a projection of the current velocity vector on the radar beam) can be retrieved from the Doppler shift Δf of the observed Bragg peaks, compared to their theoretical values, by using the Doppler formula $v_r = \Delta f/2\lambda$.

These principles are the basis for the mapping of surface currents by oceanographic HFRs. Since then, many validation studies comparing HF radar-derived surface currents with in situ observations have contributed to the growing acceptance and use of HFRs for measuring surface currents (e.g., [4, 16, 21, 23]). Stewart and Joy [26] assumed that a HF radar samples vertically integrated current velocity. Earlier work by Kirby and Chen [12] and more recent studies [21, 23] allowed us to

validate this assumption experimentally and conclude that the HFR-derived velocity corresponds to exponentially weighted velocity profile and that the equivalent velocity value is found at the effective depth $z = \lambda/8\pi$. The effective depth is of the order of 0.75 m for 16 MHz and 2.3 m for 5 MHz radar.

Mapping the surface velocity field Individual radar samples the surface velocity over a large area which can extend up to 200 km offshore (5 MHz long-range radar). This area is divided into a number of directional beams and each beam contains a large number of bins (range cells). A good quality Doppler signal requires each range cell to be larger than at least 150 Bragg wavelengths. For this reason, the bin size becomes larger as the radar frequency decreases. A HFR cannot unambiguously identify the arrival directions of backscattered signals. Currently, HFR systems for mapping surface currents may be classified into two types, beam-forming (BF) and direction-finding (DF), based on the method used to determine arrival directions.

Beam-forming radars use a linear array of antenna elements that are steered by adjusting the amplitudes and phases of received signals (e.g., [27]). An early *beam-forming* (BF) radar, used extensively for mapping surface currents, was the ocean surface current radar (OSCR) [20] with an azimuthal resolution of 10°. More recently, the Wellen radar (WERA) was developed by Gürgel et al. [10] using beam-forming configurations with up to 16 antenna elements offering an azimuthal resolution of ~3–4°. Such an array often occupies more than 150 m of coastline which is not always practical.

To achieve smaller footprints, *direction-finding* (DF) configurations are used. These radars compare the phases and amplitudes of radio signals received by closely spaced antenna elements coupled with various direction-finding inversion algorithms. The most common direction-finding radar is the CODAR SeaSonde [1]. Both algorithms (DF and BF) can be used to identify the arrival direction by linear-array antenna WERA system. The azimuthal resolution produced by each method can play a crucial role in analysis of surface circulation patterns in the regions, where high horizontal velocity gradients are expected. Figure 1 provides an example of such a difference in spatial structure of current velocity sampled by 12.4 MHz WERA system around the Ushant Island, off the western Brittany coast of France, by using both DF and BF algorithms. The DF allows small-scale structures of tidal currents to be better resolved. On the contrary, DF radars often produce coverage gaps (cf. eastern and southern margin of the radar coverage area in Fig. 1).

A single HFR measures projection of the current velocity vector on the beam direction. Therefore, velocity measurements by at least two distant radars are required for reconstruction of the vector velocity map. At a point of the beam intersection of two radars, the east–west (u) and north–south (v) components of the surface current vector may be found by solving a system of two equations involving measured radial velocities and known beam angles. If a point on the sea surface is in the coverage area of three or more radars, then u and v can be obtained from a least-squares solution of the corresponding system of linear equations. These are the

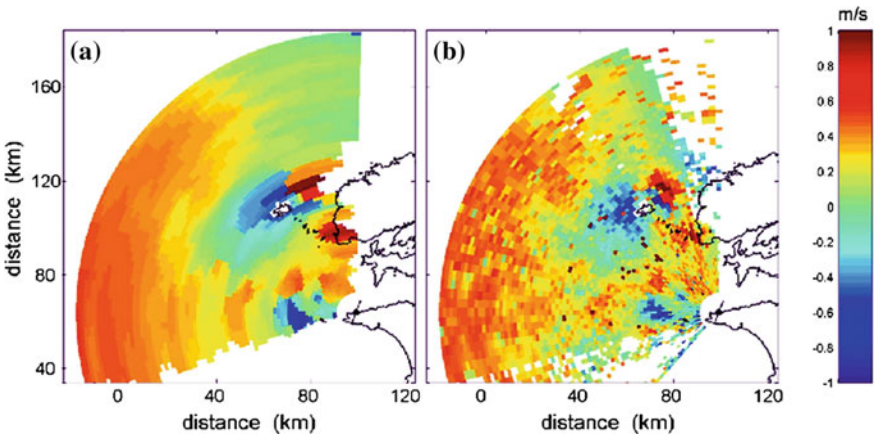


Fig. 1 Radial velocities of surface currents retrieved by the DF (**a**) and BF (**b**) algorithms from HF radar measurements off the W. Brittany coast

methods of local reconstruction of the velocity field. More sophisticated global interpolation methods use linear combinations of radial velocities at all measurement points in conjunction with additional kinematic and/or dynamical constraints to reconstruct the entire velocity field (e.g., [11, 33, 34]).

The Open-boundary Mode Analysis (OMA) method of Kaplan and Lekien [11] performs an expansion of the interpolated velocity field in the eigenfunctions of the 2D Laplacian operator in the interior of the domain with the additional control at the open boundaries by the eigenfunctions of the 1D Lapalcian.

Yaremchuk and Sentchev [33] proposed a more general and versatile 2D variational method of interpolation (2dVar) based on the direct minimization of the cost function with respect to the interpolated velocity field $\mathbf{v}(\mathbf{x})$. Apart from attracting $\mathbf{v}$ to the available observations, the cost function penalizes high-frequency components of the curl and divergence of $\mathbf{v}(\mathbf{x})$. This feature can help to overcome some limitations related to a lack of data. Additional smoothness of the interpolated field is enforced by penalizing the squared Laplacian of velocity, thus providing the algorithm with flexibility and efficient noise control. Kinematic constraints (zero flux on rigid boundary) incorporated into 2dVar, appear particularly useful in regions complicated topography. Yaremchuk and Sentchev [34] also developed a gap-filling technique for HFR observations based on the EOF analysis of the observational errors. In the subsequent presentation, the 2dVar interpolation scheme developed by the authors was used for the reconstructing the velocity fields.

HFR observations in the English Channel In this section, we describe applications of the long-term monitoring technique by HFRs in two regions where implementation of traditional in situ observational platforms is difficult due to extremely strong tidal currents or due to busy shipping traffic, economic and environmental constraints.

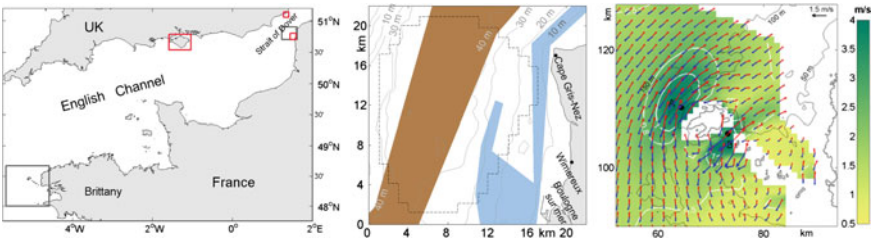


Fig. 2 Left: The English Channel (left) and areas covered by HFR (gray rectangles) and towed ADCP (red rectangles) velocity observations. Middle: HFR coverage zone (gray dashed line) off the Opal coast of France. The HFR sites are shown by black dots. The navigation routes of vessels and fishing areas are shown by brown and blue shading respectively. Contour interval of the bathymetry is 10 m (grey solid lines). Right: Study area off the W. Brittany coast (in the Iroise Sea) with radar coverage zone (grey shading) and radar sites (grey circles). Contour interval of the bathymetry is 50 m (black solid lines)

The first region is located in the Dover Strait off the Opal coast of France, where tidal amplitude can be as high as 9 m and currents can reach 3 m/s. The site is relatively shallow water with depth less than 50 m. In the middle of the Strait, there are sandbanks oriented alongshore (Fig. 2). Currents are predominantly semidiurnal with a pronounced fortnightly modulation due to the interference of the M_2, S_2, and N_2 tidal constituents. In May–June 2003, two Very High Frequency radars (VHFR) were deployed for a 35-day period to monitor tidal currents. Radar sites were located on the Cape Gris Nez and in Wimereux, 12 km southward (Fig. 2). The radars were operating at frequencies 45 MHz and 47.8 MHz, and measured velocities over a distance up to 25 km offshore at 600 m radial and 10° azimuthal resolution, with time discretization of 20 min. The radial velocities measured by the two radars were interpolated on a regular grid with 1 km spacing by variational method with the gap filling capability. As a result, one month long sequence of current velocity maps was generated and used for assessing tidal circulation in the French sector of the Dover Strait [24].

The second study site is located at the entrance to the English Channel, off the western Brittany coast of France, and has larger extension (Fig. 2). The bottom topography is complicated by the group of small islands, islets, and rocks of the Molène archipelago separated from the mainland by more than 15 km wide strait. The larger Ushant Island is separated from the Molène Islands by the 2 km wide and 60 m deep Fromveur Strait. Tidal currents in the strait are very strong with velocities often exceeding 4 m/s.

Two high-frequency Wellen Radars (WERA) operating at 12.4 MHz have been permanently operating on the W. Brittany coast since July 2006. Individual radar sites are located at Cape Garchine (site G), and Cape Brezellec (site B) (Fig. 2). The HFRs were configured to measure velocity in the surface layer 1 m thick at 1.5 km radial and 2° azimuthal resolutions, with a temporal resolution of 20 min. Radial velocities were interpolated on regular 1 km grid using 2dVar interpolation technique. The accuracy of the radar-derived velocities has been estimated by SHOM

(Oceanographic Division of the French Navy) through a comparison with surface drifters and ADCP current measurements for a period of 7 months. In the majority of situations, the discrepancy in velocity measured by different instruments did not exceed 0.15 m/s. A detailed description of the experimental settings, the methods of HFR data processing in the Iroise Sea can be found in Sentchev et al. [22].

Tidal current dynamics off the W. Brittany coast Velocity time series derived from HF radar measurements during the period 04/2007–09/2008 were used for assessing tidal circulation around Ushant Island and estimating the major properties of the flow. The maximum sustained velocity represents the maximum current observed. Tidal flow asymmetry accounts for the asymmetry of velocity magnitude during the tidal cycle. This imbalance between the strength of flood and ebb current speeds is quantified as: $a = \langle V \rangle_{flood} / \langle V \rangle_{ebb}$, where brackets denote time averaging of velocity values on flood and ebb flow respectively. The principal component analysis (PCA) [31] technique was applied to determine the principal direction of the flow during the ebb and flood tide phases. The direction asymmetry $\Delta\theta$ shows deviation of the flow from a straight line (dominant direction). This metric is defined as: $\Delta\theta = |\theta_{flood} - \theta_{ebb} - 180°|$. Current velocity spectrum was also estimated. It can help to explain the interaction between the principal tidal constituents resulting in a strong distortion of the velocity curve.

Spatial variability of tidal currents is quantified by assessing the axes of the tidal current ellipses derived from the PCA (Fig. 3). The length and orientation of ellipse's axis provide information about the tidal current strength and direction of flood and ebb flow and indicate the regions with the most powerful flow. The maximum and time averaged current velocity distribution shows significant spatial variations in the range from 0.75 to 4 m/s for the maximum velocity (Fig. 3, colour shading) and from 0.5 to 2 m/s for the mean spring tide velocity (Fig. 3, white

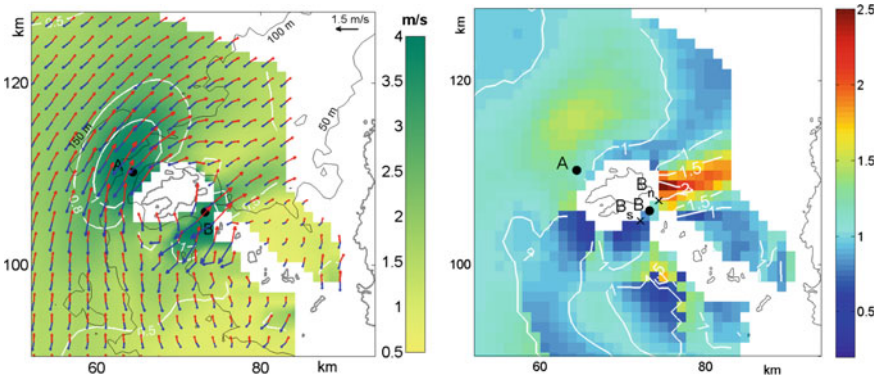


Fig. 3 Left panel: Radar derived current velocities of the flood (red) and ebb (blue) flow during spring tide conditions. Time average and maximum surface current velocity during the study period are shown by white contours and color shading respectively. Black circles mark location of grid points where the maximum velocities are observed (locations A and B). Right panel: Current velocity asymmetry a. B, B_n, B_s mark locations used for the analysis of current velocities

contours). The strongest currents are identified in two areas: west of Ushant Island and in the Fromveur Strait (locations A and B in Fig. 3). Bathymetry gradients and the presence of islands cause flow acceleration, tend to tighten the streamlines and provide the maximum velocities in these two sectors.

Figure 3 also shows a significant (up to 90°) misalignment in the current direction during flood (red) and ebb tide (blue arrows) in the eastern and southern sectors of Ushant Island. Such a deviation of the flow during a tidal cycle is known as direction asymmetry. Sentchev et al. [22] documented the presence of transient eddies in these areas which might explain such a large difference in the direction of the ebb and flood flow. In the north-western sector of the domain, $\Delta\theta$ ranges from 0° to 10°, indicating that ebb and flood flows are mostly aligned.

Another interesting feature of the tidal dynamics around Ushant Island is the velocity asymmetry *a*. The asymmetry varies in a wide range, from 0.5 to 2.5, and shows a large spatial pattern with $a > 1$ in the west and $a < 1$ in the south (Fig. 3). The strongest variation of asymmetry is observed in the Fromveur Strait. Here, the asymmetry attains 2.5 in the onshore sector, indicating that flood flow velocities are by far larger than the ebb flow velocities. The asymmetry decreases toward the centre of the strait (point B in Fig. 3), where the tidal flow reaches a balance ($a = 1$). Abrupt changes in the flow regime occur at very short distance in the strait. The asymmetry values reveal very strong variation, from 1.9 to 0.8, between two locations, B_n and B_s, separated by only 2.5 km at either side of the centre of strait (location B). This gives flood flow velocities up to 2 m/s larger than ebb flow velocities in B_n, and ebb flow velocities larger by ~1 m/s in B_s. Such a strong asymmetry variation, to the best of our knowledge, has never been documented elsewhere.

Different properties of the tidal flow in nearshore environment can explain asymmetry variations. Non-linear interactions between the principal semi-diurnal constituents yield overtides and compound tides. It was established, that the phase difference between semidiurnal and quarter-diurnal tidal constituents governs the tidal flow asymmetry. In particular, the following relationship can be used to diagnoze tidal asymmetry: $\gamma = 2\Phi_{M2} - \Phi_{M4}$, where Φ_{M2} and Φ_{M4} are the phases of the M_2 and M_4 tidal current constituents [6]. Tidal current curve is symmetric when $\gamma = 90°, 270°$. Maximum asymmetry is achieved at $\gamma = 0°, 180°$.

Harmonic analysis of velocity time series in the Fromveur Strait revealed the presence of overtides and compound tides in the spectra. In the central point of the strait (point B), the value of γ was ~270°, thus indicating that the velocity curve is balanced here ($a \approx 1$). On the contrary, in the onshore and offshore sectors of the Strait, the value of γ is close to 215° and 200° respectively, thus providing extreme asymmetry values and a large imbalance of the current flow [29]. Non-linear interactions between the principal semi-diurnal constituents generate residual currents which also contribute to tidal flow asymmetry.

Tidal currents in the Dover Strait The spatial variability of tidal currents in the southern sector of the Dover Strait are quantified using parameters of synthesized tidal current ellipses (Fig. 4) derived from the PCA. Spring tide period (May 16,

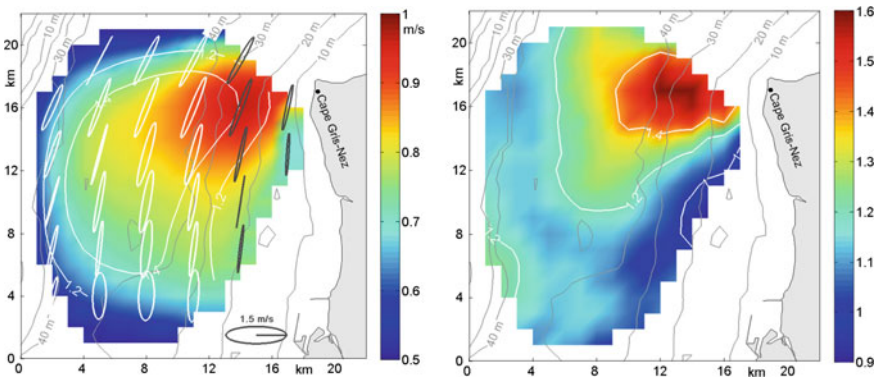


Fig. 4 Left panel: Mean surface current velocity (m/s) during the study period (color) and during the spring tide period in May 2003 (white contours). Shown also tidal current ellipses derived from PCA on May 16, 2003 (every third ellipse is shown). Black white ellipses denote counter-clockwise and clockwise rotating currents. Right panel: Current velocity asymmetry a

2003) was selected to illustrate the location and magnitude of the strongest currents observed. The ellipse orientation shows that the tidal currents are controlled by the topography, producing alignment of the semi-major axes along the depth contours. Anisotropy in current velocity field with a relatively high ellipticity is observed over the sandbanks, in the middle and in the southern part of the study domain. Two distinct zones with opposite sign of current vector rotation are clearly identified (black and white ellipses). They are separated by a line roughly following the 30-m isobath. This suggests that tidal currents are alternatively divergent or convergent during the current reversal. The time averaged current velocity distribution shows low spatial variations with values ranging from 0.5 to 1 m/s, with the highest velocity observed off the Cape Gris Nez. The maximum current speed observed by the radars does not exceed 2 m/s in the study site. The current asymmetry a exceeded unity in almost entire domain, indicating the flood flow dominance (Fig. 4). The maximum current asymmetry ($a = 1.6$) is observed 6 km west of the Cape Gris Nez, indicating the effect of the cape on the spatial distribution of phase and amplitude of the principal (M_2) tidal constituent and its higher order harmonics (M_4, MS_4).

Measurements of Tidal Currents with a Towed ADCP System

Although HFRs provide an efficient tool for long-term monitoring of tidal currents on a larger scale, their spatial resolution could be insufficient for many practical applications, such as identifying precise location of free-stream tidal turbines at tidal energy sites. Much higher spatial resolution can be provided by ship-borne ADCP

observations. In contrast to the bottom-mounted or moored ADCPs, this method of surveying does not require deployment of numerous instruments to obtain adequate spatial resolution and it is not vulnerable to harsh environmental conditions and extreme tidal velocities. Extensive ADCP surveys demonstrated that tidal currents are characterized by significant variations at the scales of a few hundred meters [9, 14], and thus, require a very high resolution surveying during multiple tidal cycles. At the same time, finite speed of the surveying vessel imposes certain limitations on the accuracy of reconstruction of the 4-dimensional velocity field. To resolve the issue, observed velocities should be constrained by the relationships governing tidal dynamics.

This approach was pursued by Goddijn-Murphy et al. [9], who synthesized vessel mounted ADCP transects in the Pentland Firth (Scotland) with the output of the Orkney 2D tidal Model (ORKM) and obtained snapshots of the currents during different tidal phases. Recently, Sentchev and Yaremchuk [25] proposed a 4D optimal interpolation (OI) technique for processing underway velocity measurements. The method employs statistics of a tidal model to infer the velocity error covariances required by the interpolation algorithm.

In this section, we discuss the results of high resolution current mapping with a towed ADCP at two sites in the English Channel characterized by strong tidal currents. They are of certain interest for industrial companies developing energy conversion systems, and thus require extended assessment of the local hydrodynamics.

Methodology High resolution current mapping is performed using an experimental platform carrying a broadband ADCP (1200 kHz or 600 kHz Teledyne RDI WorkHorse Sentinel) towed by a light boat (Fig. 5). The distance from the boat is controlled by an adjustable side fin allowing to keep the platform at a side of the boat and thus to avoid contamination of the observed velocities by the wake. Two additional rear fins installed on the hulls assure stability of the forward propulsion. The ADCP is set to operate at a pinging rate of 1 Hz providing one velocity profile per second with an accuracy of 0.05 m/s. The velocities are sampled with 0.5 m vertical resolution; the first bin centred at 0.75 m. Bottom tracking enables to correct for boat's movement and the recorded velocities form a current vector in the fixed frame relative to the bottom.

The boat speed typically varies within 2–4 m/s during a survey. The recorded ADCP data are merged with high resolution GPS data. The geolocalisation system GENEQ SXblue, mounted on the side of the platform (Fig. 5) and operating at 1 Hz, provides positioning accuracy of 40–80 cm at towing speeds of 2–4 m/s. The above described system, called Koursk, was used for high resolution current mapping at two perspective sites of marine renewable energy in the English Channel: in the Solent and off the Dover harbour.

Assessment of tidal dynamics in the Solent and off the Dover harbor Tidal motions in the Solent (Fig. 5, right panel) are very complex and well known for the unusual phenomena such as the "Double High Water", the "Young Flood Stand", and the short duration of the ebb tide. These peculiarities result from a deformation

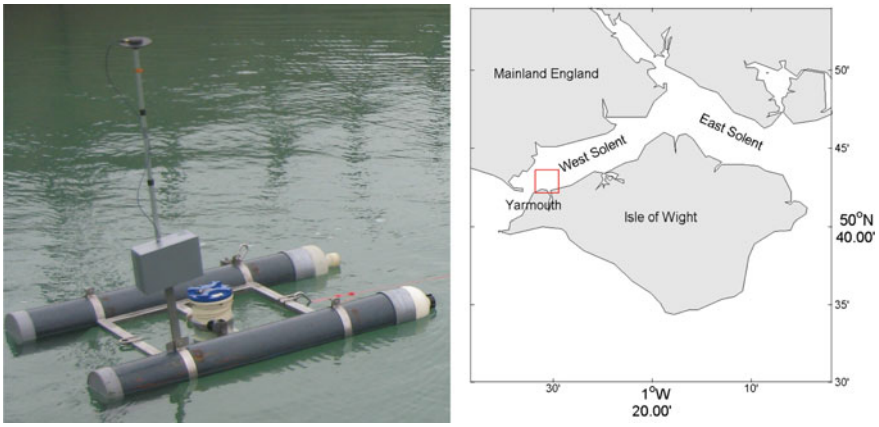


Fig. 5 Left panel: The experimental towing platform Koursk with ADCP profiler (in the centre), high precision GPS with GPS antenna and data acquisition module. Right panel: The Solent and the area covered by towed ADCP surveys in the West Solent (red rectangle)

of the tidal wave travelling through the Solent, generation of overtides and compound tides [3].

A total of five towed and one fixed point ADCP surveys were performed in July and October 2015 to assess the circulation in the West Solent in an area of approximately 3.5 km^2 at 100 m horizontal resolution (Fig. 6). Each survey lasted 1–1.5 h and covered different stages of a tidal cycle. The transects were oriented roughly in the south-north direction and separated by approximately 500 m. Figure 6 (left top panel) shows an example of the surface current velocities recorded by the system during the spring flood tide, one hour before High Water (HW) in Southampton. Significant spatial variations of velocity are observed with a local current intensification in the southwestern and northeastern sectors with steeper bottom topography. The highest current speed, ranging from 2.5 to 3 m/s, was recorded there both on ebb and flood tide. High resolution mapping allowed identifying small scale features of tidal circulation. These include a slacking current in the vicinity of the Yarmouth harbour and in the shallow northern sector, and a transient anticyclonic eddy, 0.5 km in diameter in the southern part of the surveyed zone. The eddy is generated during the flood flow by bottom friction in the vicinity the Sconce Point, which induces negative vorticity downstream.

The towed ADCP survey made possible assessing current variations with depth. Velocity distribution on a cross-section in the western part of the study area on flood flow shows the location of the tidal jet (Fig. 6, bottom panel). Large horizontal gradients of current speed are observed in the southern part of the westernmost cross-section. Velocity profiles appear rather homogeneous throughout the water column with the exception of the relatively thin boundary layer where the vertical velocity shear is large.

Typical flood tide velocities south of the Dover Harbor are shown in Fig. 6 (top right panel). A total of five towed ADCP surveys were performed during three days in July and September 2014 here. The current speed is higher all over the domain

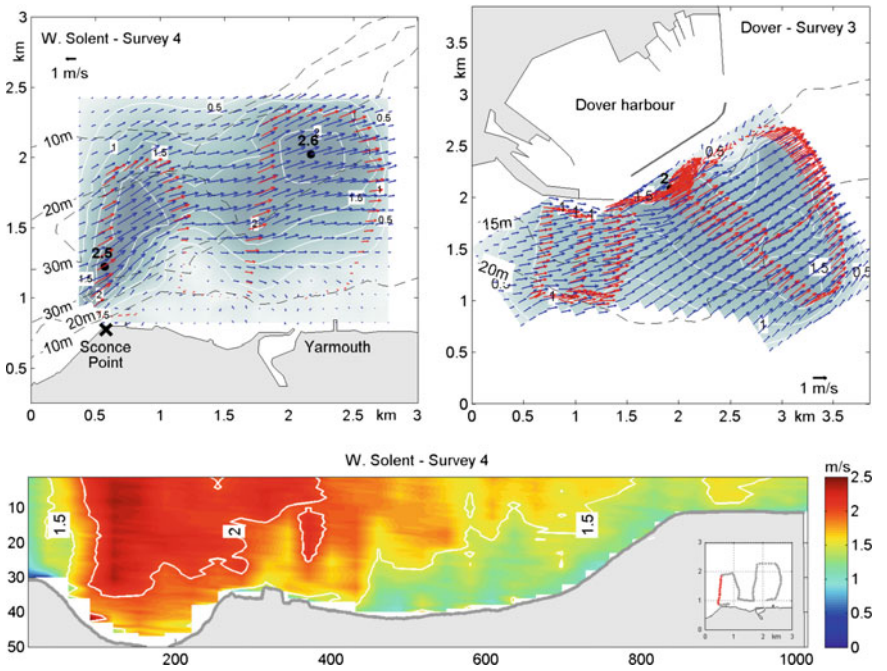


Fig. 6 Surface layer velocities in the West Solent during spring flood tide on 16-10-2015, one hour before HW in Southampton (top left) and off the port of Dover during mean flood tide on 5-7-2014, 2.5 h after LW in Dover (top right). Current vectors derived from underway ADCP velocity measurements, 30-s averaged, are shown in red. Bottom panel: Current speed over the cross-section in the West Solent during ebb tide on 1-9-2014. Cross-section location is given in red in the insert

during flood tide than during ebb tide, but the flow pattern looks similar. As can be expected, tidal currents are generally constrained by shoreline, port dykes, and bathymetry contours. Highest velocities are found in the vicinity of the breakwater, and then gradually decrease seaward. A small size cyclonic vorticity feature is identified in the northeastern part of the domain. It indicates the presence of a large recirculation area with relatively weak currents east of the breakwater.

4D interpolation of the underway velocity measurements Uncertainties in spatial patterns of the tidal flow reconstructed under the assumption of instantaneous observations can be significant since evolutions of tidal currents at time scales of a few hours are often comparable to the surveying period. To avoid distortion of the results caused by temporal variation of the tidal flow during the survey, very short (20–30 min) transects are repeatedly made, usually around the peak tidal flow (e.g., [5]). This time limitation prevents continuous studies of larger domains, which require surveying times comparable with the characteristic time scale of the velocity field ($T \sim 2$–4 h).

In this section we present 4D optimal interpolation technique, in both space and time, allowing retrieving the entire evolution of tidal currents from the survey data. The technique employs tight space-time correlations in the tidal flow field that can be accurately simulated by the existing state-of-the-art numerical models. A brief description of the method and its application to assessing tidal circulation in a limited size area (Boulogne Harbour) are given below. The harbour is located in the French sector of the Dover Strait, in the vicinity of the area covered by the HFR observations (Fig. 2).

A straightforward method of statistically consistent spatial interpolation of a vector field is the well-known optimal interpolation (OI) pioneered by Gandin [7]. Since then, the approach was widely adopted in geosciences (e.g., [2, 30]). The OI technique can be easily extended to include time dimension by using the space-time correlation functions. In this approach, the optimal correction to the evolution of a background vector field $\boldsymbol{u}^m(\boldsymbol{x}, t)$ defined on a regular (model) grid is represented by a linear combination of the weighted differences between the background trajectory and the observed velocities. The weights a_i are chosen so as to minimise the mean square difference between observations $\boldsymbol{u}_i^*$ and the background field values $\boldsymbol{u}^m$, interpolated into the space-time locations of the observations by the (linear) operator $\boldsymbol{H}^i$, projecting gridded velocity values onto the i-th observation point from the apexes of the enveloping grid cell:

$$J_u = \left\langle \left[u^m + \sum_i \mathrm{a}_i (H^i u^m - u_i^*) \right]^2 \right\rangle \rightarrow \min(\mathrm{a}_i) \tag{1}$$

Here, angular brackets denote the statistical (ensemble) average, and summation is made over all (distributed in space and time) velocity values measured during the survey period. Given the space-time covariance matrices of the model $\boldsymbol{B} = \langle \boldsymbol{u}^m(\boldsymbol{x}, t)\, \boldsymbol{u}^m(\boldsymbol{x}', t') \rangle$ and observations $\boldsymbol{R}_{ij} = \langle \boldsymbol{u}_i^* \boldsymbol{u}_j^* \rangle$, and assuming that observation errors are not correlated with the model (background) errors, the OI interpolation formula takes the form:

$$u_{opt} = u^m + \sum_{ij} B H_j^T (H_i B H_j^T + R_{ij})^{-1} (H^i u^m - u_i^*) \tag{2}$$

In order to apply Eq. (2) for interpolating the surveyed velocities, $\boldsymbol{u}^m(\boldsymbol{x}, t)$ and $\boldsymbol{B}$ are specified using the output statistics from the regional model MARS-3D [13] configured for high resolution simulations in the surveyed area [25].

The result of interpolation allows accurate tracking of the motion of the anti-cyclonic eddy derived from underway velocity measurements by towed ADCP within the Boulogne harbor on March 27, 2012 (Fig. 7a). Figure 7 also shows two snapshots of the velocity field: 20 min after the beginning of the survey, at HW − 1.2 h, and at the end of the survey, at HW + 0.5 h. As it can be seen, the optimal interpolation enables relatively accurate assessment of the circulation pattern: at the beginning of the period (Fig. 7b), the eddy center was located just south

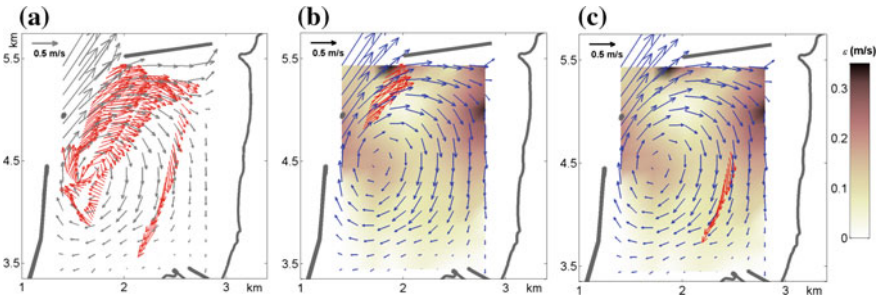


Fig. 7 **a** Observed (red) and modelled (gray) surface velocity fields during the towed ADCP surveys in Boulogne Harbour on March 27, 2012 1200–1400 GMT. Evolution of the anticyclonic eddy from HW − 1.2 h (**b**) to HW + 0.5 h (**c**). Red arrows show 30-s averaged ADCP velocities recorded along the track for the respective tidal stages. Blue arrows show interpolated velocities. Background shading shows the difference between the survey velocities interpolated in space and time and model velocities at mid-time of the survey

of the harbor entrance, 300 m east of the western seawall. In the northern part, the water inflow is observed with mean velocities ranging from 0.5 to 1 m/s. Later, at HW + 0.5 h, the eddy migrated northeastward and expanded in size to occupy the entire area of the harbor (Fig. 7c). The background color in Fig. 7 indicates the difference in the velocity magnitude between the space-time interpolated currents and the respective currents of the background model run $\boldsymbol{u}^m(\boldsymbol{x}, t)$. Most of the data-induced changes are observed in the northern part of the harbor with the major difference (≈0.30 m/s) at the periphery of the eddy, showing significant modifications of the circulation pattern provided by the interpolation.

Summary

Currents of several meters per second can never be experienced in the open ocean but could be encountered in straits and nearshore regions due to the interaction of tidal waves with topography. Although regions of such extreme currents are sparse, in situ observations of local flows are very difficult to perform due to extreme environmental conditions. Static velocity measurements performed by either bottom-mounted or moored ADCPs do not allow to adequately characterize neither the spatial structure of the flow nor its long-term variation. In this chapter, we presented two complimentary flow monitoring techniques that could be suitable for continuous observations of the extreme currents.

The first type of reported observations was provided by two HFR networks operated along the French coast, in the western and eastern sectors of the English Channel. The HFR-derived velocities of surface currents were analyzed by means of various numerical and statistical techniques. In particular, processing long-term HFR observations in the W. Brittany provided maps of surface currents in

unprecedented detail due to the application of the DF technique in conjunction with variational 2dVar interpolation. The approach was essential for accurate assessment of circulation around the islands and in the straits were high horizontal velocity gradients occur.

Our analysis revealed velocities exceeding 4 m/s and a pronounced asymmetry between the flood and ebb flow varying in a wide range, 0.5–2.5, around Ushant Island. The strongest variation of asymmetry was found in the Fromveur Strait. Neill et al. [15] investigated spatial variation of asymmetry around the Orkney Islands, where strong tidal currents occur, and documented the values ranging from 0.7 to 1.3. Four times larger asymmetry variation has been found around Ushant Island, which, to the best of our knowledge, has never been documented at any other place in the World Ocean.

HFR velocity monitoring in the Strait of Dover documented the strongest currents (up to 2.5 m/s at spring tide) occurring west of the Cape Gris Nez. This highly energetic area (power density ~1 kW/m^2), located outside the northward navigation routes and fishing zone, appears suitable for testing energy conversion devices there [28].

Although HFR-based observations present an adequate tool for monitoring extreme currents at larger scale, their spatial resolution is often insufficient for the task of accurate deployment of tidal turbines. In that respect, the presented high resolution current mapping system Koursk can be seen as complementary to other coastal observing systems, such as HFRs and moored ADCP arrays. The system was successfully tested in several sites (the Solent, ports of Dover and Boulogne) demonstrating a good combination of accuracy, portability and low cost of surveying. The system is equipped with an interpolation algorithm, which allows reconstructing space-time evolution of the velocity field for surveys whose duration is comparable or larger than the time scale of tidal variability (1–2 h).

Application of the technique in Boulogne Harbor demonstrated significant (30–60%) reduction of the model-data misfit for the velocity field obtained as a result of space-time optimal interpolation. Although the method was applied to recover surface circulation, it can be extended for assessment of the full 4D tidal flow dynamics using the data recorded throughout the entire water column. We believe that the proposed current mapping system may advance understanding and assessing coastal circulation in tidal environments, especially when used in combination with moored ADCP and/or HFR observations. The data acquired by HFRs and towed ADCP, combined with the presented methodologies of their processing could improve the efficiency of ongoing tidal energy conversion projects in France and UK. The HFR data can be also used for numerical model validation and for data assimilation in numerical models, which may improve their forecast capability.

Acknowledgements The authors acknowledge the support of the Interreg IVB (NW Europe) "Pro-Tide" project and support from the US Office of Naval Research. We also acknowledge the Oceanographic Division of the French Navy (SHOM) for providing ADCP and HF radar data in the Iroise Sea. The authors thank skipper Eric Lecuyer (LOG) and also Philippe Forget and Yves Barbin (MIO, Toulon) for their contribution to radar data processing.

References

1. Barrick, D. E. (2008). 30 years of CMTC and CODAR. In *IEEE/OES 9th Working Conference on Current Measurement Technology, 2008. CMTC 2008* (pp. 131–136). IEEE.
2. Bretherton, F. P., Davis, R. E., & Fandry, C. B. (1976). A technique for objective analysis and design of oceanographic experiments applied to MODE-73. In *Deep-Sea Research and Oceanographic Abstracts* (Vol. 23, pp. 559–582). Elsevier.
3. Dyer, K. R., & King, H. L. (1975). The residual water flow through the Solent, South England. *Geophysical Journal of the Royal Astronomical Society, 42*(1), 97–106.
4. Emery, B. M., Washburn, L., & Harlan, J. A. (2004). Evaluating radial current measurements from CODAR high-frequency radars with moored current meters. *Journal of Atmospheric and Oceanic Technology, 21*(8), 1259–1271.
5. Evans, P., Mason-Jones, A., Wilson, C., Wooldridge, C., O'Doherty, T., & O'Doherty, D. (2015). Constraints on extractable power from energetic tidal straits. *Renewable Energy, 81,* 707–722.
6. Friedrichs, C. T., & Aubrey, D. G. (1988). Non-linear tidal distortion in shallow well-mixed estuaries: a synthesis. *Estuarine, Coastal and Shelf Science, 27*(5), 521–545.
7. Gandin, L. S., & Hardin, R. (1965). *Objective analysis of meteorological fields* (Vol. 242). Israel Program for Scientific Translations. Jerusalem.
8. Geyer, W. R., & Signell, R. (1990). Measurements of tidal flow around a headland with a shipboard acoustic Doppler current profiler. *Journal of Geophysical Research: Oceans, 95* (C3), 3189–3197.
9. Goddijn-Murphy, L., Woolf, D. K., & Easton, M. C. (2013). Current patterns in the inner sound (Pentland Firth) from underway ADCP data. *Journal of Atmospheric and Oceanic Technology, 30*(1), 96–111.
10. Gurgel, K. W., Antonischki, G., Essen, H. H., & Schlick, T. (1999). Wellen Radar (WERA): A new ground-wave HF radar for ocean remote sensing. *Coastal Engineering, 37*(3), 219–234.
11. Kaplan, D. M., & Lekien, F. (2007). Spatial interpolation and filtering of surface current data based on open-boundary modal analysis. *Journal Geophysical Research, 112,* C12007.
12. Kirby, J. T., & Chen, T. M. (1989). Surface waves on vertically sheared flows: approximate dispersion relations. *Journal Geophysical Research, 94*(C1), 1013–1027.
13. Lazure, P., & Dumas, F. (2007). An external-internal model coupling for a 3D hydrodynamical model for applications at regional scale (MARS). *Advances in Water Resources*, *31*, 233–250.
14. Li, C., Armstrong, S., & Williams, D. (2006). Residual eddies in a tidal channel. *Estuaries and Coasts, 29*(1), 147–158.
15. Neill, S. P., Hashemi, M. R., & Lewis, M. J. (2014). The role of tidal asymmetry in characterizing the tidal energy resource of Orkney. *Renewable Energy, 68,* 337–350.
16. Ohlmann, C., White, P., Washburn, L., Emery, B., Terrill, E., & Otero, M. (2007). Interpretation of coastal HF radar–derived surface currents with high-resolution drifter data. *Journal of Atmospheric and Oceanic Technology, 24*(4), 666–680.
17. Paduan, J. D., & Washburn, L. (2013). High-frequency radar observations of ocean surface currents. *Annual Review of Marine Science, 5,* 115–136.
18. Polagye, B., & Thomson, J. (2013). Tidal energy resource characterization: Methodology and field study in Admiralty Inlet, Puget Sound, WA (USA). *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*.
19. Prandle, D. (1991). A new view of near-shore dynamics based on observations from HF radar. *Progress in Oceanography, 27*(3), 403–438.
20. Prandle, D., & Ryder, D. K. (1985). Measurement of surface currents in Liverpool Bay by high-frequency radar. *Nature, 315,* 128–131.

21. Röhrs, J., Sperrevik, A. K., Christensen, K. Ha., Broström, G., & Breivik, Ø. (2015). Comparison of HF radar measurements with Eulerian and Lagrangian surface currents. *Ocean Dynamics*, *65*(5), 679–690.
22. Sentchev, A., Forget, P., Barbin, Y., & Yaremchuk, M. (2013). Surface circulation in the Iroise Sea (W. Brittany) from high resolution HF radar mapping. *Journal of Marine Systems, 109,* S153–S168.
23. Sentchev, A., Forget, P., & Fraunié, P. (2017). Surface current dynamics under sea breeze conditions observed by simultaneous HF radar, ADCP and drifter measurements. *Ocean Dynamics, 67*(3–4), 499–512.
24. Sentchev, A., & Yaremchuk, M. (2007). VHF radar observations of surface currents off the northern Opal coast in the eastern English Channel. *Continental Shelf Research, 27*(19), 2449–2464.
25. Sentchev, A., & Yaremchuk, M. (2016). Monitoring tidal currents with a towed ADCP system. *Ocean Dynamics, 66*(1), 119–132.
26. Stewart, R. H., & Joy, J. W. (1974). HF radio measurements of surface currents. In: *Deep Sea Research and Oceanographic Abstracts* (Vol. 21, pp. 1039–1049). Elsevier.
27. Teague, C. C., Vesecky, J. F., & Fernandez, D. M. (1997). *HF radar instruments, past to present*. *Oceanography, 10*, 40–44.
28. Thiébaut, M., & Sentchev, A. (2016). Tidal stream resource assessment in the Dover Strait (eastern English Channel). *International Journal of Marine Energy, 16,* 262–278.
29. Thiébaut, M., & Sentchev, A. (2017). Asymmetry of tidal currents off the W. Brittany coast and assessment of tidal energy resource around the Ushant Island. *Renewable Energy, 105,* 735–747.
30. Thiébaux, H. J., & Pedder, M. A. (1987). *Spatial objective analysis with applications in atmospheric science*. London: Academic Press.
31. Thomson, R. E., & Emery, W. J. (2001). *Data analysis methods in physical oceanography*. Elsevier.
32. Vennell, R. (1994). Acoustic Doppler current profiler measurements of tidal phase and amplitude in Cook Strait, New Zealand. *Continental Shelf Research, 14*(4), 353–364.
33. Yaremchuk, M., & Sentchev, A. (2009). Mapping radar-derived sea surface currents with a variational method. *Continental Shelf Research, 29*(14), 1711–1722.
34. Yaremchuk, M., & Sentchev, A. (2011). A combined EOF/variational approach for mapping radar-derived sea surface currents. *Continental Shelf Research, 31*(7), 758–768.

Part IV
Scientific Contributions: Arctic Oceanography

Analytical Solutions Describing Zonal and Circular Wind Drift of Sea Ice with Elastic-Plastic Rheology

Aleksey Marchenko

Introduction

Modelling of large scale dynamics of drifting ice is based on the consideration of ice rheology at scales of several tens kilometres and greater. Since experiments are not possible over such scales, the rheological properties are formulated using intuitive ideas, results of field measurements, and analysis of satellite data. A model with elastic-plastic rheology was considered in AIDJEX project (e.g., [1]), and a model with viscous-plastic rheology was formulated by Hibler [2]. Both of the models use normal flow rule with closed yield curve transforming in self-similar manner depending on the ice thickness and compactness. In the elastic-plastic model, the ice is considered as an isotropic elastic continuum when the stresses are inside the yield curve. In the viscous-plastic model the ice is modelled as a linear compressible viscous continuum when the stresses are inside the yield curve.

A comprehensive review of sea ice rheological models is performed by Lepparanta [6]. He also analysed analytical steady-state solutions describing ice drift along the coastal boundaries of rectilinear and circular shapes within the viscous-plastic model of Hibler [2]. The ice thickness, compactness profiles, and drift velocity were analysed depending on the wind direction and distance to the shoreline. Axially symmetric solutions describing the drift of floating ice with elastic-plastic rheology were investigated by Schwaegler and Pritchard [11] numerically for the validation of rheological characteristics of drifting ice in the Arctic and for the investigation of quasi-steady-state response of an axisymmetric

A. Marchenko (✉)
The University Centre in Svalbard, Longyearbyen, Norway
e-mail: aleksey.marchenko@unis.no

A. Marchenko
Sustainable Arctic Marine and Coastal Technology (SAMCoT),
Centre for Research-Based Innovation (CRI), Norwegian University of Science and Technology, Trondheim, Norway

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_34

ice model driven by a prescribed atmospheric high-pressure system. They determined that the quasi-steady-state response of ice disk with a diameter of 1000 km is established within 6 h of the initial air stress application, regardless of the initial conditions.

Formation of the zones with "solid body" motion is a common property of the elastic-plastic solutions. In plasticity theory, the motion of a continuum is performed as a combination of motions of solid blocks separated by shear plastic zones or slip lines (e.g., [10]). Korsnes [4] considered elements of rigid motion of ice floes using the ERS-1 SAR data. Sharp discontinuities separating the regions of uniform ice motion in the Arctic drift sea ice were investigated by Kwok [5]. The analysis was performed using the high-resolution deformation fields of the Arctic ice cover produced by RADARSAT Geophysical Processor System. Goldstein et al. [3] used RADARSAT images to record solid body motion of ice bands in Bay of Bothnia.

Marchenko [8] investigated analytically steady-state drift of elastic-plastic ice forced by the drag of sea current unlimited in the longitudinal direction and localized in the transversal direction. It was shows that the solution consists of the elastic kernel moving with constant velocity in the direction of the sea current and plastic boundary zones located at different sides of the elastic kernel. Ice velocity inside the plastic zones is parallel to the sea current, it equals to the velocity of the elastic kernel along the contact lines and tends to zero at the periphery of the plastic zones. Circular drift of elastic-plastic ice caused by the drag force of wind vortex was investigated by Marchenko [9] analytically. Circular wind vortex is approximated by the polynomial function of the polar radius. It was shown that the drag force from cyclonic or anticyclonic wind vortex influences circular drift of ice consisting of the solid body rotation of the elastic kernel and plastic shear zone in the form of a ring surrounding the elastic kernel. Outside the plastic ring, the ice is not moving. Solutions with the elastic kernel and plastic ring exist when the maximum wind speed is greater than the critical speed. If the wind speed is below the critical value the ice stresses are in the elastic range. It was found that the width of the plastic ring decreases as the maximum wind speed becomes close to the critical value.

The main goal of this contribution consists in the further analysis of the analytical solutions of sea ice dynamics model based on AIDJEX elastic-plastic rheology taking into account sea surface tilt and wind velocity component transversal to ice drift direction. In the second section, we consider examples of ice drift features appearing under the influence of wind on the ice near shore line of linear shape, and under the influence of cyclonic wind on the ice. In the third section, the model equations are formulated. In the fourth section, the characteristics of plastic shear zones in drift ice are investigated in the plane and axially symmetric cases. In the fifth section, zonal wind drift of ice cover near rectilinear shoreline is investigated. In the sixth section, circular wind drift of ice cover is investigated. The main results of the paper are discussed and conclusions are formulated in the last section.

Ice Drift Features in the Zonal and Circular Drift of Sea Ice

Images of MYD09Q1 with a 250-m resolution in an 8-day gridded level provided by Google Earth Engine (https://explorer.earthengine.google.com/#workspace) are used to consider selected features of the ice drift in the Fram Strait, near the east coast of Edgøya (Spitsbergen), and near the north coast of Alaska. Figure 1 shows long linear faults extended over about 500 km along the direction of the ice drift in the Fram Strait. Dominant winds (https://earth.nullschool.net/) on April 22–30, 2016, were directed along the shoreline. The faults separate ice strips, which width is about 50 km and smaller.

Similar faults approaching the steady ice are recognized near the east coast of Edgøya (Fig. 2). Dominant winds on May 25–June 2, 2015, were from the north-northeast (https://earth.nullschool.net/). Winds of onshore directions influence ice compression and ice rubble build up in the shear zone. Visual observation performed from the board of the RV "Lance" showed that the ice rubble extended along the shore line over several tens of kilometres and the width of the ice rubble exceeded 10 km. The free board of the rubble reached 3–4 m and the external boundary of the rubble consisted of linear segments of almost vertical ice walls (Fig. 3). More detailed description of the ice rubble characteristics is given in (Marchenko and Marchenko 2017).

Ice faults of circular shape are observed in the spring season near the northern coast of Alaska. Analysis of the local winds (https://earth.nullschool.net/) shows

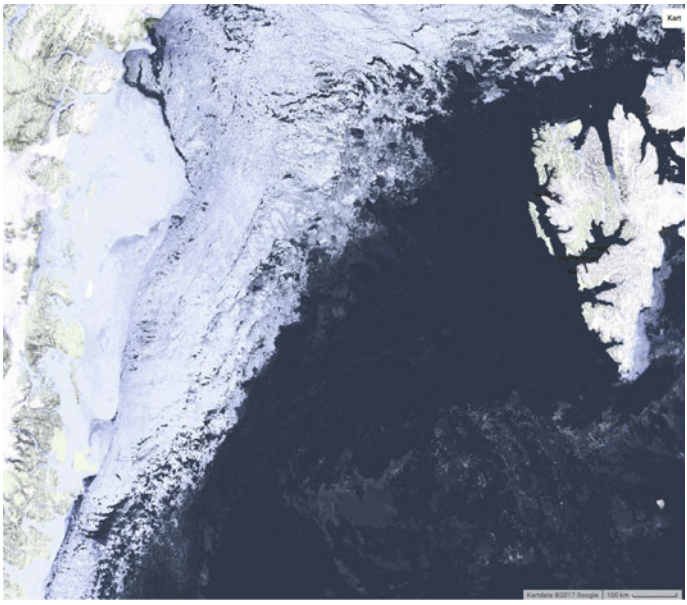


Fig. 1 Formation of longitudinal shear zones in drift ice in the Fram Strait; April 22–30, 2016

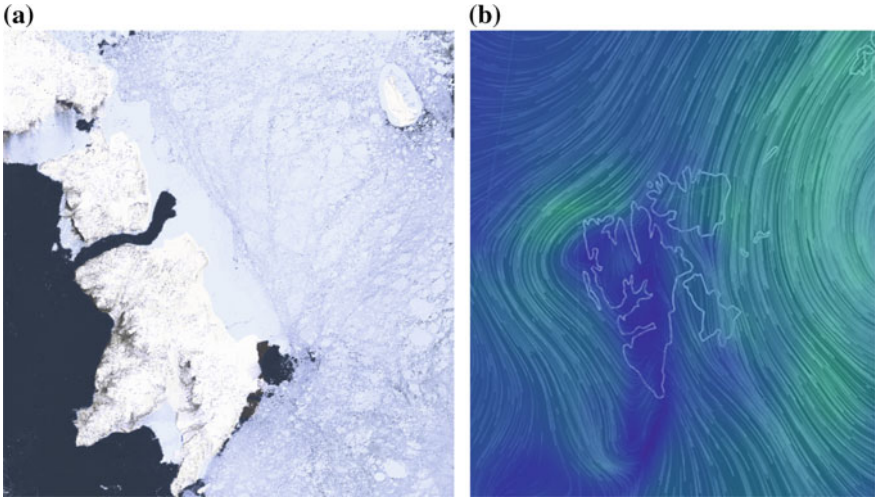


Fig. 2 Formation of shear zone near the east coast of Edgøya; May 25–June 2, 2015 (**a**). Wind field near Spitsbergen on May 30, 2015 (**b**)

that these faults are formed when local cyclone is in the region (Fig. 4). The faults are not perfectly of the round shape, but their fragments have circular or elliptic shapes. The radius of the faults is estimated at approximately 500 km. The centres of the circles or ellipses are around the centre of the cyclone. Figure 5 shows the development of the circular faults over 49 days from the end of March to the beginning of May, 2016. It shows stability and spatial localization of the faults.

Model Equations

One dimensional steady motion of solid drift ice with 100% surface concentration is considered in the plane and axially symmetric cases. The motion can be caused by ice stresses, wind and sea surface tilt. It is assumed that ice velocity is perpendicular to the spatial direction where the ice characteristics (thickness, stresses, and velocity) and the drag forces are changed. It means that pure shear motion of drift ice is considered. In this case the equation of mass balance is satisfied. The momentum balance equations are written as follows:

$$\frac{d\sigma_{xx}}{dx} = \rho_i h(-fv_y + g\eta_x) - F_x, \ \frac{d\sigma_{xy}}{dx} = -F_y, \tag{1}$$

$$\frac{d\sigma_{rr}}{dr} + \frac{\sigma_{rr} - \sigma_{\theta\theta}}{r} = \rho_i h(fv_\theta + g\eta_r) - F_r, \ \frac{d\sigma_{r\theta}}{dr} + \frac{2\sigma_{r\theta}}{r} = -F_\theta. \tag{2}$$

Fig. 3 Research Vessel "Lance" near the shear zone at the east coast of Edgøya; May 2, 2016

Equations (1) and (2) describe the momentum balance of the drift ice in the cases of steady plane and axially symmetric drift. Here, σ_{xx}, σ_{yy}, and σ_{xy} are the components of ice stresses in the Cartesian coordinates x and y; σ_{rr}, $\sigma_{\theta\theta}$, and $\sigma_{r\theta}$ are the components of ice stresses in polar coordinates r and θ; v_y and v_θ are the y-component and angular component of ice drift velocity, ρ_i and h are the density and the thickness of the ice cover, f is the Coriolis parameter; η_x and η_r are the sea surface tilts in the Cartesian and polar coordinates, $F_{x,y}$ and $F_{r,\theta}$ are the components of the drag forces applied to the ice by wind and water. It is assumed that sea surface tilts in y and θ-directions are small. Therefore, they are not included in Eqs. (1) and (2).

The Coriolis parameter is greater than zero $f > 0$ in the Northern Hemisphere. It is assumed that in the Northern Hemisphere positive direction of the y-axis is

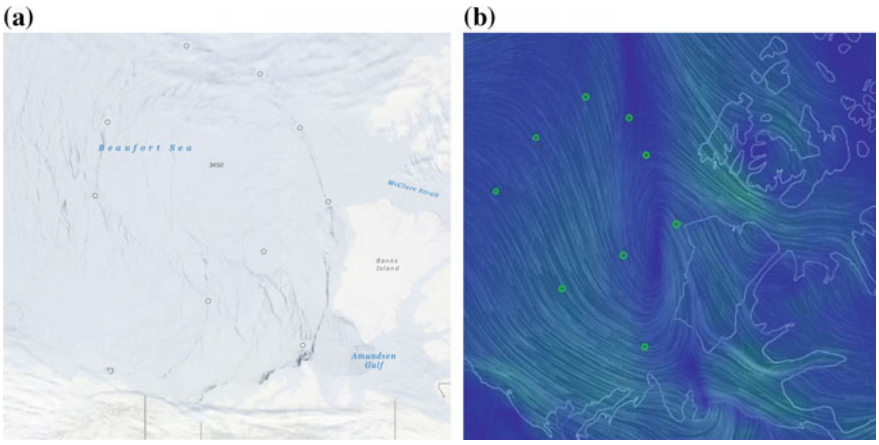


Fig. 4 Circular fault in drift ice near the north coast of Alaska; March 29–April 6, 2016 (**a**). Cyclonic wind field. Green points correspond to the contour of the circular fault (**b**)

associated with the direction to the north, positive direction of the x-axis is associated with the direction to the east, and polar angle θ increases in the clockwise direction.

Elastic-plastic model of the ice cover is considered. The yield surface (YC) is specified by the equation

$$f_Y(\sigma_I, \sigma_{II}, h) = 0, \tag{3}$$

where $2\sigma_I = \sigma_1 + \sigma_2$, $2\sigma_{II} = \sigma_1 - \sigma_2$, σ_1, and σ_2 are the maximal and minimal principal stresses, $-\sigma_I$ is the pressure, and σ_{II} is the maximal shear stress.

Normal flow rule is satisfied when the stresses belong to the YC:

$$e_I = \lambda \frac{\partial f_Y}{\partial \sigma_I},\ e_{II} = \lambda \frac{\partial f_Y}{\partial \sigma_{II}}, \lambda \geq 0, \tag{4}$$

where $e_I = e_1 + e_2$, $e_{II} = e_1 - e_2$, and e_1 and e_2 are the maximal and minimal principal strain rates.

The YC shown in Fig. 6 has a shape of the closed curve located in the region of positive pressure ($\sigma_I < 0$). The YC is characterized by two rheological constants (P^* and γ). It is described by the following formula [7]:

$$\gamma \sigma_{II}^2 = \sigma_I^2 \left(1 + \frac{\sigma_I}{P^* h}\right),\ P^* = 5\ \text{kPa}, \gamma = 3, \tag{5}$$

Point O_s with horizontal tangent line has coordinates $\sigma_I/h = -2P^*/3$ and $\sigma_{II}/h = -2P^*/(3\sqrt{3\gamma})$ (Fig. 6). According to the normal flow rule (4), plastic

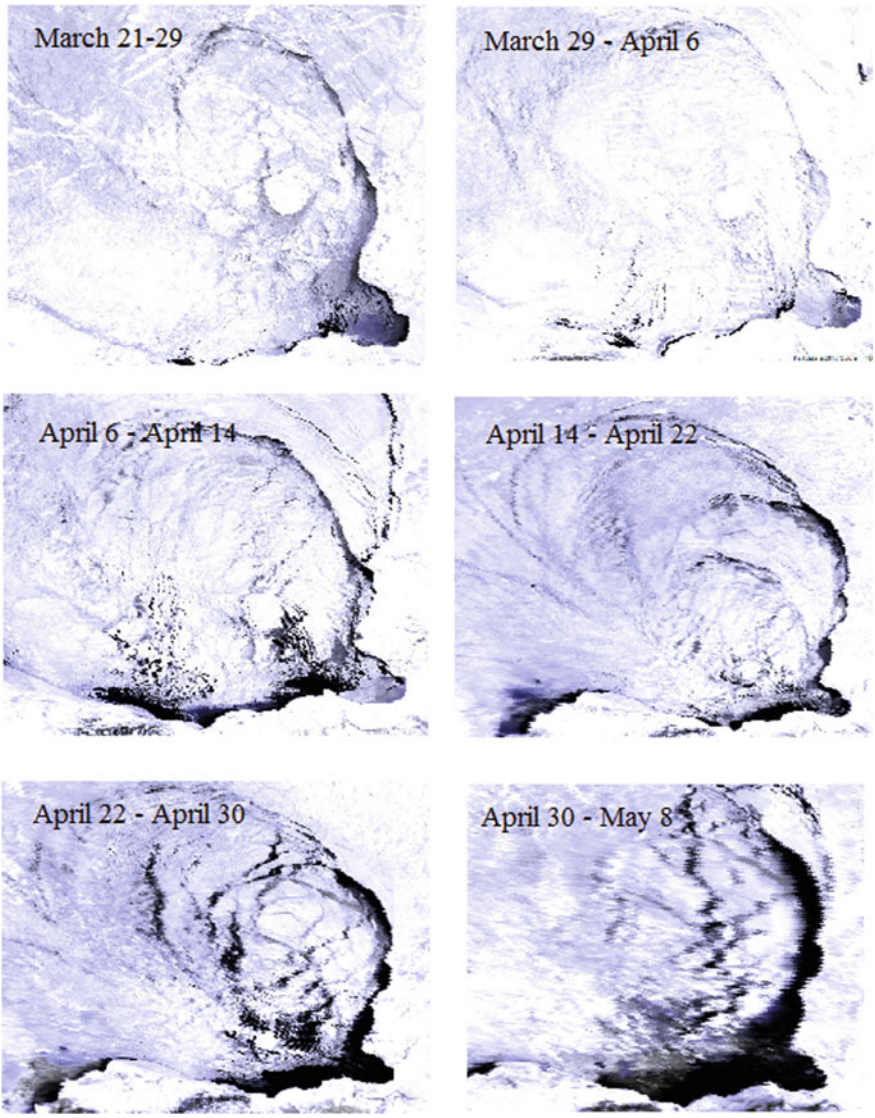


Fig. 5 Development of the circular faults in drift ice near the Alaska coast in 2016

deformations are accompanied by tension ($e_I > 0$) when $\sigma_I/h > -2P^*/3$, and by compression ($e_I < 0$) when $\sigma_I/h < -2P^*/3$.

Pure shear plastic deformations occur when the ice stresses divided by the ice thickness σ_I/h and σ_{II}/h equal to the coordinates of point O_s (Fig. 6). Slip lines analysis shows that the angle between the slip lines is equal to 90° in the case of pure shear deformations (e.g., [12]). In the case of plane motion the slip lines are

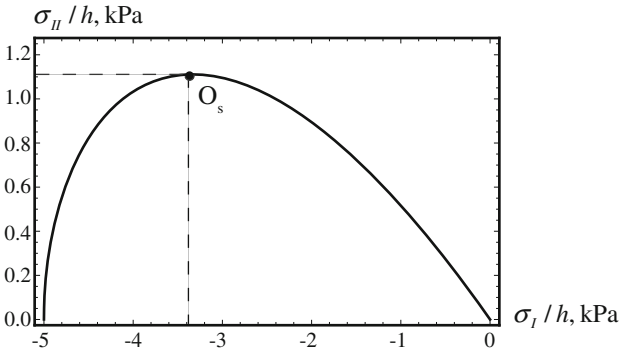


Fig. 6 Yield curve of solid ice cover

Table 1 Representative numerical values of physical constants

ρ_i (kg/m^3)	ρ_w (kg/m^3)	ρ_a (kg/m^3)	C_w	C_a	f (s^{-1})
920	1020	1.27	0.005	0.002	1.454×10^{-4}

parallel to the coordinate axes x and y. In the axially symmetric case the slip lines are parallel to the coordinate lines r (radial lines) and θ (concentric circles). Since normal stresses are the same on the slip lines passing through the same point then

$$\sigma_{xx} = \sigma_{yy},\ \sigma_{rr} = \sigma_{\theta\theta}. \tag{6}$$

Therefore, in the plane case $\sigma_I = \sigma_{xx} = \sigma_{yy}$ and $\sigma_{II} = |\sigma_{xy}|$, and in the axially symmetric case $\sigma_I = \sigma_{rr} = \sigma_{\theta\theta}$ and $\sigma_{II} = |\sigma_{r\theta}|$.

Drag force ($\mathbf{F}$) is equal to a sum of wind ($\mathbf{F}_a$) and water ($\mathbf{F}_w$) drag forces

$$\mathbf{F} = \mathbf{F}_a + \mathbf{F}_w,\ \mathbf{F}_a = \rho_a C_a V_a \mathbf{V}_a,\ \mathbf{F}_w = -\rho_w C_w v \mathbf{v}, \tag{7}$$

where $\mathbf{V}_a = (V_{ax}, V_{ay})$ and $\mathbf{v} = (0, v_y)$ are related to the plane motion, $\mathbf{V}_a = (V_{ar}, V_{a\theta})$ and $\mathbf{v} = (0, v_\theta)$ to the axially symmetric motion, $V_a = |\mathbf{V}_a|$ and $v = |\mathbf{v}|$, ρ_a and ρ_w are the air and water densities, C_a and C_w are the air-ice and water-ice drag coefficients. Representative values of physical constants are given in Table 1.

Ice Drift Induced by Shear Stresses

Equations (1) and (2) in the case when the wind velocity and sea surface tilt are absent are written as follows:

$$\frac{2P^*}{3}\frac{dh}{dx} = \rho_i f h v_y,\ \frac{2P^*}{3\sqrt{3\gamma}}\frac{dh}{dx} = \delta\rho_w C_w v_y^2, \tag{8}$$

$$-\frac{2P^*}{3}\frac{dh}{dr} = \rho_i f h v_\theta,\ \frac{2P^*}{3\sqrt{3\gamma}}\left(\frac{dh}{dr} + \frac{2h}{r}\right) = -\delta\rho_w C_w v_\theta^2, \tag{9}$$

where $\delta = \pm 1$ specifies ice drift direction.

Length scale L is constructed from the coefficients of Eqs. (8) and (9) according to the formula:

$$L^2 = \frac{2\sqrt{3\gamma}\rho_w C_w P^*}{3(\rho_i f)^2}. \tag{10}$$

Using representative values given in Table 1 and rheological constants given by (5) we find that $L \approx 1.7$ km.

It follows from Eq. (8) that:

$$h = \frac{h_0}{1 - \delta(x - x_0)h_0/L^2},\ v_y = \frac{\delta\rho_i f h}{\sqrt{3\gamma}\rho_w C_w}, \tag{11}$$

where h_0 is the ice thickness at $x = x_0$. Assuming $x_0 = 0$ we find that $h \to \infty$ when $x \to L^2/h$. This situation is not realistic because L^2/h is very large; e.g., $L^2/h = 2890$ km when $h = 1$ m.

Equation (9) are transformed to the form

$$\frac{dh}{dr} = -\frac{3\delta h^2}{2L^2}\left[1 - \delta\sqrt{1 - \frac{8\delta L^2}{hr}}\right],\ v_\theta = \frac{\delta\rho_i f h}{2\sqrt{3\gamma}\rho_w C_w}\left[1 - \delta\sqrt{1 - \frac{8\delta L^2}{hr}}\right]. \tag{12}$$

In the real situation, $8L^2/(hr) >> 1$. In this case, Eq. (12) describe only anticyclonic motion with $\delta = -1$. They are approximated as follows

$$\frac{dh}{dr} = \frac{3h}{L}\sqrt{\frac{2h}{r}},\ v_\theta = -\frac{\rho_i f L\sqrt{2h}}{\sqrt{3\gamma}\rho_w C_w}. \tag{13}$$

Solution of (13) has the following form:

$$h = \frac{h_0}{\left[1 - 3\sqrt{2h_0}(\sqrt{r} - \sqrt{r_0})/L\right]^2}. \tag{14}$$

In the above considered examples, the ice motion is initiated by shear stresses at the periphery and by rotational momentum in the origin in the case of the axially symmetric motion. Ice thickness increases in the direction of the Coriolis force. The

constructed solutions do not reflect any real situations but can explain some features of more realistic solutions describing the wind drift of ice.

Wind Induced Zonal Ice Drift

The steady solution describing wind induced steady ice drift consists of ice strips ES and PS related to the elastic and plastic states, respectively (Fig. 7a). Both of the strips move in the y-direction. The ES is extended from $x = 0$ to $x = \Delta x$, and the PS is extended to the region $x > \Delta x$. Drift speed v_y of the ES is constant, and drift speed of the PS decreases with approaching shoreline (Fig. 7b). Ice thickness h is constant within the ES and increases with the approaching shoreline within the PS: $h = h(x)$, $x > \Delta x$ (Fig. 7b). The ice stresses inside the PS are associated with the coordinates of point O_s in Fig. 6. The friction force at the shoreline is in a balance with the drag forces applied to the ice by the water and air. It is assumed that there are no stresses at the ice edge.

The boundary conditions at line $x = \Delta x$ state that the stresses, ice thickness, and ice velocity are continuous over the line separating the elastic and plastic strips. They are specified by the formulas:

$$\sigma_{xx} = -\frac{2}{3}hP^*,\ \sigma_{xy} = -\frac{2}{3\sqrt{3\gamma}}\delta hP^*,\ \lim_{x \to \Delta x - 0} h = \lim_{x \to \Delta x + 0} h,\ \lim_{x \to \Delta x - 0} v_y = \lim_{x \to \Delta x + 0} v_y, \tag{15}$$

where $\delta = 1$ corresponds to the ice drift to the "north", and $\delta = -1$ corresponds to ice drift to the "south". In the case of ice drift to the "north", the Coriolis force influences ice pressure on the shoreline. In the case of the ice drift to the "south", the Coriolis force is directed offshore.

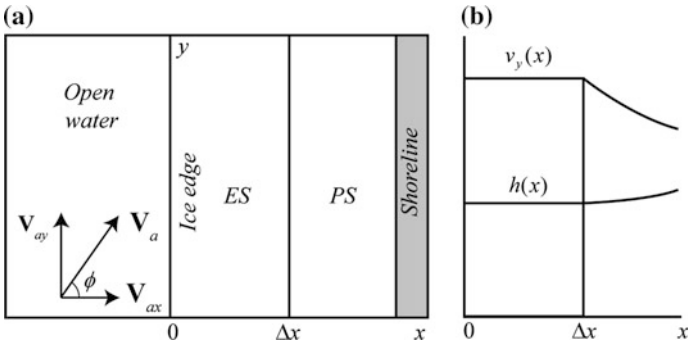


Fig. 7 Configuration of the elastic and plastic regions in steady solutions describing zonal ice drift (**a**). Schematic of ice drift velocity and ice thickness profiles over the elastic and plastic ice strips (**b**)

Since the ice motion is caused by the wind it is assumed that

$$v_y V_{ay} \geq 0. \tag{16}$$

The drift velocity v_y and the ES width Δx are calculated from the momentum balance integrated over the ES width

$$\frac{2}{3} hP^* = \left(\rho_i h f v_y + \rho_a C_a V_a V_{ax} - \rho_i g h \eta_x\right) \Delta x, \tag{17}$$

$$\frac{2}{3\sqrt{3\gamma}} hP^* = \left(\rho_a C_a V_a |V_{ay}| - \rho_w C_w v_y^2\right) \Delta x. \tag{18}$$

It follows from Eq. (17) that there is the only solution describing the ice drift to the "north" when $V_{ax} = 0$ and $\eta_x = 0$. The ES width is found as a root of Eq. (18) solved with respect to Δx after the substitution of v_y from Eq. (17). Locations of the roots are shown by thin lines in Fig. 8 on plane $(\Delta x, \eta_x)$ for different values of the wind speed and direction. The bold lines are described by equations:

$$S = \frac{2}{3} \frac{hP^*}{\Delta x} - \rho_a C_a V_a V_{ax} + \rho_i g h \eta_x = 0. \tag{19}$$

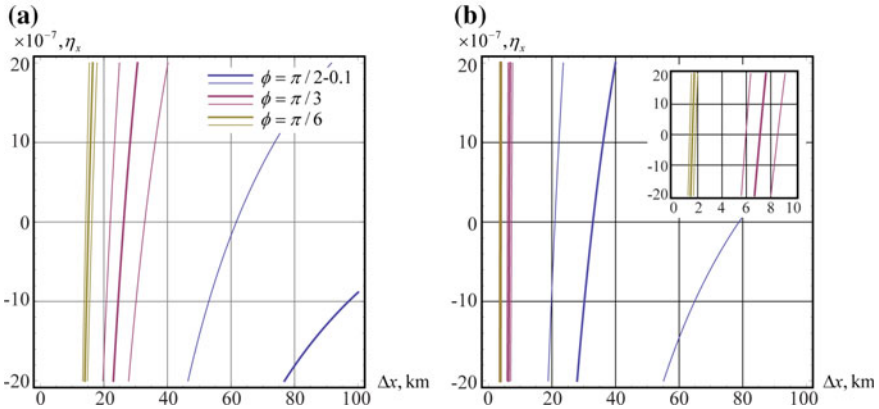


Fig. 8 The ES widths versus the sea surface tilts are shown with thin lines. Thick lines show locations of roots of Eq. (19). Colors of the lines correspond to different angles of the wind direction. Wind speeds are $V_a = 10$ m/s (**a**) and $V_a = 20$ m/s (**b**). The coordinate axes in the inset and figures are similar

Regions where $S>0$ and $S<0$ are located respectively to the left and to the right from the bold line in Fig. 8. They are related to ice drifts to the "north" and to the "south" respectively. Figure 8 show the existence of the roots of Eq. (19) in both regions $S>0$ and $S<0$ when $\phi<\pi/2$. The angle of wind direction ϕ is specified in Fig. 8a.

It follows from Eqs. (17) and (18) that:

$$v_y=\frac{\delta\rho_i hf}{2\sqrt{3\gamma}\rho_w C_w}\left(\sqrt{1+\Delta}-\delta\right),\ \Delta=\frac{4\sqrt{3\gamma}\rho_w C_w\left(\rho_a C_a V_a\left(\sqrt{3\gamma}|V_{ay}|-V_{ax}\right)+\rho_i g h\eta_x\right)}{\left(\rho_i hf\right)^2}. \tag{20}$$

Simple estimates show that $\Delta>60$ when $h\le 1$ m, $V_a>5$ m/s, $\phi\in(0.5,\pi/2)$ and $|\eta_x|<2\times10^{-6}$. In this case, $\sqrt{1+\Delta}\approx\Delta$, and the drift velocity and the ES width are approximated with the formulas

$$v_y=\frac{\delta\rho_i hf\sqrt{\Delta}}{2\sqrt{3\gamma}\rho_w C_w},\ \Delta x=\frac{2hP^*}{\delta h^2P^*/L^2+3\rho_a C_a V_a V_{ax}-3\rho_i g h\eta_x}. \tag{21}$$

It follows from Eq. (1) that inside the PS the ice thickness changes according to the equation

$$\frac{2P^*}{3}\frac{dh}{dx}=\rho_i hfv_y+\rho_a C_a V_a V_{ax}-\rho_i g h\eta_x, \tag{22}$$

where v_y is determined by formula (20).

Further, the analysis of the ice drift to the "north" ($\delta=1$) is performed with more details. Blue and pink lines in Fig. 10 are described respectively by the equations

$$\Delta=0,\ \Delta x=0. \tag{23}$$

The regions, in which conditions $\Delta>0, \Delta x>0$ are satisfied, are located between blue and pink lines shown on the plane (ϕ,η_x) in Fig. 9. The lines are constructed with different values of the wind speed V_a shown in the figures. The effect of ice thickness is visible from the comparison of Fig. 9a ($h=0.5$ m) and Fig. 9b ($h=1$ m). Vertical distances between blue and pink lines constructed with the same values of V_a and ϕ decrease proportionally to the ice thickness. The distances reach maximum when $\phi\approx\pi/2$, i.e. when the wind is parallel to the y-axis, and tend to zero when $\phi\to\pi$ and $\phi\to0$. In the case of the onshore winds with $\phi\in(0.3,\pi/2)$ conditions $\Delta>0, \Delta x>0$ are satisfied with the positive and negative sea surface tilts. In the case when onshore wind is almost perpendicular to the shoreline and $\phi\in(0,0.3)$ these conditions are satisfied only when the sea surface tilt is positive. In the case of offshore winds with $\phi\in(\pi/2,\pi)$ the conditions $\Delta>0, \Delta x>0$ are satisfied only when the sea surface tilt is negative.

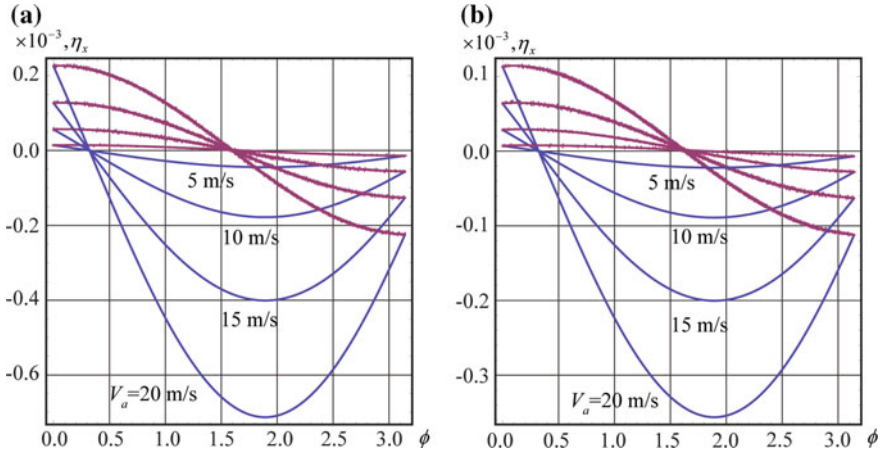


Fig. 9 Blue and pink lines are described by the Eq. (23). Ice thickness $h = 0.5$ m (**a**) and $h = 1$ m (**b**). Wind speed is shown in the figures

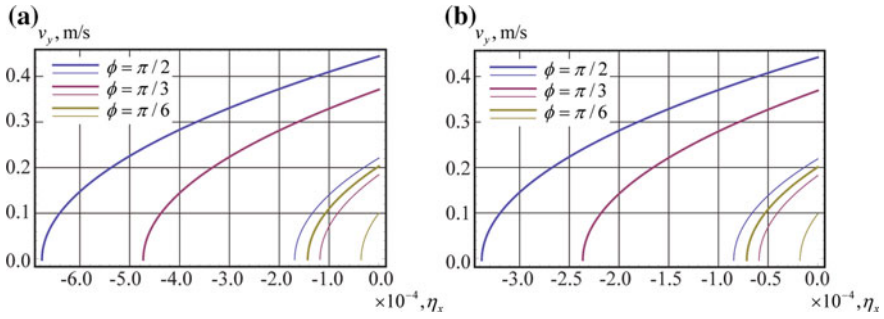


Fig. 10 Drift velocity of ES versus sea surface tilt. Thick and thin lines are constructed with wind speeds $V_a = 20$ m/s and $V_a = 10$ m/s. Ice thickness $h = 0.5$ m (**a**) and $h = 1$ m (**b**). Angles of the wind direction ϕ are shown with different colours

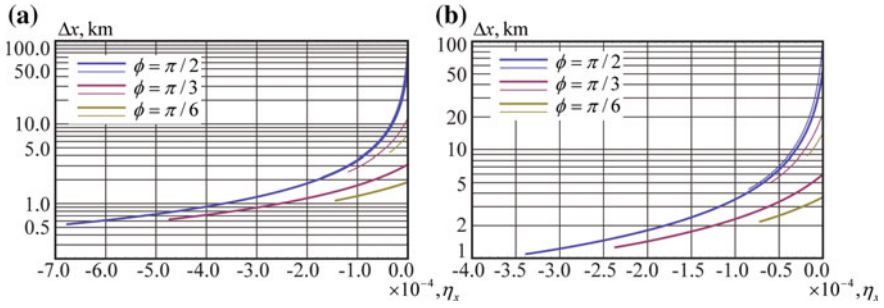


Fig. 11 The ES width versus sea surface tilt. Thick and thin lines are constructed with wind speeds $V_a = 20$ m/s and $V_a = 10$ m/s. Ice thickness $h = 0.5$ m (**a**) and $h = 1$ m (**b**). Angles of the wind direction ϕ are shown with different colours

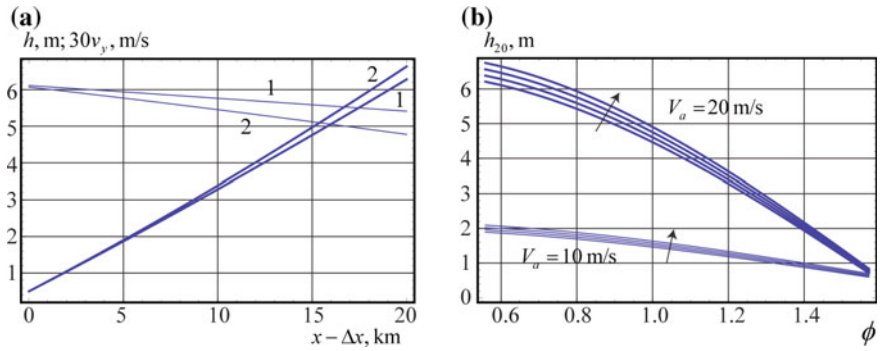


Fig. 12 Ice thickness (bold lines) and velocity (thin lines) profiles inside the PS (**a**). Ice thickness at a distance of 20 km from the boundary between the ES and PS versus the angle of wind direction (**b**)

The ES speed v_y depends only on the rheological constant γ and doesn't depend on P^*. The ES width Δx depends on both rheological constants. The ES speed v_y and width Δx are shown respectively in Fig. 10 and Fig. 11 versus sea surface tilt. They are calculated with different values of the onshore wind speed and wind direction. The ES speed v_y decreases with the decrease of the wind speed and sea surface tilt. The ice thickness increase influences the decrease of the ice ES speed. The ES width increases with the increase of the sea surface tilt and ice thickness and decreases with the increase of the wind speed. The maximum values of the ES speed and width are reached when the wind is parallel to the ice edge.

Profiles of the ice thickness and ice drift velocity over the 20 km distance inside the PS are shown in Fig. 12a. The ice thickness inside the ES is $h = 0.5$ m, the wind speed is $V_a = 20$ m/s, the angle of wind direction is $\phi = \pi/6$. Thin and thick lines correspond to sea surface tilts $\eta_x = 0$ and $\eta_x = -2 \times 10^{-6}$. Figure 12b shows the ice thickness at a distance of 20 km from the boundary between the ES and PS versus the angle of wind direction ϕ. Two sets of the dependencies are constructed with wind speed $V_a = 10$ m/s and $V_a = 20$ m/s. Each of the sets includes dependencies calculated with sea surface tilts $\eta_x = -n \times 10^{-6}$, $n = 0, 1, 2, 3$. Arrows corresponds to the increase of the sea surface tilt.

Wind Induced Circular Ice Drift

Circular ice drift is caused by localised circular wind field or wind gyre. Steady axially symmetric solutions describing circular ice drift consist of the elastic kernel (EK) if the ice rotates as a solid body with angular velocity ω_{ek}, and plastic ring (PR), if the ice is in the pure plastic shear state. There is a shoreline or steady ice in the elastic state outside the PR. The EK is extended from $r = 0$ to $r = R_{ek}$, and the

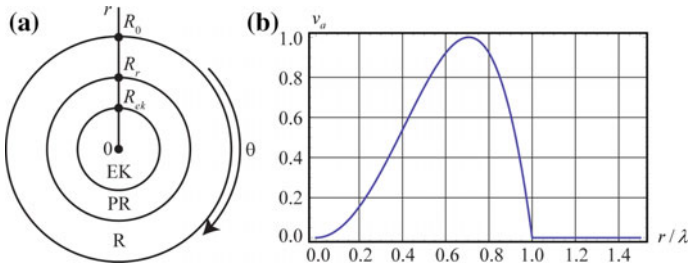


Fig. 13 Configuration of elastic and plastic regions in steady solutions describing circular ice drift (**a**). Schematic of dimensionless profile of wind velocity (**b**)

PR is located by $r \in (R_{ek}, R_p)$ (Fig. 13a). The ice stresses inside the PR are associated with coordinates of point O_s in Fig. 6.

The boundary conditions at line $r = R_{ek}$ state that the stresses, ice thickness, and ice velocity are continuous over the line separating the elastic kernel and the plastic ice ring. They are specified by the formulas

$$\sigma_{rr} = -\frac{2}{3}hP^*, \quad \sigma_{r\theta} = -\frac{2}{3\sqrt{3\gamma}}\delta hP^*, \quad \lim_{r \to R_{ek}-0} h = \lim_{r \to R_{ek}+0} h, \quad \lim_{r \to R_{ek}-0} v_\theta = \lim_{r \to R_{ek}+0} v_\theta, \tag{24}$$

where $\delta = 1$ corresponds to the anticyclonic ice drift in the clockwise direction, and $\delta = -1$ corresponds to the cyclonic ice drift in the counter clockwise direction. In the case of the cyclonic ice drift the Coriolis force is directed from the origin. In the case of the anticyclonic ice drift the Coriolis force is directed to the origin.

Since the ice motion is caused by the wind it is assumed that

$$v_\theta V_{a\theta} \geq 0. \tag{25}$$

Further, it is assumed that $V_{ar} = 0$ and $V_a = |V_{a\theta}|$. The radial and shear stresses are determined by the formula [9]

$$\sigma_{rr} = -\frac{2}{3}P^*h + \frac{1}{2}\rho_i hf\omega_{ek}(r^2 - R_{ek}^2) + \rho_i gh\eta_r(r - R_{ek}), \tag{26}$$

$$\sigma_{r\theta} = -\delta\rho_a C_a r^{-2}\int_0^r V_a^2 r^2 dr + \delta\rho_w C_w \omega_{ek}^2 r^3/5. \tag{27}$$

It follows from the second boundary condition (24) that:

$$G = \frac{2}{3\sqrt{3\gamma}} hP^* - \rho_a C_a R_{ek}^{-2} \int_0^{R_{ek}} V_a^2 r^2 dr + \rho_w C_w \omega_{ek}^2 R_{ek}^3 / 5 = 0. \tag{28}$$

It follows from Eq. (2) that inside the PS the ice thickness changes according to the equation

$$\frac{2}{3} P^* \frac{dh}{dr} = -\rho_i h (f\omega r + g\eta_r), \tag{29}$$

where the angular velocity of ice ω inside the PR is determined by the formulas

$$\omega r = \frac{\rho_i h f}{2\sqrt{3\gamma}\rho_w C_w} \left(1 \pm \sqrt{1+\Delta}\right), \; \Delta = \frac{4\sqrt{3\gamma}\rho_w C_w (\sqrt{3\gamma}\rho_a C_a V_a^2 + \rho_i g h \eta_r - 4P^* h/3r)}{(\rho_i h f)^2}. \tag{30}$$

It follows from the last boundary condition (24) that:

$$\omega_{ek} R_{ek} = \frac{\rho_i h f}{2\sqrt{3\gamma}\rho_w C_w} \left(1 \pm \sqrt{1+\Delta}\right)\Big|_{r=R_{ek}}. \tag{31}$$

Formula (31) gives the values of the ice drift velocity $v_{ek} = \omega_{ek} R_{ek}$ at the periphery of the EK. Positive and negative roots existing at $\Delta > 0$ correspond to the anticyclonic and cyclonic ice drifts. In the case when $\Delta \in (-1, 0)$ both of the roots are positive and correspond to the anticyclonic ice drift. Condition (28) is used for the calculation of the EK radius R_{ek}. We substitute the values of v_{ek} determined by formula (31) into condition (28) and get equations $G_+ = 0$ and $G_- = 0$ for the calculation of R_{ek}. Subscripts "±" correspond to the signs "±" in formula (31).

Analysis of solutions describing circular ice drift is performed with wind velocity specified by formula $V_a = V_{\max} v_a$, where $V_{\max}$ is the maximal wind speed within the wind gyre. The dimensionless wind velocity v_a shown in Fig. 13b is determined by the formula

$$v_a = 4\frac{r^2}{\lambda^2}\left(1 - \frac{r^2}{\lambda^2}\right) r \in (0, L);\; v_a = 0,\; v_a = 0, \tag{32}$$

where λ is the radius of the wind gyre. Numerical simulations were performed at $\lambda = 200$ km.

The roots of equations $G_+ = 0$ and $G_- = 0$ are shown in Fig. 14a with the bold and thin black lines, respectively for three values of sea surface tilt $\eta_r = 0$ and $\eta_r = \pm 2 \times 10^{-6}$. One can see that the EK radii in the anticyclonic drift are greater than the EK radii in the cyclonic drift when the maximal wind speeds are the same. The increase of the sea surface tilt from the negative to positive values influences the increase of the EK radii. The tips of black lines in Fig. 14a are in the vicinity of

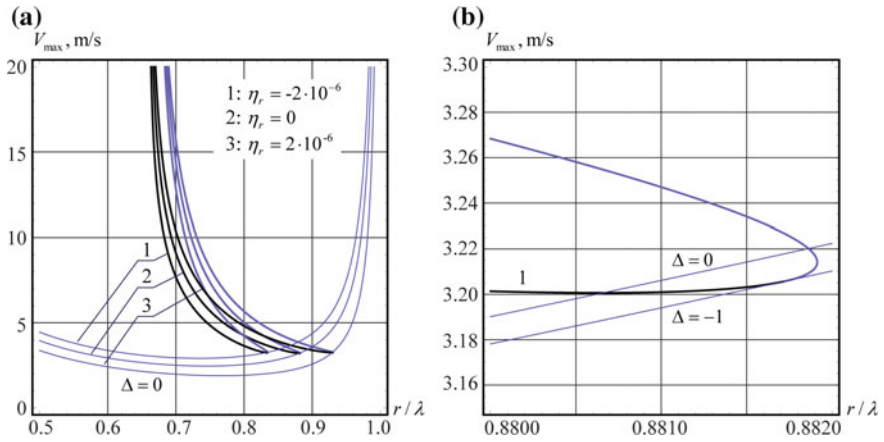


Fig. 14 **a** Thick and thin black lines show locations of the EK radii calculated from equations $G_+ = 0$ and $G_- = 0$ versus the maximal wind speed. Blue thin lines are described by equation $\Delta = 0$. **b** Enlarged picture around the tip of the line 2

the blue lines described by equation $\Delta = 0$. Enlarged Fig. 14b constructed with the zero tilt angle shows that this line crosses both the thin and thick black lines. The thin and thick black lines are connected at the point where $\Delta = -1$. The points of both black lines located in the region $\Delta \in (-1, 0)$ correspond to the anticyclonic ice drift.

Each of the thick black lines in Fig. 14 has a point where $V_{\max}$ reaches minimum. This value is equal to the critical wind velocity, below which ice stresses in the steady-state solution are inside the YC. In the example considered here with the wind field specified by formula (32) and $\lambda = 200$ km, the critical wind speed is below 3.3 m/s.

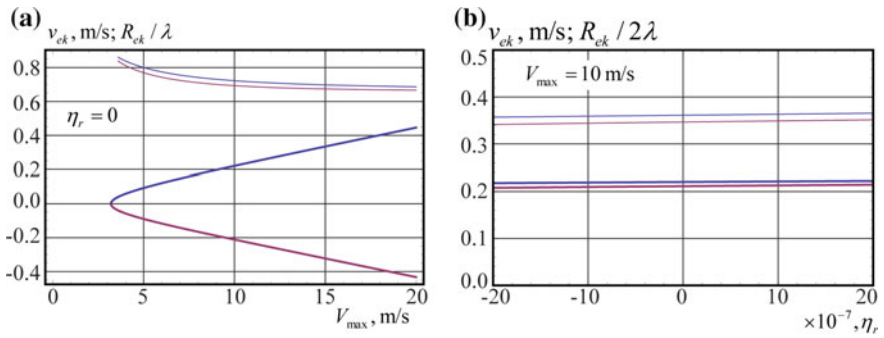


Fig. 15 Dependencies of the EK drift velocities on the maximal wind speed (**a**) and absolute values of the EK drift velocities from sea surface tilt (**b**) are shown with thick lines. Dependencies of the EK radii on the maximal wind speed (**a**) and sea surface tilt (**b**) are shown with thin lines. Blues and pink colors correspond to the anticyclonic and cyclonic ice drift

The EK drift velocities and radii are shown in Fig. 15a versus the maximal wind velocity and in Fig. 15b versus the sea surface tilt. Blue and pink colors correspond to anticyclonic ice drift and cyclonic ice drift. The absolute values of the EK drift velocity increase proportionally to the maximal wind speed. Ice drift speeds $|v_{ek}|$ in the anticyclonic gyre are slightly higher than the drift speeds in the cyclonic gyre. The EK radii R_{ek} in the anticyclonic gyres are also slightly greater than the EK radii in the cyclonic gyres. The EK radii decrease with the increase of the maximal wind speed. The EK radii are slightly greater than the value of the r-coordinate where the absolute wind velocity is maximal. The EK radii and ice drift speeds $|v_{ek}|$ insignificantly increase as the sea surface tilt increases.

Conclusions

Analytical solutions describing steady-state wind drift of sea ice with elastic-plastic rheology are constructed and analyzed. Solutions describing zonal ice drift consist of the elastic ice strip with the solid body motion and the plastic ice strip where pure shear drift are realized. There is the only solution describing the ice drift to the "north" when the wind is parallel to the shoreline and the shoreline is on the "eastern" side of the drift ice. If the wind has onshore component then there are solutions describing both zonal ice drifts in the "northern" and "southern" directions. Depending on the sea surface tilt and onshore wind component the width of the elastic strip changes from 1 to 100 km. The drift speed of the elastic strip is below 45 cm/s, when the wind speed is 20 m/s, and below 25 cm/s, when the wind speed is 10 cm/s. The mean thickness of the ice rubble in the shear zone of 20 km width exceeds 6 m when the speed of onshore wind is 20 m/s.

The analytical solutions describing circular wind drift of ice are investigated in the case when velocity profile of the wind gyre is described by the polynomial function considered in [9]. The solutions describing cyclonic and anticyclonic drift are constructed taking into account sea surface tilt. The solutions consist of elastic circular kernel with the solid body rotation and plastic shear zone of the ring shape. The solutions are localized in the region of the wind gyre. Solutions with the plastic shear zone exist when the maximal wind velocity is greater than the critical value. The general property of the solutions is the increase of the elastic kernel radius, and the decrease of the shear zone width with the decrease of the maximal wind velocity. It is shown that cyclonic and anticyclonic solutions with the plastic shear zone exist both when the maximal wind velocity is greater than the critical value. The influence of the sea surface tilt on the characteristics of the ice drift is much smaller than the influence of the wind velocity. The maximal drift speed at the periphery of the elastic kernel is below 45 cm/s, when the maximal wind speed is 20 m/s, and below 25 cm/s when the maximal wind speed is 10 cm/s.

The comments for the future analysis are as follows. Rheological constants $P^* = 5$ kPa and $\gamma = 3$ give reasonable characteristics of the constructed solutions. More detailed comparison of analytical solutions and observed features of ice drift

may give local corrections of these values. From the analysis of the zonal ice drift it follows that the influence of radial component of wind velocity on the properties of solutions describing circular ice drift can be significant.

References

1. Coon, M. D., Maykut, G. A., Pritchard, R. S., Rothrock, D. A., & Thorndike, A. S. (1974). Modeling the pack ice as an elastic-plastic material. *AIDJEX Bull, 24,* 1–105.
2. Hibler, W. D., III. (1979). A dynamic thermodynamic sea ice model. *Journal of Physical Oceanography, 9,* 815–846.
3. Goldstein, R., Osipenko, N., & Lepparanta, M. (2001). On the formation of large scale structural features. In J. P. Dempsey & H. H. Shen (Eds.), *IUTAM Symposium on Scaling Laws in Ice Mechanics and Ice Dynamics* (pp. 323–334). Dordrecht, The Netherlands: Kluwer Academic.
4. Korsnes, R. (1994). An ice drift series from the Fram Strait January–March 1992 based on ERS-1 SAR data. *Polar Research, 13,* 55–58.
5. Kwok, R. (2001). Deformation of the Arctic Ocean sea ice cover between November 1996 and April 1997: A qualitative survey. In J. P. Dempsey & H. H. Shen (Eds.), *IUTAM Symposium on Scaling Laws in Ice Mechanics and Ice Dynamics* (pp. 315–322). Dordrecht, The Netherlands: Kluwer Academic.
6. Lepparanta, M. (2011). *The drift of sea ice* (347 pp.). Springer, Berlin, Heidelberg.
7. Pritchard, R. S. (1975). An elastic-plastic constitutive law for sea ice. *Journal of Applied Mechanics, 42E,* 379–384.
8. Marchenko, A. V. (1994). A model of a drifting ice cover. *Journal of Applied Mathematics and Mechanics (PMM), 58*(1), 43–58.
9. Marchenko, A. (2013). Axially symmetric solutions of sea ice dynamics models with elastic-plastic and viscous-plastic rheology. In *Proceedings of International Conference* on *Port and Ocean Engineering Under Arctic Conditions*, *POAC13-169* (11 pp.), Espoo, Finland.
10. Rabotnov, Y. N. (1985). *Mechanics of deformable solid body* (712 pp.). Moscow, Nauka (in Russian).
11. Schwaegler, R. T., & Pritchard, R. S. (1980). AIDJEX model response to axisymmetric loadings. In Pritchard (Ed.), *Proceedings of the Arctic Ice Dynamics Joint Experiment. International Commission on Snow and Ice Symposium* (pp. 134–139). Seattle and London: University of Washington Press.
12. Sokolovskii, V. V. (1990). *Statics of granular media* (272 pp.). Moscow, Nauka (in Russian).

Arctic Ocean Modeling: The Consistent Physics on the Path to the High Spatial Resolution

Nikolay G. Iakovlev

Introduction

The Arctic Ocean (AO) is one of the less explored regions of the World because of its geographical location. The keen interest to the AO (and to the Arctic in general) is due both to the purely academic problems: the observed structural changes of the climate system, the polar system stability, the nature of the Polar Amplification [25], and the practical issues of navigation and petroleum/gas production on the shelf. The ongoing structural changes in the polar climate system make the statistical approaches, based on historical data, irrelevant. Despite the impressive progress, the observational net is still too sparse, especially in the deep ocean under ice. That is why the role of numerical models, based on the "first principles", is important in solving the problems of the AO state analysis and forecast.

Multiyear experience of the AO modeling in the frame of the AOMIP (Arctic Ocean Modeling Intercomparison Project, now—FAMOS, The Forum for Arctic Ocean Modeling and Observational Synthesis, http://www.whoi.edu/projects/famos/, [23]) shows, that many of the problems formulated 17 years ago are still on the agenda. For example, the first coordinated experiment, carried out in 1999, showed that not all the modern models were able to simulate the observed salinity structures, and only a few models reproduced the "classical" cyclonic circulation scheme in the Atlantic water layer at the depths of 300–500 m [24]. Now, thanks to the cooperation between modelers and field oceanographers, both models and circulation schemes are refined.

N. G. Iakovlev (✉)
Institute of Numerical Mathematics, Russian Academy of Sciences,
Gubkina 8, Moscow 119333, Russia
e-mail: nick_yakovlev@mail.ru

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_35

On the other hand, the recent modeling experience shows that the model ability to reproduce observed phenomena depends on the spatial resolution, while it is possibly controlled not by the resolution only [1]. It may be treated as the manifestation of the complex nonlinear physics of the ocean and sea ice, which should be investigated more carefully.

Hydrostatic and Non-hydrostatic Ocean Dynamics

In order to look into the causes of the faults in the AO modeling it is worth recalling the basics of the modern large-scale ocean models. Traditionally hydrostatic, Boussinesq, and incompressibility ($div\vec{u} = 0$) approximations are used. Apart of them, the special form of equations is used in the so-called "shallow water approximation" in the z-coordinate system, when "horizontal" motions at the surface normal to the sum of the local gravity+centrifugal acceleration (we denote them by index h) and "vertical" ones in the direction of this sum, are specified. The initial system of equations appears to be specified on the manifold called "cylinder over sphere". A series of simplifications has to be performed to ensure analogs of energy and angular momentum conservation laws.

The common approach is to use the Boussinesq approximation with the constant reference density ρ_c. It may be done if the pressure in the state equation will be redefined $p \rightarrow p_0 = \rho_c g z$: otherwise, the energy will be imbalanced [10]. Moreover, if the deviations of density e and pressure p' from the reference values ρ_c, p_0 are defined, then the system of equations of the ocean hydro thermodynamics (in the terms of potential temperature θ and salinity S) is written as (vector $\vec{k}$ is directed upwards, f is the Coriolis parameter):

$$
\begin{aligned}
&\tfrac{D\vec{v}_h}{Dt} + f\vec{k} \times \vec{v}_h + \tfrac{1}{\rho_c}\nabla_h p' = \vec{F},\\
&\varepsilon_{nh}\tfrac{Dw}{Dt} + \left[\tfrac{g\rho'}{\rho_c} + \tfrac{1}{\rho_c}\tfrac{\partial p'}{\partial z}\right] = \varepsilon_{nh} F_w,\\
&\vec{\nabla}_h \cdot \vec{v}_h + \tfrac{\partial w}{\partial z} = 0,\\
&\rho' = \rho(\theta, S, p_0) - \rho_c,\ p_0(z) = \rho_c g z,\\
&p' = p - p_{0,}\\
&\tfrac{D\theta}{Dt} = Q_\theta,\\
&\tfrac{DS}{Dt} = Q_S.
\end{aligned}
$$

The form of the equation for vertical velocity w differs from the usual used in the large-scale modeling, and "non-hydrostatic" parameter ε_{nh} is introduced. Often in large-scale ocean modeling it is assumed that vertical accelerations are much smaller than the *reduced* gravity acceleration $\frac{g\rho'}{\rho_c}$, and turbulent viscosity is negligible. In this case $\varepsilon_{nh} = 0$, one gets the hydrostatic approach (but it is better to say almost *quasi-static*, because the vertical velocity is not equal to zero). The typical

values in the ocean are $\frac{\rho'}{\rho_c} = O(10^{-3})$; hence, usually, this approach is valid in the major part of the World Ocean. The quasi-static equation is written in square brackets. Sometimes confusion arises because usually primes in density ρ' and pressure p' are omitted, and it looks like the vertical acceleration should be compared with the gravity acceleration, but not with the *reduced* gravity one.

One of the possible sources of modeling errors in this approach is inaccurate approximation of the processes associated with relatively high frequency vertical motions, say with the internal waves. This is important in winter, when the Brunt-Väisälä frequency is small and much less then the internal wave frequency. Note, that on the Siberian shelf in the Kara Sea, near the river estuary, even the annual-mean Brunt-Väisälä frequency exceeds 0.05 s^{-1}, and should be even greater in summer, hence actually quasi-static approximation is valid even for very fine horizontal model resolution.

Traditionally the limit of the quasi-static approximation is also associated with the ratio of vertical scale H to horizontal scale L; it is commonly assumed that this limit is about 0.01. It means that a horizontal resolution of 1 km is the limit to describe the upper layer processes with a vertical scale of 10 m. Our modeling experience shows that this criterion is too rough and the H/L limit may be increased up to 0.1 in the case of strong steady stratification mentioned above. It means that the AO is a very complicated region for high resolution modeling, because various regimes exist there, which should be taken into account in the same frame.

The other problem is associated with the unstable vertical motions like convection, which is extremely important in the high-latitude regions where cooling occurs from above. Usually vertical convection caused by static instability is parameterized with the assumption that convective plumes have horizontal diameters of the order of 100 m, and there are sufficiently many plumes on the model characteristic horizontal scale to be taken into account statistically. In the case when the model grid is fine enough (the size of an individual plume) one may expect that the hydrostatic approximation will be incorrect.

And, the last but not the least source of uncertainty is the choice of turbulent viscosity/diffusion terms $\vec{F}, Q_\theta, Q_S$, which are process dependent. Here, we mention some of such processes.

Mesoscale Eddies

A small diameter of quasi-geostrophic eddies is a characteristic feature of the Arctic Ocean, which may be measured by the Rossby deformation radius. In the deep Canada Basin the baroclinic Rossby deformation radius ($R_{i,n} = \frac{NH}{fn}$, N is the Brunt-Väisälä frequency, H is the ocean depth, n is the vertical mode [9] is about 10 km for the first mode and 5 km for the second one [21], it is even smaller on the Siberian shelf. This contrasts with the mid-latitudes of the World Ocean, where these scales are about 50–100 km, and a model resolution of 15 km is quite enough

to reproduce the eddy induced transport of temperature and salinity, and other tracers correctly [7].

It is quite possible that namely the small Rossby radius (eddy effective diameter) is the cause of the fact that refining of the model resolution over some interval will not lead to the expected improvement in the transport of heat and salt (or other conservative tracers) by narrow jets [1]. If the model resolution is about 5 km, the model spectrum is truncated on the scale of the individual eddy, thus the energy and enstrophy fluxes are treated incorrectly. Thus, an increase in the resolution from 20 to 5 km may lead to even worse results, because the jets are disintegrated into the chains of eddies with poorly approximated dynamics. Finally, on the climate time scales, it leads to the wrong reproduction of the mean large-scale state of the ocean.

In the case of low resolution, when the statistical approach for an ensemble of eddies is valid, some of parameterizations of the so-called "eddy transport of scalar" were developed [4, 27]. The eddy transport of the scalar τ may be associated with the additional velocity $\vec{u}_{GM}$, and may be defined as "skew diffusion" with the anti-symmetric tensor of diffusion coefficients κ_{GM} and K_{GM}:

$$\vec{u}_{GM}\tau = -\kappa_{GM}\mathbf{K}_{GM}\overrightarrow{\nabla}\tau,$$

$$\mathbf{K}_{GM} = \begin{pmatrix} 0 & 0 & -S_x \\ 0 & 0 & -S_y \\ S_x & S_y & 0 \end{pmatrix}.$$

Here S_x, S_y are tangents of the iso-neutral surface slope with respect to the horizontal plane. Coefficient κ_{GM} is usually specified as function of density [27] and some length scale, which may be associated with the Rossby radius. Under the AO conditions with the small Rossby radius this effect seems to be small, but actually as far as the background transports are relatively low, the eddy transport is important on the large (climatic) time scales, being responsible for the large-scale thermohaline features formation. Indeed, this parameterization proved to be helpful in the case of modeling the transport of Atlantic water in the Central Arctic [6] in the case of the low horizontal resolution.

We keep the above mentioned in mind, and expect that the eddy induced transport parameterization will be consistent down to the horizontal resolution of about 20 km. Working with the higher resolution we will get into the zone, in which the eddies will be taken into account twice: explicitly and by parameterization (see Fig. 1). A sufficient resolution would be 1 km in the most areas of the AO, but at this resolution we should expect problems with the adequate description of the *submesoscale* eddies, which we will discuss below.

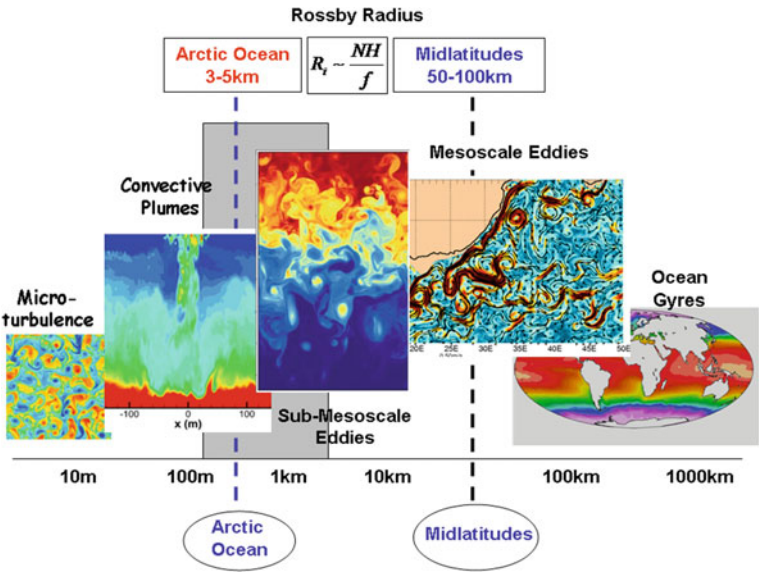


Fig. 1 The scales, typical for the ocean at mid-latitudes and in the Arctic Ocean. The resolutions appropriate to describe submesoscale eddies are depicted by captures in ovals. Note the "gray zone" between meso- and submesoscale eddy diameters in the Arctic Ocean, which roughly coincides with the transition zone from the hydrostatic to non-hydrostatic regimes

Submesoscale Eddies

Recent observations in the upper layer of the Canada Basin [26] show that there are small eddies with the sizes of about 2–4 km and orbital velocities up to 20 cm/s so that the Rossby number is $Ro = \frac{U}{fL} = O(1)$, where U, L are characteristic scales of velocity and length. These numbers obviously differ from the ones in the case of mesoscale eddies, in which the Rossby number is in the range 0.1–0.01. Such eddies are classified as "submesoscale" ones; it was established that they are responsible for the re-stratification of the upper layer during winter. An important feature of these eddies is the relatively high vertical velocity $W \sim 10^{-4}$–$10^{-3} \cdot U \sim$ 1–10 meters per day, in contrast with the mesoscale eddies with the vertical velocities about $W \sim 10^{-4}$–$10^{-3} \cdot U \sim$ 1–10 meters per day.

It is interesting that the very existence of the ocean submesoscale eddies was found mainly due to the numerical modeling and ocean biochemistry satellite remote sensing development [12].

The features of the submesoscale eddies, their role in the large-scale temperature and salinity structures and the possibility of using the existing parameterizations [3] remain absolutely uninvestigated in the Arctic Ocean.

It is important for us to know that the spatial scales of the observed submesoscale eddies at high latitudes are close to the mesoscale eddies (2 and 5 km, respectively). Hence, these two processes cannot be separated and treated independently, by some specific parameterizations valid for each single process, as it is used at mid-latitudes (see the "gray zone" depicted on Fig. 1).

This fact exhibits the *fundamental problem* for the AO modeling with the horizontal resolution of about 1 km. From the physical point of view it will be attractive to model submesoscale eddies explicitly with the model resolution of the order of the first hundreds of meters, but in this case we will get in the *non-hydrostatic* regime and should use another numerical model, not to mention the growing computational cost.

Internal Waves and Tides

The tidal force in the AO is small and the tides are generated by the incident waves, mainly from the Atlantic. The M_2 semidiurnal tide generates the internal tide of the same frequency over the slopes of the bottom topography. Singularity exists in the equation for the normal modes of vertical velocity w [9]:

$$\frac{d^2w}{dz^2} + \frac{N^2(z)}{g}\frac{dw}{dz} + \frac{N^2(z)-\omega^2}{\omega^2-f^2}k^2w = 0,$$

where k is the wave vector and ω is the frequency. The propagation of the internal tide in the form of free wave is bounded by the critical latitude. It means that intense vertical mixing due to the M_2 internal tide collapse may occur at the critical latitude ~75° N. This collapse may be enhanced by the bottom topography as well. Theoretical and modeling aspects of non-hydrostatic nature of the ocean dynamics over steep slopes in the straits were examined in [28].

The other source of internal waves is the decay of mesoscale and submesoscale eddies. Internal waves may be produced directly by mesoscale eddies, along with the submesoscale eddies in the same spatial-temporal scales range [12].

Moving sea ice may also generate internal waves in the AO, especially in summer, when the fresh and warm upper layer exists. These internal waves change air-ocean drag, and their instability leads to enhanced turbulent mixing at the mixed layer [11].

As indicated in [2] who analyzed observations over the Yermak Plateau, the "Dissipation of near-inertial wave energy in the upper ocean is of crucial importance in the Arctic Ocean and must be properly accounted for. If sufficiently elevated, the dissipation can play a significant role in irreversible vertical mixing in the Arctic Ocean, in the cold halocline layer in particular." The mixing should be parameterized and taken into account in the terms of $\vec{F}$, Q_θ, Q_S. Thus, mixing (say, in the popular K-profile parameterization scheme [8]) is specified by background

viscosity and diffusivity with constant coefficients 1 $cm^2\ s^{-1}$ and 0.1 $cm^2\ s^{-1}$, respectively. The AO modeling experience shows that these coefficients should be at least 10 times smaller in the most parts of the Central Arctic to maintain the observed vertical stratification. This conclusion agrees with the theoretical assumptions that the internal waves are generally weak in the central parts of the AO under the pack ice, but it seems to be oversimplified taking into account the recent observational data.

The important research of internal waves both theoretical and observational in the World Ocean [15] and especially in the high-latitude regions were carried out by E.G. Morozov and his coauthors. In a series of papers [13, 14, 16–19] they showed the importance of internal waves not only for the Arctic Ocean mixing, but also for the sea ice formation [20].

The understanding of the role of internal waves in large-scale or climatic Arctic Ocean state formation is still beyond the scope of ocean climate modeling, but it is likely that it will be demanded soon, as the spatial resolution of the models will be fine enough and the requirements to the accuracy will rise due to the improvements in the observing systems in the AO. One may detect that in the models of the general atmosphere circulation and weather forecast there are comprehensive parameterizations employed for a long time both for orographic [22] and non-orographic internal waves (induced by vertical shift of the wind or by convection [5]), which proved to be important for the mean stratosphere wind formation. The experience of ocean modeling shows that it lags behind the atmosphere modeling approximately by 15–20 years, therefore, now we at the very moment to take comprehensive parameterizations of internal waves into account in the ocean models as well. The obvious problem for the parameterization of the internal waves and tides in the AO is that their lengths are close to the scales of mesoscale and submesoscale eddies; hence we again get into the “gray zone” in Fig. 1 and should be aware of double describing the processes.

Summary

Numerical modeling in the Arctic Ocean exhibited an impressive progress in the last 20 years. It happened mostly with the assistance of the ocean general circulation models based on the “traditional” approximations of incompressibility, Boussinesq, and hydrostatics. These models were designed initially for the conditions of low- and mid-latitudes, in which the horizontal scale of mesoscale eddies is large, and the vertical stratification is generally stable. The progress was provided mainly by the *evolution* of the program codes towards higher horizontal resolution. Now, at the resolution of the order of a few kilometers accessible even for the global setups, the *revolution* is at hand. The bunch of mutually interacting processes like mesoscale and submesoscale eddies, internal waves and vertical convection, all approximately of the same horizontal scale, demand the new approaches to model formulation, new parameterizations, and new numerical design. This work may be

done only by means of the close cooperation between theoretical and observational oceanographers, numerical mathematicians and computer programmers.

In this contribution we did not discuss the problems of sea ice modeling, as well as the non-homogeneous air-ice and ice-ocean boundary layers, which also arise on the way to the very fine horizontal and vertical resolution. All the modern large-scale sea ice models, despite the adopted sea ice rheology, are based on the assumption of the "sufficiently large" number of floes on the scale of the model mesh to use the granular media mechanics achievements. This assumption, obviously, becomes wrong for the model resolution of about 1 km, having in mind that the mean floe diameter in the central parts of the AO of about 300 m. The correct formulation of the ocean model in the case of ice ridges with the depths keels up to 25 m is also a challenge.

Acknowledgements The study was performed at the Institute of Numerical Mathematics, Russian Academy of Sciences and supported by the Russian Science Foundation, grant 14-27-00126.

References

1. Aksenov, Y., Karcher, M., Proshutinsky, A., Gerdes, R., de Cuevas, B., Golubeva, E., et al. (2016) Arctic pathways of Pacific Water: Arctic Ocean model intercomparison experiments. *Journal of Geophysical Research Oceans*, *121*, 27–59.
2. Fer, I. (2014). Near-inertial mixing in the central Arctic Ocean. *Journal of Physical Oceanography, 44,* 2031–2049. https://doi.org/10.1175/JPO-D-13-0133.1.
3. Fox-Kemper, B., Ferrari, R., & Hallberg, R. (2008). Parameterization of mixed layer eddies. Part I: Theory and diagnosis. *Journal of Physical Oceanography, 38,* 1145–1165.
4. Gent, P. R., & McWilliams, J. C. (1990). Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography, 20*(1), 150–155.
5. Hines, C. O. (1997). Doppler spread parameterization of gravity wave momentum deposition in the middle atmosphere. Part 2. Broad and quasimonochromatic spectra, and implementation. *Journal of Atmospheric and Solar-Terrestrial Physics, 59,* 387–400.
6. Iakovlev, N. G. (2012). On the simulation of temperature and salinity fields in the Arctic Ocean. *Izvestiya Atmospheric and Oceanic Physics, 48*(1), 86–101. https://doi.org/10.1134/S0001433812010136.
7. Iakovlev, N. G., Volodin, E. M., & Gritsun, A. S. (2016). Simulation of the spatiotemporal variability of the World Ocean sea surface height by the INM climate models. *Izvestiya Atmospheric and Oceanic Physics, 52*(4), 376–385. https://doi.org/10.1134/S0001433816040125.
8. Large, W. G., McWilliams, J. C., & Doney, S. C. (1994). Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. *Reviews of Geophysics, 32,* 363–403.
9. LeBlond, P. H., & Mysak, L. A. (1978). *Waves in the ocean* (p. 602). Amsterdam: Elsevier Oceanographic Series, Elsevier Scientific Publishing Company.
10. Marshall, J., Hill, C., Perelman, L., Adcroft, A. Hydrostatic, quasi-hydrostatic, and nonhydrostatic ocean modeling. *Journal of Geophysical Research, 102*(C3), 5733–5752.
11. McPhee, M. G., & Kantha, L. H. (1989). Generation of internal waves by sea ice. *Journal Geophysical Research, 94*(C3), 3287–3302.
12. McWilliams, J. C. (2016). Submesoscale currents in the ocean. *Proceedings of the Royal Society of London A, 472,* 20160117. https://doi.org/10.1098/rspa.2016.0117.

13. Morozov, E. G., & Pisarev, S. V. (2002). Internal tides at the Arctic latitudes (numerical experiments). *Oceanology, 42*(2), 153–161.
14. Morozov, E. G., & Paka, V. T. (2010). Internal waves in a high-latitude region. *Oceanology, 50*(5), 668–674. https://doi.org/10.1134/S0001437010050048.
15. Morozov, E. G. (1995). Semidiurnal internal wave global field. *Deep-Sea Research, 42*(1), 135–148. https://doi.org/10.1016/0967-0637(95)92886-C.
16. Morozov, E. G., Paka, V. T., Bakhanov, V. V. (2008) Strong internal tides in the Kara Gates Strait. *Geophysical Research Letters 35*(16). https://doi.org/10.1029/2008gl033804.
17. Morozov, E. G., & Marchenko, A. V. (2012). Short-period internal waves in an arctic Fjord (Spitsbergen). *Izvestiya Atmospheric Oceanic Physics, 48*(4), 401–408. https://doi.org/10.1134/S0001433812040123.
18. Morozov, E. G., Kozlov, I. E., Shchuka, S. A., & Frey, D. I. (2017). Internal tide in the Kara Gates Strait. *Oceanology, 57*(1), 8–18. https://doi.org/10.1134/S0001437017010106.
19. Morozov, E. G., Pisarev, S. V., Neiman, V. G., & Erofeeva, S. Y. (2003). Internal tidal waves in the Barents Sea. *Doklady Earth Sciences, 393*(8), 1124–1126.
20. Morozov, E. G., & Pisarev, S. V. (2003). Internal waves and polynya formation in the Laptev Sea. *Doklady Earth Sciences, 398*(7), 983–986.
21. Nurser, A. J. G., & Bacon, S. (2014). The Rossby radius in the Arctic Ocean. *Ocean Science, 10,* 967–975.
22. Palmer, T. N., Shutts, G. J., & Swinbank, R. (1986). Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity drag parameterization. *Quarterly Journal of the Royal Meteorological Society, 112,* 1001–1031.
23. Proshutinsky, A., Steele, M., & Timmermans, M.-L. (2016). Forum for Arctic modeling and observational synthesis (FAMOS): past, current, and future activities. *Journal of Geophysical Research Oceans, 121,* 3803–3819. https://doi.org/10.1002/2016JC011898.
24. Rudels, B., Jones, E. P., Anderson, L. G., Kattner, G. (1994) On the intermediate depth waters of the Arctic Ocean. In O. M. Johannessen, R. D. Muench, J. E. Overland (Eds.), *The Polar Oceans and their role in shaping the global environment*. *Geophysical monograph*. 85: 33–46.
25. Serreze, M. C., & Barry, R. G. (2011). Processes and impacts of Arctic amplification: a research synthesis. *Global and Planetary Change, 77,* 85–96.
26. Timmermans, M.-L., Toole, J., Proshutinsky, A., Krishfield, R., & Plueddemann, A. (2008). Eddies in the Canada Basin, Arctic Ocean, observed from ice-tethered profilers. *Journal of Physical Oceanography, 38,* 133–145.
27. Visbeck, M., Marshall, J., Haine, T., & Spall, M. (1997). Specification of eddy transfer coefficients in coarse resolution ocean circulation models. *Journal of Physical Oceanography, 27,* 381–402.
28. Voltzinger, N. E., & Androsov, A. A. (2016). Nonhydrostatic dynamics of straits of the World Ocean. *Fundamentalnaya i prikladnaya gidrofizika, 9*(1), 26–40. (in Russian).

Numerical Modeling of Internal Wave Generation at High Latitudes

Oxana E. Kurkina, Tatiana G. Talipova, Efim N. Pelinovsky and Andrey A. Kurkin

Introduction

It is well known from the theory of linear waves in the ocean that the propagation of very long tidal internal waves ceases near the critical latitudes [1]. The critical latitude existing due to the Earth's rotation is determined as the latitude φ where the tidal frequency ω is equal to the local inertial frequency $f = 2\Omega_E \sin\varphi$, where $\Omega_E = 0.00007292$ 1/s is the frequency of the Earth's rotation. The linear theory of the baroclinic tide predicts that baroclinic waves of frequency ω cannot propagate as free waves to the north of this critical latitude [1]. This latitude for the semidiurnal tidal period $M_2 = 12.4$ h is 75° N, and only about 30° for the diurnal tidal period. Nevertheless, observations show the existence of short internal waves (IWs) of very large amplitudes and strong baroclinic fluxes here [2–6].

A significant contribution to the study of IWs in the Arctic basin has been done by the Russian scientist Eugene Morozov [7–12]. We reproduce briefly the description of internal waves in the Barents Sea based on the papers [5, 9]. Special measurements of IWs with distributed temperature sensors were carried out from drifting ice

O. E. Kurkina (✉) · A. A. Kurkin
Nizhny Novgorod State Technical University n.a. R.E.Alekseev,
Nizhny Novgorod, Russia
e-mail: Oksana.Kurkina@mail.ru

A. A. Kurkin
e-mail: aakurkin@gmail.com

T. G. Talipova · E. N. Pelinovsky
Federal Research Center Institute of Applied Physics of the Russian Academy
of Sciences, Nizhny Novgorod, Russia
e-mail: tgtalipova@mail.ru

E. N. Pelinovsky
e-mail: pelinovsky@gmail.com

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_36

in 1983 in the Saint Anna Trough and in 1986–1989 in the Franz Victoria Trough. The latitude of the measurements was 80°–81° N and the depths were in the range 200–500 m. In addition to the other phenomena, very short and localized semidiurnal IWs were found. Their lengths were within 2–6 km. They propagated from the slope of the trough and their amplitudes reached 40 m. When the measuring instruments drifted from the slopes of the trough to a distance of 20–40 km toward flat bottom regions, the amplitude of these waves significantly decreased. Under certain stipulations, these semidiurnal waves were classified as internal tides.

The possibility of the generation of internal tidal waves by the topography north of the critical latitude was demonstrated in [13] on the basis of a fully nonlinear non-hydrostatic model. These waves are generated and trapped by the barotropic tidal flow at the lee side of a sill. Such waves are called unsteady lee waves. Their amplitudes depend on the amplitude of the tidal currents and the steepness of the bottom topography. Dramatically high, up to 500 m, internal lee waves were observed in the Luzon Strait and South China Sea [14]. Nonlinear generations of IWs near the critical latitude and north of it were first modeled in the Barents Sea conditions (near Bear Island) in [9, 15]. Propagation of short nonlinear IWs with amplitudes of about 20 m was shown.

The goal of this study is to explain the nature of large-amplitude baroclinic motions observed over the critical latitude in the Franz Victoria Trough, Barents Sea in the framework of the fully nonlinear Euler equations for stratified water, and to forecast dangerous underwater events in this region. Some preliminary results of internal wave modeling in this region were reported in [16]. The numerical model based on 2D Euler equations is described in section "Numerical Model". The results of numerical simulations are presented in section "Numerical Modelling". The results are summarized in section "Summary".

Numerical Model

The numerical model developed by Lamb [17] is applied to study the generation and propagation of IWs in the Barents Sea. The model equations are two-dimensional (vertical section) fully nonlinear Euler equations on a rotating *f*-plane with the Boussinesq approximation. The coordinates are defined as follows: *x*-axis is along the section, *y* is perpendicular to the section, and *z* is the vertical axis. The equations of the model are

$$\begin{aligned}
&\overrightarrow{V}_t + \left(\overrightarrow{V}\nabla\right)\overrightarrow{V} - f\overrightarrow{V}\times\overrightarrow{k} = -\nabla P - \overrightarrow{k}\,\rho\, g,\\
&\rho_t + \overrightarrow{V}\nabla\rho = 0,\\
&\nabla\overrightarrow{V} = 0,
\end{aligned}$$

where $\vec{V}(u, v, w)$ is the 3D velocity vector, but all variables are functions of *x*, *z,* and *t*, ∇ is the 2D gradient operator in the vertical plane (*x*, *z*), subscript *t* denotes the

time derivative, ρ is the density, P is the pressure, g is the gravitational acceleration, $f = 1.425 \times 10^{-4}$ 1/s is the Coriolis parameter at a latitude of 78.5° and $\vec{k}$ is the unit vector along the z-direction.

The equations are transformed to the terrain-following coordinate system (sigma-coordinates). Thus, it possible to increase the vertical resolution over the shallow regions. The equations are solved over a domain bounded below by the topography and a rigid lid above. The flow is forced by the semidiurnal barotropic tidal inflow at the left boundary of the form $V_T \sin(\omega_{M2} t)$, where ω_{M2} is the M_2 tidal frequency with period $T = 12.4$ h. At the right boundary an outflow condition is used. The water column is initially at rest wit a horizontally uniform density $\rho_0(z)$. The horizontal grid size is 90 m and the vertical sigma-coordinate resolution consists of 80 grid points. The time step is tied to the Courant-Friedrichs-Levy condition and varies from 2 to 5 s.

The modeling efforts were performed over a cross-section in the Barents Sea close to the Franz Victoria Trough (Fig. 1). The entire wave path is located north of the critical latitude (74.5° N). The bathymetry along this cross-section was taken from the GEBCO database (solid lines in Fig. 2). Its 8-sin fit is used for a better description of the basin depth, $H(x)$ in the middle portion of the domain is

$$H(x) = \sum_{m=1}^{8} [a_m \sin(b_m x + \varphi_m)], \tag{1}$$

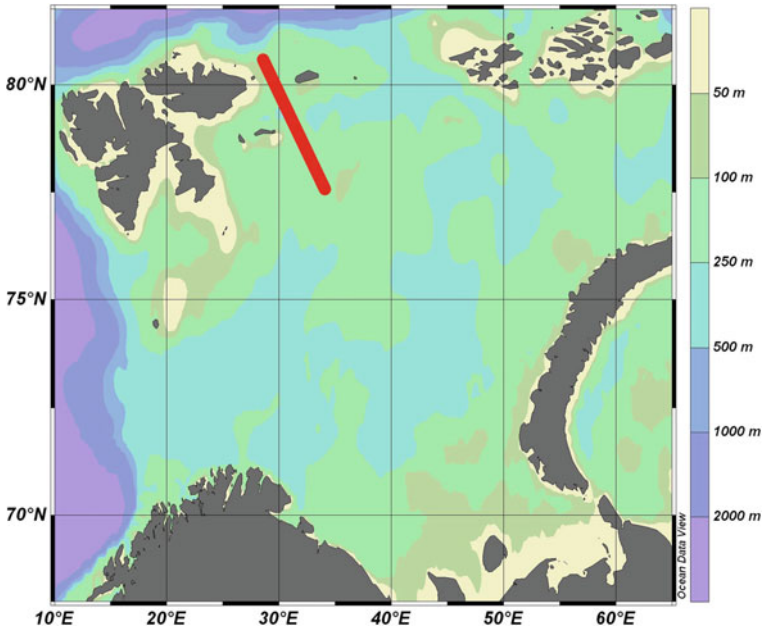


Fig. 1 Bottom topography of the study site with the cross-section between 34.11° E, 77.57° N and 28.61° E, 80.58° N

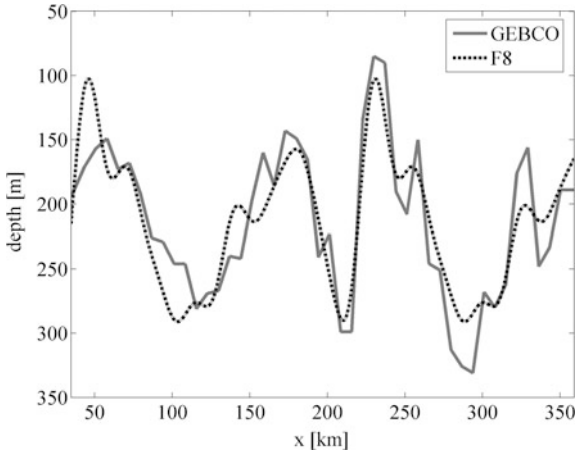


Fig. 2 Profile of the bottom topography along the section (solid line denotes the GEBCO data, the dotted line is the F8 approximation)

where a_m, b_m and φ_m are the fit coefficients. Bottom profile, $H(x)$ is shown by the dotted line in Fig. 2. The double humped sills along the modeling path are similar to the double ridge structure in the Luzon Strait where tidally generated IWs were observed [14].

The typical vertical density profile in this region is taken from the GDEM climatology for July. It is approximated by the bi-exponential function

$$\rho_0(z) = \tilde{\rho}_0 + q_1 \exp(r_1 z) + q_2 \exp(r_2 z),$$

with $\tilde{\rho}_0 = 1000$ (kg m^{-3}), $q_1 = 27.90$ (kg m^{-3}), $r_1 = 2.022 \times 10^{-6}$ (m^{-1}), $q_2 = -1.363$ (kg m^{-3}), $r_2 = -0.03054$ (m^{-1}), and the buoyancy frequency $N(z)$ (Fig. 3) is then calculated.

The study site is characterized by strong tidal currents and underwater mountain ridges. A map of the mean tidal fluxes in the Arctic Ocean (with 8 tidal components) calculated with the help of the Arctic Ocean Tidal Inverse Model (AOTIM) is presented in [18]. Four major tidal components M_2, S_2, K_1, and O_1 contain up to 79, 10, 5 and 1% of the total (8-components) potential energy of the tide, respectively. Thus, the semidiurnal tide M_2 dominates in the tidal velocities in this region of the Arctic Ocean.

The mean values of the tidal flux of the M_2 barotropic tide over the shelf of the Barents Sea are 20–60 m^2/s. The maximal velocities can exceed these values of the mean velocities as much as twice. The barotropic tidal wave in our model numerical experiments is represented only by the semidiurnal tide M_2. As the mean tidal current in the study site is about 0.5 m/s, the maximal barotropic tidal velocity can reach 1 m/s, and we use this value for V_T to set up the model.

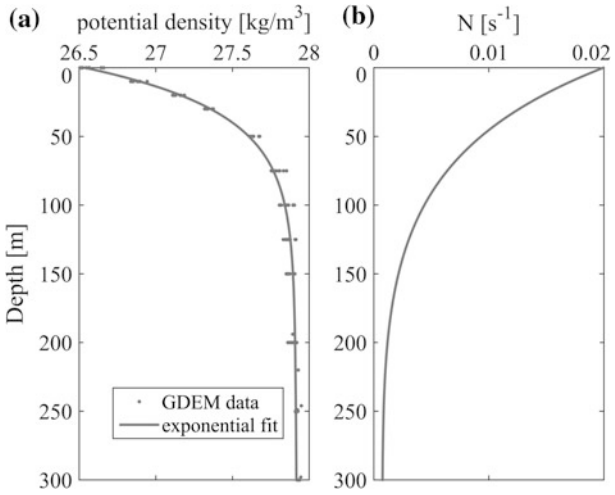


Fig. 3 Typical vertical mean profile of the potential density anomaly (1026–1028 kg/m^3) $\rho_0(z)$ (**a**); Brunt–Väisälä frequency (**b**) in the study site taken from the GDEM climatology for July

Numerical Modelling

The calculated normalized density perturbation fields $(\rho - \rho_0(z))/\tilde{\rho}_0$ and "baroclinic" $u - u_{\mathrm{bt}}$ horizontal velocity values are shown in Fig. 4 for various tidal phases. Linear barotropic tidal velocity is given by relations: $u_{\mathrm{bt}} = V_T \sin(\omega_{M_2} t)$ $H(x_0)/H(x)$. Vlasenko et al. [19] mentioned that full periodicity of the process is formed after 2–3 tidal periods T_{M_2}; therefore, the snapshots for $t \in [3\ T_{M_2}; 4\ T_{M_2}]$ are shown.

Time moment $3T_{M_2}$ corresponds to the end of the ebb phase and the beginning of the flood phase. The velocity of the tide at $3T_{M_2}$ is zero and there are large perturbations of density (and corresponding displacements of isopycnals) over the left (western) side of the largest hill. Intense IWs are generated over the slopes of the highest central hill (located between 200 and 300 km, while the crest is at 230 km). Their amplitudes decrease when they propagate out of this zone. After a quarter of the tidal period, at $t = 3.25\ T_{M_2}$, the tidal flood phase starts; the velocities reach the maxima, and the wave generation intensifies at the right lee slope of the central hill. At time moment $t = 3.5\ T_{M_2}$, the phase of the flood finishes, and the barotropic tidal velocities vanish; the IWs are developed well on the right side of the central hill. When the ebb phase is maximal (at $t = 3.75\ T_{M_2}$), the zone of generation is also shifted to the other (left) lee side of the hill, and the waves gain the largest amplitude at $t = 4\ T_{M_2}$ when the velocity of the barotropic tide vanishes again.

The density perturbation shown in the left column in Fig. 4 perfectly illustrates the phenomena of upwelling and dowelling caused by the interaction of the barotropic tidal flow with the bottom topography. The difference between densities of

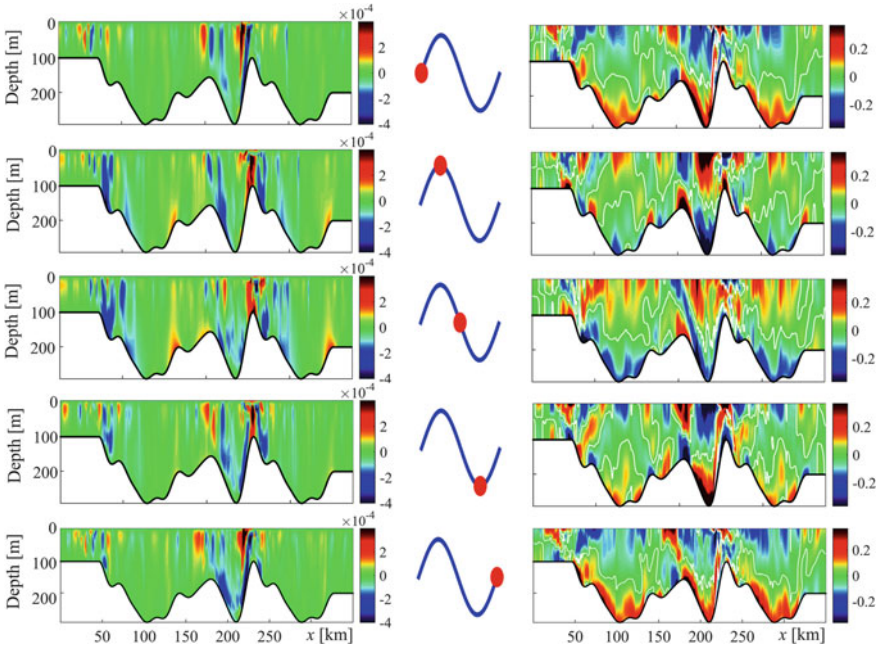


Fig. 4 Snapshots of the normalized density perturbation $(\rho - \rho_0(z))/\tilde{\rho}_0$ (left panel) and perturbation of horizontal velocity $u - u_{bt}$ (m/s), zero contour is shown by white line (right panel) as the reference. The middle panel shows the phase of the barotropic tide

displaced and surrounding (unperturbed) water masses can reach 0.5 kg/m^3, which may correspond to a temperature difference of 5 °C (in the case of fixed salinity).

The baroclinic horizontal velocity zero contours (white lines in the right column in Fig. 4) let one roughly estimate the vertical mode composition of the wave field by counting the number of zeros by the vertical at each point along horizontal coordinate x. The number of modes in the solution increases due to the influence of nonlinearity in the shallower regions of the section.

Thus, the presented IWs are mainly unsteady lee waves that are generated by the tide and ebb currents over the lee slope of the hill with the phase shift of about a quarter of the tidal period between the maximal tide velocity and maximal isopycnal displacement. The wave velocities are evident at time moments 3 T_{M_2}, 3.5 T_{M_2}, and 4 T_{M_2}, when the tidal current is zero. Their values do not exceed 0.4 m/s. The process is well illustrated in the Hovmuller plot (or x–t diagram) for the displacement of the isopycnal located at a depth of 70 m in the undisturbed state (Fig. 5). The more intense IWs with amplitudes of 50 m and total heights (crest-to-trough height) of about 80 m appear regularly in the band of about 40 km wide and the central position approximately at 230 km. Their lengths may be evaluated as 6–12 km using the displacements in Fig. 6.

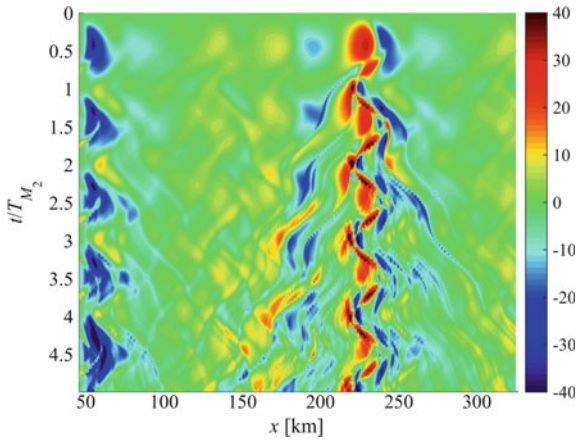


Fig. 5 Contour plot in the space-time domain for the displacement (m) of the isopycnal located at a depth of 70 m in the undisturbed state

The typical frequencies/periods of the modeled waves can be estimated from their amplitude spectrum shown in Fig. 7. The most pronounced components of the wave field are localized in space by the semidiurnal baroclinic motions of the period T_{M_2}, and waves with multiple frequencies $\omega = 2\ \omega_{M_2}, 3\ \omega_{M_2}$, etc. The amplitudes of the spectral peaks of the second and third harmonics are of the order of half of the first harmonic amplitude. The higher harmonics are still visible, but their spectral peaks are much lower and narrower.

Within the context of the linear non-hydrostatic model (e.g., [20]) the oscillatory flow over an uneven bottom produces resonant baroclinic motions of frequency ω, when condition $\alpha \leq \gamma$ is satisfied for inclination α of the characteristic lines of the hyperbolic wave equation

$$\alpha^2(z) = \frac{\omega^2 - f^2}{N^2(z) - \omega^2}$$

and slope γ of the bottom relief $\gamma = dH/dx$. We analyzed this condition for $\omega = 2\omega_{M2}$ in our model setup ($H(x)$ given by (1) and $N(z)$ shown in Fig. 3) and selected intervals of horizontal coordinate x, where $\alpha \leq \gamma$. These 'supercritical' regions, where the motion is essentially baroclinic in the entire water column, and radiated waves can propagate, creating wave "beams" are marked by red circles on the bottom profile in Fig. 8. Though we solve a nonlinear problem, it is clearly seen that the regions selected with the use of linear criterion $\alpha \leq \gamma$, perfectly correlate with the locations of the strongest density perturbations shown in the left column in Fig. 4. Rotation suppresses the generation of IWs at high latitudes; therefore, the amplitude of radiated waves rapidly decreases.

Vlasenko et al. [19] suggest another classification of the processes of IW generation based on the magnitude of the Froude number $Fr = u_{\max}/c$. Here, $u_{\max}$ is the maximal velocity of the barotropic tidal flow in the x-cross-section ($u_{\max} = V_7 H(x_0)/H(x)$) and c is the phase speed of the long lowest-mode IWs. In the

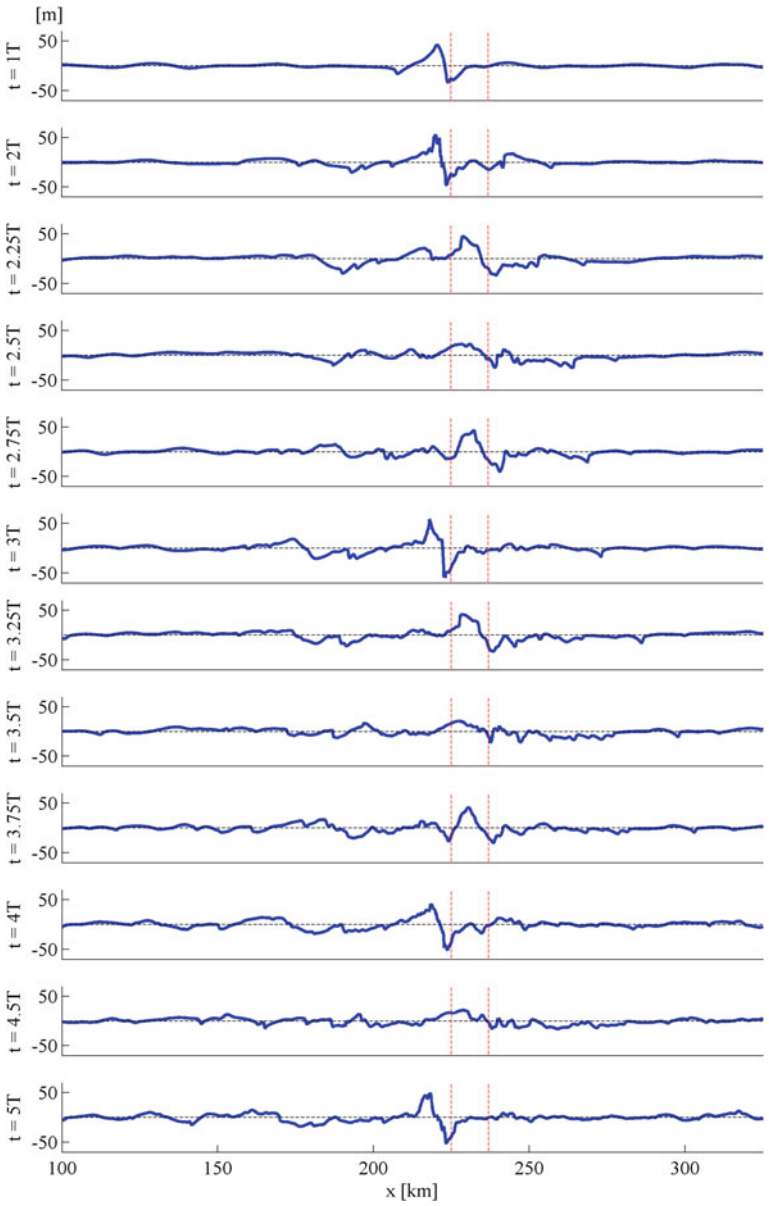


Fig. 6 Displacement of the 70 m isopycnal surface at different time moments

'supercritical' regime $Fr > 1$, the generation of strong nonlinear internal lee waves by tide is possible at any latitude, whereas there is no IW generation when $Fr \ll 1$ in the region north of the critical latitude. Therefore, the waves shown in Figs. 4, 5, and 6 can be interpreted as internal lee waves.

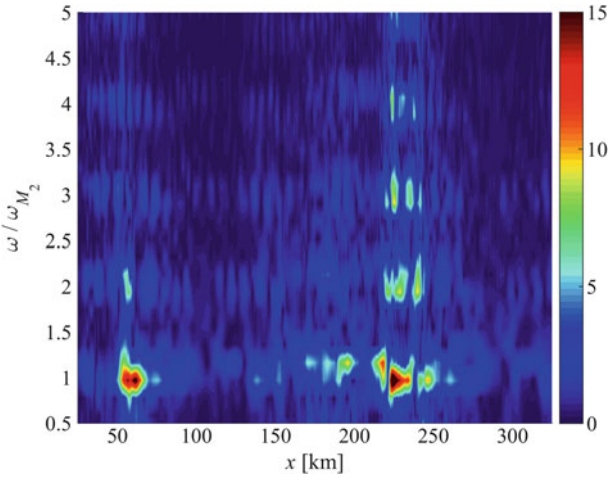


Fig. 7 Contour plot in the space—normalized frequency domain for a single-sided amplitude spectrum (m) of the isopycnal located at a depth of 70 m in the undisturbed state

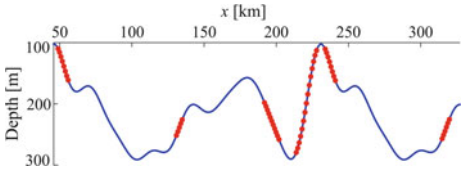


Fig. 8 Supercritical regions (red points) of the sea bottom for wave frequency $2\omega_{M2}$ according to the linear nonhydrostatic theory

The spatial variation of the Froude number along the study section is shown in Fig. 9 together with the other characteristic quantities: linear phase speed of IWs and maximal tidal velocity in the study site. The zone of $Fr > 1$ is located between points at 225 and 237 km. The most intense IWs appeared in the numerical runs exactly within this interval.

It is interesting to note that such a complex fully nonlinear process of the barotropic tide interaction with the bottom topography generates almost symmetric (from the point of view of horizontal direction) field of total (barotropic and baroclinic) near-bottom velocity (see Fig. 10), because the supercritical zones are located over almost symmetrical flanks of the underwater hill (Fig. 8).

Internal wave records are rather irregular, and the waves of large amplitude appear and disappear over a finite time; their number is also highly variable. Therefore, the methods of mathematical statistics can be applied to them, determining the frequency of the occurrence of flow velocities with a value greater than the specified one. In the case of a random ergodic process this quantity is known as $1 - P(u)$, where $P(u)$ is the probability of occurrence of the current velocity values in the range 0–u. Doing so, we separately calculate the positive and negative values of the flow velocity (that is, the different directions of the flow). The variability of this characteristic is shown in Fig. 11. It is clearly seen that the share of high

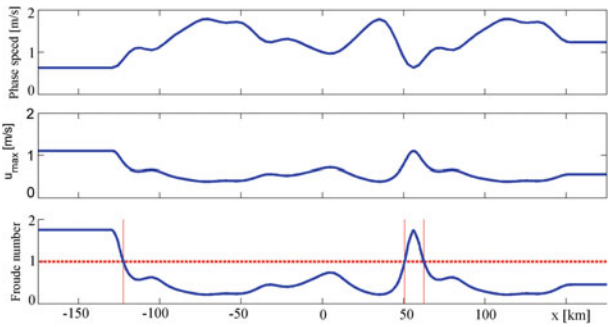


Fig. 9 The Froude number and auxiliary characteristic parameters versus distance along the section

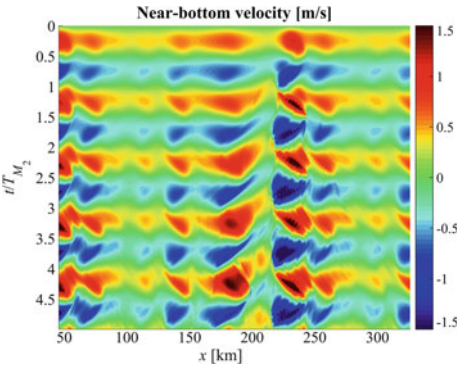


Fig. 10 Contour plot in the space-time domain of the total horizontal velocity near the bottom

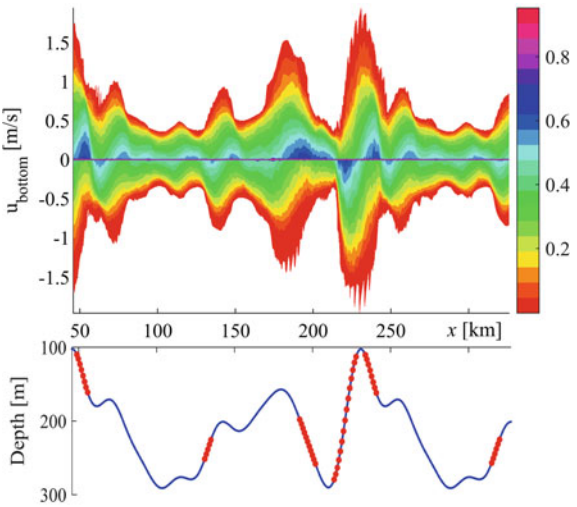


Fig. 11 Frequency of occurrence of large flow velocity values along the *x*-axis

velocities (about 5–10% of all waves) varies in space in the amplitude range 0.5–1.5 m/s, and the frequency of occurrence of flow velocities of different signs is approximately the same at every spatial point.

Summary

The problem of generation of small-scale intense internal waves in the Arctic regions north of the so-called critical latitude is discussed. It is well known, that baroclinic waves at such latitudes cannot propagate as free waves. However, due to the nonlinear generation of higher harmonics, which can propagate, highly nonlinear small-scale internal waves are excited. This process has been studied in the regions of the Franz-Victoria Trough in the Barents Sea and in the Arctic Ocean, where such waves, generated by the interaction of the barotropic tidal flow with uneven bottom, were observed several times (with the participation of Professor Eugene Morozov). The adapted numerical model of K. G. Lamb, solving the two-dimensional Euler equations on the rotating Earth, is used for calculations. Modeling results confirmed that, similarly to the in situ records, 40 m-amplitude waves are generated in this region and their wavelengths (6–12 km) also agree with the observations. These waves conserved significant amplitudes of isopycnal displacements at a distance of about 20 km from the zone of generation. They appear twice per the semidiurnal tide period and conserve their state for about 6.2 h, slowly propagating out of the peak of the underwater hill. More intense waves were generated in the ebb phase and when the ebb velocity vanished. The largest observed height between the crest and trough (about 80 m) appeared at the end of the ebb phase each 12.4 h period. The time evolution of the horizontal velocity field is also modeled, and significant values of flow velocities are obtained, the maximal value is 1.5 times greater than the amplitude of the barotropic tidal flow at the time moments of the maximal tide. A spectrum of isopycnal displacement corresponding to undisturbed density at 70 m depth reveals significant values to the fifth harmonic. The amplitudes of spectral peaks of the second and third harmonics are half of the order of the first harmonic amplitude. The zones of internal wave generation over the flanks of the hills are found using two methods, and both give similar results. The frequency of the appearance of large peak values in the velocity field is also calculated. In particular, the fraction of large velocities in the range 0.5–1.5 m/s (approximately 5–10% of all waves) varies in space significantly growing in the supercritical regions. The frequency of the occurrence of the flow velocities of different signs is approximately the same at each spatial point.

Acknowledgements This study was initiated in the framework of the state task programme in the sphere of scientific activity of the Ministry of Education and Science of the Russian Federation (projects No. 5.4568.2017/6.7 and No. 5.1246.2017/4.6) and financially supported by this programme, grant of the President of the Russian Federation (NSh-2685.2018.5) and Russian Foundation for Basic Research (grant No. 16-05-00049).

References

1. LeBlond, P. H., & Mysak, L. A. (1978). *Waves in the ocean*. Amsterdam: Elsevier.
2. D'Asaro, E. A., & Morison, J. H. (1992). Internal waves and mixing in the Arctic Ocean. *Deep-Sea Research, 39*(2), S459–S484. https://doi.org/10.1016/s0198-0149(06)80016-6.
3. Guthrie, J. D., Morison, J. H., & Fer, I. (2013). Revisiting internal waves and mixing in the Arctic Ocean. *Journal of Geophysical Research: Oceans, 118*(8), 3966–3977. https://doi.org/10.1002/jgrc.20294.
4. Pisarev, S. V. (1995). Internal waves measurements with distributed temperature sensors near Arctic Ocean continental shelf margin. In *Challenges of Our Changing Global Environment, Conference Proceedings*, OCEANS'95 MTS/IEEE. https://doi.org/10.1109/oceans.1995.528542.
5. Pisarev, S. V. (1996). Low-frequency internal waves near the shelf edge of the Arctic Basin. *Oceanology, 36,* 771–778.
6. Støylen, E., & Fer, I. (2014). Tidally induced internal motion in an Arctic fjord. *Nonlinear Processes in Geophysics, 21*(1), 87–100. https://doi.org/10.5194/npg-21-87-2014.
7. Marchenko, A. V., Morozov, E. G., Muzylev, S. V., & Shestov, A. S. (2010). Interaction of short internal waves with the ice cover in an Arctic fjord. *Oceanology, 50*(1), 18–27. https://doi.org/10.1134/s0001437010010029.
8. Morozov, E. G., Paka, V. T., & Bakhanov, V. V. (2008). Strong internal tides in the Kara Gates Strait. *Geophysical Research Letters, 35,* L16603. https://doi.org/10.1029/2008GL033804.
9. Morozov, E. G., & Pisarev, S. V. (2002). Internal tides at the arctic latitudes (numerical experiments). *Oceanology, 42*(2), 153–161.
10. Morozov, E. G., Parrilla-Barrera, G., Velarde, M. G., & Scherbinin, A. D. (2003). The Straits of Gibraltar and Kara gates: A comparison of internal tides. *Oceanologica Acta, 26*(3), 231–241.
11. Morozov, E. G., & Paka, V. T. (2010). Internal waves in a high latitude region. *Oceanology, 50*(5), 668–674.
12. Morozov, E. G., & Marchenko, A. V. (2012). Short-period internal waves in an Arctic fjord (Spitsbergen). *Izvestiya, Atmospheric and Oceanic Physics, 48*(4), 401–408. https://doi.org/10.1134/s0001433812040123.
13. Nakamura, T., Awaji, T., Hatayama, T., Akimoto, K., Takizawa, T., Koho, T., et al. (2000). The generation of large-amplitude unsteady lee waves by subinertial tidal flow: A possible vertical mixing mechanism in the Kuril Straits. *Journal of Physical Oceanography, 30,* 1601–1621.
14. Alford, M. H., Peacock, T., MacKinnon, J. A., Nash, J. D., Buijsman, M. C., Centuroni, L. R., et al. (2015). The formation and fate of internal waves in the South China Sea. *Nature, 521*(7550), 65–69. https://doi.org/10.1038/nature14399.
15. Vlasenko, V., Stashchuk, N., Hutter, K., & Sabinin, K. (2003). Nonlinear internal waves forced by tides near the critical latitude. *Deep-Sea Research, 50*(1), 317–338.
16. Kurkina, O., & Talipova, T. (2011). Huge internal waves in the vicinity of the Spitsbergen Island (Barents Sea). *Natural Hazards and Earth Systems Sciences, 11,* 981–986. https://doi.org/10.5194/nhess-11-981-2011.
17. Lamb, K. (1994). Numerical experiments of internal wave generation by strong tidal flow across a finite amplitude bank edge. *Journal Geophysical Research, 99*(C1), 843–864.
18. Padman, L., & Erofeeva, S. (2004). A barotropic inverse tidal model for the Arctic Ocean. *Geophysical Reseach Letters, 31,* L02303.
19. Vlasenko, V., Stashchuk, N., & Hutter, K. (2005). *Baroclinic tides: Theoretical modeling and observational evidence*. Cambridge: Cambridge University Press.
20. Baines, P. G. (1995). *Topographic effects in stratified flows*. Cambridge: Cambridge University Press.

Internal Wave Frequency Spectrum in the Amundsen Basin of the Arctic Ocean Inferred from Ice Tethered CTD Instruments

Sergey V. Pisarev

Introduction

Reduction of the extent (area of the ocean with at least 15% of sea ice), average age, and thickness of the drifting sea ice are the widely known manifestations of the natural processes, which have received the title "warming of the Arctic". Changes in the characteristics of ice are most pronounced in the summer period, but they are also significant in winter. The change in the characteristics of the drifting sea ice in the Arctic, despite a set of other consequences must apparently cause an increase in the energy of internal waves in the Arctic Ocean. The fact is that the existence of the drifting ice is considered among the two main causes explaining the well known low energy of internal waves in the high-latitude Arctic Ocean when compared with the ice-free ocean [1–3]. The current understanding is that internal wave energy is reduced in the ice-covered basins by (among other causes) the dissipation of internal wave energy in the surface boundary layer immediately below the ice [4, 5].

Recent findings demonstrate the sensitivity of the Arctic Ocean circulation to the background deep-ocean mixing. Results using a large-scale coupled ice-ocean model [6] suggest the appropriate model background mixing in the Arctic Ocean is one order of magnitude lower than in the ice-free oceans. Direct measurements of oceanic mixing are difficult, but decades of research suggest that deep background mixing is a consequence of the dissipation of internal wave energy. The exact nature of this relationship is an open question, but it makes it possible to infer mixing from relatively simple observations of internal wave energy. Consequently, if the ice cover is reduced, we may see increased internal wave energy, mixing, and heat flux in the deep ocean because less internal wave energy would be lost in the

S. V. Pisarev (✉)
Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
e-mail: pisarev@ocean.ru

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_37

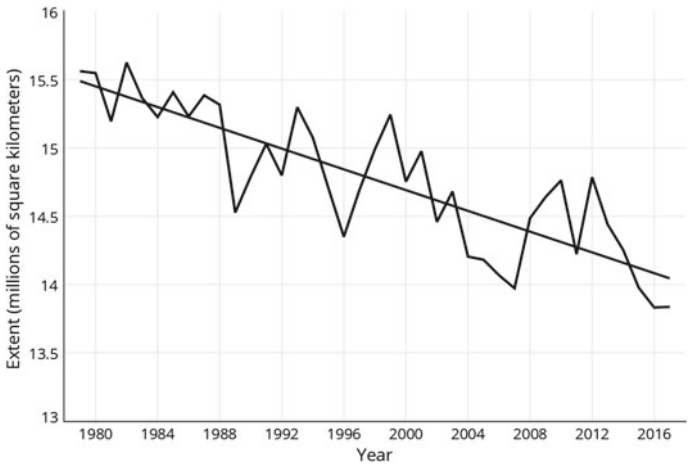


Fig. 1 Arctic Ocean monthly April sea ice extent in 1979–2017 shows a decline of 2.6% per decade.
Credit National Snow and Ice Data Center

under-ice boundary layer. This would tend to result in the increased heat flux to the Arctic sea ice from the warmest Arctic Ocean water masses: Atlantic and Summer Pacific; hence, a positive climate feedback appears that would melt more ice.

To estimate the energy of internal waves of the Arctic in the years, in which the extent of ice clearly decreased and the records of low extent followed one by one, a series of measurements with the use of ice tethered CTD (Conductivity-Temperature-Depth) instruments has been carried out in April of 2007, 2009, 2010, and 2012. The estimates are compared in this research with the estimates of the same type, in the measurements in the 1960s–1980s when the extent of sea ice was definitely larger (Fig. 1) than the so-called climatic norm defined as the average extent in 1981–2010.

Data

The measurements of temperature-salinity (TS) variability at 3 levels within the 60–120 m of the water column as well as vertical TS profiles in the upper 500 m were carried out in the Russian seasonal drifting ice camp in the central Arctic Ocean in 2007, 2009, 2010, and 2012. Observations of the basic meteorological parameters and ice drift accompanied the measurements of the TS characteristic.

All measurements were carried out in April (the winter Arctic season) when the ice camp drifted through the northern part of the Amundsen Basin over the flat oceanic bottom with depths of ~4300 m. The directions of the drifts were in accordance with the Transpolar Drift in 2007, 2009, 2012: from north to

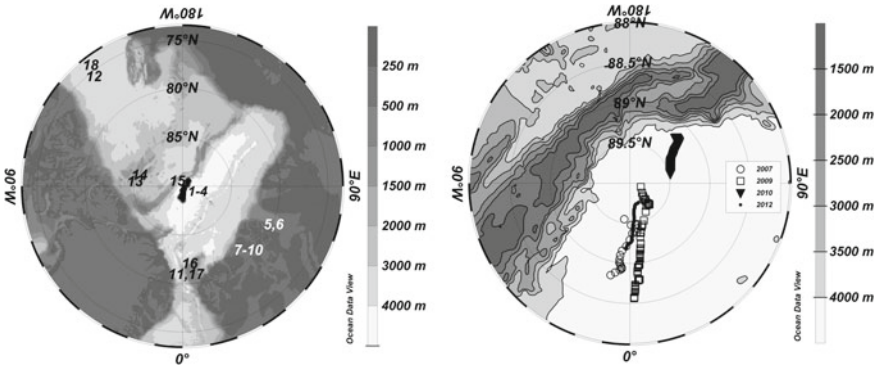


Fig. 2 Maps showing the positions of the ice camp drift (digits 1–4 and black spots on the left map and symbols according to the legend on the right map) together with the bottom relief during the measurements of internal waves in April of 2007, 2009, 2010, and 2012. Numbers 5–18 on the left map correspond to the coordinates of field measurements in 1966–1989. The frequency spectra of internal waves were determined from these data. The estimates of the energy of internal waves based on the field measurements 1–18 are summarized in Table 1

south-southwest. On the contrary, in 2010, the drift was in unusual direction for this region to the southeast (Fig. 2). In all cases, the drift was determined mainly by the wind. The median drift speed varied from year to year within 7–13 cm/s. The maximum values of the drift velocity (40 cm/s) occurred very rarely and for a short time.

Three SBE37 CTD instruments were used to get T and S time series, while the SBE25 profilers were used to measure vertical TS profiles. The SBE37 instruments were vertically distributed over a distance of 25 m between the instruments within the halocline or thermocline. Variability of T or S at specific depths was transformed into vertical displacements of "water particles" using the vertical gradient in the neighborhood of the level of measurements. (It was assumed, that the calculated vertical displacements in the range of frequencies from inertial to buoyancy, are determined by internal waves.) The gradient for the central instrument among the three SBE37 at every time moment of measurements was determined using the values of the upper and lower instruments and from the vertical TS profiles at the nearest time. The vertical soundings were made with irregular time interval but the TS observations at fixed levels were carried out with a constant time interval, which was 1 min in 2010 and 2012, and 3 min in 2007 and 2009.

The distance between the hole in the ice floe, through which the CTD instruments were deployed at three levels, and the hole for vertical soundings varied from 50 to 150 m in different years. While the "moored" T and S variations were transformed into displacements, it was assumed, that there was no horizontal difference in the magnitude of the vertical gradient determined at distances up to 150 m.

Any characteristic, either T or S (or both) were used for the calculation of the vertical displacements. The main criterion for choosing a particular characteristic

was the presence of a linear gradient on the vertical profiles within the water layer 10 m above and below the level that was under the consideration. Part of measurements at the levels was not transformed to vertical displacements and was not used in the calculation of frequency spectra. It happened in the cases when there were such reorganizations of the structure of vertical gradients on consecutive vertical profiles, which eventually formed a nonlinear gradient in the neighborhood of these levels. It should be noted that only about 2% of the records at the levels were subjected to such rejection since the measurements were made within the northern part of the Amundsen Basin where the horizontal TS fields are comparatively uniform because of their location far from the frontal zones or others causes of abrupt changes in the halocline or thermocline.

Some 2009, the data are presented here as an example to demonstrate the results of field observations and their post-processing preceding the calculation of the frequency spectra of internal waves (Figs. 3, 4 and 5). It follows from Fig. 5 that there was no linear gradient in the 50–110 m layer of T profiles, thus the S profile was chosen as the indicator of the vertical motions. The gradient of S was not very linear also; hence, one gradient was used with a decrease of S according to the indications of the middle SBE37 instrument (upward motion of the water particle), and the other with increasing S (downward motion). The vertical profiles of buoyancy frequency in 2009 (Fig. 5) show, that the sampling period 3 min, used for the SBE37, did not allow measuring internal waves with a frequency of 20 cycle per hour (cph), the existence of which is possible in a narrow waveguide, but provided the resolution of waves with the maximum frequencies up to 10 cph.

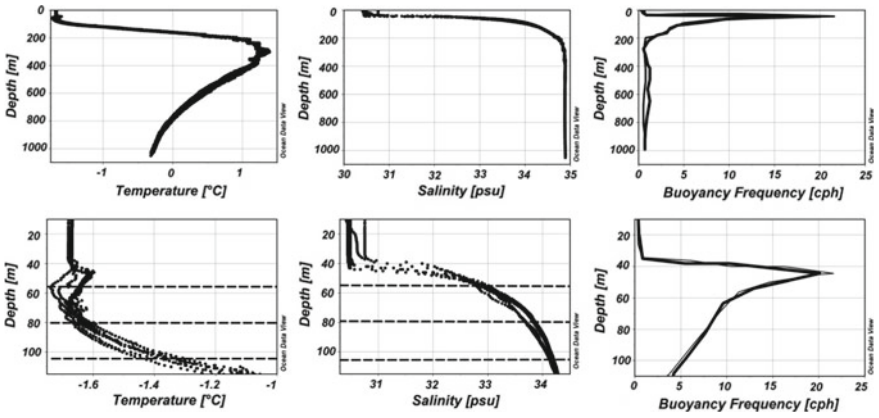


Fig. 3 Eight vertical profiles of T (left column) and S (central column) measured in the internal wave experiment in 2009. Vertical profiles of buoyancy frequency calculated using low-pass filtering from the TS profiles (right column). The top panel demonstrates the full range profiles, while the bottom panel presents the upper part of the profiles, where the measurements at 3 levels (horizontal dotted lines) were carried out

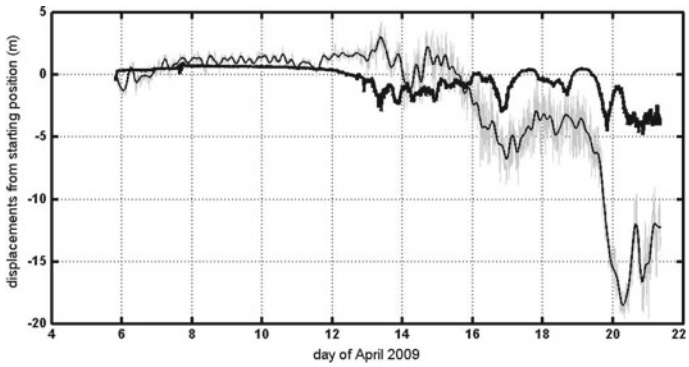


Fig. 4 Vertical displacements from starting position in April 2009 based on the depth of the central SBE37 (black thick line), "water particle" (gray thin line) and "water particle" after low-pass filtering (black thin line)

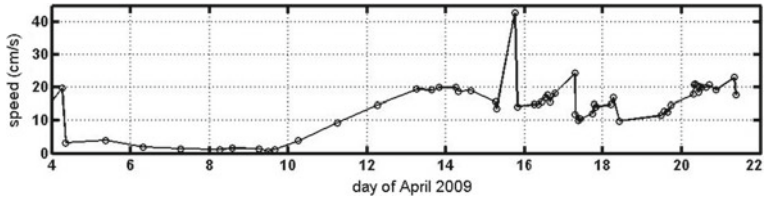


Fig. 5 Ice camp drift speed in April 2009 determined from irregular GPS observations of the geographical coordinates (circles)

It follows from Fig. 4 that the SBE37 instruments were not at the same level, but were subjected to vertical motions. These motions were explained by the deviation from the vertical position of the wire with fixed instruments when an increase in the drift speed of the ice floe occurred (Fig. 5). Despite the relatively small magnitude, the vertical movements, associated with the instrument's own oscillations, were algebraically subtracted from those vertical motions that were calculated using changes in the salinity and vertical salinity gradients to calculate resulting vertical displacements of "water particles".

It should be noted that the amplitude of internal waves increased significantly (Fig. 4) shortly after the drift increased (Fig. 5). Without discussing in detail the possible mechanisms for such an ocean response to increase in the wind and drift, it should be noted, that if we estimate the energy of internal waves only on the basis of measurements before April 13 or only after this date, the two estimates will differ by one order of magnitude.

Frequency Spectra

An observed spectrum of vertical displacement (gray line in Fig. 4) during the period from April 6, 2009 19:30 UTC to April 22, 2009 8:36 UTC is shown in Fig. 6. The Garrett-Munk (GM) model spectrum is plotted in Fig. 6 for comparison. The GM spectrum is a good representation of the internal wave field at lower latitudes away from irregularities of bottom relief or ocean surface. The GM model provides the statistical description of the "typical" internal wave observation because it is empirically derived.

According to Desaubies' formulation of the GM model presented by Munk in 1981 [7]; links in [1], the GM spectra of vertical displacement are given by

$$S(w,z) = \frac{2}{\pi} r \frac{f}{N(z)} \frac{\sqrt{(w^2 - f^2)}}{w^3}$$

where

$S(w, z)$	is the spectral density in m^2/cph;
w, $N(z)$, f	are the current, local buoyancy, and local inertial frequencies in cph;
$r = Eb^2N_0 = 320$ m^2/cph	is one of the two parameters of the GM model in Desaubies' formulation.

Both GM spectra, and hence the total energy, are scaled by parameter r;

$E = 6.3 \times 10^{-5}$	is the nondimensional energy level;
$b = 1300$ m	is the vertical depth scale of N;
$N_0 = 3$ cph	is the buoyancy frequency scale.

The local values of N and f are used to scale the level of the GM spectra plotted in Fig. 6. The values of N at ~80 m were taken from Fig. 2 and found to be near 7 cph. At high frequency, $w \gg f$, the frequency dependence of the GM spectra of vertical displacement is w^{-2}.

Our observed spectrum (Fig. 6) demonstrates both specific particularities of such spectra in the Arctic that was for the first time reported in the second half of the 1980s [1, 2]. The total internal wave energy is lower when comparing with GM while the observed spectral slope is substantially less steeper with the frequency dependence close to w^{-1} at high frequencies and near $w^{-1.5}$ at $0.5f \geq w > f$ The observed spectra of 2007, 2010, 2012 (not shown here) have the same form as in 2009, although at a lower level, especially at the frequencies greater than 1 cph.

With the goal in mind to compare our data with the historical Arctic internal wave spectra, a quantitative measure of the spectral levels observed in 2007–2012 was calculated by fitting the GM spectrum shape to the observed spectra and adjusting parameter r so that the model variance equaled the variance of the data over the frequency band 0.25–1.5 cph (far from the inertial and buoyancy frequencies). The best-fit value for r is 70 m^2 cph for the 2009 spectrum (Fig. 6). The

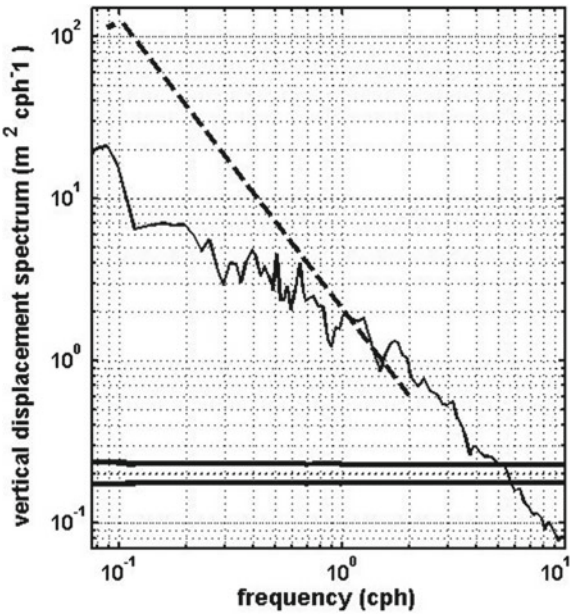


Fig. 6 Spectrum of the vertical displacement at a depth of ~80 m (solid line). Vertical displacement was inferred from salinity by means of dividing by the current vertical salinity gradients. The GM spectrum is shown for comparison (dashed line). The 95% confidence limits are indicated bellow

same type of adjustment was performed for the other observed spectra: the best fits for parameter r were 40, 45, and 35 m^2 cph in 2007, 2010, and 2012, respectively. If we compare the levels of the spectra of the Arctic internal waves measured in 1966–1989 and 2007–2012, it is easy to see that there is no trend in the spectral level (Table 1, Fig. 2). Indeed, because of the causes that are not completely clear yet, the spectra levels determined on the basis of relatively short measurements in different years can differ by one order of magnitude even at close geographic coordinates and under similar meteorological and ice conditions. Bearing this in mind, however, it can be argued that until 2012 there has been clearly no increase in the energy of internal waves in the Arctic.

The same conclusion about the absence of a trend in the energy of internal waves was recently obtained on the basis of another type of measurement [10]. Internal wave energy was estimated from expendable current probes (XCP) shear variance and CTD derived buoyancy frequency. A total of 43 vertical profiles of the currents were used to characterize internal wave energy in April of 2007, 2008, 2010, and 2012, while five vertical profiles were used in the summer of 2007. These data were compared with 11 profiles measured during the MIZEX 83 experiment in June–July 1983, 20 profiles from the Polarstern 87 cruise in July–August 1987 and 40 profiles of the AIWEX experiment in March–April 1987. At least one vertical current profile in 2007, 2010, and 2012 was measured on the same ice floe, on which our observations of internal waves were performed.

Table 1 Parameter r of the GM model adjusted for the recent and historical measurements of the frequency spectra of internal waves in the Arctic (T is the water temperature, S is salinity, V is horizontal velocity)

Number in Fig. 2	Experiment	Data type	Duration (days)	Instrument depth (m)	Bottom depth (m)	References	**r** (m^2cph)
1	SPIW 2007	TS	7.9	75	4300	This article	40
2	SPIW 2009	S	15.6	80	4300	This article	70
3	SPIW 2010	S	5.5	80	4300	This article	45
4	SPIW 2012	TS	11.9	75–110	4300	This article	35
5	SPIW 1983 slope	T	5	100	450	[8]	143
6	SPIW 1983 flat bottom	T	3	100	500	[8]	5
7	SPIW 1985	T	4	80	300–400	[3]	38
8	SPIW 1986	T	6	80	370–460	[3]	140, 160
9	SPIW 1988	T	14	80	220–550	[9]	15–42
10	SPIW 1989	T	15	80	310–450	[9]	3–55
12	Fram -3, 1981	T	4	50–160	800	[1]	40–85
13	T-3 1966	T	4	60	3000	[1]	2, 5
14	T-3 1972	V, T	70	40–60, 200–250	3500	[1]	10, 90
15	LOREX 1981	V	35	1240, 1415	1500	[1]	51
16	MIZEX-83	T	7	15–65	500–1000	[1]	41
17	MIZEX-84	T	<1	0–100	1680–2220	[1]	8
18	AIWEX 1985	T	15	250	3700	[1]	23

Summary and Conclusions

Four series of internal wave measurements were carried out in April 2007, 2009, 2010, and 2012 to determine whether internal wave energy and, hence, deep-ocean mixing has increased with the decline of the Arctic sea ice when the Russian

seasonal ice camp drifted through the northern part of the Amundsen Basin of the Arctic Ocean. The measurements of temperature-salinity (TS) variability were performed at 3 levels within the 60–120 m of the water column as well as vertical TS profiles were measured in the upper 500 m. Observations of the basic meteorological parameters and ice drift accompanied the measurements of the TS characteristic. The total duration of the collected internal wave data was more than 40 days. The results of the recent experiments were compared with 14 previous measurements of the same type that had a total duration of about 180 days. They were carried out in the Arctic Ocean in 1966–1989. The comparison of the data processing performed using the same method by adjusting parameter r of the GM76 model [7] reveal no trend evident over the 40 year period in spite of the drastic decline of the sea ice.

The observed absence of any trend in the energy of internal waves can be explained on the basis of two considerations. First, the existing ideas about the strong dissipation of energy of internal waves just in the layer under ice as an important mechanism that determines the well-known low energy of these waves in the Arctic are not entirely true. In this case, the only cause for the low energy and the specific form of the spectrum of internal waves in the Arctic remains: this is the impossibility of the existence of the M_2 tidal semidiurnal internal waves north of 75° N due to inertial frequency limitation. The importance of the absence of tidal internal waves is evidenced by the fact that the regions of the Arctic, in which the shape of relatively high internal wave spectral level (for the Arctic) corresponds to the GM spectrum (see experiments 5–10 in Table 1), are only those regions, in which the forced waves of tidal periods exist [9]. The second possible cause is that, instead of ice, highly freshened layers of water are formed at the surface. Hence internal waves dissipate within these layers as a result of a not yet determined mechanism.

Acknowledgements The research leading to these results has received funding from the European Union 7th Framework Programme (FP7 2011–2016) under Grant Agreements No.: 265863 ACCESS. R. Schlitzer's Ocean Data View has been used for making Figs. 2 and 3.

References

1. Levine, M. D., Paulson, C. A., & Morison, J. H. (1985). Internal waves in the Arctic Ocean—Comparison with lower-latitude observations. *Journal of Physical Oceanography, 15*(6), 800–809.
2. Levine, M. D., Paulson, C. A., & Morison, J. H. (1987). Observations of internal gravity-waves under the Arctic Pack Ice. *Journal Geophysical Research, 92*(C1), 779–782. https://doi.org/10.1029/JC092iC01p00779.
3. Pisarev, S. V. (1991). Some measurements of the spatial and temporal characteristics of internal waves in an ice-covered high-latitude basin. *Oceanology, 31*(1), 42–46.
4. Halle, C., & Pinkel, R. (2003). Internal wave variability in the Beaufort Sea during the winter of 1993/94. *Journal Geophysical Research, 108,* 3210. https://doi.org/10.1029/2000JC000703.

5. Pinkel, R. (2005). Near-inertial wave propagation in the Western Arctic. *Journal of Physical Oceanography, 35,* 645–665.
6. Zhang, J, Steele, M. (2007). Effect of vertical mixing on the Atlantic water layer circulation in the Arctic Ocean. *Journal of Geophysical Research: Oceans, 112,* C04S04. https://doi.org/10.1029/2006jc003732.
7. Desaubies, Y. J. F. (1976). Analytical representation of internal wave spectra. *Journal of Physical Oceanography, 6*(6), 976–981.
8. Pisarev, S. V. (1988). Experimental frequency spectra of internal waves in an ice-covered high-latitude basin. *Oceanology, 28*(5), 577–580.
9. Pisarev, S. V. (1992). Spatial and temporal characteristics of internal waves at the edge of the continental shelf in the Arctic Basin. *Oceanology, 32*(5), 579–583.
10. Guthrie, J. D., Morison, J. H., & Fer, I. (2013). Revisiting internal waves and mixing in the Arctic Ocean. *Journal of Geophysical Research: Oceans, 118,* 3966–3977. https://doi.org/10.1002/jgrc.20294.

Experimental Studies of Sea and Model Ice Fracture Mechanics

Marina Karulina, Alexey Marchenko, Alexandr Sakharov, Evgeny Karulin and Peter Chistyakov

Introduction

The water motion in the Arctic Ocean is strongly influenced by the ice cover. The modern state of the economical development in the Arctic seas requires our knowledge of the ice properties because of the need for navigation in the ice and especially in the construction of icebreakers. Development of oil and gas industry and widening of the engineering constructions in the open sea (oil platforms) and in the shallow coastal regions (bridges) needs the knowledge of the ice properties to protect these constructions from the ice forcing. This contribution is related to the field and laboratory research of the ice properties at the present day of technology and fracture mechanics.

In the terms of fracture mechanics the natural sea ice is a complex material whose behavior under load is not always predictable with confidence. It applies both to the qualitative description of ice fracture mechanism (failure pattern) and to the quantitative estimate of the breaking load. Sea ice is a complex structural material in view of its specific structure, inhomogeneity and considerable depen-

M. Karulina (✉) · E. Karulin
Krylov State Research Centre, Moskovskoe shosse, 44, 196158 Saint Petersburg, Russia
e-mail: marina@ek17106.spb.edu

M. Karulina · A. Marchenko · A. Sakharov · E. Karulin · P. Chistyakov
Sustainable Arctic Marine and Coastal Technology (SAMCoT),
Centre for Research-Based Innovation (CRI), Norwegian
University of Science and Technology, Trondheim, Norway

A. Marchenko
The University Centre in Svalbard, 156 N-9171, Longyearbyen, Norway
e-mail: aleksey.marchenko@unis.no

A. Sakharov · P. Chistyakov
Moscow State University, ul. Leninskiye Gory, 1, 119991 Moscow, Russia

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_38

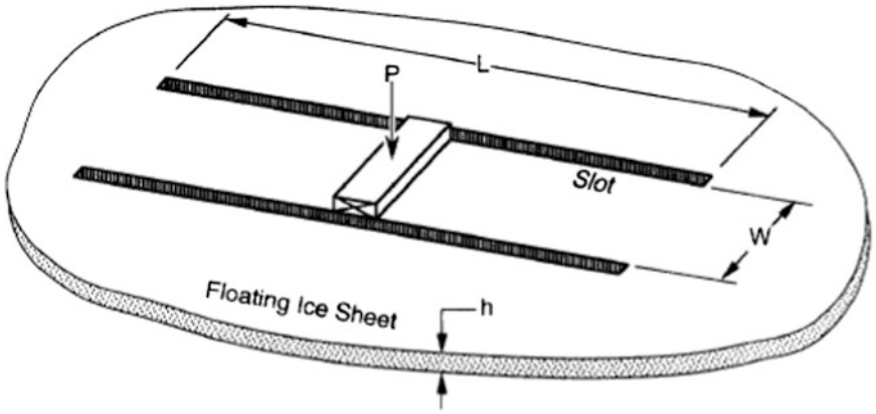


Fig. 1 Test set-up: vertical loading of the ice beam with clamped ends (Sodhi tests)

dence of its strength on external factors, primarily temperature, salinity, porosity as well as time history of ice cover formation [1–6].

One of the interesting tests in ice failure mechanism studies is loading of floating ice beams with fixed ends. The force is applied at mid-length of an ice beam and normally to its upper or side surface. Such tests provide good insights because the ice beam behavior under load is governed by the relationships of various ice strength parameters (flexural, compressive, crushing or shear strength) and the beam geometry.

The first systematic loading tests on beams with fixed ends were carried out by Sodhi [7]. These tests were combined with the other experiments, which were primarily intended for appraisal of ice sheet bearing capacity. The tests were conducted in an ice basin with freshwater columnar ice. The ice sheet was made by seeding ice nuclei over water surface at 0 °C, after that the ice cover was grown to the specified thickness at an ambient temperature of −25 °C. The set-up of Sodhi tests with vertical loading of clamped ice beams is shown in Fig. 1. Ice thickness ranged from 66 to 145 mm, the ice beam length-to-thickness ratio was in the range of $7 < L/h < 44$, while the width-to-thickness ratio was in the range $2 < w/h < 5$. The test results were used to obtain the relationship between the flexural and compressive ice strength.

In 2013–2014, the authors of this contribution conducted systematic loading tests on ice beams with fixed ends in sea ice near the Svalbard archipelago coast and in fresh-water ice of its inland lake. There are no published data on any earlier experiments of similar kind conducted in full-scale ice. The tests were conducted as part of the international SAMCoT project (Sustainable Arctic Marine and Coastal Technology). Unlike the above mentioned Sodhi experiment, the clamped ice beams were mainly loaded horizontally (Fig. 2) so that the visual patterns of crack initiation and propagation were obtained. The geometry of full-scale ice beams was outside the range of Sodhi tests: relative beam length was approximately

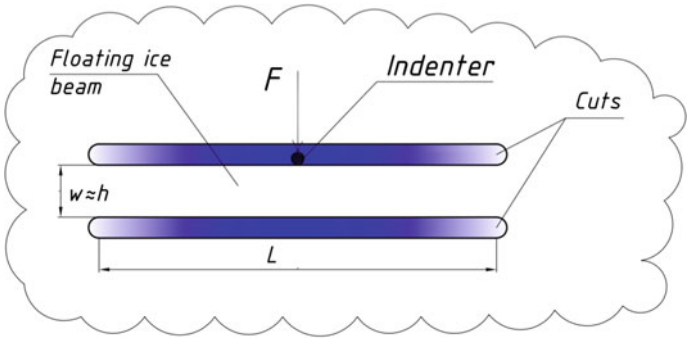


Fig. 2 Test setup for horizontal loading of floating ice beam with clamped ends

$4 < L/h < 15$, while the width-to-height ratio was $1 < w/h < 2$. Some test data and results of analysis for sea- and fresh-water ice were published earlier by the authors in Marchenko et al. [8] and in Sakharov et al. [9]. Marchenko et al. [8] give the first ever description of various beam failure scenarios, while the Sodhi experiment examined only one failure scenario typical for relatively long beams. The failure processes of floating fixed-end ice beams are described and analysed in Sakharov et al. [9], where it is shown how geometrical dimensions of beams influence the failure pattern in case of seawater and freshwater ice.

Following a series of full-scale loading tests on ice beams with fixed ends, which showed a range of various failure patterns, it was decided to carry out similar tests in the model ice in the ice basin and compare the model test results with the field data.

Physical modeling of structure/ice interaction processes in ice model basins is a potential tool for studying ice loads on engineering structures. Preparation of model ice is itself a challenging experimental task [10–14]. Currently there is no technology so far to prepare ideal environment for such experiments, i.e. model ice that would provide ice failure patterns identical to the full scale field conditions. The point is that it is not feasible to simultaneously meet similarity the criteria for different physical and mechanical parameters of ice: flexural strength, compressive strength, crushing strength, shear strength, elasticity modulus.

The goal of this study is to compare failure patterns of the natural sea ice and two model ice types both qualitatively and quantitatively. These investigations are based on the tests of floating fixed-end ice beams loaded at the middle; this study considers horizontal load applied normally to the beam side (Fig. 2). The tests were carried out on sea ice and in model ice of two types: fine granule (FG) and columnar ice. All tests were video filmed, and the video records were synchronized with the force records for further detailed analysis of the processes to relate these to specific load values.

Model Ice Preparation

The tests on model ice were performed in the Ice basin of the Krylov State Research Centre in St. Petersburg, Russia, with dimensions 35.0 m × 6.0 m × 1.8 m.[1] Two types of model ice were used in the tests: (1) FG model ice with the granular structure, and (2) columnar model ice with a thin top layer of granular structure and main layer of the columnar structure. These model ice types reveal various behaviors under identical loading. The ice basin of the Krylov Centre is capable to generate both types of model ice to suit best specific experimental purposes.

The procedure for preparation of fine granule (FG) ice is as follows: a service carriage is running forward and back in a cold room (ambient air temperature is about −20 °C) spraying droplets of brine freezing in the air, which are settling as fine grains layer by layer on the water surface. The spraying is continued until the required ice thickness is reached. The advantage of the FG model ice is that the flexural bending strength of ice can be reproduced in a rather wide range with a close elasticity modulus to flexural strength ratio, which is characteristic of the natural sea ice. Among shortcomings of this method are low values of crushing and compression strength of the model ice (lower than that required according to the similarity criterion) as well as the absence of brittle fracture in some loading conditions.

Unlike the FG ice, the columnar ice is prepared by seeding ice crystals on the water surface to form crystallization nuclei with further growth of vertical columnar crystals at low air temperature. As a rule, this type of model ice is characterized by relatively high compressive strength, which allows brittle failure of ice, but often the ratio between the flexural and compressive strengths and modulus of ice elasticity are different from the full scale field values and, therefore, ice failure patterns do not reproduce the actual processes.

Description of the Test Procedure and the Equipment

In both full-scale and model ice tests two parallel cuts were made in the ice cover to obtain an ice beam with fixed ends (Fig. 2). The distance between these cuts w is roughly equal to the ice thickness, while the beam length in different tests varied approximately from 2 to 8 ice thickness. A horizontal force was applied at the mid-length of the ice beam normally to the beam side. Vertical indenters were used in these experiments: in the full scale tests it was a semi-cylinder indenter with a diameter of 0.15 m, while in the ice basin it was a cylinder indenter of 0.02 m diameter. The indenters were high enough to ensure that the load was applied across the entire ice thickness by the horizontal displacement of the vertical indenter.

[1]The basin has been demounted in 2014.

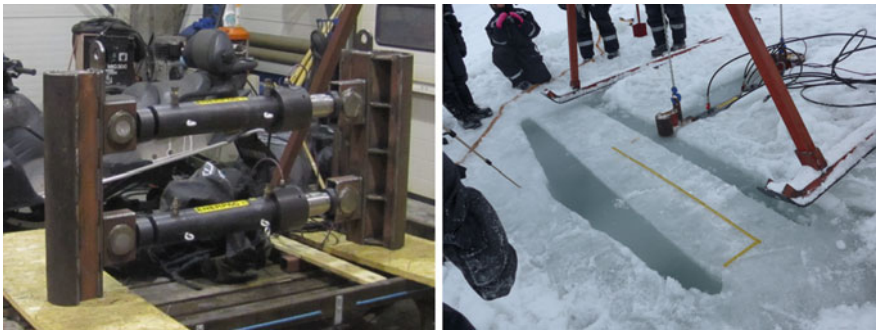


Fig. 3 The rig for tests on sea ice (in the air and submerged into the water before the tests)

Fig. 4 Loading of the ice beam with fixed ends by a vertical cylinder indenter in the ice basin (FG ice)

In the full-scale experiments the indenter was driven by two horizontal hydraulic cylinders of the maximal force about 300 kN equipped with a displacement sensor and a load cell. The measurement system was controlled by a field computer providing data sampling with a frequency of 100 Hz. The rig is shown in Fig. 3, and its detailed description is given in [15].

In the model ice tests the indenter was driven by a towing carriage travelling at a constant speed with the indenter rigidly connected to the carriage via a dynamometer (Fig. 4). The towing carriage speed was 0.005 m/s.

Tests Matrix and Tests Conditions

Table 1 contains the matrix of tests. The scope of work included variation of the beam length to the thickness ratio for all ice types under consideration.

The flexural strength and Young modulus of the sea ice were determined using cantilever beam bending tests when breaking force and the beam vertical

Table 1 Tests matrix

Ice type	Beam#	Thickness h (m)	Length L (m)	Width W (m)	W/h	L/h
Sea ice	Sea1_L	0.530	4.050	0.500	0.9	7.6
	Sea2_M	0.550	3.000	0.500	0.9	4.6
	Sea3_S	0.420	1.600	0.500	1.2	3.8
Model ice/FG fine granule	FG1	0.050	0.345	0.050	1.0	6.9
	FG2	0.050	0.260	0.050	1.0	5.2
	FG3	0.050	0.180	0.050	1.0	3.6
Model ice/columnar	Col3	0.053	0.355	0.045	0.8	6.7
	Col4	0.053	0.260	0.045	0.8	4.9
	Col5	0.053	0.100	0.060	1.1	1.9
	Col6	0.053	0.175	0.052	1.0	3.3

Table 2 Ice properties (mean values)

Ice type	Flexural strength (kPa)	Indentation pressure (kPa)	Indentation strain rate (1/s)	Young Modulus (MPa)	Stress intensity factor (kPa $m^{0.5}$)
Sea ice	345.0	4000	3.5E−03	1200	115–200
Model ice FG fine granule	15.6	55.0	1.25E−01	16.2	0.18
Model ice columnar	26.6	528.8	1.25E−01	24.0	5.84

displacement were measured. Model ice flexural strength was also determined using cantilever beams. The displacement of infinite ice plate under the weight of specified mass was measured for the determination of the Young modulus. The mean values of the ice properties are given in Table 2. Effective strain rate during indentation tests is calculated as $\dot{\varepsilon} = U/2D$, where U is the horizontal velocity of the indenter, D is the indenter diameter.

The stress intensity factor K_{1c} was found using the method described in the KSRI reports [16, 17]. A cantilever beam with a cut in the root section was loaded by a horizontal force applied to the beam end (Figs. 5 and 6). The force P that broke the beam was recorded. The stress intensity factor was found from relation $K_{Ic} = \frac{2PL\sqrt{\pi}}{h_i(W-a)^{1.5}}$, where L is the beam length, W is its width, a is the length of a cut in the beam root.

The stress intensity factor was not determined for the sea ice by the described procedure. In accordance with the data from various sources [2, 6] K_{Ic} for the one-year sea ice is within 115–200 kPa $m^{0.5}$.

The sea ice salinity and temperature averaged through the ice thickness were 6.0 ppt and −4 °C, respectively. The temperature and salinity profiles across the model ice thickness were not determined. The mean salinity of the model ice was 5.0 ppt.

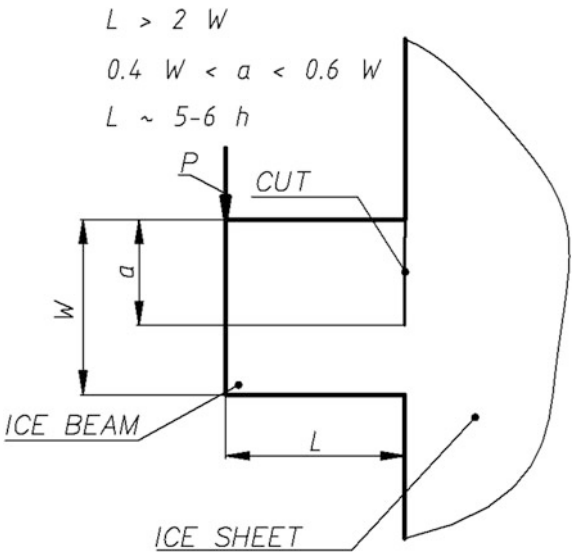


Fig. 5 Diagram for estimating the stress intensity factor

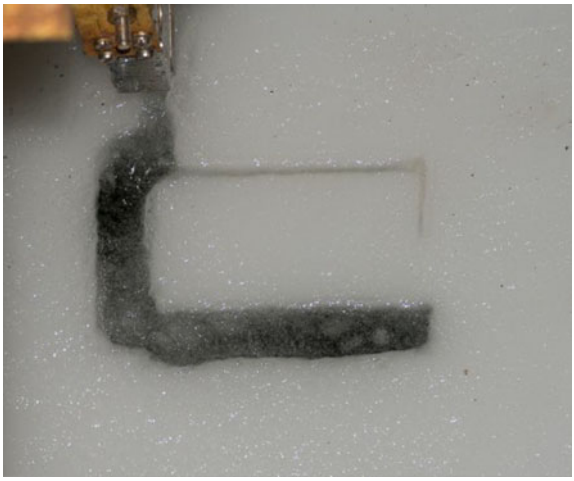


Fig. 6 An ice sample prepared for measuring the stress intensity factor before the test

The ice properties of various ice types under consideration are given in Table 2.

The above mentioned parameters determined for each type of the ice enabled us to find a relationship between the ice beam failure pattern, its geometry, and ice strength characteristics.

Ice Failure Modes

Ice beams loaded as described above in the set up (Fig. 2) are characterized by different failure patterns, the main of them are considered below.

Crushing on indenter. The first phase of the ice beam loading is penetration of the indenter and local fracture (crushing) of ice in the indenter contact area. Figure 4 shows this process during the tests in model ice of the FG type. The crushing force is estimated from

$$F_{ind} = p_{ind} \cdot A \tag{1}$$

where p_{ind} is the ice pressure averaged over the ice/indenter contact area, A is the nominal contact area calculated as product of the ice thickness and the indenter diameter. Table 2 gives the pressure on indenter based on indentation tests in the infinite ice field for each ice type.

Flexural cracks. When the ice beam is loaded, flexural cracks may occur due to the tensile forces at the mid-length and root sections of the beam (Fig. 7a). The process of flexural cracking is described by Sodhi [7] for vertical loading of the beam with fixed ends, and by Sakharov et al. [9] for the horizontal loading of the beam. For

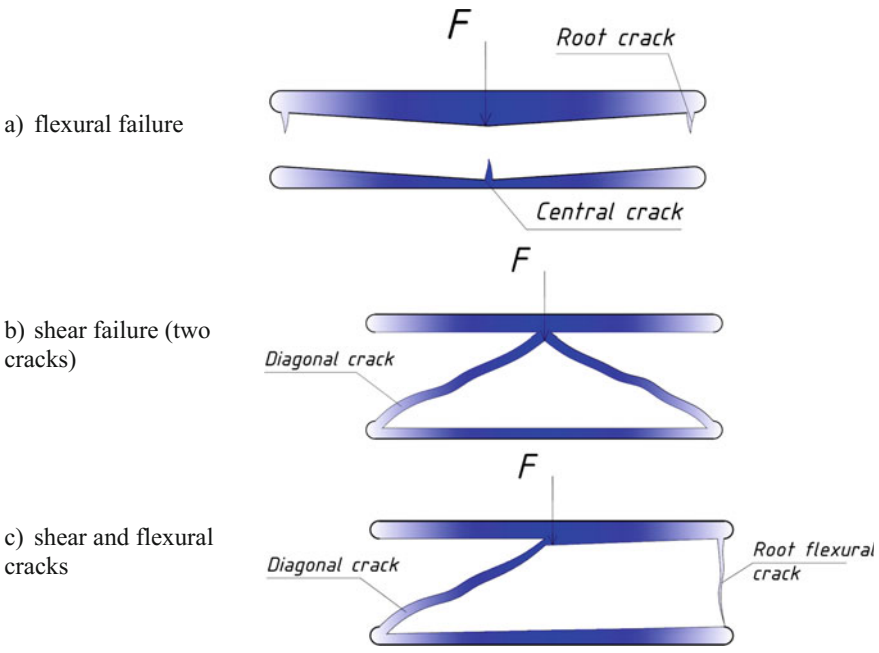


Fig. 7 Various failure types during the tests

estimation of the force causing flexural cracks according to the test shown in Fig. 7a, a formula from the beam bending theory can be applied [18]:

$$F_f = \frac{4\sigma_f h w^2}{3L} \tag{2}$$

where σ_f is the flexural strength of ice.

Table 2 gives the ice strength found from cantilever beam tests with vertical loading for each ice type. In formula (2) the flexural strength corresponds to the horizontal loading of beams.

Ice compression and crushing. Propagation of flexural ice cracks across the entire width is prevented by compression areas formed at mid-length and root sections as can be seen in Fig. 7a. Based on the analysis of the vertical loading tests with fixed-end ice beams, Sodhi [7] suggested a formula for estimating the maximum bearing strength of a beam with compression areas:

$$F_c = \frac{4\beta(1-\beta)\sigma_c w h^2}{L} \tag{3}$$

where σ_c is ice compression strength. Parameter β characterizes the length of the compression area across the beam width. Sodhi [7] estimated $\beta \cong 1/3$ based on the tests of vertical loading beams with fixed ends. Formula (3) is applied in the analysis of ice compression strength in the case of horizontal loading of beams using corresponding replacement of the beam width with ice thickness and of the ice thickness with ice width. Parameter β for each test was determined using photo- and video records.

Diagonal shear cracks. When tangential stresses in some sections of the beam reach their ultimate values (ice shear strength limit) the ice beam is subject to shear failure across these sections and a block of ice is pressed out. In the tests discussed here, different shear cracking patterns are observed: two diagonal cracks propagating from the mid-length section of the beam to the fixed end corners (Fig. 7b); one formed diagonal crack combined with flexural crack in the corner (Fig. 7c); and diagonal cracks not reaching the fixed end corner. The force giving rise to cracking as in the tests shown in Fig. 6b can be estimated from:

$$F_{shear} = 2\tau_s w h \tag{4}$$

where τ_s is the ice shear strength.

Not all types of ice failure modes mentioned above are represented or identified in each individual test. Combinations of these ice failure modes are possible. Prevalence of one or another failure process at each time moment is determined by the ratio of ice strength limits (compression, crushing, tensile, shear) as well as beam proportions: failure occurs when the limit in one mode is exceeded.

Therefore, diverse behaviour of the beam under load can be expected for different ice types, as well as different influence of beam dimensions on failure patterns.

It should be noted that both sea and model ice have non-homogeneous structure and non-uniform properties across the ice thickness. In this study the influence of this factor on the results is not analyzed. Since this study is using averaged integral characteristics (referred to the total ice thickness), we can expect that if the effect due to ice non-uniformities across thickness is taken into account, it might somewhat change the quantitative estimates, but the conclusions will remain principally the same.

The Test Results

Sea Ice

The tests in the sea ice have generally confirmed the conclusions made in the earlier study [9]. In the test with loading relatively long beams ($\frac{L}{h} = 7.6$ and $\frac{L}{h} = 5.5$), the flexural cracking in the mid-length and root sections of the beam were observed (according to the scheme in Fig. 2a), then the load level was governed by the ice compression process. Figure 8 shows a typical time history of force for a relatively long beam $\frac{L}{h} = 7.6$. Local peaks on initial stage of the beam loading correspond to the formation of flexural cracks.

In the tests with the sea ice, parameters β were measured (Figs. 9 and 10); they appeared equal ca. 0.5 for the beam with $\frac{L}{h} = 7.6$, and 0.4 for the beam with $\frac{L}{h} = 5.5$.

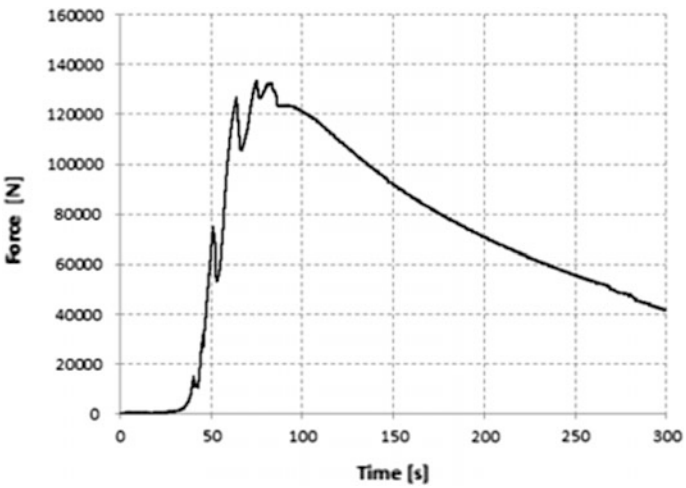
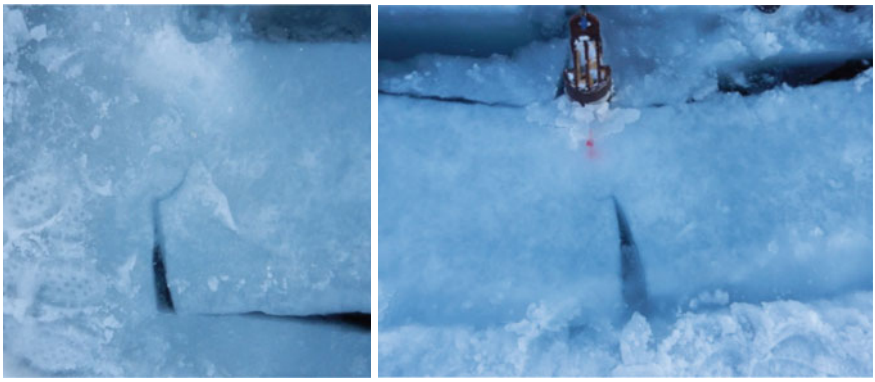


Fig. 8 Time history of loading force for the beam #Sea1_L, $\frac{L}{h} = 7.6$ (Table 1)



View of flexural crack in one of the fixed end of the beam

View of flexural crack in the middle

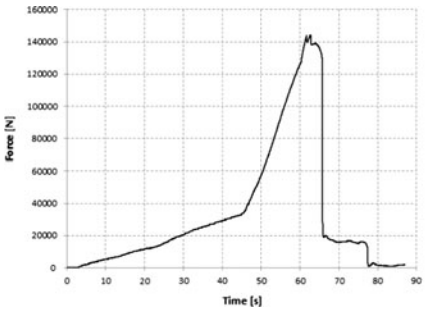
Fig. 9 Fracture pattern after the test. Beam #Sea1_L, $\frac{L}{h} = 7.6$ (Table 1)

Fig. 10 Fracture pattern after the test. Beam #Sea2_L, $\frac{L}{h} = 5.5$ (Table 1)

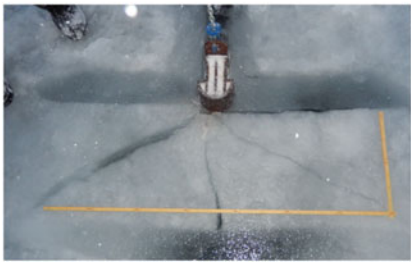
These values differ from the value $\beta \cong 1/3$, which was determined by Sodhi [7] from similar laboratory tests with longer beams.

In the tests with loading of a short beam ($\frac{L}{h} = 3.8$), the flexural cracking and diagonal cracking due to the shear processes were identified. Video record of the test and force time history (Fig. 11a) shows that flexural and diagonal cracks were formed practically simultaneously: two corresponding peak forces are approximately equal to 144 kN. Then, a small shelf can be seen in the force record with a value of 139 kN when governing force is the ice compression. The beam view after failure is shown in Fig. 11b.

From the force record and using formula (4) the sea ice shear strength can be estimated as $\tau \cong 0.343$ MPa.



Time history of force



View of the broken beam after the test. The cracks due to shear propagate from the middle to corners in the fixed ends

Fig. 11 Results of the test on sea ice. Beam #Sea3_S; $\frac{L}{h} = 3.8$ (Table 1)

FG Model Ice

Tests of the FG model ice beams with length ratios of $3.6 \leq \frac{L}{h} \leq 6.9$ have revealed two characteristic features of their failure patterns:

(1) ice crushing under indenter forcing in the initial phase of loading;
(2) absence of diagonal cracks and shear.

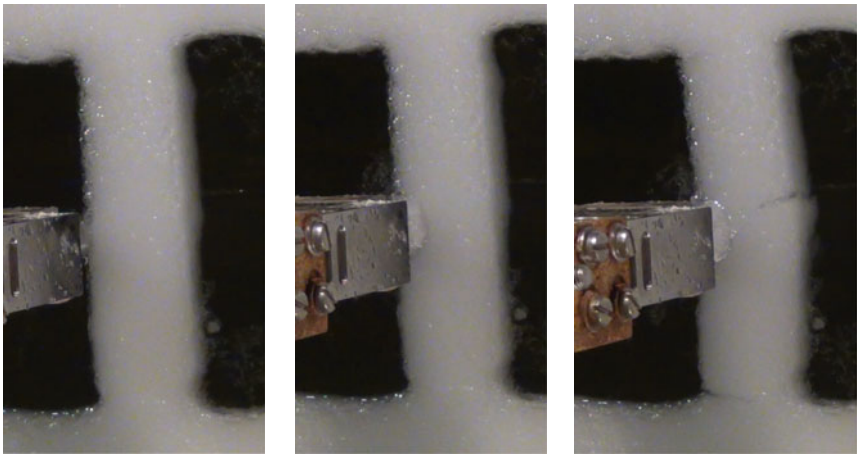
The first feature is related to lower crushing strength of the model ice as compared to the required value according to the similitude law. In the case of relatively long beams the indenter penetration (until flexural cracking) lasted almost 0.5 s; the tests with a shorter beam ($\frac{L}{h} = 3.6$) showed a vivid penetration pattern during 2.5 s, as can be seen from the load time history (Fig. 12) and the force record (Fig. 13).

In all cases, the failure pattern involved flexural cracking at mid-length and root sections of the beams followed by compression at further deformation of beams. All the failure processes are illustrated in Fig. 14, in which broken beam #FG2 is shown.

Values of parameter β are based on the visual data (photos and video records). The values are $\beta \cong 0.4$ for the relatively long beam ($\frac{L}{h} = 6.9$) and $\beta \cong 0.5$ for the other two beams.

The ice strength values given in Table 4 show that:

- the ice flexural strength under horizontal loading of the beam is much smaller than the one under vertical bending of the beam;
- the ice compressive strength is too low, and the ratio of compression strength to flexural strength is close to unity.



Initial stage of penetration of indenter
Time moment: 15.5 sec

Indentation process
Time moment: 17.7 s

Flexural cracks in mid-section and in cornes
Time moment: 18.1 s

Fig. 12 Loading and failure of beam #FG3

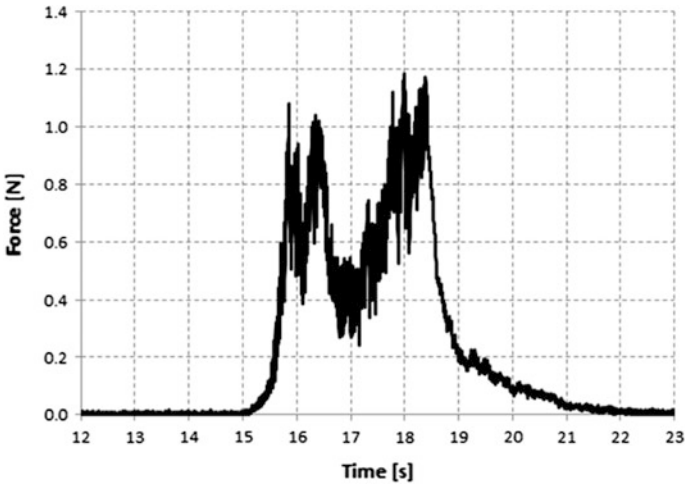


Fig. 13 Time history of force. Beam #FG3

Columnar Model Ice

The tests of the columnar model ice beams have shown three typical patterns of the beam failure depending on the relative lengths of the beams: failure pattern of long beam ($\frac{L}{h} = 6.7$) featured flexural cracks (Fig. 15a); medium length beams ($\frac{L}{h} = 4.9$ and $\frac{L}{h} = 3.3$) showed a combination of cracking types: flexural crack at a fixed end

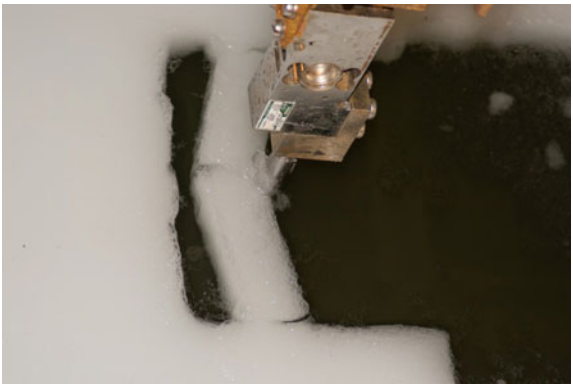


Fig. 14 View of broken beam #FG2 after the test

and diagonal shear crack (Fig. 15b); short beam ($\frac{L}{h} = 1.9$) failed after diagonal shear cracking (Fig. 15c).

In comparison with the processes observed in sea ice tests, it can be noted that compression at beam deformation after initiation of the flexural cracks played a significantly less important role. It is explained by the fact that the model ice in these tests was strongly non-uniform across its thickness, i.e. the failure processes were mainly governed by fracturing of the hard surface layer of small thickness. At cracking (flexural or shear) the surface layer was not only fractured but also flaked, and a softer underlayer started to take the load. From the force records of beam tests with #Col3 ($\frac{L}{h} = 6.7$) and #Col4 ($\frac{L}{h} = 4.9$) shown in Fig. 16 it can be seen that compression forces are much less than the forces causing ice failure due to bending or shear.

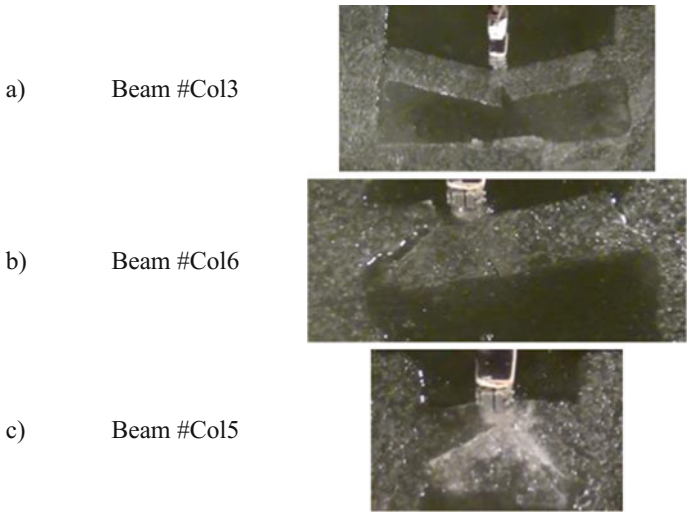


Fig. 15 Failure patterns of columnar model ice

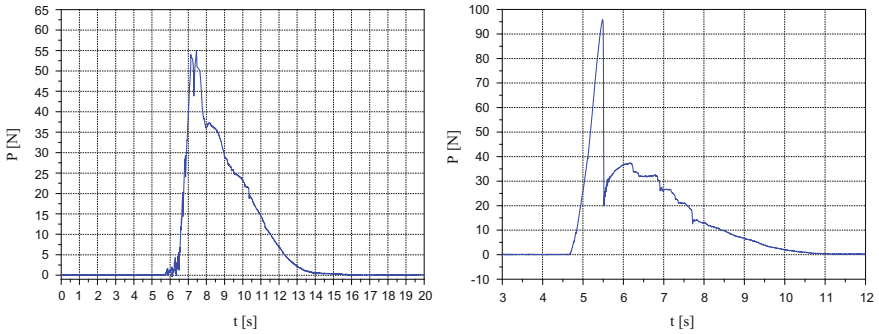


Fig. 16 Examples of force records during the beam tests in columnar model ice

Analysis of the Tests Data

The forces measured in the experiments with reference and various types of ice beam failure are given in Table 3. The table presents only those force values that were identified in the joint analysis of force records and videos of the ice beam loading and failure processes. Table 4 contains the estimates of ice strength parameters using formulas (1)–(4). It can be noted that not all parameters are determined in each test because the formulas were used only when the ice beam failure pattern matched the scenario behind the formula.

Table 3 Failure forces

Ice type	Beam#	F_{ind} (N)	F_f (N)	F_c (N)	F_{shear} (N)	Failure modes
Sea ice	Sea1_L		15300	133660		Flexural cracks in middle and in roots, compression
	Sea2_M		29500	158810		
	Sea3_S		143800	139000	144000	Diagonal shear cracks
Model ice/ FG fine granule	FG1		0.72	0.55		Crushing on indenter, flexural cracks in middle and in roots, compression
	FG2	0.65	0.91	0.95		
	FG3	1.05	1.12	1.19		
Model ice/ columnar	Col3		54.0	37.0		Flexural cracks in middle and in roots, compression
	Col4		97.0	37.0	97.0	Flexural crack at fixed end and diagonal shear crack
	Col6	179.0	230.0		230.0	Flexural crack at fixed end and diagonal shear crack
	Col5	240.0			260.0	Diagonal shear cracking

Table 4 Strength parameters of ice (according to beam tests with fixed ends)

Ice type	Beam#	p_{ind} formula (1) (kPa)	σ_f formula (2) (kPa)	σ_c formula (3) (kPa)	τ_s formula (4) (kPa)	$\frac{\sigma_c}{\sigma_f}$
Sea ice	Sea1_L		351	4085		11.6
	Sea2_M		483	3609		7.5
	Sea3_S				343	
Model ice/FG fine granule	FG1		1.5	1.6		1.1
	FG2	0.7	1.4	2.0		1.4
	FG3	1.1	1.2	1.7		1.4
Model ice/ columnar	Col3		134	122		0.9
	Col4		176	140		0.8
	Col6	169			42	
	Col5	226			41	

In view of the above-mentioned specifics of columnar ice modeling, the flexural and shear strength values given in Table 4 should be referred to the upper consolidated layer, while the compressive strength should be referred to the lower layer.

It is interesting to compare the data in Table 4 with the ice strength properties obtained by conventional methods (Table 2). Both tables give flexural ice strength parameters and indenter pressures, though not at all tests.

We emphasize the following results:

- The sea ice flexural strength values are quite close both under the vertical loading of a cantilever ice beam and under horizontal loading of a clamped beam despite different test procedures applied. The tests performed by the authors to find the flexural strength of seawater ice in the horizontal plane using cantilever ice beams have not revealed any major differences.
- It was found in the tests with fine-grained ice that the ice flexural strength in the horizontal plane decreases substantially (almost 10 times) as compared with the vertical flexural strength, while the model columnar ice feature a substantial (almost 4 times) increase in this parameter. It should be taken into account that different test procedures were applied during vertical and horizontal bending of ice beams, and conclusions should reasonably be drawn from the data obtained by similar test procedures. However, this issue was beyond the scope of the study.
- Lower indenter pressures for the model ice presented in Table 4 as compared with Table 2 are generally explained by the differences in the boundary conditions: free boundaries in the case of indenter penetration into an ice beam and without such boundaries in the case of boundless ice sheet.

- Compressive strength of sea ice is in the range of 3.6–4.1 MPa and practically coincides with the indenter pressure (4.0 MPa) given in Table 2. It should be noted that in the clamped ice beam case we have constrained compression rather than uniaxial compression.
- The last column of the table contains compressive/flexural strength ratios found from formulas (3) and (2), respectively. It is seen that in the sea ice this strength ratio is close to 10, while in the model ice it is about 1.

Ice Failure Diagrams

Failure modes of various ice types observed during the tests are presented as the plots of conventional strength parameters versus relative beam lengths (Figs. 17, 18 and 19). The conventional strength parameter is defined as the ratio of the failure (governing) force applied to the beam cross-section area $\frac{F}{wh}$ where F is the indentation force (F_{ind}), bending (F_f), compression (F_c), or shear (F_{shear}) forces. The same beam is subject to more than one failure mode because these failure processes occur at different moments of time. The conventional strength parameter allows comparison of failure patterns for various ice types, and the failure forces can be compared without a link to the absolute dimensions of the beams and the failure mode.

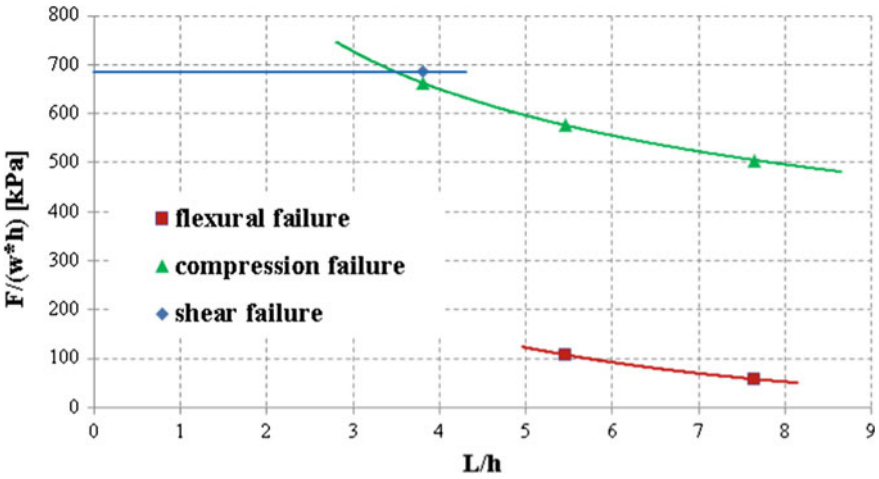


Fig. 17 Beam failure diagram in sea ice

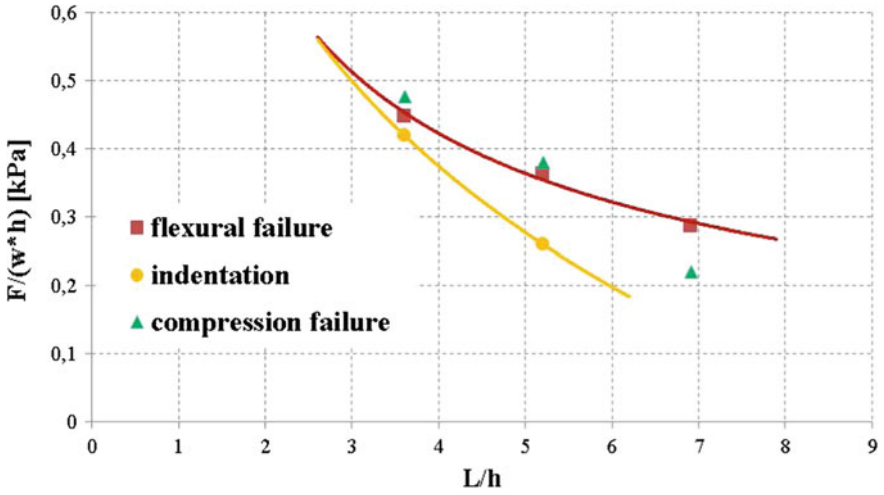


Fig. 18 The beam failure diagram in the model ice of FG type

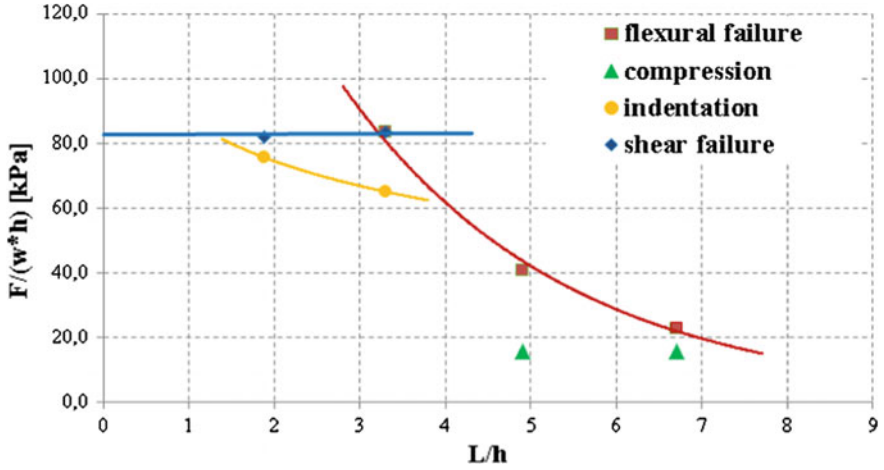


Fig. 19 The beam failure diagram in model ice of columnar type

The failure diagrams show that the studied ice types demonstrate various behavior during the test under consideration and variations in the beam length. It can be seen that:

- Ice crushing under indenter forcing in the initial stage of the beam loading is almost not distinguished from the force record in the sea ice; it can be identified in the model columnar ice at some tests and it clearly appears in the model fine-granule ice (FG type).

- Flexural cracks were observed in all ice types under study.
- In the natural sea ice case, the maximum load was produced by compression process, while in the model ice this process was not so well pronounced.
- Diagonal shear cracks were observed during the beam tests in the natural sea ice and in the columnar model ice.
- As the relative length of the beams ($\frac{L}{h}$) decreased to values 3–4, the growth of the failure force was observed: in the sea ice it was due to compression, in the model ice due to bending.
- In the shorter beams of the sea ice we can expect shear failure: the shear force as given in formula (4) is independent of the relative ice length; it is governed by the shear strength of ice and cross-sectional area, i.e. remains constant (Fig. 17). The same is valid for the model columnar ice beams (Fig. 19). In the fine-grained ice case, the ice crushing under indenter forcing is growing as the relative beam length is reduced, and this force becomes the governing factor of the ice beam failure.

Conclusions

Our investigations have shown that the test on loading of floating ice beam with fixed ends is representative in the terms of study of the ice failure mechanisms at various relations of its strength parameters (crushing, compressive, flexural, shear strength). The ice types used for the described tests (sea ice, model ice of the FG type, and columnar model ice) have different relationships of the strength parameters, primarily, the ratio of the ice flexural strength to crushing strength. This, in turn, causes various ice failure patterns when loading geometrically similar beams.

The tests in the Ice Basin were performed only with two model ice sheets (one of the FG type and one of the columnar type), which were prepared according to the technology used in the Krylov Centre. It is interesting to spread the number of the ice sheets, as well to carry out the similar tests of ice beams with fixed ends in the other Ice Basins with their specific technology of model ice preparation.

In the current analysis the effects of indentation rate, the Young modulus, and ice inhomogeneity as functions of ice thickness were not considered. Additional tests are required to study these factors.

Acknowledgements The authors acknowledge the support from UNIS and the Research Council of Norway through the Centre for Research-based Innovation (SAMCoT project).

References

1. Vershinin, S. A. (1988). Ice action on offshore structures. In *Results of science and technology, issue: Water Transport* (Vol. 13). Moscow.
2. Sanderson, T. J. O. (1988). *Ice mechanics—Risks to offshore structures.* Graham and Trotman.

3. Schulson, E. M., & Duval, P. (2009). *Creep and fracture of ice* (p. 417). Cambridge University Press.
4. Schwarz, J., & Weeks, W. (1977). Engineering properties of sea ice. *Journal of Glaciology, 19*(81), 499–531.
5. Sodhi, D. S. (2001). Crushing failure during ice-structure interaction. *Engineering Fracture Mechanics, 68,* 1889–1921.
6. Timco, G. W., & Weeks, W. F. (2010). A review of the engineering properties of sea ice. *Cold Regions Science and Technology, 60,* 107–129.
7. Sodhi, D. S. (1998). Vertical penetration of floating ice sheets. *International Journal of Solids and Structures, 35*(31–32), 4275–4294.
8. Marchenko, A., Karulin, E., Chistyakov, P., Sodhi, S., Karulina, M., & Sakharov, A. (2014). Three dimensional fracture effects in tests with cantilever and fixed ends beams. In *Proceedings of the 22nd IAHR ice symposium*. Singapore, ICE14-1178.
9. Sakharov, A., Karulin, E., Marchenko, A., Karulina, M., Sodhi, D., & Chistyakov, P. (2015). Failure envelope of the brittle strength of ice in the fixed-end beam test (two scenarios). In *Proceedings of the 23rd international conference on port and ocean engineering under Arctic conditions*. Trondheim, Norway.
10. Enkvist, E., & Makinen, S. (1984). A fine-grain model-ice. In *Proceedings IAHR ice symposium* (Vol. 2, pp. 217–227). Hamburg, Germany.
11. Evers, K.-U., & Jochmann, P. (1993). An advanced technique to improve the mechanical properties of model ice developed at the HSVA ice tank. In *The 12th international conference on port and ocean engineering under arctic conditions, 17–20 August 1993* (Vol. 2, pp. 877–888). Hamburg.
12. Nortala-Hoikkanen, A. (1990). FGX model ice at the Masa-Yards Arctic Research Centre. In *IAHR ice symposium* (pp. 247–259). Espoo, Finland.
13. Timco, G. W. (1986). A new type of model ice for refrigerated towing tanks. *Cold Regions Science and Technology, 12,* 175–195.
14. Von Bock und Polach, R., Ehlers, S., & Kujala, P. (2013). Model-scale ice—Part A: Experiments. *Cold Regions Science and Technology, 94,* 74–81.
15. Karulin, E., Marchenko, A., Karulina, M., Chistyakov, P., Sakharov, A., Ervik, A., et al. (2014). Field indentation tests of vertical semi-cylinder on first-year ice. In *Proceedings of the 22nd IAHR ice symposium 2014*. Singapore, ICE14-1125.
16. Experimental verification of theoretical approach for model ice failure mechanism in ice model basin. (1998). *Technical report of KSRI*. Issue 39545.
17. Method of ice cover simulation taking into account fracture toughness of ice. (1990). *Technical report of KSRI*. Issue 33141.
18. Shimansky, Yu. A. (1958). Handbook on ships mechanics (Vol. 1, p. 627). Leningrad.

Seasonal Freezing of a Subwater Ground Layer at the Laptev Sea Shelf

Peter V. Bogorodskii, Andrey V. Pnyushkov and Vasilii Yu. Kustov

Introduction

Fast ice is a regular congelation ice, formed when initial forms of drifting ice (e.g., frazil and grease ice, slash, and shuga) are frozen in shallow coastal waters, where the natural roughness of the coast contributes to their cohesion to a shore. Regular ice observations at polar stations suggest that the formation of fast ice occurs when sea ice thickness reaches 5–10 cm, usually within 10–15 days after the onset of steady ice formation. At the initial stages, fast ice appears at the coastline in the form of ice shores [26]. At shallow depths (<2–3 m), the layer of water beneath the ice can freeze completely and the newly formed ice begins to directly affect the physical and mechanical properties of bottom sediments, including freezing of porous water there and the appearance of frozen rocks [3, 13, 30].

Fast (immobile) ice along the continental and insular coasts of the Laptev Sea represents a characteristic feature of the Arctic ice landscape and belongs to the most developed ice types in the Siberian shelf seas. The nearly wholly fast ice in the Laptev Sea is formed in the southern and southeastern parts of the sea, covering about 40% of the sea area at the end of ice development season: i.e., April–May [11]. Such a large area covered with fast ice suggests that the southeastern Laptev Sea has the widest fast ice extent in the Arctic Ocean [22]. Interannual changes in fast ice edge position in the Laptev Sea are less dependent on fluctuations of hydrometeorological conditions (as was evident, for example, from low correlations

P. V. Bogorodskii (✉) · V. Yu. Kustov
Arctic and Antarctic Research Institute, St. Petersburg, Russia
e-mail: bogorodski@aari.ru

A. V. Pnyushkov
International Arctic Research Centre, University of Alaska, Fairbanks, USA

A. V. Pnyushkov
Arctic Research Centre, University of Hokkaido, Sapporo, Japan

M. G. Velarde et al. (eds.), *The Ocean in Motion*, Springer Oceanography,
https://doi.org/10.1007/978-3-319-71934-4_39

with the atmospheric air temperature and pressure time series), compared to the other Arctic seas [11] and, to a high degree, are controlled by the intensity and redistribution of river runoff [8].

Studies of Laptev Sea fast ice began almost simultaneously with the establishing of the polar station network in the Arctic in the 1930s, and continue through the present time. The bibliography of relevant papers, focused primarily on phenomenological descriptions and features of the fast ice regime in the Laptev Sea, is a list hundreds long. Since the late 1970s, a substantial portion of these studies has been devoted to mathematical modeling of fast ice. Among recent studies, we should mention a paper by [29], who examined the annual evolution of fast ice in the Laptev Sea. Despite imperfect reproduction of the atmospheric precipitations in the reanalysis model used in the performed simulations as forcing, the authors estimated fast ice growth rates by applying different values for heat fluxes from the sea to the bottom of the ice and the surface albedo. The authors came to the obvious conclusion about the substantial effect of these parameters on the rates of ice thickness changes.

The freezing of bottom sediments in the Laptev Sea has been much less studied than water layer freezing in the coastal areas of the sea. Despite the urgency for understanding of sediment layer freezing/melting processes, in particular, for quantifying the rates and volume of methane release from the eastern Arctic seas (see [23, 24] for details), and numerous applied (engineering) aspects of this problem [3], our knowledge of under-sea ground thermodynamics is still fragmentary. For example, until recently only spatial distributions of various types of deposits, concurrent with general information about their thermo-physical properties were available in the Laptev Sea [15]. The situation was improved by the implementation of international projects on the study of underwater permafrost by drilling from immobile ice [10, 18, 21]. Up to now, the most detailed studies of bottom sediments at the Laptev Sea shelf have been around the River Lena delta and the western part of the Buor-Khaya Bay, near the Tiksi township [5, 6]. These studies suggest that bottom soils can stay both in a cooled state (without inclusions of ice) and partly or completely frozen. In addition, they have analyzed the first reliable data on the composition, cryogenic structure, and thermophysical properties of samples of frozen and thawed soils and their changes to depths exceeding 50 m.

Despite the availability of experimental data, mathematical models of heat and salt transfer in bottom soils are rare; the available ones, as a rule, do not describe the interactions and feedbacks of the energy and mass transfers in the boundary layers of the sea, atmosphere, and bottom [9]. In addition, there is no common view about the mechanisms playing a role in bottom sediment freezing [30]. For instance, it is believed that this bottom freezing begins only after the complete freezing of the overlying water column, notwithstanding reported evidence for the opposite point [13] for discussion.

In this contribution, we describe the fast ice formation and seasonal freezing of the sub-water ground layer in the Tiksi Gulf (a coastal part of the Buor-Khaya Bay) during the winter of 2014–2015 based on joint analysis of experimental data, theoretical studies, and modeling. For assessment of the effects of fast ice growth on

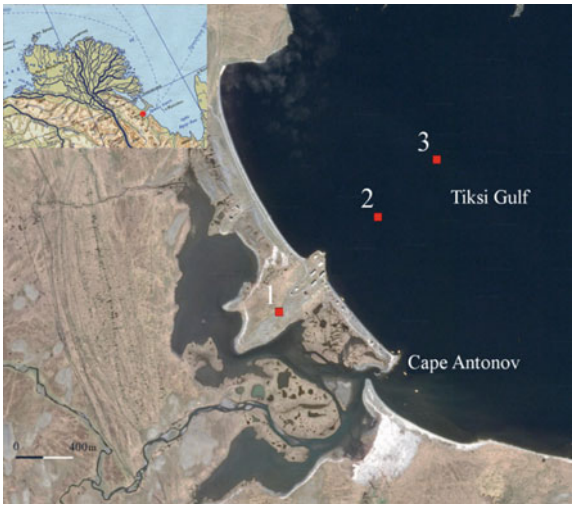


Fig. 1 Maps showing locations of standard meteorological observations (1), sea ice thickness and snow depth measurements (2), and the thermistor chain (3) in the Tiksi Gulf in 2014–2015

seasonal freezing of bottom sediments, a one-dimensional thermodynamic model is proposed. This model represents contacting layers of air, snow, ice, water, and soil as a single system connected by thermal interactions, fundamentally distinguishing it from analogs developed for the individual media. The paper is organized as follows. We first describe the utilized model and basic data sets used for simulations and validation of the model. We then describe the results of simulations and conclude with a general discussion and summary of results (Fig. 1).

Observational Data

For model simulations and further validation of model results, we utilized standard (3-h) meteorological and ice observations performed at the Tiksi polar station. The collected data set includes regular measurements of sea ice; they snow thicknesses with daily (for thin < 0.2-m ice) and ten-day (for thick > 1-m ice) temporal resolutions according Nastavlenie po gidrometeorologicheskim stanciyam i postam [16]. Recognizing the natural heterogeneities of fast ice and snow cover, the average error of these measurements can be estimated as 0.5 cm.

Meteorological and ice observations at the Tiksi station were accompanied by temperature measurements within the freezing water layer, carried out at a distance of 0.5 km from the coast at a depth of about 3 m, from December 10, 2014 through June 10, 2015, with an hourly time interval. For this purpose, a GP5W-Shell thermistor chain consisting of ten sensors was deployed in water through a hole in the young ice (Fig. 2a). The distance between thermal sensors was ~0.5 m. To minimize the effect of penetrating solar radiation on the thermistors, the dark sections of the instrumental cable were protected with a foil tape. According to the

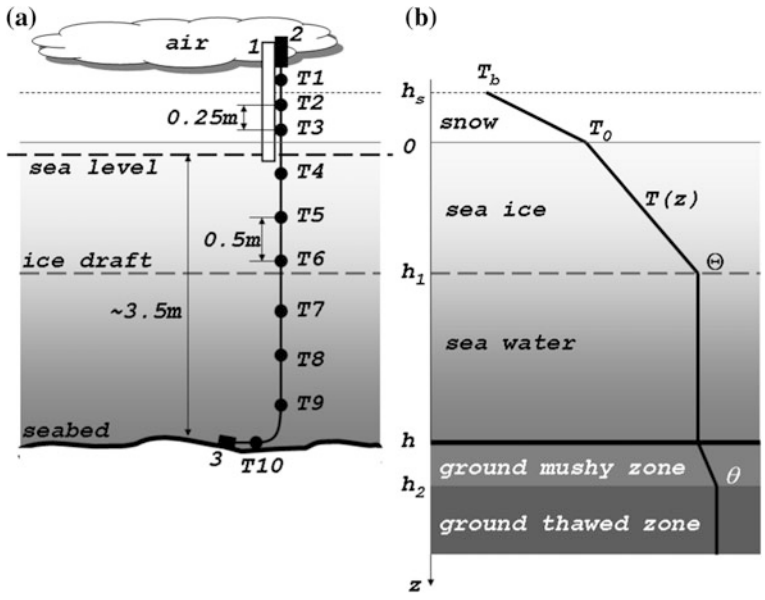


Fig. 2 **a** Schematics of the thermistor chain installed in growing fast ice in the Gulf of Tiksi in 2014–2015, and **b** piecewise linear temperature profile $T(z)$ in the simulated "ice-water-ground" system. Numbers indicate a data logger (1), a wooden pole (2), an anchor (3), and thermistor sensors (T1–T10)

manufacturer, the accuracy of temperature registration with the GP5W-Shell thermistors is ±0.25 °C, in a temperature range from −10 to +30 °C.

Model Description

The freezing of water and sub-water sediment layers in the Gulf of Tiksi was studied by employing the modified one-dimensional thermodynamic model originally proposed by Bogorodskii and Pnyushkov [4] to describe ground freezing over the Laptev Sea shelf. The applied model simulates the heat and salt transfer accompanied by phase transitions in a system consisting of a semi-infinite layer of initially thawed saline bottom soil (subscript "th") underlying the layer of sea water (subscript "w"). The formation of ice (subscript "i") in the water layer is described by the classical Stefan problem, in which phase transitions are localized at the ice-water interface (h_i). In addition to the atmospheric heat fluxes used in this model as forcing, the model takes into account snow (subscript "s") accumulation, derived from meteorological observations collected at the Tiksi weather station. The freezing (crystallization) of the bottom sediment layer occurs within an extended area (two-phase or mushy zone; subscript "m"), the liquid phase of which shows

crystallization in the temperature spectrum: the range determined by the local condition of liquidus (see [2, 25, 27, 28] for details).

The model assumes a two-layer structure for freezing soils, consisting of a two-phase (partially frozen) zone and an underlying, completely thawed layer (Fig. 2b). The layers of fast ice and water are considered a monophase, meaning there is no liquid water in the ice layer, and vice versa. In addition, ice does not exist in the thawed ground layer. It is assumed that the freezing soil is a porous medium, saturated only with ice and sea water; the ice skeleton (subscript "g") is incompressible and immobile, and all dissolved salt formed during freezing is rejected in the volume of unfrozen liquid. The properties of sea and pore ice are considered the same, and the heat fluxes in all layers of the system are stationary. In the mushy zone, phase composition is determined by the thermodynamic equilibrium (liquidus condition) of ice and an unhardened solution. This solution is characterized by water fraction ν, whose value at the interface between the mushy and thawed zones changes discontinuously.

The propagation of heat in the snow $(-h_s \le z \le 0)$, ice $(-h_s \le z \le 0)$, and thawed soil $(h < z < h_2)$ layers are described by the equations of heat transfer:

$$(\rho C)_{s,i,th} \frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left(k_{s,i,th} \frac{\partial T}{\partial z}\right). \tag{1}$$

In the two-phase zone $(h < z < h_2)$, the heat and mass transfer equations can be written as:

$$\frac{\partial}{\partial t}\left[(\rho C)_m T\right] + m\rho_i L \frac{\partial \nu}{\partial t} = \frac{\partial}{\partial z}\left(k_m \frac{\partial T}{\partial z}\right), \tag{2a}$$

$$\frac{\partial}{\partial t}(\nu S) = D \frac{\partial}{\partial z}\left(\nu \frac{\partial S}{\partial z}\right), z \in (h, h_2), T = T_{eq} - \alpha S. \tag{2b, c}$$

where $(\rho C)_m = (1-m)(\rho C)_g + m\nu (\rho C)_w + m(1-\nu)(\rho C)_i$, $(\rho C)_{th} = (1-m)(\rho C)_g + m(\rho C)_w$, $k_m = (1-m)k_g + m\nu k_w + m(1-\nu)k_i$; $k_{th} = (1-m)k_g + m k_w$, T is temperature, S is salinity, C is specific heat, ρ is density, k is the heat conduction coefficient, L is the latent heat of fusion of ice, D is salt diffusivity, and m is porosity.

The sub-ice water layer $(h_1 < z < h)$ is supposed to be homogeneous due to convective mixing with temperatures equal to the freezing point, which decreases with ice growth due to salt rejection.

$$T \equiv \Theta = T_{eq} - \alpha S, S = S_0(h - h_{1,0})/(h - h_1(t)), \quad z \in (h_1, h), \tag{3a, b}$$

where Θ is the freezing point of seawater, α and T_{eq} are constants, and subscript "0" indicates the initial time $t = 0$.

At the upper interface of the system (subscript "b"), the boundary conditions describe total heat fluxes to the atmosphere:

$$k_s \frac{\partial T}{\partial z} = EH, \quad z = h_b, \tag{4}$$

where $EH = H + LE + R$ is the thermal flux through the ice cover; H and LE are the vertical turbulent fluxes of sensible and latent heat, respectively; and R is the longwave radiation balance (the shortwave radiation is neglected). H and LE are calculated using bulk aerodynamic equations widely used to estimate energy exchange between the sea and the atmosphere in the presence of ice [14]. The calculation of R was performed according to the Ångström's equation, linearized with respect to $(T_b - T_a)$, where the subscript "a" indicates air.

At the interface between snow and ice (index "0"), we assume continuity for temperatures and heat fluxes:

$$T^- = T^+ \equiv T_0, k_i \frac{\partial T^+}{\partial z} - k_s \frac{\partial T^-}{\partial z} = 0, z = 0, \tag{5a, b}$$

where the negative and positive signs denote upper and lower sides of the interface.

At the moving interface between the water and ice phases (subscript "1"), and at $z = h$, we use temperature continuity and thermodynamic equilibrium (liquidus) conditions, alongside the classical Stefan condition for prescribing sea ice growth:

$$T^- = T^+ = T_{eq} - \alpha S \equiv T_1, z \in (h_1, h_2), \tag{6a}$$

$$\rho_i L \frac{dh_1}{dt} = k_i \frac{\partial T^-}{\partial z}, t > 0 \tag{6b}$$

Similarly, the Stefan and the continuity for temperatures and water fraction conditions, together with liquidus condition, were implemented at the interface between the two-phase and thawed zones ($z = h_2(t)$, subscript "2"):

$$m(1 - \nu_3)\rho_i L \frac{dh_2}{dt} = k_m \left(\frac{\partial T}{\partial z}\right)^- - Q_{th}, \tag{7a}$$

$$\nu^- = \nu_2, \nu^+ = 1, T^- = T^+ \equiv \theta = T_{eq} - \alpha S_0, t > 0 \tag{7b}$$

where θ is the sea water freezing temperature, and Q_{th} is the geothermal heat flux. Diffusion of salt is neglected here.

We use the following initial conditions for the model:

$$t = 0: \quad T(z, 0) = T_0, S(z, 0) = S_0, z \in [0, h],$$

$$\nu = \nu_\infty = 1, h_{1,0} = 0, h_{2,0} = h, T = T_0, S = S_0, z \in [h, \infty),$$

The equivalent water content ν_{eq}, which serves as a measure of the thermodynamic state of the two-phase zone [12], can be estimated by integrating (2b) within the mushy layer over time.

Table 1 Thermophysical properties of the system [14, 17, 25]

	Property	Units	Snow	Ice	Water	Ground
ρ	Density	kg/m^3		910	1000	1300
k	Thermal conductivity	W/(мK)	0.31	2.23	0.58	2
L	Latent heat	J/kg			3.33×10^5	
c	Specific heat	J/(kgK)		2000	4190	1920
α	Constant	°C/‰			0.054	
T_{eq}	Constant	°C			0	
m	porosity					0.6

$$\nu_{eq} = 2\,\theta\,\nu_2/(\Theta + \theta), \tag{8}$$

In this case, the lower boundary moisture ν_2 is a free model parameter, prescribed a priori.

We summarize major model parameters used in the simulations in Table 1. The water content of the frozen and thawed zones interface was set to 0.1.

Ice Temperature Observations

In this section, we describe 2014–2015 ice/water temperature observations collected by the deployed thermistor chain (Fig. 2 for location), to provide insight into the evolution of ice thermal characteristics during the ice growth period. Atmospheric conditions in the winter of 2014–2015 were close to their climatic averages for the southern Laptev Sea, but characterized by substantial temporal (synoptic) variability (Fig. 3). According to the air temperature record, steady negative temperatures over the Gulf of Tiksi were observed from mid-October 2014 through the end of May 2015 (i.e., across ~230 days). Taking into account the delayed start of ice formation after establishing negative air temperatures, in this study we use the

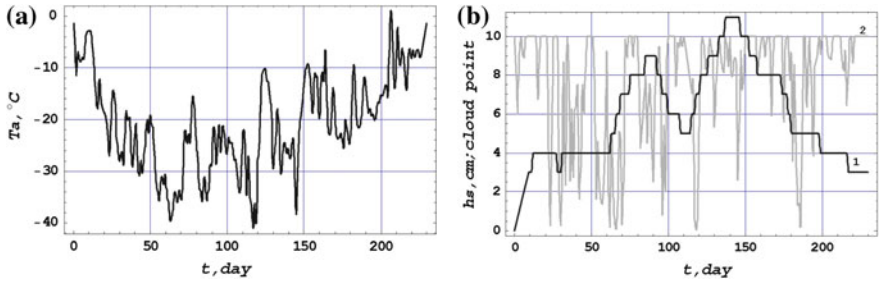


Fig. 3 **a** Time series of daily air temperatures, and **b** snow thickness on the ice (1) and total cloudiness (2) at the Tiksi station. The zero on the time axis corresponds to December 10, 2014

period from October 12, 2014 through June 1, to perform simulations with the thermodynamic 1D model (see section "Model Description" for details). The choice of October 12 as the start date for ice formation in the model agrees well with ice observations carried out daily in October–November in the Gulf of Tiksi. The extended period of negative air temperatures, which dropped below −40 °C in January and February 2015, results in substantial ice production in the Gulf, so that at the end of ice growth season in the Laptev Sea (late May), fast ice thickness reaches 2.08 m.

Vertical temperature profiles measured by the thermistor chain within the contacting layers of snow, ice, and water during winter 2014–2015 were close to linear, indicating rather constant heat fluxes there (Fig. 4). Different temperature gradients (vertical slopes of temperature distribution) in the snow and ice layers likely suggest differences in the thermal conductivity of these layers. We found that the linearity of the temperature profiles increases with the growth of fast ice thickness. This finding is contrary to the theoretical conclusion that the linear profile is more typical for thin (up to 0.5 m) ice [14]. The difference can be theoretically explained by stronger variability of the atmospheric forcing at the initial period of ice formation, compared with the observations in late winter. Highly variable thermal fluxes at the top of young ice result in variations of the rates of phase transitions in the body of newly formed ice saturated with brine. During the initial period of ice growth, the fraction of liquid phase within the ice, one of the major contributors to a non-linear temperature distribution, is quite large, but decreases when the solid phase begins to grow due to internal crystallization. In addition, during this period there was still no snow cover on top of the sea ice, which serves as a powerful thermal insulator. The formation of steady snow cover gradually reduces temperature contrasts between the atmospheric boundary layer and the snow/ice surface and, consequently, decreases vertical heat fluxes through the ice. In the water column below the fast ice, temperature distribution was always uniform due to convective mixing; the temperatures were close to the freezing point over the entire period of ice growth in the Gulf of Tiksi (i.e., from December to May).

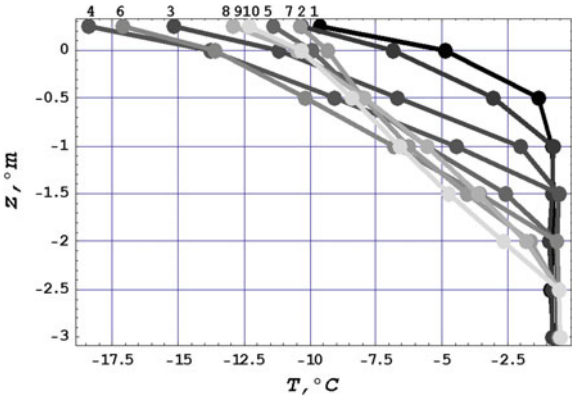


Fig. 4 Temperature profiles in the layers of snow, ice, and water measured by the thermistor chain in the Gulf of Tiksi on 15 December (1), 30 December (2), 15 January (3), 30 January (4), 15 February (5), 28 February (6), 15 March (7), 30 March (8), 15 April (9), and 30 April (10)

Simulation Results

We performed simulations of the seasonal growth of fast ice and freezing of bottom sediments using the model described in section "Model Description". To do so, model (1)–(8) was integrated over 230 days, from October 12, 2014 through June 1, 2015, for the initial water salinity of 10 psu and 1-, 1.5-, and 3-m bottom depths. The geothermal heat fluxes at the interface between the mushy zone and the underlying, completely melted layer of bottom sediment were taken as $Q_{th} = 0.06\,\mathrm{Wt/m^2}$. The date of the onset of ice formation was determined from regular ice observations carried out at the Tiksi station. Three-hour measurements of air temperature, relative humidity, wind velocity, atmospheric pressure, total cloudiness, and snow thickness over the ice surface averaged over daily intervals were utilized as an atmospheric forcing for the model.

The estimated growth of fast ice at various depths is shown in Fig. 5a. The differences between simulated ice thicknesses during the first two months of ice growth (i.e., for thin ice) were very small at all depths. We note good agreement between simulated fast ice thicknesses at 3-m depth with those measured by regular ice thickness surveys carried out at the Tiksi station (<8 cm on the first thirty days of simulations; and <4 cm after thirty days; Fig. 5). Some discrepancies with the in situ measurements at the end of the ice growth season can be explained through

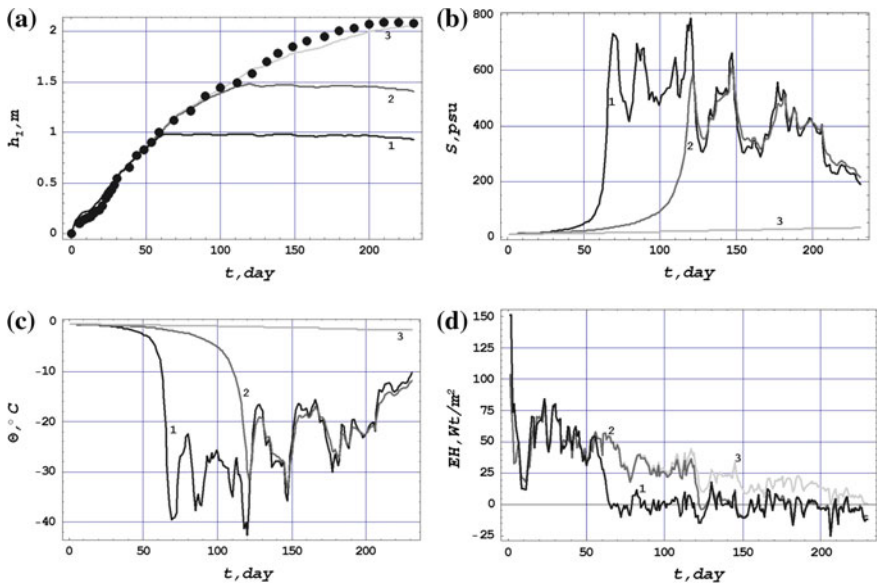


Fig. 5 Simulated fast ice thickness (**a**), salinity (**b**), and temperature (**c**) of the under-ice layer, and heat fluxes through the ice cover (**d**) in 2014–2015. Numbers indicate the basin depths of 1 (1), 1.5 (2) and 3 (3) meters used in the model simulations. Black circles in (**a**) show the fast ice thicknesses measured at the Tiksi station

potential impact from radiation heating not described by the model, as well as by natural irregularities in the thickness of the ice cover.

One of the important features of ice formation in shallow waters is the stabilization of ice thickness caused by a substantial water salinity increase [3, 4]. The calculations performed suggested that strong winter cooling maintains a very thin under-ice brine layer (a few mm thick), in which water salinity reaches 800 psu (Fig. 5b), corresponding to water freezing temperatures below −40 °C (Fig. 5c). This makes the liquid in the under-ice layer similar to unfrozen water in the soil. The presence of the thin brine layer theoretically prevents the consolidation of fast ice with the bottom sediment layer, even at very low Arctic temperatures. This finding does not contradict the results of laboratory experiments on ice adhesion in salt water (see [7] for details). As follows from the simulations, the restraining effect of the salinity increase reduces as bottom depths exceed 3 m; at depths greater than 3 m, this effect becomes insignificant.

Simulated heat fluxes through the ice (EH) demonstrate a clear tendency to decrease from autumn through spring, resulting in the corresponding reduction in ice growth rate for thick fast ice (Fig. 5d). In the second half of winter, temperature at the lower ice interface becomes higher than at the upper one, due to an increase of the air temperature (Fig. 3a). This temperature distribution suggests a change in the direction of heat fluxes through the ice cover. Following Stefan's condition (6b), these heat fluxes result in ice melting despite continued negative air temperatures. This effect is especially strong in the shallow water, where cooling of the brine layer is more intense. In these zones, the fast ice will melt, while in the deeper parts of the Gulf it will continue to grow at the same air temperatures.

The features of fast ice growth and salinization of the under-ice brine layer determine the evolution of the mushy zone in the bottom sediment layer (Fig. 6). The consequent decrease in water temperatures in winter leads to deeper freezing of the soil, accompanied by a decrease in its moisture content. Simulations performed with the model demonstrate the partial freezing of the upper layers of the soil (Fig. 6). As expected, the layer of bottom sediments is frozen most rapidly and deeply at the minimum depth (i.e., in the model experiment with 1-m bottom depth). The contents of unfrozen liquid in the two-phase zone is strongly affected by the depth of the basin (Fig. 6d). For instance, when bottom depth decreases from 3 to 1 m, the simulated water content equivalent of the two-phase zone decreases by almost an order of magnitude (from ∼0.07 to ∼0.007). This decrease is explained by a faster transition of liquid pore moisture into the solid state in the upper soil layer at shallow depths.

According to the conditions of thermodynamic equilibrium, ice formation in the layers of sea water and the bottom sediments begins simultaneously at all depths. The rates of ice formation in these two layers vary considerably over time. For example, if the highest growth rates of fast ice occur during the initial period of ice formation, the fastest consolidation of the bottom sediment layer coincides with the maximal development of unfrozen sub-ice brine layer (Fig. 6c).

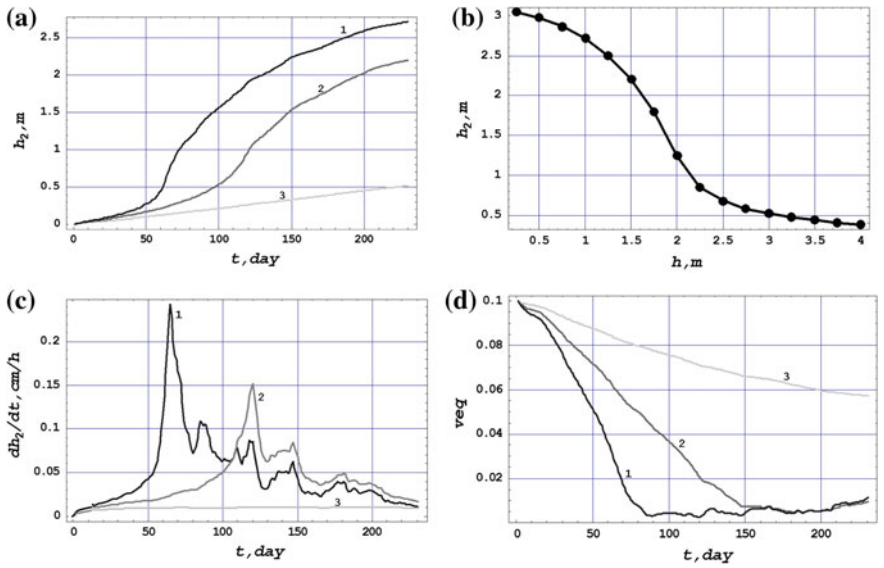


Fig. 6 **a** Simulated thickness, **b** maximal vertical extension, **c** growth rate, and **d** equivalent water content of two-phase zone reproduced with the thermodynamic model. Numbers indicate basin depths of 1 (1), 1.5 (2) and 3 (3) meters used in the model simulations

The initial salinity of sea water also has a significant effect on the growth of permafrost at the depths exceeding the thickness of fast ice. According to simulations performed for a water depth of 3 m, a salinity increase from 5 to 20 psu results in the decrease of the maximum ice thickness from 2.07 m to 1.98 m, and corresponding changes in the permafrost layer from 0.37 to 0.7 m. We argue that these changes result from a substantial reduction in water temperatures beneath the fast ice layer due to salinization, creating favorable conditions for deeper freezing of the sediment layer. At the same time, the effect of the salt flux at the interface between water and soil layers for the growth of permafrost is almost negligible, due to weak salt diffusion.

Discussion and Conclusions

Numerical simulations performed with the 1D thermodynamic model plausibly reproduce seasonal evolution in the "ice-brine-ground" system as follows, for example, from the good agreement between simulated and measured fast ice thicknesses in 2014–2015. Fast ice growth in the winter of 2014–2015 is also well represented by the empirical equation $h_1 = a N^b$ by Vieze [26] proposed for the Laptev Sea, where N is the sum of freezing-degree-days (the sum of daily

temperatures below the freezing point); $a = 0.00633$ and $b = 0.692$ are the empirical coefficients that take into account the other meteorological factors. We speculate that this circumstance may be the evidence of quite stable conditions of fast ice formation over the Laptev Sea shelf, or, most likely, the existence of some compensation mechanisms that reduce the effect of increased air temperatures in the Arctic [1].

Despite the plausibility of major simulated properties in the "ice-brine-ground" system generally agreeing with the current knowledge of the mechanisms of fast ice and the ground freezing, they should be considered with caution, due to several limitations used in the model. For instance, water salinity in the sub-ice brine layer may not reach the simulated values (~800 psu) due to horizontal mixing and salt advection, which are completely neglected in the one-dimensional model. Therefore, the simulated salinization effect is most pronounced in the stagnant areas of the Gulf with typical depths of 2–4 m and weak tidal and surge dynamics.

Advective heat and salt fluxes induced by the barotropic tides in the Buor-Khaya Bay are small, as evident from simulations using a linear inverse model of Arctic Ocean barotropic tides by Padman and Erofeeva [19], with assimilation of the coastal and benthic tide gauge measurements, TOPEX/Poseidon, and ERS altimetry. A relatively coarse (~5 km) spatial resolution of this model for coastal applications does not allow us to resolve all particular features of tidal dynamics in shallow parts of the Gulf, but still provides valuable information about the general pattern of tides in the region.

According to the model, the highest tidal elevations (~30 cm) in the Buor-Khaya Bay were simulated in semi-enclosed shallow (<5-m) lagoons near the coast of the Bykovsky Peninsula. However, tidal velocities in these lagoons are small and usually do not exceed 1 cm/s. Tidal currents gradually rise toward the deep part of the Gulf, so they increase up to 2 cm/s towards the gulf opening. The strongest barotropic tidal velocities were simulated for the semi-diurnal constituents M_2 and S_2, the most energetic tidal harmonics at the Laptev Sea shelf and slope [20]. The diurnal components of tides (K_1 and O_1 constituents) were substantially (more than an order of magnitude) weaker. Low amplitudes of the tidal currents suggest unsubstantial lateral tidal mixing in this region, which can be negligible in the simulations of the under-ice water salinization reproduced by our thermodynamic model. The average slopes of the bottom of the investigated water area are also insignificant (of the order of 10^{-3}), which allows us to neglect salt advection caused by the dense brine outflow.

An additional source of the discrepancy between the observations and modeling is that the model does not take into account potential supercooling of pore liquid in the bottom sediments. At the same time, this supercooling is not always observed in the bottom sediment layer, at salinities exceeding 5 psu, the supercooling of pore liquid usually does not occur [30].

The following conclusions can be drawn as a summary of this study:

1. The linear temperature profiles measured by the thermistor chain in the snow, ice, and water layers suggest constant vertical thermal fluxes in these media.

The linearity of the temperature profiles in ice increases with the growth of fast ice, likely indicating a reduction in the liquid phase fraction in the ice layer due to crystallization.

2. The liquid sub-ice layer formed by salt rejection is an important element of the "ice-brine-ground" system. This layer cannot completely freeze, even at very low air temperatures typical for the shelf of the Siberian seas. The salinity increase in the sub-ice brine layer can cause melting of fast ice in shallow areas at negative air temperatures, alongside its simultaneous growth for offshore regions.
3. The bottom depth is one of the most important limiting factors of ice formation in water and bottom sediment layers. The deepest freezing in the sediment layer was simulated for shallow depths, where the grown fast ice approaches the bottom.

Acknowledgements This study was supported by the Russian Foundation for Basic Research (Project #17-05-01221 "Investigation of atmospheric boundary layer in the Arctic region with the data of the Russian polar observatories measurements"), the Ministry of Science and Education of the Russian Federation (Project "Changing Arctic Transpolar System" # 2017-14-588-0005-003), Roshydromet (Target scientific and technical program 1.5.3.2 "Environmental monitoring at the Tiksi Hydrometeorological Observatory in the frame of international project of joint researches"), and NSF (grants AON-1203473 and AON-1338948).

References

1. Bekryaev, R. V., Polyakov, I. V., & Alexeev, V. A. (2010). Role of polar amplification in long-term surface air temperature variations and modern Arctic warming. *Journal of Climate, 23,* 3888–3906.
2. Bogorodsky, P. V., & Pnyushkov, A. V. (2007). A simple model for seawater crystallization in the temperature spectrum. *Oceanology, 47*(4), 500–506.
3. Bogorodskii, P. V., Marchenko, A. V., Pnyushkov, A. V., & Ogorodov, A. S. (2010). Formation of fast ice and its impact on the coastal zone of the Arctic seas. *Oceanology, 50*(3), 345–354.
4. Bogorodskii, P. V., & Pnyushkov, A. V. (2011). Impact of Arctic land fast-ice growth on subwater ground freezing. *Problemy Arktiki i Antarktiki. Problems of Arctic and Antarctic*, *3* (89), 69–77 (in Russian).
5. Cheverov, V. G., Vidyapin, I. Y., & Tumskoj, V. E. (2007). The composition and properties of thermokarst lagoons deposits at the Bykovsky Peninsula. *Kriosfera Zemli. Earth Cryosphere, 11*(3), 44–50 (in Russian).
6. Chuvilin, E. M., Buhanov, B. A., Tumskoj, V. E., Shahova, N. E., & Semiletov, I. P. (2013). Thermal conductivity of bottom sediments in the region of Buor-Haya Bay (shelf of the Laptev Sea). *Kriosfera Zemli. Earth Cryosphere, 17*(2), 32–40 (in Russian).
7. Grishin, P. A. (1963). The freezing temperature of saline soils. *Trudy SOYUZMORNIIproekta. The works of the SOYUZMORNIIproekt, 3*(9), 84–91 (in Russian).
8. Dmitrenko, I. A., Gribanov, V. A., Volkov, D. L., Kassens, H., & Eicken, H. (1999). Impact of river discharge on the fast ice extension in the Russian Arctic shelf area. In *Proceedings of 15th International Conference on Port and Ocean Engineering Under Arctic Conditions (POAC99)*, Helsinki, 23–27 August 1999 (Vol. 1, pp. 311–321).

9. Harrison, W. D., & Ostercamp, T. E. (1978). Heat and mass transport processes in subsea permafrost. 1. An analysis of molecular diffusion and its consequences. *Journal Geophysical Research, 83*(C9), 4707–4712.
10. Junker, R., Grigoriev, M. N., & Kaul, N. (2008). Non-contact infrared temperature measurements in dry permafrost boreholes. *Journal Geophysical Research, 113,* B04102. https://doi.org/10.1029/2007JB004946.
11. Karklin, V. P., Karelin, I. D., Yulin, A. V., & Usoltseva, E. A. (2013). Peculiarities of the landfast ice formation in the Laptev sea. *Problemy Arktiki i Antarktiki. Problems of Arctic and Antarctic, 3*(97), 5–14 (in Russian).
12. Kerr, R. S., Woods, A. N., Worster, M. G., & Huppert, H. E. (1990). Solidification of an alloy cooled from above. Part 1. Equilibrium growth. *Journal of Fluid Mechanics, 216,* 323–342.
13. Khimenkov, A. N., & Brushkov, A. V. (2003). *Oceanic cryolite genesis* (336 pp.). Nauka (in Russian).
14. Makshtas, A. P. (1984). *Teplovoy balans arkticheskih ldov v zimniy period. The heat balance of the Arctic sea ice in the winter period* (66 pp.). Leningrad: Gidrometeoizdat (in Russian).
15. Molochushkin, E. N., & Gavriliev, Z. I. (1970). Structure, phase composition and thermal regime of rocks composing the bottom of the coastal zone of the Laptev sea. In *The Arctic Ocean and its coast in the Cenozoic* (pp. 503–509). Gidrometeoizdat (in Russian).
16. Nastavlenie po gidrometeorologicheskim stanciyam i postam. Manual to Hydrometeorological Stations and Posts (Vol. 9. Issue 1, 311 pp.). (1984). Leningrad: Gidrometeoizdat (in Russian).
17. Nazintsev, Y. L., & Panov, V. V. (2000). Phase composition and thermal characteristics of the sea ice (83 pp.). Gidrometeoizdat (in Russian).
18. Overduin, P. P., Wetterich, S., Günther, F., Grigoriev, M. N., Grosse, G., Schirrmeister, L., et al. (2016). Coastal dynamics and submarine permafrost in shallow water of the central Laptev Sea, East Siberia. *The Cryosphere, 10,* 1449–1462. https://doi.org/10.5194/tc-10-1449-2016.
19. Padman, L., & Erofeeva, S. (2004). A barotropic inverse tidal model for the Arctic Ocean. *Geophysical Research Letters, 31*(2), L02303. https://doi.org/10.1029/2003GL019003.
20. Pnyushkov, A., & Polyakov, I. (2012). Observations of tidally induced currents over the continental slope of the Laptev Sea, Arctic Ocean. *Journal of Physical Oceanography, 42,* 78–94.
21. Rachold, V., Bolshiyanov, D. Y., Grigoriev, M. N., Hubberten, H.-W., Junker, R., Kunitsky, V. V., et al. (2007). Nearshore Arctic subsea permafrost in transition. *EOS: Transactions, American Geophysical Union, 88*(13), 149–156.
22. Selyuzhenok, V., Krumpen, T., Mahoney, A., Janout, M., & Gerdes, R. (2015). Seasonal and interannual variability of fast ice extent in the southeastern Laptev Sea between 1999 and 2013. *Journal of Geophysical Research: Oceans, 120,* 7791–7806. https://doi.org/10.1002/2015JC011135.
23. Sergienko, V. I., Lobkovskii, L. I., Semiletov, I. P., et al. (2012). The degradation of submarine permafrost and the destruction of hydrates on the shelf of east Arctic seas as a potential cause of the "methane catastrophe". *Doklady Earth Science, 446*(3), 1132. https://doi.org/10.1134/S1028334X12080144.
24. Shakhova, N., Semiletov, I., Leifer, I., Sergienko, V., Salyuk, A., Kosmach, D., et al. (2014). Ebullition and storm-induced methane release from the East Siberian Arctic Shelf. *Nature Geoscience, 7*(1), 64–70. https://doi.org/10.1038/ngeo2007.
25. Vasiliev, V. I., Maksimov, A. M., Petrov, E. E., & Tsypkin, G. G. (1997). Heat and mass transfer in freezing and thawing grounds. *Physical and Mathematical Literature, 1997,* 224 (in Russian).
26. Vieze, V. Y. (1994). *Osnovy dolgosrochnyh ledovyh prognozov dlya arkticheskih morey. The basis of long-term ice extent forecasts for the Arctic seas* (Vol. 159, 274 pp.). Transactions of the Arctic Research Institute. Leningrad: Publishing of Glavsevmorput' (in Russian).

27. Wettlaufer, J. (1998). Introduction to crystallization phenomena in natural and artificial sea ice. In M. Leppäranta (Ed.), *Physics of ice-covered seas* (Vol. 1, pp. 105–194). Helsinki: Helsinki University Press.
28. Worster, M. G. (1986). Solidification of an alloy from a cooled boundary. *Journal of Fluid Mechanics, 167,* 481–501.
29. Yang, Y., Leppäranta, M., Li, Z. J. et al. (2015). Model simulations of the annual cycle of the landfast ice thickness in the East Siberian Sea. *Advanced in Polar Science*, *26*(2), 168–178. https://doi.org/10.13679/j.advps.2015.2.00168.
30. Zhigarev, L. A. (1997). *Oceanocheskaya kriolitozona. The Cryo-litosperic zone of ocean* (320 pp.). Moscow: Moscow State University (in Russian).