
Light
Scattering
Reviews 2
Alexander A. Kokhanovsky
Editor
Springer
PRAXIS

# Light Scattering Reviews 2

## Remote Sensing and Inverse Problems

Alexander A. Kokhanovsky (Editor)

# Light Scattering Reviews 2

## Remote Sensing and Inverse Problems

Springer

Published in association with
Praxis Publishing
Chichester, UK

Editor
Dr Alexander A. Kokhanovsky
Institute of Environmental Physics
University of Bremen
Bremen
Germany

SPRINGER–PRAXIS BOOKS IN ENVIRONMENTAL SCIENCES (*LIGHT SCATTERING SUB-SERIES*)
SUBJECT *ADVISORY EDITOR*: John Mason B.Sc., M.Sc., Ph.D.
EDITORIAL *ADVISORY BOARD MEMBER*: Dr Alexander A. Kokhanovsky, Ph.D. Institute of Environmental Physics, University of Bremen, Bremen, Germany

ISBN 10: 3-540-30932-2 Springer Berlin Heidelberg New York
ISBN 13: 978-3-540-30932-1 Springer Berlin Heidelberg New York

Springer is part of Springer-Science + Business Media (springer.com)
Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available from the Internet at http://dnb.ddb.de

Library of Congress Control Number: 2005935704

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

© Praxis Publishing Ltd, Chichester, UK, 2007
Printed in Germany

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: Jim Wilkie
Project copy editor: Mike Shardlow
Author-generated LaTex, processed by EDV-Beratung Frank Herweg, Germany

Printed on acid-free paper

# Contents

# List of Contributors

**A. J. Baran**
Met Office
Cordouan 2
C2-5, Fitzroy Road
Exeter
EX1 3PB
UK
anthony.baran@metoffice.gov.uk

**H. W. Barker**
Cloud Physics and Severe Wearther Research Division
Meteorological Service of Canada
4905 Dufferin Street
Downsview
Ontario M3H 5T4
Canada
Howard.Barker@ec.gc.ca

**L. P. Bass**
Keldysh Institute of Applied Mathematics
Russian Academy of Sciences
Miusskaya Sq. 4
125047 Moscow
Russia
bass@kiam.ru

**O. Hasekamp**
Netherlands Institute for Space Research
Sorbonnelaan 2
3584 CA, Utrecht
The Netherlands
O.Hasekamp@sron.nl

**H. Iwabuchi**
Frontier Research Center for Global Change
Japan Agency for Marine-Earth Science and Technology
3173-25 Showa-machi
Kanazawa-ku
Yokohama
Kanagawa 236-0001
Japan
hiro-iwabuchi@jamstec.go.jp

**A. A. Kokhanovsky**
Institute of Environmental Physics
University of Bremen
Otto Hahn Alee 1
D-28334 Bremen
Germany
alexk@iup.physik.uni-bremen.de

**V. S. Kuznetsov**
Research Scientific Center 'Kurchatov Institute'
Kurchatov Sq. 1
123182 Moscow
Russia
lri@bk.ru

**J. Landgraf**
Netherlands Institute for Space Research
Sorbonnelaan 2
3584 CA, Utrecht
The Netherlands
J.Landgraf@sron.nl

**A. V. Malinka**
Institute of Physics
National Academy of Sciences of Belarus
Pr. Nezavisimosti 68
Minsk 220072
Belarus
mal@zege.bas-net.by

**O. V. Nikolaeva**
Keldysh Institute of Applied Mathematics
Russian Academy of Sciences
Miusskaya Sq. 4
125047 Moscow
Russia
nika@kiam.ru

**A. V. Rozanov**
Institute of Environmental Physics
University of Bremen
Otto Hahn Alee 1
D-28334 Bremen
Germany
alex@iup.physik.uni-bremen.de

**V. V. Rozanov**
Institute of Environmental Physics
University of Bremen
Otto Hahn Alee 1
D-28334 Bremen
Germany
rozanov@iup.physik.uni-bremen.de

**T. Wriedt**
Stiftung Institut für Werkstofftechnik
Badgasteiner Str. 3
D-28359 Bremen
Germany
thw@iwt.uni-bremen.de

# Notes on the contributors

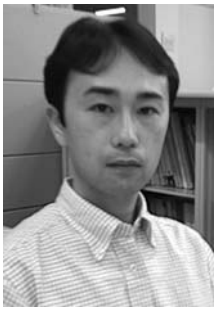
Before entering the field of atmospheric science **Anthony J. Baran** was actively engaged in various other fields such as astrophysics, theoretical laser-plasma physics and stochastic theory applied to wind energy. In 1990 he joined the Met Office, England, and in 1995 further studied for a Ph.D. in the atmospheric sciences at University College London, London University, on the radiative and remote sensing properties of cirrus, which he gained in 1997. Since then he has remained at the Met Office. His research interest covers the areas of scattering and absorption from non-spherical ice crystals, cirrus remote sensing and radiative transfer. Dr Baran has actively been engaged in research covering those three principal areas of interest. More lately, he has become interested in relating the macrophyscal properties of cirrus to their scattering properties, very high resolution measurements of cirrus in the far IR, and polarimetric measurements of cirrus. Dr Baran has also been the principal investigator of airborne cirrus campaigns. He has authored and co-authored over 50 peer-reviewed publications.

**Howard Barker** obtained his Ph.D. from McMaster University in 1991. His thesis work addressed solar radiative transfer for inhomogeneous cloudy atmospheres above reflecting surfaces. Since then he has worked as a research scientist in the Global Climate Modelling and Cloud Physics Research Divisions of Environment Canada. His current scientific interests deal with modelling radiative transfer for realistic cloudy atmosphere–surface systems for the purposes of climate modelling and remote sensing of Earth.

**Leonid P. Bass** is senior scientist at the Keldysh Institute of the Applied Mathematics (KIAM) of Russian Academy of Sciences. He graduated from Moscow Geodesic Institute in 1956. Dr Bass received the M.S. degree from Moscow State University, where he studied at the Department of Mechanics and Mathematics. Dr Bass obtained Ph.D. at Keldysh Institute of Applied Mathematics in Moscow, Russia, in 1976. Since 1960 he has worked at the KIAM with problems connected to the physics of reactors and radiation shielding. His main efforts were aimed at the development of effective computational methods for the estimation of radiation fields in different media with complicated structures such as nuclear reactors and similar power plants. For example, he has participated in the development of the special software for the solution of the multi-group steady-state transport equation for neutrons and gamma quanta in two- and three-dimensional geometries by the discrete ordinates method. Dr Bass is well-known as an expert in the field of parallel algorithms for supercomputers. Current interests include the calculation of solar radiation fields in spatially non-uniform media for the estimation of atmospheric parameters.

The scientific activity of **Professor T. Germogenova** (10.04.1930–27.02.2005) was associated with the solving of mathematical, physical and computational problems of radiation transport theory and reactor physics. She started working on transport theory problems in 1953 during the postgraduate course of the Physical Department of Moscow State University. Professor E. S. Kuznetsov was her postgraduate work supervisor. T. Germogenova was a staff member of the Institute of Applied Mathematics of the Russian Academy of Sciences for almost 50 years. Her Ph.D. thesis (1957) was dedicated to the solution of the transport equation for highly peaked phase functions. In 1962 she proved the principle of maximum for the linear transport equation. This quite general result can be used, particularly, in the study of convergence properties of some difference schemes for transport equation. Her Doctor of Mathematical Sciences degree thesis (1972) was dedicated to boundary problems of the transport equation and local properties of its solutions. The results obtained by T. Germogenova in the mathematical study of resolvability, smoothness properties and singularities of the transport equation solutions in dependence on medium and source properties are collected in her monograph, 'The Local Properties of Transport Equation Solutions' (1986, in Russian). T. Germogenova has been involved in the solution of a number important atmospheric optics problems (e.g. 3D radiative transport in cloudy media). She proved that the set of physically realizable states of polarized light in the Stokes–Poincaré representation is a cone in an appropriate functional space of four-dimensional vector-functions (1978). Next, this property was used for strict formulation both the 'non-negativity' property of the scattering matrix and the mathematical theory of the characteristic equation for the polarized light transport equation. T. Germogenova also has derived the set of asymptotical approximations for transport problems in optically thick inhomogeneous finite size regions (1961). Some of them are used for solving remote sensing and inverse atmospheric problems. She also developed a number of widely used numerical and analytical techniques, such as, the method of averaged fluxes for acceleration of inner iterations convergence (1968, 1969), Fourier analysis of stability of the WDD scheme, accuracy and stability analysis of the family of weighted nodal schemes (1994), and eigenfunctions of the finite moments method analysis (1996). Under her guidance a set of codes for 1D, 2D and 3D transport calculations has been developed for serial and parallel computers.

**Otto Hasekamp** obtained his Ph.D. in physics from the Free University of Amsterdam in 2002. His PhD work concentrated on radiative transfer theory and inverse problems in the context of ozone profile retrieval from satellite measurements of backscattered sunlight. He extended the forward-adjoint radiative perturbation theory with the inclusion of polarization, and proposed a new instrument concept for measuring tropospheric ozone using polarization measurement. His current work at the Netherlands Institute for Space Research (SRON) is focused on the retrieval of aerosol properties from satellite measurements of intensity and polarization. He developed a novel approach to the retrieval of aerosol properties, including a linearization of vector radiative transfer with respect to microphysical aerosol properties. Additionally he is involved in several other research projects related to GOME and SCIAMACHY, such as the retrieval of cloud properties and the retrieval of ozone profiles.



**Hironobu Iwabuchi** graduated from the Faculty of Science of Tohoku University, Japan, in 1996. He received a Doctor of Science degree in Geophysics from Tohoku University in 2001. His Ph.D. work examined three-dimensional radiative effects on satellite-based remote sensing of cloud properties. He is now a research scientist at the Frontier Research Center for Global Change of the Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan. His main research interests are modelling of three-dimensional radiative transfer and its applications in meteorology and in the remote sensing of gases, aerosols and clouds.

**Alexander A. Kokhanovsky** graduated from the Physical Department of the Belarusian State University, Minsk, Belarus, in 1983. He received his Ph.D. in optical sciences from the Institute of Physics, National Academy of Sciences of Belarus, Minsk, Belarus, in 1991. The Ph.D. work was devoted to modelling the light scattering properties of aerosol media and foams. Alexander Kokhanovsky is currently a member of the SCIAMACHY/ENVISAT algorithm development team (Institute of Environmental Physics, University of Bremen). His research interests are directed towards modelling light propagation and scattering in the terrestrial atmosphere. Dr Kokhanovsky is the author of the books *Light Scattering Media Optics: Problems and Solutions* (Chichester: Springer-Praxis, 1999, 2001, 2004), *Polarization Optics of Random Media* (Berlin: Springer-Praxis, 2003), and *Cloud Optics* (Berlin: Springer, 2006). He has published more than one hundred papers in the field of environmental optics, radiative transfer, and light scattering. Dr Kokhanovsky is a member of the European and American Geophysical Unions and also he is a member of the Belarusian Physical Society.

**Victor S. Kuznetsov** graduated from the Theoretical Physics Department of the Moscow Engineering and Physics Institute in 1965. Since 1976 he has worked at the Russian Research Center 'Kurchatov Institute' as senior scientist. His work is concerned with particle transport in different reactors plants, their safety and radiation shielding. Recently he developed the Discrete Ordinates Method algorithms for the solar radiation transport through the atmosphere and for radiation transport from different sources with singularities.

**Jochen Landgraf** is senior scientist at the Netherlands Institute for Space Research (SRON) and his research is focused on atmospheric radiative transfer, inverse modelling and photochemistry modelling. He obtained a Ph.D. at the Max-Planck Institute for Chemistry in Mainz, where he developed a module for efficient online calculations for photodissociation rates, which is implemented in several three-dimensional chemistry models. In addition, he has been involved in studying the effects of 3D radiative transport in inhomogeneous cloud fields and their relevancy for global chemistry modelling. His recent research has focused on remote sensing from satellite measurements, particularly from GOME and SCIAMACHY observations. He is involved in several national and international research projects, which are related to stratospheric and tropospheric ozone profile retrieval, aerosol retrieval, modelling of inelastic Raman scattering in atmospheric radiative transfer, and the development of a linearized radiative transfer model.

**Aleksey Malinka** graduated from the Belarusian State University, Department of Physics, in 1997. Since 1997 he has been a researcher in the B. I. Stepanov Institute of Physics of the National Academy of Sciences of Belarus in Minsk. He received a Ph.D. for his work on the theory of Raman lidar sounding with account for multiple light scattering effects in 2005. Dr Malinka's studies are aimed at the better understanding of light scattering phenomena in various natural environments. He is also active in oceanic and atmospheric lidar sounding research.

**Olga V. Nikolaeva** received the B.A. and M.S. degrees in applied mathematics from the Moscow State University, Russia, in 1993, and the Ph.D. degree in radiation transport investigation from Keldysh Institute of Applied Mathematics in 2002. Her research interest is focused on the simulation of radiation propagation and scattering in complicated multidimensional regions by parallel computers and investigation of discrete ordinates methods schemes for the transport equation. She is a member of I3RC group.

**Alexei V. Rozanov** graduated from the Physical Department of St. Petersburg State University, Russia, in 1994. He received a Ph.D. degree in natural sciences from the Institute of Environmental Physics, University of Bremen, Germany, in 2001. His Ph.D. work was devoted to the modelling of radiative transfer through a spherical planetary atmosphere and also to the retrieval of vertical distributions of atmospheric trace gases from occultation and limb measurements of backscattered solar light. Alexei Rozanov is currently a member of the SCIAMACHY/ENVISAT algorithm development team at the Institute of Environmental Physics of the University of Bremen, where he is active in the development of radiative transfer models and retrieval algorithms intended to obtain vertical distributions of atmospheric trace gases from satellite measurements. He has contributed to the publication of more than thirty scientific papers in the field of radiative transfer modelling and the remote sensing of the Earth's atmosphere.

**Vladimir V. Rozanov** graduated from the University of St. Petersburg, Russia, in 1973. He received his Ph.D. degree in physics and mathematics from the University of St. Petersburg, Russia, in 1977. From 1973 until 1991 he was a research scientist at the Department of Atmospheric Physics of the University of St. Petersburg. In 1990–1991 he worked at the Max-Planck Institute of Chemistry, Mainz, Germany. In July 1992 he joined the Institute of Remote Sensing at the University of Bremen, Germany. The main directions of his research are atmospheric radiative transfer and remote sensing of atmospheric parameters (including aerosols, clouds and trace gases) from space-borne spectrometers and radiometers. He is author and co-author of about one hundred papers in peer-reviewed journals.

**Thomas Wriedt** has been the head of the Particle Technology and Particle Characterization Department at the Institut für Werkstofftechnik, Bremen, Germany, since 1989. He studied Electrical Engineering at the University of Applied Science, Kiel, and also at the University of Bremen, and obtained his Dr.-Ing. degree on the numerical design of microwave antennas and components in 1986. From 1986 to 1989 he did research on signal processing with phase Doppler anemometry at the University of Bremen. His current research is mainly focused on optical particle characterization and light scattering theory.

# Preface

Light scattering is used in many applications, ranging from optical particle sizing of powders to interstellar dust studies. At the moment there is no a specialized journal aimed at studies of exclusively light scattering problems. Instead, different aspects of the problem and also different applications are considered in a variety of specialized journals covering several scientific disciplines such as chemistry, physics, biology, medicine, astrophysics, and atmospheric science, to name a few.

The Light Scattering Reviews (LSR) series started in 2006 with the aim of facilitating interaction between different groups of scientists working in diverse scientific areas but using the same technique, namely light scattering, for solution of specific scientific tasks. This second volume of LSR is devoted mostly to applications of light scattering in atmospheric research. The book consists of eight contributions prepared by internationally recognized authorities in correspondent research fields.

The first paper prepared by Howard Barker deals with the recent developments in solar radiative transfer in the terrestrial atmosphere and global climate modelling. In particular, methods to compute radiative transfer characteristics needed for numerical global climate models are discussed in a great depth. Their deficiencies are addressed as well. The problem of 3D radiative transfer in cloudy atmospheres, a hot topic in modern climate modelling, is also considered.

Anthony Baran prepared a comprehensive review aimed to studies of radiative characteristics of cirrus clouds. The global coverage of these clouds is quite large – up to 30% (70% in tropics). So cirrus plays an important but poorly defined role in the climate system. Methods of computing local optical characteristics of cirrus such as extinction coefficient, single scattering albedo, phase function and phase matrix, are given as well. He outlined traditional and novel methods to probe cirrus using airborne and satellite measurements.

Hironobu Iwabuchi addresses an important question of 3D radiative transfer in satellite cloud remote sensing. Up-to-date cloud remote sensing operational satellite remote sensing techniques are based on the model of a homogeneous cloud layer, which is never the case in reality. Clouds are inhomogeneous on all scales. Therefore, it is of importance to quantify errors, which are due to the use of 1D theory in retrievals. More importantly, new techniques must be developed, which account for 3D effects in retrieval procedures and also use 3D effects (e.g., shadows, cloud brightening and darkening) for the development of

new aerosol/cloud remote sensing techniques. The paper of Hironobu Iwabuchi is an important step in this direction.

Aleksey Malinka reviews the physical principles and ideas behind Raman lidar remote sensing of clouds and other geophysical media. The scattered light frequency can shift to the value, equal to an eigenfrequency of a scattering substance molecule. This allows the establishing of the presence of substances whose eigenfrequencies correspond to lines in the measured spectrum. The review is focused on techniques to account for Raman multiple light scattering in inverse problems of light scattering media optics. The described approximate analytical theory has an important advantage in terms of speed of calculations and no doubt will be used in future for a number of applications in different branches of light scattering media optics. In particular, a solution which explicitly relates the Raman lidar return to the medium local optical characteristics and the lidar parameters is derived. This is of importance for the solution of inverse problems.

Otto Hasekamp and Jochen Landgraf consider the application of the forward-adjoint perturbation theory for the solution of selected inverse problems of atmospheric optics taking account of the polarization of scattered light. In particular, the authors perform an analytical linearization of the vector radiative transfer equation with respect to atmospheric scattering parameters. Based on the developed approach, the authors describe a very effective approach to retrieve microphysical aerosol characteristics from spectral measurements of multiply scattered light intensity and polarization. The proposed theoretical technique has a potential for applications beyond the area of atmospheric research (e.g., in ocean and tissue optics).

Vladimir Rozanov and his co-authors demonstrate the role of derivatives of scattered light intensity in the formulation and solution of inverse problems. In particular, a relationship between the partial and variational derivatives of the intensity of radiation with respect to atmospheric parameters and the weighting functions is described. The basic equations for the direct and adjoint radiative transfer are reviewed. The solutions of these equations are used for the calculation of the weighting functions needed for the determination of atmospheric parameters from backscattered solar light measurements.

The last part of the book is aimed to the description of advanced numerical techniques of light scattering media optics. In particular, Thomas Wriedt describes the null-field method with discrete sources widely used for the calculation of scattering and absorption characteristics of scatterers having nonspherical shapes (spheroids, fibres, disks, Cassini ovals, hexagonal prisms, clusters of spheres, etc.). The method was originally developed to solve the stability problems in the standard T-matrix technique for the case of elongated and flat scatterers.

Olga Nikolaeva and co-workers review numerical grid schemes of the 3D radiative transfer equation solution. In particular, RADUGA code designed for multiprocessor computations is described. The code can be used to study light scattering and transport in finite light scattering objects of complicated shapes such as broken clouds and aerosol plumes.

This volume of Light Scattering Reviews is dedicated to the memory of Yoram Kaufman (01.06.1948–31.05.2006) and Kirill Ya. Kondratyev (14.06.1920–01.05. 2006), who made extremely valuable contributions to modern atmospheric research.

Bremen, Germany
October, 2006

Alexander A. Kokhanovsky

Part I

# Remote Sensing and Radiative Transfer

# 1 Solar radiative transfer and global climate modelling

H. W. Barker

## 1.1 Introduction

Over a period of time, Earth's outer shell (atmosphere, hydrosphere, cryosphere, biosphere) winds along a unique trajectory toward an ever-changing, and hence elusive, radiative equilibrium. It is elusive partly because Earth's overwhelming external boundary condition, solar irradiance, is never constant thanks to continuous variations in both orbit about the Sun and solar output. As such, the best Earth, and any other planet, can do is achieve a sequence of states that are in quasi- (radiative) equilibrium over a period of time that spans at least several annual cycles. Even if boundary conditions were static, it is now recognized that Earth's climate would not settle down to a single state or even a fixed cycle. Instead, it would execute a non-repeating sequence of, potentially very diverse, states that approximate radiative equilibrium. This chaotic character is supported by the inexorable intertwining of internal processes that operate at radically different time-scales. Indeed, the life giving/supporting character of Earth's climate system, that begins with absorption of solar radiation and ends with infrared emission to space, owes much of its richness, and worthiness of study, to the four-dimensional interaction between radiation and the three phases of water.

Given the fundamental role of large-scale radiation budgets in setting the character of climate and climatic change, and the fact that large-scale budgets are governed by conditions at smaller scales, it is essential that radiation–water interactions be accounted for as accurately as possible in numerical global climate models (GCMs). In GCMs, however, physical processes that occur at scales less than several hundred kilometres are often unresolved and so must be parametrized in terms of resolved conditions and assumptions about the state of unresolved conditions. This includes radiative transfer and many atmospheric fluctuations as their characteristic scales are typically less than a few kilometres.

While radiative fluxes at specific wavelengths are important for certain processes, such as photosynthesis, the primary role played by radiation within the climate system is heating and cooling. Hence, for the most part, modelling the flow of radiation in a dynamical model of Earth involves integrations of fluxes

over fairly broadbands. Broadband irradiances, or fluxes, are generally defined for incoming solar radiation and terrestrial radiation emitted by the Earth–atmosphere system. While these sources overlap, they are generally considered to be exclusive and to span wavelengths between $[0.2, 5]$ μm for solar and $[5, 50]$ μm for terrestrial. Like their counterparts in the spatial domain, atmospheric spectral properties have to be defined at fairly course resolutions for radiation calculations in GCMs.

Given that the central topic of this volume is scattering of light, and the fact that solar radiation is scattered in the Earth–atmosphere system to a much greater extent than terrestrial radiation[1], this chapter focuses overwhelmingly on the treatment of solar radiative transfer within GCMs. Hence, the purpose of this chapter is to briefly review methods for computing solar radiative transfer in GCMs and to speculate on both deficiencies in these methods and how far we should be going to address these deficiencies.

The second section of this chapter gives a brief overview of how radiation figures into the climate system and the important role that radiation plays in diagnosing both real and modelled climates. The third section presents the basic method of representing solar radiative transfer in GCMs; two-stream approximations. It also discusses methods that have been proposed to extend two-streams in order that they capture the essence of solar transfer through unresolved inhomogeneous cloudy atmospheres. In the fourth section, 1D (i.e., two-stream-based) solar transfer is contrasted with 3D transfer and it is asked whether global modellers should be concerned about systematic differences. The fifth section highlights some recent work with remote sensing of cloudy atmospheres. Concluding remarks are made in the final section.

## 1.2 Earth's radiation budget and feedbacks

This section provides an overview of the central role played by radiation in global climatology. It discusses the impact of clouds on Earth's radiation budget and the prominent role of radiation in cloud–climate feedback processes.

### 1.2.1 Earth's radiation budget and climatic variables

The fundamental working hypothesis in analysis of global climate is that over a sufficiently long period of time $T$ ($> 1$ year), the Earth–atmosphere system is in radiative equilibrium such that

$$\int_T \left\{ \frac{S_\odot(t)}{4} \left[1 - \alpha_p(t)\right] - \mathcal{I}(t) \right\} \mathrm{d}t = 0, \tag{1.1}$$

[1]Rayleigh scattering is effectively nil at terrestrial wavelengths and the imaginary part of the refractive index for water in the heart of the solar spectrum is several orders of magnitude smaller that it is in the atmospheric thermal window.

where $S_\odot$ is incoming normal solar irradiance (according to the SORCE radiometer, $S_\odot \approx 1361 \pm 0.5\,\mathrm{W\,m^{-2}}$ at 1 AU; see http://lasp.colorado.edu/sorce/tsi_data.html), the factor of 1/4 arises because Earth is spherical, $\alpha_p$ is top of atmosphere (TOA) albedo, and $\mathcal{I}$ is outgoing longwave radiation (OLR). All of the quantities in (1.1) are spectrally-integrated. Climatic variables are often grouped into two classes: *external* and *internal* variables. External variables are those that evolve independently of climate. That is, they affect an effect, but there are no return effects. The ultimate external variable is of course $S_\odot$. Through variations in Earth's orbit about the Sun and variations in solar output (both spectrally and integrated), $S_\odot$ varies constantly (see the SORCE page listed earlier). Volcanic activity is often cited as an external variable. On sub-geologic time-scales this is accurate as volcanic emissions can impact $\alpha_p$ significantly with no discernible reverse impact on volcanic activity. On much longer time-scales, however, volcanoes alter atmospheric composition and it has been speculated that this alters life, ocean sediments, the lithosphere, plate tectonics, and hence volcanic activity (Lovelock 1988).

If a fluctuating variable exhibits excessive intermittency it can fall into a grey zone between internal and external. For instance, desert dust storms, which can alter the vertical distribution of radiation significantly over large regions, are highly intermittent. One could argue that dust storms only impact weather and climate and that their rate and frequency of occurrence does not depend on themselves. But where does one draw the line and impose a threshold? Since the distribution of dust storms depends on local conditions that are naturally tied to large-scale conditions, then since dust storms influence weather, there must be a direct link back to their occurrence. While many factors conspire to bring about a certain distribution of dust storms, dust storms themselves must play a role and so dust storms are not external variables, despite the seeming gulf between their immediate impacts and their sources.

Another grey variable is human activity. On one hand, climate determines greatly the distribution of human settlement and activity. It has been known for many decades that human activity can directly impact climate, as exemplified in studies of desertification (Charney et al. 1977), deforestation (Snyder et al. 2004), and greenhouse gas emissions (Houghton et al. 2001). Desertification and deforestation result in direct alterations to surface albedo, surface roughness, and evapotranspiration. This has direct impacts on local partition of energy available for heat and water fluxes. These alter circulation and moisture patterns and hence local and global climate. Emissions of greenhouse gases alter atmospheric opacity and hence $\mathcal{I}$. The short to medium term impacts on global climate have been studied widely, and while the gross details are recognized generally, specifics (i.e., regional impacts) are still unclear. So, for example, if desertification and deforestation alter regional and global climate, and these alterations force changes in agricultural and silvicultural practices (i.e., human activity), human activity will then be undoubtedly an internal variable. Likewise, if the impacts of global warming via increased levels of greenhouse gases alter human activity, humans are again an internal variable. Indeed, it is difficult to imagine how life in general can be anything other than an internal variable.

An indisputable example of an internal climatic variable is cloud. Clouds are highly intermittent, like dust storms, exist over a wide range of spatial and temporal scales, and demonstrably influence, and get influenced by, their environment. While each cloud, and field that it is associated with, is an individual example, their properties are generally discussed in climatology in terms of spatial and temporal integrals and hence as members of a population. Clearly there is an overall population of clouds, but often clouds are categorized into subpopulations in terms of the meteorological conditions in which they live. As climate fluctuates, correlations of distributions of occurrence of cloud subpopulations can change and so too can the internal characteristics of subpopulations. In the former case, clouds are clouds and it is frequencies of occurrence of subpopulations that fluctuate in space and time. For example, a region subject to a changing climate might experience more cumuliform clouds and fewer stratiform clouds, yet the characteristics of each subpopulation might remain unchanged. In the latter, detailed properties of clouds might change presumably along with changes to frequency distributions of subpopulations. For example, a drier climate might have more wind-blown dust which might alter particle size, turbulent structure, and vertical extent of cumuliform clouds. Changes to subtle properties like these will impact local energy and moisture budgets, circulation, dust occurrence, subtle cloud properties, and ultimately frequencies of occurrence of cloud subpopulations. Either way, the role of clouds, as internal variable, in climate and climatic change is still very murky. Hence, uncertainty about representation of their structural and radiative properties in climate models diminishes confidence in climate model predictions (Houghton et al. 2001).

### 1.2.2 Radiation and climate feedbacks

Conventionally, climate sensitivity has been assessed by assuming that (1.1) holds and that perturbations to (1.1) by amounts $\Delta R$, arising from changing either an internal or external variable, are sudden and followed by restoration of equilibrium so that (1.1) holds again. In actuality, $\Delta R$ are generally time-dependent (i.e., forcings are generally transient) thereby inexorably intertwining forcing(s), restoration of equilibrium, and internal chaotic behaviour.

Aires and Rossow (2003) developed a general multivariate expression for the time evolution of changes to TOA net flux $F$. For brevity and simplicity, discussion of their formulation picks up after several basic assumptions have been made that limit the nature, and make for more tractable analyses, of the climate system. Assume that a time-dependent external forcing $\Delta R$ is applied to the climate system and that it acts only on $F$. Thus, $\Delta R$ represents a radiative perturbation such as, for example, a change to the solar constant or a volcanic eruption. This results in changes to internal climatic variables $x_i$ that interact with one another and thus alter $F$ further. Aires and Rossow expressed the change in $F$ as a function of time $t$ as

$$\Delta F\left(t_0+2\Delta t\right) \approx \Delta R\left(t_0+2\Delta t\right)+\sum_i \frac{\partial F\left(t_0+2\Delta t\right)}{\partial x_i\left(t_0+\Delta t\right)}\Delta x_i\left(t_0+\Delta t\right) + \sum_i\sum_j \frac{\partial F\left(t_0+2\Delta t\right)}{\partial x_i\left(t_0+\Delta t\right)}\frac{\partial x_i\left(t_0+\Delta t\right)}{\partial x_j\left(t_0\right)}\Delta x_j\left(t_0\right)\,, \tag{1.2}$$

where $\Delta t$ is timestep. A large step towards classical analysis of climate models is to assume that the external forcing $\Delta R$ on $F$ directly impacts a single *diagnosed* variable, considered here (and most often) to be surface air temperature $T_s$, with negligible direct impact on other variables. This simplifies (1.2) to

$$\Delta F\left(t_0+2\Delta t\right) \approx \Delta R\left(t_0+2\Delta t\right)+\frac{\partial F\left(t_0+2\Delta t\right)}{\partial T_s\left(t_0+\Delta t\right)}\Delta T_s\left(t_0+\Delta t\right) + \underbrace{\sum_i \underbrace{\frac{\partial F\left(t_0+2\Delta t\right)}{\partial x_i\left(t_0+\Delta t\right)}}_{\text{radiative sensitivities}}\underbrace{\frac{\partial x_i\left(t_0+\Delta t\right)}{\partial T_s\left(t_0\right)}}_{\text{state relations}}}_{\text{feedbacks}}\Delta T_s\left(t_0\right)\,, \tag{1.3}$$

where it is seen that the feedbacks, or cause and effect relations, consist of a state relation and a radiative sensitivity. Going further and assuming that $\Delta R$ and feedbacks are independent of $t$, and that the system actually makes it to equilibrium (i.e., $\Delta F \to 0$), (1.3) collapses to the classical expression (e.g., Schlesinger and Mitchell 1987)

$$\Delta T_s \approx \frac{-\Delta R}{\dfrac{\partial F}{\partial T_s}+\displaystyle\sum_i \frac{\partial F}{\partial x_i}\frac{\partial x_i}{\partial T_s}} \tag{1.4}$$

which upon expansion of $F$ into its solar and terrestrial components yields the familiar form

$$\Delta T_s \approx \frac{-\Delta R}{\underbrace{\dfrac{\partial \mathcal{I}}{\partial T_s}+\dfrac{S_\odot}{4}\dfrac{\partial \alpha_p}{\partial T_s}}_{\text{initial}}+\underbrace{\displaystyle\sum_i \underbrace{\left[\dfrac{\partial \mathcal{I}}{\partial x_i}+\dfrac{S_\odot}{4}\dfrac{\partial \alpha_p}{\partial x_i}\right]}_{\text{radiative sensitivities}}\underbrace{\dfrac{\partial x_i}{\partial T_s}}_{\text{state relations}}}_{\text{feedbacks}}}\,. \tag{1.5}$$

The term labelled *initial* is also referred to as system gain. In the absence of feedbacks, it is the gain that brokers the response of $T_s$ to $\Delta R$. For example, doubling $[CO_2]$ would increase Earth's atmospheric opacity and reduce net longwave radiation at the tropopause resulting in $\Delta R \approx -4\,\mathrm{W\,m^{-2}}$ (Cess et al. 1993). Assuming that $T_s \approx 287\,\mathrm{K}$ and $\alpha_p \approx 0.3$, Earth's effective emissivity is

$$\varepsilon \approx \frac{\frac{S_\odot}{4}\left(1-\alpha_p\right)}{\sigma T_s^4} \approx 0.62, \tag{1.6}$$

and since changing $[CO_2]$ has a negligible impact on $\alpha_p$,

$$\frac{\partial F}{\partial T_s} \approx \frac{\partial \mathcal{I}}{\partial T_s} \approx 4\varepsilon\sigma T_s^3 \approx 3.3\,\mathrm{W\,m^{-2}\,K^{-1}} \tag{1.7}$$

where $\sigma$ is the Stefan–Boltzmann constant. Substituting these numbers into (1.4) implies that in the absence of feedbacks, doubling $[CO_2]$ would result in roughly $\Delta T_s \approx -\Delta R/4\varepsilon\sigma T_s^3 \approx 1.2$ K. The fact that GCM estimates of $\Delta T_s$ for doubling $[CO_2]$ range from $\sim$1 K to $\sim$5 K means that internal climatic variables work together to affect anything from a modest attenuation to a strong enhancement of the gain.

This attenuation and enhancement of system gain arises through feedbacks, or cause and effect relations, between internal variables. These relations are represented by the term in (1.5) labelled *feedbacks*. This is where almost all of the uncertainty about, and research into, climate prediction rests. Grouping feedback processes and gain together into a *global feedback parameter* as

$$\widehat{\Lambda} = \frac{\partial F}{\partial T_s} + \sum_i \frac{\partial F}{\partial x_i}\frac{\partial x_i}{\partial T_s} \tag{1.8}$$

and defining $\widehat{s} = -1/\widehat{\Lambda}$ as the *climate sensitivity parameter*, (1.4) becomes

$$\Delta T_s \approx \widehat{s}\Delta R. \tag{1.9}$$

This formulation is actually applicable to equilibria states only and there is no sound reason to believe that $\widehat{s}$ (or $\widehat{\Lambda}$) is independent of regime or time. Hence the motivation behind Aires and Rossow's (2003) formalism: $\widehat{s}$ depends on climatic state and simply boiling down a GCM simulation to one number is certainly an oversimplification, and at worst misleading.

Nevertheless, it has been proposed (e.g., Gregory et al. 2004; Stowasser et al. 2006) that useful information can be obtained by studying the phase trajectory of a perturbed model's recovery of radiative equilibrium. Begin by defining $\langle R'(t)\rangle$ as a model's time-dependent radiative imbalance at the TOA, where the initial perturbation is $\langle R'(0)\rangle = \Delta R$ and $\langle R'(\infty)\rangle = 0$, and $\langle T_s'(t)\rangle$ as a model's time-dependent change in mean surface temperature, where initially $\langle T_s'(0)\rangle = 0$ and equilibrium temperature change is $\langle T_s'(\infty)\rangle$. Then, plotting $\langle R'(t)\rangle$ vs. $\langle T_s'(t)\rangle$, one often finds that for long stretches of time, following an adjustment period where the GCM recovers from the shock of having $\Delta R$ administered to it suddenly,

$$\langle R'(t)\rangle \approx a + b\,\langle T_s'(t)\rangle \tag{1.10}$$

is a fair approximation in which $b = \widehat{\Lambda}$. In general, however, one could fit the results with some curvilinear function and analyze

$$\frac{\partial\,\langle T_s'\rangle}{\partial\,\langle R'\rangle} = -\widehat{s}(t)\,. \tag{1.11}$$

As discussed by Aires and Rossow and Barker and Räisänen (2005), isolating individual feedback relations and obtaining an estimate of $\widehat{\Lambda}$ from observations

is a daunting task relative to studying only radiative sensitivities (i.e., $\partial F/\partial x_i$). Once one has profiles of information such as cloud fraction, mean water paths, variance of water paths, they can, however, be varied thereby estimating radiative sensitivities numerically. This goes for both model data and data inferred from observations that provide vertical profiles and horizontal transects of cloud properties (such as, for example, profiles obtained from the Atmospheric Radiation Measurement (ARM) Program's surface sites and active/passive satellite systems such as the A-train and EarthCARE).

In addition to sensitivities, radiative uncertainties can be defined as

$$\Delta F_{x_i} \approx \frac{\partial F}{\partial x_i} \Delta x_i \,, \tag{1.12}$$

where $\Delta x_i$ is uncertainty for an internal variable $x_i$. Given a reasonable estimate of $\Delta x_i$, study of radiative uncertainties could help guide the extent to which effort should be expended on developing subgrid-scale parametrizations of radiative properties. For instance, if one is parametrizing $x_i$, yet it turns out that $\partial F/\partial x_i$, and ultimately $\Delta F_{x_i}$, is very small for a realistic $\Delta x_i$, a parametrization as simple as a judicious global setting, as opposed to a detailed parametrization, that could be years in the making, may be sufficient.

The purpose of this discussion on climate sensitivities and analyses is to point out the central role of radiation and radiative transfer in both the climate system and models that attempt to capture some of its characteristics. In light of this, it is interesting to note that radiative transfer model intercomparison studies reported by Fouquart et al. (1991) and Barker et al. (2003) indicate that when several different radiative transfer models act on identical clear and cloudy atmospheres, the range of responses can be surprisingly large. Thus, it is still unclear how much of the disparity among GCM feedback parameters is due to different treatments of clouds, their optical properties, and different treatments of radiative transfer (particularly for cloudy atmospheres). Similarly, Collins (pers. comm., 2005) show that the much more straightforward radiative forcings due to changes in trace gas concentrations are still in question. This leads to GCM differences right off the top as *standardized* forcings used in GCM intercomparisons differ.

To summarize, representation of radiative transfer is crucial for confident prediction of climate and assessment of climate models. As such, the following sections discuss some current issues facing modelling of solar radiation in climate models.

## 1.3 Solar radiative transfer for global models

For reasons of tractability and justifiability in the face of numerous assumptions and uncertainties, essentially all global models employ two-stream approximations to solve for atmospheric radiative transfer. Since it appears that this will be the case for some time to come, barring the occasional jump to more sophisticated models (e.g., Gu and Liou 2001), this section gives a brief account of

two-stream approximations for plane-parallel, homogeneous conditions, methods for extending two-streams, and finally solar transfer for Earth surfaces.

### 1.3.1 The Independent Column Approximation (ICA)

If one is provided with a surface–atmosphere domain $\mathcal{D}$ whose light attenuation properties can be described in three dimensions, domain-average albedo $\langle R \rangle$ (or transmittance, or flux in general) can be computed with an *exact* solution of the radiative transfer equation such that

$$\langle R \rangle = \iint_{\mathcal{D}} R_{3\mathrm{D}}\,(x, y)\ \mathrm{d}x\,\mathrm{d}y \ \Big/ \iint_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,, \tag{1.13}$$

where $R_{3\mathrm{D}}$ is albedo from a solution that accounts for the 3D flow of radiation. Now divide $\mathcal{D}$ into subcolumns and assume that radiation flows through each subcolumn independently of all other subcolumns, regardless of the cross-sectional area of the subcolumns, and that flow through each subcolumn can be described by 1D transport theory. This is the independent column approximation (ICA) of (1.13) which can be expressed as

$$\langle R \rangle = \iint_{\mathcal{D}} R_{1\mathrm{D}}\,(x, y)\ \mathrm{d}x\,\mathrm{d}y \ \Big/ \iint_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,, \tag{1.14}$$

where $R_{1\mathrm{D}}$ is from a 1D radiative transfer model that can range from a two-stream approximation to a Monte Carlo algorithm. Alternatively, (1.14) can be expressed as

$$\langle R \rangle = \sum_{n=1}^{\mathcal{N}} a(n) R_{1\mathrm{D}}\,(n) \Big/ \sum_{n=1}^{\mathcal{N}} a(n)\,, \tag{1.15}$$

where $\mathcal{D}$ is now recognized as consisting of $\mathcal{N}$ subcolumns of cross-sectional area $a(n)$. In most cases, $a$ are equal for all $n$, and so (1.15) becomes simply

$$\langle R \rangle = \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} R_{1\mathrm{D}}\,(n)\ . \tag{1.16}$$

On several occasions the ICA has performed very well for many different cloud regimes (Cahalan et al. 1994; Chambers et al. 1997; Barker et al. 1999; Benner and Evans 2001). Where (1.13) and (1.14) differ most is for large solar zenith angles $\theta_0$ when energy input is small. The ICA tends to become an increasingly better approximation of (1.13) as $\mathcal{D}$ increases in size and temporal integration lengthens (Benner and Evans 2001). Thus, the ICA seems to be a reasonable standard for less rigorous models to aim for, especially when descriptions of the multi-point statistics of $\mathcal{D}$ are uncertain or unknown.

For GCMs, the cross-sectional area of $\mathcal{D}$ generally exceeds $10^4$ km$^2$. Moreover, descriptions of unresolved fluctuations in optical properties inside $\mathcal{D}$ are, almost by definition, lacking. Hence, GCMs can justify using only 1D radiative transfer models within the ICA framework. The current paradigm is to apply

1D codes based on the two-stream approximation in which unresolved variability is either reduced to fractional coverage of homogeneous clouds that overlap according to extremely idealized configurations or incorporated directly into 1D transport solvers. Thus, the following subsections review two-stream approximations, their limitations, and some of the more popular attempts to extend their use in conjunction with the ICA.

### 1.3.2 Fluxes for single layers: the two-stream approximation

The steady-state, elastic radiative transfer equation can be written as

$$\mathbf{\Omega} \cdot \nabla I(\mathbf{x}, \mathbf{\Omega}) = \sigma(\mathbf{x}) I(\mathbf{x}, \mathbf{\Omega}) - \sigma_s(\mathbf{x}) \int p(\mathbf{\Omega} \cdot \mathbf{\Omega}') I(\mathbf{x}, \mathbf{\Omega}') \, \mathrm{d}\mathbf{\Omega}' - f(\mathbf{x}, \mathbf{\Omega}) \quad (1.17)$$

where $\mathbf{x}$ is position, $\mathbf{\Omega}$ is direction, $I$ is radiance, $\sigma$ is extinction coefficient, $\sigma_s$ is scattering coefficient, $p$ is scattering phase function describing the probability of radiation incident from direction $\mathbf{\Omega}$ being scattered into direction $\mathbf{\Omega}'$, and $f$ is the attenuated source term. Exact solution of this equation for a general medium requires both much information regarding the nature of the medium and much computational power. Therefore, approximations are required for global models with the basic requirement being computation of reflectance and transmittance for individual model layers.

First, it is assumed that

$$\frac{\partial_{\substack{\text{optical} \\ \text{properties}}}}{\partial x} = \frac{\partial_{\substack{\text{optical} \\ \text{properties}}}}{\partial y} = \frac{\partial I}{\partial x} = \frac{\partial I}{\partial y} = 0 \,. \quad (1.18)$$

which eliminates horizontal fluctuations in the atmosphere, surface, and radiation field. This simplifies (1.17) to the azimuthally-averaged 1D equation of transfer that can be written as

$$\begin{aligned} \mu \frac{\mathrm{d}I(\tau, \mu)}{\mathrm{d}\tau} &= I(\tau, \mu) - \frac{\omega_0}{2} \int_{-1}^{1} p(\mu; \mu') I(\tau, \mu') \, \mathrm{d}\mu' \\ &\quad - \frac{F_0}{4} \omega_0 p(\mu; \mu_0) \, \mathrm{e}^{-\tau/\mu_0} \,, \end{aligned} \quad (1.19)$$

where all terms have been azimuthally-averaged, $F_0$ is incoming solar at the top of atmosphere (TOA), $\mu$ is cosine of zenith angle, $\mu_0 = \cos\theta_0$, and

$$\mathrm{d}\tau = \sigma \, \mathrm{d}s; \qquad \omega_0 = \sigma_s / \sigma \,, \quad (1.20)$$

where $s$ is geometric distance, $\tau$ is optical thickness, and $\omega_0$ is single scattering albedo. Defining

$$F^{\pm}(\tau, \mu_0) = \int_0^1 \mu I(\tau, \pm\mu) \, \mathrm{d}\mu \quad (1.21)$$

as upwelling and downwelling irradiances, and applying the operators $\int_0^1 I(\tau,\mu) \, \mathrm{d}\mu$ and $\int_{-1}^0 I(\tau, \mu) \, \mathrm{d}\mu$ to (1.19) yields the coupled equations

$$\begin{cases} \dfrac{\mathrm{d}F^+(\tau,\mu_0)}{\mathrm{d}\tau} &= \gamma_1 F^+(\tau,\mu_0) - \gamma_2 F^-(\tau,\mu_0) - \dfrac{F_0}{4}\omega_0\gamma_3\,\mathrm{e}^{-\tau/\mu_0} \\ \dfrac{\mathrm{d}F^-(\tau,\mu_0)}{\mathrm{d}\tau} &= \gamma_2 F^+(\tau,\mu_0) - \gamma_1 F^-(\tau,\mu_0) + \dfrac{F_0}{4}\omega_0\gamma_4\,\mathrm{e}^{-\tau/\mu_0} \end{cases} \tag{1.22}$$

that can be solved readily by standard methods subject to specific boundary conditions where the coefficients $\gamma_1, \ldots, \gamma_4$ depend on assumptions made about $I$ and $p$, as well as on $\mu_0$ and optical properties. The general two-stream solution to (1.22) that describes layer reflectance and transmittance for absorbing layers irradiated by a collimated-beam from above with no upwelling or downwelling diffuse irradiances on the boundaries (Meador and Weaver 1980) is

$$R_{\mathrm{pp}}(\tau) = \frac{\omega_0}{\alpha}\frac{r_+\,\mathrm{e}^{k\tau} - r_-\,\mathrm{e}^{-k\tau} - r\,\mathrm{e}^{-\tau/\mu_0}}{\mathrm{e}^{k\tau} - \beta\,\mathrm{e}^{-k\tau}} \tag{1.23}$$

and

$$T_{\mathrm{pp}}(\tau) = \mathrm{e}^{-\tau/\mu_0}\left\{1 - \frac{\omega_0}{\alpha}\frac{t_+\,\mathrm{e}^{k\tau} - t_-\,\mathrm{e}^{-k\tau} - t\,\mathrm{e}^{-\tau/\mu_0}}{\mathrm{e}^{k\tau} - \beta\,\mathrm{e}^{-k\tau}}\right\}, \tag{1.24}$$

where

$$r_\pm = (1 \mp k\mu_0)\,(\gamma_1\gamma_3 - \gamma_2\gamma_4 \pm k\gamma_3)\;;\; r = 2k\,[\gamma_3 - (\gamma_1\gamma_3 - \gamma_2\gamma_4)\,\mu_0]\,,$$

$$t_\pm = (1 \pm k\mu_0)\,(\gamma_1\gamma_4 - \gamma_2\gamma_3 \pm k\gamma_4)\;;\; t = 2k\,[\gamma_4 - (\gamma_1\gamma_4 - \gamma_2\gamma_3)\,\mu_0]\,,$$

$$\alpha = \left[1 - (k\mu_0)^2\right](k + \gamma_1)\;;\; k = \sqrt{\gamma_1^2 - \gamma_2^2}\;;\; \beta = -\frac{k - \gamma_1}{k + \gamma_1}\,.$$

Corresponding solutions when only isotropic diffuse irradiance is incident from above are

$$r_{\mathrm{pp}}(\tau) = \frac{\gamma_2(1 - \mathrm{e}^{-2k\tau})}{k + \gamma_1 + (k - \gamma_2)\,\mathrm{e}^{-2k\tau}} \tag{1.25}$$

and

$$t_{\mathrm{pp}}(\tau) = \frac{2k\,\mathrm{e}^{-k\tau}}{k + \gamma_1 + (k - \gamma_2)\,\mathrm{e}^{-2k\tau}}\,. \tag{1.26}$$

It can be verified that there is a removable singularity in (1.23) and (1.24) as $\omega_0 \to 1$. Thus, a separate solution to (1.22) for conservative scattering, $\omega_0 = 1$, has to be obtained which for collimated irradiance leads to (see Meador and Weaver 1980)

$$R_{\mathrm{pp}}(\tau) = \frac{\gamma_1\tau + (\gamma_3 - \gamma_1\mu_0)\left(1 - e^{-\tau/\mu_0}\right)}{1 + \gamma_1\tau} = 1 - T_{\mathrm{pp}}(\tau)\,, \tag{1.27}$$

while for diffuse irradiance

$$r_{\mathrm{pp}}(\tau) = \frac{\gamma_1\tau}{k + \gamma_1} = 1 - t_{\mathrm{pp}}(\tau)\,. \tag{1.28}$$

Numerous two-stream approximations can be defined depending on assumptions made about the nature of phase functions and the scattered radiance field

(see Meador and Weaver 1980; Zdunkowski et al. 1980; King and Harshvardhan 1986; Appendix of this chapter). Over the past two decades, however, the approximation of choice among modelling groups has been delta-two-stream approximations. For these approximations, $p$ is expressed as a combination of a smoothly varying portion and a sharp forward scattering peak:

$$p_\delta(\mu,\mu') \equiv 2g^2\delta(\mu-\mu') + \left(1-g^2\right)\left(1+\frac{3g\mu\mu'}{1+g}\right), \tag{1.29}$$

where

$$g = \frac{1}{2}\int_{-1}^{1} \mu' p(\mu;\mu')\,\mathrm{d}\mu' \tag{1.30}$$

is asymmetry parameter, and $\delta$ is the Dirac distribution (Joseph et al. 1976). Figure 1.1 shows that for cloud droplets this approximation is good. It is straightforward to show that when this transformation is applied, all of the above presented equations are recovered, but layer optical properties are scaled as

$$\begin{aligned} \tau' &= (1-\omega_0 g^2)\tau \\ \omega_0' &= \frac{(1-g^2)\omega_0}{(1-\omega_0 g^2)} \\ g' &= \frac{g}{1+g}\,. \end{aligned} \tag{1.31}$$

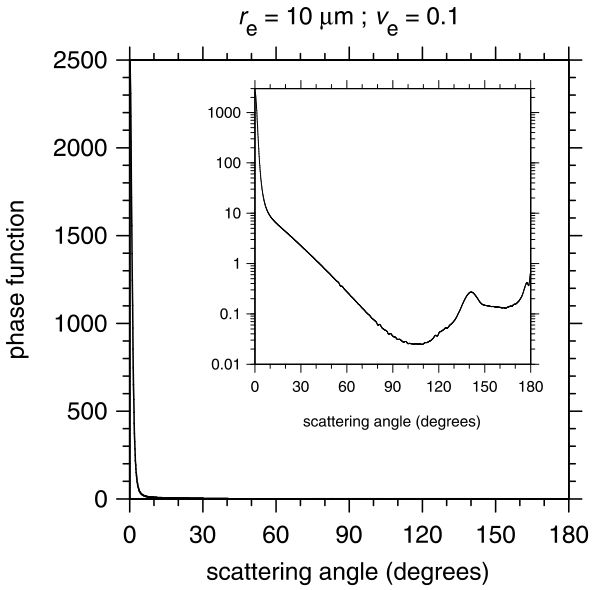



**Fig. 1.1.** Inset plot shows a log plot of the Mie scattering phase function as a function of scattering angle for a droplet size distribution with effective radius of 10 µm and effective variance of 0.1 at wavelength 0.6 µm. This is the usual way these functions are presented. The outer plot is the same as the inner except it is on a linear scale. Shown this way it is immediately apparent that the *delta* approximation in (1.29) is perfectly adequate for cloud droplets.

Delta-two-streams generally perform well for energetically-important high Sun conditions, but they scatter too little in near-forward directions and so for low Sun conditions they yield systematic underestimates of cloud and aerosol albedo. This is illustrated in Fig. 1.2 and may be of some concern in polar regions where during summer $\mu_0$ is always small yet solar input is large (given long Sun-up periods). This deficiency in delta-two-streams can be largely rectified by shifting to the more computationally demanding delta-four-stream approximation (e.g., Li and Ramaswamy 1996).

Twenty years ago, King and Harshvardhan (1986) concluded that because of the assumptions upon which two-streams are based, no single two-stream approximation performs well under all conditions all the time. While this is still true today, (delta-) two-streams are unlikely, however, to be replaced, by another analytic model, as the model of choice in global models. This is because they are computationally inexpensive and given the crude cloud properties they operate on, it makes little sense to replace them with more sophisticated models such as, for example, four-streams or diffusion approximations.

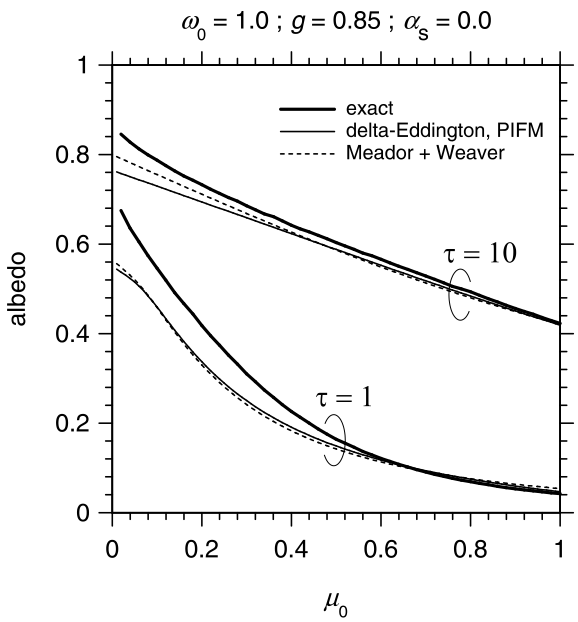



**Fig. 1.2.** Albedos for single layer, conservative scattering, homogeneous clouds (two optical depths as listed on the plot) as functions of cosine of solar zenith angle $\mu_0$ using $g = 0.85$ and a black underlying surface. Heavy solid lines are *exact* solutions computed by DISORT (Stamnes et al. 1988), and other lines are two-stream approximations as listed. Note that for $\omega_0 = 1$, the delta-Eddington and Practical Improved Flux Method (PIFM) (Zdunkowski et al. 1980) are equivalent.

### 1.3.3 Linking layers

With reflectances and transmittances for individual layers in hand, they must be linked vertically in order to provide what global models are really after: heating rate profiles and surface absorption. This is achieved best with adding methods. For a simple two-layer system, total reflectance and transmittance are defined as

$$R_{1,2}(\mu_0) = R_1(\mu_0) + \frac{t_1 \left\{ \left[ T_1(\mu_0) - T_1^{\mathrm{dir}} \right] r_2 + T_1^{\mathrm{dir}} R_2(\mu_0) \right\}}{1 - r_1 r_2} \tag{1.32}$$

and

$$T_{1,2}(\mu_0) = T_1^{\mathrm{dir}} T_2(\mu_0) + \frac{t_2 \left\{ \left[ T_1(\mu_0) - T_1^{\mathrm{dir}} \right] + T_1^{\mathrm{dir}} R_2(\mu_0) r_1 \right\}}{1 - r_1 r_2} \,, \tag{1.33}$$

where $T(\mu_0)$ and $R(\mu_0)$ are equivalent to $T_{\mathrm{pp}}$ and $R_{\mathrm{pp}}$ as derived above,

$$T_1^{\mathrm{dir}} = \mathrm{e}^{-\tau_1/\mu_0} \,, \tag{1.34}$$

and $t$ and $r$ are layer transmittances and reflectances for isotropic diffuse irradiance (see Liou 1992). The terms in these equations are portrayed schematically in Fig. 1.3. Note that the geometric sum formula has been applied under the assumption that internal reflectances and transmittances are all equal for all reflections. This is not the case for inhomogeneous media (see Barker and Davies 1992).

The expressions in (1.32) and (1.33) can be generalized to $N$ number of layers. Flux profiles are then constructed by working up and down through the atmosphere computing reflectances and transmittances for collections of layers using

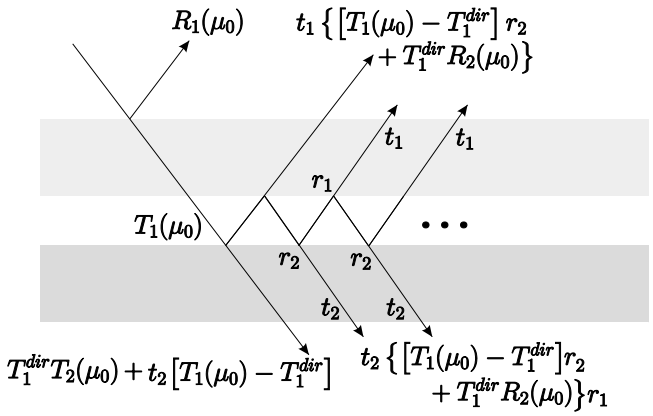



**Fig. 1.3.** Schematic showing how layers are linked in a SW adding scheme.

$$
\begin{aligned}
F_i^{\uparrow} &= \mu_0 S \left\{ \frac{T_{1,i-1}^{\mathrm{dir}} R_{i,N}(\mu_0) + \left[ T_{1,i-1}(\mu_0) - T_{1,i-1}^{\mathrm{dir}} \right] r_{i,N}}{1 - r_{i-1,1} r_{i,N}} \right\} \\
F_i^{\downarrow} &= \mu_0 S \left\{ T_{1,i-1}^{\mathrm{dir}} + \frac{T_{1,i-1}^{\mathrm{dir}} R_{i,N}(\mu_0) r_{1,i-1} + \left[ T_{1,i-1}(\mu_0) - T_{1,i-1}^{\mathrm{dir}} \right]}{1 - r_{i-1,1} r_{i,N}} \right\} ,
\end{aligned}
\tag{1.35}
$$

where double-subscripted diffuse reflectances represent values for collections of layers as indicated by the subscripts, and

$$
T_{1,i-1}^{\mathrm{dir}} = \exp\left[ \sum_{k=1}^{i-1} \tau_k / \mu_0 \right] . \tag{1.36}
$$

Expressions like (1.35) can be computed for both clear and cloudy portions of an atmosphere. Layers can then be linked depending on the desired nature of vertical overlap of cloud. The most common way to proceed is to assume that clouds in adjacent layers are maximally overlapped and that collections of layers containing contiguous clouds that are separated by cloudless layers are randomly overlapped (e.g., Geleyn and Hollingsworth 1979). This approach has become the paradigm despite it being an extreme approximation that depends on model vertical resolution, and systematically underestimating total cloud fraction and atmospheric reflectance (see Barker et al. 2003).

### 1.3.4 When is the two-stream approximation applicable?

As mentioned at the beginning of the previous section, the heart of two-stream approximations is the assumption that the medium and boundary conditions are uniform. Since this applies to individual layers, vertical inhomogeneity is not an issue, at least when layers are homogeneous slabs, as the atmosphere can be partitioned vertically into as many homogeneous layers as one sees fit (Wiscombe 1977). Likewise, it is usually appropriate to use two-stream approximations for computation of flux profiles for cloudless atmospheres. This is because horizontal variations in air density across regions the size of GCM cells are typically very small, and hence so too are variations in scattering efficiencies. Moreover, variations in absorbing gases are generally of second-order importance too, and so use of mean mixing ratios to compute extinction coefficients, and optical depths, are appropriate. The same goes for aerosols in cloudless atmospheres as their concentrations are generally in the linear-response regime for reflectance and transmittance. If the underlying surface albedo is inhomogeneous, simple linear averaging of albedo and using the result in a two-stream approximation will lead to errors (see Barker 2005), but again, these errors are most often second-order.

Where standard two-streams do encounter serious troubles, however, are with realistic cloudy atmospheres. This is because most portions of cloudy atmospheres that span domains the size of typical GCM cells are horizontally inhomogeneous with fluctuations that often bridge the nonlinear response regime for reflectance and transmittance. As such, assumptions upon which two-streams

are based are often seriously violated. Since these are extreme assumptions, one can expect systematic errors which are the most offensive errors for a dynamical model as they manifest themselves as *phantom* forcings. That is, their presence goes undetected because while the two-stream itself might be working perfectly, the dynamical model may demonstrate unreasonable behaviour that could prompt adjustments to be made to unrelated parameters. To make matters worse, clouds are neither totally uncorrelated nor perfectly correlated in the vertical and this confounds attempts to vertically link horizontally inhomogeneous cloudy layers. Furthermore, the layers that a GCM is partitioned into are, for all intents and purposes, arbitrary and so cloud layers in a GCM are not layered clouds that get reported by observers.

Regarding horizontal fluctuations, the bias resulting from use of a regular two-stream can be demonstrated easily. Figure 1.4 shows albedo for two values of optical depth $\tau_1$ and $\tau_2$. If these values occur with equal probability, mean optical depth is simply $(\tau_1 + \tau_2)/2$. If one operates on this mean value with the two-stream one gets an albedo that is sytematicaly greater than the result obtained by operating on each value separately and averaging the responses. This off-set is known as the *homogeneous bias*. It can be stated generally by Jensen's integral inequality (Gradshteyn and Ryzhik 1980) as

$$\int_0^\infty p(\tau) R_{\rm pp}(\tau)\,\mathrm{d}\tau \leqslant R_{\rm pp}\left(\int_0^\infty \tau p(\tau)\,\mathrm{d}\tau\right) = R_{\rm pp}(\overline{\tau})\,. \tag{1.37}$$

For transmittance the inequality is reversed because the concavity of the response function is opposite that for reflectance.

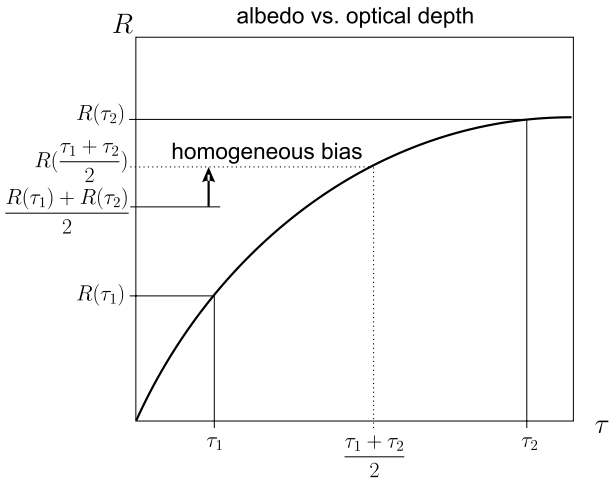



**Fig. 1.4.** Schematic showing albedo $R$ for a plane-parallel, homogeneous cloud layer as a function of optical depth $\tau$. Two equally likely values $\tau_1$ and $\tau_2$ yield responses $R(\tau_1)$ and $R(\tau_2)$ which when averaged produce a mean albedo that is less than that obtained by $R$ operating on mean optical depth $(\tau_1 + \tau_2)/2$.

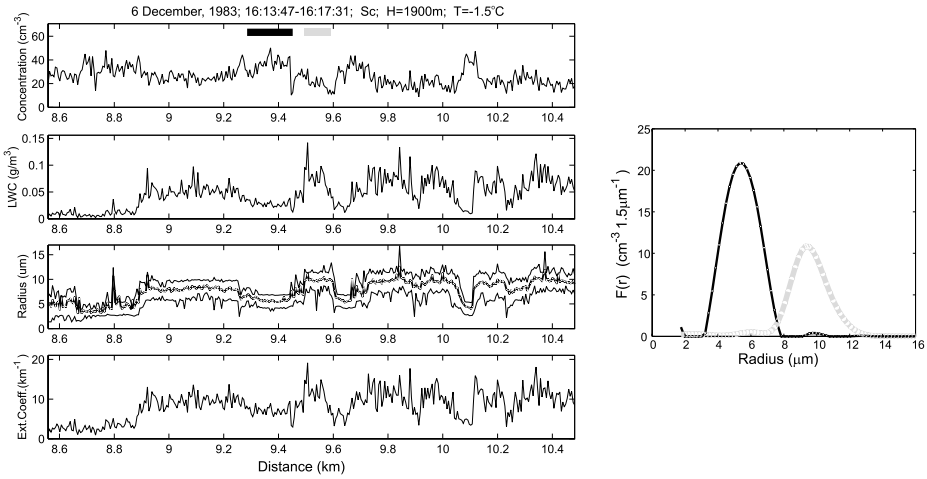



**Fig. 1.5.** Plot in the upper left shows concentration of cloud droplets recorded by an aircraft flying 1900 m above the former Soviet Union on Dec. 6, 1983. Plots beneath this one show cloud liquid water content (LWC), droplet effective radius, and extinction coefficient. Plot on the right shows the droplet size distributions for two sections of the flight as shade-coded on the plot in the upper left. (Data courtesy of A. Korolev, 2002.)

Figure 1.5 shows an example of data collected along a 2.5 km level transect by an instrumented aircraft. It shows that cloud droplet concentration, liquid water content, droplet size distribution, and hence extinction coefficient all vary at all scales. Since situations like this have be expected to occur at unresolved scales in GCMs, it is clear that the homogeneous assumption is untenable (cf. Clothiaux et al. 2005).

Demonstrating limitations of two-streams, or plane-parallel models in general, to vertical overlap of fractional cloud is not as simple. The most common assumption made in GCMs is that when clouds are separated by a cloudless layer they are randomly overlapped, and that when they are in contiguous layers they are maximally overlapped. The latter condition can be confusing. Take for example the three layer system shown in Fig. 1.6 with layer cloud amounts $A_i$. The top left shows true maximal overlap for contiguous clouds. Here there are three possible combinations of total cloud optical depth. The top right shows what happens when one adheres to the maximum-random overlap rule where cloud common to all three layers are maximally overlapped but the overhanging portions in layers 1 and 3 are randomly overlapped. Now there are four combinations of total cloud optical depth. Clearly these scenarios have different total cloud fractions and distributions of vertically integrated optical depth, and so their radiative responses will differ too (see Barker et al. 1999).

In general, however, there is no reason to assume that clouds cut into arbitrary layers, as they are in a GCM, will abide by the vertical resolution-dependent maximum-random overlap rule, whatever one's interpretation of it might be. Rather, one can expect something more general like that shown in the

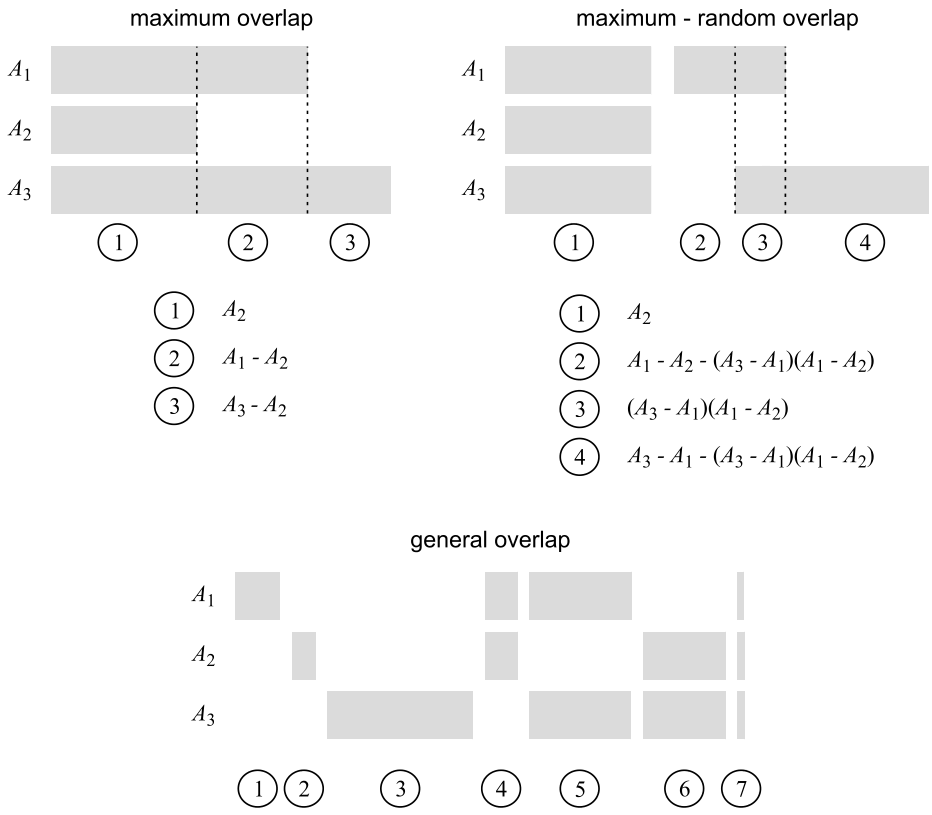



**Fig. 1.6.** Schematic in the upper left shows three contiguous layers of cloud in maximum overlap configuration. Cloud fractions are $A_1$, $A_2$, and $A_3$. Fractional amounts for the three distinct vertical integrals are listed beneath the schematic. The schematic in the upper right shows the same layer cloud fractions in maximum-random overlap mode. Lower schematic shows an example of the same clouds overlapping in a general manner where all seven possible vertical combinations of cloud are realized.

lower part of Fig. 1.6. While this might not differ much from maximum overlap if the three layers are thin, one can expect large differences when there are many more than three layers that extend over a significant fraction of the lower atmosphere; such as with towering convective clouds where a spectrum of clouds in various stages of life can be expected to occur for domains the size of GCM cells (e.g., Hogan and Illingworth 2000; Mace and Benson-Troth 2002; Stephens et al. 2004).

### 1.3.5 Strategies to extend two-stream approximations

Again, because the two-stream is reasonably accurate and computationally very efficient, several attempts have been made to extend its range of application to include horizontally inhomogeneous clouds and various vertical overlap assump-

tions. In this section, the most popular methods, all of which have been used in GCMs, are reviewed and critiqued.

All of these methods are based on the assumption that domain-average reflectance (or transmittance, absorptance, or flux in general) can be computed as

$$\langle R \rangle = (1 - A_{\rm c}) \langle R_{\rm clr} \rangle + A_c \langle R_{\rm cld} \rangle \,, \tag{1.38}$$

where $A_{\rm c}$ is layer cloud fraction, and $\langle R_{\rm clr} \rangle$ and $\langle R_{\rm cld} \rangle$ are mean reflectances associated with the clear and cloudy portions of the layer, respectively. The basic idea here is that radiation that interacts with the cloudy portion of a layer does not interact with the clear portion, and vice versa (cf. Stephens 1988). In this sense, (1.38) is fundamentally an ICA. Moreover, (1.38) relies completely on the concept of cloud fraction which is often discussed and presented in passing without question or hesitation yet as soon as one attempts to describe it with any formality, one recognizes that it is fraught with confusion and misinterpretation. The only reason we passively accept it in (1.38) is because we are approaching the problem with a plane-parallel vision of atmospheric layers and have computation of average fluxes for large domains in mind.

#### 1.3.5.1 Gamma-weighted two-stream approximation (GWTSA)

This model is an example of an explicit independent column approximation (ICA). Its starting point is to rewrite the area integral in (1.14) in its distribution form as

$$\langle R_{\rm cld} \rangle = \int_0^\infty p(\tau) \, R_{\rm 1D}(\tau) \, {\rm d}\tau \,, \tag{1.39}$$

where $p(\tau)$ is a normalized density function that describes variations in $\tau$ over a domain (Ronnholm et al. 1980; Cahalan 1989; Stephens et al. 1991). There are several ways to solve (1.39) depending on the functional forms of $p(\tau)$ and $R_{\rm 1D}(\tau)$. Clearly, if the forms are intractable and require numerical integration, (1.39) is not tenable for use in GCMs. Several studies (e.g., Barker et al. 1996) using satellite-inferred values of $\tau$ and cloud-resolving model data have shown that for domains the size of those used in typical GCMs it is reasonable to represent $p(\tau)$ by a gamma distribution defined as

$$p_\gamma(\tau) = \frac{1}{\Gamma(\nu)} \left( \frac{\nu}{\overline{\tau}} \right)^\nu \tau^{\nu-1} \, {\rm e}^{-\nu\tau/\overline{\tau}} \,, \tag{1.40}$$

where $\nu$ is related to the variance of $\tau$, and $\Gamma(\nu)$ is the gamma function. Note that if particle size distribution is assumed to be constant, (1.40) applies to cloud water content and water path too.

By substituting the generalized, non-conservative scattering, two-stream approximation for collimated irradiance given by (1.23) and (1.24) along with (1.40) into (1.39) leads to the gamma-weighted two-stream approximation (GWTSA) (Barker 1996) in which

$$\langle R_{\rm cld} \rangle = \phi_1^\nu \frac{\omega_0}{\alpha} \left[ r_+ \mathcal{F}(\beta, \nu, \phi_1) - r_- \mathcal{F}(\beta, \nu, \phi_2) - r \mathcal{F}(\beta, \nu, \phi_3) \right] \tag{1.41}$$

and

$$\langle T_{\rm cld}\rangle = \left(\frac{\nu}{\nu+\overline{\tau}/\mu_0}\right)^{\nu} - \phi_1^{\nu}\frac{\omega_0}{\alpha}\left[t_+\mathcal{F}\left(\beta,\nu,\phi_4\right) - t_-\mathcal{F}\left(\beta,\nu,\phi_5\right) - t\mathcal{F}\left(\beta,\nu,\phi_6\right)\right] \tag{1.42}$$

where

$$\mathcal{F}\left(\beta,\nu,\phi\right) = \sum_{n=0}^{\infty}\frac{\beta^n}{\left(\phi+n\right)^{\nu}}\ ;\ |\beta|\le 1,\ \beta\ne 1,\ \nu>0,$$

which is known formally as the Lerch transcendent, and

$$\begin{aligned}
\phi_1 &= \frac{\nu}{2k\overline{\tau}} \quad ;\ \phi_4 = \phi_1 + \frac{1}{2k\mu_0}\\
\phi_2 &= \phi_1 + 1 \quad ;\ \phi_5 = \phi_4 + 1\\
\phi_3 &= \phi_4 + \frac{1}{2} \ ;\ \phi_6 = \phi_1 + \frac{1}{2}.
\end{aligned}$$

As $\omega_0 \to 1.0$, $\beta \to 1$ and one approaches the removable singularity in (1.41) and (1.42). Therefore, substituting (1.27) and (1.40) into (1.39) yields the conservative scattering GWTSA as

$$\begin{aligned}
\langle T_{\rm cld}\rangle &= \left(\frac{\nu}{\gamma_1\overline{\tau}}\right)^{\nu}\left[\left(\gamma_1\mu_0+\gamma_4\right)\mathcal{G}\left(1-\nu,\ \frac{\nu}{\gamma_1\overline{\tau}}\right)\right.\\
&\qquad\left. -\left(\gamma_1\mu_0-\gamma_3\right)\mathcal{G}\left(1-\nu,\ \frac{\nu\mu_0+\overline{\tau}}{\gamma_1\mu_0\overline{\tau}}\right)\right]\\
&= 1-\langle R_{cld}\rangle
\end{aligned} \tag{1.43}$$

where

$$\mathcal{G}\left(1-\nu,x\right) = \mathrm{e}^{x}\Gamma\left(1-\nu,x\right).$$

For diffuse irradiance, the non-conservative solutions are

$$\langle r_{\rm cld}\rangle = \phi_1^{\nu}\frac{\gamma_2}{k+\gamma_1}\left[\mathcal{F}\left(\beta,\nu,\phi_1\right) - \mathcal{F}\left(\beta,\nu,\phi_2\right)\right] \tag{1.44}$$

and

$$\langle t_{\rm cld}\rangle = \phi_1^{\nu}\frac{2k}{k+\gamma_1}\mathcal{F}\left(\beta,\nu,\phi_6\right), \tag{1.45}$$

while the conservative scattering solution is (Oreopoulos and Barker 1999)

$$\langle r_{\rm cld}\rangle = 1-\left(\frac{\nu}{\gamma_1\overline{\tau}}\right)^{\nu}\mathcal{G}\left(1-\nu,\ \frac{\nu}{\gamma_1\overline{\tau}}\right) = 1-\langle t_{\rm cld}\rangle. \tag{1.46}$$

Naturally, other solutions are possible for different representations of $p(\tau)$. For instance, if $p(\tau)$ is approximated by a beta distribution, one ends up with rather intractable solutions for $\langle R_{\rm cld}\rangle$ and $\langle T_{\rm cld}\rangle$ involving hypergeometric functions. A closed-form solution using a lognormal distribution for $\tau$ has not been found as yet. Likewise, it is not known whether closed-form solutions exist for a gamma-weighted four-stream.

To conclude, the attraction of this approach is that as long as one is willing to accept that the underlying distribution of $\tau$ can be represented by $p_\gamma(\tau)$, and that the ICA is a reasonable benchmark for GCM-style radiative transfer models to aim for, the GWTSA represents, by definition, an exact solution for single layers. The GWTSA, for both SW and LW radiation, is being used currently in the CCCma GCM (Li and Barker 2002; Li et al. 2005). It should be noted, however, that the GWTSA encounters problems when it comes to linking layers (see Oreopoulos and Barker 1999).

#### 1.3.5.2 Effective thickness approximation (ETA)

Following from (1.37), an obvious approach to approximate $\langle R_{\rm cld}\rangle$ is

$$\langle R_{\rm cld}\rangle \approx R_{\rm pp}\left[f(\overline{\tau})\right] ,$$

where $f(\overline{\tau}) \leqslant \overline{\tau}$ represents some transformation to $\overline{\tau}$. What makes this approach desirable is that it utilizes directly the efficient two-stream solution with, ideally, only a minor adjustment to its input.

Davis et al. (1990) hypothesized that $\langle R_{\rm cld}\rangle$ could be approximated as

$$\langle R_{\rm cld}\rangle \approx R_{\rm pp}(\overline{\tau}^{\delta}) , \tag{1.47}$$

where $\delta$ was referred to as a co-packing factor. The authors determined $\delta$ from Monte Carlo simulations for very heterogeneous fractal cloud models based on singular cascades. This parametrization found its way into at least one operational GCM (McFarlane et al. 1992). In general $\delta \leq 1$, but for $\overline{\tau} < 1$ $\delta$ has to exceed 1 or its purpose is defeated. This produces the desired effect by reducing the value of $R_{\rm pp}(\overline{\tau})$, but given its simplicity and high level of parametrization, results can be expected to be, at best, very approximate.

Cahalan et al. (1994) advanced the potential applicability of the ETA, in principle for stratocumulus layers only, by noting that expansion of $R_{\rm pp}$ in a Taylor series about $\overline{\log_{10}\tau}$ and averaging over all cells of a bounded cascade model yields

$$\langle R_{\rm cld}\rangle = R_{\rm pp}\left(\eta\overline{\tau}\right) + \sum_{n=1}^{\infty} M_{2n}\frac{\partial^{2n} R_{\rm pp}\left(\eta\overline{\tau}\right)}{\partial\left(\log_{10}\tau\right)^{2n}} , \tag{1.48}$$

where $M_{2n}$ is related to the $2n$th moment of $\log_{10}\tau$, and

$$\eta = \frac{e^{\overline{\ln\tau}}}{\overline{\tau}} \leqslant 1 \tag{1.49}$$

is the reduction factor. Cahalan et al. presented results for (1.48) in its simplest form of

$$\langle R_{\rm cld}\rangle \approx R_{\rm pp}(\eta\overline{\tau}). \tag{1.50}$$

Cahalan et al. (1994, 1995) estimated $\eta$ to be roughly 0.6 to 0.7 for marine boundary layer clouds off the coast of California and on Porto Santo Island. These estimates were obtained by compositing several days' worth of 30 s or

1 min observations of cloud LWP inferred from microwave radiometer data. By compositing data, however, the PPH bias can be made arbitrarily large as more and more variability becomes unresolved and subsumed into $\eta$, or any other measure of variance. Hence, values of $\eta$ that are arrived at by compositing several days of data are applicable to GCMs that call their ETA just once a day or so. Most GCMs call their radiation codes at least once per hour, so if they used values of $\eta$ derived from compositing over extended periods, they would effectively double-count the impact of cloud variability. Subsequent analyses indicate that for most overcast marine boundary layer clouds, $\eta \approx 0.9$ (Barker et al. 1996; Pincus et al. 1999; Rossow et al. 2002). Nevertheless, (1.50) has been used in operational GCMs. For example, Tiedtke (1996) used $\eta \approx 0.7$ for all clouds all the time.

#### 1.3.5.3 ETA vs. GWTSA and the assumption of underlying distributions for $\tau$

Based on the previous paragraph, one might conclude, correctly, that there is still confusion regarding description of unresolved cloud structure in GCMs. This is compounded by the fact that GCMs require information on a per-layer basis, not for entire cloud fields as is often estimated from passive satellite imagery (e.g., Barker et al. 1996; Oreopoulos and Davies 1998). Moreover, even if one felt comfortable about setting values of $\nu$ and $\eta$, there is still the question: what is the underlying distribution of $\tau$? In this subsection we explore the importance of this question.

Basically, there are two situations: a non-analytic representation for $p(\tau)$, or an assumed analytic form for $p(\tau)$. In either case, one could compute, at potentially great computational expense, the ICA solution using (1.16) or (1.39), respectively. On the other hand, it is easy to compute corresponding statistics such as $\overline{\tau}$, $\overline{\ln \tau}$, $\sigma$, $\nu$, and $\eta$. Once computed, they can be used directly in either the ETA or the GWTSA. For the latter, one makes the explicit assumption that for computation of mean radiative fluxes, the gamma distribution is suitable regardless of what $p(\tau)$ the parameters come from. At the same time, the gamma distribution could also drive estimates of $\eta$ to be used in the ETA.

For tractability, $R_{\mathrm{pp}}$ is represented here by Coakley and Chýlek's (1975) 'model 1' approximation for $\omega_0 = 1$ in which

$$R_{\mathrm{pp}}(\tau) = \frac{\beta(\mu_0)\tau}{\mu_0 + \beta(\mu_0)\tau} = 1 - T_{\mathrm{pp}}(\tau)\,, \tag{1.51}$$

where $\beta(\mu_0)$ is the zenith angle dependent backscatter function (Wiscombe and Grams 1976). When an analytic distribution for $\tau$ is used, it is assumed to be $p_\gamma(\tau)$. On the other hand, non-analytic distributions of $\tau$ are realized by the bounded cascade (BC) model of Cahalan et al. (1994). For the most part, the variability of $\tau$ for the BC model depends on the *fractal parameter* $f_0$ which is related to $\eta$ and $M_2$ [see (1.48)] as

$$\eta = \sqrt{\prod_{n=0}^{\infty} (1 - f_0^2 c^{2n})}$$

and

$$M_2 = \sum_{n=1}^{\infty} \left[ \frac{1}{2} \log \left( \frac{1 + f_0 c^n}{1 - f_0 c^n} \right) \right]^2 ,$$

where $c \in (0, 1]$ but is generally set to $2^{-1/3}$.

Consider first the distributions produced by the BC model. These distributions resemble lognormal distributions, and for $f_0 \lesssim 0.3$ they resemble gamma distributions too. It is most likely that for layers with dimensions resembling those found in GCMs $f_0 \in [0.2, 0.7]$. Figure 1.7 shows albedo as a function of $f_0$ for several models. ICA values were computed by (1.16) and differences between ICA and homogeneous are known as the homogeneous bias. The ETA proper is represented by (1.50) where $\eta$ were computed directly from BC data using (1.49). For $\mu_0 = 0.5$ and $\overline{\tau} = 10$, this model does extremely well, but for $\mu_0 = 1$ it overestimates the homogeneous bias systematically. Inclusion of the second term in (1.48), with $M_2$ computed directly from data, improves estimates notably. For $\mu_0 = 1$, however, it still underestimates the ICA systematically.

At this point, assume that the underlying distribution of $\tau$ is $p_\gamma(\tau)$. It can be shown that (1.48) can be expressed as

$$\langle R_{\rm cld} \rangle = R_{\rm pp}(\eta\overline{\tau}) + \left[ \frac{1}{2 \ln 10} \psi_1(\nu) \right] \frac{\partial^2 R_{\rm pp}(\eta\overline{\tau})}{\partial (\ln \tau)^2} + \cdots , \tag{1.52}$$

where

$$\eta = \frac{e^{\psi_0(\nu)}}{\nu} , \tag{1.53}$$

and

$$\psi_n(\nu) = \frac{d^{n+1}}{d\nu^{n+1}} \ln \Gamma(\nu)$$

is the polygamma function. Although $\psi_n(\nu)$ are easy to parameterize, multiple derivatives of $R_{\rm pp}$ are tedious in general. Substituting (1.51) into (1.52) and retaining just the first two terms gives

$$\langle R_{\rm cld} \rangle \approx R_{\rm pp}(\eta\overline{\tau}) \left\{ 1 + \frac{\mu_0 \left[ \mu_0 - \beta(\mu_0)\eta\overline{\tau} \right]}{2 \ln(10) \left[ \mu_0 + \beta(\mu_0)\eta\overline{\tau} \right]^2} \psi_1(\nu) \right\} . \tag{1.54}$$

Even for (1.51) the third term is already too intractable to be of much use. Figure 1.7 shows that when only the leading term in (1.54) is used with $\eta$ computed by (1.53) and $\nu$ by the method of moments (mom), which is defined as

$$\nu_{mom} = \left( \frac{\overline{\tau}}{\sigma} \right)^2 ,$$

where $\sigma$ is standard deviation of $\tau$, the situation worsens greatly. This is because $\nu_{mom}$ is impacted too much by extreme values of $\tau$ and so corresponding $\eta$ are too small as are albedos.

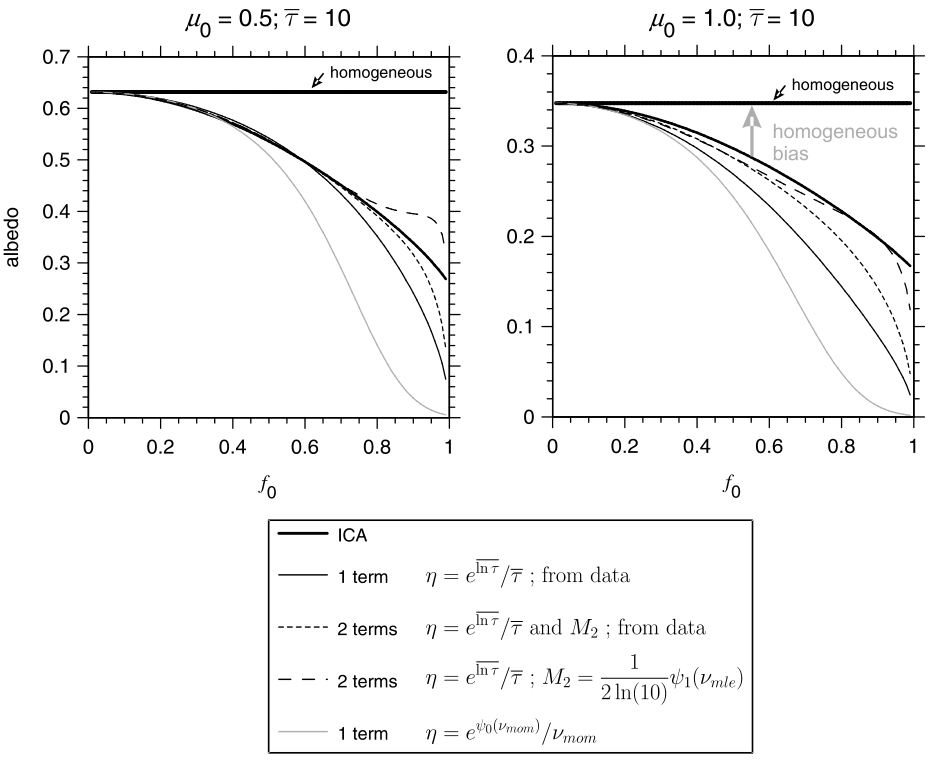



**Fig. 1.7.** Albedo for non-absorbing clouds as a function of the bounded cascade model's fractal parameter $f_0$ ($f_0 = 0$ homogeneous) for two solar zenith angles and $\overline{\tau} = 10$. See the text for a description of the various model results shown here.

The maximum likelihood estimate (MLE) of $\nu$ is

$$\psi_0(\nu_{mle}) + \ln\left(\frac{\overline{\tau}}{\nu_{mle}}\right) - \overline{\ln\tau} = 0, \tag{1.55}$$

which upon rearranging as

$$\frac{e^{\psi_0(\nu_{mle})}}{\nu_{mle}} = \frac{e^{\overline{\ln(\tau)}}}{\overline{\tau}}$$

verifies that explicit computation of $\eta$ from data yields the same reduction factor as does use of the gamma distribution's MLE value for $\nu$ in (1.53). In general, $\nu_{mle}$ differs from $\nu_{mom}$. Interestingly, Fig. 1.7 shows that when one uses $\eta$ computed directly from data, or by (1.53) with $\nu_{mle}$, in conjunction with the second term of (1.54), results generally improve over the use of two terms with parameters set directly from data (especially for $\mu_0 = 1$).

The other end of the spectrum is to assume that distributions of $\tau$ are in actuality gamma distributions. In this case $\nu_{mle} = \nu_{mom}$. Substituting (1.40)

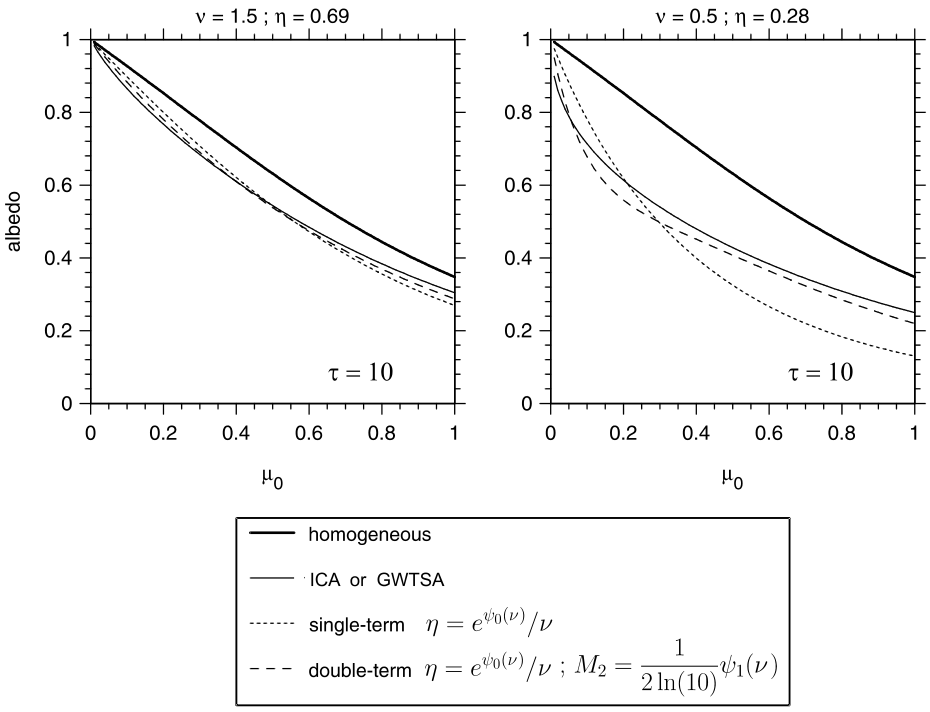



**Fig. 1.8.** Conservative scattering albedo of an overcast cloud layer with $\overline{\tau} = 10$ as a function of cosine of solar zenith angle $\mu_0$. Results are shown for two inhomogeneous clouds with $\nu$ and $\eta$ as listed at the top of each plot, as well as for their homogeneous counterpart which corresponds to $\nu \to \infty$ and $\eta = 1$.

and (1.51) into (1.39) and evaluating the integral yields

$$\langle R_{\rm cld} \rangle = 1 - e^{\xi + \nu \ln \xi} \Gamma\left(1 - \nu,\ \xi\right) = 1 - \langle T_{\rm cld} \rangle \ , \tag{1.56}$$

where

$$\xi = \frac{\nu \mu_0}{\beta(\mu_0)\overline{\tau}},$$

and $\Gamma(1 - \nu,\ \xi)$ is the incomplete gamma function. Equation (1.56) represents one of the simplest forms of the GWTSA. By definition, the GWTSA is now equivalent to the ICA and so, in this situation, produces perfect estimates of the homogeneous bias. Now the tables are turned and, as Fig. 1.8 shows, it is the single term ETA that produces poor estimates of albedo. In fact, it performs well only when $\eta\overline{\tau} \approx \mu_0/\beta(\mu_0)$ or when $M_{2n}$ are small (i.e., variability is weak). Again, however, inclusion of the second term improves estimates significantly over those of the simple ETA.

#### 1.3.5.4 Renormalization of optical properties

Cairns et al. (2000) developed an approximate solution that is based on assumptions similar to Stephens's (1988) method. First, they assume that the number concentration of scatters can be described by

$$n(\mathbf{x}) = \overline{n} + n'(\mathbf{x}) \tag{1.57}$$

where $\overline{n}$ is domain-average number concentration, $n'$ is a local fluctuation at position $\mathbf{x}$, and that fluctuations are isotropic in three-dimensions. Averaging the 3D radiative transfer equation over the domain leads them to domain-average intensity $\overline{I(\mathbf{x}, \mathbf{\Omega})}$ given by

$$\begin{aligned} \mathbf{\Omega} \cdot \nabla \overline{I(\mathbf{x}, \mathbf{\Omega})} \; + \; \sigma \overline{n} \int_{4\pi} B(\mathbf{x} \cdot \mathbf{\Omega}') \overline{I(\mathbf{x}, \mathbf{\Omega}')} \, \mathrm{d}\mathbf{\Omega}' \\ = -\sigma \int_{4\pi} B(\mathbf{\Omega} \cdot \mathbf{\Omega}') \overline{n'(\mathbf{x}) I(\mathbf{x}, \mathbf{\Omega}')} \, \mathrm{d}\mathbf{\Omega}' \, , \end{aligned} \tag{1.58}$$

where

$$B(\mathbf{s} \cdot \mathbf{s}') = \delta(\mathbf{\Omega} \cdot \mathbf{\Omega}' - 1) - \omega_0 p(\mathbf{\Omega} \cdot \mathbf{\Omega}'),$$

and $\mathbf{\Omega}$ is angular direction. In an attempt to close (1.58), Cairns et al. rewrite (1.58) as an integral equation involving a Green's function, perform a perturbation expansion, and re-sum the series. This effectively decouples the term $\overline{n'(\mathbf{x}) I(\mathbf{x}, \mathbf{\Omega}')}$ in (1.58). They then invoke the nonlinear approximation, and assuming the effects of fluctuations are local (i.e., not long-range), they are able to recover the LHS of (1.58) which can be solved as though the medium was homogeneous with the following transformed domain-average optical properties:

$$\begin{aligned} \sigma' &= \sigma \, (1 - \varepsilon) \, , \\ \omega_0' &= \omega_0 \left[ 1 - \frac{\varepsilon}{1 - \varepsilon} \, (1 - \omega_0) \right] , \\ \omega_0' g' &= \omega_0 g \left[ 1 - \frac{\varepsilon}{1 - \varepsilon} \, (1 - \omega_0 g) \right] , \end{aligned} \tag{1.59}$$

where

$$\varepsilon = \frac{1}{2} \left( q - \sqrt{q^2 - 4V} \right) ,$$

and

$$q = \frac{1 + \sigma l_c}{\sigma l_c} \, ,$$

where $V$ is relative variance of $n$, and $l_c$ is effective correlation length of the variations. When $\sigma l_c \approx 1$, particle density fluctuations follow a lognormal distribution. In this case, however, only moderate fluctuations are allowed (i.e., $V < 1$). For a more thorough assessment of this method, see Barker and Davis (2005).

From Cairns et al.'s (2000) initial formulation, it would appear that long-range fluctuations in $n(\mathbf{x})$ are neglected thereby rendering the transformations

in (1.59) applicable on relatively small scales such as individual cells of a stratocumulus or individual cumuli. Cairns et al. allude to the ICA being more suitable at describing the effect of fluctuations larger than the mean diffusion length. Nevertheless, Rossow et al. (2002) applied (1.59) to ISCCP (International Satellite Cloud Climatology Project) data, which has a horizontal resolution of $\sim$5 km, and defined $\varepsilon$ as

$$\varepsilon = 1 - \frac{\widehat{\tau}}{\overline{\tau}}, \tag{1.60}$$

where

$$\widehat{\tau} = R_{\mathrm{pp}}^{-1}\left[\frac{1}{N}\sum_{n=1}^{N} R_{1D}\left(\tau_n\right)\right]$$

in which $N$ is the number of satellite pixels in a large domain, and $\tau_n$ is cloud optical depth inferred for the $n$th pixel. In the appendix to Rossow et al. (2002), it was shown that an accurate approximation relating $\nu_{mle}$ [see (1.55)] and $\varepsilon$ is

$$\nu_{mle} = \frac{1}{\varepsilon - \ln\left(1-\varepsilon\right)}. \tag{1.61}$$

While Cairns et al.'s (2000) model does not suffer from the same ailment that Stephens's (1988) does (i.e., potential violation of the conservation of energy; see Barker and Davis 2005), it would appear from their initial presentation that it is meant to be applied at small scales; perhaps superimposed onto another model designed to account for fluctuations at larger scales, such as the GWTSA.

#### 1.3.5.5 The Monte Carlo Independent Column Approximation (McICA)

Several studies have shown that differences in estimates of domain-averaged flux profiles predicted by the ICA and 3D radiative transfer models are usually small (Cahalan et al. 1994; Barker et al. 1999, 2003; Benner and Evans 2001). So, in light of sketchy descriptions of unresolved clouds that are available to GCM radiation codes, the ICA seems to be a suitable and tractable standard for 1D radiation codes despite its neglect of 3D transfer. The 1D models that address unresolved horizontal fluctuations that have been discussed thus far, and other operational methods that involve vertical linking of layers, lead to: i) additional computation relative to the straight-up, multi-layer two-stream model; ii) limited, and unrealistic, descriptions of unresolved cloud fluctuations (fluctuations for other components are rarely, if ever, addressed); and iii) most important, biases relative to the full ICA. In an attempt to circumvent these limitations, Barker et al. (2002) and Pincus et al. (2003) introduced the Monte Carlo Independent Column Approximation (McICA) which segregates descriptions of surface–atmosphere structure from the GCM's radiative transfer algorithms.

The essence of McICA is stochastic sampling of subcolumns that are unresolved by a GCM as one sweeps across the necessary spectral integral. To begin, full ICA, domain-average monochromatic radiative fluxes are computed as

$$\overline{F} = \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} F_n \,, \tag{1.62}$$

where $F_n$ is monochromatic radiative flux for the $n$th subcolumn of the domain. With the correlated $k$-distribution (CKD) method (e.g., Fu and Liou 1992), broadband radiative fluxes for the $n$th subcolumn are computed as

$$\mathcal{F}_n = \sum_{k=1}^{\mathcal{K}} F_{n,k} \,, \tag{1.63}$$

where $F_{n,k}$ is the contribution from the $k$th quadrature point in $k$-space. Combining (1.62) with (1.63) yields domain-average, broadband fluxes for the full ICA as

$$\overline{\mathcal{F}} = \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \sum_{k=1}^{\mathcal{K}} F_{n,k}. \tag{1.64}$$

GCM radiation codes typically have $\mathcal{K} \sim 30$–$100$ so even for a modest value of $\mathcal{N}$, the double sum in (1.64) is intractable in terms of CPU usage.

The McICA method approximates (1.64) simply as

$$\widehat{\overline{\mathcal{F}}} = \sum_{k=1}^{\mathcal{K}} F_{n_k,k} \,, \tag{1.65}$$

where $F_{n_k,k}$ designates a monochromatic radiative flux for a *randomly selected* subcolumn, denoted as $n_k$, and the circumflex signifies a single sample. The McICA solution (1.65) equals the ICA solution only when all $\mathcal{N}$ columns are identical or when $\mathcal{N} = 1$. In general, McICA's incomplete pairing of subcolumns and spectral intervals ensures that its solution will contain conditional random, but unbiased, errors.

Before proceeding with operational details of this method, it is instructive to show that in the limit of taking $T \to \infty$ samples of (1.65), the ensemble average $\left\langle \widehat{\overline{\mathcal{F}}} \right\rangle$ is identically equal to the ICA. This is easily seen using the CKD method: if there are $N_c$ cloudy columns to select from, then as $T \to \infty$, the $n$th cloudy column and $k$th quadrature point will be paired $f_{n,k}$ times such that the expectation of $\left\langle \widehat{\overline{\mathcal{F}}} \right\rangle$ is

$$E\left(\left\langle \widehat{\overline{\mathcal{F}}} \right\rangle\right) = \lim_{T\to\infty} \frac{1}{T} \left\{ f_{1,1} F_{1,1} + \cdots + f_{N_c,\mathcal{K}} F_{N_c,\mathcal{K}} \right\}. \tag{1.66}$$

Since samples are drawn uniformly, all $f_{n,k} = T/N_c$ which reduces (1.66) to

$$E\left(\left\langle\widehat{\overline{\mathcal{F}}}\right\rangle\right) = \frac{1}{N_c}\left\{F_{1,1}+\cdots+F_{N_c,\mathcal{K}}\right\} = \frac{1}{N_c}\sum_{k=1}^{\mathcal{K}}\sum_{n=1}^{N_c}F_{n,k}\,, \tag{1.67}$$

which is the ICA. Hence, McICA is entirely unbiased with respect to the ICA. Since McICA completely decouples the transfer solver from descriptions of unresolved fluctuations, a GCM using McICA is capable of efficiently performing sensitivity studies for a wide range of subgrid-scale assumptions in an unbiased manner.

Equation (1.62) can be partitioned into clear and cloudy contributions, as with all models discussed thus far, and rewritten as

$$\widehat{\overline{\mathcal{F}}} = (1-C_{tot})\,\mathcal{F}^{clr} + C_{tot}\sum_{k=1}^{\mathcal{K}}F^{cld}_{n_k,k}\,, \tag{1.68}$$

where $C_{tot}$ is total cloud fraction for the GCM column, and *clr* and *cld* refer to cloud-free and cloudy subcolumns respectively. Note that when aerosols, gases, and surfaces are assumed to be horizontally homogeneous, $\mathcal{F}^{clr}$ is noise-free. Now all $\mathcal{K}$ samples are devoted to cloudy subcolumns as opposed to, on average, $C_{tot}\mathcal{K}$ samples for (1.65). Hence, sampling noise for (1.68) is smaller than that for (1.65). Although in principle, computation of $\mathcal{F}^{clr}$ makes (1.68) more expensive than (1.65), this is a moot point as most GCMs routinely compute $\mathcal{F}^{clr}$ diagnostically in order to compute cloud radiative effects (forcings).

McICA variance $\sigma^2$ can be reduced further by taking more than $\mathcal{K}$ samples. This transforms (1.68) into the more general form presented originally by Barker et al. (2002):

$$\widehat{\overline{\mathcal{F}}} = (1-C_{tot})\,\mathcal{F}^{clr} + C_{tot}\sum_{k=1}^{\mathcal{K}}\left[\frac{1}{N_k}\sum_{n=1}^{N_k}F^{cld}_{n,k}\right]\,, \tag{1.69}$$

where the total number of samples is

$$\mathcal{N}' = \mathcal{K} + \sum_{k=1}^{\mathcal{K}}(N_k-1) = \mathcal{K}+\mathcal{M}\,.$$

As such, radiative fluxes for point $k$ are computed for $N_k$ randomly selected cloudy subcolumns, and results are averaged for regular CKD summation.

Räisänen and Barker (2004) presented a procedure to find the optimal set of $N_k$ when McICA noise arises overwhelmingly from clouds. The optimal set of $N_k$ depends on the quantity whose random errors are to be minimized. In a GCM, different predefined sets of $N_k$ could be used based on the state of the Earth–atmosphere column. The most obvious distinction is between land and ocean surfaces. Land surfaces respond rapidly to changes in net solar radiation during daytime, whereas sea surface temperatures change slowly. Therefore, for

the SW, it is probably best to base optimization of $N_k$ on CRE at the surface when over land during Sun-up periods but on CRE for lower atmospheric heating when over ocean. Moreover, separate sets of $N_k$ could be defined based on land surface type (e.g., vegetated or snow and ice surfaces). In the LW, optimization of $N_k$ should probably be based on CRE of atmospheric heating regardless of location and time.

Given the size of GCM grid-spacings and the paucity of information regarding subgrid-scale cloud structure, it is natural to expect domain-averaged radiative heating profiles to be characterized by conditional distributions whose widths are wider than that of the Dirac delta distribution. This *natural noise*, or uncertainty, buys some leeway for random noise to be present in subgrid-scale parametrizations. Beyond this natural noise, it may be that GCMs can *consume* additional unbiased random noise generated by subgrid-scale parametrizations with little, or no, statistically significant impact on performance. The obvious, and simple, argument for this is that noise of this type, at or near the spatial and temporal inner-scales of a GCM, is incapable of spawning organized *structures* that significantly affect the trajectory, and perhaps even one-point statistics, of the overall simulated climate.

Figure 1.9 shows the impact of McICA noise on screen temperature simulated by two GCMs. The simulations were 15 days long and 10-member ensembles were run for each rendition of McICA. The benchmark simulation used (1.69) with $\mathcal{M} = 1000$ which almost squelches random noise entirely. This plot shows that the CCCma GCM is insensitive to McICA noise as the fraction of the globe exhibiting differences, relative to the benchmark simulation, that are statistically significant at the $\alpha\%$ confidence level is always close to $1-\alpha$ regardless of McICA noise level. These fractions of statistically significant differences are expected for two samples drawn at random from the same population. For the CAM-3 GCM it is clear that it is impacted by noise as noisy renditions of McICA exhibit statistically significant differences over a large portion of the global. However, this sensitivity can be reduced to effectively nothing by using (1.69) with $\mathcal{M} = \mathcal{K}/2$.

### 1.3.6 Surface albedo

When discussing problems related to fluctuations that are unresolved by conventional GCMs, clouds receive, by far, the most attention. But fluctuations in surface type deserve mention too. Spectral surface albedos $\alpha_s(\lambda)$ are used as lower boundary conditions for atmospheric radiative transfer models. What is usually buried under the convenient blanket of *surface albedo* is actually the fraction of photons not absorbed by anything below either a water surface, a soil surface, or a vegetation canopy. Several studies have examined the reflectance properties of surfaces in detail (e.g., Hapke 1981; Ross 1981; Preisendorfer and Mobley 1986; Verstraete 1987). The purpose of this section is to briefly comment on just a few of the myriad of problems facing specification of surface albedo in global models.

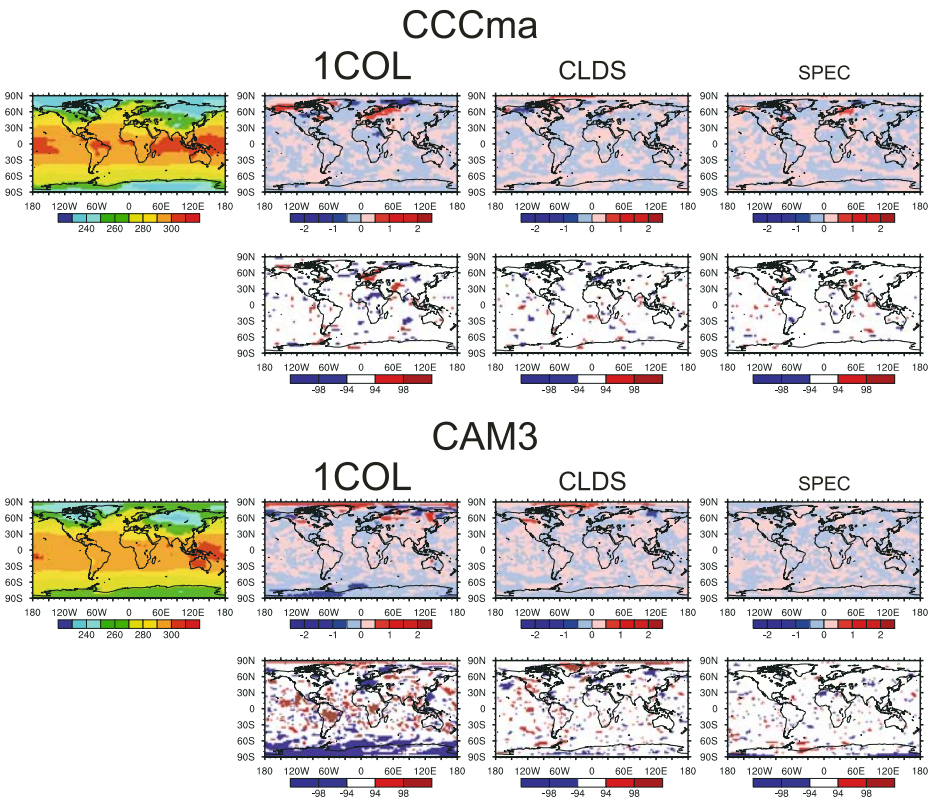



**Fig. 1.9.** Plots on the left show mean screen temperature (K) for the CCCma GCM and the CAM3 GCM for 15-day simulations using fixed SSTs and sea-ice cover. These results are for benchmark simulations that used almost noiseless versions of McICA. Upper plots in the two sets of plots to the right show screen temperature differences between the benchmarks and three versions of McICA; going from left to right, and as indicated by the font size of the map titles, McICA noise begins large with 1COL [a single sample is drawn to represent all $N_k$ in (1.65)], diminishes to intermediate values for CLDS [which uses (1.68)], and finally to small values for SPEC [using (1.69) with $\mathcal{M} = \mathcal{K}/2$]. See Räisänen and Barker (2004) for details. Lower plots on indicates area where differences were significant at the 95% and 99% confidence levels.

Consider first albedo for land surfaces. Figure 1.10 shows visible and near-IR $\alpha_s(\lambda)$ for various surface types at $\theta_0$ near 60°. Clearly, broadband radiative transfer models face the same problem with surface albedo as they do with atmospheric constituents: they require spectrally-weighted values, but the necessary spectral irradiances needed to compute mean values, for relatively broad bands, are not available. Generally, simple uniform weightings are applied and surfaces are assumed to be perfectly Lambertian. In many applications, only a gross distinction is drawn between albedo in a few bands across the solar spectrum (e.g., $<0.7\,\mu$m and $>0.7\,\mu$m). Likewise, dependencies on $\theta_0$ are generally crude

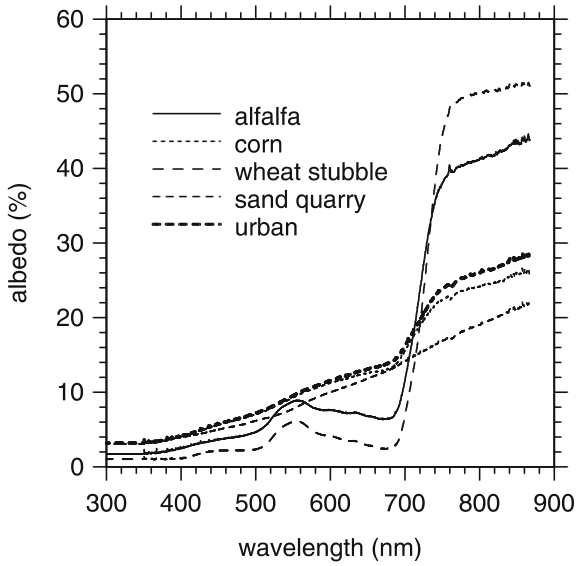



**Fig. 1.10.** Examples of spectral surface albedos for various surface types as functions of wavelength. These data were collected during spring in OK, USA. (Data courtesy of Z. Li 2000.)

and often do not distinguish between direct and diffuse albedos (but again, this partition actually requires solution of the radiative transfer algorithm).

Owing to the fact that roughly two-thirds of Earth's surface is covered by open water, albedo of water surfaces is an important quantity in a global model. There have been several studies that examined how the albedo of water surfaces with wind-generated waves depends on $\mu_0$ and wind speed $|\mathrm{v}|$ (e.g., Cox and Munk 1956; Payne 1972; Preisendorfer and Mobley 1986). Cox and Munk's method assumes that the slopes of wave facets follow a Gram–Chevalier (i.e., Gaussian-like) distribution that depends on $|\mathrm{v}|$ and that only Fresnel reflection need be considered. Despite it being a single-reflection model that does not account for spatial correlation of wave facets, it has been used successfully to infer $|\mathrm{v}|$ from observations of sun glint (its original intention).

Figure 1.11 shows a curve fit to Payne's observations (Briegleb et al. 1986) along with results from Cox and Munk's model (Hansen et al. 1983). Clearly the Cox and Munk parametrization captures the essence of ocean albedo. It can, however, be augmented slightly to account for the effects of whitecaps (Monahan and O'Muircheartaigh 1987), suspended particulates, and bottoms (in shallow areas). Multiple reflection effects, which depend directly on spatial structure of wave forms, are likely to be important for radiances, inside certain ranges of illumination and viewing, but are second-order as far as fluxes are concerned.

Obviously, if one believes there is an important feedback to capture between ocean absorption of solar radiation and near-surface winds, it is essential that a wind-speed dependent description of ocean albedo be employed in global models.

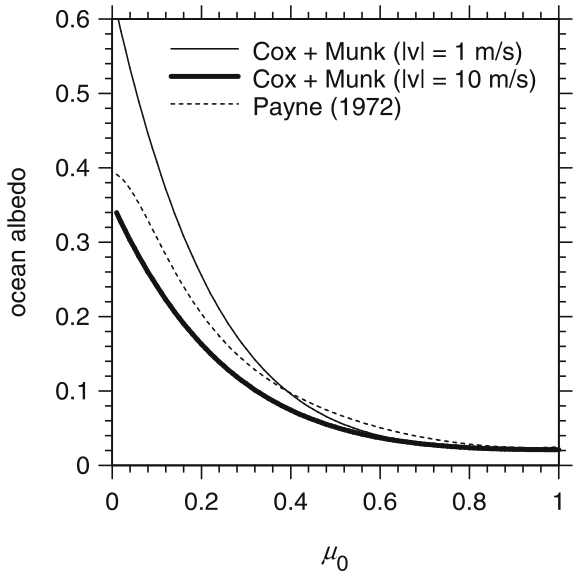



**Fig. 1.11.** Dashed line shows a curve fit to observed ocean albedo (Payne 1972). Other lines are predictions by the Cox and Munk wave slope approach for two wind speeds $|\mathbf{v}|$ assuming Fresnel reflection.

Likewise, at a similar level of concern, there should be an explicit dependence on the quality of illumination; that is, a distinction between direct- and diffuse-beam surface irradiance.

Computation of albedo for a totally snow-covered surface is a fairly straightforward problem because snow is optically dense. As a result, plane-parallel, homogeneous conditions can be satisfied fairly easily which makes for reliable application of two-stream approximations provided one has a reasonable idea of crystal size and amounts of impurities (see Warren 1982). Complications and biases arise, however, when dealing with snow amidst vegetation. For example, masking of snow by vegetation, and vice versa, exhibits a strong dependence on illumination angle (e.g., Otterman 1984). Another example is snow in mountain regions. Assume that almost all the snow exists in shaded areas. If the sky is overcast with only diffuse surface irradiance, mean albedo would be close to a linear weighting based on fractional area of snow and exposed rock. If the majority of surface irradiance is direct-beam, mean surface albedo would be close to that of the rock if the majority of snow is on the shaded side. This points to a more general issue: the proper mean albedo to be used in a 1D radiative transfer model is not the areal-weighted mean, but rather the irradiance- and areal-weighted mean. To do this properly, however, requires a solution for surface irradiance beforehand, which does not exist, as well as consideration of surface tilt geometry.

It is expected that albedo of sea-ice will require increasing attention as representation of sea-ice in GCMs continues to improve (Kreyscher et al. 2000; Bitz

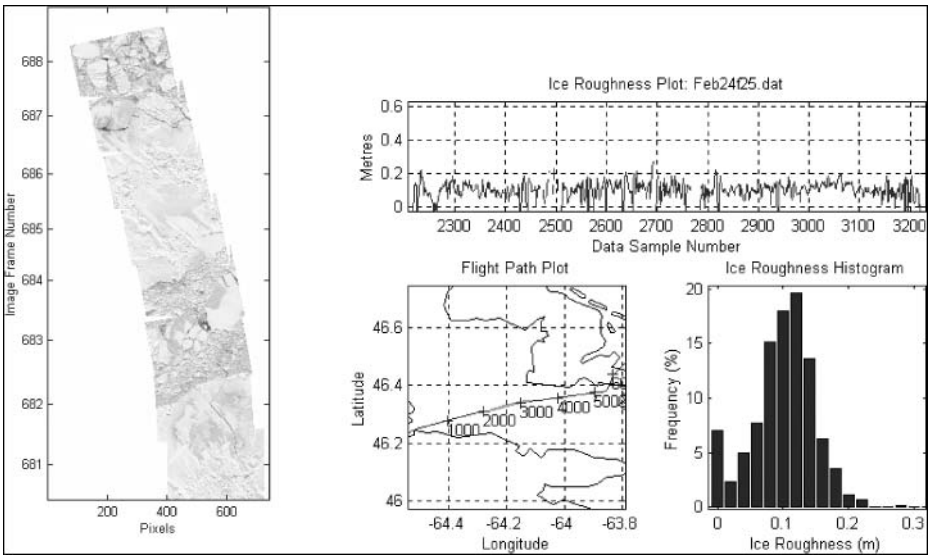



**Fig. 1.12.** Plot on the left shows imagery of sea-ice in Northhumberland Strait, Canada (image is about 0.5 km wide). Upper plot on the right shows normalized ice altitudes as detected by a helicopter-mounted laser altimeter tracking down the centre of the image on the left. Lower plots on the right show a map of the transect and frequency distribution of ice altitude. (Data courtesy of I. Peterson, 2001.)

et al. 2002). It may be that different types of ice can be diagnosed and a $\mu_0$ dependence of albedo assigned to each type. Figure 1.12 shows an example of the complexity of sea-ice. Presumably the roughness characteristics of these surfaces impacts albedo? The same goes for wind-generated sastrugi that exist over vast tracts of Antarctica (see Warren et al. 1998). However, we are undoubtedly a long way off performing actual on-line radiative transfer calculations for surfaces as complex as these.

Regarding representing radiative transfer for vegetated surfaces inside GCMs, much work has been done by Pinty et al. (2006). The essence of their work is to utilize the standard two-stream solutions with optical properties suitably modified to represent the discrete scattering elements found in vegetation canopies and to alter these quantities in such a way that the two-stream mimics corresponding results obtained by 3D Monte Carlo simulations. The attraction of this approach is that the surface system becomes an extra layer at the base of the regular atmospheric column. So instead of an atmospheric radiative transfer code using $N$ layers, it uses $N + 1$ layers with an underlying description of ground (not collective surface) albedo. Moreover, if there is a distribution of surface-vegetation types inside a GCM column, they can be included in a stochastic subgrid-scale generator (cf. Räisänen et al. 2004) and used in conjunction with the McICA approach.

## 1.4 1D vs. 3D radiative transfer for cloudy atmospheres: should global modellers be concerned?

Thus far only results from 1D radiative transfer codes have been shown and discussed. The assumption all along has been that the ICA is a justifiable standard for GCM-style radiation solvers to aim for. This goal has been obtained by the Monte Carlo class of ICA codes. The lingering question, however, is: are systematic flux and heating rate differences between the ICA and full 3D important for numerical simulation of climate? We are just entering an era with supercomputing that enables us to address this question, and progress has already been made. Two points of concern are reiterated here: i) are domain-average fluxes provided by the ICA sufficient; and ii) how important are interactions between cloud and radiation at scales that are unresolved by conventional GCMs? The former addresses 3D radiative transfer squarely, while the latter, which is essentially a 3D issue, can be addressed to some extent by the ICA.

### 1.4.1 Domain-average fluxes

When contrasting 1D and 3D radiative transfer it is helpful to study details at scales finer than large domain-averages to help appreciate differences. As an example, consider the 2D cloud shown in Fig. 1.13. This stratiform cloud is moderately inhomogeneous for its range of liquid water path (LWP) is fairly large yet its physical aspect ratio is not. Vertically-projected cloud fraction over this 10 km domain is 0.9, mean visible optical depth is 18.9, and corresponding $\nu_{mom}$ is 2.6. Even water vapour content, as expected, varies in the horizontal. Figure 1.14 shows heating rates for 3D transfer and the ICA at several different solar zenith angles. In a plot like this differences between 3D and ICA transfer are obvious. Most notably, ICA casts cloud shadows vertically regardless of sun angle. A more subtle difference is in cloud heating; clouds absorb more for 3D transfer at small $\mu_0$ and less at large $\mu_0$.

Figure 1.15 shows corresponding domain-average values. Domain-average atmospheric absorptances for 3D and ICA are virtually identical despite the clear spatial differences seen in Fig. 1.14. Albedo is overestimated very slightly by the ICA for most $\mu_0$ (but much less so than outright neglect of variability), and underestimated only at very small $\mu_0$ on account of side illumination for 3D transfer (testified to in Fig. 1.14 by elimination of direct-beam transmittance at small $\mu_0$). The centre plot of Fig. 1.15 shows that mean photon pathlengths differ little between ICA and 3D, except again at small $\mu_0$ where ICA pathlengths are slightly longer. The rightmost plot shows that the mean numbers of times reflected photons are scattering by droplets are essentially identical for ICA and 3D. For transmittance, however, photons in the 3D simulation experience fewer scattering events for all $\mu_0$, especially small $\mu_0$ where photons frequently exit cloud sides in downward directions after few scatterings. Despite these differences, it is still interesting to note again that domain-average atmospheric absorptances are almost identical (see Barker et al. 1998).

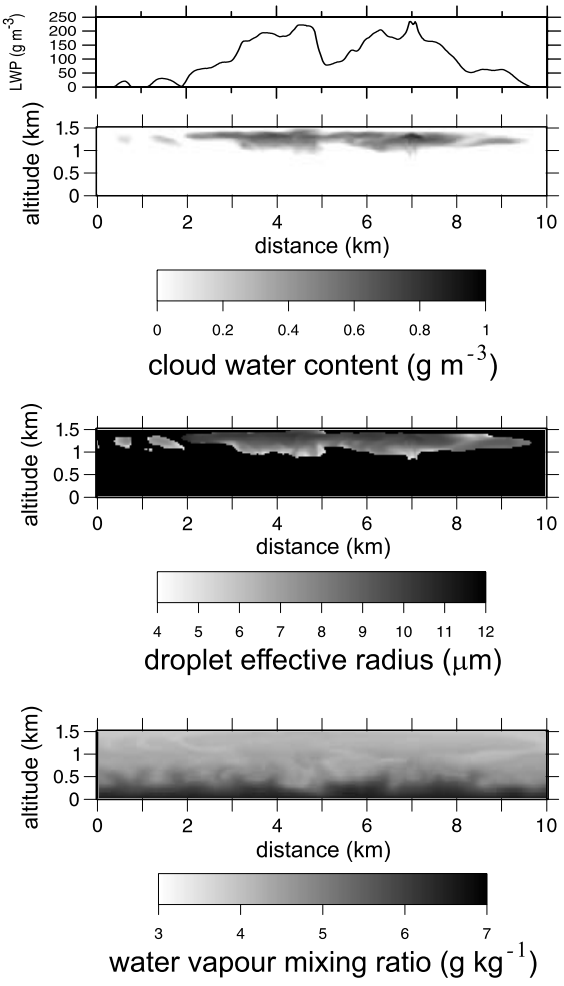



**Fig. 1.13.** Top plot shows visible optical depth of cloud for a 10 km section of cloud simulated by a cloud system-resolving model with 25 m horizontal and vertical grid-spacings. Beneath it is liquid water content, droplet effective radius, and water vapour mixing ratio. (Data courtesy of J.-P. Blanchet, 2001).

This example serves to demonstrate what several studies have found: there are obvious local differences between 3D and ICA, but as soon as the domain size exceeds a few characteristic cloud cell dimensions, differences between 3D and ICA domain-average radiative quantities diminish rapidly.

In the past, small numbers of cloud configurations were used to address differences between 3D radiative transfer and approximate solutions (e.g., McKee and Cox 1974; Welch and Wielicki 1985; O'Hirok and Gautier 1998; Barker et al. 1999). Barker et al. (2003) intercompared domain-average, broadband irradiances for cloudy atmospheres as computed by several 1D and 3D solar transfer

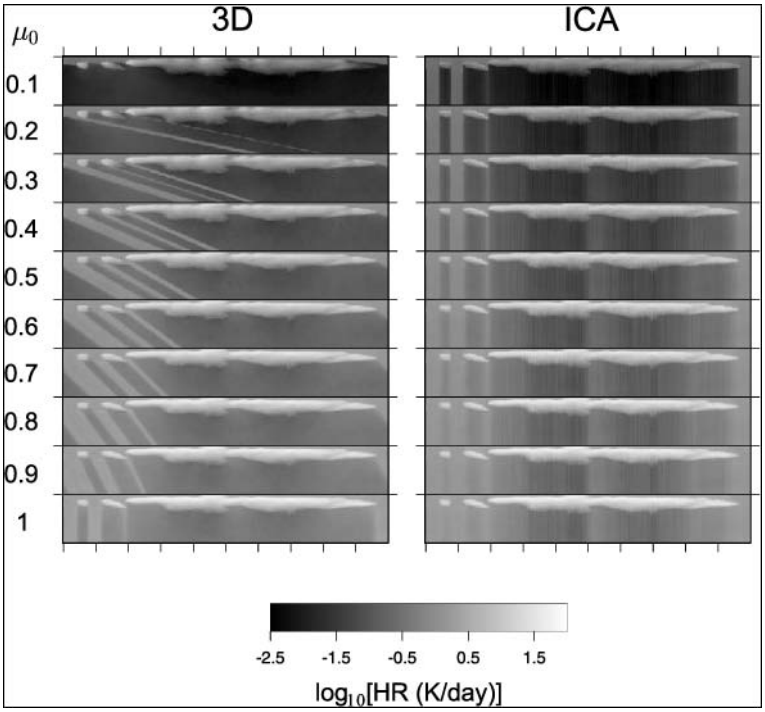



**Fig. 1.14.** Heating rates for the cloud field shown in Fig. 1.13 for various solar zenith angles $\theta_0$ ($\mu_0 = \cos\theta_0$ are listed on the left). Left column is for 3D radiative transfer while right column is for the ICA model. All calculations were performed with a Monte Carlo photon transport algorithm; the 3D simulations used 25 m horizontal grid-spacings while the ICA used an extremely large setting.

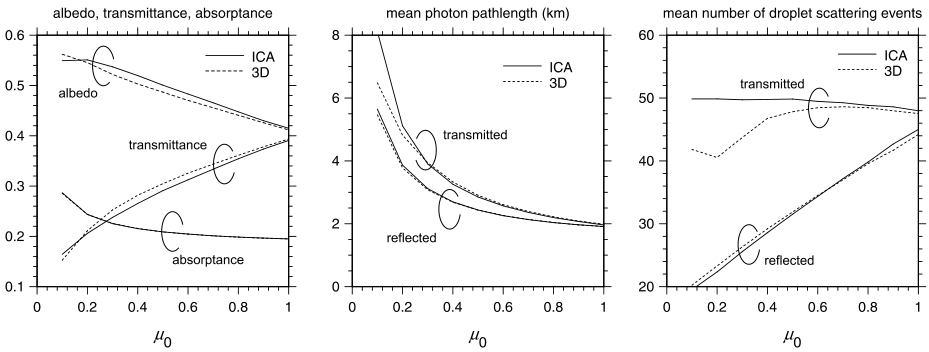



**Fig. 1.15.** Left plot shows domain-average albedo, transmittance, and atmospheric absorptance as functions of cosine of solar zenith angle $\mu_0$ for 3D and ICA simulations performed on the field shown in Fig. 1.13 Middle and right plots show corresponding mean photon pathlengths (below cloudtop) and mean number of droplet scattering events for photons that are reflected to space and transmitted to the surface.

codes, but again used only a half-dozen or so cases. Cole et al. (2005a), however, used thousands of cloud fields that were produced by a GCM whose conventional cloud parametrization was replaced by a 2D cloud system-resolving model (CSRM). Such super-parametrized GCMs, known officially as Multi-scale Modeling Framework (MMF) GCMs, are developing rapidly (see Randall et al. 2003). For Cole et al.'s (2005a) study, CSRM domains had 64 columns with 4 km horizontal grid-spacing $\Delta x$, 24 vertical layers and a timestep of 20 s, and were aligned east to west (Khairoutdinov and Randall 2003). Each CSRM was forced by large-scale tendencies updated every GCM timestep, and provided horizontally-averaged tendencies back to the GCM. The CSRM prognostic thermodynamic variables included liquid/ice water moist static energy, total non-precipitating water, and total precipitating water. All simulations started on September 1, 2000. Global arrays of CSRM data were sampled and saved every 9 model hours. Allowing the model a short spin-up period, radiation calculations were performed on model output for December 2000. This amounted to solar calculations being performed on over 300,000 fields.

Figure 1.16 shows monthly-mean differences between 2D radiative transfer and the ICA for upwelling SW at the TOA. The largest TOA differences are associated with tropical deep convective clouds with secondary maxima across the southern ocean storm belt partly because of excessive cloudiness and large solar inputs. The adjacent plot shows the distribution of flux differences as a function of latitude and $\mu_0$. Suppression of photon leakage out the sides of convective clouds in the ITCZ by the ICA at large $\mu_0$ explains why it overestimates reflected flux. Conversely, the ICA does not account for illumination onto cloud

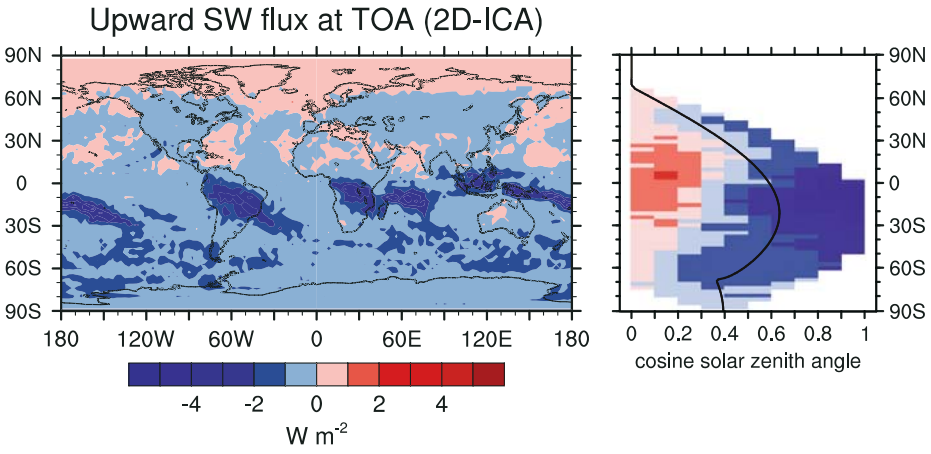



**Fig. 1.16.** Differences for monthly-mean upwelling SW flux at TOA when radiation calculations are done with the ICA and 2D radiative transfer. Global mean value is 104.7 W m$^{-2}$ for the 2D case. Plot on the right shows corresponding mean values as a function of $\mu_0$ and latitude. Solid line indicates monthly-mean $\mu_0$.

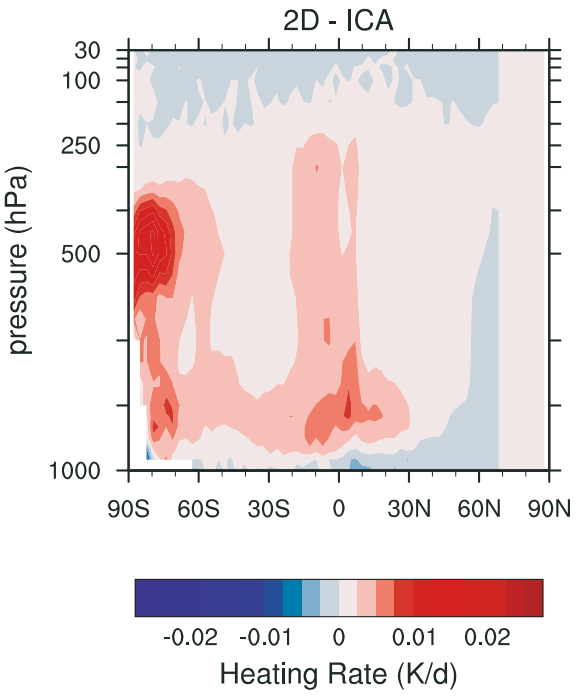



**Fig. 1.17.** Monthly-mean, zonal cross-section of differences in SW heating rate between 2D radiative transfer and the ICA as a function of latitude and altitude.

sides, so at small $\mu_0$ it underestimates reflected flux (e.g., Welch and Wielicki 1985; O'Hirok and Gautier 1998).

Figure 1.17 shows a vertical cross-section of monthly-mean SW heating rate differences between 2D and ICA. This shows clearly the impact of cloud side illumination at cloud-bearing altitudes. Namely, when significant side illumination occurs, as with 2D transfer, clouds absorb more radiation relative to the ICA approach.

Using data from O'Hirok and Gautier (2005), Cole et al. (2005a) also addressed the question of drawing too many conclusions from their results given $\Delta x = 4$ km. They concluded that if $\Delta x$ was reduced to about 0.5 km, values shown in Fig. 1.16 could roughly double in certain areas due to cloud fluctuations becoming increasingly resolved. At that point, differences between 3D radiative transfer and the ICA would begin, at times, to rival those that occur between ICA and conventional GCM treatments (see Cole et al. 2005a; Stephens et al. 2004).

The point to remember with conventional GCMs is that only gross descriptions of cloud structure are available. Often this amounts to just mass of condensed water in a layer and a corresponding estimate of cloud fraction. We are just beginning to parametrize other details like droplet concentration and size, variance of water content, and vertical overlap rates. While one-point distribu-

tions of clouds can be generated stochastically and used with the McICA method (e.g., Räisänen et al. 2004), it is still an open question as to whether it is worth going the extra step and assuming (i.e., imposing) horizontal structure which is required to perform 3D radiative transfer. It may be that the differences discussed in this section are significant and that these extra assumptions are well worth it, but it will require a suitable and efficient algorithm for generating subgrid-scale clouds and an equally efficient 3D radiative transfer model. Nevertheless, all that would be required would be domain-averaged and spectrally-integrated flux profiles. We know already that random noise is a minor issue for weather forecasting and climate modelling (cf. section 1.3.5.5 on McICA), so reasonable numbers of photons in a 3D Monte Carlo simulation (especially for SW transfer; see Evans and Marshak 2005) would likely suffice.

### 1.4.2 Unresolved cloud–radiation interactions

Even if a GCM computes domain-average fluxes based on 3D radiative transfer, the remaining question is: do conventional GCMs, with horizontal grid-spacings on the order of 100 km or more, resolve cloud-radiation interactions sufficiently well? This crucial question was one of the prime factors behind the push toward MMF-GCMs (Grabowski 2001; Randall et al. 2003). By definition this question cannot be answered with a conventional GCM. One must be content to use either stand-alone CSRMs or an MMF-GCM. Obviously the latter is more preferable, but more expensive, as large-scale circulation comes into play explicitly.

Presumably, this question can be addressed to a great extent using the ICA (e.g., Fu et al. 1995); that is, allow the CSRMs to experience local radiative heating rates regardless of whether they are computed using 3D transfer or not. This is what Cole et al. (2005b) did using an MMF GCM with a CSRM whose horizontal grid-spacing was 4 km. In a series of experiments, they demonstrated that allowing the CSRMs to respond to cloud–radiation interactions at the 4 km scale was roughly as important as getting domain-average fluxes correct. Incorrect domain-average fluxes were provided by a standard two-stream model with maximum-random overlap of homogeneous clouds. Hence, domain-average errors were close to what could be expected if a common GCM radiative transfer algorithm were to be replaced by the McICA algorithm.

Figure 1.18 shows the impact on cloud radiative effects at the TOA due to inclusion of unresolved interactions between clouds and radiation (pers. comm., J. Cole 2005). Experiment 1 served as the benchmark for it utilized heating rates at 4 km and also provided the GCM with proper (ICA) domain-average fluxes. Experiment 4 utilized heating rates at 4 km too but incorrect domain-averages were passed back to the GCM. Evidently this was not important for this seasonal simulation. For Experiments 2 and 3, on the other hand, heating rates were averaged horizontally across the domains and used by the CSRMs. Experiment 2 passed correct domain-average fluxes back to the GCM while Experiment 3 did not. Clearly, omission of heating rates at the CSRMs' inner-scale are most important for this experiment. The significance of this result rests in the realization that getting proper domain-averages in a GCM is one

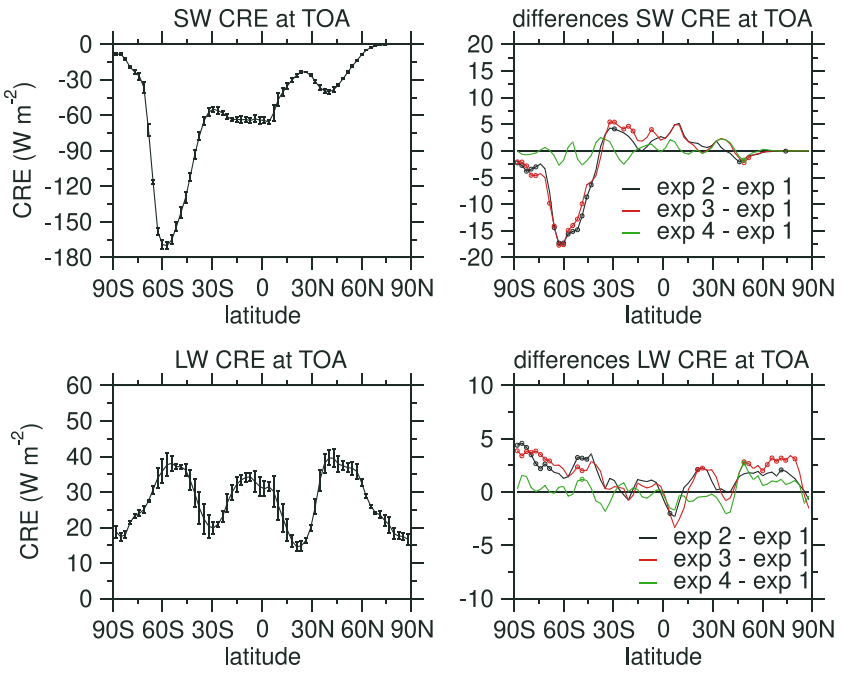



**Fig. 1.18.** In the left column are plots of zonal-mean, time-averaged SW and LW cloud radiative effects (CRE), also known as cloud radiative forcing, for December as computed by an MMF GCM. Error bars indicate one standard deviation as realized by a five-member ensemble of the benchmark simulation (exp 1). Plots on the right show differences between various simulations and exp 1. See text for details. Dots indicate when a difference is significant at the 95% confidence level. (Data courtesy of J. Cole, 2005.)

thing, but how can one expect to account for unresolved interactions between clouds and radiation in a conventional GCM cloud parametrization where clouds are not resolved anywhere near fundamental cloud formation-dissipation scales? The answer is far from obvious, and we might run out of time trying solve it, for eventually either MMF-GCMs will become commonplace or GCMs will simply become, in essence, global CSRMs. In either case, the issue of subgrid-scale parametrization will diminish over time. Then again, conventional GCMs will likely continue to serve purposes, and if they continue to be used in important roles, as they are today, progress will be required on cloud-radiation interactions of the kind discussed here.

## 1.5 Remote sensing of cloudy atmospheres and global climate modelling

For some time now the most valuable data for assessing GCMs have come from the Earth radiation budget satellite missions, namely Earth Radiation Budget Experiment (ERBE) and Clouds and the Earth's Radiant Energy System (CERES). For the most part, these datasets are used to determine monthly-mean radiation budgets and cloud radiative effects for diagnostic comparison against corresponding values predicted by GCMs. Yet even here there are still issues and uncertainties such as radiance-to-flux conversion and sparse sampling (e.g., Loeb et al. 2003). Recently these Earth radiation budget datasets have been merged with other datasets and radiative quantities have been sorted according to dynamical regime in an attempt to identify the conditions responsible for the wide disagreement observed among GCMs with respect to estimates of cloud feedback (cf. Bony et al. 2004). Additionally, attempts have been made to bring models and observations closer together via dataset emulators such as the so-called ISCCP simulator (Webb et al. 2001) where GCM fields, like cloud fraction and optical depth, are converted, without actually performing radiative transfer calculations, to resemble products inferred from satellite data. The alternative to this approach is to apply more sophisticated radiative transfer algorithms to GCM data, thereby producing radiance estimates that can be compared directly to satellite measurements.

For the past 15 years or so, surface-based cloud-profiling radars (CPRs) have provided much information about the vertical and horizontal structural characteristics of clouds (e.g., Clothiaux et al. 1999). Recently, NASA's CPR-bearing CloudSat satellite was launched into a Sun-synchronous orbit (see Stephens et al. 2002). CloudSat is flying in close formation with several other satellites in the so-called *A-train* constellation. When data from these satellites are merged, they have the strong potential to yield invaluable insights into the structure of clouds at the global scale. There are, however, many issues that have to be grappled with first.

The CPR emits pulses of electromagnetic radiation that are scattered by particles in proportion to the sixth power of particle radius. Hence, CPRs are quite sensitive to large cloud droplets, but have difficulty sensing small droplets. It is immediately clear then that CPR data have to be augmented with data from measurements that are sensitive to other moments of particle size distributions. Hence the need to collocate CPR data with other data.

In addition, satellite-based CPRs suffer from the same sampling issues as do aircraft and instruments fixed at the surface. Astin et al. (2001) have shown that in broken cloud conditions, transects may have to be several hundred kilometres long before significant reductions can be realized in the confidence intervals for estimates of cloud fraction of the domain from which the transect was drawn. Figure 1.19 illustrates the situation one will encounter with both CPRs fixed on the surface (as this one is), or on aircraft or satellite. It shows the 512-km domain of a CSRM which is taken here to be a grid-cell of a GCM. What the GCM (or the CSRM) requires is a time series of the domain. What the CPR

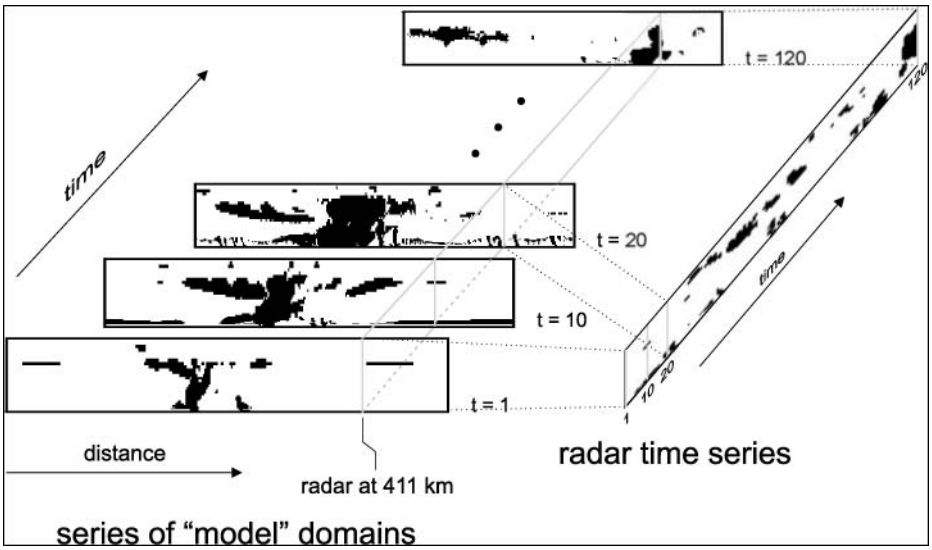



**Fig. 1.19.** On the right is a sampled time series of cloud masks for a 512-km domain produced by a CSRM (Fu et al. 1995). Imagine that a surface-based CPR is fixed at 411 km along the domain. Stringing together the 120 snapshots at 411 km leads to the CPR-domain shown on the right. A GCM or NWP model would want what is on the left, but the radar produced what is on the right. Clearly, in this particular case, what the radar samples is a rather poor representation of the sequence of domains. Satellite-based CPRs will be subject to this sampling issue as well.

(fixed here at 411 km across the domain) gives is a time series of cloud that drifts over it. As Fig. 1.19 shows, the time series of cloud obtained at 411 km appears to contain only a slim semblance of the domains.

Despite these complications with CPR data, NASA's CloudSat satellite, with its 94 GHz CPR in conjunction with the lidar on CALIPSO, and the passive sensors on AQUA, should provide an interesting global view of the coincidental vertical and horizontal structure of clouds. Figure 1.20 shows an example of CloudSat data along with near-simultaneous 1-km GOES visible and thermal imagery and NEXRAD surface precipitation radar data. Clearly CloudSat has pinpointed precipitating clouds and appears to have even sensed the shallow clouds immediately east of the Florida–Georgia border.

Data from CloudSat and the A-train will enable global-scale calculations of information such as, for example, radiative sensitivities for climate model parameters like cloud overlap decorrelation lengths which are required by the more sophisticated GCM parametrizations. This is because cloud fractions in distinct layers will be reported. To illustrate, if $c_k$ and $c_l$ are fractional amounts of cloud in layers $k$ and $l$ as reported by CloudSat, and if $A_c$ is total cloud fraction, define

$$A_c = \alpha_{k,l} c_{k,l}^{\max} + (1 - \alpha_{k,l}) c_{k,l}^{\mathrm{ran}} , \tag{1.70}$$

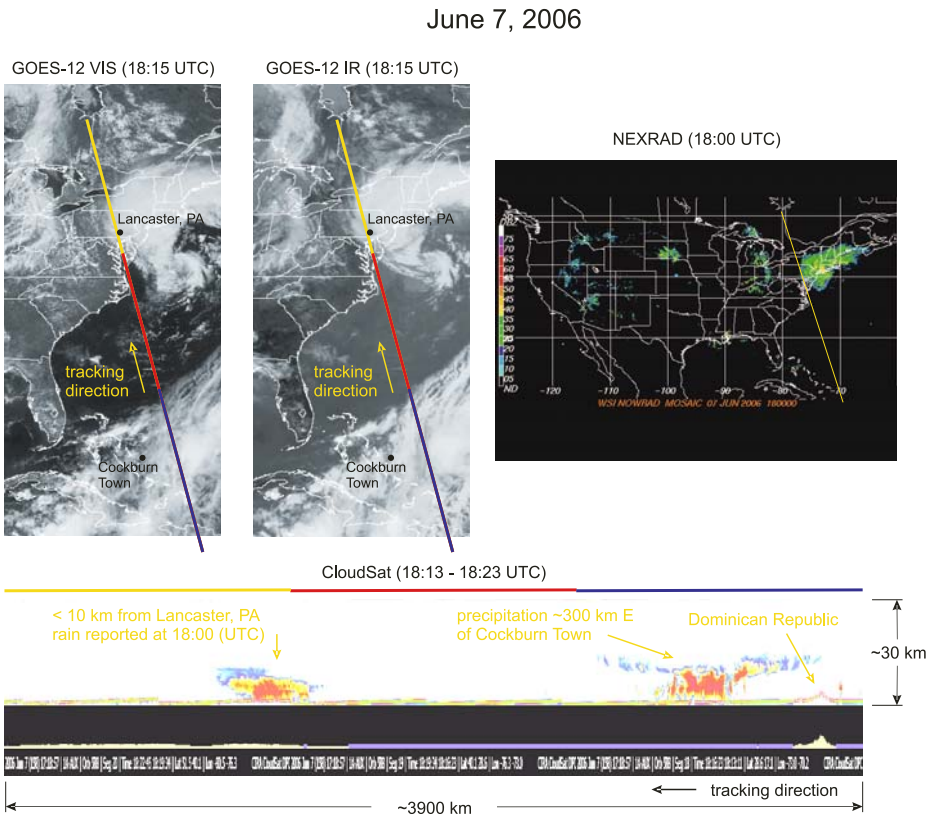



**Fig. 1.20.** Lower plot shows an almost 4000 km long transect of CPR reflectivities measured by CloudSat. These uncalibrated data were produced only a week after CloudSat's radar was switched on. Nevertheless, it shows an unprecedented view of clouds that up until then would have been restricted to the passive GOES images shown in the upper left and the NEXRAD surface precipitation radar composite shown in the upper right. CloudSat's colour-coded cross-section is shown on the GOES imagery and the total trajectory is shown on the NEXRAD image.

where

$$\begin{aligned} c_{k,l}^{\max} &= \max\left(c_k, c_l\right) , \\ c_{k,l}^{\mathrm{ran}} &= c_k + c_l - c_k c_l . \end{aligned} \tag{1.71}$$

The overlap parameter $\alpha_{k,l}$ in (1.70) is defined as

$$\alpha_{k,l} = \exp\left[-\int_{z_k}^{z_l} \frac{\mathrm{d}z}{\mathcal{L}_{cf}(z)}\right] , \tag{1.72}$$

where $\mathcal{L}_{cf}$ is decorrelation length for overlapping fractional cloud, and $z$ is altitude (Hogan and Illingworth 2000; Bergman and Rasch 2002).

Changes to $\mathcal{L}_{cf}$ impact SW fluxes primarily through changes to $A_c$ and $p(\tau)$. As shown in section 1.3.5.3, the first two moments of $p(\tau)$ are often sufficient to capture domain-average SW fluxes. Therefore, following Barker and Räisänen (2005), the radiative sensitivity for $\mathcal{L}_{cf}$ can be expressed as

$$\frac{\partial F}{\partial \mathcal{L}_{cf}} \approx \frac{\partial F}{\partial A_c}\frac{\partial A_c}{\partial \mathcal{L}_{cf}} + \frac{\partial F}{\partial \overline{\tau}}\frac{\partial \overline{\tau}}{\partial \mathcal{L}_{cf}} + \frac{\partial F}{\partial \sigma_\tau}\frac{\partial \sigma_\tau}{\partial \mathcal{L}_{cf}}\,, \tag{1.73}$$

where $\overline{\tau}$ is mean cloud optical depth for the cloudy part of the domain, and $\sigma_\tau$ is corresponding relative standard deviation of $\tau$. Each derivative in (1.73) depends on cloud structure inside the domain as well as other parameters such as water vapour and temperature profiles, surface conditions, and $\theta_0$. Note also that the leading term in the three terms on the RHS of (1.73) are themselves sensitivities that can be computed separately holding all else constant, including $\mathcal{L}_{cf}$ (cf. Schneider 1972).

With the stochastic subcolumn generator developed by Räisänen et al. (2004) it is straightforward to compute how cloud properties and $F$ depend on $\mathcal{L}_{cf}$ from which corresponding derivatives can be estimated numerically. Figure 1.21 shows global estimates of $\partial F/\partial \mathcal{L}_{cf}$ using data for a single day (January 1) generated by an MMF GCM (Khairoutdinov and Randall 2001). The global-mean value for net radiation (SW + LW) is $1.76\,\mathrm{W\,m^{-2}\,km^{-1}}$ with $1.99\,\mathrm{W\,m^{-2}\,km^{-1}}$ in the SW and just $-0.23\,\mathrm{W\,m^{-2}\,km^{-1}}$ in the LW. Perhaps as expected, the largest sensitivities are in the tropics where minor deviations to the overlap rate of

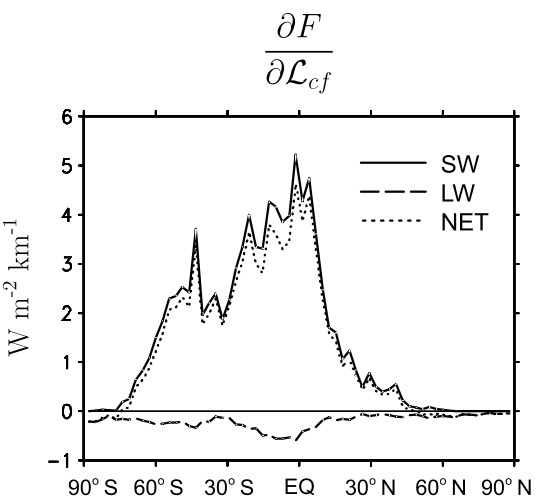



**Fig. 1.21.** Radiative sensitivities $\partial F/\partial \mathcal{L}_{cf}$ for cloud fraction vertical decorrelation length $\mathcal{L}_{cf}$ computed with all else held constant. SW, LW, and NET (SW + LW) components are shown as functions of latitude. These estimates were computed using data from an MMF GCM simulation. It is anticipated that plots like this will be available from data collected by CloudSat and the other A-train satellites.

towering clouds coupled with high solar irradiance conspire to affect large TOA albedo changes. The hope is that a corresponding analysis can be conducted using inferences from A-train data.

## 1.6 Concluding remarks

This central thesis of this volume is scattering of radiation (light). Given limitations on length, this chapter focused on treatment of solar radiative transfer in global climate models (GCMs). The intention was to give readers, especially upper-level undergraduates and entry-level graduates, a *taste* of some of the light scattering-related issues facing computation of radiative fluxes in global models. Emphasis was on some of the macroscopic issues regarding solar radiative transfer for cloudy atmospheres.

After discussing the central role played by radiation in global climatology and analyses of climate, both real and simulated, the mainstay of GCM radiative transfer solvers was discussed; two-stream approximations. Limitations of two-streams were mentioned and some techniques were reviewed for extending the application of basic two-streams to address fluctuations in cloud density that are unresolved by conventional GCMs. The obvious limitation is the fact that clouds are never homogeneous in the horizontal or the vertical, and so the basic assumptions upon which two-streams are built are violated repeatedly in GCMs. Regarding the macroscopic aspects of clouds, this translates into descriptions of horizontal fluctuations of cloud properties across a GCM layer as well as how these fluctuations stack up in the vertical. This discussion finished with the Monte Carlo Independent Column Approximation (McICA) which decouples description of unresolved fluctuations in the medium from the radiative transfer solver and so eliminates biases that analytic two-stream extensions are bound to possess. Hence, McICA is able, to the same extent as the full ICA, to address all descriptions of horizontal fluctuations and vertical overlap patterns. The problem is that conventional GCM cloud parametrizations are still in their infancy when it comes to describing the nature of unresolved cloud structure, and so at the moment, McICA within a GCM can be provided only with limited information. The hope is that a synthesis of data from a host of sensors will help guide production and assessment of these much needed aspects of cloud parametrization. Data from the active sensors in the A-train of satellites (see Stephens et al. 2002) as well as from the Atmospheric Radiation Measurement (ARM) program (Stokes and Schwartz 1994) and the BALTEX Bridge Cloud (BBC) campaign, should help immensely, but it is still too early to tell as extensive numerical end-to-end simulations of what we know these instruments are sensing have yet to take place. Having said this, the unbiased nature of McICA is, however, only as good as one's willingness to neglect 3D radiative transfer effects on fluxes.

On this last point there are two levels of concern. First, are domain-average fluxes provided by the ICA sufficient for climate simulation? Global estimates

of differences between the ICA and 3D radiative transfer were provided in section 1.4.1 (see Cole et al. 2005a). In the framework of conventional GCMs, with only limited amounts of information regarding unresolved cloud structure, the answer is probably: yes, the ICA is sufficient. This is because to go beyond the ICA requires explicit cloud fields as opposed to descriptions of one-point probability distributions for horizontal variations and statistical descriptions for vertical overlap. GCMs are barely in a position to provide these descriptions let alone make the leap to conjuring up full 3D, even 2D, unresolved cloud fields. Again, data like those from the satellite A-train should help address this concern. Nevertheless, it might be worthwhile performing some GCM sensitivity experiments that employ a simple stochastic cloud field generator [like that developed by Räisänen et al. (2004) but including horizontal structure] to see if the fairly subtle, yet systematic, effects of 3D transfer impact global climate simulations.

The second question regarding 3D effects is: are interactions between cloud and radiation at scales that are unresolved by conventional GCMs important for climate simulation? As mentioned in section 1.4.2, this effect can be addressed partially using the ICA method. This is how Cole et al. (2005b) investigated this issue using an MMF GCM in which the cloud parametrization consisted of a 2D cloud system-resolving model (CSRM). In essence, the CSRM provides explicit cloud fields (i.e., no need to invoke parametrizations to describe unresolved fluctuations) so the ICA can be applied directly to the CSRM fields and the CSRMs can either evolve based on local radiative heating rates or domain-averaged heating rates. Likewise, either ICA domain-average fluxes or incorrect values based on a conventional GCM radiation solver (e.g., one that makes simple assumptions about clouds at scales unresolved by the host GCM) can be passed back to the host GCM. There is nothing, except computational limitations perhaps, stopping one from repeating Cole et al.'s experiments using 3D radiative transfer models. The disconcerting conclusion they came to was that, for their experiments in particular, cloud-radiation interactions at scales unresolved by the host GCM appear to be approximately as important as getting the domain-average flux profiles correct (i.e., via the ICA or McICA). This is disconcerting because it is very difficult to see how these interactions can be parametrized in a conventional GCM to an extent that would be considered to be even remotely satisfying.

There are many aspects to GCM radiative transfer calculations that were not addressed explicitly in this chapter. Some of them have been discussed elsewhere in this volume. For example, scattering by ice crystals is dictated by the size, geometric structure, and orientation of crystals, but as yet there is little consensus on any of these properties. As such, representation of scattering by ice crystals in GCM radiation codes is still at the stage of fiddling with gross properties, like asymmetry parameter, in order to satisfy model simulation of global radiation budgets. On the other hand, treatment of absorption of radiation by atmospheric constituents has been addressed only in passing throughout this volume. Likewise, relatively little has been said about measurement of Earth's radiation budget by satellites as a means of assessing GCM simulations, though an entire chapter could easily be devoted to this subject.

## A Appendix: Two-stream approximations

The purpose of this appendix is to clarify two-stream approximations. Begin by restating the azimuthally-averaged 1D equation of transfer [see (1.19)] as

$$\begin{aligned}\mu\frac{\mathrm{d}I(\tau,\mu)}{\mathrm{d}\tau} &= -I(\tau,\mu)+\frac{\omega_0}{2}\int_{-1}^{1}p(\mu;\mu')I(\tau,\mu')\,\mathrm{d}\mu' \\ &\quad +\frac{F_0}{4}\omega_0 p(\mu;\mu_0)\,\mathrm{e}^{-\tau/\mu_0}\,,\end{aligned} \tag{A1.1}$$

where $I$ is radiance, $F_0$ is incoming solar at the TOA, $\mu$ is cosine of zenith angle, $\mu_0$ is cosine of solar zenith angle, $\tau$ is optical thickness, and $\omega_0$ is single scattering albedo. Define

$$F^{\pm}(\tau)=\int_0^1 \mu I(\tau,\pm\mu)\,\mathrm{d}\mu \tag{A1.2}$$

as upwelling and downwelling irradiances. Going a step further, let $I(\tau,\pm\mu)$ be defined as

$$I(\tau,\pm\mu)=\sum_{m=0}^{\infty} i_m(\tau)\,P_m(\mu) \tag{A1.3}$$

where $P_m(\mu)$ is the $m$th-order Legendre polynomial. Hence, (A1.2) becomes

$$F^{\pm}(\tau)=\int_0^1 \mu\sum_{m=0}^{\infty} i_m(\tau)\,P_m(\mu)\,\mathrm{d}\mu. \tag{A1.4}$$

By applying the hemispheric operators $\int_0^1 \mathrm{d}\mu$ and $\int_{-1}^0 \mathrm{d}\mu$ to (A1.1), using (A1.4), and dropping explicit notation of dependence on $\tau$ yields the coupled equations:

$$\begin{cases}\dfrac{\mathrm{d}F^{+}}{\mathrm{d}\tau} &= \displaystyle\int_0^1\sum_{m=0}^{\infty} i_m P_m(\mu)\,\mathrm{d}\mu-\frac{\omega_0}{2}\int_0^1\int_{-1}^1 p(\mu;\mu')\sum_{m=0}^{\infty} i_m P_m(\mu)\,\mathrm{d}\mu'\,\mathrm{d}\mu \\ &\quad -\dfrac{F_0}{4}\omega_0\,\mathrm{e}^{-\tau/\mu_0}\displaystyle\int_0^1 p(\mu;\mu_0)\,\mathrm{d}\mu \\ \dfrac{\mathrm{d}F^{-}}{\mathrm{d}\tau} &= \displaystyle\int_{-1}^0\sum_{m=0}^{\infty} i_m P_m(\mu)\,\mathrm{d}\mu-\frac{\omega_0}{2}\int_{-1}^0\int_{-1}^1 p(\mu;\mu')\sum_{m=0}^{\infty} i_m P_m(\mu)\,\mathrm{d}\mu'\,\mathrm{d}\mu \\ &\quad +\dfrac{F_0}{4}\omega_0\gamma_4\,\mathrm{e}^{-\tau/\mu_0}\displaystyle\int_{-1}^0 p(\mu;\mu_0)\,\mathrm{d}\mu.\end{cases} \tag{A1.5}$$

Decomposing the zenith-to-nadir integrals into integrals over the up and down hemispheres leads to

$$
\begin{cases}
\dfrac{\mathrm{d}F^+}{\mathrm{d}\tau} = \displaystyle\int_0^1 \sum_{m=0}^{\infty} i_m P_m(\mu)\,\mathrm{d}\mu + \underbrace{\frac{\omega_0}{2}\int_0^1\int_0^1 p(\mu;\mu')\sum_{m=0}^{\infty} i_m P_m(\mu)\,\mathrm{d}\mu'\,\mathrm{d}\mu}_{\text{diffuse-beam forescatter}} \\
\qquad\qquad - \underbrace{\frac{\omega_0}{2}\int_0^1\int_{-1}^0 p(\mu;\mu')\sum_{m=0}^{\infty} i_m P_m(\mu)\,\mathrm{d}\mu'\,\mathrm{d}\mu}_{\text{diffuse-beam backscatter}} - \underbrace{\frac{F_0}{4}\omega_0\,\mathrm{e}^{-\tau/\mu_0}\int_0^1 p(\mu;\mu_0)\,\mathrm{d}\mu}_{\text{direct-beam backscatter}}. \\
\dfrac{\mathrm{d}F^-}{\mathrm{d}\tau} = \displaystyle\int_{-1}^0 \sum_{m=0}^{\infty} i_m P_m(\mu)\,\mathrm{d}\mu + \underbrace{\frac{\omega_0}{2}\int_{-1}^0\int_0^1 p(\mu;\mu')\sum_{m=0}^{\infty} i_m P_m(\mu)\,\mathrm{d}\mu'\,\mathrm{d}\mu}_{\text{diffuse-beam backscatter}} \\
\qquad\qquad + \underbrace{\frac{\omega_0}{2}\int_{-1}^0\int_{-1}^0 p(\mu;\mu')\sum_{m=0}^{\infty} i_m P_m(\mu)\,\mathrm{d}\mu'\,\mathrm{d}\mu}_{\text{diffuse-beam forescatter}} + \underbrace{\frac{F_0}{4}\omega_0\,\mathrm{e}^{-\tau/\mu_0}\int_{-1}^0 p(\mu;\mu_0)\,\mathrm{d}\mu}_{\text{direct-beam forescatter}}.
\end{cases}
\tag{A1.6}
$$

Equation (A1.6) can be simplified to

$$
\begin{cases}
\dfrac{\mathrm{d}F^+(\tau)}{\mathrm{d}\tau} &= \gamma_1 F^+(\tau) - \gamma_2 F^-(\tau) - \dfrac{F_0}{4}\omega_0\gamma_3\,\mathrm{e}^{-\tau/\mu_0} \\
\dfrac{\mathrm{d}F^-(\tau)}{\mathrm{d}\tau} &= \gamma_2 F^+(\tau) - \gamma_1 F^-(\tau) + \dfrac{F_0}{4}\omega_0\gamma_4\,\mathrm{e}^{-\tau/\mu_0}
\end{cases}
\tag{A1.7}
$$

by approximating the integrals. As such, (A1.7) is in essence the general two-stream approximation where $\gamma_1, \ldots, \gamma_4$ depend on assumptions made about $I$ and $p$, as well as on $\mu_0$ and optical properties. General solutions to (A1.7) are given by (1.23) through (1.28) in the mainbody of this chapter.

Some two-stream approximations (e.g., Coakley and Chýlek 1975) work directly with the backscattered fractions in (A1.6). For diffuse-beam, the backscattered fraction is

$$
\overline{\beta} = \frac{1}{2}\int_0^1\int_0^1 p(\mu;\mu')\sum_{m=0}^{\infty} i_m P_m(\mu)\,\mathrm{d}\mu'\,\mathrm{d}\mu\,. \tag{A1.8}
$$

Expanding $p(\mu;\mu')$ using the addition theorem for spherical harmonics and

$$
P_n(-\mu) = (-1)^n P_n(\mu)\,, \tag{A1.9}
$$

(A1.8) becomes

$$
\overline{\beta} = \frac{1}{2}\sum_{n=0}^{\infty}(-1)^n\omega_n\int_0^1 P_n(\mu)\,\mathrm{d}\mu\left\{\sum_{m=0}^{\infty} i_m\int_0^1 P_n(\mu')P_m(\mu')\,\mathrm{d}\mu'\right\}, \tag{A1.10}
$$

where $\omega_n$ are phase function expansion coefficients [note that $\omega_0 = 1$ as single scattering albedo appears explicitly in (A1.1)]. For isotropic irradiance, (A1.10) reduces to

$$\overline{\beta} = \frac{1}{2}\sum_{n=0}^{\infty}(-1)^n\omega_n\int_0^1\int_0^1 P_n(\mu)P_n(\mu')\,\mathrm{d}\mu\,\mathrm{d}\mu' \tag{A1.11}$$

which can be shown to equal (Wiscombe and Grams 1976)

$$\overline{\beta}_{\mathrm{iso}} = \frac{1}{2} - \frac{1}{8\pi}\sum_{n=0}^{\infty}\left[\frac{\Gamma\left(n+\frac{1}{2}\right)}{\Gamma\left(n+2\right)}\right]^2\omega_{2n+1}\,. \tag{A1.12}$$

For non-isotropic irradiance it can be shown (Barker 1994) that

$$\overline{\beta} = i_0\overline{\beta}_{\mathrm{iso}} + \frac{1}{4\sqrt{\pi}}\sum_{m=0}^{\infty}(-1)^m\left[\frac{\Gamma\left(m+\frac{1}{2}\right)}{\Gamma\left(m+2\right)}\right]\left(1-\frac{\omega_{2m+1}}{4m+3}\right)i_{2m+1}. \tag{A1.13}$$

Correspondingly, via a similar development (Wiscombe and Grams 1976), the backscatter function for direct-beam irradiance can be expressed as

$$\beta\left(\mu_0\right) = \frac{1}{2} - \frac{1}{4\sqrt{\pi}}\sum_{n=0}^{\infty}(-1)^n\left[\frac{\Gamma\left(n+\frac{1}{2}\right)}{\Gamma\left(n+2\right)}\right]\omega_{2n+1}P_{2n+1}(\mu_0). \tag{A1.14}$$

As another example, the Eddington approximation (Shettle and Weinman 1970) uses the first two Legendre terms and approximates $I$ and $p$ as

$$I(\mu) = i_0 + i_1\mu \tag{A1.15}$$

and

$$p\left(\mu_0,\mu'\right) = 1 + 3g\mu_0\mu'\,, \tag{A1.16}$$

where $g$ is the asymmetry parameter. This is often paired with the delta approximation (see Joseph et al. 1976) where the phase function is assumed to be

$$p\left(\mu_0,\mu'\right) = 2f\delta\left(\mu_0-\mu'\right) + \left(1-f\right)\left(1+3g\mu_0\mu'\right)\,, \tag{A1.17}$$

where $\delta$ is Dirac's distribution, and it is often adequate to set $f = g^2$. With these approximations, standard solutions to the two-stream [(1.23) through (1.28)] can be used with the basic input transformed as

$$\begin{aligned}\tau' &= \left(1-\omega_0 g^2\right)\tau\,,\\ \omega_0' &= \frac{1-g^2}{1-\omega_0 g^2}\omega_0\,,\\ g' &= \frac{g}{1+g}\,.\end{aligned} \tag{A1.18}$$

Table 1.1 gives parameter expressions for various common two-streams.

**Table 1.1.** References and parameter values for four common two-stream approximations. Note that by conservation of energy, $\gamma_3 = 1 - \gamma_4$

| Method | Reference | $\gamma_1$ | $\gamma_2$ | $\gamma_3$ |
|---|---|---|---|---|
| Eddington | Shettle + Weinman (1970) | $\frac{7-\omega_0(4+3g)}{4}$ | $-\frac{1-\omega_0(4-3g)}{4}$ | $\frac{2-3g\mu_0}{4}$ |
| delta-Eddington | Joseph et al. (1976) | $\frac{7-\omega_0'(4+3g')}{4}$ | $-\frac{1-\omega_0'(4-3g')}{4}$ | $\frac{2-3g'\mu_0}{4}$ |
| Coakley + Chýlek | Coakley + Chýlek (1975) | $2\left[1-\omega_0(1-\overline{\beta})\right]$ | $2\omega_0\overline{\beta}$ | $\beta\left(\mu_0\right)$ |
| PIFM | Zdunkowski et al. (1980) | $\frac{8-\omega_0(5+3g)}{4}$ | $\frac{3}{4}\omega_0'(1-g')$ | $\frac{2-3g'\mu_0}{4}$ |

## References


Aires, F. and W. B. Rossow, 2003: Inferring instantaneous, multivariate and nonlinear sensitivities for the analysis of feedback processes in a dynamical system: The Lorenz model case study. *Quart. J. Roy. Meteor. Soc.*, **129**, 239–275.

Astin, I., L. Di Girolamo, and H. M. van de Poll, 2001: Bayesian confidence intervals for true fractional coverage from finite transect measurements: Implications for cloud studies from space. *J. Geophys. Res.,* **106**, 17,303–17,310.

Barker, H. W. and J. A. Davies, 1992: Solar radiative fluxes for broken cloud fields above reflecting surfaces. *J. Atmos. Sci.*, **49**, 749–761.

Barker, H. W., 1994: A paremeterization and generalization of backscatter functions for two-stream approximations. *Beitr. Phys. Atmosph.*, **67**, 195–199.

Barker, H. W., 1996: A parameterization for computing grid-averaged solar fluxes for inhomogeneous marine boundary layer clouds. Part I: Methodology and homogeneous biases. *J. Atmos. Sci.*, **53**, 2289–2303.

Barker, H. W., B. A. Wielicki, and L. Parker, 1996: A parameterization for computing grid-averaged solar fluxes for inhomogeneous marine boundary layer clouds. Part II: Validation using satellite data. *J. Atmos. Sci.*, **53**, 2304–2316.

Barker, H. W., J.-J Morcrette, and G.D. Alexander, 1998: Broadband solar fluxes and heating rates for atmospheres with 3D broken clouds. *Q. J. R. Meteorol. Soc.*, **124**, 1245–1271.

Barker, H. W., G. L. Stephens, and Q. Fu, 1999: The sensitivity of domain-averaged solar fluxes to assumptions about cloud geometry. *Q. J. R. Meteorol. Soc.*, **125**, 2127–2152.

Barker, H. W., R. Pincus, and J.-J. Morcrette, 2002: The Monte Carlo Independent Column Approximation: Application within large-scale models. Proceedings from the GCSS workshop, Kananaskis, Alberta, Canada, May 2002. Available at: http://www.met.utah.edu/skrueger/gcss-2002/Extended-Abstracts.pdf.

Barker, H. W., and co-authors, 2003: Assessing 1D atmospheric solar radiative transfer models: Interpretation and handling of unresolved clouds. *J. Climate*, **16**, 2676–2699.

Barker, H. W. and P. Räisänen, 2005: Radiative sensitivities for cloud structural properties that are unresolved by conventional GCMs. *Q. J. R. Meteorol. Soc.*, **131**, 3103–3122.

Barker, H. W. and A. B. Davis, 2005: Approximation methods in atmospheric 3D radiative transfer, Part 2: Unresolved variability and climate applications. In *3D Radiative Transfer in Cloudy Atmospheres*, A. Marshak and A. B. Davis, Eds., Springer, Heidelberg, 686 pp.

Barker, H. W., 2005: Broadband irradiances and heating rates for cloudy atmospheres. In *3D Radiative Transfer in Cloudy Atmospheres*, A. Marshak and A. B. Davis, Eds., Springer, Heidelberg, 686 pp.

Benner, T. C. and K. F. Evans, 2001: Three-dimensional solar radiative transfer in small tropical cumulus fields derived from high-resolution imagery. *J. Geophys. Res.*, **106**, 14975–14984.

Bergman, J. W. and P. J. Rasch, 2002: Parameterizing vertically coherent cloud distributions. *J. Atmos. Sci.*, **59**, 2165–2182.

Bitz, C. M., J. C. Fyfe, and G. M. Flato, 2002: Sea ice response to wind forcing from AMIP models. *J. Climate*, **15**, 522–536.

Bony, S., J.-L. Dufresne, H. Le Treut, J.-J. Morcrette, and C. Senior, 2004: On dynamic and thermodynamic components of cloud changes. *Clim. Dyn.*, **22**, 71–86.

Briegleb, B. P., P. Minnis, V. Ramanathan, and E. Harrison, 1986: Comparison of regional clear-sky albedos inferred from satellite observations and model computations. *J. Clim. and Appl. Meteor.*, **25**, 214–226.

Cahalan, R. F., 1989: Overview of fractal clouds. In *Advances in Remote Sensing*. Deepak Publ., p. 371–389, 515 pp.

Cahalan, R. F., W. Ridgway, W. J. Wiscombe, T. L. Bell, and J. B. Snider, 1994: The albedo of fractal stratocumulus clouds. *J. Atmos. Sci.*, **51**, 2434–2455.

Cahalan, R. F., D. Silberstein, and J. Snider, 1995: Liquid water path and plane-parallel albedo bias during ASTEX. *J. Atmos. Sci.*, **52**, 3002–3012.

Cairns, B, A. A. Lacis, and B. E. Carlson, 2000: Absorption within inhomogeneous clouds and its parameterization in general circulation models. *J. Atmos. Sci.*, **57**, 700–714.

Cess, R. D., and others, 1993: Intercomparison of CO2 radiative forcing in atmospheric general circulation models. *Science*, **262**, 1252–1255.

Chambers, L. H., B. A. Wielicki, and K. F. Evans, 1997: Accuracy of the independent pixel approximation for satellite estimates of oceanic boundary layer cloud optical depth. *J. Geophys. Res.*, **102,** 1779–1794.

Charney, J., W. J. Quirk, S.-H. Chow, and J. Kornfield, 1977: A comparative study of the effects of albedo change on drought in semi-arid regions. *J. Atmos. Sci.*, **34**, 1366–1385.

Clothiaux, E. E., and co-authors, 1999: The atmospheric radiation measurement program cloud radars: Operational modes. *J. Atmos. and Ocean. Tech.*, **16**, 819–827.

Clothiaux, E. E., H. W. Barker, and A. V. Korolev, 2005: Observing clouds and their optical properties. In *3D Radiative Transfer in Cloudy Atmospheres*, A. Marshak and A. B. Davis, Eds., Springer, Heidelberg, 686 pp.

Coakley, J. A., Jr., and P. Chýlek, 1975: The two-stream approximation in radiative transfer: Including the angle of the incident radiation. *J. Atmos. Sci.*, **32**, 409–418.

Cole, J. N. S., H. W. Barker, W. O'Hirok, E. E. Clothiaux, M. F. Khairoutdinov, and D. A. Randall, 2005a: Atmospheric radiative transfer through global arrays of 2D clouds. *Geophys. Res. Lett.,* **32**, L19817, doi: 10.1029/2005GL023329.

Cole, J. N. S., H. W. Barker, D. A. Randall, M. F. Khairoutdinov, and E. E. Clothiaux, 2005b: Global consequences of interactions between clouds and radiation at scales unresolved by global climate models. *Geophys. Res. Lett.*, **32**, L06703, doi:10.1029/2004GL020945.

Cox, C. and W. Munk, 1956: Slopes of the sea surface deduced from photographs of the sun glitter. *Bull. Scripps Inst. Oceanog.*, **6**, 401–488.

Davis, A., P. Gabriel, S. Lovejoy, D. Schertzer, and G. Austin, 1990: Discrete angle radiative transfer – Part III: Numerical results and applications. *J. Geophys. Res.*, **95,** 11,729–11,742.

Evans, K. F. and A. Marshak, 2005: Numerical methods. In *3D Radiative Transfer in Cloudy Atmospheres*, A. Marshak and A. B. Davis, Eds., Springer, Heidelberg, 686 pp.

Fouquart, Y., B. Bonnel, and V. Ramaswamy, 1991: Intercomparing shortwave radiation codes for climate studies. *J. Geophys. Res.*, **96**, 8955–8968.

Fu, Q. and K.-N. Liou, 1992: On the correlated $k$-distribution method for radiative transfer in inhomogeneous atmospheres. *J. Atmos. Sci.*, **49**, 2139–2156.

Fu, Q., S. K. Krueger, and K.-N. Liou, 1995: Interactions of radiation and convection in simulated tropical cloud clusters. *J. Atmos. Sci.*, **52**, 1310–1328.

Geleyn, J. -F. and A. Hollingsworth, 1979: An economical analytical method for the computation of the interaction between scattering and line absorption of radiation. *Cont. Atmos. Phys.*, **52**, 1–16.

Grabowski, W. W., 2001: Coupling cloud processes with the large-scale dynamics using the cloud-resolving convection parameterization (CRCP), *J. Atmos. Sci.*, **58**, 978–997.

Gradshteyn, I. S. and Ryzhik, I. M. 1980: *Table of Integrals, Series, and Products: Corrected and enlarged edition.* Academic Press, London, 1160 pp.

Gregory, J. M. and co-authors, 2004: A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.*, **31**, L03205, doi:10.1029/2003GL018747.

Gu, Y. and K.-N. Liou, 2001: Radiation parameterization for three-dimensional inhomogeneous cirrus clouds: Application to climate models. *J. Climate*, **14**, 2443–2457.

Hansen, J. E., D. Russell, D. Rind, P. Stone, A. Lacis, L. Travis, S. Lebedeff, and R. Ruedy, 1983: Efficient three-dimensional global models for climate studies: Models I and II. *Mon. Wea. Rev.*, **111**, 609–662.

Hapke, B. W., 1981: Bidirectional reflectance spectroscopy I. Theory. *J. Geophys. Res.*, **86**, 3039–3054.

Hogan, R. J. and A. J. Illingworth, 2000: Derived cloud overlap statistics from radar. *Q. J. R. Meteorol. Soc.*, **126**, 2903–2909.

Houghton, J. T., and co-editors, 2001: *Climate Change 2001: The Scientific Basis.* Cambridge University Press, 881 pp.

Joseph, J. H., W. J. Wiscombe, and J. A. Weinman, 1976: The delta-Eddington approximation for radiative transfer. *J. Atmos. Sci.*, **33**, 2452–2459.

Khairoutdinov, M. F. and D. A. Randall, 2001: A cloud-resolving model as a cloud parameterization in the NCAR Community Climate System Model: Preliminary results. *Geophys. Res. Lett.,* **28,** 3617–3620.

Khairoutdinov, M. F. and D. A. Randall, 2003: Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties, and sensitivities, *J. Atmos. Sci.*, **60**, 607–625.

King, M. D. and Harshvardhan, 1986: Comparative accuracy of selected multiple scattering approximations, *J. Atmos. Sci.*, **43**, 784–801.

Kreyscher, M., M. Harder, P. Lemke, and G. M. Flato, 2000: Results of the Sea Ice Model Intercomparison Project: Evaluation of sea-ice rheology schemes for use in climate simulations. *J. Geophys. Res.*, **105**, 11299–11320.

Li, J. and V. Ramaswamy, 1996: Four-stream spherical harmonic expansion approximation for solar radiative transfer. *J. Atmos. Sci.*, **53**, 1174–1186.

Li, J. and H. W. Barker, 2002: Accounting for unresolved clouds in a 1D infrared radiative transfer model. Part II: Horizontal variability of cloud water path. *J. Atmos. Sci.*, **59**, 3321–3339.

Li. J., J. S. Dobbie, P. Räisänen, and Q. Min, 2005: Accounting for unresolved cloud in solar radiation. *Q. J. Royal Meteorol. Soc.*, **131**, 1607–1629.

Liou, K. -N., 1992: *Radiation and Cloud Processes in the Atmosphere*. Oxford University Press, New York, 487 pp.

Loeb, N. G., K. Loukachine, B. A. Wielicki, and D. F. Young, 2003: Angular distribution models for top-of-atmosphere flux estimation from the Clouds and the Earth's Radiant Energy System instrument on the Tropical Rainfall Mission Satellite. Part II: Validation. *J. Appl. Meteorol.*, **42**, 1748–1769.

Lovelock, J., 1988: *The Ages of Gaia*. W. W. Norton and Co., New York, USA, ISBN 0-393-02583-7, 252 pp.

Mace, G. G. and S. Benson-Troth, 2002: Cloud-layer overlap characteristics derived from long-term cloud radar data. *J. Climate*, **15**, 2505–2515.

McFarlane, N. A., G. J. Boer, J.-P. Blanchet, M. Lazare, 1992: The Canadian Climate Centre second-generation general circulation model and its equilibrium climate. *J. Climate*, **5**, 1013–1044.

McKee, T. B. and S. K. Cox, 1974: Scattering of visible radiation by finite clouds. *J. Atmos. Sci.*, **31**, 1885–1892.

Meador, W. E. and W. R. Weaver, 1980: Two-stream approximations to radiative transfer in planetary atmospheres: A unified description of existing methods and a new improvement. *J. Atmos. Sci.*, **37**, 630–643.

Monahan, E. C. and I. G. O'Muircheartaigh, 1987: Comments on 'Albedos and glitter patterns of a wind-roughened sea surface'. *J. Phys. Oceanogr.*, **17**, 549–550.

O'Hirok, W. and C. Gautier, 1998: A three-dimensional radiative transfer model to investigate the solar radiation within a cloudy atmosphere. Part I: Spatial effects. *J. Atmos. Sci.*, **55**, 2162–2179.

O'Hirok, W. and C. Gautier, 2005: The impact of model resolution on differences between independent column approximation and Monte Carlo estimates of shortwave surface irradiance and atmospheric heating rate, *J. Atmos. Sci.*, **62**, 2939–2951.

Oreopoulos, L. and R. Davies, 1998: Plane parallel albedo biases from satellite observations. Part I: Dependence on resolution and other factors. *J. Climate*, **11**, 919–932.

Oreopoulos, L. and H. W. Barker, 1999: Accounting for subgrid-scale cloud variability in a multi-layer, 1D solar radiative transfer algorithm. *Q. J. R. Meteorol. Soc.,* **125**, 301–330.

Otterman, J., 1984: Albedo of a forest modeled as a plane of dense protrusions. *J. Atmos. Sci.*, **23**, 297–307.

Payne, R. E., 1972: Albedo of the sea surface. *J. Atmos. Sci.*, **29**, 959–970.

Pincus, R., S. A. MacFarlane, and S. A. Klein, 1999: Albedo bias and the horizontal variability of clouds in subtropical marine boundary layers: Observations from ships and satellites. *J. Geophys. Res.*, **104**, 6183–6191.

Pincus, R., H. W. Barker, and J.-J. Morcrette, 2003: A new radiative transfer model for use in GCMs. *J. Geophys. Res.*, **108**(D13), 4376, doi:10.1029/2002JD003322.

Pinty, B., T. Lavergne, R. E. Dickenson, J.-L. Widlowski, N. Gobron, and M. M. Verstrate, 2006: Simplifying the interaction of land surfaces with radiation for related remote sensing products to climate models. *J. Geophys. Res.*, **111**, D02116, doi:10.1029/2005JD005952.

Preisendorfer, R. W. and C. D. Mobley, 1986: Albedos and glitter patterns of a wind-roughened sea surface. *J. Phy. Oceanogr.*, **16**, 1293–1316.

Räisänen, P., H. W. Barker, M. Khairoutdinov, J. Li, and D. A. Randall, 2004: Stochastic generation of subgrid-scale cloudy columns for large-scale models. *Quart. J. Roy. Meteor. Soc.*, **130**, 2047–2067.

Räisänen, P. and H. W. Barker, 2004: Evaluation and optimization of sampling errors for the Monte Carlo independent column approximation. *Quart. J. Roy. Meteorol. Soc.*, **130**, 2069–2086.

Räisänen, P., H. W. Barker and J. Cole, 2005: The Monte Carlo Independent Column Approximation's conditional random noise: Impact on simulated climate. *J. Climate*, **17**, 4715–4730.

Randall, D., M. Khairoutdinov, A. Arakawa and W. Grabowski, 2003: Breaking the cloud parameterization deadlock. *Bull. Am. Met. Soc.*, **84**, 1547–1564.

Ronnholm, K., M. B. Baker, and H. Harrison, 1980: Radiative transfer through media with uncertain or variable parameters. *J. Atmos. Sci.*, **37**, 1279–1290.

Ross, J., 1981: *The Radiation Regime and Architecture of Plant Stands.* Dr. W. Junk, Norwell, MA, 391 pp.

Rossow, W. B., C. Delo, and B. Cairns, 2002: Implications of the observed mesoscale variations of clouds for the Earth's radiation budget. *J. Climate*, **15**, 557–585.

Schlesinger, M. E. and J. F. B. Mitchell, 1987: Climate model simulations of the equilibrium climatic response to increased carbon dioxide. *Review of Geophys.*, **4**, 760–798.

Schneider, W. H., 1972: Cloudiness as a global climatic feedback mechanism: The effects on the radiation balance and surface temperature of variations in cloudiness. *J. Atmos. Sci.*, **29**, 1413–1422.

Shettle, E. P. and J. A. Weinman, 1970: The transfer of solar irradiance through inhomogeneous turbid atmospheres evaluated by Eddington's approximation. *J. Atmos. Sci.*, **27**, 1048–1055.

Snyder, P. K., J. A. Foley, M. H. Hitchman, and C. Delire, 2004: Analyzing the effects of complete tropical foret removal on the regional climate using a detailed three-dimensional energy budget: An application to Africa. *J. Geophys. Res.*, **109**, D211102, doi: 10.1029/2003JD004462.

Stamnes, K., S.-C. Tsay, W. J. Wiscombe, and K. Jayaweera, 1988: A numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media. *Appl. Opt.*, **27**, 2502–2509.

Stephens, G. L., 1988: Radiative transfer through arbitrary shaped optical media: Part II: Group theory and simple closures. *J. Atmos. Sci.*, **45**, 1837–1848.

Stephens, G. L., P. M. Gabriel, and S.-C. Tsay, 1991: Statistical radiative transfer in one-dimensional media and its application to the terrestrial atmosphere. *Trans. Theory Stat. Phys.*, **20**, 139–175.

Stephens, G. L., and co-authors, 2002: The CloudSat mission and the A-train: A new dimension of space-based observations of clouds and precipitation. *Bull. Amer. Meteorol. Soc.*, **83**, 1771–1790.

Stephens, G. L., N. B. Wood, and P. M. Gabriel, 2004: An assessment of the parameterization of subgrid-scale cloud effects on radiative transfer. Part I: Vertical overlap. *J. Atmos. Sci.*, **61**, 715–732.

Stokes, G. M. and S. E. Schwartz, 1994: The atmospheric radiation measurement (ARM) program: Programmatic background and design of the cloud and radiation test bed. *Bull. Amer. Meteor. Soc.*, **75**, 1201–1221.

Stowasser, M., K. Hamilton, and G. J. Boer, 2006: Local and global climate feedbacks in models with differing climate sensitivities. *J. Climate*, **19**, 193–209.

Tiedtke, M., 1996: An extension of cloud-radiation parameterization in the ECMWF model: The representation of subgrid-scale variations of optical depth. *Mon. Wea. Rev.*, **124**, 745–750.

Verstraete, M. M., 1987: Radiation transfer in plant canopies: Transmission of direct solar radiation and the role of leaf orientation. *J. Geophys. Res.*, **92**, 10,985–10,995.

Warren, S. G., 1982: Optical properties of snow. *Rev. Geophys. Space Phys.*, **20**, 67–89.

Warren, S. G., R. E. Brandt, and P. O'Rawe-Hinton, 1998: Effect of surface roughness on bidirectional reflectance of Antarctic snow. *J. Geophys. Res.*, **103**, 25,789–25,807.

Webb, M., C. Senior, S. Bony, and J.-J. Morcrette, 2001: Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF, and LMD atmospheric climate models. *Clim. Dyn.*, **17**, 905–922.

Welch, R. M and B. A. Wielicki, 1985: A radiative parameterization of stratocumulus cloud fields. *J. Atmos. Sci.*, **42**, 2888–2897.

Wiscombe, W .J. and G. W. Grams, 1976: The backscattered fraction in two-stream approximations. *J. Atmos. Sci.,* **33**, 2440–2451.

Wiscombe, W. J., 1977: The delta-Eddington approximation for a vertically inhomogeneous atmosphere. NCAR Tech. Note, TN-121+STR, 66 pp.

Zdunkowski, G. W., R. M. Welch, and G. Korb, 1980: An investigation of the structure of typical two-stream methods for the calculation of solar fluxes and heating rates. *Contr. Atmos. Phys.*, **53**, 147–166.

# 2 On the remote sensing and radiative properties of cirrus

Anthony J. Baran

## 2.1 Introduction

Imagine an evening sky just before sunset as one gazes into the dark blue sky whilst lying in a country field surrounded by bird song, there often appears high in the sky wispy thin fibrous clouds. These innocuous-looking clouds are called cirrus. A non-specialist might be forgiven for thinking that such insubstantial-looking clouds are unimportant to the climate system. In fact, nothing could be further from the truth, as this chapter will demonstrate.

Cirrus is high-level cloud and appears at altitudes usually greater than about 6 km occurring at all latitudes and during all seasons (Wylie et al., 1994). Cirrus can cover substantial parts of the Earth's surface; recent estimates suggest the coverage to be 20–30% with 60–70% coverage in the tropics (Liou, 1986; Hartmann et al., 1992; Wylie et al., 1994). With such a spatial and temporal coverage cirrus has an important impact on the Earth atmosphere radiative balance (Stephens and Webster, 1981; Mitchell et al., 1989; Liou and Takano, 1994; Lohmann and Roeckner, 1995; Donner et al., 1997; Kristjánsson et al., 2000; Hong et al., 2006; Edwards et al., 2007). This impact can manifest itself in several ways. Cirrus exists at low temperatures and is optically thin high-level cloud, which generally transmit solar radiation and absorb long-wave radiation. Since cirrus is cold little infrared radiation is emitted back to space, thereby warming the Earth's surface, this is a positive feedback. However, ice cloud associated with deep convection or fronts can reflect significant amounts of solar radiation back to space, thereby tending to cool the Earth's surface, this is a negative feedback. In the paper by Zhang et al. (1999) it is shown that the net radiative forcing of cirrus can vary for a fixed optical depth (i.e., the cloud extinction multiplied by its vertical geometric depth) between net warming and net cooling by changing the size of ice crystals from large to small, respectively. The overall sign of the net cirrus radiative forcing is crucial to determine if climate models are to realistically simulate future climate change (IPCC, 2001). In order to achieve this it is necessary to understand the basic microphysical and macrophysical composition of cirrus in terms of ice crystal size, ice crystal shape and Ice Water Content (IWC). The IWC is an important macrophysical variable in radiative

transfer (see, for example, Foot, 1988; Francis et al., 1994; Mitchell, 2002) and it is defined as the mass of ice present per unit volume and has units of $gm^{-3}$. For a complete review of cirrus microphysical and macrophysical properties see Lynch et al. (2002).

Since cirrus occurs at high altitudes these clouds predominantly consist of nonspherical ice crystals. The sizes and shapes of these nonspherical ice crystals can vary significantly. In terms of size cirrus ice crystals can vary between less than 50 μm to several thousand micrometres and the shapes can take on many different geometric forms (Garrett et al., 2005; Connolly et al., 2004; Gallagher et al., 2004; Heymsfield and Miloshevich, 2003; Korolev et al., 2000; McFarquhar and Heymsfield, 1996). Typically the shapes of nonspherical ice crystals that appear in cirrus can range from simple hexagonal ice columns, hexagonal ice plates, single bullets, bullet-rosettes having varying numbers of branches, to complex aggregates composed of roughened and/or distorted hexagonal columns. More recent observations demonstrate that chains of aggregates consisting of plates can also exist in tropical anvil cloud (Connolly et al., 2004) and Lawson et al. (2003) also found evidence of aggregate chains in continental anvils. Interestingly, no evidence of aggregate chains was found in anvils generated by maritime convection as reported by Lawson et al. (2003). It is remarked by Connolly et al. (2004) that the aggregate chains observed by them are strikingly similar to aggregate chains found in the laboratory under the influence of electric fields (Wahab, 1974; Saunders and Wahab, 1975).

Typical examples of ice crystal ensembles of varying shapes and sizes that might exist in cirrus are shown as a function of height in Fig. 2.1 and Fig. 2.2 (the images were provided by Andrew Heymsfield). The images shown in both figures were obtained using the Cloud Particle Imager (CPI) instrument and the CPI is described in the paper by Lawson et al. (2001). The crystal sizes shown in Fig. 2.1 are greater than 100 μm (Heymsfield and Miloshevich, 2003) and it is evident from the figure that there is little evidence of the more pristine shapes such as hexagonal columns or plates, the most common shapes appear to be rosettes and chains of rosettes and the rosettes appear spatial rather than compact. Although in Fig. 2.1 there does appear to be the odd hexagonal column, there is evidence of air inclusions, both in the single columns and in some branches of the rosettes. The shapes shown in Fig. 2.2 are dominated by bullet-rosettes or aggregates of rosettes for crystal sizes larger than 100 μm with again very little evidence of pristine ice crystal shapes such as hexagonal columns or hexagonal plates. The shape of ice crystals less than 100 μm in size is currently unknown due to the limiting resolving power of the CPI. As shown in Fig. 2.2 these shapes of less than 100 μm in size can appear as quasi-spherical or spheroidal but nonetheless may still be irregular. It is very important to characterize the size and shapes of ice crystals smaller than 100 μm as these may exist in significant concentrations and can have a large impact on the radiative properties of cirrus (Ivanova et al., 2001; Yang et al., 2001). Since ice crystals less than 100 μm in size appear 'quasi-spherical' it is often the case that such crystals are modelled as spheres as suggested by McFarquhar et al. (1999) or Chebyshev polynomials (McFarquhar et al., 2002), spheroids were suggested by Asano and Sato (1980).

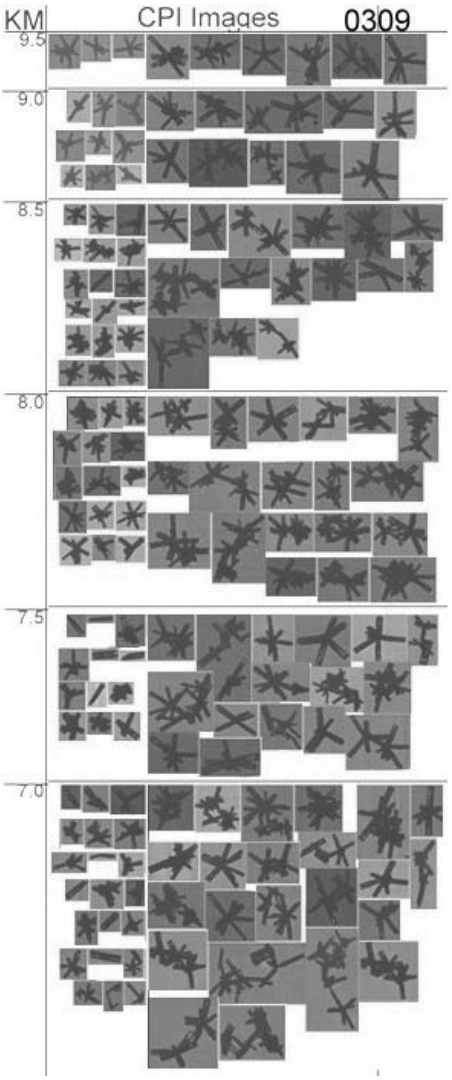



**Fig. 2.1.** A set of ice crystal images shown as a function of height. The images were obtained using the CPI instrument (courtesy A. Heymsfield).

Other possibilities are the Gaussian random sphere as suggested by Nousiainen and McFarquhar (2004) or droxtals as proposed by Yang et al. (2003). Although it is not currently possible to say which of these representations truly represents ice crystals less than 100 µm in size, it is, however, likely that small ice crystals are faceted as commented by Heymsfield and Platt (1984). In order to distinguish the shapes of small ice crystals an enhanced CPI is required or a new approach. A new approach could be based on two-dimensional scattering patterns as proposed by Clarke et al. (2006). In the paper by Clarke et al. (2006) it is shown that

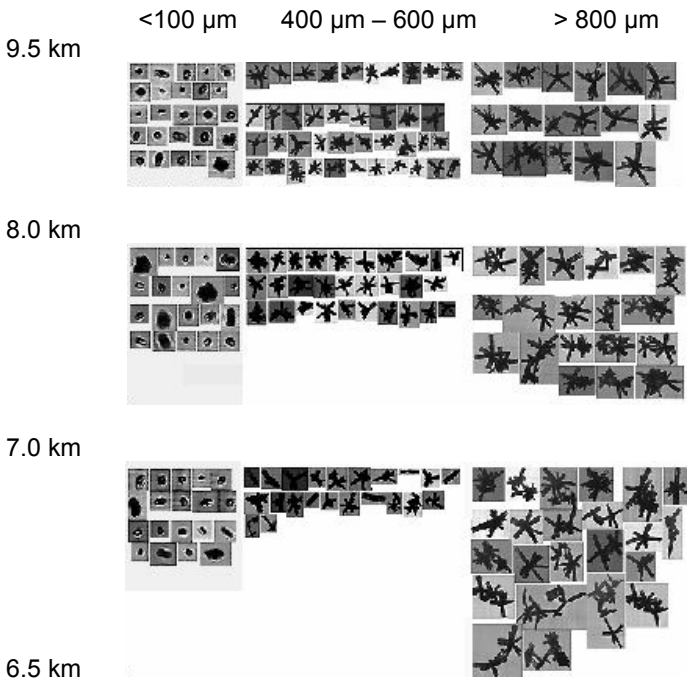



**Fig. 2.2.** Same as Fig. 2.1 but ice crystal size is shown on the top along the $x$-axis (courtesy A. Heymsfield).

the 2D scattering patterns between a small hexagonal column, hexagonal plate, rosette and droxtal are quite different and these differences could be used to potentially classify ice crystal shapes less than 100 µm in size.

From the currently available evidence it can be said that the most common type of ice crystal that inhabits synoptically generated cirrus is bullet-rosettes whilst anvil cirrus is chiefly populated by non-symmetric irregulars. This is further supported by observations made by Korolev et al. (2000) in the mid-latitudes and McFarquhar and Heymsfield (1996) whom made measurements of ice crystal size and shape in deep tropical convection.

With such a variability of ice crystal size and shape, adequately modelling and computing cirrus scattering and absorption properties is problematic; but this problem must be addressed if the net radiative impact of cirrus is to be quantified. The rest of this chapter will be devoted to reviewing the current modelling approaches to representing cirrus ice crystal shapes and the current methodologies adopted in computing their scattering and absorption properties. The chapter ends by reviewing how such ice crystal scattering and absorption properties can be tested by remotely sensing, cirrus using both airborne and space-based instruments. The chapter attempts to bring together the importance of cirrus microphysical and macrophysical properties to the light scatter-

ing problem and how these two equally important components can be combined to improve understanding of the net radiative impact of cirrus.

## 2.2 Cirrus ice crystal models

As can be seen from Fig. 2.1 and Fig. 2.2 the shapes of cirrus ice crystals are complex and so, in order to model their scattering and absorption properties, idealized geometric shapes are sought. The typical range of ice crystal models currently used is shown in Fig. 2.3. The modelling of shapes which have symmetric properties such as the hexagonal column, hexagonal plate and bullet-rosette is straightforward since these have a well defined three-dimensional geometry as shown in Figs. 2.3 (b), 2.3 (c), and 2.3 (d), respectively. However, as the complexity of ice crystal shape increases then there are a number of possibilities as

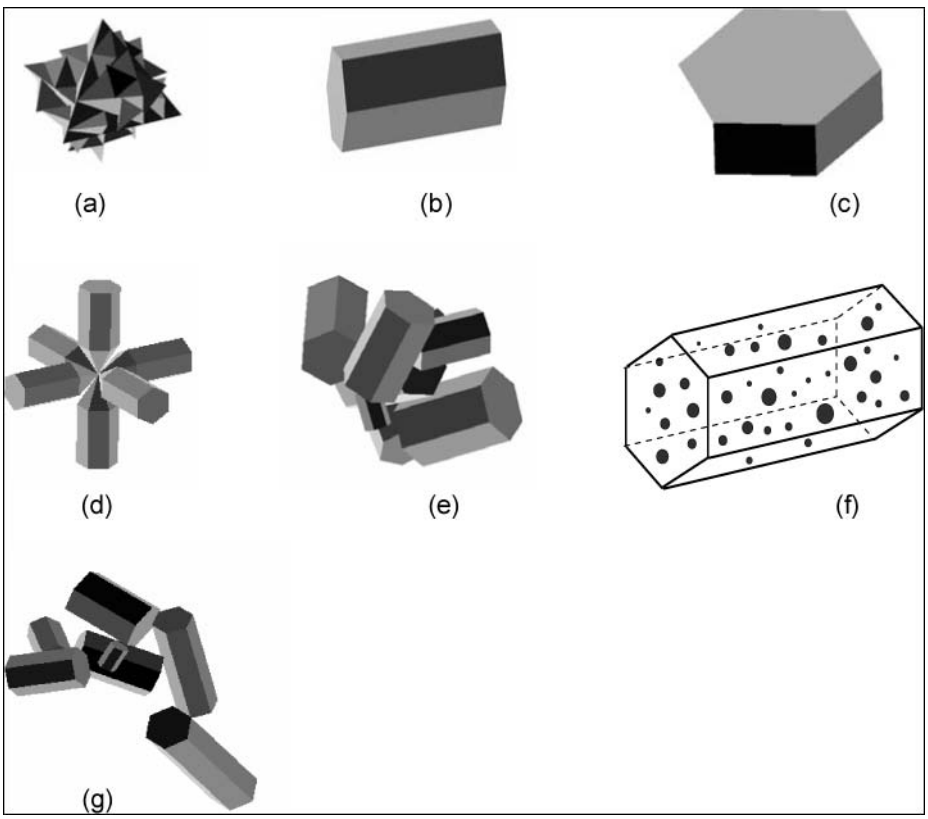



**Fig. 2.3.** Geometrical realizations of ice crystal shapes showing (a) randomized polycrystal, (b) pristine hexagonal ice column, (c) hexagonal ice plate, (d) six-branched bullet-rosette, (e) randomized hexagonal ice aggregate, (f) inhomogeneous hexagonal monocrystal, (g) chain-like aggregate.

to how to represent such a complexity. In Fig. 2.3 four such realizations of complex ice crystals are illustrated. In Fig.2.3 (a) the 'polycrystal' due to Macke et al. (1996) is a randomization of the second-generation triadic Koch fractal, the basic element of which is the tetrahedron and the polycrystal remains invariant with respect to size. The polycrystal is supposed to represent in one ice crystal model the variability of shape observed in Fig. 2.1 and Fig. 2.2. Figure 2.3 (e) illustrates the hexagonal ice aggregate introduced by Yang and Liou (1998); the aggregate is composed of eight hexagonal elements, the surfaces of which can be roughened. The hexagonal ice aggregate also remains invariant with respect to size. Figure 2.3 (f) shows the Inhomogeneous Hexagonal Monocrystal (IHM), which was introduced by Labonnote et al. (2001) in order to retain the simplicity of the hexagonal column but introducing randomization of the ice crystal by adding inclusions such as air bubbles and aerosol. Figure 2.3 (g) shows the chain-like aggregate introduced by Baran and Labonnote (2006) based on the Yang and Liou (1998) aggregate but with two of the original hexagonal elements elongated and re-transformed into a chain. This model is supposed to capture the more spatial and chain-like properties shown in Fig. 2.1 and Fig. 2.2.

The common feature of the polycrystal and hexagonal ice aggregate is that their aspect ratio remains invariant with respect to size. Rather than arbitrarily constructing some ice crystal model it would be desirable to predict resulting crystal shapes from an initial monomer crystal, which is allowed to aggregate. This approach has been applied by Westbrook et al. (2004), where a distribution of monomer ice crystals such as single six-branched rosettes are allowed to collide until a distribution of aggregates is produced. An example of such a fully grown ice aggregate is shown in Fig. 2.4. As can be seen from the figure the resulting aggregate has an aspect ratio greater than unity and is also spatial. These are the two properties which are common to Fig. 2.1 and Fig. 2.2. It has also been demonstrated by Westbrook et al. (2004) that the resulting aspect ratio of the ice aggregate asymptotes to 1.54 and is independent of assumptions regarding the initial monomer.

The geometric ice crystal representations illustrated by Fig. 2.3 are single model realizations but, as shown in Fig. 2.1 and Fig. 2.2, in reality ensembles of different shapes occur which Westbrook et al. (2004) attempt to emulate. It is becoming more common to construct ensembles of geometric shapes rather than assume one single geometric shape over the entire particle size distribution function. Such an approach has been utilized by Rolland et al. (2000) and McFarquhar et al. (1999, 2002). More recently, Baum et al. (2005) have demonstrated that a mixture of shapes can better represent the bulk IWC than single shape models such as the hexagonal ice aggregate. The mixture of ice crystal shapes proposed by Baum et al. (2005) comprise droxtals, hexagonal plates, solid hexagonal columns, hollow columns, bullet-rosettes and aggregates.

In this chapter another approach to representing the distribution of cirrus ice crystal size and shape by some distribution of idealized shapes is presented and has been described in Baran (2006). As Fig. 2.1 and Fig. 2.2 illustrate, ice crystal shape appears to become more progressively spatial and complex lower in the cloud. In order to mimic this change in shape as a function of crystal

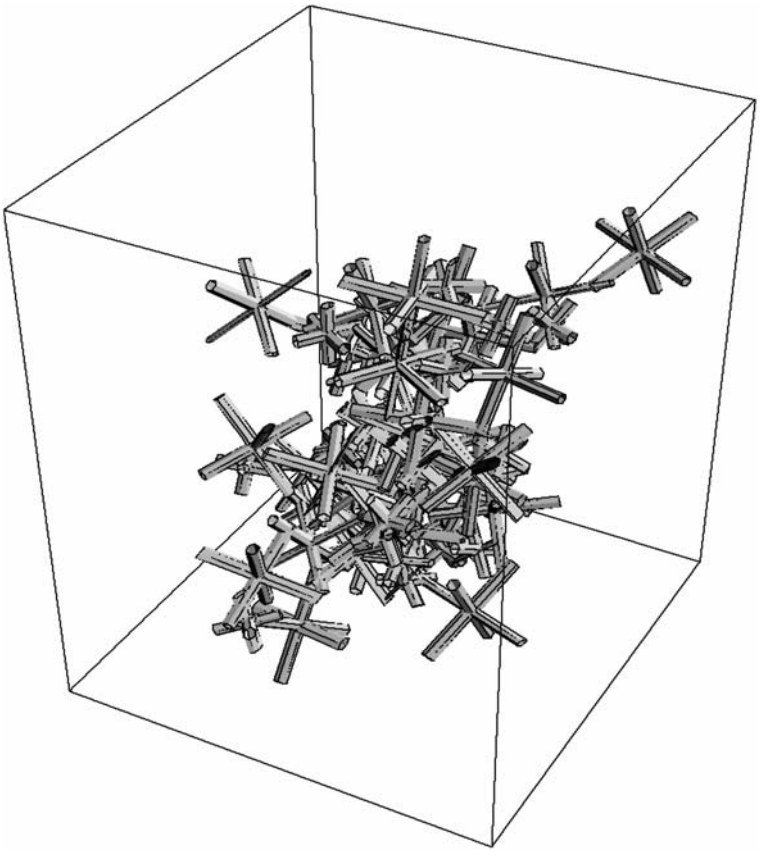


**Fig. 2.4.** A realization of an ice aggregate ice crystal grown from an initial six-branched hexagonal rosette. The aspect ratio of the fully grown aggregate is 1.54 (reproduced, with permission, from Baran, 2003a).

maximum dimension (literally the largest extent of the crystal) an ensemble model, as illustrated in Fig. 2.5, has been constructed. The smallest ice crystals consist of solid hexagonal ice columns assuming an aspect ratio of unity (i.e., ratio between column length and diameter). As the maximum dimension of ice crystals increase, the shapes become progressively more complex and spatial by arbitrarily attaching other column elements. One important aspect of the ensemble model shown in Fig. 2.5 is that the overall aspect ratio does not remain invariant with respect to the ice crystal maximum dimension. The various shapes in Fig. 2.5 are assumed to be distributed equally throughout the particle size distribution function. In the paper by Baum et al. (2005) the particle size distribution functions are obtained from many different field campaigns with no clear relationship between IWC and the cloud temperature ($T_c$). It would be desirable to relate the particle size distribution function (PSD) to macroscopic variables such as IWC and $T_c$ such that the PSD can be generated from any given val-

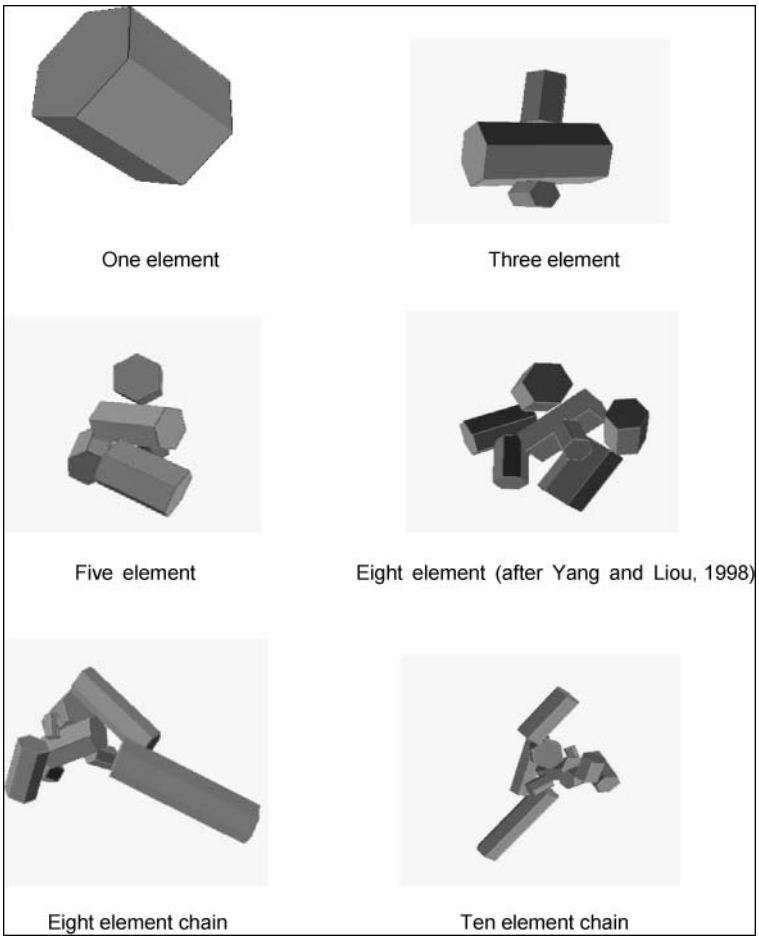



**Fig. 2.5.** The ensemble model. The one element model represents the smallest ice crystals taken to be a hexagonal ice column assuming an aspect ratio of unity, whilst the ten element model represents the largest ice crystals. All elements are assumed to be equally distributed in the particle size distribution function.

ues of these two variables. Such a parametrization has been realised by Field et al. (2005) and the parametrization should be of value in climate and numerical weather prediction models where IWC and $T_c$ are important variables. The parametrization due to Field et al. (2005) is based on many *in situ* measured PSD obtained in mid-latitude stratiform ice cloud at temperatures between 0°C and −60°C. The paper demonstrates importantly that the many PSD may be represented by a single underlying PSD from which the initial PSD can be retrieved from knowledge of two moments. Field et al. (2005) make use of the IWC (second moment of the PSD) and by using $T_c$ they obtain power laws to link IWC to any moment. However, the paper by Field et al. (2005) does not quan-

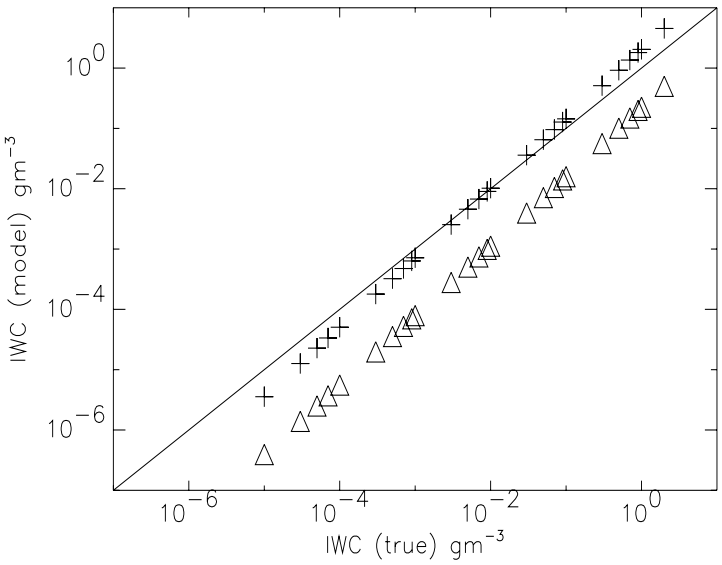



**Fig. 2.6.** The predicted IWC assuming the ensemble model (plus signs) and the hexagonal ice aggregate (triangles) plotted as a function of true IWC for $T_c = -30°C$. The full line shows the one-to-one relationship.

tify the contribution of ice crystals smaller than 100 µm to the PSD; however, the parametrization is independent of ice crystal shape assumptions. Therefore, from given values of IWC and $T_c$ the PSD can be generated. From the generated PSD using the parametrization due to Field et al. (2005) the IWC can then be predicted from the ensemble model shown in Fig. 2.5. The predicted IWC from the ensemble model can then be compared with the true IWC used to generate the PSD. The results of comparing the predicted IWC from the ensemble model with the true IWC are shown in Fig. 2.6 for $T_c = -30°C$, also shown in the figure are the results for the hexagonal ice aggregate. The figure shows that the hexagonal ice aggregate is not a good predictor of the true IWC, which is consistent with Baum et al. (2005). In contrast, the ensemble model prediction of the true IWC is good. In general the ensemble model prediction of the true IWC is generally well within a factor 2. Results of comparison assuming $T_c = -60°C$ are shown in Fig. 2.7. In this case the hexagonal ice aggregate under-predicts the true IWC by significant factors whilst the ensemble model under predicts by about a factor 2 though this improves with increasing true IWC. Typical ranges of measured IWC at $T_c = -30°C$ and $-60°C$ are 0.01–1.0 gm$^{-3}$ and 0.001–0.1 gm$^{-3}$, respectively (see Fig. 2.2 in Field et al., 2005, top scale). Considering the measured range of IWC, the ensemble model prediction of IWC at $T_c = -30°C$ is in excellent agreement with the true IWC and at $T_c = -60°C$ the agreement is satisfactory. It should be remarked that the measurements made by Field et al. (2005) at temperatures of $-60°C$ were at the limits of instrumental capability and the contribution of ice crystals less than 100 µm in size was

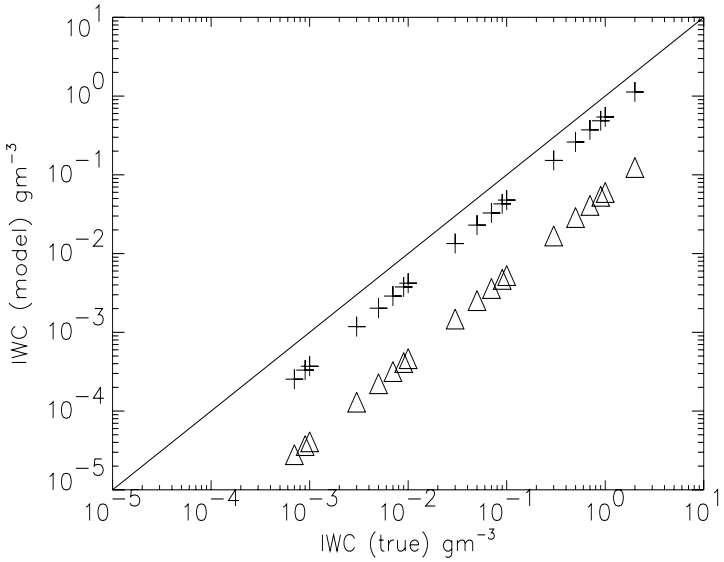



**Fig. 2.7.** Same as Fig. 2.6 but for $T_c = -60°C$.

ignored. Figure 2.6 and Fig. 2.7 demonstrate that it is possible to construct an ensemble ice crystal model which from the PSD predicts reasonable values for the IWC. Linking the ice crystal model to important variables such as IWC and $T_c$, via the PSD, is important if cirrus parametrization in climate models is to be further improved. Given an ensemble ice crystal model such as that shown in Fig. 2.5 the question of computing its scattering and absorption properties arise. The next section reviews computational methods currently used to obtain scattering and absorption characteristics of nonspherical ice crystals.

## 2.3 Computational methods applied to nonspherical ice crystals

As can be seen from Fig. 2.1 and Fig. 2.2 the range of ice crystal size and shape is significant and computation of their single scattering properties is demanding. The problem is to be able to apply some computational method to a tractable geometry resulting in stable and convergent solutions which represent the scattering and absorption properties of real ice crystals. The basic description of incident light being scattered from a collection of randomly oriented ice crystals suspended in the Earth's atmosphere is briefly outlined below.

Assuming an incident unpolarized beam of light upon an ensemble of randomly oriented ice crystals which each posses a plane of symmetry, the Stokes vector of the incident light ($I_{inc}$, $Q_{inc}$, $U_{inc}$, $V_{inc}$) is linearly related to the Stokes vector of the scattered light ($I_{sca}$, $Q_{sca}$, $U_{sca}$, $V_{sca}$) by a $4 \times 4$ scattering matrix, for each scattering angle, $\theta$, given by (van de Hulst, 1957):

$$\begin{pmatrix} I_{\text{sca}} \\ Q_{\text{sca}} \\ U_{\text{sca}} \\ V_{\text{sca}} \end{pmatrix} = \frac{c_{\text{sca}}}{4\pi r^2} \begin{pmatrix} P_{11} & P_{12} & 0 & 0 \\ P_{21} & P_{22} & 0 & 0 \\ 0 & 0 & P_{33} & P_{34} \\ 0 & 0 & P_{43} & P_{44} \end{pmatrix} \begin{pmatrix} I_{\text{inc}} \\ Q_{\text{inc}} \\ U_{\text{inc}} \\ V_{\text{inc}} \end{pmatrix} \tag{2.1}$$

where in Eq. (2.1) $C_{\text{sca}}$ is the ice crystal scattering cross-section (extinction efficiency multiplied by the ice crystal geometric cross-section) and r is the distance of the scattering particle from some observer. The $4\times4$ matrix shown in Eq. (2.1) is called the phase matrix. Due to the assumed symmetry properties of the system then, out of the eight elements shown in Eq. (2.1) only six are independent due to $P_{21} = P_{12}$ and $P_{43} = -P_{34}$ (see van de Hulst, 1957). Since incident unpolarized light has been assumed (which is the case for incident sunlight) the first element in Eq. (2.1) is proportional to the scattered light and $P_{11}$ is called the scattering phase function. It is normalized as follows:

$$\frac{1}{2}\int_{-1}^{1} P_{11}(\theta)\sin\theta\,\mathrm{d}\theta = 1 \tag{2.2}$$

Moreover, $-P_{12}/P_{11}$ describes the degree of linear polarization (DLP). The $P_{11}$ element and DLP are useful quantities in the remote sensing of cirrus since $P_{11}$ and DLP depend on the shape and size of ice crystals.

In this chapter the orientation of ice crystals is assumed to be random. This assumption leads to the following question. What is the current evidence that atmospheric ice crystals are randomly oriented in space? In the paper by Chepfer et al. (1999) it was reported that at least 40% of their space-based measurements of cirrus suggested that the ice crystals were horizontally oriented. However, in more recent papers by Bréon and Dubrulle (2004) and Noel and Chepfer (2004) they conclude that the actual fraction of horizontally oriented ice crystals is more likely to be about $10^{-2}$. Therefore, given this information the assumption of randomly oriented ice crystals can be generally applied, at least for solar and infrared measurements of cirrus.

In terms of radiative transfer the single scattering properties that are required to compute the radiative properties of cirrus are the volume extinction coefficient, $K_{\text{ext}}$, volume scattering coefficient, $K_{\text{sca}}$, the single scattering albedo (the ratio of the scattered energy to the total amount of attenuated energy), $\omega_0$, and the asymmetry parameter, $g$. The volume extinction/scattering coefficient is defined as

$$K_{\text{ext,sca}} = \int Q_{\text{ext,sca}}(\boldsymbol{q})\,\langle S(\boldsymbol{q})\rangle\,n(\boldsymbol{q})\,\mathrm{d}\boldsymbol{q} \tag{2.3}$$

where $Q_{\text{ext,sca}}(\boldsymbol{q})$ is the extinction/scattering efficiency factor (defined as the ratio between the scattering/extinction cross-section and geometric area of the ice crystal), $\langle S(\boldsymbol{q})\rangle$ is the orientation averaged geometric area and $n(\boldsymbol{q})$ is the PSD. Each term in Eq. (2.3) is expressed as a function of the vector parameter, $\boldsymbol{q}$, which characterizes the shape and size of ice crystal. The single scattering albedo is given by

$$\omega_0 = K_{\text{sca}}/K_{\text{ext}} \tag{2.4}$$

and the asymmetry parameter is a parametrization of the $P_{11}$ element into a single number which describes how much incident radiation is scattered into the backward and forward hemispheres and can take on values between $-1$ and 1 depending on the size, shape, and refractive index of the scatterer. The formal definition of $g$ is the average cosine of the scattering angle:

$$g = \langle \cos\theta \rangle = \int_{-1}^{1} \mathrm{d}(\cos\theta) P_{11}(\cos\theta) \cos\theta \,. \tag{2.5}$$

In order to compute or measure $g$ from Eq. (2.5) the angular dependence of the $P_{11}$ element must be known. The asymmetry parameter is a very important quantity in climate models as choice of $g$ determines the radiative impact of cirrus (Stephens et al., 1990). The reason why choice of $g$ is so important is because the backward reflection of incident sunlight assuming conservative scattering depends on $1 - g$, so for small and large values of $g$ reflection of sunlight back to space increases and decreases, respectively. It is, therefore, necessary to constrain the value of $g$ so that the most representative value can be applied to climate models. Calculations of $g$ for the various ice crystal models described in section 2.2 range from 0.74 for the polycrystal (Macke et al., 1996), 0.77 for the hexagonal ice aggregate (Yang and Liou, 1998) and 0.75–0.84 for the solid hexagonal column (Takano and Liou, 1989a). It is possible to 'measure' $g$ using the Cloud Integrating Nephelometer (CIN) described in Gerber et al. (2000). The $P_{11}$ element measured by CIN is truncated at a scattering angle of 10°, thus the full phase function is not utilized in the 'measurement' of $g$. The CIN instrument has measured $g$ in an Arctic ice cloud consisting of bullet-rosettes at visible wavelengths, suggesting values around 0.74. In tropical cirrus CIN measured g values of $0.75 \pm 0.01$ as reported in Garrett et al. (2003). In tropical anvil cloud CIN generally measured $g$ values ranging between 0.70 and 0.74 as discussed in Garrett et al. (2005). Baran et al. (2005) estimated the asymmetry parameter using ground-based Polar Nephelometer measurements of Antarctic ice crystals as they fell into a light scattering chamber, and the irregular ice crystal ensemble was found to have a $g$ value of $0.74 \pm 0.02$. In the paper by Field et al. (2003) the asymmetry parameter in mid-latitude cirrus was estimated to be 0.76 using an airborne light scattering probe by fitting a phase function to the angular intensity measurements that depends on $g$ only. Currently, there appears to be some convergence of measured $g$ values of around $0.74 \pm 0.02$ for atmospheric ice crystals according to the literature. However, further measurements of $g$ are required both in the laboratory of actual ice crystals and in the field using new instrumentation that captures a near-complete phase function before one can say that true convergence has been achieved.

The computation of Eqs (2.1)–(2.5) over the entire spectrum of ice crystal size is not an easy task since typical cirrus ice crystal size parameters (product of particle characteristic size and wavenumber, where the wavenumber is $2\pi/\lambda$ and $\lambda$ is the incident wavelength) can range between less than unity to thousands. Currently, there is no one single computational method that is capable of covering the whole cirrus size parameter space. As a result approximate methods

are still required to bridge the gap between small size parameters, intermediate size parameters and large size parameters.

Electromagnetic methods are usually applied to ice crystals covering small size parameter space of less than about 40. The Finite-Difference-Time-Domain (FDTD) method has been applied to compute the single scattering properties of the hexagonal ice aggregate up to a size parameter of about 20 (Yang and Liou, 1995; Yang et al., 2000, 2004) and Havemann et al. (2003) have applied the T-matrix method to the solid finite hexagonal ice column at size parameters of up to about 40. Sun et al. (1999) have applied the FDTD method to the ice sphere at size parameters of up to 40. The accuracy of the FDTD method has been tested by Baran et al. (2001a) against the T-matrix method for a finite randomly oriented solid hexagonal ice column and it was found that the relative errors for $C_{\text{ext}}$, $\omega_0$, and $g$ were less than 1%. The FDTD method has also been applied to more general shapes such as the bullet-rosette (Baum et al., 2000), Gaussian random particles (Sun et al., 2003) and droxtals (Yang et al., 2003). The T-matrix method has now been applied to more general polyhedral prisms by Kahnert et al. (2001). Other electromagnetic methods that have been applied to ice crystals include the Separation of Variables Method (SVM) developed by Rother et al. (2001) to compute the scattering matrix elements of the infinite hexagonal ice column. Kokhanovsky (2005a,b, 2006) has used the discrete dipole approximation (DDA) to compute the scattering properties of hexagonal and cubic ice crystals. By combining desirable properties of the T-matrix and DDA, Mackowski (2002) has developed the Discrete Dipole Method of Moments (DDMM) approach, which shows promise for computing the single scattering properties of ice crystals for size parameters of about 40. The boundary-element method has been successfully applied by Mano (2000) to compute the single scattering properties of oriented finite hexagonal ice columns for size parameters of 50. An extensive review of electromagnetic methods can be found in Mishchenko et al. (2002) and Kahnert (2003).

For the intermediate size parameter range ($\sim$20 to 60), there are a number of physical optics based approaches which fill the gap between 'exact' and approximate methods. In the paper by Yang and Liou (1996) it is shown that the method of Improved Geometric Optics (IGO) applied to the geometry of the solid hexagonal ice column converges to FDTD solutions for the extinction cross-section and single scattering albedo at size parameters around $\sim$20. This holds for the $P_{11}$ element in Eq. (2.1) as well (Yang and Liou, 1995). The physical optics approach of Muinonen (1989) could also be applied to ice crystals. More recently, a computationally fast edge diffraction method has been proposed by Hesse and Ulanowski (2003) and further developed in Clarke et al. (2006). This new approach has been specifically developed for ice crystals with facets, though in principle it could be applied to any arbitrary dielectric faceted object. Diffraction on facets has been compared with SVM in computing the $P_{11}$ element and the asymmetry parameter, assuming oriented hexagonal columns of size parameters 50 and 100 (Hesse et al., 2003). Borovoi and Grishin (2003) have developed a proper ray-tracing method for computation of the Jones scattering matrix inclusive of diffraction and phase information is accounted for exactly,

which has been applied to compute the backscattering properties of large hexagonal ice columns. For size parameters much larger than 60, then ray-tracing can be applied to any arbitrary ice crystal shape. The first solution of the 3D problem assuming hexagonal columns was achieved by Wendling et al. (1979) but without polarization. Polarization was incorportated by Cai and Liou (1982) and further refinements such as including birefringence and particular ice crystal orientations were reported in Takano and Liou (1989a,b). The ray-tracing method was applied to more complex shapes most notably by Macke (1993) and Macke et al. (1996) in which it was shown for the first time that the polycrystal could produce asymmetry parameters as low as 0.74 at non-absorbing wavelengths. Borovoi et al. (2000) computed the backscattering cross-section of arbitrarily oriented hexagonal ice columns at visible wavelengths using ray-tracing. It was found that for a tilt angle of 32.5° a very large backscattered intensity peak occurs, which is explained by a corner-reflector effect. The authors suggest that this finding could be used to discriminate between aligned hexagonal ice plates and hexagonal ice columns by using slant lidar. More recently, Borovoi et al. (2005) proposed an optical model for cirrus clouds by parametrizing the phase functions for a variety of randomly oriented ice crystal particles by means of weight coefficients for the wedges occurring in each ice crystal shape.

Other methods that have been suggested to compute the single scattering properties of nonspherical ice crystals are modified anomalous diffraction theory proposed by Mitchell et al. (1996), which has been further developed in Mitchell (2002) and Mitchell et al. (2006). The modified anomalous diffraction theory is based on the Bryant and Latimer (1969) approximation (BL), which approximates the original van de Hulst (1957) anomalous diffraction theory (ADT). The original ADT assumes that the size of particle is much greater than the incident wavelength and that the refractive index is close to unity. The BL approach attempts to apply ADT to nonspherical particles by taking the ratio of the particle volume-to-averaged cross-sectional area as a size and phase shift parameter. The original ADT and the BL approximation do not incorporate internal reflection, surface waves or large angle diffraction. The approach of Mitchell et al. (1996) was to incorporate this missing physics into the BL approximation. However, Sun and Fu (2001) compared the Bryant and Latimer approximation against exact ADT for computing the extinction coefficient of the finite hexagonal column and showed that BL could be in significant error. Though Mitchell et al. (2001, 2006) showed that modified ADT was in good agreement with laboratory-based experiments of hexagonal column ice crystal extinction efficiency between the wavelengths of 2.2 and 16.0 μm. Since ADT has no angle dependence (except at $\theta \rightarrow 0°$) then neither the $P_{11}$ element at arbitrary $\theta$ nor the asymmetry parameter can be calculated.

The application of the electromagnetic methods to the geometries outlined in section 2.2 is a complex task. Therefore, Baran (2003b) proposed that it might be possible to simulate the infrared properties of more complex ice crystals by representing the complex ensemble by some ensemble of symmetric ice crystals of varying aspect ratio. In the paper by Baran (2003b) the T-matrix method due to Mishchenko and Travis (1998) was applied to an ensemble of circular ice

cylinders and by conserving the volume-to-projected area ratio of the hexagonal ice aggregate it was shown that the total optical properties ($C_{\text{ext}}$, $\omega_0$ and $g$) of the more complex shape could be simulated to well within 4% when compared against solutions from FDTD. The method in principle could be applied to any ice crystal shape at infrared wavelengths and, since $C_{\text{ext}}$ and $\omega_0$ depend largely on the volume-to-area ratio, it is expected to work well for those quantities. Similar approaches of representing scattering from ensembles of complex non-symmetric shapes by ensembles of more symmetric shapes have also been applied to scattering by aerosols (Kahnert et al., 2002a). The difficulty with using ensembles of simpler shapes to represent scattering from more complex shapes is that angle-dependent quantities such as the matrix elements in Eq. (2.1) are more difficult to simulate as demonstrated by Kahnert et al. (2002b). It was shown by Lee et al. (2003) that randomly oriented finite circular cylinders could be used to simulate the single scattering properties of randomly oriented hexagonal ice columns at infrared wavelengths to within a few percent.

To bridge the gap between small and large size parameter space Liou et al. (2000) proposed the 'unified' method where FDTD and IGO are combined to calculate the single scattering properties over the whole ice crystal PSD, and in principle this method can be applied to any ice crystal shape. The approach proposed by Fu et al. (1999) is similar to the 'unified' method.

The methods outlined in this section can be used to compute the single scattering properties of the ice crystal realizations outlined in section 2.2. However, the problem in computing the radiative properties of cirrus with such a diversity of ice crystal shape is how best to represent the single scattering properties by some single dimension? Should that dimension be maximum size, chord length, or facet length? What is required is a common dimension such that the radiative properties can be computed independently of ice crystal shape and shape of the PSD. For example, water clouds are more straightforward (Slingo, 1989) since these are composed of water spheres and their PSDs are not as dispersed as cirrus. In recent years there appears to have been a consensus of opinion as to which dimension to apply in computing the radiative properties of cirrus. Initially, it was proposed by Foot (1988) that cirrus radiative properties might well be represented if the distribution of ice crystal shapes and sizes was expressed as a ratio of the distribution volume-to-distribution averaged cross-section. This proposition of an effective dimension has now been adopted by a number of authors when computing cirrus solar bulk single scattering properties or retrieving cirrus microphysical/macrophysical properties (Francis et al., 1994; Fu, 1996; Yang et al., 1997; McFarquhar and Heymsfield, 1998; Wyser and Yang, 1998; Mitchell, 2002; McFarquhar et al., 2002; Baran et al., 2003; Baran and Havemann, 2004). Thus throughout the rest of this chapter the cirrus PSD is characterized by an effective dimension called the effective diameter, $D_{\text{e}}$, defined as,

$$D_{\text{e}} = 3/2 \int V(D_{\text{m}}) n(D_{\text{m}})\, \mathrm{d}D_{\text{m}} \Big/ \int \langle S(D_{\text{m}}) \rangle\, n(D_{\text{m}})\, \mathrm{d}D_{\text{m}} \tag{2.6}$$

where $V(D_{\mathrm{m}})$ is the geometric volume of the ice crystal and $\langle S(D_{\mathrm{m}})\rangle$ is the orientationally averaged geometrical cross-section of ice crystals in a unit volume of a cloud. For monodisperse spheres, $D_{\mathrm{e}}$ equates to their diameters.

It follows for large convex ice crystals in random orientation (Kokhanovsky, 2004),

$$K_{\mathrm{ext}} = 3f/D_{\mathrm{e}} \tag{2.6a}$$

where $f = IWC/\rho$ is the volumetric concentration of ice crystals, $\rho$ is the density of ice. This confirms that Eq. (2.6) is a useful parameter to characterize size/shape distributions with respect to calculations of light extinction in cirrus. How universal is Eq. (2.6a) for calculations of $K_{\mathrm{ext}}$? Can this concept be applied at all wavelengths? In the papers by Mitchell (2002) and Baran (2005) it is shown that the concept breaks down at infrared wavelengths. This is demonstrated in Fig. 2.8 where the mass extinction coefficient ($K_{\mathrm{ext}}$/IWC) for 30 PSDs is plotted as a function of $D_{\mathrm{e}}$ for six wavelengths in the infrared. The figure from Baran (2005) shows that the mass extinction coefficient for the shorter wavelengths is still inversely proportional to $D_{\mathrm{e}}$ as given by Eq. (2.6a), but as the wavelength increases this relationship begins to break down. At wavelengths between 20 µm to 30 µm the concept outlined above (see Eqs. (2.6) and (2.6a)) cannot be generally applied to compute the radiative properties of cirrus. Eq. (2.6a) only has physical meaning when the ice crystal size is much larger than the incident wavelength; this is not surprising since it is fundamentally based on the principle of geometric optics. It should be pointed out here that the simple optical parameter – effective diameter relationship does not hold for the asymmetry parameter since this fundamentally depends on the shape of ice crystals as demonstrated by Kokhanovsky and Macke (1997) and Wyser and Yang (1998).

Given the geometric ice crystal realizations described in section 2.2 and the means to compute their single scattering properties the next section describes how these ice crystal models are tested using remote sensing.

## 2.4 Airborne and satellite remote sensing of cirrus at solar and infrared wavelengths

As pointed out in section 2.3 the two most useful quantities to use from Eq. (2.1) in the remote sensing of cirrus are the phase function and the degree of linear polarization. Calculations of the phase function and the ratio $P_{12}/P_{11}$ are shown in Fig. 2.9 and Fig. 2.10, respectively. The calculations assume a bullet-rosette (Fig. 2.3 (d)) and a distorted bullet-rosette each having a maximum dimension of 100 µm using a complex refractive index for ice taken from Warren (1984) at the wavelength of 0.865 µm. The reason why distortion is applied in the calculations shown in Figs. 2.9 and 2.10 to the ice crystal geometry is to randomize the ice crystal such that the symmetry properties that are responsible for optical features such as halos are removed, thereby producing a featureless phase function. The method of ray-tracing due to Macke et al. (1996) was applied to the bullet-rosette geometry to calculate the scattering matrix elements. Also,

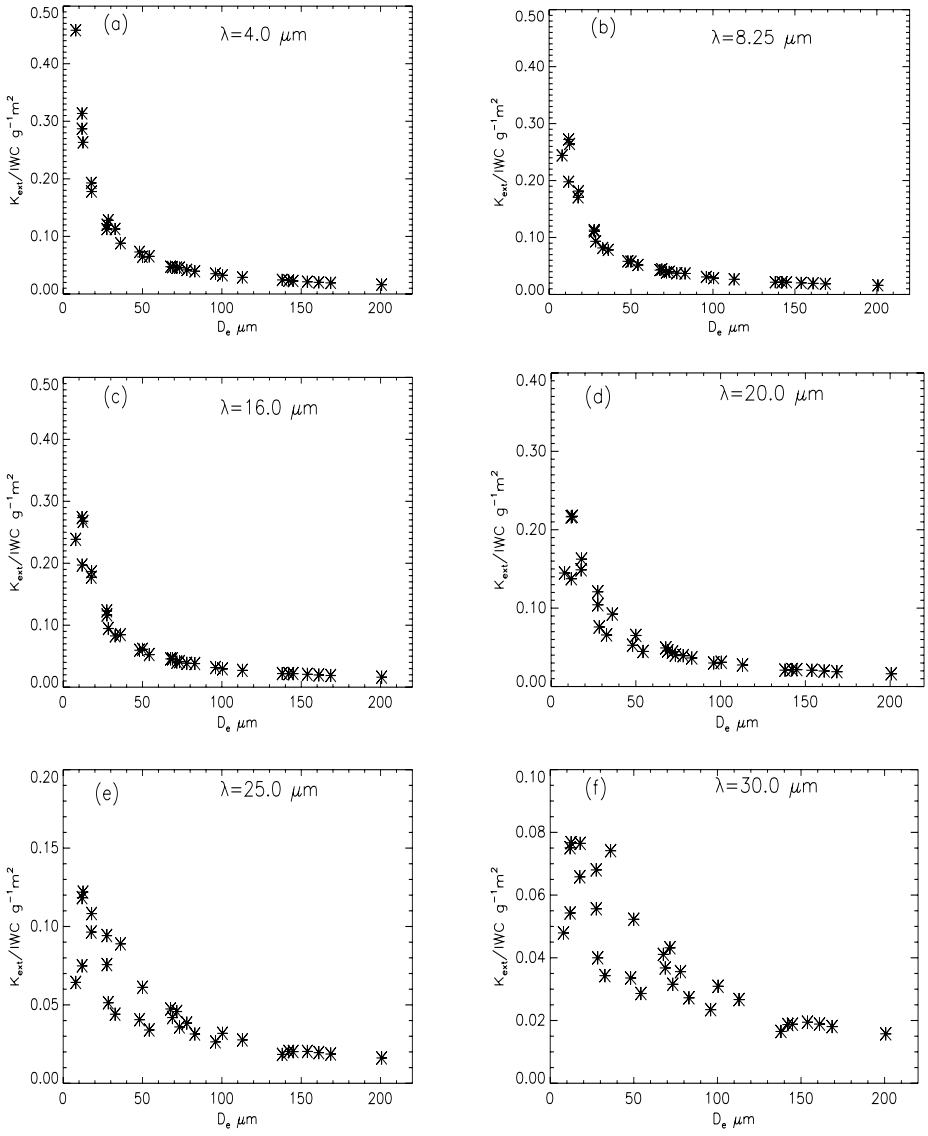



**Fig. 2.8.** The mass extinction coefficient, $K_{\rm ext}$/IWC in units of $\rm g^{-1}m^2$, plotted as a function of effective diameter, $D_{\rm e}$, at the wavelengths of: (a) 4.0 µm, (b) 8.25 µm, (c) 16.0 µm, (d) 20.0 µm, (e) 25.0 µm, (f) 30.0 µm. The mass extinction coefficient has been calculated using size distribution functions from Fu (1996) and Mitchell et al. (1996) (reproduced, with permission, from Baran, 2005).

shown in Fig. 2.9 for comparison is the analytic phase function due to Baran et al. (2001b). This function is defined by different expressions depending on the asymmetry parameter. In particular, it follows

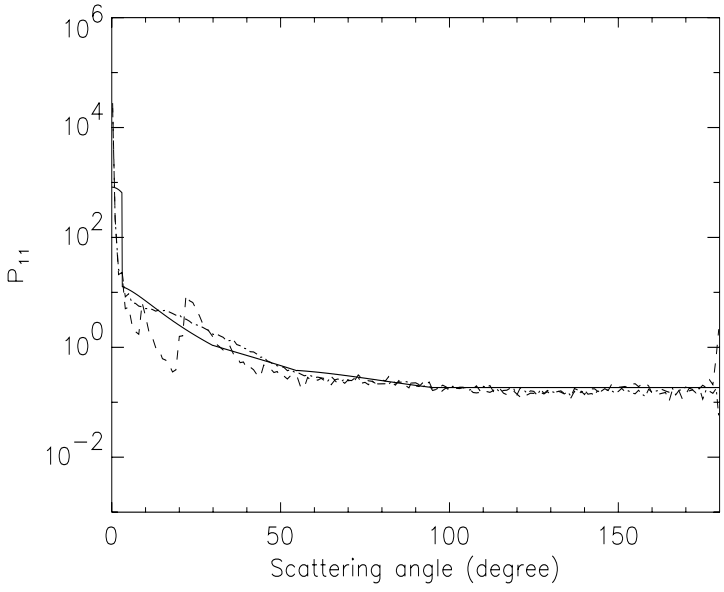



**Fig. 2.9.** The phase function plotted as a function of scattering angle assuming a randomly oriented six-branched bullet-rosette shown as the dashed line. The phase function of the randomized six-branch bullet-rosette is shown as the dashed-dotted line and the full line represents the analytic phase function calculated assuming an asymmetry parameter value of 0.78. The six-branched bullet-rosette is assumed to have a size of 100 µm and the incident wavelength is assumed to be 0.865 µm with an associated complex refractive index for ice of $1.304 + \mathrm{i}2.40 \times 10^{-7}$ (reproduced, with permission, from Baran and Labonnote, 2006).

$$P_{11} = \begin{cases} \dfrac{1-g^2}{(1+g^2-2g\cos\theta)^{3/2}}\alpha\cos\theta : \theta \leq 54.8^\circ \\ \dfrac{1-g^2}{(1+g^2-1.5g\cos\theta\sin\theta)^{3/2}} : \theta > 54.8^\circ \end{cases} \tag{2.7}$$

at $g < 0.7$ and

$$P_{11} = \begin{cases} \dfrac{1-g^2}{(1+g^2-2g\cos\theta)^{3/2}}\alpha\cos^{128}\theta : \theta \leq 3^\circ \\ \dfrac{1-g^2}{1+g^2-2g\cos\theta(1.3\theta)^{1.2}}\cos\theta : 3^\circ < \theta \leq 30^\circ \\ \dfrac{1-g^2}{1+g^2-2g\cos\theta(\Delta\theta)^{\sigma}}\cos\theta : 30^\circ < \theta \leq 54.8^\circ \\ \dfrac{1-g^2}{(1+g^2-1.5g\cos\theta\sin\theta)^{3/2}} : 54.8^\circ < \theta \leq 95^\circ \\ P_{11} = 95^\circ : \theta > 95^\circ \end{cases} \tag{2.8}$$

**Table 2.1.** Values of the coefficients $A$, $B$, $C$, and $\sigma$ for various values of the asymmetry parameter, $g$

| $g$ | $A$ | $B$ | $C$ | $\sigma$ |
|---|---|---|---|---|
| $0.70 \leq g \leq 0.80$ | 148.1 | 202.5 | 49.49 | 0.68 |
| $0.80 \leq g \leq 0.90$ | 277.1 | 510.2 | 232.9 | 0.68 |
| $g \geq 0.90$ | 421.9 | 827.1 | 406.3 | 0.71 |

**Table 2.2.** Values of the coefficient $\beta$ for various values of the asymmetry parameter, $g$

| $g$ | $\beta$ |
|---|---|
| $g < 0.30$ | 1.25 |
| $0.30 \leq g < 0.45$ | 1.50 |
| $0.45 \leq g < 0.60$ | 1.23 |
| $0.60 \leq g < 0.70$ | 1.095 |

at $g \geq 0.70$; Here $\sigma$ is given in Table 2.1 and

$$\Delta = \left(\frac{1-g}{4.6}\right) + g \tag{2.9}$$

The value of $\alpha$ is given by the following equations for the values of $g$ shown,

$$\alpha = \begin{cases} \dfrac{\beta}{\sqrt{1-g}} : g < 0.3 \\ \dfrac{1}{\sqrt{\beta g}} : 0.3 \leq g < 0.7 \\ \dfrac{N}{\sqrt{g}} : g \geq 0.7 \end{cases} \tag{2.10}$$

In Eq. (2.10), $N$ is a polynomial fit to the asymmetry parameter to ensure that $P_{11}$ is correctly normalized to $4\pi$, and $N = A - Bg + Cg^2$. The values for each of the coefficients $A$, $B$, $C$ and $\sigma$ for various ranges of $g$ are given in Table 2.1 and values for $\beta$ are given in Table 2.2.

This is a featureless phase function modelled on a laboratory phase function obtained from an ensemble of nonspherical ice crystals (Volkovitskiy et al., 1980, referred to as the VPP phase function); it is a linear-piecewise parametrization of the Henyey–Greenstein phase function (Henyey and Greenstein, 1941) and is entirely generated by the asymmetry parameter.

As can be seen from Figs. 2.1 and 2.2, atmospheric ice crystals are not pristine. They may be distorted or roughened, or contain inclusions. All these processes would remove or reduce any angular features present on the scattering phase function or degree of linear polarization (Macke et al., 1996; Mishchenko and Macke, 1997; Yang and Liou, 1998; Labonnote et al., 2001; Sun et al., 2004; Ulanowski et al., 2006). Fig. 2.9 shows that with no distortion the bullet-rosette

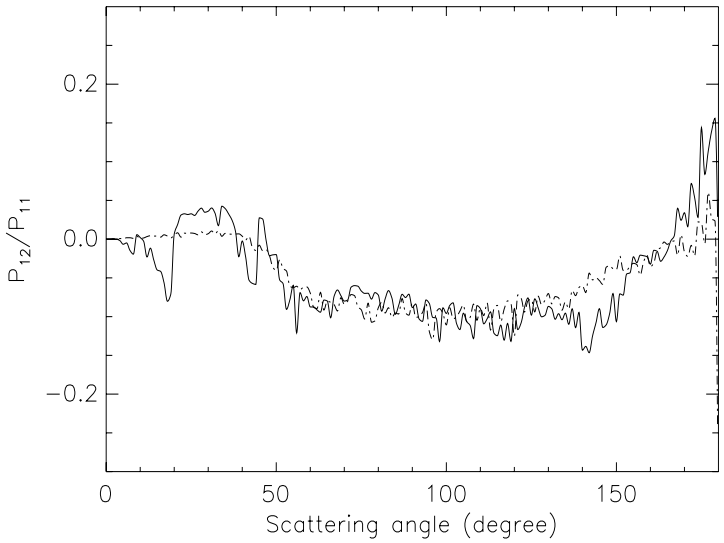



**Fig. 2.10.** Same as Fig. 2.9 but for the degree of linear polarization and the six-branched bullet-rosette and randomized six-branched bullet-rosette are represented by the full line and dashed-dotted line, respectively (reproduced, with permission, from Baran and Labonnote, 2006).

exhibits typical halo features present on the phase function at scattering angles of about 10°, 22°, and 42° with the 'ice bow' appearing at about 150° and retro-reflection peak at 180°. All pristine faceted ice crystals such as hexagonal columns, and hexagonal plates would also exhibit typical halos and enhanced backscattering intensities (Borovoi et al., 2000). The distorted bullet-rosette appears featureless with distinct halos and backscattering enhanced intensities removed and by a scattering angle of 50° matches the analytic phase function. The angular features present in the undistorted bullet-rosette phase function are reflected in the degree of linear polarization shown in Fig. 2.10. Therefore, Fig. 2.9 and Fig. 2.10 suggest that remote sensing instrumentation may be used to test whether cirrus is chiefly composed of pristine faceted ice crystals or more complex particle shapes. The rest of this chapter concentrates on how the idealized model geometries presented in section 2.2 and their predicted single scattering properties described in section 2.3 can be tested using aircraft and satellite data.

### 2.4.1 Airborne remote sensing of cirrus

To test ice crystal model predictions of the scattering phase function multi-angular radiometric measurements are required of cirrus from below and above the cloud. In the paper by Foot (1988) an airborne-based method of testing model phase functions is described where the aircraft flies in an orbit below or above the cirrus at a particular banking angle, at a distance of 1 km or sev-

eral km from the cirrus base or top. With the solar zenith angle fixed, and the azimuthal angle varying, the scattered radiance from the same section of cirrus is measured as the orbit is completed, thereby describing the scattered radiance, in principle, between the scattering angles of about 5° to 180°, depending on banking angle and solar geometry. An example of measuring the phase function of cirrus using this airborne technique is shown in Fig. 2.11 from Baran et al. (2001b, Fig. 4). The figure is derived from aircraft-measured multi-angle radiance measurements obtained at the wavelength of 0.87 µm from the aircraft orbit below the cirrus on the 9th November 1995 off the north-east coast of England. The solar zenith angle at the time of the orbit was measured to be 74° and the aircraft was banked at an angle of about 53°, which enables the phase function to be sampled between the scattering angles of 21° to 127° relative to the Sun. Each set of results shown in Fig. 2.11 were offset by a factor 10 to aid clarity. The model ice crystals assumed in Fig. 2.11 are the small hexagonal ice column, small six-branched bullet-rosette, large six-branched bullet-rosette, small hexagonal ice aggregate, large hexagonal ice aggregate, the VPP and analytic phase functions. The single scattering properties for each of the ice crystal models were obtained from the method of improved geometric optics due to Yang and Liou (1996) and are listed in Table 2.3 in the form of $D_e$, $\omega_0$ and $g$. The asymmetry parameter value assumed to generate the analytic phase function in Fig. 2.11 was 0.80 and the VPP phase function asymmetry parameter was estimated to be 0.85 by Foot (1988). The measured solar radiances were simulated using a Monte Carlo multiple-scattering model due to Kite (1987), with different values of extinction optical thickness, $\tau_{ext}$, $D_e$, and solar zenith angle as input parameters. The figure shows that single model phase functions representing hexagonal ice columns, hexagonal ice plates, bullet-rosettes or the hexagonal ice aggregate do not describe the measured angular radiometric data well between the scattering angles of 20° to about 125° at the wavelength of 0.87 µm. However, phase functions such as the Volkovitskiy et al. (1980) or analytic that represent scattering from an ensemble collection of nonspherical ice crystals do represent the measured angular radiometric data well. This finding is consistent with the result found by Foot (1988) and Francis et al. (1999). The papers by Francis et al. (1999) and Baran et al. (2001b) demonstrate that phase functions,

**Table 2.3.** Single-ice-crystal models assuming the small hexagonal ice column (Small column), small six-branched bullet-rosette (Small bullet/rosette), large six-branched bullet-rosette (Large bullet/rosette), small hexagonal ice aggregate (Small aggregate), and large hexagonal ice aggregate (Large aggregate)

| Ice crystal model | $D_e$ (µm) | $\omega_0$ | $g$ |
|---|---|---|---|
| Small column | 9.3 | 1.0 | 0.67 |
| Small bullet/rosette | 4.0 | 1.0 | 0.73 |
| Large bullet/rosette | 79.0 | 1.0 | 0.83 |
| Small aggregate | 5.30 | 1.0 | 0.76 |
| Large aggregate | 134.0 | 1.0 | 0.77 |

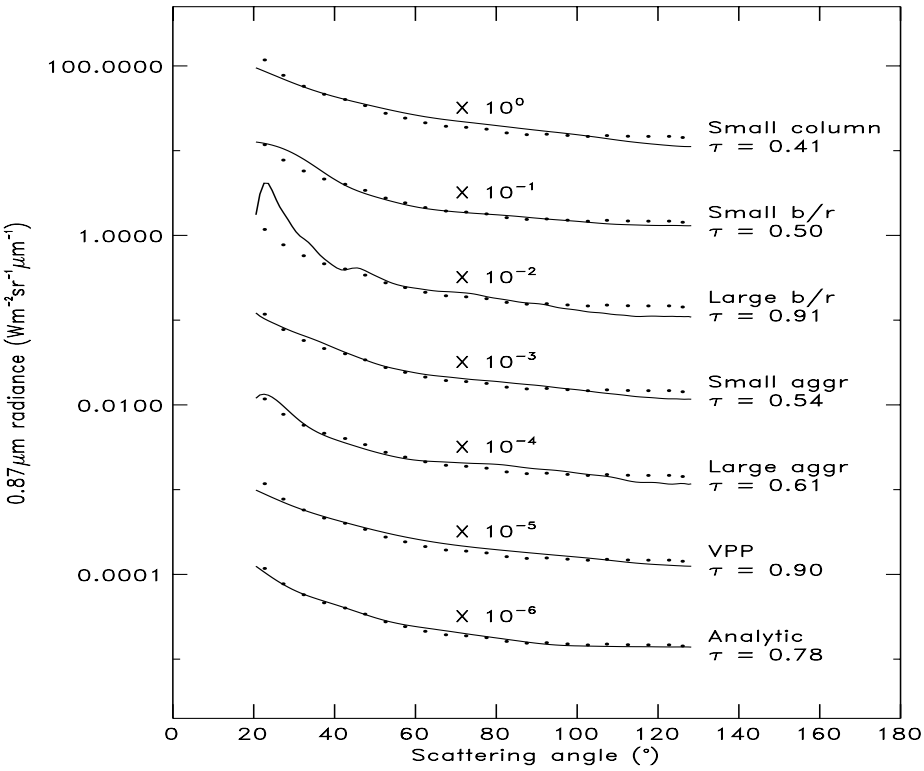



**Fig. 2.11.** The measured transmitted 0.87 μm radiance plotted against scattering angle with the filled circles representing the airborne radiance measurements and the full lines representing the predicted intensity assuming various ice crystal models. The model ice crystals assumed were the small hexagonal ice column, Small b/r (small bullet-rosette), Large b/r (large bullet-rosette), Small aggr (hexagonal aggregate), Large aggr (hexagonal aggregate), the Volkovitskiy et al. 1980 and analytic phase functions, respectively. The values on the right-hand side of the figure are the optimally derived optical thickness, $\tau$, derived for each of the assumed ice crystal models. The mean experimental value found for $\tau$ was $\tau = 0.75 \pm 0.08$ (reproduced, with permission, from Baran et al., 2001).

such as the analytic, representing an ensemble of ice crystals rather than single ice crystals best represent multi-angular radiometric data obtained from below cirrus.

The infrared radiative properties of cirrus are also very important when parametrizing cirrus for climate models as demonstrated by Edwards et al. (2007). Recently, simultaneous airborne high-resolution measurements of cirrus at solar and infrared wavelengths have become available and examples of such measurements are shown in Baran and Francis (2004). In that paper the high-resolution radiometric measurements were obtained in eleven sections above a

piece of semi-transparent cirrus located north of Scotland during October 2000. Since both solar and infrared measurements were made simultaneously the optical thickness above the cloud was retrieved at the wavelengths of 0.87 µm and 11.0 µm assuming hexagonal ice columns and randomized hexagonal ice aggregates. To simulate the solar and infrared high-resolution measurements and retrieve the optical thickness the radiative transfer model used was due to Edwards and Slingo (1996) assuming a plane-parallel cloud, which has been extended to radiance space by using the spherical harmonic method. This method fully takes into account the strong forward scattering peak of the ice crystal phase function (Ringer et al., 2003). The angular distribution of the radiance is decomposed into a series of spherical harmonics, with the order at which the infinite series is truncated determining the accuracy of the calculated radiance. For the radiance calculations presented in this chapter the truncation of the direct radiance has been set to 399, with the diffuse truncation being set to 21. For the infrared calculations, the diffuse truncation has been set to 19. Applying this radiative transfer model to the data the $D_e$ (see Eq. (2.6)) values found for the hexagonal ice column that best fits both the solar and infrared high-resolution measurements were 67 µm and 87 µm. The aspect ratio of the hexagonal ice columns is based on the tabulations from Mitchell and Arnott (1994). The hexagonal ice aggregate is randomized by roughening the mantle surfaces as described in Yang and Liou (1998) and the best-fit $D_e$ value found for this model was 78 µm. Figure 2.12 shows the retrieved optical thickness for all eleven sections assuming the two model ice crystals. In the case of the hexagonal ice column consistency in the retrieved optical thickness could not be found for all eleven sections. However, for the randomized ice aggregate consistency was found for all eleven sections. The figure demonstrates that for this case the radiative properties of the cirrus were best represented by complex randomized ice crystals. An example of high-resolution infrared measurements obtained from the optically thinnest section is shown in Fig. 2.13 (from Baran and Francis, 2004, Fig. 9c). Figure 2.13 shows brightness temperature differences between model and measurements; the dotted line shows the scene variability (i.e., the cloud was not uniform) in the high-resolution measurements. The figure shows that the cirrus radiative properties between the wavelengths of 3.3 µm and 16.0 µm are well predicted within the bounds of the scene variability for this particular case. In the case of airborne remote sensing of cirrus the generality of testing the predictions of phase functions or single scattering properties by assuming some ice crystal model is limited to relatively few cases. For this reason it is important to obtain space-based measurements of cirrus through the use of satellites, which are able to sample cirrus over many cases.

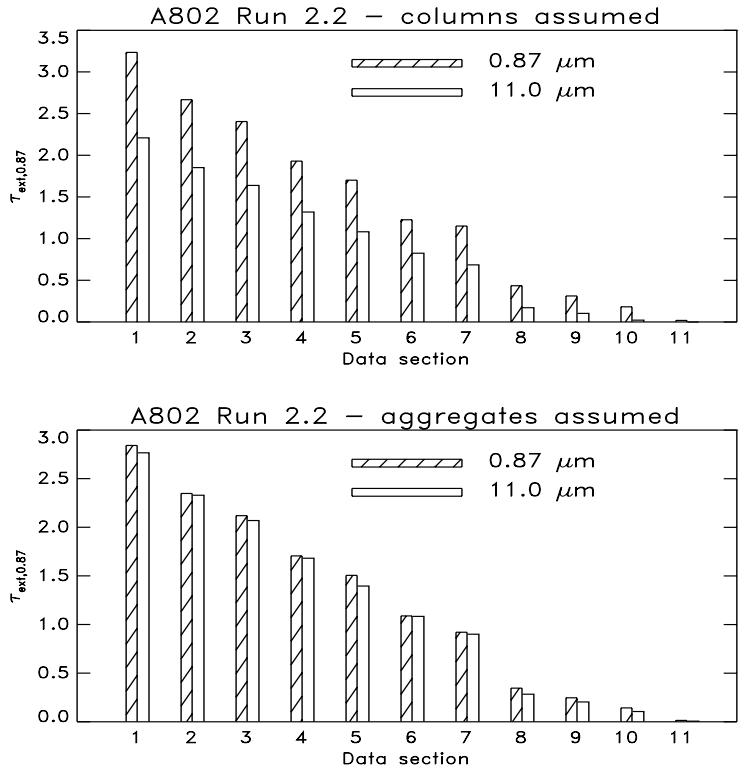



**Fig. 2.12.** Comparison of the best-fit optical thickness (referenced to $\tau_{\text{ext}}$ at 0.87 μm), retrieved from the airborne radiance measurements at 0.87 μm and infrared measurements at 11.0 μm, assuming (a) pristine hexagonal ice columns and (b) hexagonal ice aggregates (reproduced, with permission, from Baran and Francis, 2004).

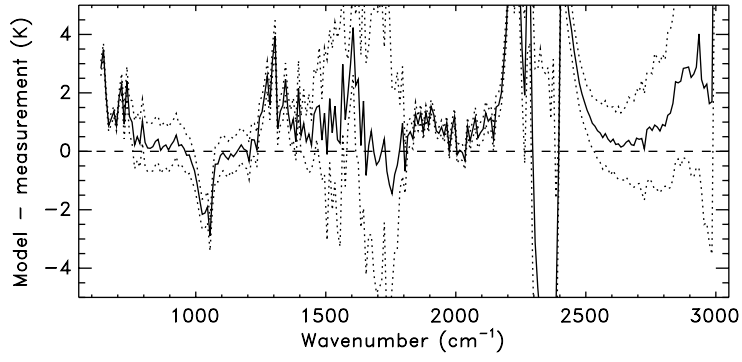



**Fig. 2.13.** An example of high-resolution radiometric data showing brightness temperature differences plotted against wavenumber in units of $\text{cm}^{-1}$ ($\nu = 10000/\lambda$; so $\nu = 1000\,\text{cm}^{-1}$ corresponds to $\lambda = 10.0\,\mu\text{m}$) between simulations assuming the hexagonal ice aggregate model and the measurements. The dotted line in the figure represents the scene variability (after Baran and Francis, 2004).

### 2.4.2 Satellite remote sensing of cirrus

So that the single scattering properties of model ice crystals can be tested it is necessary to have space-based instruments which are able to sample the model phase function and/or polarization properties at a number of scattering angles. Currently, there are three satellites that are capable of testing model phase functions. One such instrument is the Along Track Scanning Radiometer (ATSR-2) described in Baran et al. (1999). This instrument is dual-viewing and has been used to infer ice crystal shape at non-absorbing and absorbing wavelengths (Baran et al., 1999, 2003). More recently, McFarlane et al. (2005) have made use of an instrument called the Multiangle Imaging Spectroradiometer (MISR) to infer ice crystal habit. Combining MISR with the Moderate Resolution Imaging Spectroradiometer instrument (MODIS) McFarlane et al. (2005) are able not only to estimate ice crystal shape but also to retrieve ice crystal size, since both instruments are located on the same satellite. The MISR instrument measures at solar wavelengths and has up to nine viewing angles, whilst the single-view MODIS instrument has its channels located at non-absorbing and absorbing wavelengths, which makes the retrieval of ice crystal size possible (Baum et al., 2000). The third instrument that can be utilized to study the angular reflection properties of cirrus is called the Polarization and Directionality of the Earth's Reflectances (POLDER) and a description of this instrument can be found in Buriez et al. (1997). The unique feature of POLDER is that it measures not only light reflection from cirrus but also the polarized reflectance defined as the ratio between the normalized polarized radiance and the solar zenith angle. The POLDER instrument can measure cirrus reflection function and polarization properties at up to 14 different viewing directions and can sample the phase function and the degree of linear polarization between the scattering angles from 60° to 180° dependent on latitudinal position as described in the paper by Labonnote et al. (2001). This simultaneous combination of measurements in both reflection and polarization space is very important for inferring information about the complexity of cirrus particle habits, as demonstrated by Baran and Labonnote (2006). The importance of this combination in terms of reflection is shown in Fig. 2.14 (taken from Baran and Labonnote (2006, Fig. 6), which shows the POLDER measured spherical albedo (SA) differences (i.e., measurements – model) for a variety of randomized ice crystal models plotted as a function of scattering angle. The methodology of utilizing SA to test ice crystal model phase functions has been previously given in Labonnote et al. (2001), but a brief description of this approach is given here. To compute SA the cloud bi-directional reflection, $\mathrm{R}(\mu, \mu_0, \phi\text{-}\phi_0)$, is found by:

$$\mathrm{R}(\mu, \mu_0, \phi\text{-}\phi_0) = \pi \mathrm{I}(\mu, \mu_0, \phi\text{-}\phi_0)/\mu_0 \mathrm{E}_0 \tag{2.11}$$

where $\mathrm{I}(\mu, \mu_0, \phi\text{-}\phi_0)$ is the reflected solar radiance from cloud-top, $\mathrm{E}_0$ is the incident solar flux density, $\mu$ and $\mu_0$ are cosines of the zenith view and solar zenith angles, respectively. The relative azimuth is given by $\phi\text{-}\phi_0$.

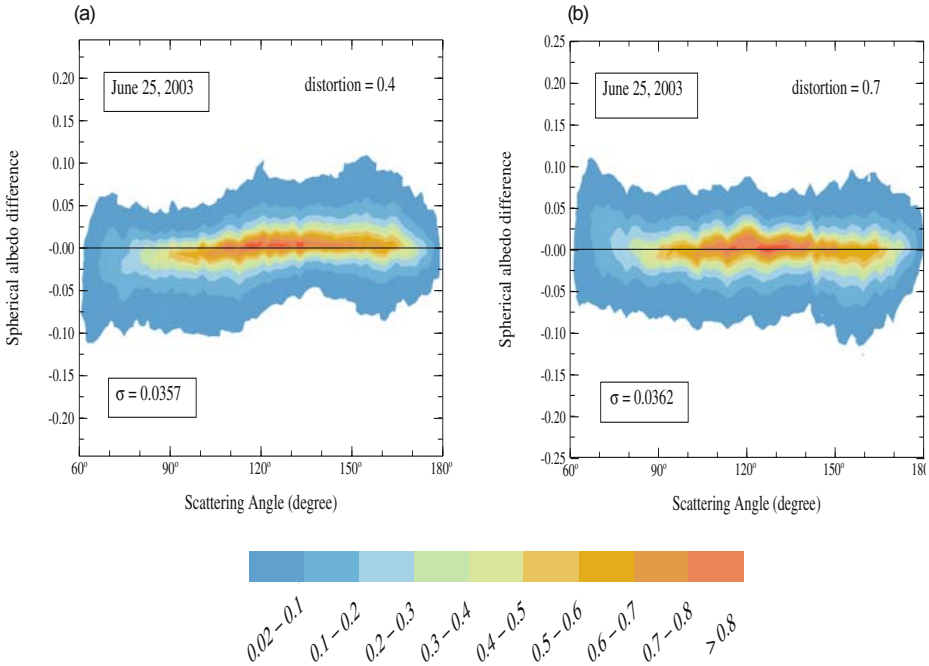



**Fig. 2.14.** Normalized density of selected pixel directions (the red colour represents more than 80% of the pixels) against scattering angle, showing differences between the retrieved spherical albedo obtained at the wavelength of 0.87 μm and the directionally averaged spherical albedo for the (a) chain-like hexagonal aggregate (Fig. 2.3 (g)) and (b) hexagonal ice aggregate (Fig. 2.3 (e)]. The assumed ice crystal parameters used in the calculations are given in Table 2.4. The degree of randomization is shown in the upper right-hand side of the figures and the standard deviation of the residual spherical albedo represented by $\sigma$ is also shown in the figures (reproduced, with permission, from Baran and Labonnote, 2006).

**Table 2.4.** The physical dimensions of each ice crystal model assumed in the POLDER radiative transfer calculations

| Model | Maximal dimension, μm |
|---|---|
| Six-branched bullet-rosette | 100 |
| Chain-like aggregate | 100 |
| Ice aggregate | 100 |
| Polycrystal | 100 |
| IHM | 220 |

Eq. (2.11) can be integrated over all $\mu$ and $\phi$-$\phi_0$ to give the plane albedo, $\mathrm{A}(\mu_0)$, given by

$$A(\mu_0) = \frac{1}{\pi} \iint \mathrm{R}(\mu, \mu_0, \phi\text{-}\phi_0)\mu \, \mathrm{d}\mu \, \mathrm{d}\phi \tag{2.12}$$

and Eq. (2.12) can be integrated over all solar zenith angles to give the cloud spherical albedo (SA) $a$, given by

$$a = 2 \int \mathrm{A}(\mu_0)\mu_0 \, \mathrm{d}\mu_0 \tag{2.13}$$

The measurements of bi-directional reflectance at the wavelength of 0.670 µm are used to retrieve the cloud optical thickness in each viewing direction, which, assuming a black underlying surface, is equivalent to SA (see Doutriaux-Boucher et al., 2000; Labonnote et al., 2001). The measured SAs are simulated using a radiative transfer model based on the discrete-ordinates method due to Stamnes et al. (1988). The radiative transfer model assumes a homogeneous plane-parallel cloud and uses as input the satellite–Sun geometry, assumed ice crystal model phase function, the optical thickness, and the single scattering albedo, which is unity. If the phase function model were a perfect representation of scattering from cirrus then the retrieved SA would be independent of scattering angle. It is this aspect that Fig. 2.14 is testing.

The POLDER data shown in the figure was obtained during one day on the 25th June 2003 and the POLDER pixels were globally distributed and only pixels located over the sea were included. As can be seen from the figure both the randomized chain-like aggregate (Fig. 2.3 (g)) and randomized hexagonal ice aggregate (Fig. 2.3 (e)) do minimize the POLDER spherical albedo measurements well, with the standard deviations for each of the ice crystal models appearing quite similar. The reason for this similarity is that if the ice crystals are sufficiently randomized then the phase functions appear featureless, as shown in Fig. 2.9. However, the geometrical forms of each of the ice crystal models are very different, as shown in Fig. 2.3, but using reflection measurements alone is not sufficient to distinguish which of the randomized ice crystals best explains the POLDER measurements. However, the POLDER instrument also measures the polarized reflectance and, as shown in Fig. 2.10, polarization properties depend strongly on ice crystal shape. Perhaps this measurement can be used to distinguish between different types of randomized ice crystal? The results for the measured polarized reflectance assuming a variety of ice crystal models described in Fig. 2.3 are shown in Fig. 2.15 (a) and 2.15 (b) (from Baran and Labonnote (2006, Figs. 7a and Fig. 7b). Figure 2.15 (a) shows the POLDER-measured polarized reflectance plotted as a function of scattering angle with each of the ice crystal models with and without randomization represented by the various lines shown in the figure. The figure shows that when randomization is extreme the fit to the polarized reflection becomes worse; this is evident for the hexagonal ice aggregate and polycrystal models. However, the hexagonal ice aggregate did minimize the spherical albedo differences with such a randomization but this same model does not describe the POLDER-measured polarized

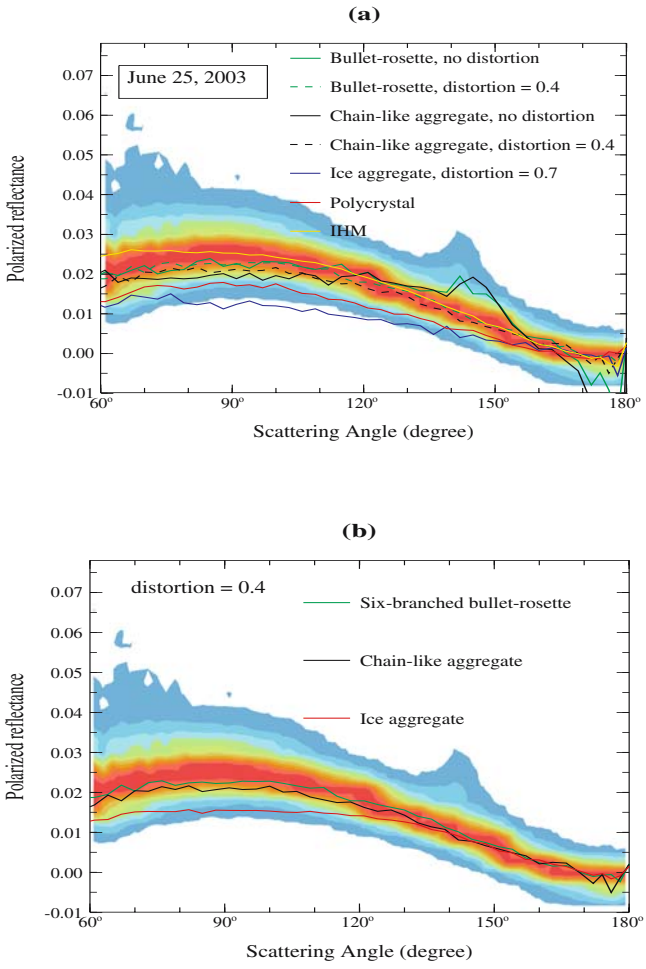



**Fig. 2.15.** (a) Same as Fig. 2.14 but for the polarized reflectance plotted against scattering angle for a variety of ice crystal models. The ice crystal parameters assumed in the calculations are given in Table 2.4. The ice crystals models are shown on the right-hand side of the figure together with the degree of distortion. (b) Same as (a) but for a single randomization showing polarization results for the six-branched bullet-rosette (green line), chain-like aggregate (black line), and the hexagonal ice aggregate (red line) (reproduced, with permission, from Baran and Labonnote, 2006).

reflectance. Therefore, in order to eliminate such models, intensity measurements alone are insufficient and additional information such as polarized reflectance is required.

Fig. 2.15 (b) shows that more spatial ice crystals such as the randomized chain-like aggregate and randomized bullet-rosette do satisfactorily explain the measured polarized reflectance, whilst more compact ice crystals such as the hexagonal ice aggregate do not. Again the geometrical form of the bullet-rosette

and chain-like aggregate are very different (see Fig. 2.3), but not even a combination of reflection and polarized reflection can distinguish between them. The reason for this is that if sufficient randomization is applied to ice crystal models then all elements of the scattering matrix become featureless thus making distinction between more complex habits problematic. In order to distinguish between more complex randomized ice crystal models, further information is required, which could take the form of two-dimensional scattering patterns described by Ulanowski et al. (2006) or an enhanced version of the CPI. The utility of using the first three Stokes parameters ($I$, $Q$, and $U$) in the remote sensing of cirrus is further demonstrated by Ou et al. (2005). They show that by using simulated measurements at the wavelengths of 0.865 µm and 2.25 µm there is sensitivity to ice crystal shape, size and surface roughness.

## 2.5 Summary

This chapter has reviewed the current understanding of the optical and radiative properties of cirrus and it has demonstrated the importance of this cloud to climate modelling and remote sensing. The populations of nonspherical ice crystals that exist in cirrus are diverse; however, there is now sufficient evidence to say that pristine ice crystals such as hexagonal ice columns and hexagonal ice plates are uncommon. The most common nonspherical ice crystal type that inhabits synoptically generated cirrus is bullet-rosettes whilst anvil cirrus is mostly populated by non-symmetric irregulars. Representing these types of crystals by some geometric model such that the full single scattering properties can be solved is problematic. The current consensus appears to be that representing the variability of complex shapes by one single ice crystal model geometry does not appear to be supported either by *in situ* measurements of the IWC or airborne remote sensing of the scattering phase function. Representing the actual diversity of shapes by some ensemble model of ice crystal shapes, which are individually randomized, and that ensemble is able to replicate the measured IWC to a reasonable degree of accuracy is the better way forward. This approach reconciles the single scattering properties of the ensemble with macroscopic quantities such as IWC for any given particle size distribution function. This link between the cirrus single scattering properties and the amount of ice mass or IWC is the fundamental problem to be solved.

In recent years there have been significant advances in the development of electromagnetic methods to solve the single scattering properties of nonspherical ice crystals. This is especially true for the T-matrix and FDTD methods; however, there is still no one method that can solve the complete light scattering problem over the entire cirrus particle size distribution function. There is still reliance on approximations such as physical and geometric optics to bridge the gap between the so-called 'exact' methods and approximations. Though there are methods that can in principle be applied to any ice crystal shape given appropriate computational resources. The scattering properties of cirrus appear to be best represented by phase functions which are smooth and featureless; this

is also true of the other elements of the scattering matrix. The reason for this featureless nature of the scattering phase function is due to the most common types of ice crystals having a non-symmetric form which may also be distorted or roughened, and/or contain inclusions of air or aerosol. In essence, the complete scattering properties from cirrus ice crystal ensembles are of a simple functional form, what is required is a computational method that can reproduce this simplicity from a given ensemble.

In order to solve the cirrus problem as outlined above, further airborne field campaigns are certainly required that are able to further quantify the cirrus particle size distribution function, and especially the role of small ice crystals less than 100 μm in size, and the most common geometrical form of these small ice crystals. Quantification of ice crystal shapes from different locations, heights and seasons, and instrumentation to measure their masses is also required. As regards airborne remote sensing, further measurements of the scattering phase function at non-absorbing and absorbing wavelengths over a wide range of scattering angle are needed. The development of high-resolution spectrometers that are able to measure cirrus radiances at both solar and infrared wavelengths simultaneously should enable rigorous testing of the modelled cirrus single scattering properties. As this chapter has shown, combining intensity measurements with polarization measurements is important when trying to distinguish between complex cirrus ice crystal models. However, as shown in this chapter, due to the simplicity of the scattering matrix elements for complex randomized ice crystals with very different geometrical forms, using both intensity and polarization to distinguish between such ice crystals is still problematic. For instance, it follows from Fig. 2.14 (a) (or Fig. 2.6 of Baran and Labonnote, 2006), and Fig. 2.15 that both the distorted chain-like aggregates (Fig. 2.3 (g)) and six-branched bullet-rosettes (Fig. 2.3 (d)) well represent the radiative and polarization properties of cirrus. These shapes are also similar to those shown in Fig. 2.1. Certainly, the use of intensity alone measurements is insufficient. Distinguishing between complex non-symmetric ice crystals might be achieved by using 2D scattering patterns or an enhanced version of the CPI. This distinction is important since at infrared wavelengths the absorption properties of different non-symmetric ice crystals will differ and will therefore not have the same radiative responses in climate models.

## Acknowledgements

The author would like to thank A. J. Heymsfield for providing Fig. 2.1 and Fig. 2.2, P. N. Francis for providing the aircraft data, and C. Westbrook for providing Fig. 2.4. The UK Department of the Environment, Fisheries and Rural Affairs is also acknowledged for part supporting some of the work presented in this chapter.

## References


Asano, S., and M. Sato, 1980: Light scattering by randomly oriented spheroidal particles, *Appl. Opt.*, **19**, 962–974.

Baran, A. J., P. D. Watts, and P. N. Francis, 1999: Testing the coherence of cirrus microphysical and bulk properties retrieved from dual-viewing multispectral satellite radiance measurements, *J. Geophys. Res.*, **104**, 31673–31683.

Baran, A. J., P. Yang, and S. Havemann, 2001a: Calculation of the single scattering properties of randomly oriented hexagonal ice columns: a comparison of the T-matrix and the finite-difference time-domain methods, Appl. Opt. **40**, 4376–4386.

Baran, A. J., P. N. Francis, L.-C. Labonnote, and M. Doutriaux-Boucher, 2001b: A scattering phase function for ice cloud: Tests of applicability using aircraft and satellite multi-angle multi-wavelength radiance measurements of cirrus, *Q. J. R. Meteorol. Soc.*, **127**, 2395–21416.

Baran, A. J., S. Havemann, P. N. Francis, and P. D. Watts, 2003: A consistent set of single scattering properties for cirrus cloud: A test using radiance measurements from a dual-viewing multi-wavelength satellite-based instrument, *J. Quant. Spectr. Rad. Transfer*, **79–80**, 549–567.

Baran, A. J., 2003a: On the scattering and absorption properties of cirrus cloud, *J. Quant. Spectr. Rad. Transfer*, **89**, 17–36.

Baran, A. J., 2003b: Simulation of infrared scattering from ice aggregates by use of a size-shape distribution of circular ice cylinders, *Appl. Opt.*, **42**, 2811-2818.

Baran, A. J., and S. Havemann, 2004: The dependence of retrieved ice-crystal effective dimension on assumed ice crystal geometry and size-distribution function at solar wavelengths, *Q. J. R. Meteorol. Soc.*, **130**, 2153–2167.

Baran, A. J., and P. N. Francis, 2004: On the radiative properties of cirrus cloud at solar and thermal wavelengths: A test of model consistency using high-resolution airborne radiance measurements, *Q. J. R. Meteorol. Soc.*, **130**, 763–778.

Baran, A. J., 2005: The dependence of cirrus infrared radiative properties on ice crystal geometry and shape of the size-distribution function, *Q. J. R. Meteorol. Soc.*, **131**, 1129–1142.

Baran, A. J., V. N. Shcherbakov, B. A. Baker, J. F. Gayet and R. P. Lawson, 2005: On the scattering phase-function of non-symmetric ice crystals, *Q. J. R. Meteorol. Soc.*, **131**, 2609–2616.

Baran A. J., and L. C. Labonnote, 2006. On the reflection and polarization properties of ice cloud, *J. Quant. Spectrosc. Radiat. Transfer*, **100**, 41–54.

Baran, A. J., 2006: The suitability of assuming ice aggregates as a predictor of ice water content, *OBR Tech. Note. No. 56*, Met Office, Exeter.

Baum, B. A., D. P. Kratz, P. Yang, S. C. Ou, Y. X. Hu, P. F. Soulen, and S. C. Tsay, 2000: Remote sensing of cloud properties using MODIS airborne simulator imagery during SUCCESS 1. Data and models, J. Geophys. Res. **105**, 11767–11780.

Baum, B. A., A. J. Heymsfield, P. Yang, and S. T. Bedka, 2005: Bulk scattering properties for the remote sensing of ice clouds. Part I. Microphysical data and models, *J. App. Met.*, **44**, 1885–1895.

Borovoi, A., I. Grishin, E. Naats, and U. Oppel, 2000: Backscattering peak of hexagonal ice columns and plates, *Opt. Lett.*, **25**, 1388-1392.

Borovoi, A. G., and I. A. Grishin, 2003: Scattering matrices for large ice crystal particles, *JOSA*, **A 20**, 2071–2080.

Borovoi, A., A. Cohen, N. Kustova, and U. Oppel, 2005: Light scattering by large ice crystal particles, Proc. 8th *Conference on Electromagnetic and Light Scatter-*

*ing by Nonspherical Particles: Theory, Measurements and Applications*, Salobreña, Granada, Spain, May 16th–20th, 44–47.

Bréon, F.-M., and B. Dubrulle, 2004: Horizontally oriented plates in clouds, *J. Atmos. Sci.*, **61**, 2888–2898.

Bryant, F. D., and P. Latimer, 1969: Optical efficiencies of large particles of arbitrary shape and orientation, *J. Colloid Interface Sci.*, **30**, 291–304.

Buriez, J.-C., C. Vanbauce, F. Parol, P. Gouloub, M. Herman, B. Bonnel, Y. Fouquart, P. Couvert, and G. Séze, 1997: Cloud detection and derivation of cloud properties from POLDER, *Int. J. Remote. Sensing*, **18**, 2785-2813.

Cai, Q., and K. N. Liou, 1982: Polarized light scattering by hexagonal ice crystals: Theory, *Appl. Opt.*, **21**, 3569–3580.

Chepfer, H., G. Brogniez, P. Goloub, F. M. Bréon, and P. H. Flamant, 1999: Observations of horizontally oriented ice crystals in cirrus clouds with POLDER-1/ADEOS-1, *J. Quant. Spectr. Rad. Transfer*, **63**, 521–543.

Clarke, A. J. M., E. Hesse, Z. Ulanowski, and P. H. Kaye, 2006: A 3D implementation of ray tracing combined with diffraction on facets: Verification and a potential application, *J. Quant. Spectrosc. Radiat. Transfer*, **100**, 103–114.

Connolly, P. J., C. P. R. Saunders, M. W. Gallagher, K. N. Bower, M. J. Flynn, T. W. Choularton, J. Whiteway, and R. P. Lawson, 2004: Aircraft observations of the influence of electric fields on the aggregation of ice crystals, *Q. J. R. Meteorol. Soc.*, **128**, 1-1-9.

Donner, L., C. J. Seman, B. J. Soden, R. S. Hemler, J. C. Warren, J. Strom, and K. N. Liou, 1997: Large-scale ice clouds in the GFDL SKYHI general circulation model, *J. Geophys. Res.*, **102**, 460–470.

Doutriaux-Boucher, M., J.-C. Buriez, G. Brogniez, L.-C. Labonnote, and A. J. Baran, 2000: Sensitivity of retrieved POLDER directional cloud optical thickness to various ice particle models, Geophys. *Res. Lett.*, **27**, 109–112.

Edwards, J. M., and A. Slingo, 1996: Studies with a flexible new radiation code. 1. Choosing a configuration for a large-scale model, Q. J. R. Meteorol. Soc. **122**, 689–719.

Edwards, J. M., S. Havemann, J.-C. Thelen, and A. J. Baran, 2007: A new parametrization for the radiative properties of ice crystals: Comparison with existing schemes and impact in a GCM, *Atmos. Res.*, **83**, 19–34.

Field, P. R., A. J. Baran, P. H. Kaye, E. Hirst, and R. Greenway, 2003: A test of cirrus ice crystal scattering phase functions, *Geophys. Res. Lett.*, **30**, 1752, doi: 10.1029/2003GL017482.

Field, P. R., R. J. Hogan, P. R. A. Brown, A. J. Illingworth, T. W. Choularton, and R. J. Cotton, 2005: Parametrization of ice-particle size distribution functions for mid-latitude stratiform cloud, *Q. J. R. Meteorol. Soc.*, **131**, 1997–2017.

Foot, J., 1988: Some observations of the optical properties of clouds. II: Cirrus, *Q. J. R. Meteorol. Soc.*, **114**, 141–164.

Francis, P. N., A. Jones, R. W. Saunders, K. P. Shine, A. Slingo, and Z. Sun, 1994: An observational and theoretical study of the radiative properties of cirrus: Some results from ICE'89, *Q. J. R. Meteorol. Soc.*, **120**, 809–848.

Francis, P. N., J. S. Foot, and A. J. Baran, 1999: Aircraft measurements of the solar and infrared radiative properties of cirrus and their dependence on ice crystal shape, *J. Geophys. Res.*, **104**, 31685–31695.

Fu, Q., 1996: An accurate parameterization of the solar radiative properties of cirrus clouds for climate models, *J. Climate*, **9**, 2058–2082.

Fu, Q., W. B. Sun, and P. Yang, 1999: Modeling of scattering and absorption by nonspherical cirrus ice particles at thermal infrared wavelengths, *J. Atmos. Sci.*, **56**, 2937–2947.

Gallagher, M. W., J. Whiteway, M. J. Flynn, P. J. Connolly, D. Figueros-Nieto, T. W. Choularton, K. N. Bower, C. Cook, R. Busen, and J. Hacker, 2004: An overview of the microphysical structure of cirrus clouds observed during emerald-1, *Q. J. R. Meteorol. Soc.*, **131**, 1143–1169.

Garrett, T. J., H. Gerber, D. G. Baumgardner, C. H. Twohy, and E. M. Weinstock, 2003: Small, highly reflective ice crystals in low-latitude cirrus, *Geophys. Res. Lett.*, **30**, Art. No. 2132.

Garrett, T. J., B. C. Navarro, C. H. Twohy, E. J. Jensen, D. G. Baumgardner, P. T. Bui, H. Gerber, R. L. Herman, A. J. Heymsfield, P. Lawson, P. Minnis, L. Nguyen, M. Poellot, S. K. Pope, F. P. J. Valero, and E. M. Weinstock, 2005: Evolution of a Florida anvil, *J. Atmos. Sci.*, **62**, 2352–2372.

Gerber H., Y. Takano, T. J. Garrett, P. V. Hobbs, 2000: Nephelometer measurements of the asymmetry parameter, volume extinction coefficient, and backscatter ratio in Arctic clouds, *J. Atmos. Sci.*, **57**, 3021–3034.

Hartmann, D. L., M. E. Ockert-Bell, and M. L. Michelsen, 1992: The effect of cloud type on earth's energy balance: Global analysis, *J. Clim.*, **5**, 1281–1304.

Havemann, S., A. J. Baran, and J. M. Edwards, 2003: Implementation of the T-matrix method on a massively parallel machine: A comparison of hexagonal ice cylinder single scattering properties using the T-matrix and improved geometric optics methods, *J. Quant. Spectr. Rad. Transfer*, **79–80**, 707–720.

Henyey, L., and J. Greenstein, 1941: Diffuse radiation in the galaxy, *Astrophys. Journal*, **93**, 70–83.

Hesse, E., and Z. Ulanowski, 2003: Scattering from long prisms computed using ray tracing combined with diffraction on facets, *J. Quant. Spectr. Rad. Trans.*, **79–80**, 721–732.

Hesse, E., Z. Ulanowski, and S. Havemann, 2003: Scattering from long prisms: a comparison between ray tracing combined with diffraction on facets and SVM, In *Proc 7th Int. Conf. on Electromagnetic and Light Scattering bu Bon-spherical particles*, Bremen, 119–122.

Heymsfield, A. J., and C. M. R. Platt, 1984: A parametrization of the particle size spectrum of ice clouds in terms of the ambient temperature and ice water-content, *J. Atmos. Sci.*, **41**, 846–855.

Heymsfield, A. J., A. Bansemer, P. R. Field, S. L. Durden, J. L. Stith, J. E. Dye, W. Hall, C. A. Grainger, 2002: Observations and parameterizations of particle size distributions in deep tropical cirrus and stratiform precipitating clouds: Results from in situ observations in TRMM field campaigns, *J. Atmos. Sci.*, **59**, 3457–3491.

Heymsfield, A. J., and L. M. Miloshevich, 2003. Parametrizations for the cross-sectional area and extinction of cirrus and stratiform ice cloud particles, *J. Atmos. Sci.*, **60**, 936–956.

Hong, G., G. Heygster, and C. A. M. Rodriguez, 2006: Effect of cirrus clouds on the diurnal cycle of tropical deep convective clouds, *J. Geophys. Res.*, **111**, D06209, doi:10.1029/2005JD006208.

IPCC, 2001: Climate change 2001: The scientific basis. Third assessment report of the Intergovernmental Panel on Climate Change, Eds. J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. Van der Linden, X. Dai, K. Maskell and C. A. Johnson, Cambridge University Press, Cambridge and New York.

Ivanova, D., D. L. Mitchell, W. P. Arnott, and M. Poellot, 2001: A GCM parameterization for bimodal size spectra and ice mass removal rates in mid-latitude cirrus clouds, *Atmos. Res.*, **59**, 89–113.

Kahnert, F. M., J. J. Stamnes, and K. Stamnes, 2001: Application of the extended boundary condition method to particles with sharp edges: a comparison of two surface integration approaches, *Appl. Opt.*, **40**, 3101–3109.

Kahnert, F. M., J. J. Stamnes, and K. Stamnes, 2002a: Can simple particle shapes be used to model scalar optical properties of an ensemble of wavelength-sized particles with complex shapes, *J. Opt. Soc. Am.*, **A 19**, 521–531.

Kahnert, F. M., J. J. Stamnes, and K. Stamnes, 2002b: Using simple particle shapes to model the stokes scattering matrix of ensembles of wavelength-sized particles with particles with complex shapes: possibilities and limitations, *J. Quant. Spectr. Rad. Transfer*, **74**, 167–182.

Kahnert, F. M., 2003: Numerical methods in electromagnetic scattering theory, *J. Quant. Spectrosc. Radiat. Transfer*, **79–80**, 775–824.

Kite, A. J., 1987: The albedo of broken cloud fields, *Q. J. R. Meteorol. Soc.*, **113**, 517–531.

Kokhanovsky, A., and A. Macke, 1997: The dependence of the radiative characteristicss of optically thick media on the shape of particles, *J. Quant. Spectr. Rad. Transfer*, **63**, 393–407.

Kokhanovsky, A., 2004: Optical properties of terrestrial clouds, *Earth Sci. Revs.*, **64**, 189–241.

Kokhanovsky, A., 2005a: Phase matrix of ice crystals in noctilucent clouds, *Proc. SPIE*, **5829**, 44–52.

Kokhanovsky, A., 2005b: Microphysical and optical properties of noctilucent clouds, *Earth Sci. Rev.*, **71**, 127–146.

Kokhanovsky, A., 2006: Modeling of light depolarization by cubic and hexagonal particles in noctilucent clouds, *Atmos. Res.*, **79**, 175–181.

Korolev, A., G. A. Isaac, and J. Hallett, 2000: Ice particle habits in stratiform clouds, *Q. J. R. Meteorol. Soc.*, **126**, 2873–2902.

Kristjánsson, J. E., J. M. Edwards, and D. L. Mitchell, 2000: The impact of a new scheme for the optical properties of ice crystals on the climate of two GCMs, *J. Geophys. Res.*, **105**, 10063–10079.

Labonnote, L.-C., G. Brogniez, J. C. Buriez, M. Doutriaux-Boucher, J. F. Gayet, and A. Macke, 2001: Polarized light scattering by inhomogeneous hexagonal monocrystals: Validation with ADEOS-POLDER measurements, *J. Geophys. Res.*, **106**, 12139–12153.

Lawson, P., B. A. Baker, C. G. Schmitt, T. L. Jensen, and K. N. Liou, 2001: An overview of microphysical properties of artic clouds observed in May and July 1998 during fire ace, *J. Geophys. Res.*, **106**, 14989–15014.

Lawson, P., B. A. Baker, and B. L. Pilson, 2003: In situ measurements of microphysical properties of mid-latitude and anvil cirrus and validation of satellite retrievals, In *Proc 30th International Symposium on Remote Sensing of Environment*. Honolulu, Hawaii.

Lee, Y. K., P. Yang, M. I. Mishchenko, B. A. Baum, Y. X. Hu, H. L. Huang, W. J. Wiscombe, and A. J. Baran, 2003: Use of circular cylinders as surrogates for hexagonal pristine ice crystals in scattering calculations at infrared wavelengths, *Appl. Opt.*, **42**, 2653–2664.

Liou, K. N., 1986: Influence of cirrus clouds on weather and climate process: A global perspective, *Mon. Weather Rev.*, **114**, 1167–1199.

Liou, K. N., and Y. Takano, 1994: Light scattering by nonspherical particles: Remote sensing and climatic implications, *Atmos. Res.*, **31**, 271–298.

Liou, K. N., Y. Takano, and P. Yang, 2000: Light scattering and radiative transfer by ice crystal clouds: Applications to climate research, in *Light Scattering by Nonspherical Particles: Theory, Measurements, and Geophysical Applications*, Eds. M. I. Mishchenko, J.W. Hovenier, and L. D. Travis, Academic Press, San Diego, CA, 417–449.

Lohmann, U., and E. Roeckner, 1995: Influence of cirrus radiative forcing on climate and climate sensitivity in a general-circulation model, *J. Geophys. Res.*, **100**, 16305–16323.

Lynch, D., K. Sassen, D. Star and G. L. Stephens (Eds.), 2002: *Cirrus*, Oxford University Press, Oxford.

Macke, A., 1993: Scattering of light by polyhedral ice crystals, *Appl. Opt.*, **32**, 2780–2788.

Macke, A. J. Mueller, and E. Raschke, 1996: Single scattering properties of atmospheric ice crystal, *J. Atmos. Sci.*, **53**, 2813–2825.

Mackowski, D. W., 2002: Discrete dipole moment method for calculation of the T matrix for nonspherical particles, *J. Opt. Soc. Am.*, **A 19**, 881–893.

Mano, Y., 2000: Exact solution of electromagnetic scattering by a three-dimensional hexagonal ice column obtained with the boundary-element method, *Appl. Opt.*, **39**, 5541–5546.

McFarlane, S. A., R. T. Marchand, and T. P. Ackerman, 2005: Retrieval of cloud phase and crystal habit from multiangle imaging spectroradiometer (MISR) and moderate resolution imaging spectroradiometer (MODIS) data, *J. Geophys. Res.*, **110**, D14201, doi:10.1029/2004JD004831.

McFarquhar, G. M., and A. J. Heymsfield, 1996: Microphysical characteristics of three anvils sampled during the central equatorial pacific experiment, *J. Atmos. Sci.*, **53**, 2401–2423.

McFarquhar, G. M., and A. J. Heymsfield, 1998: The definition and significance of an effective radius for ice clouds, *J. Atmos. Sci.*, **55**, 2039–2052.

McFarquhar, G. M., A. J. Heymsfield, A. Macke, J. Iaquinta, and S. M. Aulenbach, 1999: Use of observed ice crystal sizes and shapes to calculate the mean-scattering properties and multispecrtral radiance, CEPEX April 4 1993 case study, *J. Geophy. Res.*, **104**, 31763–31779.

McFarquhar, G. M., P. Yang, A. Macke and A. J. Baran, 2002: A new parameterization of single scattering solar radiative properties for tropical anvils using observed ice crystal size and shape distributions, *J. Atmos. Sci.*, **59**, 2458–2478.

Mishchenko, M. I., and A. Macke, 1997: Asymmetry parameters of the phase function for isolated and densely packed spherical particles with multiple internal inclusions in the geometric optics limit, *J. Quant. Spectr. Rad. Transfer*, **57**, 767–794.

Mishchenko, M. I., and L. D. Travis, 1998: Capabilities and limitations of a current FORTRAN implementation of the T-matrix method for randomly oriented, rotationally symmetry scatterers, *J. Quant. Spectrosc. Radiat. Transfer*, **60**, 309–324.

Mishchenko, M. I., L. D. Travis, and A. A. Lacis, 2002: *Scattering, Absorption, and Emission of Light by Small Particles*, Cambridge University Press, Cambridge, UK.

Mitchell, J. F. B., C. A. Senior, and W. J. Ingram, 1989: CO2 and climate: A missing feedback? *Nature*, **341**, 132–134.

Mitchell, D. L., and W. P. Arnott, 1994: A model predicting the evolution of ice particle size spectra and radiative properties of cirrus clouds, II, Dependence of absorption and extinction on ice crystal morphology, *J. Atmos. Sci.*, **51**, 817–832.

Mitchell, D. L., Y. Liu, and A. Macke, 1996: Modeling cirrus clouds. Part II: Treatment of radiative properties, *J. Atmos. Sci.*, **53**, 2967–2988.

Mitchell, D. L., W. P. Arnott, C. Schmitt, A. J. Baran, S. Havemann, and Q. Fu, 2001: Photon tunneling contributions to extinction for laboratory grown hexagonal columns, *J. Quant. Spectr. Rad. Transfer*, **70**, 761–776.

Mitchell, D. L., 2002: Effective diameter in radiation transfer, *J. Atmos. Sci.*, **59**, 2330–2346.

Mitchell, D. L., A. J. Baran, W. P. Arnott, and C. Schmitt, 2006: Testing and comparing the Modified Anomalous Diffraction Approximation, *J. Atmos. Sci.*, **63**, 2948–2962.

Muinonen, K., 1989: Scattering of light by crystals: a modified Kirchhoff approximation, *Appl. Opt.*, **28**, 3044–3050.

Noel, V., and H. Chepfer, 2004: Study of ice crystal orientation in cirrus clouds based on satellite polarized radiance measurements, *J. Atmos. Sci.*, **61**, 2073–2081.

Nousiainen, T., and G. M. McFarquhar, 2004: Light scattering by small quasi-spherical ice crystals, *J. Atmos. Sci.*, **61**, 2229–2248.

Ou, S. C., K. N. Liou, Y. Takano, and R. L. Slonaker, 2005: Remote sensing of cirrus cloud particle size and optical depth using polarimetric sensor measurements, *J. Atmos. Sci.*, **62**, 4371–4383.

Ringer, M. A., J. M. Edwards, and A. Slingo, 2003: Simulation of satellite channel radiances in the Met Office Unified Model, *Q. J. R. Meterol. Soc.*, **129**, 1169–1190.

Rolland, P., K. N. Liou, M. D. King, S. C. Tsay, and G. M. McFarquhar, 2000: Remote sensing of optical and microphysical properties of cirrus clouds using MODIS channels: methodology and sensitivity to assumptions, *J. Geophys. Res.*, **105**, 11,721–11,738.

Rother, T., K. Schmidt, and S. Havemann, 2001: Light scattering on hexagonal ice columns, *J. Opt. Soc. Am*, **A 18**, 2512–2517.

Saunders, C. P. R., and N. M. A. Wahab, 1975: The influence of electric fields on the aggregation of ice crystals, *J. Meteorol. Soc. Japan*, **53**, 121–126.

Slingo, A., 1989: A GCM parameterization for the shortwave radiative properties of water clouds, *J. Atmos. Sci.*, **46**, 1419–1427.

Stamnes, K., S.-C. Tsay, W. Wiscombe, and K. Jayaweera, 1988: Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media, *Appl. Opt.*, **27**, 2502–2509.

Stephens, G. L., and P. J. Webster, 1981: Clouds and climate: sensitivity of simple systems, *J. Atmos. Sci.*, **38**, 235–247.

Stephens, G. L., S.-C. Tsay, P. W. Stackhous, and P. J. Flatau, 1990: The relevance of the microphysical and radiative properties of cirrus clouds to climate and climate feedback, *J. Atmos. Sci.*, **47**, 1742–1753.

Sun, W. B., Q. Fu, and Z. Chen, 1999: Finite-difference time-domain solution of light scattering by dielectric particles with perfectly matched layer absorbing boundary conditions, *Appl. Opt.*, **38**, 3141–3151.

Sun, W. B., and Q. Fu, 2001: Anomalous diffraction theory for randomly oriented nonspherical particles: A comparison between original and simplified solutions, *J. Quant. Spectr. Rad. Transfer*, **70**, 737–747.

Sun, W. B., T. Nousiainen, K. Muinonen, Q. Fu, N. G. Loeb, and G. Videen, 2003: Light scattering by Gaussian particles: a solution with finite-difference time-domain technique, *J. Quant. Spectr. Rad. Transfer*, **79**, 1083–1090.

Sun, W. B., N. G. Loeb, G. Videen, and Q. Fu, 2004: Examination of surface roughness on light scattering by long ice columns by use of a two-dimensional finite-difference time-domain algorithm, *Appl. Opt.*, **43**, 1957–1964.

Takano, Y., and K. N. Liou, 1989a: Solar radiative transfer in cirrus clouds. Part I. Single-scattering and optical properties of hexagonal ice crystals, *J. Atmos. Sci.*, **46**, 3–19.

Takano, Y., and K. N. Liou, 1989b: Radiative transfer in cirrus clouds. II. Theory and computation of multiple scattering in an anisotropic medium, *J. Atmos. Sci.*, **46**, 20–36.

Ulanowski, Z., E. Hesse, P. H. Kaye and A. J. Baran, 2006: Light scattering by complex ice-analogue crystals, *J. Quant. Spectr. Rad. Transfer*, **100**, 382–392.

van de Hulst, H. C., 1957: Light Scattering by Small Particles, Wiley, New York.

Volkovitskiy, O. A., L. N. Pavlova, and A G. Petrushin, 1980: Scattering of light by ice crystals, *Atmos. Ocean. Phys.*, **16**, 90–102.

Wahab, N. M. A., 1974: Ice crystal interactions in electric fields, Ph.D. thesis, UMIST.

Warren, S. G. 1984: Optical constants of ice from the ultraviolet to the microwave, *Appl. Opt.*, **23**, 1206–1224.

Wendling, P., R. Wendling, and H. K. Weickmann, 1979: Scattering of solar radiation by hexagonal ice crystals, *Appl. Opt.*, **18**, 2663–2671.

Westbrook, C. D., R. C. Ball, P. R. Field, and A. J. Heymsfield, 2004: Theory of growth by differential sedimentation, with application to snowflake formation, *Phys. Rev. E*, **70**, Art. No. 021403.

Wylie, D. P., W. P. Menzel, H. M. Woolf, and K. I. Strabala, 1994: Four years of global cirrus statistics using HIRS, *J. Climate*, **7**, 1972–1986.

Wyser, K., and P. Yang, 1998: Average ice crystal size and bulk short-wave single scattering properties of cirrus clouds, *Atmos. Res.*, **49**, 315–335.

Yang, P., and K. N. Liou, 1995: Light scattering by hexagonal ice crystals: Comparison of finite-difference time domain and geometric optics methods, *J. Opt. Soc. Amer. A*, **12**, 162–176.

Yang, P., and K. N. Liou, 1996: Geometric-optics-integral-equation method for light scattering by nonspherical ice crystals, *Appl. Opt.*, **35**, 6568–6584.

Yang, P., K. N. Liou, and W. P. Arnott, 1997: Extinction efficiency and single scattering albedo for laboratory and natural cirrus clouds, *J. Geophys. Res.*, **102**, 21825-21835.

Yang, P., and K. N. Liou, 1998: Single-scattering properties of complex ice crystals in terrestrial atmosphere, *Contr. Atmos. Phys.*, **71**, 223–248.

Yang, P., K. N. Liou, M. I. Mishchenko, and B.-C. Gao, 2000: An efficient finite-difference time domain scheme for light scattering by dielectric particles: Application to aerosols, *Appl. Opt.*, **39**, 3727–3737.

Yang, P., B.-C. Gao, B. A. Baum, W. Wiscombe, Y. Hu, S. L. Nasiri, A. Heymsfield, G. McFarquhar, and L. Miloshevich, 2001: Sensitivity of cirrus bidirectional reflectance in MODIS bands to vertical inhomogeneity of ice crystal habits and size distributions, *J. Geophys. Res.*, **106**, 17267-17291.

Yang, P., B. A. Baum, A. J. Heymsfield, Y. X. Hu, H.-L. Huang, S.-Chee Tsay, and S. Ackerman, 2003: Single-scattering properties of droxtals, *J. Quant. Spectrosc. Radiat. Transfer*, **79–80**, 1159–1180.

Yang, P., G. W. Kattawar, and W. J. Wiscombe, 2004: Effect of particle asphericity on single scattering parameters: Comparison between Platonic solids and spheres, *Appl. Opt.*, **43**, 4427–4435.

Zhang, Y., A. Macke, and F. Albers, 1999: Effect of crystal size spectrum and crystal shape on stratiform cirrus radiative forcing, *Atmos. Res.*, **52**, 59–75.

# 3 Retrieval of cloud optical thickness and effective radius using multispectral remote sensing and accounting for 3D effects

Hironobu Iwabuchi

## 3.1 Introduction

Understanding spatial and temporal variations in cloud properties is crucial to determine the radiation balance on Earth. Remote sensing from satellites provides valuable information on cloud physical properties at global scales (e.g., Rossow and Schiffer, 1991). Recent Earth-observing sensors, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) and Global Imager (GLI), have well-designed spectral channels and horizontal resolutions between 250 m and 1000 m. Compared to earlier sensors, these sensors allow improved derivations of atmospheric and land surface properties. Operational products include the cloud optical thickness and effective particle radius, which are very useful for studying aerosols' indirect effects (Radke et al., 1989; Rosenfeld, 2000).

Clouds in the real atmosphere generally exhibit three-dimensional (3D) inhomogeneity. However, clouds have commonly been assumed to be plane-parallel and homogeneous (PPH) in applications that use one-dimensional (1D) radiative transfer theory (e.g., remote sensing). Nakajima and King (1990) developed a two-channel algorithm to retrieve the cloud optical thickness and effective particle radius, using remote measurement data of solar-reflected radiances at visible and near-infrared wavelengths. The same algorithm has been applied to satellite data at regional and global scales (Han et al., 1994; Nakajima and Nakajima, 1995). In the algorithm, observed radiances of individual pixel are compared with 1D radiative transfer calculations. Such retrievals implicitly use the independent pixel approximation (IPA; Cahalan et al., 1994a,b). The IPA method uses 1D radiative transfer theory at local scales; net horizontal transport of radiation is ignored. Of course, real radiative transfer in a cloudy atmosphere is 3D, and 3D radiative effects influence the radiance actually observed from space. Many studies have suggested that cloudy pixels are not independent at the resolution of satellite data, for both visible and near-infrared wavelengths (e.g., Marshak et al., 1995a, 1999, 2006; Chambers et al., 1997; Titov, 1998; Zuidema and Evans, 1998).

If 1D radiation theory is applied to the retrieval of cloud properties, then estimates of cloud parameters are biased because of 3D effects. Not only aver-

age radiance but also amplitude of radiance fluctuation is different from that of the IPA calculation. Thus, statistics of optical thickness retrieved with the IPA are also biased. Several studies have sought to correct the statistics of retrieved optical thickness by accounting for 3D effects (Barker and Liu, 1995; Chambers et al., 1997; Iwabuchi and Hayasaka, 2002). Marshak et al. (2006) investigated theoretically how the 3D effects work on the retrieval of effective radius of cloud droplets and suggested the 3D effects indeed appear in actual observations with MODIS data. At low resolution (e.g., 1 km), information about sub-pixel inhomogeneity in individual pixels cannot be obtained, even though that information is necessary to correctly interpret pixel reflectance. Thus, accurate pixel-by-pixel retrieval cannot be expected from low-resolution data. It is possible, however, to improve conventional IPA retrievals by correcting statistical moments of the retrieved parameters.

A difficulty in remote sensing of inhomogeneous cloud properties arises from the decorrelation between 3D radiances and cloud properties. For example, Fig. 3.1 compares pixel-averaged nadir radiances simulated at visible wavelengths and pixel-averaged cloud optical thickness at 250-m resolution. A strong relationship between IPA radiance and pixel-averaged optical thickness is obvious. Conventional IPA retrieval assumes a one-to-one relationship between the radiative quantity and the physical quantity to be retrieved. However, the 3D radiance (or the observed radiance) is poorly associated with pixel optical thickness. The 3D radiance of every pixel of inhomogeneous clouds is one of multiple solutions of 3D radiative transfer for various spatial arrangements of cloud elements. Regardless, available satellite data are two-dimensional (2D). Satellite-based remote sensing of clouds is clearly an ill-posed problem, and an accurate retrieval should never be expected. A common strategy for this kind of problem is to increase

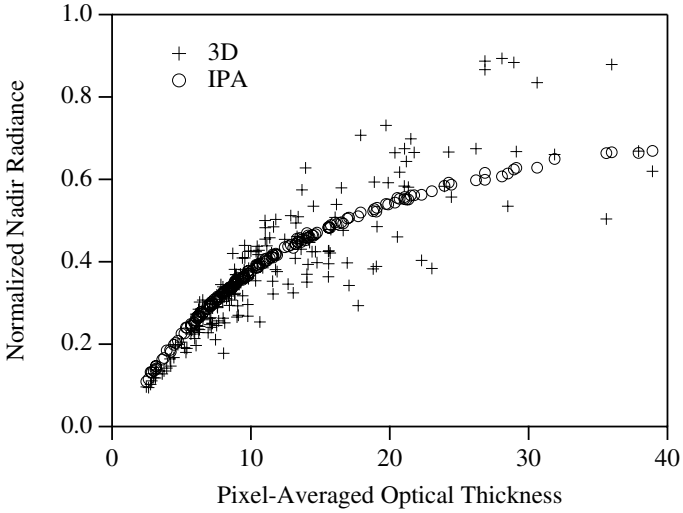



**Fig. 3.1.** The 3D and IPA radiances as functions of pixel-averaged optical thickness.

the data information used in the retrieval algorithm. To that end, multi-angle, multi-spectral data, which are available from recent Earth-observation sensors, can be used.

If satellite data at high resolution (10–50 m) are available, advanced retrieval methods that include pixel-by-pixel retrieval to account for 3D effects can be used. A fundamental approach to improve retrievals is to base the retrieval model on realistic 3D cloud models and accurate 3D radiation models, rather than on 1D models. Horizontal inhomogeneity in pixels can be ignored for radiances at high resolution, but pixel radiances are greatly influenced by the 3D distribution of cloud elements in the adjacent pixels. Therefore, adjacent pixel information could be used to estimate cloud parameters (Marshak et al., 1998; Oreopoulos et al., 2000; Faure et al., 2002; Iwabuchi and Hayasaka, 2003). For example, if a de-convolution is applied to smoothed data, information lost through radiative diffusion (a 3D effect due to multiple scattering) can be recovered. Marshak et al. (1998) used this technique and proposed an inverse non-local IPA (NIPA) model to retrieve the cloud optical thickness from visible wavelength data. Subsequent demonstrations showed that a similar method could be applied to multi-spectral data to retrieve both optical thickness and effective radius (Faure et al., 2001, 2002; Iwabuchi and Hayasaka, 2003). Zinner et al. (2006) recently applied direct 3D radiative transfer calculations in the optical thickness retrieval algorithm that was based on a standard iteration method. New algorithms using high-resolution data are based on realistic 3D radiative transfer models instead of 1D models, which may increase credibility of the retrieved quantities.

This paper considers the effects of 3D radiative transfer on the retrieval of optical thickness and effective particle radius for boundary layer clouds. Section 3.2 describes the cloud model used to simulate radiances and assess retrieval algorithms. Section 3.3 briefly describes the 3D radiative effects found in the reflected radiances. Such effects are important for understanding 3D artifacts in the retrieved cloud parameters. In Section 3.4, 3D effects in optical thickness retrievals that use 1-km resolution visible reflectance data are discussed. Section 3.5 presents algorithms for pixel-by-pixel retrieval of optical thickness and effective radius; these algorithms account for 3D effects. Section 3.6 includes concluding remarks.

## 3.2 The stochastic cloud model

Radiances are typically simulated using a 3D radiation model and a model that allows 3D cloud realizations in studies of 3D radiative effects. Realistic assumptions of optical and geometric properties are vital if an accurate retrieval algorithm is to be developed. Cloud/eddy resolving models, observations (e.g., using satellites, cloud radar, or lidar), and artificial stochastic models can provide cloud data. This study uses a stochastic model because it can easily generate cloud distributions with arbitrary cloud parameters. Spectrum-based stochastic models have been used for this purpose (Evans, 1993; Titov, 1998; Iwabuchi and Hayasaka, 2002, 2003). The power spectrum of the cloud optical thickness obeys

the power law $P \sim k^{-\beta}$, where $k$ denotes the wavenumber, and the spectral exponent $\beta$ is approximately 1.5, yielding values similar to those in observations of stratocumulus clouds (e.g., Barker and Davies, 1992; Davis et al., 1997; Oreopoulos et al., 2000). The scaling exponent is not a dominant parameter for 3D radiation if the exponent is in the range of observed values (Davis et al., 1997; Iwabuchi and Hayasaka, 2002).

Past studies have suggested that the frequency distribution of optical thickness is not Gaussian but has a positive skew (Hayasaka et al., 1994; Cahalan et al., 1994a; Oreopoulos and Davies, 1998b). If the frequency distribution of optical thickness is lognormal, then the probability density function, $p(\tau)$, can be represented as

$$p(\tau) = \frac{1}{\sqrt{2\pi V}\tau \ln 10} \exp\left[-\frac{1}{2}\frac{(\log\tau - M)^2}{V}\right], \tag{3.1}$$

where $M$ and $V$ denote the mean and variance of $\log\tau$, respectively. The mean and variance of optical thickness, $\bar{\tau}$ and $\sigma_\tau^2$, can be determined from $M$ and $V$ in analytic form:

$$\log\bar{\tau} = M + \frac{\ln 10}{2}V \tag{3.2a}$$

$$\rho_\tau \equiv \frac{\sigma_\tau}{\bar{\tau}} = \sqrt{\exp\left[(\ln 10)^2 V\right] - 1} \tag{3.2b}$$

where $\rho_\tau$ is the inhomogeneity parameter (a measure of the degree of horizontal inhomogeneity) presented by Davis et al. (1997) and Szczap et al. (2000). Thus, $\rho_\tau$ has a one-to-one relation with $V$ when the frequency distribution of optical thickness is lognormal. Analyses of six years of satellite data over northeastern Asian seas revealed variability in the inhomogeneity parameter with respect to season and geographical region; for boundary layer clouds, $V = 0.001 - 0.15$ in a $9 \times 9\,\mathrm{km}^2$ domain (Iwabuchi, 2000).

The adiabatic cloud assumption (e.g., Brenguiter et al., 2000) suggests that the geometrical thickness $h$ (m) of a cloud column is related to optical thickness as follows:

$$h(x, y) = B\sqrt{\tau(x, y)} \tag{3.3}$$

where $B$ is a coefficient depending on the cloud type. Minnis et al. (1992) reported that the above proportional relationship could also be derived from satellite observations of marine stratocumulus clouds. Vertical variations in the extinction coefficient and effective droplet radius increase with height, as in the adiabatic parameterization of Brenguiter et al. (2000).

Fig. 3.2 shows one spatial distribution of cloud parameters. There is no orientation to the horizontal distribution of optical thickness; any fluctuations are isotropic, unlike for cloud bands. In the example in Fig. 3.2, the base and top of cloud columns are equally rough. The cloud droplet size is large in regions associated with large optical thickness. The positive correlation between local optical thickness and effective radius is reasonable for boundary layer clouds with few drizzle droplets (Bower et al., 1994).

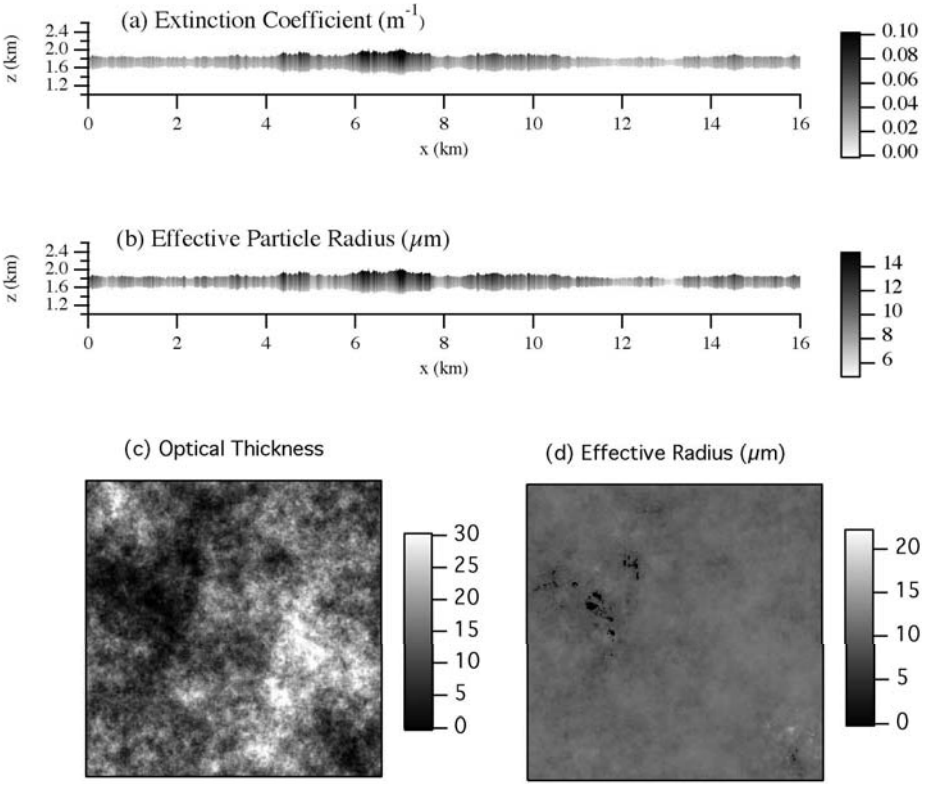



**Fig. 3.2.** Artificially generated input cloud data: (a) and (b) vertical cross-section of the extinction coefficient at 0.55 µm wavelength and the effective droplet radius; (c) and (d) as in (a) and (b) but for horizontal distributions.

## 3.3 Properties of high-resolution radiance

This section describes the 3D effects on high-resolution solar reflected radiance. Differences in the 3D radiance and IPA radiance are important for the inverse problem, as will be shown later. A 3D radiation model that uses Monte Carlo methods was applied to compute solar reflected radiances. The algorithms used in the model have been described by Iwabuchi (2006).

Fig. 3.3 shows horizontal distributions of normalized nadir radiance (reflection function $R$) from sample computations of 3D radiative transfer ($R_{\mathrm{3D}}$) and IPA ($R_{\mathrm{IPA}}$). The reflection function (normalized radiance) is defined as $R = \pi I/(F_0 \cos\theta_0)$, where $I$ is the radiance, $F_0$ is the solar irradiance at the top of atmosphere, and $\theta_0$ is the solar zenith angle. The 3D radiance $R_{\mathrm{3D}}$ is notably smoother than the IPA radiance $R_{\mathrm{IPA}}$ for high solar elevations and rougher for low solar elevations because of smoothing and roughening effects, respectively. These 3D effects have been investigated well by previous studies and known as caused by net horizontal transport of radiation (Marshak et al., 1995a; Várnai

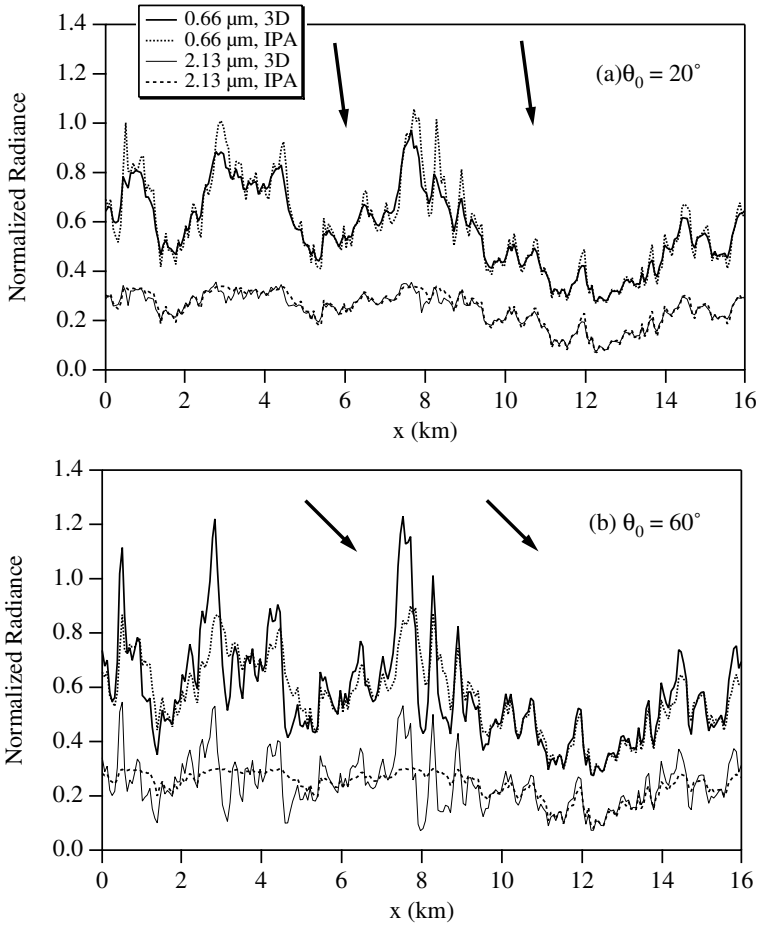



**Fig. 3.3.** High-resolution normalized nadir radiances simulated by the 3D radiative transfer model and the IPA. For clarity, radiances for visible wavelength (0.66 μm) were shifted by 0.2. Vertical homogeneity of liquid water was assumed for the experiment. The effective droplet radius was fixed at 10 μm for the entire domain.

and Marshak, 2003). Smoothing dominates for visible wavelengths and is caused by horizontal divergence of multiply scattered photons; smoothing is therefore most effective in optically thick regions. In optically thin regions, $R_{3D}$ is similar to $R_{IPA}$, i.e., IPA closely approximates the 3D radiance. Less smoothing occurs at near-infrared wavelengths. At low solar elevations, optically thick regions show enhanced roughening that is introduced by shadowing and illumination of direct and low-order scattering beams. At near-infrared wavelengths, $R_{IPA}$ shows small variability in optically thick regions, but $R_{3D}$ shows large fluctuations. However, shadowing has smaller effects in optically thin regions, which helps to explain why $R_{3D}$ is approximately the same as $R_{IPA}$ at both visible and near-infrared wavelengths.

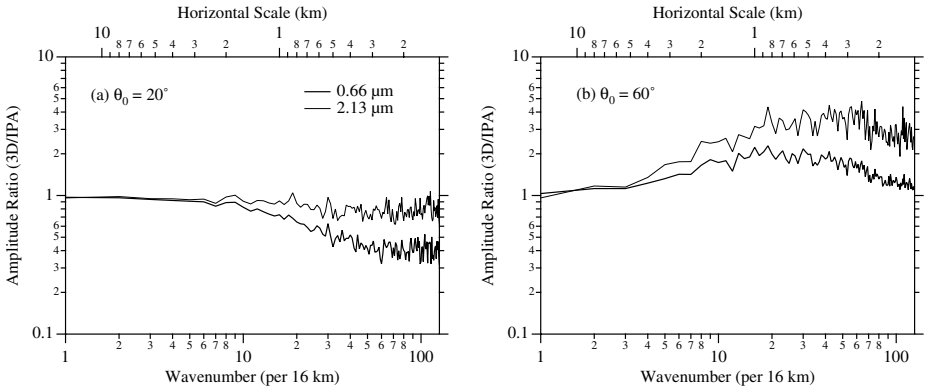



**Fig. 3.4.** Amplitude ratios of Fourier spectra of 3D radiance to those of IPA, at visible and near-infrared wavelengths.

The 3D radiative effects depend on the horizontal scale. This could be well known from power spectra of 3D and IPA radiances. Fig. 3.4 shows amplitude ratios of Fourier spectra between 3D and IPA nadir radiances. Each plot includes an ensemble average of ten cloud realizations. Amplitude ratios less than or greater than unity correspond to smoothing or roughening phenomena, respectively. At large horizontal scales (>5 km), the ratio is close to unity; IPA can approximate the 3D radiative transfer because net horizontal radiative transport can be neglected. At visible wavelengths, smoothing dominates for high solar elevations and small horizontal scales (<2 km), and roughening dominates for low solar elevations and intermediate horizontal scales (0.1–5 km). The 3D effects depend on the single scattering albedo (i.e., on wavelength and effective droplet radius). Less smoothing occurs for absorbing wavelengths (especially for larger effective radii). Thus, multi-spectral rather than single-spectral data can improve the retrieval accuracy of cloud parameters.

Fig. 3.5 shows amplitude ratios of visible-wavelength nadir radiances for values of $M = 0.5$, 1, and 1.5 with fixed inhomogeneity $V = 0.07$. The corresponding domain-averaged optical thicknesses are 3.9, 12, and 39, respectively, for each case. The geometric thickness given by (3.3) increases for increasing $M$, although the geometric roughness of the cloud top and bottom remains constant for fixed values of $V$. Smoothing is effective for moderate optical thickness at $M = 1$ but does not strongly depend on $M$. When the solar elevation is low, roughening is very sharp for optically thick clouds because shadowing and illumination of direct and forward-scattering beams are more effective for optically and geometrically thicker clouds. The amplitude of 3D radiance for $M = 1.5$ is four times the IPA radiance at horizontal scales of approximately 0.5–1 km. The 3D radiative effects thus vary substantially with mean optical thickness. The dependence of 3D effects on $M$ at near-infrared wavelengths is similar to that at visible wavelengths (not shown).

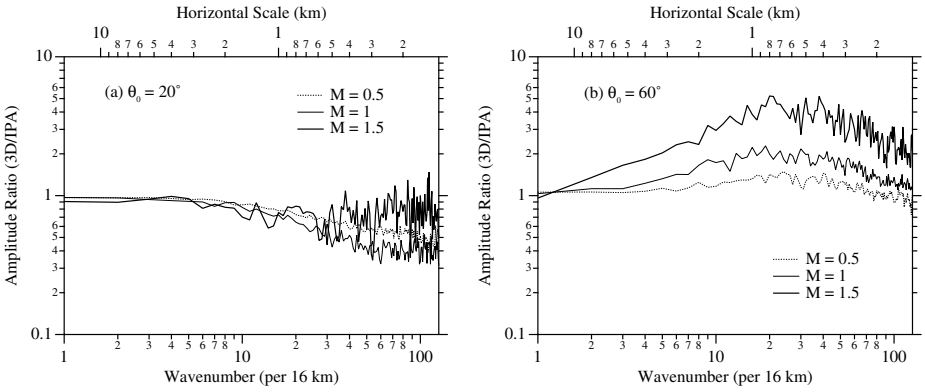



**Fig. 3.5.** Amplitude ratios of the Fourier spectra of 3D radiance to those of IPA, for various averaged logarithms of optical thickness, $M$, with a fixed inhomogeneity parameter, $V = 0.07$. Results are shown for visible wavelength (0.66 μm).

Fig. 3.6 is the same as Fig. 3.5, but with $V = 0.03$, 0.07, and 0.11 for a fixed $M = 1$; corresponding domain-averaged optical thicknesses are 11, 12, and 14, respectively. No significant difference exists between the three cases because the effects of smoothing and roughening are independent of the inhomogeneity parameter $V$ if $M$ is the same. The inhomogeneity parameter roughly determines the amplitude of the IPA radiance fluctuation. Relative changes in the amplitude of the 3D radiance depend on the horizontal scale, wavelength, and cloud optical thickness, but not on horizontal inhomogeneity. If, for example, the 3D radiance for an inhomogeneous cloud field exhibits five-times smoother (or rougher) fluctuation than the IPA radiance, then that is also true for more homogeneous cloud field. Such a similarity of 3D radiative effects can be found between inhomogeneous and relatively homogeneous cloud fields. The 3D radiation problem and remote sensing retrieval algorithms can be simplified based on the similarity of 3D radiative effects related to inhomogeneity.

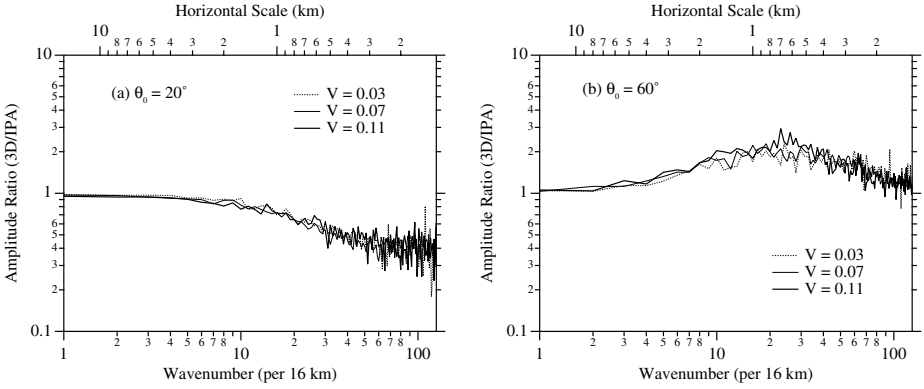



**Fig. 3.6.** As in Fig. 3.5, but for various values of $V$, with $M$ fixed at 1.

## 3.4 Statistical analysis of the 3D effects and correction

This section describes the use of visible reflected radiance data with 1-km resolution to evaluate 3D effects on moments of cloud optical thickness retrievals. Optical thickness is often retrieved using only visible data because visible radiance is not highly dependent on the effective droplet radius. Results are shown for a wavelength of 0.64 μm that is the center of the Advanced Very High Resolution Radiometer (AVHRR) visible channel. The statistical quantities of the retrieved cloud optical thickness in the $8 \times 8\,\mathrm{km}^2$ domain were investigated. The goal was to correct the statistical properties of optical thickness retrieved with the IPA.

### 3.4.1 The influence on the statistics of retrieved optical thickness

The reflected radiance is roughly linear in $\log\tau$ especially when $\tau$ is between 3 and 30. When net horizontal radiative transport is absent, i.e., when the IPA works perfectly, moments of radiance are closely associated with moments of $\log\tau$ rather than moments of $\tau$. Consider the mean and variance, $M$ and $V$, respectively, of $\log\tau$. The pixel-average radiance of the IPA ($R_{\mathrm{IPA}}$) is roughly approximated as the 1D radiance for PPH cloud with $\log\tau = M$, i.e.,

$$R_{\mathrm{IPA}} \sim R_{\mathrm{1D}}(\log\tau = M)\,. \tag{3.4}$$

This is equivalent to the effective thickness approximation of Cahalan et al. (1994a), who used it in calculating domain-averaged albedo. To improve the linear relationship, the following function can be used instead of $\log\tau$:

$$\chi = \frac{\gamma\tau}{1+\gamma\tau}\,,$$

where the constant $\gamma$ is $1-g$ ($g$ is the asymmetry parameter and is approximately 0.86 for water clouds in the visible). However, this paper is devoted to analyses of mean and variance of $\log\tau$ for simplicity.

A cloud field can be characterized by the two statistical quantities, $M$ and $V$. If conventional IPA inversion is used to retrieve the optical thickness from satellite measurements, then the retrieved value is biased because 3D radiative effects influence the observed radiance. Here, $M_{\mathrm{IPA}}$ and $V_{\mathrm{IPA}}$ are the mean and variance, respectively, of the logarithm of retrieved optical thickness. Fig. 3.7 shows the frequency distribution of initial and retrieved optical thickness for a cloud field with inhomogeneity $V = 0.09$ for the $8 \times 8\,\mathrm{km}^2$ region. The initial field of the optical thickness was artificially generated using the cloud model (section 3.2). For this field, 3D radiances were simulated and subsequently used to retrieve the optical thickness with the IPA inversion. Thus, we can estimate the error in the retrieved value by comparing with the initial (truth) data. It is shown that both $M_{\mathrm{IPA}}$ and $V_{\mathrm{IPA}}$ are biased, with the biases defined as

$$\Delta M \equiv M_{\mathrm{IPA}} - M \tag{3.5a}$$

$$\Delta V \equiv V_{\mathrm{IPA}} - V\,. \tag{3.5b}$$

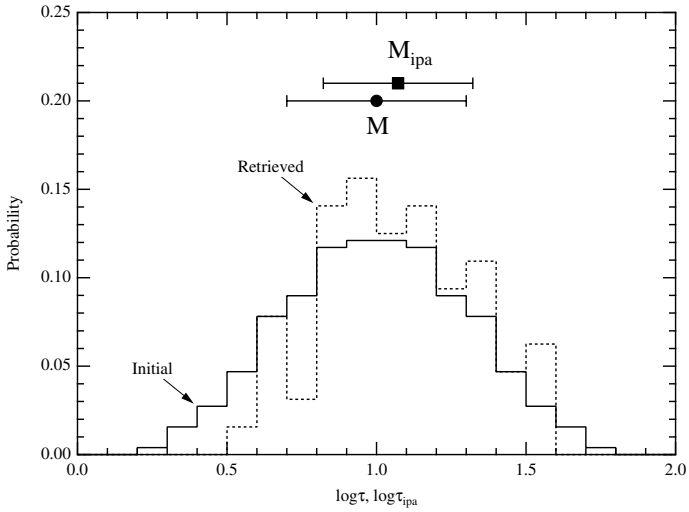



**Fig. 3.7.** Frequency distributions of initial optical thickness ($(\tau)$ and retrieved optical thickness ($\tau_{\mathrm{IPA}}$) for an inhomogeneous cloud field. Marks in the upper part denote mean values of $\log \tau$ (circles) and $\log \tau_{\mathrm{ipa}}$ (squares), and bars correspond to respective standard deviations. This example is for a case of backscattering viewing geometry ($\theta_0 = 60°$, $\mu = 0.7$–$0.8$, $\phi = 150°$–$180°$).

These biases imply differences between statistical properties of 3D radiance and IPA radiance. For a homogeneous cloud field, both biases should be zero because the 1D radiative transfer can accurately approximate the observed radiance. Biases will be large for inhomogeneous cloud fields.

### 3.4.2 Biases in the statistics of the optical thickness

Biases in $M_{\mathrm{IPA}}$ and $V_{\mathrm{IPA}}$ were investigated under various boundary layer cloud conditions. Fig. 3.8 relates the biases with the viewing angle for four cloud models with different assumptions in geometric roughness. Results are shown for angular averages of radiance for bins of $\mu = \cos\theta$ ($\theta$ is viewing zenith angle) and relative azimuth angle, which is 0°–30° or 150°–180°. The angular averaging does not significantly affect this kind of results because the radiances were calculated at the level of highest cloud top and the spatial resolution being considered here is as low as 1 km. The bias in $M_{\mathrm{IPA}}$ has small negative values for an overhead sun (the solar zenith angle $\theta_0 = 0°$). This bias is ascribed to the tilted cloud surface that reflects photons in the off-nadir direction and leads to the decrease of the near-nadir reflection. For an oblique sun ($\theta_0 = 60°$), the bias in $M_{\mathrm{IPA}}$ is negative for the forward view due to cloud side shadowing and positive for the backward view due to cloud side illumination. The tendency of the cloud 3D effect is similar among the four cloud models, but its magnitude differs substantially and increases for clouds with rough tops. Cloud top bumps allow for large horizontal transport of incident and reflected solar radiation. In addi-

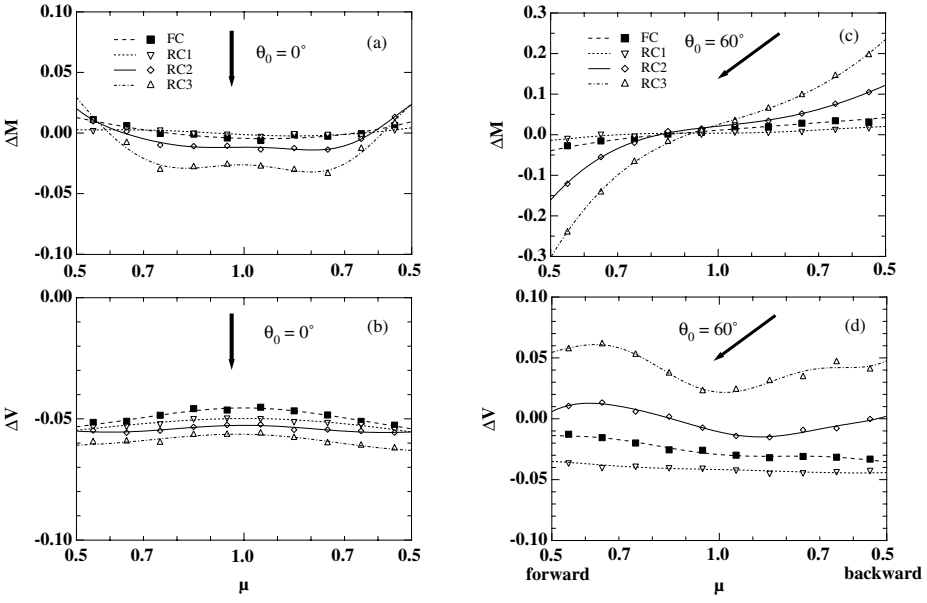



**Fig. 3.8.** The biases $\Delta M$ (a) and (c) and $\Delta V$ (b) and (d) as functions of cosine of viewing zenith angle ($\mu = \cos\theta$), for four cloud models: flat cloud (FC), cloud with a rough bottom and flat top (RC1), cloud with a rough top and bottom (RC2), and cloud with a flat bottom and rough top (RC3). The domain-averaged geometric thickness and the local optical thickness of each column are the same for the four models. Results are shown for angular averages of radiance for bins of $\mu$ and the relative azimuth angle ($\phi$), which is 0°–30° for the left hand side of each panel or 150°–180° for the right hand side.

tion, most of the photons are reflected from the upper part of the cloud layer, so that cloud top inhomogeneity is important. Many observations of boundary layer clouds using lidar, radar, and stereo-photography have shown rough cloud tops and cloud bottoms (e.g., Boers et al., 1998; Kikuchi et al., 1993; Vali et al., 1998). Such an assumption therefore is reasonable in evaluating the 3D effects.

Fig. 3.8 also shows bias in the IPA-based retrieved inhomogeneity ($V_{\mathrm{IPA}}$). Bias in $V_{\mathrm{IPA}}$ is large and negative when the solar elevation is high, which indicates that the spatial variability of the reflected radiance is small because of smoothing by photon diffusion in multiple scattering processes, as shown in section 3.3. An enhanced smoothing effect occurs in the model that includes bumpy cloud tops, which allow greater horizontal radiative transport than flat cloud tops. When the solar elevation is low, the opposite effect, roughening, is caused by enhanced variability of direct and low-order scattering of radiation incident to the inhomogeneous cloud column. The roughening effect is large for bumpy cloud tops and closely associated with the cloud top structure. Figure 3.8 (d) shows a tendency for sharper roughening in the forward view than in the backward view. Both bright and dark regions of the cloud surface are viewed in the for-

ward view; direct solar beams illuminate the bright regions, and optically dense parts shadow the dark regions. However, bright parts are mainly viewed in the backward direction, which decreases the reflectance contrast.

The viewing angle dependence of the 3D effect suggests that satellite-derived optical thickness is systematically larger for backward views than for forward views. Loeb and Coakley (1998) reported a systematic decrease in observed optical thickness with increasing viewing angle in the forward view, results that are consistent with results presented here. Similarly, Oreopoulos and Davies (1998a) showed a solar zenith angle dependence on the variance of $\log \tau_{\mathrm{IPA}}$ that was remotely sensed using AVHRR data. In their results, the variance of $\log \tau_{\mathrm{IPA}}$ systematically increased as the solar zenith angle increased between $\theta_0 = 50°$–$80°$, results that are consistent with the results presented here.

Figure 3.9 compares $\Delta M$ and $\Delta V$ to the inhomogeneity parameter $V$. Large values of $V$ are associated with large variability in optical and geometrical thicknesses and the cloud top height. The standard deviation of the cloud top height is large for a large $V$. The absolute value of $\Delta M$ increases as $V$ increases so that cloud 3D effects on brightness (e.g., brightening and darkening) are governed by the parameter $V$. In fact, $\Delta M$ is nearly proportional to $V$. The parameter $V$ is important in describing the radiative effects of cloud inhomogeneity. In addition, $\Delta V$ is roughly proportional to $V$, and the proportionality is good except for off-nadir views with oblique sun. In other words, the relative bias $\Delta V/V$ is nearly constant with respect to the bi-directional angle. Thus, smoothing and roughening phenomena are almost independent of the degree of horizontal inhomogeneity and the two-parameter representation (with $M$ and $V$) simplifies the 3D radiative effect dependence on the degree of inhomogeneity.

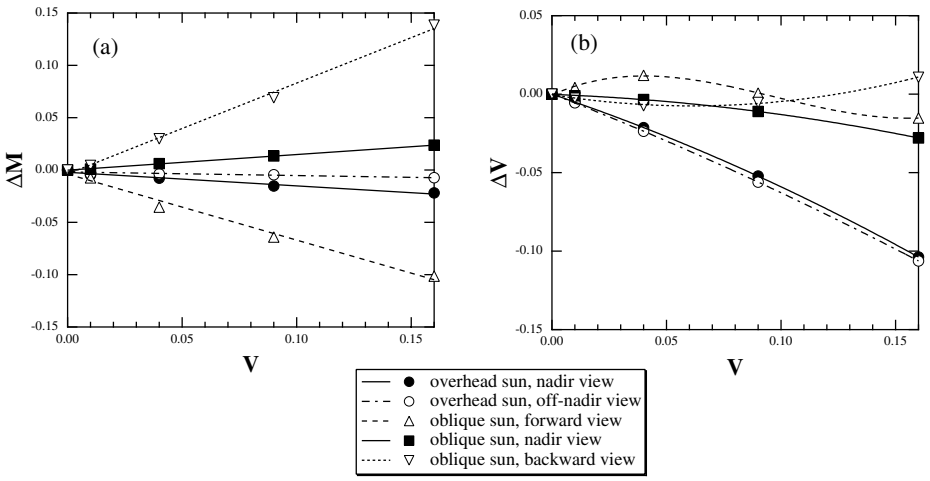



**Fig. 3.9.** IPA biases (a) $\Delta M$ and (b) $\Delta V$ as functions of the inhomogeneity parameter $V$ for overhead sun ($\theta_0 = 0°$) and oblique sun ($\theta_0 = 60°$). Geometric parameters are 'nadir view,' $\mu = 0.9$–$1.0$, $\phi = 0°$–$30°$; 'off-nadir view,' $\mu = 0.6$–$0.7$, $\phi = 0°$–$30°$; 'forward view,' $\mu = 0.6$–$0.7$, $\phi = 0°$–$30°$; 'forward view,' $\mu = 0.6$–$0.7$, $\phi = 150°$–$180°$.

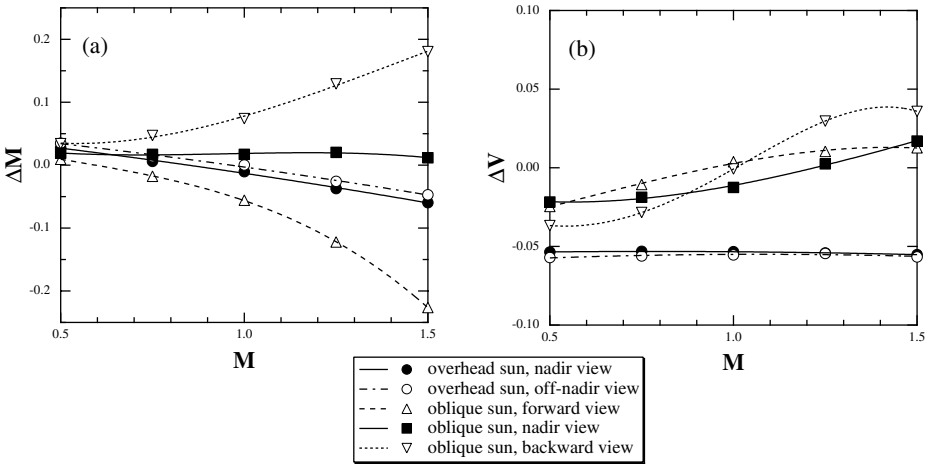



**Fig. 3.10.** As in Fig. 3.7, but for a function of $M$.

Fig. 3.10 compares $\Delta M$ and $\Delta V$ to $M$. A high sensitivity of $\Delta M$ to $M$ is obvious, especially for off-nadir views with oblique sun. The difference in $\Delta M$ between the forward and backward views increases for optically thick cloud fields. The 3D effects on brightness are sharp for optically thick cloud fields. The bias $\Delta V$ increases with increasing $M$ for oblique sun but is almost independent of $M$ for overhead sun, because the roughening is sharp for optically thick clouds, as shown in Fig. 3.5. In thick clouds, photons travel short trajectories (on average). Thus the horizontal distance between the incident point at the cloud surface and the reflected exit point is relatively close, reducing the smoothing effect.

The solar zenith angle dependence of the IPA-retrieved optical thickness may be the most remarkable artifact of neglecting 3D effects. The average of the retrieved optical thickness can be roughly estimated as

$$\log \bar{\tau}_{\mathrm{IPA}} \cong M_{\mathrm{IPA}} + \frac{\ln 10}{2} V_{\mathrm{IPA}} \,.$$

Figure 3.11 shows angular distributions for a typical case with $M = 1$ and $V = 0.09$. For this case, the above equation for nadir-viewing geometry ($\mu = 1$) yields estimated mean optical thicknesses of 11, 13, and 17, respectively, for $\theta_0 = 0°$, $60°$, $70°$. The true mean optical thickness is 12.7. Such a solar angle dependence could appear in cloud climatology. For example, the annual zonal-mean optical thickness of low clouds increases with increasing latitude from about 5 at the Equator to about 12 at 60° north and south in the International Satellite Cloud Climatology Project (ISCCP) product (Tselioudis et al., 1992; Drake, 1993). Seasonal changes have also been reported. Latitudinal and seasonal changes might be partly affected by 3D effects. Therefore, global observations of the inhomogeneity parameter and correction of the IPA retrieved optical thickness to account for 3D radiative effects are clearly warranted.

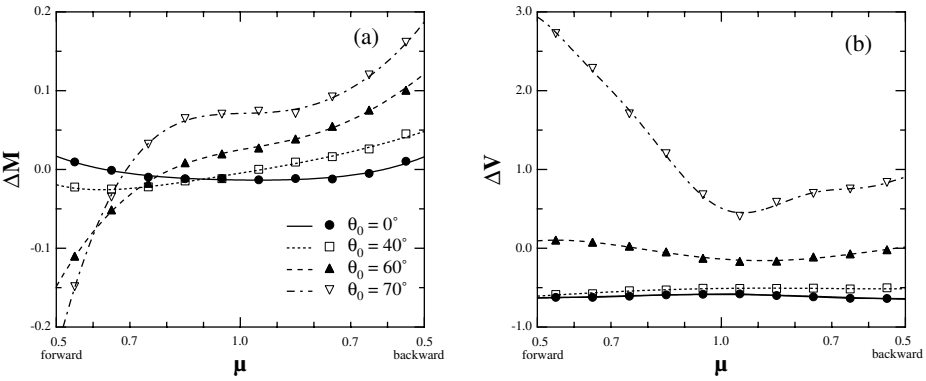



**Fig. 3.11.** Viewing angle distributions of the IPA biases $\Delta M$ and $\Delta V$ for the four solar zenith angles $\theta_0 = 0°$, 40°, 60°, 70°.

### 3.4.3 Bias removal

To correct for the 3D effect on retrieved optical thickness, some empirical assumptions are required because it is difficult to retrieve several properties of clouds from optical remote sensing (e.g., cloud top roughness at sub-pixel scales). If the vertical profile of liquid water is prescribed (e.g., as the adiabatic parcel) and geometrical parameters (thickness and cloud top/bottom hight) are assumed to link to the optical thickness solely, then biases in the mean and variance of the logarithm of retrieved optical thickness can be expressed as

$$\frac{\Delta M}{V} = f(\mu_0, \mu, \phi; M, \alpha_g) \tag{3.6a}$$

$$\frac{\Delta V}{V} = g(\mu_0, \mu, \phi; M, \alpha_g) \tag{3.6b}$$

where $\mu_0$ and $\mu$ are cosines of solar and satellite zenith angles, respectively, and $\alpha_g$ is the surface albedo. The proportionality of the biases to the inhomogeneity parameter $V$ in (3.6) simplifies the parameterization. The formulation could be further modified to include additional minor factors including the relationship between optical and geometrical thickness in cloudy columns, vertical inhomogeneity in columns, effective particle radius, and azimuth-dependent orientation of cloud inhomogeneity. The functions $f$ and $g$ can be tabulated and computed by interpolation using a look-up table. Such computations are better than the function fitting shown by Iwabuchi and Hayasaka (2002).

Equations (3.5) and (3.6) can easily be applied to correct the retrieved optical thickness from satellite observations. They form a nonlinear system of equations for $M$ and $V$ that can be solved by the method of iterations:

(i) Initial estimates: $M = M_{\mathrm{IPA}}$, $V = V_{\mathrm{IPA}}$
(ii) Biases $\Delta M$, $\Delta V$ computed using (3.6)
(iii) Biases removed: $M = M_{\mathrm{IPA}} - \Delta M$, $V = V_{\mathrm{IPA}} - \Delta V$
(iv) If results have converged, then end. Otherwise, return to step (ii)

If needed, the mean and standard deviation of the optical thickness can be estimated as a by-product from $M$ and $V$, using Eqs. (3.2a) and (3.2b). If the statistics of $\tau$ are more important than those of $\log\tau$, then a parameterization of the bias in the statistics of $\tau$ can be developed, with a form similar to that of (3.6). A main advantage of this type of correction method is that it can be used after conventional IPA retrieval; no reanalysis that uses the radiance is required. The above algorithm corrects the statistics; however, pixel optical thickness could be corrected following a similar path. After the corrected statistics $M$ and $V$ are computed, pixel optical thickness can be estimated by

$$\log\tau = M + \sqrt{\frac{V}{V_{\mathrm{IPA}}}}\,(\log\tau_{\mathrm{IPA}} - M_{\mathrm{IPA}}) \tag{3.7}$$

This algorithm may be useful for correcting any operational product currently derived using IPA.

Another possible method is to use a parameterization of the statistics of radiances. Differences in the average and variance of the 3D radiance from that of IPA radiance can be expressed as

$$\bar{R}_{\mathrm{3D}} - \bar{R}_{\mathrm{IPA}} = f'(\mu_0, \mu, \phi; M, V, \alpha_g) \tag{3.8a}$$

$$\sigma^2_{\mathrm{3D}} - \sigma^2_{\mathrm{IPA}} = g'(\mu_0, \mu, \phi; M, V, \alpha_g) \tag{3.8b}$$

The retrieval algorithm first corrects the observed (3D) radiance and then estimates the IPA radiance using the parameterization. Then, optical thickness (or its moments) can be estimated from the IPA radiance. This algorithm does require analyses of observed radiances and cannot be used to correct operational products.

## 3.5 Pixel-by-pixel retrieval

Pixel-by-pixel retrievals of the pixel-averaged optical thickness and droplet effective radius are discussed in this section. These two parameters are usually retrieved with the IPA; the parameters are inverted from observed visible and near-infrared radiances and theoretical simulations that use a 1D radiative transfer model (Fig. 3.12). The use of data with resolution of 250 m (i.e., the same resolution as for MODIS visible channels) is considered first. The target spectral wavelengths in this study were 0.66 and 2.13 μm, which are the center wavelengths of visible and near-infrared MODIS bands, respectively.

Figure 3.1 shows that the 3D radiance is poorly associated with local cloud properties, mainly because 3D radiance can be affected by net horizontal radiative transport from neighboring pixels. Unfortunately, the quantity observed is the 3D radiance rather than the IPA radiance. Once multi-spectral IPA radiances are estimated, optical thickness and effective radius can be retrieved conventionally by interpolation from a look-up table (Fig. 3.12). Thus, retrieval of inhomogeneous cloud parameters can be restated regarding how to estimate

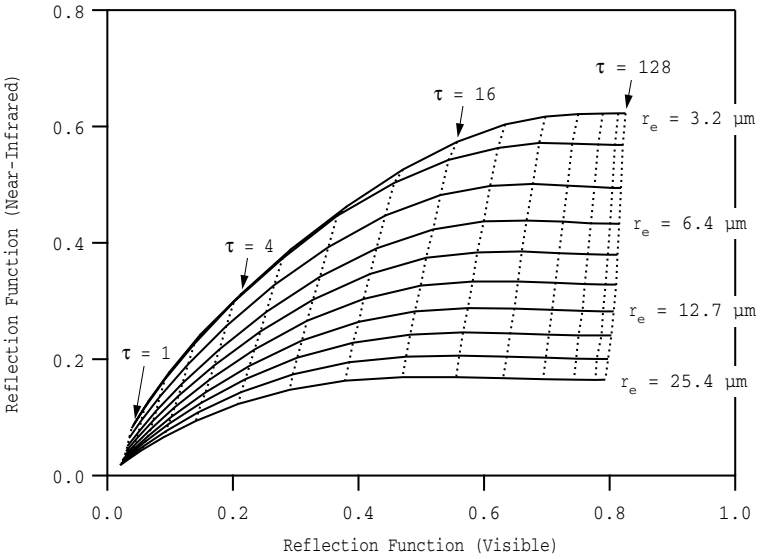



**Fig. 3.12.** Relationships between cloud parameters (optical thickness and effective particle radius) and reflection functions simulated by the 1D radiative transfer model for plane-parallel clouds.

the IPA radiance from (observed) 3D radiance. Therefore, neighboring pixel data are used to estimate the IPA radiance, which has a one-to-one relationship with the cloud property. This is the basis of non-local retrieval methods such as NIPA (Marshak et al., 1998). One can consider an alternative approach, as shown in section 3.5.4; if the 3D radiance can be estimated from IPA radiance using some method, then that method can also be used to retrieve cloud parameters.

The 3D-to-IPA estimate or IPA-to-3D estimate can be done using a convolution (or filtering). Estimated radiance is then a summation of all contributions from adjacent pixels. This technique is so flexible that a filter can be designed to treat 3D radiative effects, which work differently by solar zenith angle, average optical thickness, single scattering albedo that depends on wavelength and effective particle radius, as shown in section 3.3. An empirical model can be used to determine the filter; this model could be based on regressions using the least-squares method, neural nets, genetic algorithms, or a Bayesian method.

### 3.5.1 Retrieval method using adjacent pixel information

In the method presented here, multi-spectral IPA radiances were estimated from observed radiances at target and neighboring pixels within 750 m of the target, using a technique similar to a convolution. The coefficients for the convolution kernel can be determined by a regression that uses 3D radiances simulated under various conditions (e.g., average optical thickness and effective radius).

The IPA radiances $Y$ that depend strongly on the physical quantities (i.e. optical thickness and effective radius) of the target pixel are represented in the

following regression formula:

$$Y = a + \sum_{i=1}^{m} \left\{ b_i \tilde{X}_i + \sum_{j=-n}^{-1} c_{ij} \frac{X_{i,j} - X_{i,j+1}}{\tilde{X}_i} + c_{i0}(X_{i,0} - \tilde{X}_i) + \sum_{j=1}^{n} c_{ij} \frac{X_{i,j} - X_{i,j+1}}{\tilde{X}_i} \right\} \tag{3.9}$$

where $m = 2$, $X_{i,j}$ denotes a 3D radiation function at the $i$th wavelength and the $j$th pixel, $\tilde{X}_i$ is the averaged 3D radiation function over $j$ from $-n$ to $n$, and $a$, $b_i$, and $c_{ij}$ are coefficients. The pixel index $j$ is 0 for the target pixel; $j < 0$ and $j > 0$ denote neighboring pixels. The number of neighboring pixels was set to $n = 3$, and data from pixels within 750 m of the target pixel were used. Preliminary tests showed that neighboring pixels beyond $n = 3$ did not significantly improve the retrieval accuracy. Table 3.1 summarizes definitions adopted for $Y$ and the corresponding $X_{i,j}$. Coefficients and 3D radiative functions ($X_{i,j}$) are defined independently for respective IPA radiances (visible and near-infrared), as in Table 3.1. Once the IPA radiances are estimated from observed 3D radiation functions ($X_{i,j}$) using the above equations, cloud properties can be retrieved conventionally from the one-to-one relationship between IPA radiances and cloud properties using a lookup table and interpolation as in Fig. 3.12.

**Table 3.1.** Radiative quantities used for the regression model of Eq. (3.9)

| $Y$ | $X_{1,j}$ | $X_{2,j}$ |
|---|---|---|
| $R_{\mathrm{ipa}}(\lambda = 0.66, \tau, r_{\mathrm{e}})$ | $R_{\mathrm{3d}}(\lambda = 0.66, j)$ | $R_{\mathrm{3d}}(\lambda = 0.66, j)/R_{\mathrm{3d}}(\lambda = 0.66, j)$ |
| $R_{\mathrm{ipa}}(\lambda = 2.13, \tau, r_{\mathrm{e}})$ | $R_{\mathrm{3d}}(\lambda = 2.13, j)$ | $R_{\mathrm{3d}}(\lambda = 2.13, j)/R_{\mathrm{3d}}(\lambda = 0.66, j)$ |

The coefficients, $a$, $b_i$, and $c_{ij}$, were determined using the least-squares method. The regression used training data sets of IPA and 3D radiances for several cloud realizations. One set of coefficients was determined for $R_{\mathrm{ipa}}$ for each wavelength and for each solar and view directions. The estimation formula (3.9) was used to recover information lost by the 3D radiative effects including radiative smoothing and shadowing. Faure et al. (2002) showed that a de-smoothing process such as image enhancement is required for smoothed data. In addition, smoothing should operate in the solar azimuth direction for data roughened by the shadowing that works in that direction. The horizontal distributions of the coefficients $c_{ij}$ can be considered as a filter that operates on 3D radiance data. Section 3.3 showed that radiative smoothing and roughening work differently for average optical thickness and effective particle radius. Different sets of regression coefficients were therefore prepared for different average optical thickness and effective particle radius values.

The method described above can be simplified to a method that uses single-spectral and/or local data. If local data are used to estimate the IPA radiance,

then $n = 0$, and equation (3.9) becomes

$$Y = a + \sum_{i=1}^{m} b_i X_{i,0} \,. \tag{3.10}$$

Similarly, a method might use single-spectral visible radiance to retrieve optical thickness. The next sections investigate performances of the single-spectral local (SSL) retrieval method, the single-spectral non-local (SSN) method, and the multi-spectral local (MSL) method. Tables 3.2 and 3.3 summarize data used in the five methods, including IPA.

**Table 3.2.** Numbers of spectral wavelengths ($m$) and neighboring pixels ($2n$) used to retrieve cloud optical thickness

| Method | $m$ | $n$ |
|---|---|---|
| IPA | 1 (*) | 0 |
| SSL | 1 (*) | 0 |
| SSN | 1 (*) | 3 |
| MSL | 2 | 0 |
| MSN | 2 | 3 |

Independent pixel approximation (IPA), single-spectral local (SSL), single-spectral non-local (SSN), multi-spectral local (MSL), and multi-spectral non-local (MSN) methods. (*) Visible wavelength data were used.

**Table 3.3.** Number of spectral wavelengths ($m$) and neighboring pixels ($2n$) used for two-parameter retrieval of cloud optical thickness and effective droplet radius

| Method | $m$ | $n$ |
|---|---|---|
| IPA | 2 | 0 |
| MSL | 2 | 0 |
| MSN | 2 | 3 |

Notation as in Table 3.2

### 3.5.2 Optical thickness retrieval

The performances of different methods of retrieving optical thickness are compared. Table 3.2 lists the five methods: IPA, SSL, SSN, MSL, and MSN. Cloud realizations were generated with the domain average optical thickness of 12 and standard deviation of 8. Quasi-observation radiances were simulated with a stochastic cloud model and a 3D radiation model. Such radiances are used to assess retrieval algorithms. The retrieval error can be estimated by considering the input distribution of cloud parameters to be true.

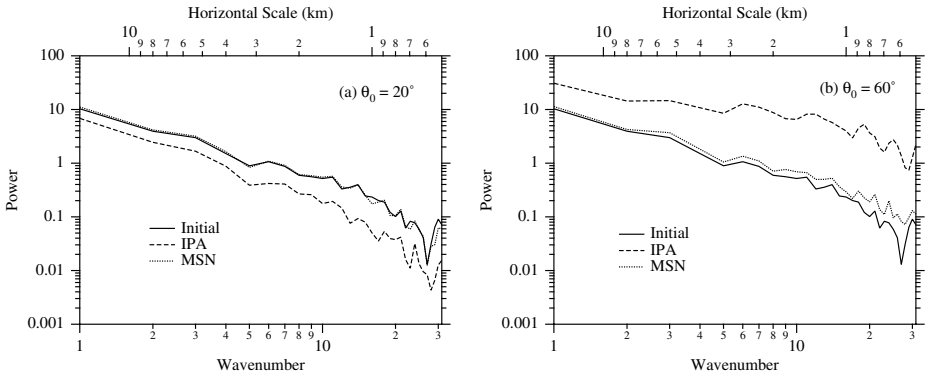



**Fig. 3.13.** Power spectra of the initial and retrieved optical thickness from the IPA and from the empirical retrieval method (MSN) that uses multi-spectral data and neighboring pixel data.

Figure 3.13 shows ensemble-average power spectra of the reference (initial) and retrieved pixel-averaged optical thickness. IPA-retrieved optical thickness is affected significantly by smoothing and roughening. In contrast, the MSN method almost perfectly reproduces the original fluctuations when the solar elevation is high. Even for low solar elevations, the MSN method retrieves a fluctuation, albeit rougher, that is very close to the original. Both multi-spectral data and neighboring pixel data improve the retrieval, as does the ability to incorporate filtering as used in image processing (e.g., image enhancement, phase correction, and smoothing).

Figure 3.14 shows the root mean square errors (RMSEs) for retrieved pixel-averaged optical thickness. The RMSE with IPA is 15–20% for $\theta_0 = 20°$ and 50–80% for $\theta_0 = 60°$. The SSL method uses the same data as the IPA method, but the SSL error is much smaller because the SSL method uses an empirical model that is based on 3D models. The retrieval performance of the SSN method is better than that of the MSL method in most cases, and both are better than the SSL method. Use of non-local data is particularly effective for high solar elevations. In addition, multi-spectral data reduce the retrieval error primarily at low solar elevations. The MSN method has the best performance. The RMSE is 4–6% for $\theta_0 = 20°$ and 7–8% for $\theta_0 = 60°$. Errors are smaller than in the IPA method by a factor of 3–10. The use of both multi-spectral and non-local data yields better estimates of cloud properties.

### 3.5.3 Retrieval of optical thickness and effective particle radius

Section 3.3 showed that 3D radiative effects depend primarily on average optical thickness and single scattering albedo (i.e., on effective radius). Therefore, regression coefficients in (3.9) were determined independently and tabulated for different combinations of average optical thickness and effective droplet radius. Regression coefficients vary strongly with the average optical thickness and effec-

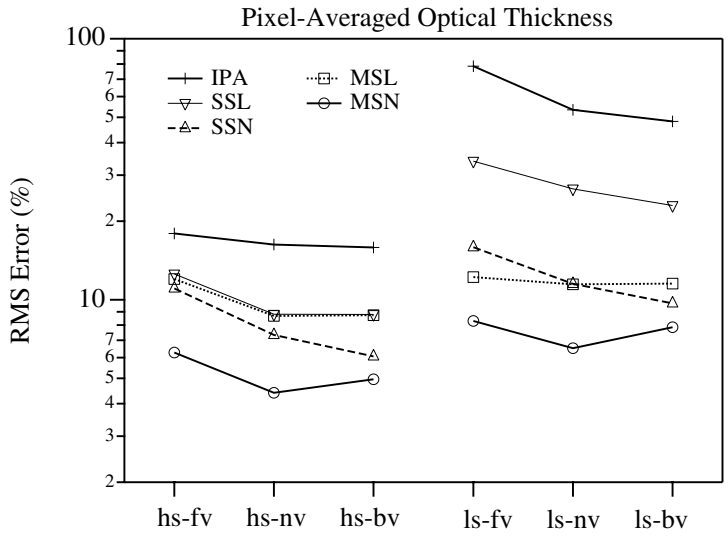



**Fig. 3.14.** Root-mean-square errors (RMSEs) in the retrieved optical thickness for various methods: IPA, single-spectral local (SSL), single-spectral non-local (SSN), multi-spectral local (MSL), multi-spectral non-local (MSN). Notations on the horizontal axis: 'hs': high solar elevation ($\theta_0 = 20°$); 'ls': low solar elevation ($\theta_0 = 60°$); 'fv': forward view; 'nv': nadir view; 'bv': backward view. For oblique views, the view zenith angle is 30°; the relative azimuth angle is 0° and 180° for the forward and backward views, respectively.

tive radius. That variability underscores the importance of training for various cases to generalize the empirical model.

Retrieved cloud parameters vary according to what set of regression coefficients for specific averages of optical thickness and effective radius are used. However, accurate estimates of the averages are unavailable when the retrieval algorithm is initially applied. Thus, optimal values of pixel-averaged optical thickness ($\bar{\tau}$) and effective radius ($\bar{r}_{\rm eff}$) were derived using an iterative procedure that included three processes:

(i) estimation of pixel averages, $\bar{\tau}$ and $\bar{r}_{\rm eff}$, from the IPA radiances that are derived using (3.9);
(ii) computation of running averages of cloud parameters within 750 m of the center of the target pixel;
(iii) computation of regression coefficients interpolated for the previously computed running averages.

A few iterations are usually sufficient to obtain a converged set of $\bar{\tau}$ and $\bar{r}_{\rm eff}$.

Figure 3.15 shows initial and retrieved cloud quantities from the IPA and MSN methods. Smoothing and roughening has a big influence on pixel-averaged optical thickness in the IPA method. Negative and positive mean biases exist for high and low solar elevations, respectively. The IPA error in $\bar{\tau}$ for $\bar{\tau} > 30$ is very large (>100%) for low solar elevations. In addition, large positive error is present

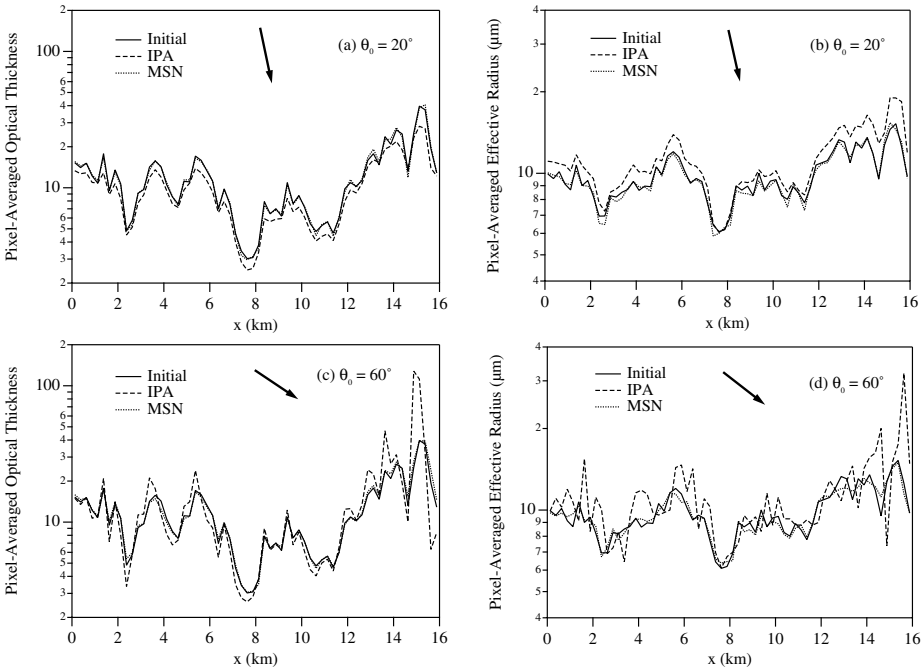



**Fig. 3.15.** Retrieved parameters from the IPA and MSN methods, and the initial distributions of (a) and (c) optical thickness and (b) and (d) effective radius. Retrieval simulations used 3D nadir radiances as virtually observed radiances.

at the sunny parts of clouds because of enhanced direct beam illumination. Large negative bias appears at shadowed sides. In contrast, the MSN method yields an optical thickness that is distributed closely around the initial value, with almost no mean bias. In addition, the MSN method successfully corrects a phase lag in optical thickness fluctuations when roughening occurs. The effective radius from the IPA method has positive bias for both solar angles, with a significant fluctuation for low solar elevations. In contrast, the MSN retrieval of effective radius is highly accurate, although accurate fluctuations at small horizontal scale are difficult to obtain even with the MSN method, especially for low solar elevations.

Figure 3.16 compares retrieval errors for the IPA, MSL, and MSN methods. The RMSE in optical thickness is similar to that in the previous case study (Fig. 3.14). The MSN method outperforms the other methods. The RMSE in the MSN method is smaller than in the IPA by a factor of 3 for $\theta_0 = 20°$ and 10 for $\theta_0 = 60°$. MSN error is about 50% of the MSL error. The relative error in effective radius with the IPA method is 15–25% and 30–40% for $\theta_0 = 20°$ and 60°, respectively. That error is reduced in the MSN method to 4–7% and 5–8%, respectively. The RMSE of the retrieved effective radius using MSN is similar to that from the MSL method when $\theta_0 = 20°$. However, the MSN error is three times smaller than the MSL error when the sun is low ($\theta_0 = 60°$). Retrieval errors

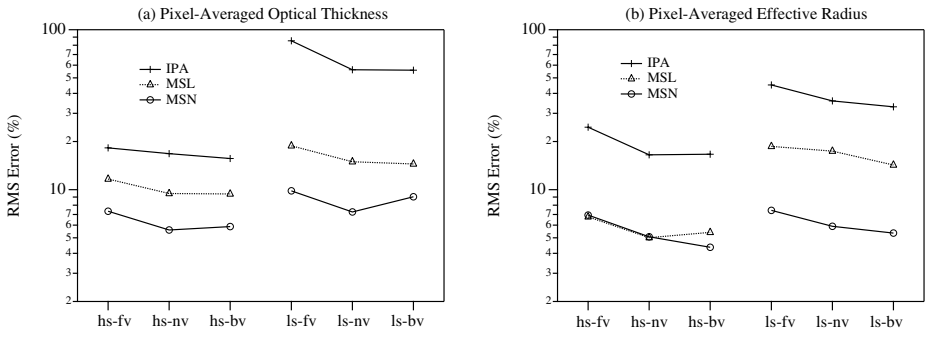



**Fig. 3.16.** As in Fig. 3.13, but for the retrieval of optical thickness and effective droplet radius.

for both the optical thickness and effective radius can be reduced significantly using the MSN method. Iwabuchi and Hayasaka (2003) assessed the influences of a few assumptions in the cloud modeling on the retrieval performance of the empirical inverse model. They found that the MSN method outperforms the IPA method, even given the uncertainties in the cloud modeling. In other words, some 3D radiative effects that vary with situations can be well modeled in the empirical inverse model.

### 3.5.4 Advanced method

The empirical model in the previous section estimates IPA radiance from observed (3D) radiances. In contrast, a model can be developed to calculate 3D radiances from the IPA radiances. This method can also be used to retrieve cloud parameters using an iterative procedure:

(i) Initial estimates of IPA radiances are equated as the observed radiances.
(ii) 3D radiance is computed from the IPA radiance using the empirical model.
(iii) IPA radiance estimates are modified by comparing 3D radiances and observed radiances.
(iv) Test for convergence (the 3D radiances are the same as the observations). If true, then finish. Otherwise, return to step (ii).
(v) Finally, cloud properties are inverted from the estimated IPA radiances.

Zinner et al. (2006) proposed a similar iterative algorithm for optical thickness retrieval, applying direct 3D radiative transfer calculations instead of step (ii) in the above. Although the 3D radiative transfer codes may be easily used in retrieval algorithms in the future, quicker calculation methods are preferable with currently limited computational power. In the IPA-to-3D conversion, a 2D filter convolution on the IPA radiances can be used. This idea is similar to one shown by Várnai and Marshak (2003). The filter expresses solar-azimuth dependence of contributions of horizontal radiative transport to the emergent radiance. The filter should be designed to include 3D radiative effects, such as

diffusion, direct beam effects, and other minor effects such as backscattering. Diffusion is assumed to work isotropically in all azimuth directions. Direct beam and backscattering effects operate along solar azimuths, and the horizontal scale length is a function of solar zenith angle. Filter coefficients were determined by the least-squares method under different cloud conditions and for different averages of optical thickness and effective radius. Radiative transfer simulations of visible and near-infrared wavelengths using the Monte Carlo radiation model were used to empirically determine the filter.

Figure 3.17 shows the approximately computed 3D radiance using the filter, at approximately 60-m pixel resolution. A horizontal segment along the solar azimuth was sampled from a 2D image and is shown in the figure. The radiance computed with filtering (filtered IPA, denoted as FIPA) is well correlated with the 3D radiance, suggesting that 3D radiative effects (such as shadowing and illumination) are well modeled by the empirical filter. FIPA accuracy is significantly higher than that of the method without filtering (i.e., just IPA). This is important because the performance of the IPA-to-3D conversion influences the retrieval performance in the inversion algorithm. Fig. 3.18 demonstrates parameter retrievals using FIPA and IPA. Significant retrieval error using IPA was absent for the FICA retrieval. The high-resolution data used (about 60 m) yields better estimates of cloud properties using a 2D filter that is adaptive to the solar azimuth of the case.

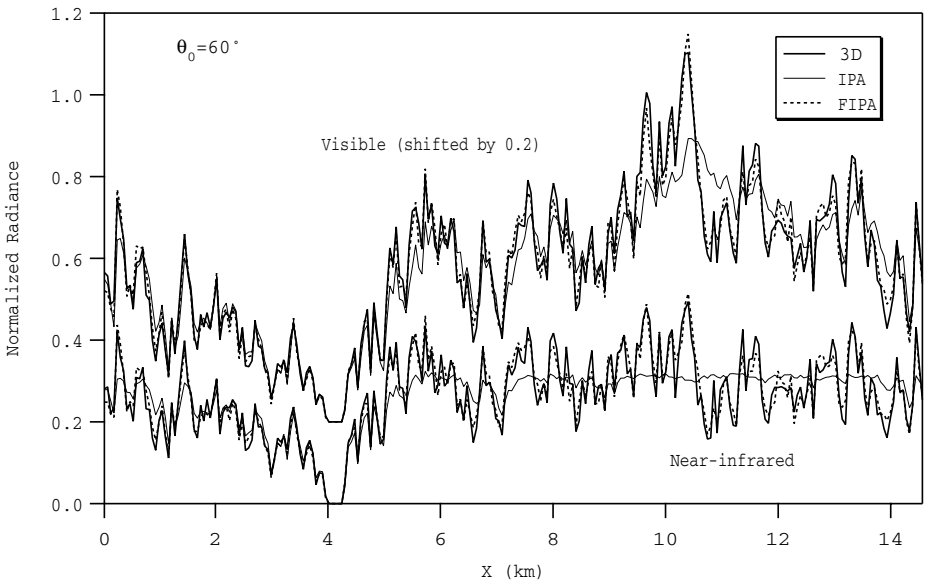



**Fig. 3.17.** Normalized nadir radiances computed by the 3D model, IPA, and by an empirical model that uses a filtering technique (FIPA), with approximately 60-m resolution.

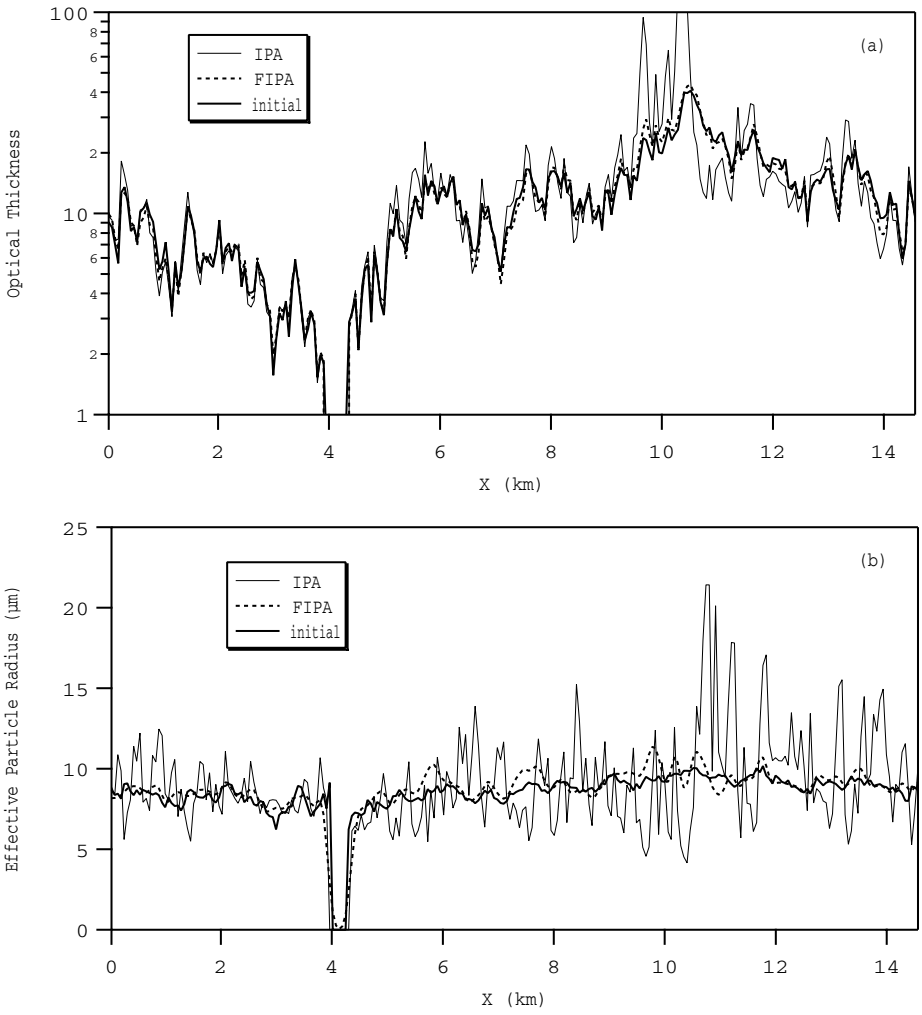



**Fig. 3.18.** Initial and retrieved cloud parameters. Retrieval simulations used the 3D radiances in Fig. 3.17 as virtually observed radiances.

## 3.6 Concluding remarks

The retrieval of optical thickness and effective radius in inhomogeneous clouds using satellite measurements has been considered. Several algorithms have been presented that yield improved estimates of cloud parameters by accounting for 3D radiative effects.

The first approach is a statistical correction of cloud parameters that are retrieved using 1D radiative transfer. We also examined the 3D radiative effects on the statistical quantities of optical thickness retrieved with independent pixel approximation (IPA) from visible-wavelength data with 1-km resolution.

Inhomogeneous clouds are characterized by two statistical quantities: the mean ($M$) and variance ($V$) of the logarithm of the optical thickness, where $V$ represents a degree of horizontal inhomogeneity. These statistical quantities are most important in determining major 3D radiative effects on the optical thickness retrieval. The optical thickness retrieved with IPA was biased, and these biases in the moments were investigated. Retrieval of optical thickness is insufficiently accurate without cloud-top geometrical roughness information, especially for off-nadir views with oblique sun. Darkening of mean reflectance occurs for forward-scattering viewing geometry because of cloud side shadowing. Similarly, brightening occurs for back-scattering-view geometry because of cloud side illumination. Therefore, optical thickness retrieval should be restricted to close-nadir-view geometry if conventional IPA is applied to oblique sun angles. The effects of radiative smoothing and roughening dominate for overhead sun and oblique sun, respectively. This change in dominance may yield an apparent solar zenith angle dependence in the cloud inhomogeneity parameter estimated from the IPA-based retrieval. Solar angle dependence may produce an unrealistic latitudinal and seasonal dependence of optical thickness in cloud climatologies. A correction for the influence of cloud horizontal inhomogeneity is therefore warranted. Pixel-by-pixel parameter estimates have errors for low-resolution data because sub-pixel information is missing in spite of the fact that the information is important for interpreting the pixel radiance. However, estimates of reasonable cloud parameter statistics are more important even if pixel-by-pixel estimates are not accurate. From this perspective, the correction method of moments is a good approach for remote sensing of inhomogeneous clouds.

The second approach considered pixel-by-pixel retrieval, increasing the data used for a retrieval. An empirical model can be used to retrieve pixel-averaged cloud optical thickness and effective droplet radius. IPA pixel radiances at visible and near-infrared wavelengths are expressed in the empirical model using regression formulae with respect to multi-spectral 3D radiances at the target pixel and neighboring pixels. Even for cloud optical thickness retrievals, which often use single-spectral visible wavelength data, multi-spectral and non-local data improved retrieval performance. Training was done under a variety of conditions (e.g., average optical thickness and effective radius) to generalize the retrieval method because 3D radiative effects vary significantly as conditions vary. The empirical inversion model performed significantly better than the IPA inversion. Although vertical cloudy columns were assumed in the cloud modeling, this type of algorithm will not predict accurate single pixel properties if shear is present in the cloud. However, the algorithm would be useful in correcting some artifacts that appear in the retrieved cloud parameters using the IPA.

Parameter retrieval that accounts for 3D effects is feasible if geometric and microphysical properties of target clouds are well modeled. The assumption of inhomogeneous clouds is obviously more realistic than the plane-parallel homogeneous assumption. Better realism results when assumptions for the geometric and microphysical properties of clouds are more sophisticated. Data from in situ observations or output from cloud/eddy resolving model simulations may be useful for this purpose. Cloud top roughness especially plays a major role in

determining 3D radiative effects (Loeb et al., 1998; Várnai and Davies, 1999). Improvements in assumptions used in the retrieval algorithm will allow more reasonable estimates of cloud properties to be obtained from satellite measurements.

## References

Barker, H. W., and J. A. Davies, 1992: Cumulus cloud radiative properties and the characteristics of satellite radiance wavenumber spectra. *Remote Sens. Environ.*, **42**, 51–64.

Barker, H. W., and D. Liu, 1995: Inferring optical depth of broken clouds from Landsat data. *J. Climate*, **8**, 2620–2630.

Barker, H. W., 1996: Estimating cloud field albedo using one-dimensional series of optical depth. *J. Atmos. Sci.*, **53**, 2826–2836.

Boers, R., J. D. Spinhirne, and W. D. Hart, 1988: Lidar observations of the fine-scale variability of marine stratocumulus clouds. *J. Appl. Meteor.*, **27**, 797–810.

Bower, K. N., T. W. Choularton, J. Latham, J. Nelson, M. B. Baker, and J. Jensen, 1994: A parameterization of warm clouds for use in atmospheric general circulation models. *J. Atmos. Sci.*, **51**, 2722–2732.

Brenguiter, J.-L., H. Pawlowska, L. Schüller, R. Preusker, J. Fischer, and Y. Fouquart, 2000: Radiative properties of boundary layer clouds: Effective droplet radius versus number concentration. *J. Atmos. Sci.*, **57**, 803–821.

Cahalan, R. F., W. Ridgway, W. J. Wiscombe, T. L. Bell, and J. B. Snider, 1994a: The albedo of fractal stratocumulus clouds. *J. Atmos. Sci.*, **51**, 2434–2455.

Cahalan, R. F., W. Ridgway, and W. J. Wiscombe, 1994b: Independent pixel and Monte Carlo estimates of the stratocumulus albedo. *J. Atmos. Sci.*, **51**, 3776–3790.

Chambers, L. H., B. A. Wielicki, and K. F. Evans, 1997: Accuracy of the independent pixel approximation for satellite estimates of oceanic boundary layer cloud optical depth. *J. Geophys. Res.*, **102**, 1779–1794.

Davis, A. B., A. Marshak, R. F. Cahalan, and W. J. Wiscombe, 1997: The Landsat scale break in stratocumulus as a three-dimensional radiative transfer effect: Implications for cloud remote sensing. *J. Atmos. Sci.*, **54**, 241–260.

Drake, F., 1993: Global cloud cover and liquid water path from ISCCP C2 data. *Int. J. Climatol.*, **13**, 581–605.

Evans, 1993: Two-dimensional radiative transfer in cloudy atmosphere: The spherical harmonic spatial grid method. *J. Atmos. Sci.*, **50**, 3111–3124.

Faure, T., H. Isaka, and B. Guillemet, 2001: Neural network retrieval of cloud parameters of inhomogeneous and fractional clouds: Feasibility study. *Remote Sens. Environ.*, **77**, 123–138.

Faure, T., H. Isaka, and B. Guillemet, 2002: Neural network retrieval of cloud parameters from high-resolution multispectral radiometric data: A feasibility study. *Remote Sens. Environ.*, **80**, 285–296.

Han, Q., W. B. Rossow, and A. A. Lacis, 1994: Near-global survey of effective droplet radii in liquid water clouds using ISCCP data. *J. Climate*, **7**, 465–497.

Hayasaka, T., M. Kuji, and M. Tanaka, 1994: Air truth validation of cloud albedo estimated from NOAA advanced very high resolution radiometer data. *J. Geophys. Res.*, **99**, 18 685–18 693.

Iwabuchi, H., 2000: Effects of cloud horizontal inhomogeneity on optical remote sensing of cloud parameters. Doctor thesis, Tohoku University, Sendai, Japan, 104 pp.

Iwabuchi, H., and T. Hayasaka, 2002: Effects of cloud horizontal inhomogeneity on the optical thickness retrieved from moderate-resolution satellite data. *J. Atmos. Sci.*, **59**, 2227–2242.

Iwabuchi, H. and T. Hayasaka, 2003: A multi-spectral non-local method for retrieval of boundary layer cloud properties from optical remote sensing data. *Remote Sens. Environ.*, **88**, 294–308.

Iwabuchi, H., 2006: Efficient Monte Carlo methods for radiative transfer modeling. *J. Atmos. Sci.*, **63**, 2324–2339.

Kikuchi, K., Y. Asuma, T. Taniguchi, M. Kanno, M. Tanaka, T. Hayasaka, T. Takeda, and Y. Fujiyoshi, 1993: Structure and reflectance of winter maritime stratocumulus clouds. *J. Meteor. Soc. Japan*, **71**, 715–731.

Loeb, N. G., and J. A. Coakley, Jr, 1998: Inference of marine stratus cloud optical depth from satellite measurements: Does 1D theory apply? *J. Climate*, **11**, 215–233.

Loeb, N. G., T. Várnai, and D. M. Winker, 1998: Influence of subpixel-scale cloud-top structure on reflectances from overcast stratiform cloud layers. *J. Atmos. Sci.*, **55**, 2960–2973.

Marshak, A., A. Davis, W. J. Wiscombe, and R. F. Cahalan, 1995a: Radiative smoothing in fractal clouds. *J. Geophys. Res.*, **100**, 26 247–26 261.

Marshak, A., A. Davis, W. J. Wiscombe, and G. Titov, 1995b: The verisimilitude of the independent pixel approximation used in cloud remote sensing. *Remote Sens. Environ.*, **52**, 71–78.

Marshak, A., A. Davis, R. F. Cahalan, and W. J. Wiscombe, 1998: Nonlocal independent pixel approximation: Direct and inverse problem s. *IEEE Trans. Geosc. and Remote Sens.*, **36**, 192–205.

Marshak, A., A. B. Davis, W. J. Wiscombe, and R. F. Cahalan, 1999: Horizontal radiative fluxes in clouds and accuracy of the independent pixel approximation at absorbing wavelengths. *Geophys. Res. Lett.*, **26**, 1585–1588.

Marshak A., S. Platnick, T. Várnai, G. Wen, R. F. Cahalan, 2006: Impact of three-dimensional radiative effects on satellite retrievals of cloud droplet sizes. *J. Geophys. Res.*, **111**, D09207, doi:10.1029/2005JD006686.

Minnis, P., P. W. Heck, D. F. Young, C. W. Fairall, and J. B. Snider, 1992: Stratocumulus cloud properties derived from simultaneous satellite and island-based instrumentation during FIRE. *J. Appl. Meteor.*, **31**, 317–339.

Nakajima, T., and M. D. King, 1990: Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements. Part I: Theory. *J. Atmos. Sci.*, **47**, 1878–1893.

Nakajima, T. Y., and T. Nakajima, 1995: Wide-area determination of cloud microphysical properties from NOAA AVHRR measurements for FIRE and ASTEX regions. *J. Atmos. Sci.*, **52**, 4043–4059.

Oreopoulos, L., and R. Davies, 1998a: Plane parallel albedo biases from satellite observations. Part I: Dependence on resolution and other factors. *J. Climate*, **11**, 919–932.

Oreopoulos, L., and R. Davies, 1998b: Plane parallel albedo biases from satellite observations. Part II: Parameterizations for bias removal. *J. Climate*, **11**, 933–944.

Oreopoulos, L., A. Marshak, R. F. Cahalan, and G. Wen, 2000: Cloud three-dimensional effects evidenced in Landsat spatial power spectra and autocorrelation functions. *J. Geophys. Res.*, **105**, 14 777–14 788.

Radke, L. F., J. A. Coakley, and M. D. King, 1989: Direct and remote sensing observations of the effects of ships on clouds. *Science*, **246**, 1146–1149.

Rosenfeld, D., 2000: Suppression of rain and snow by urban and industrial air pollution. *Science*, **287**, 1793–1796.

Rossow, W. B., and R. A. Schiffer, 1991: ISCCP cloud data products. *Bull. Amer. Meteor. Soc.*, **72**, 2–20.

Szczap, F., H. Isaka, M. Saute, B. Guillemet, and Y. Gour, 2000: Inhomogeneity effects of 1D and 2D bounded cascade model clouds on their effective radiative properties. *Phys. Chem. Earth*, **25**, 83–89.

Titov, G. A., 1998: Radiative horizontal transport and absorption in stratocumulus clouds. *J. Atmos. Soc.*, **55**, 2549–2560.

Tselioudis, G., W. B. Rossow, and D. Rind, 1992: Global patterns of cloud optical thickness variation with temperature. *J. Climate*, **5**, 1484–1495.

Vali, G., R. D. Kelly, J. French, S. Haimov, and D. Leon, 1998: Finescale structure and microphysics of coastal stratus. *J. Atmos. Sci.*, **55**, 3540–3564.

Várnai, T., and R. Davies, 1999: Effects of cloud heterogeneities on shortwave radiation: Comparison of cloud-top variability and internal heterogeneity. *J. Atmos. Sci.*, **56**, 4206–4224.

Várnai T., A. Marshak, 2003: A method for analyzing how various parts of clouds influence each other's brightness. *J. Geophys. Res.*, **108** (D22), 4706, doi:10.1029/2003JD003561.

Zinner T., B. Mayer, M. Schröder, 2006: Determination of three-dimensional cloud structures from high-resolution radiance data. *J. Geophys. Res.*, **111**, D08204, doi:10.1029/2005JD006062.

Zuidema, P., and K. F. Evans, 1998: On the validity of the independent pixel approximation for boundary layer cloud observed during ASTEX. *J. Geophys. Res.*, **103**, 6059–6074.

# 4 Raman lidar remote sensing of geophysical media

Aleksey V. Malinka

## 4.1 Introduction

Lidars are equipment, consisting of a laser and a photo-receiver, that measures the backward scattering of light. They appeared in the 1960s (Fiocco and Smullin, 1963), i.e., immediately after the invention of the laser, and since then they have been actively used in the problems of natural media monitoring. Lidars are of great use in providing atmosphere and ocean pollution control, in control of atmospheric gases, and in measuring meteorological and climate characteristics. Generation of a beam of high power and small angular divergence makes the great advantage of lidars over projector sounding, having existed before. The possibility of accurate wavelength tuning, as well as spectral return measuring, allows the determination of the chemical composition of the atmosphere and the biochemical composition of the ocean. Thanks to the measurement of scattered light polarization degree one can learn about the shape of scatterers. Furthermore, as lasers are able to generate powerful pulses of short duration, there appears the possibility of measuring time-dependent returns, i.e., measuring not only the integral optical characteristics of a medium, but also their spatial distribution. These features made lidars a powerful tool in the investigation of geophysical media.

Besides the usual, elastic, lidar sounding, recently the methods of inelastic, particularly Raman, lidar sounding, have been developed. These methods provide a wide range of new possibilities (Ansmann et al., 1990, 1992a; Reichardt et al., 1996). Raman lidar sounding implies sounding based on receiving the signal of Raman scattering, in which the scattered light shifts frequency to a value equal to the eigenfrequency of a molecule of the scattering substance. The Raman lidar return is proportional to the scattering substance concentration. Thus, measuring the Raman lidar return allows one to establish the presence of substances whose eigenfrequencies correspond to lines in the measured spectrum. On the other hand, well-known characteristics of Raman scattering by stable components (nitrogen in the atmosphere and water in the ocean) permits the use of the Raman lidar return as a reference (calibration) signal for other measurements.

The growing interest in Raman sounding methods is not only due to their advantages and their additional possibilities. It is also due to the fact that the technical capabilities to fix the weak light signal have been steadily improving. For example, thirty years ago Raman scattering by atmospheric nitrogen in a pulse lidar regime could be fixed from altitudes of up to 3 km (Cohen et al., 1978). Nowadays the Raman lidar return is accurately fixed from altitudes of 30 km and more (Ansmann et al., 1992b; Sherlock et al., 1999a). Even the possibility of spaceborne Raman lidar measurement is discussed (Girolamo et al., 2006).

Joint use of both elastic and inelastic, including Raman, scattering suggests a future trend to discover new ways of sounding and measuring properties of various turbid media.

Nowadays, most of the methods of medium characteristics retrieval use the lidar equation within the framework of single scattering approximation. They consider multiple scattering as an interference to be suppressed. However, as many authors have mentioned, in most geophysical media (such as clouds, dense aerosols or seawater) multiple scattering plays the key role in lidar return forming (Bruscaglioni et al., 1999; Eloranta, 1972; Reichardt et al., 2000; Weinman and Shipley, 1972). There are also lidar systems, in which the signal is completely defined by multiple scattering. These are multiple-field-of-view lidars (that measure light fluxes at several angle intervals) (Bissonnette and Hutt, 1990; Roy et al., 1997) and imaging lidars (that measure the irradiance distribution in the focal plane of the receiver, i.e., the angular distribution of radiance at the entrance of the receiver optics) (Muscari et al., 1996). In both cases, the problem is not to estimate the contribution of multiple scattering as some correction, but to describe it correctly in a qualitative and quantitative way.

The problem of multiple scattering in lidar measurements is a focus of interest of the international workshop MUSCLE (MUltiple SCattering in Lidar Experiments), which has succeeded both in simulation of multiple scattering and in its use in the inverse problem s for the case of elastic lidar sounding (Muscari et al., 1996; Zege et al., 2003a; Bruscaglioni et al., 1999). However, for a long time few attempts were made to simulate multiple scattering for Raman lidar sounding (Bruscaglioni et al., 1999; Wandinger, 1998) and practically no attempts were made to include it in the inverse problem.

This chapter gives a short review of the existing methods of Raman lidar sounding of geophysical media, and it presents the theory for the Raman lidar return with multiple scattering and the new methods of using multiple scattering to retrieve the microphysical characteristics of a light scattering medium.

## 4.2 Review of the existing methods of Raman lidar sounding

The idea of using Raman scattering in lidar sounding of geophysical media appeared in the 1960s. However its implementation was delayed, primarily because of the weak signal power. In the late 1970s, there appeared lidar measurements of Raman scattering from water while sounding ocean (Klyshko and Fadeev,

1978). Methods of fixing Raman scattering from atmospheric nitrogen were suggested by Egert et al. (1983). Nowadays there are a lot of lidar systems all over the world. These systems make it possible to measure Raman scattering not only from atmospheric nitrogen, but also from water vapour and other atmospheric gases of low concentration, not to mention seawater. Raman lidar is a common instrument to get information about the composition of the sounding medium, its temperature and humidity (for atmosphere), and about the presence of suspended particles (cloud droplets, aerosols, hydrosols), their chemical properties and sizes (Ansmann et al., 1990, 1992a, 1992b; Reichardt et al., 1996, 2000; Wandinger et al., 1995; Roy et al., 1997; Sherlock et al., 1999a, 1999b; Whiteman and Melfi, 1999).

This chapter describes the main modern methods of sounding and processing the Raman lidar return.

### 4.2.1 Lidar equation

The lidar equation is an expression that relates the lidar return $F(z)$ (the energy of light coming from the depth interval from $z$ to $z+\mathrm{d}z$ divided by $\mathrm{d}z$) to optical medium characteristics and lidar system parameters. Usually it is written in the framework of the single scattering approximation. For elastic lidar return it has the form:

$$F(z) = A\sigma(z,\pi)\,\exp\left(-2\int_0^z \varepsilon(z')\,\mathrm{d}z'\right)\,, \tag{4.1}$$

where $A$ is a calibration constant, including energetic characteristics of the lidar system, geometric parameters of the experiment and refractive index of sounding medium, $\sigma(z,\pi)$ is the backscattering coefficient, $\varepsilon(z)$ is the extinction coefficient, and $z$ is the sounding depth.

The time $t$ the photon arrives at the receiver is related to the depth $z$ the photon penetrates into the medium by the following expression:

$$t = 2\frac{H+z}{c}\,, \tag{4.2}$$

where $H$ is the distance from the lidar to the nearest medium border, $c$ is the speed of light in air.

The expression for the Raman lidar return in the framework of the single scattering approximation is:

$$F_R(z) = A\sigma_R(z,\pi)\,\exp\left(-\int_0^z \left[\varepsilon(z',\lambda_0)+\varepsilon(z',\lambda_R)\right]\mathrm{d}z'\right)\,, \tag{4.3}$$

where $\sigma_R(z,\pi)$ is the Raman backscattering coefficient, $\lambda_0$ is the initial wavelength, and $\lambda_R$ is the Raman shifted wavelength, the initial and shifted wavelengths being related as:

$$\lambda_R = \frac{1}{1/\lambda_0 - \delta\tilde{\nu}}\,, \tag{4.4}$$

where $\delta\tilde{\nu}$ is a Raman frequency shift, depending on an eigenfrequency of a molecule.

Unlike high-spectral-resolution lidars (Grund and Eloranta, 1991; Shipley et al., 1983; Piironen and Eloranta, 1994), Raman lidars have a receiver frequency band broader than the width of the appropriate Raman line, so all the light that is Raman scattered at the wavelength $\lambda_0$ is assumed to shift to the wavelength $\lambda_R$.

Equation (4.3) is the basic equation used to process Raman sounding data. The Raman backscattering coefficient $\sigma_R(z,\pi)$ is proportional to the concentration of scattering molecules. Therefore, the Raman lidar return is also proportional to the concentration of molecules, and so measuring the power of the Raman lidar return enables the estimation of the concentrations. However, such a straight way is not a very accurate one, because the extinction coefficient remains unknown. For real measurements, other methods are applied. They are discussed below.

### 4.2.2 The method of Raman reference signal

The idea of the method is to measure two different Raman signals simultaneously. In this case, one signal is used to measure concentration, and the other is used as a reference signal (Ansmann et al., 1992a).

Let a laser pulse propagate into the atmosphere at the wavelength $\lambda_0$ and let the receiver fix two Raman scattering signals, one, for example, from atmospheric nitrogen and the other from water vapour. The powers of these lidar returns are

$$F_{H_2O} = A\sigma_R^{H_2O}(z,\pi)\,\exp\left(-\int_0^z [\varepsilon(z',\lambda_0)+\varepsilon(z',\lambda_{H_2O})]\,\mathrm{d}z'\right), \tag{4.5}$$

$$F_{\mathrm{N}_2} = A\sigma_R^{\mathrm{N}_2}(z,\pi)\,\exp\left(-\int_0^z [\varepsilon(z',\lambda_0)+\varepsilon(z',\lambda_{\mathrm{N}_2})]\,\mathrm{d}z'\right). \tag{4.6}$$

Assuming that wavelengths $\lambda_{\mathrm{H_2O}}$ and $\lambda_{\mathrm{N}_2}$ differ slightly, i.e., the difference of the extinction coefficients at these wavelengths is negligible: $\varepsilon(z,\lambda_{\mathrm{H_2O}}) \approx \varepsilon(z,\lambda_{\mathrm{N}_2})$, we get for the ratio of these returns:

$$\frac{F_{\mathrm{H_2O}}}{F_{\mathrm{N}_2}} = \frac{\sigma_R^{\mathrm{H_2O}}(z,\pi)}{\sigma_R^{\mathrm{N}_2}(z,\pi)}\,. \tag{4.7}$$

The Raman backscattering coefficient is equal to the product of the Raman backscattering cross-section $Q(\pi)$ of one molecule by the concentration of molecules $n(z)$:

$$\sigma_R(z,\pi) = n(z)Q(\pi)\,. \tag{4.8}$$

So, we get:

$$\frac{F_{\mathrm{H_2O}}}{F_{\mathrm{N}_2}} = \frac{Q^{\mathrm{H_2O}}(\pi)}{Q^{\mathrm{N}_2}(\pi)}\,\frac{n^{\mathrm{H_2O}}(z)}{n^{\mathrm{N}_2}(z)}\,. \tag{4.9}$$

Whereas the nitrogen concentration is a stable quantity, in fact, Eq. (4.9) gives a straightforward way to measure the water vapour concentration profile (mixing

ratio). The humidity profile could be derived using the well-known formulas if an additional measurement of a temperature profile is carried out (Mattis et al., 2002).

### 4.2.3 The method of measuring an aerosol extinction profile with a Raman lidar

Raman scattering by nitrogen is also applied to investigation of aerosols (Ansmann et al., 1990; Reichardt et al., 1996). In this case, the power of the Raman lidar return can be written as:

$$F_R(z) = A\,\sigma_R^{\mathrm{N}_2}(z,\pi)\,\exp\left(-\int_0^z [\varepsilon(z',\lambda_0)+\varepsilon(z',\lambda_R)]\,\mathrm{d}z'\right), \tag{4.10}$$

backscattering being due to nitrogen (with a wavelength shift) and extinction being due to aerosol. If spectral difference of aerosol properties at $\lambda_0$ and $\lambda_R$ is negligible (it is so, if sounding wavelengths do not match the absorption lines of aerosols or atmospheric gases), then we get:

$$\varepsilon(z) = -\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}z}\ln\left(\frac{F(z)}{A\,\sigma_R^{\mathrm{N}_2}(z,\pi)}\right), \tag{4.11}$$

where $\sigma_R^{\mathrm{N}_2}(z,\pi)$ is a known function (e.g., from standard atmospheric models).

This technique has a significant advantage over the elastic scattering technique, whereas, instead of two variables (the aerosol backscattering and extinction coefficients), only one variable is unknown, namely, the aerosol extinction coefficient, which is easily found from Eq. (4.11). The backscattering coefficient is then found as a ratio of elastic lidar return to Raman one. In this way the problem uncertainty (lidar ratio) is significantly reduced. However, the appearing drawback is a logarithmic derivative, which in the presence of experimental errors is a mathematically incorrect procedure and needs regularization.

### 4.2.4 The Raman DIAL method

Raman DIAL method is analogous to the conventional DIAL method, but, instead of two elastic lidar returns at different wavelengths, one Raman lidar return is used (Reichardt et al., 1996; Tomasi et al., 2001). This method is generally used to measure the concentration of gases in atmosphere. Raman scattering is from nitrogen or oxygen (or both), and the wavelengths are tuned in such a way that the initial wavelength matches the absorption band of the gas to be investigated and the shifted wavelength matches the band where the atmosphere is clear.

Thus, for example, when investigating the ozone profile, the initial wavelength is set in the near-UV, where ozone absorption is strong, while the shifted wavelength reaches the visible range, where atmosphere does not absorb sensibly.

Lidar return is written in the form:

$$F(z) = A\,\sigma_R^{\mathrm{N_2}}(z,\pi)\,\exp\left(-\int_0^z \alpha^{\mathrm{O_3}}(z',\lambda_0)\,\mathrm{d}z' - 2\int_0^z \sigma(z')\,\mathrm{d}z'\right), \qquad (4.12)$$

where $\alpha^{\mathrm{O_3}}$ is the ozone absorption coefficient:

$$\alpha^{\mathrm{O_3}} = Q_a^{\mathrm{O_3}}(\lambda_0)\,n^{\mathrm{O_3}}(z)\,, \qquad (4.13)$$

where $Q_a^{\mathrm{O_3}}$ is the absorption cross-section of an ozone molecule.

So, the concentration $n^{\mathrm{O_3}}(z)$ can be easily found:

$$n^{\mathrm{O_3}}(z) = -\frac{1}{Q_a^{\mathrm{O_3}}(\lambda_0)}\left\{\frac{\mathrm{d}}{\mathrm{d}z}\ln\left(\frac{F(z)}{A\,\sigma^{\mathrm{N_2}}(z,\pi)}\right) + 2\sigma(z)\right\}. \qquad (4.14)$$

Reichardt et al. (1996) claim that the Raman DIAL method is less sensitive to measurement errors than the conventional DIAL method and, therefore, it is much more accurate. The Raman DIAL method is even more stable, if one measures the ratio of Raman signals from two stable components of atmosphere, namely nitrogen and oxygen.

### 4.2.5 The method of rotational Raman scattering for determining the thermodynamic characteristics of atmosphere

The use of the rotational spectrum of Raman scattering as a way of measuring temperature was apparently suggested first by Cooney (1972). In most cases of Raman lidar sounding of atmosphere the frequency shift, corresponding to the main vibrational transition, is used. The energy of vibrational transition is much greater than the energy of heat motion of molecules. This makes the vibrational transition signal independent of temperature and, therefore, convenient to use as a reference. The energy of pure-rotational transition, on the other hand, is of the same order of the heat motion energy and, therefore, the rotational transition signal can be used to measure temperature profile (Mattis et al., 2002).

Let, for example, one measure two rotational Raman scattering signals, corresponding to the quantum numbers $j_1$ and $j_2$:

$$F_R^{j_1}(z) = A\sigma_R^{\mathrm{N_2}}(j_1,z,\pi)\,\exp\left(-\int_0^z [\varepsilon(z',\lambda_0) + \varepsilon(z',\lambda_R^{j_1})]\,\mathrm{d}z'\right), \qquad (4.15)$$

$$F_R^{j_2}(z) = A\sigma_R^{\mathrm{N_2}}(j_2,z,\pi)\,\exp\left(-\int_0^z [\varepsilon(z',\lambda_0) + \varepsilon(z',\lambda_R^{j_2})]\,\mathrm{d}z'\right). \qquad (4.16)$$

Whereas the frequency shift at rotational transition is much less than that at vibrational transition, in this case the difference between extinction coefficients at initial and shifted wavelengths can be neglected. So, the ratio of signals (4.15) and (4.16) gives:

$$\frac{F_R^{j_1}(z)}{F_R^{j_2}(z)} = \frac{\sigma_R^{\mathrm{N_2}}(j_1,z,\pi)}{\sigma_R^{\mathrm{N_2}}(j_2,z,\pi)}\,, \qquad (4.17)$$

where the ratio of scattering coefficients is proportional to the ratio of concentrations of molecules, belonging to different energy levels, and is described with the Boltzmann distribution:

$$\frac{F_R^{j_1}(z)}{F_R^{j_2}(z)} = \exp\left(\frac{C_1}{T(z)} + C_2\right). \tag{4.18}$$

Constants $C_1$ and $C_2$ are usually found through calibration of lidar data with that of a radiosonde. The temperature profile is then found from the formula:

$$T(z) = \frac{C_1}{\ln\left(F_R^{j_1}(z)/F_R^{j_2}(z)\right) - C_2}. \tag{4.19}$$

It is interesting to note that due to negligible frequency shift the method of rotational Raman scattering is not affected in most cases by multiple scattering.

## 4.3 The Raman lidar return with regard to multiple scattering

### 4.3.1 Problem statement

Let us consider the following problem. The sounding medium is a plane-parallel turbid layer. A lidar is located at the distance $H$ from the nearest border of a layer. We use the Cartesian coordinate system with the OZ axis, perpendicular to the border and directed into the medium. The two-dimensional vector $\mathbf{r}$ describes the coordinates in the plane $(x, y)$. The two-dimensional vector $\mathbf{n}$ is a projection of a unit vector, pointing the direction of light propagation, onto the plane $(x, y)$. The lidar is situated at the point $(-H, 0, 0)$. The spatial-angular distribution of the source radiance and the diagram of the receiver sensitivity are given by functions $\varphi_{\rm src}(\mathbf{r}, \mathbf{n})$ and $\varphi_{\rm rec}(\mathbf{r}, \mathbf{n})$, which are normalized as following:

$$\int \varphi_{\rm src}(\mathbf{r}, \mathbf{n})\, \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{n} = 1\,, \tag{4.20}$$

$$\int \varphi_{\rm rec}(\mathbf{r}, \mathbf{n})\, \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{n} = S_{\rm rec}\Omega_{\rm rec}\,, \tag{4.21}$$

where $S_{\rm rec}$ and $\Omega_{\rm rec}$ are the area and the solid angle of the receiver, respectively.

The scattering medium is characterized by the extinction coefficient $\varepsilon(\lambda, z)$, the elastic scattering coefficient $\sigma(\lambda, z)$, the elastic scattering phase function $P(\lambda, \theta)$, the Raman scattering coefficient $\sigma_R$ (total), and the Raman scattering phase function $P_R(\theta)$. Phase functions are normalized as:

$$\frac{1}{2}\int_0^\pi P(\theta)\sin\theta\, \mathrm{d}\theta = 1\,. \tag{4.22}$$

The initial wavelength is $\lambda_0$, the Raman shifted wavelength is $\lambda_R$ that is defined by Eq. (4.4).

### 4.3.2 General solution

The model of small-angle quasi-single scattering, which accounts for multiple forward scattering and single backward scattering, has been successfully used to describe multiple scattering in elastic lidar sounding (Katsev et al., 1997; Zege et al., 1995). This approach is bound up with the fact that phase functions of real geophysical media, such as clouds, aerosols, and seawater, are strongly peaked in the forward direction. As a result, the angular spectrum of scattered light from a laser source is peaked in the forward direction too. This means that in most of the problems of lidar sounding one can consider just the only one event of scattering to large angles. According to this approximation, the expression for the lidar return has a form (Katsev et al., 1997):

$$F(z) = W_0 \frac{\sigma(\lambda, z)}{4\pi} \int \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{n}'\, \mathrm{d}\mathbf{n}''\, I_{\mathrm{src}}(z, \mathbf{r}, \mathbf{n}') P(z, \pi - |\mathbf{n}' - \mathbf{n}''|) I_{\mathrm{rec}}(z, \mathbf{r}, \mathbf{n}'')\,, \tag{4.23}$$

where $W_0$ is a laser pulse energy, $I_{\mathrm{src}}(z, \mathbf{r}, \mathbf{n})$ and $I_{\mathrm{rec}}(z, \mathbf{r}, \mathbf{n})$ are the angular distributions of the radiance at the point $(z, \mathbf{r})$, due to the *real* and *fictitious* continuous wave sources of unit power with the spatial-angular diagrams $\varphi_{\mathrm{src}}(\mathbf{r}, \mathbf{n})$ and $\varphi_{\mathrm{rec}}(\mathbf{r}, \mathbf{n})$, respectively. Functions $I_{\mathrm{src}}(z, \mathbf{r}, \mathbf{n})$ and $I_{\mathrm{rec}}(z, \mathbf{r}, \mathbf{n})$ satisfy the radiative transfer equation and can be found within the small-angle approximation.

A real lidar records temporal dependencies of the lidar return power $F(t)$, $t$ being the photon arrival time. This approach neglects the temporal spread of small-angle photons and associates the photon arrival time $t$ with the depth $z$, which the photons return from, by Eq. (4.2). As a matter of fact, the forward pulse stretching can violate the relation (4.2) and affect the lidar return (McLean et al., 1998). However, both theoretical estimations and computer simulations show that this feature has to be regarded only for sounding layers of large optical thicknesses (Zege et al., 2001).

In order to find the Raman lidar return under strong influence of multiple scattering, we should note first that the Raman scattering coefficient is several orders less than the elastic scattering coefficient, for example, for clouds, their ratio is about $10^{-6}$ or even less (Reichardt et al., 1996). This means that only one Raman scattering event is enough to account for. On the contrary, the elastic scattering should be considered as multiple one.

By the manner of forming light field, elastic scattering can be divided into small-angle scattering and backward (large-angle) scattering. According to Katsev et al. (1997), because of the forward peak of elastic scattering, backscattering can be considered just once in lidar sounding. Hence, multiple scattering appears in small-angle elastic scattering only.

Having separated in such a way the process of forming light field into forward elastic multiple scattering (FES), backward elastic scattering (BES), forward Raman scattering (FRS), and backward Raman scattering (BRS), and considering just one event of Raman and one event of backward elastic scattering, we arrive at the conclusion that the Raman lidar return could be formed by the following processes:

1. FES – BRS – FES,
2. FES – FRS – FES – BES – FES,
3. FES – BES – FES – FRS – FES.

The second and the third processes are obviously equivalent, so we will merge them into one process: forward Raman – backward elastic (FRS – BES). The first process is principally different: it includes the Raman backscattering event and there is no elastic backscattering in it. As it happens, the FRS – BES process can be neglected comparing to the first process (BRS). Let us display it, using the double-scattering pattern. The relative contribution of double-scattering $F^{(2)}(z)$, as compared with the single scattering $F^{(1)}(z)$, in the case of a homogeneous medium equals (Cohen et al., 1978; Eloranta, 1972; Samokhvalov, 1979):

$$\frac{F^{(2)}(z)}{F^{(1)}(z)} = \frac{\sigma(\lambda)\ (z+H)}{P(\pi)}\ \gamma_{\text{rec}} \int_0^{\pi/2} P(\gamma)P(\pi-\gamma)\,\mathrm{d}\gamma\,, \tag{4.24}$$

where $\gamma_{\text{rec}}$ is the receiver field-of-view (FOV).

Here we imply that the source is a mono-directional one and FOV is small enough:

$$\frac{z+H}{z}\gamma_{\text{rec}} \ll 1\,. \tag{4.25}$$

Noting that in our case double scattering consists of one elastic and one Raman scattering event, we get:

$$\frac{F^{(\text{el},R)}(z)}{F^{(R)}(z)} = \frac{(z+H)}{P_R(\pi)}\ \gamma_{\text{rec}} \left( \sigma(\lambda_R, z) \int_0^{\pi/2} P_R(\gamma)P(\pi-\gamma)\,\mathrm{d}\gamma + \sigma(\lambda_0, z) \int_0^{\pi/2} P(\gamma)P_R(\pi-\gamma)\,\mathrm{d}\gamma \right). \tag{4.26}$$

The first term in parentheses describes the FRS – BES process, and the second one describes the BRS process. The ratio of these two terms illustrates the contribution of the FRS – BES process, in comparison with the BRS process. This ratio can be estimated as a ratio of elastic phase function integral over backward hemisphere to that over forward hemisphere:

$$\int_{\pi/2}^{\pi} P(\gamma)\,\mathrm{d}\gamma \Big/ \int_0^{\pi/2} P(\gamma)\,\mathrm{d}\gamma\,.$$

This ratio is negligible for real geophysical media, because of the forward scattering peak. It is about $10^{-3}$ for the Cloud C.1 model (Deirmendjian, 1969) in the visible range.

So, the process FES – BRS – FES is the main contribution to light field forming. Or, if we decode it, light field forming goes the following way (see Fig. 4.1):

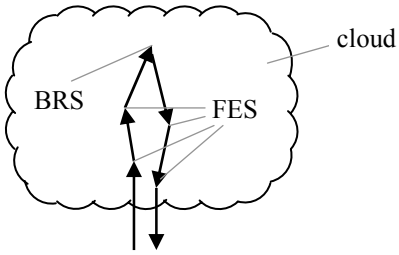



**Fig. 4.1.** The scheme of forming the Raman lidar return.

1. elastic small-angle multiple scattering, when photons travel out of the lidar into the medium,
2. the single event of Raman scattering in the backward direction,
3. elastic (but at the shifted wavelength) small-angle multiple scattering, when photons travel from the medium back to the lidar.

This scheme is a complete paradigm of the model (4.23) for elastic lidar sounding.

Thus the model (4.23) can be easily generalized to the case of Raman lidar sounding. To do so, one should

- replace the elastic scattering coefficient $\sigma(\lambda, z)$, followed by the integral, by the Raman scattering coefficient $\sigma_R(z)$,
- substitute the Raman scattering phase function $P_R(\pi - \theta)$ for the elastic backscattering phase function $P(\pi - \theta)$,
- use medium characteristics at wavelengths $\lambda_0$ and $\lambda_R$, while calculating functions $I_{\text{src}}(z, \mathbf{r}, \mathbf{n})$ and $I_{\text{rec}}(z, \mathbf{r}, \mathbf{n})$, respectively.

Finally we get the expression for the Raman lidar return:

$$F_R(z) = W_0 \frac{\sigma_R(z)}{4\pi} \int \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{n}'\, \mathrm{d}\mathbf{n}''\, I_{\text{src}}(\lambda_0, z, \mathbf{r}, \mathbf{n}') P_R(\pi - |\mathbf{n}' - \mathbf{n}''|) I_{\text{rec}}(\lambda_R, z, \mathbf{r}, \mathbf{n}'')\,. \tag{4.27}$$

This expression can be rewritten more compactly by introducing the effective light field radiance:

$$I_{\text{eff}}(z, \mathbf{r}, \mathbf{n}) = \int \mathrm{d}\mathbf{r}'\, \mathrm{d}\mathbf{n}'\, I_{\text{src}}(\lambda_0, z, \mathbf{r}', \mathbf{n}') I_{\text{rec}}(\lambda_R, z, \mathbf{r} + \mathbf{r}', \mathbf{n} + \mathbf{n}')\,. \tag{4.28}$$

Then Eq. (4.27) can be presented as:

$$F_R(z) = W_0 \frac{\sigma_R(z)}{4\pi} \int \mathrm{d}\mathbf{n} P_R(\pi - |\mathbf{n}|) I_{\text{eff}}(z, \mathbf{r} = 0, \mathbf{n})\,. \tag{4.29}$$

This expression is more compact and demonstrative. However we need to figure out the physical meaning of the effective radiance $I_{\text{eff}}$.

To do so, let us introduce the Fourier transform of the function $I_{\text{eff}}$ and the Hankel transform of the Raman backscattering function $P_R(\pi-\theta)$:

$$I_{\text{eff}}(z,\boldsymbol{\nu},\mathbf{p}) = \int \mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{n}\, I_{\text{eff}}(z,\mathbf{r},\mathbf{n}) \exp(-i\,\boldsymbol{\nu}\cdot\mathbf{r} - i\,\mathbf{p}\cdot\mathbf{n})\,, \tag{4.30}$$

$$P_R(p) = \frac{1}{2}\int_0^\infty P_R(\pi-\theta)\, J_0(p\,\theta)\,\theta\,\mathrm{d}\theta\,, \tag{4.31}$$

where $J_k(x)$ is the Bessel function of the $k$th order.

(Here, as always in the small-angle approximation, we assume that backscattering phase function $P_R(\pi-\theta)$ is equal to zero for $\theta$, which is greater than the prescribed value of the scattering angle $\theta_0$. We choose $\theta_0 = \pi/2$.)

According to the Parseval equality, we have:

$$F_R(z) = W_0\sigma_R(z)\int \frac{\mathrm{d}\boldsymbol{\nu}\,\mathrm{d}\mathbf{p}}{(2\pi)^4} P_R(p) I_{\text{eff}}(z,\boldsymbol{\nu},\mathbf{p})\,, \tag{4.32}$$

where

$$I_{\text{eff}}(z,\boldsymbol{\nu},\mathbf{p}) = I^*_{\text{src}}(\lambda_0,z,\boldsymbol{\nu},\mathbf{p})\, I_{\text{rec}}(\lambda_R,z,\boldsymbol{\nu},\mathbf{p})\,. \tag{4.33}$$

Here $I_{\text{src}}(\lambda_0,z,\boldsymbol{\nu},\mathbf{p})$ and $I_{\text{rec}}(\lambda_R,z,\boldsymbol{\nu},\mathbf{p})$ are the Fourier transforms of the functions $I_{\text{src}}(\lambda_0,z,\mathbf{r},\mathbf{n})$ and $I_{\text{rec}}(\lambda_R,z,\mathbf{r},\mathbf{n})$, respectively (sign * represents complex conjugation).

Within the small-angle approximation, the Fourier transform of light radiance, generated by the source with the diagram $\varphi(\mathbf{r},\mathbf{n})$ in a scattering medium, is equal to (Zege et al., 1991):

$$I(z,\boldsymbol{\nu},\mathbf{p}) = \varphi(\boldsymbol{\nu},\mathbf{p}+\boldsymbol{\nu}\,(z+H))\,\exp\!\left(-\int_0^z \left[\varepsilon(\xi)-\sigma(\xi)P^f(|\mathbf{p}+\boldsymbol{\nu}(z-\xi)|)\right]\,\mathrm{d}\xi\right) \tag{4.34}$$

where $\varphi(\boldsymbol{\nu},\mathbf{p})$ is the Fourier transform of the source diagram, and $P^f(p)$ is the Hankel transform of the forward scattering phase function:

$$P^f(p) = \frac{1}{2}\int_0^\infty P^f(\theta)\, J_0(p\,\theta)\,\theta\,\mathrm{d}\theta \tag{4.35}$$

Substitution of (4.33) and (4.34) into Eq. (4.32) gives:

$$\begin{aligned} F_R(z) \;=\; & W_0\sigma_R \int \frac{\mathrm{d}\boldsymbol{\nu}\,\mathrm{d}\mathbf{p}}{(2\pi)^4} P_R(p)\varphi_{\text{eff}}(\boldsymbol{\nu},\mathbf{p}+\boldsymbol{\nu}\,(z+H)) \\ & \times \exp\!\left(-\int_0^z \left[\varepsilon_{\text{eff}}(\xi)-\sigma_{\text{eff}}(\xi)P^f_{\text{eff}}(|\mathbf{p}+\boldsymbol{\nu}(z-\xi)|)\right]\,\mathrm{d}\xi\right), \end{aligned} \tag{4.36}$$

where

$$\varphi_{\text{eff}}(\boldsymbol{\nu},\mathbf{p}) = \varphi^*_{\text{src}}(\boldsymbol{\nu},\mathbf{p})\varphi_{\text{rec}}(\boldsymbol{\nu},\mathbf{p})\,, \tag{4.37}$$

$$\varepsilon_{\text{eff}}(z) = \varepsilon(\lambda_0, z) + \varepsilon(\lambda_R, z)\,, \tag{4.38}$$

$$\sigma_{\text{eff}}(z) = \sigma(\lambda_0, z) + \sigma(\lambda_R, z)\,, \tag{4.39}$$

$$P^f_{\text{eff}}(z,\theta) = \frac{\sigma(\lambda_0, z)P^f(\lambda_0, z, \theta) + \sigma(\lambda_R, z)P^f(\lambda_R, z, \theta)}{\sigma(\lambda_0, z) + \sigma(\lambda_R, z)}\,. \tag{4.40}$$

So, as seen from (4.36), the effective intensity $I_{\text{eff}}$ is the light field radiance, produced by the effective source with the diagram

$$\varphi_{\text{eff}}(\mathbf{r}, \mathbf{n}) = \int \mathrm{d}\mathbf{r}'\,\mathrm{d}\mathbf{n}'\varphi_{\text{src}}(\mathbf{r}', \mathbf{n}')\varphi_{\text{rec}}(\mathbf{r}' + \mathbf{r}, \mathbf{n}' + \mathbf{n}) \tag{4.41}$$

in some effective medium with the extinction coefficient, the scattering coefficient, and the forward scattering phase function, defined by Eqs (4.38)–(4.40), respectively.

Formulas (4.36)–(4.41) are the solution of the direct problem of the Raman lidar return simulation within the framework of small-angle approximation. It is to be noted that, unlike other approaches (such as Monte–Carlo (Bruscaglioni et al., 1999) or Shipley (Weinman and Shipley, 1972) methods), this solution explicitly relates the Raman lidar return to the medium optical characteristics and the lidar parameters. This relation is of great importance while solving the inverse problem. Moreover, this approach makes it possible to obtain the solution for the Raman lidar return in the same form as it is for the elastic lidar return. This plays an important role while developing the methods of joint use of the elastic and Raman scattering signals.

### 4.3.3 Isotropic backscattering approximation

Equation (4.29) can be simplified if one takes into account that the functions in the integrand behave in strongly different ways. Radiance $I_{\text{eff}}(z, \mathbf{r} = 0, \mathbf{n})$ has a sharp peak in the direction $\mathbf{n} = 0$ and is negligible at large angles. In contrast, the Raman phase function is the Rayleigh function (Mobley et al., 1993):

$$P_R(\theta) = \frac{3}{4}\,\frac{1+3w}{1+2w}\left(1 + \frac{1-w}{1+3w}\cos^2\theta\right)\,, \tag{4.42}$$

where $w$ is a depolarization ratio.

The phase function (4.42) is smooth in a neighbourhood of the point $\theta = \pi$ and, therefore, it can be replaced by a constant:

$$P_R(\pi - \theta) \approx P_R(\pi)\,. \tag{4.43}$$

In order to estimate the error of this 'isotropic' approximation, let us expand the Rayleigh function into series at the point $\theta = \pi$:

$$P_R(\pi - \theta) \approx P_R(\pi)\left(1 - \frac{1-w}{1+w}\,\frac{\theta^2}{2}\right)\,. \tag{4.44}$$

The relative error $\delta$ of changing the function to the isotropic one can be estimated as:

$$\delta = \frac{1-w}{1+w} \frac{\int \mathrm{d}\mathbf{n} \frac{\mathbf{n}^2}{2} I_{\mathrm{eff}}(z, \mathbf{r}=0, \mathbf{n})}{\int \mathrm{d}\mathbf{n}\, I_{\mathrm{eff}}(z, \mathbf{r}=0, \mathbf{n})} .$$

The ratio of the integrals in this expression is nothing but the variance $V_\theta$ of the angular distribution, described by the function $I_{\mathrm{eff}}(z, \mathbf{r}=0, \mathbf{n})$:

$$V_\theta = \frac{\int \mathrm{d}\mathbf{n} \frac{\mathbf{n}^2}{2} I_{\mathrm{eff}}(z, \mathbf{r}=0, \mathbf{n})}{\int \mathrm{d}\mathbf{n}\, I_{\mathrm{eff}}(z, \mathbf{r}=0, \mathbf{n})} . \tag{4.45}$$

Hence, the relative error of isotropic approximation is equal to

$$\delta = \frac{1-w}{1+w} V_\theta .$$

Therefore, the isotropic backscattering approximation can be used if

$$\frac{1-w}{1+w} V_\theta \ll 1 . \tag{4.46}$$

Let us note that the isotropic approximation requires no restrictions in addition to the small-angle approximation validity condition, which is (Zege et al., 1991):

$$V_\theta \ll 1 . \tag{4.47}$$

Computer simulations show that the error of the isotropic approximation is less then 0.03% for a typical geometry of Raman lidar measurements.

So, the expression for the Raman lidar return (4.29) takes the form:

$$F_R(z) = W_0 \sigma_R(\pi) \int \mathrm{d}\mathbf{n}\, I_{\mathrm{eff}}(z, \mathbf{r}=0, \mathbf{n}) , \tag{4.48}$$

where $\sigma_R(\pi)$ is the Raman backscattering coefficient, related to the total Raman scattering coefficient $\sigma_R$ as following:

$$\sigma_R(\pi) = \frac{\sigma_R P_R(\pi)}{4\pi} . \tag{4.49}$$

It follows in the Fourier space:

$$F_R(z) = W_0 \sigma_R(\pi) \int \frac{\mathrm{d}\boldsymbol{\nu}}{(2\pi)^2} I_{\mathrm{eff}}(z, \boldsymbol{\nu}, \mathbf{p}=0) . \tag{4.50}$$

The isotropic approximation results in a strong simplification of expression for the Raman lidar return. The four-dimensional integral in Eq. (4.32) is reduced to the two-dimensional integral in Eq. (4.50), that is of great importance in numeric simulations. Furthermore, the expression (4.48) has a clear physical meaning: the Raman lidar return is proportional to the irradiance on the axis of

the effective light beam in the effective medium with characteristics defined by Eqs (4.38)–(4.40).

So, the problem of the Raman lidar return calculation is reduced to a standard problem of finding the irradiance in a medium with a peaked phase function. A wide range of known small-angle scattering methods could be applied to the solution of this problem (Zege et al., 1991).

### 4.3.4 The case of axially symmetric source and receiver patterns

A particular case when both source and receiver diagrams have axially symmetric angular-spatial patterns, which can be represented as

$$\varphi_j(\mathbf{r},\mathbf{n}) = \varphi_j^{\mathrm{sp}}(|\mathbf{r}-\mathbf{r}_j^0|)\varphi_j^{\mathrm{ang}}(|\mathbf{n}-\mathbf{n}_j^0|)\,, \quad j=\mathrm{scr},\mathrm{rec}\,, \tag{4.51}$$

is of a great practical interest. In this case, the Fourier transform of the effective source diagram takes the form:

$$\varphi^{\mathrm{eff}}(\boldsymbol{\nu},\mathbf{p}) = \varphi_{\mathrm{scr}}^{\mathrm{sp}}(\nu)\,\varphi_{\mathrm{scr}}^{\mathrm{ang}}(p)\,\varphi_{\mathrm{rec}}^{\mathrm{sp}}(\nu)\,\varphi_{\mathrm{rec}}^{\mathrm{ang}}(p)\exp(-i\,\boldsymbol{\nu}\cdot\mathbf{R}-i\,\mathbf{p}\cdot\boldsymbol{\Omega})\,, \tag{4.52}$$

where $\mathbf{R}=\mathbf{r}_{\mathrm{rec}}^0-\mathbf{r}_{\mathrm{scr}}^0$ is the vector, connecting the source and receiver centres, $\boldsymbol{\Omega}=\mathbf{n}_{\mathrm{rec}}^0-\mathbf{n}_{\mathrm{scr}}^0$ is the vector, determining the angle between the receiver and source axes. Besides, it follows:

$$\begin{aligned}\varphi_j^{\mathrm{ang}}(p) &= 2\pi\int_0^\infty \varphi_j^{\mathrm{ang}}(\theta)\,J_0(p\theta)\theta\,\mathrm{d}\theta\,,\\ \varphi_j^{\mathrm{sp}}(\nu) &= 2\pi\int_0^\infty \varphi_j^{\mathrm{sp}}(r)\,J_0(\nu r)r\,\mathrm{d}r\,;\quad j=\mathrm{scr},\mathrm{rec}.\end{aligned} \tag{4.53}$$

Equation (4.52) makes possible the simplification of the expression (4.50) by integrating over the azimuth and reducing the integral (4.50) to the one-dimensional integral:

$$F_R(z) = W_0\sigma_R(\pi)\int\frac{\nu\,\mathrm{d}\nu}{2\pi}\;J_0(\nu|\mathbf{R}+(z+H)\boldsymbol{\Omega}|)I_{\mathrm{eff}}(z,\nu,\mathbf{p}=0)\,. \tag{4.54}$$

Here, the Fourier transform of the effective intensity is:

$$I_{\mathrm{eff}}(z,\nu,\mathbf{p}=0)=\varphi'(\nu,\nu(z+H))\,\exp\left(-\int_0^z\left[\varepsilon_{\mathrm{eff}}(\xi)-\sigma_{\mathrm{eff}}(\xi)P_{\mathrm{eff}}^f(\nu(z-\xi))\right]\mathrm{d}\xi\right) \tag{4.55}$$

and

$$\varphi'(\nu,p) = \varphi_{\mathrm{scr}}^{\mathrm{sp}}(\nu)\,\varphi_{\mathrm{scr}}^{\mathrm{ang}}(p)\,\varphi_{\mathrm{rec}}^{\mathrm{sp}}(\nu)\,\varphi_{\mathrm{rec}}^{\mathrm{ang}}(p)\,. \tag{4.56}$$

Equations (4.51)–(4.56), along with the effective medium properties (4.38)–(4.40), give the explicit solution to the problem for the case of axially symmetric source and receiver patterns.

Figure 4.2 presents the Raman lidar return simulation example in comparison with the data of Wandinger (1998).

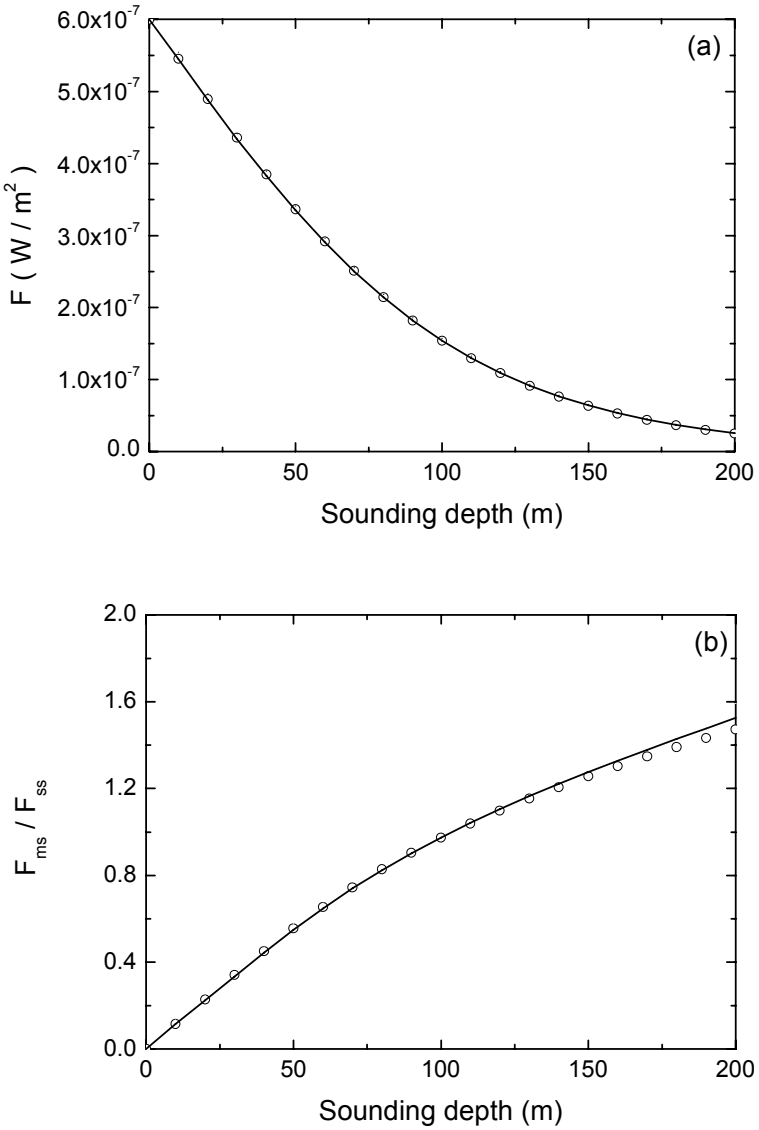



**Fig. 4.2.** The total signal (a) and the multiple-to-single scattering ratio (b). Data by Wandinger (1998) (○) and the solution (4.54) (curve). The cloud C.1 model (Deirmendjian, 1969) with an extinction coefficient of 0.01 m$^{-1}$. The lidar -cloud distance is 5000 m, the source divergence is 0.1 mrad, the receiver's FOV is 0.4 mrad (full angles).

## 4.4 Spatial-angular pattern of the Raman lidar return

### 4.4.1 Introduction to the problem

The Raman lidar return is usually used in the sounding of aerosols to measure the extinction coefficient profile, because within the framework of the single scattering approximation the Raman lidar return depends on the extinction coefficient alone. Owing to the spectral dependence of the extinction coefficient, the microstructure characteristics of a medium (e.g., sizes of particles) can be retrieved. However, for the sounding of clouds, consisting of large particles, the spectral behaviour of the extinction coefficient contains no information about microstructure parameters. On the other hand, when sounding clouds, the effect of multiple scattering becomes significant, and lidar return depends not only on the extinction, but also on the phase function, which, in its turn, depends on the size of the scatterers.

Multiple scattering becomes crucial when measuring angular patterns of lidar return (e.g., with a multiple-field-of-view (Bissonnette and Hutt, 1990) or imaging (Muscari et al., 1996) receiver). In that case the signal at some angles is exclusively determined by multiple scattering.

Bissonnette and Hutt (1990) and Roy et al. (1997) apparently were the first to suggest using a multiple-field-of-view (MFOV) receiver to determine the effective particle size using elastic lidar sounding of warm clouds and aerosols. However, the absence of the appropriate theoretical base made them use the semi-empirical methods in the retrieval procedure. The theoretical investigation of an angular pattern of the elastic lidar return from warm clouds is given by Polonsky et al. (2001). This investigation shows that for sounding depths, at which the scattering is mainly determined by Fraunhofer diffraction on water droplets, the angular dependence of the lidar return appears as a product of the receiving angle by the effective droplet size. Such an angular dependence enables the retrieval of cloud droplet sizes. However, the complex behaviour of the elastic backscattering phase function hinders the implementation of such methods (e.g., accounting for finite size of the receiver, which is necessary when investigating the spatial-angular pattern of lidar return). In fact, isotropic backscattering simplifies the problem greatly for the case of Raman lidar sounding.

Basing on the developed model, the investigation of spatial-angular pattern of the Raman lidar return is carried out in this chapter. Further, on its basis, a new optical particle sizing technique is proposed.

### 4.4.2 The effective medium properties

In order to investigate the spatial-angular pattern of the Raman lidar return, let us examine first the properties of the effective medium introduced above. We consider optically hard scatterers of sizes much larger than the wavelength. Warm (liquid) clouds could be a paradigm of such a medium. Further, we consider sounding depths that are not very large. These two premises allow us to regard scattering as Fraunhofer diffraction of light on water droplets. Within

the framework of the Fraunhofer diffraction approximation the scattering cross-section and, therefore, the scattering coefficient $\sigma^D$ do not depend on wavelength:

$$\sigma^D(\lambda_0, z) = \sigma^D(\lambda_R, z) = \sigma^D(z)\,. \tag{4.57}$$

So the effective phase function (4.40) is:

$$P^D_{\text{eff}}(\theta) = \frac{P^D(\lambda_0, \theta) + P^D(\lambda_R, \theta)}{2}\,, \tag{4.58}$$

where $P^D(\lambda, \theta)$ is the Fraunhofer phase function (van de Hulst, 1957):

$$P^D(\lambda, \theta) = \frac{1}{\langle r^2 \rangle} \left\langle r^2 \frac{4}{\theta^2} J_1^2 \left( \frac{2\pi r}{\lambda} \theta \right) \right\rangle\,, \tag{4.59}$$

where $r$ is a droplet radius (angle brackets denote size averaging).

As was shown by Kokhanovsky and Zege (1997), the phase function of clouds does not really depend on the exact form of the size distribution function, but rather on the value of the effective diffraction parameter (the dimensionless droplets size) $\rho_{32}$ defined as:

$$\rho_{32} = \frac{2\pi r_{32}}{\lambda}\,, \tag{4.60}$$

where $r_{32}$ is the effective droplets radius:

$$r_{32} = \frac{\langle r^3 \rangle}{\langle r^2 \rangle} \tag{4.61}$$

and $\lambda$ is the wavelength of the incident light.

As a result, whereas the angular dependence in Fraunhofer diffraction appears as a product $\rho\theta$, $\rho = 2\pi a/\lambda$ being the dimensionless radius of scatterers, the diffraction phase function should have the form:

$$P^D(\theta) = \rho_{32}^2 \Phi(\rho_{32}\theta)\,, \tag{4.62}$$

where the function $\Phi(x)$ satisfies the normalizing condition, which is the small-angle analogue of Eq. (4.22):

$$\frac{1}{2} \int_0^\infty \Phi(x) x \, \mathrm{d}x = 1\,. \tag{4.63}$$

According to the developed model, the light propagates into the medium at the initial wavelength and travels back to the lidar receiver at the shifted wavelength. Hence, with application to Raman lidar sounding, we can introduce the effective dimensionless radius:

$$\rho_{\text{eff}} = \frac{1}{2} \left( \rho_{32}(\lambda_0) + \rho_{32}(\lambda_R) \right) = \frac{\pi r_{32}}{\lambda_0} + \frac{\pi r_{32}}{\lambda_R}\,, \tag{4.64}$$

and, thereafter, the effective wavelength $\lambda_{\text{eff}}$:

$$\frac{2}{\lambda_{\text{eff}}} = \frac{1}{\lambda_0} + \frac{1}{\lambda_R}\,. \tag{4.65}$$

Then it follows:

$$\rho_{\text{eff}} = \frac{2\pi r_{32}}{\lambda_{\text{eff}}}\,. \tag{4.66}$$

Taking into account that $\lambda_R$ is related to $\lambda_0$ and $\delta\tilde{\nu}$ through Eq. (4.4), let us express the initial and shifted wavelengths through $\lambda_{\text{eff}}$:

$$\lambda_{0,R} = \frac{\lambda_{\text{eff}}}{1 \pm \lambda_{\text{eff}}\delta\tilde{\nu}/2}\,. \tag{4.67}$$

As could be easily shown by differentiation with respect to parameter $\rho_{32}$, the phase function value at the point $\theta = 0$, in the case of a fairly wide size distribution, is most sensitive to the size of scatterers, and, therefore, to the wavelength. It follows from Eq. (4.59) that:

$$P^D(\lambda, \theta = 0) = \frac{4\pi^2}{\lambda^2}\,\frac{\langle r^4\rangle}{\langle r^2\rangle} \tag{4.68}$$

and one can derive at $\theta = 0$ (see Eq. (4.58)):

$$\begin{aligned}
\frac{P^D(\lambda_0, 0) + P^D(\lambda_R, 0)}{2} &= 2\pi^2\frac{\langle r^4\rangle}{\langle r^2\rangle}\left(\frac{1}{\lambda_0^2} + \frac{1}{\lambda_R^2}\right) \\
&= 2\pi^2\frac{\langle r^4\rangle}{\langle r^2\rangle}\left(\frac{(1 + \lambda_{\text{eff}}\delta\tilde{\nu}/2)^2}{\lambda_{\text{eff}}^2} + \frac{(1 - \lambda_{\text{eff}}\delta\tilde{\nu}/2)^2}{\lambda_{\text{eff}}^2}\right) \\
&= \frac{4\pi^2}{\lambda_{\text{eff}}^2}\,\frac{\langle r^4\rangle}{\langle r^2\rangle}\left(1 + \left(\frac{\lambda_{\text{eff}}\delta\tilde{\nu}}{2}\right)^2\right) \\
&= P^D(\lambda_{\text{eff}}, 0)\left(1 + \left(\frac{\lambda_{\text{eff}}\delta\tilde{\nu}}{2}\right)^2\right).
\end{aligned} \tag{4.69}$$

That is, the half-sum of the Fraunhofer phase functions at $\theta = 0$ equals the Fraunhofer phase function at the effective wavelength within accuracy to the second order of the value $\lambda_{\text{eff}}\delta\tilde{\nu}$. If the initial wavelength is 532 nm and Raman scattering is from nitrogen, the shifted wavelength will be 607 nm and the effective one will be 567 nm. In this case the correction will be less than 1%. (For Raman lidar sounding the shorter wavelengths are often used. In that case the correction is even smaller, for example, for the initial wavelength of 308 nm it is about 0.1%.) Whereas the phase function value at $\theta = 0$ is most sensitive to the wavelength, the effective phase function (4.58) can be changed by the phase function at the effective wavelength for any $\theta$ with a correction less than the correction for $\theta = 0$. Finally, we get:

$$P_{\text{eff}}^D(\theta) = P^D(\lambda_{\text{eff}}, \theta) = \rho_{\text{eff}}^2\Phi(\rho_{\text{eff}}\theta)\,. \tag{4.70}$$

Thus, the effective forward scattering phase function depends on the effective dimensionless radius only.

It should be noted that the effective dimensionless radius and, therefore, the phase function can generally depend on the sounding depth $z$. However, in order to simplify the problem, we will consider here the case when $\rho_{\text{eff}}$ does not depend on $z$, otherwise, the effective particle size will stand for some average value over the interval $(0, z)$.

### 4.4.3 Spatial-angular patterns of Raman lidar returns and their dependence on the size of scatterers

Now, after investigating the angular pattern of the effective medium phase function, let us consider the pattern of a lidar return.

Let a source be infinitesimal and mono-directional, i.e., its diagram is:

$$\varphi_{\text{src}}(\mathbf{r}', \mathbf{n}') = \delta(\mathbf{r}')\delta(\mathbf{n}')\,, \tag{4.71}$$

where $\delta(\mathbf{x})$ is the two-dimensional Dirac $\delta$-function.

Let us investigate the spatial-angular distribution of returned radiance in the lidar plane ($z = -H$). This means that the 'receiver' has a diagram:

$$\varphi_{\text{rec}}(\mathbf{r}', \mathbf{n}') = \delta(\mathbf{r}' - \mathbf{r})\delta(\mathbf{n}' - \mathbf{n})\,, \tag{4.72}$$

and its Fourier transform is equal to:

$$\varphi_{\text{rec}}(\boldsymbol{\nu}, \mathbf{p}) = \exp(-i\boldsymbol{\nu}\cdot\mathbf{r} - i\mathbf{p}\cdot\mathbf{n})\,. \tag{4.73}$$

Then, the diagram of the effective source (4.41) is identical to the diagram (4.72) and its Fourier transform is given by (4.73). The Raman lidar return, according to Eq. (4.48), is proportional to the irradiance, produced by this source in the effective medium at the point $(z, 0, 0)$. Substituting (4.70) and (4.73) into (4.48) and keeping in mind that the scattering coefficient does not depend on wavelength, we get, after integrating over the azimuth:

$$\begin{aligned} F(z, \mathbf{r}, \mathbf{n}) \;&=\; W_0\sigma_R(\pi)\int \frac{\nu d\nu}{2\pi}\, J_0\left(\nu|\mathbf{r} + \mathbf{n}\,(z+H)|\right) \\ &\quad\times \exp\left(-\int_0^z \left[\varepsilon_{\text{eff}}(\xi) - 2\sigma^D(\xi)P^D_{\text{eff}}\left(\nu(z-\xi)\right)\right]\,\mathrm{d}\xi\right)\,. \end{aligned} \tag{4.74}$$

Here $P^D_{\text{eff}}(p)$ is the Hankel transform of the diffraction phase function $P^D_{\text{eff}}(\theta)$:

$$\begin{aligned} P^D_{\text{eff}}(p) \;&=\; \frac{1}{2}\int_0^\infty \rho^2_{\text{eff}}\Phi(\rho_{\text{eff}}\theta)\, J_0(p\,\theta)\,\theta\,\mathrm{d}\theta \\ &=\; \frac{1}{2}\int_0^\infty \Phi(x)\, J_0\left(x\frac{p}{\rho_{\text{eff}}}\right)\, x\,\mathrm{d}x = \chi(p/\rho_{\text{eff}})\,, \end{aligned} \tag{4.75}$$

where $\chi(p)$ is the Hankel transform of the function $\Phi(x)$.

Substituting (4.75) into (4.74) and changing variables in integration

$$\boldsymbol{\nu} \to \rho_{\text{eff}}\boldsymbol{\nu}\,,$$

we come to the expression

$$F(z,\mathbf{r},\mathbf{n}) = W_0\sigma_R(\pi)\rho_{\text{eff}}^2\int\frac{\nu\,\mathrm{d}\nu}{2\pi}\,J_0\left(\rho_{\text{eff}}\nu|\mathbf{r}+\mathbf{n}\,(z+H)|\right) \times\exp\left(-\int_0^z\left[\varepsilon_{\text{eff}}(\xi)-2\sigma^D(\xi)\chi(\nu(z-\xi))\right]\mathrm{d}\xi\right), \quad (4.76)$$

in which spatial and angular coordinates appear as a combination $\rho_{\text{eff}}|\mathbf{r}+(z+H)\mathbf{n}|$ only:

$$F(z,\mathbf{r},\mathbf{n}) = \rho_{\text{eff}}^2 f\left(\rho_{\text{eff}}|\mathbf{r}+(z+H)\mathbf{n}|\right). \quad (4.77)$$

If the receiver size is negligible, its spatial diagram is:

$$\varphi_{\text{rec}}(\mathbf{r}) = S_{\text{rec}}\delta(\mathbf{r}). \quad (4.78)$$

In that case one can put $\mathbf{r}=0$ in Eq. (4.77) and find that the lidar return does not depend on the azimuth, but only on the product of $\rho_{\text{eff}}$ by the receiving angle $\theta$:

$$F(z,\theta) = \rho_{\text{eff}}^2 f\left((z+H)\rho_{\text{eff}}\theta\right). \quad (4.79)$$

To illustrate the accuracy of Eq. (4.79) we simulated the angular distribution of the Raman lidar return on the basis of the initial expression (4.54) without additional approximations. The phase functions were calculated with Mie theory for warm clouds with droplets size distribution of Cloud C.1 type (Deirmendjian, 1969) with different effective radii $r_{32}$:

$$\frac{\mathrm{d}\omega}{\mathrm{d}r} = \frac{r^m}{m!}\left(\frac{r_{32}}{m+3}\right)^{-m-1}\exp\left(-\frac{r\,(m+3)}{r_{32}}\right), \quad (4.80)$$

where $r$ is the droplet radius, $\mathrm{d}\omega$ is the probability that the droplet radius lies in the interval from $r$ to $r+\mathrm{d}r$, $m$ is a Gamma-distribution parameter (here $m=6$).

The dependence of the ratio $F(z,\theta)/\rho_{\text{eff}}^2$ on the product $\rho_{\text{eff}}\theta$ is plotted in Fig. 4.3. The initial wavelength is 532 nm, the shifted one is 607 nm (the effective one is 567 nm). The cloud altitude is 1000 m, the receiver radius is 25 cm, and the pulse energy is 1 J. It is seen that, regardless of great difference in properties of media under consideration (the effective radius varies all over the range usually observed in warm clouds), the value $F(z,\theta)/\rho_{\text{eff}}^2$ depends on the product $\rho_{\text{eff}}\theta$ only, in close agreement with Eq. (4.79). The discrepancy in the range of small angles is due to the singly scattered light, for which the approximation (4.78) is inapplicable.

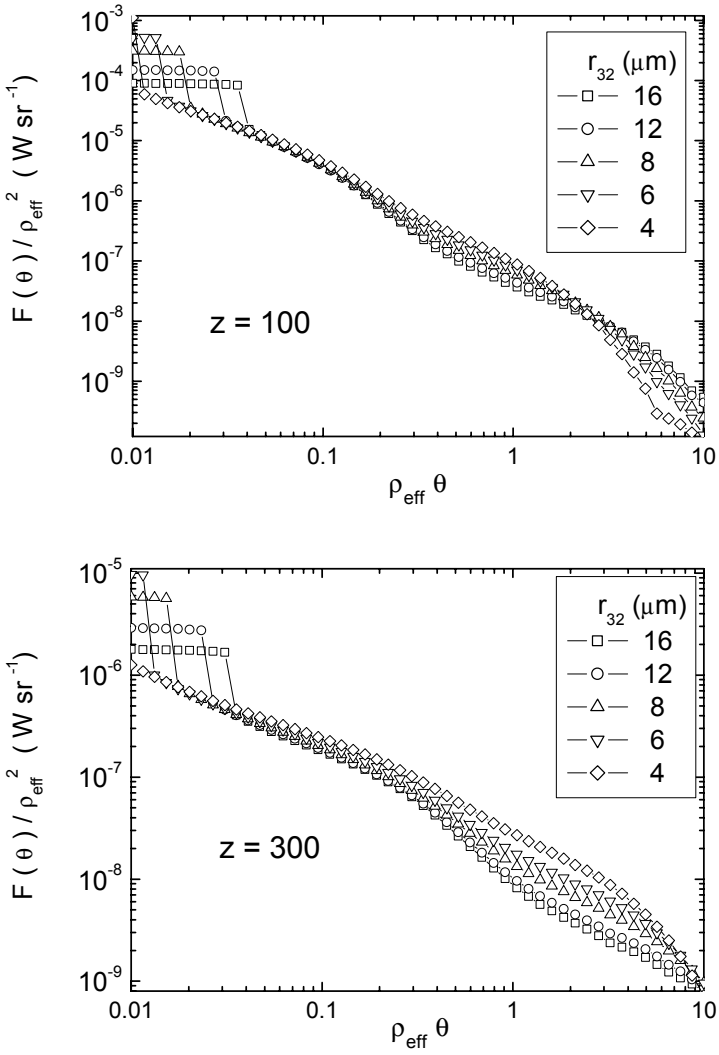



**Fig. 4.3.** Ratio $F(z,\theta)/\rho_{\text{eff}}^2$ vs. the product $\rho_{\text{eff}}\theta$ for clouds with the different effective size of droplets.

## 4.5 Retrieval of the microphysical properties of light scattering media using measurements of the Raman lidar return angular patterns

### 4.5.1 The retrieval possibilities

Equation (4.79) makes it possible to determinate the sizes of scatterers using measurements of the Raman lidar return at several angular intervals. Indeed, integrating (4.76) over the spatial coordinate from 0 to $R$ and over the angle from 0 to $\gamma$, we get:

$$F(\gamma) = W_0 \sigma_R(\pi) \frac{R\gamma}{z+H} \int \frac{2\pi \, \mathrm{d}\nu}{\nu} J_1\left(\rho_{\text{eff}} R \nu\right) J_1\left(\rho_{\text{eff}}(z+H)\gamma \nu\right) \times \exp\left( - \int_0^z \left[\varepsilon_{\text{eff}}(\xi) - 2\sigma^D(\xi)\chi(\nu(z-\xi))\right] \mathrm{d}\xi \right) , \tag{4.81}$$

where $F(\gamma)$ denotes the light flux, measured with the receiver of radius $R$ at the angle interval from 0 to $\gamma$:

$$F(\gamma) = \int_0^{\gamma} \theta \, \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi^{\text{ang}} \int_0^R r \, \mathrm{d}r \int_0^{2\pi} \mathrm{d}\phi^{\text{sp}} F(z, \mathbf{r}, \mathbf{n}) , \tag{4.82}$$

where $\phi^{\text{ang, sp}}$ is the azimuthal angle for angular and spatial coordinates, respectively.

According to Eq. (4.81), FOV $\gamma$, the receiver radius $R$, and the effective dimensionless radius $\rho_{\text{eff}}$ appear in this expression only as a combination:

$$F(\gamma) = R\gamma f\left(\rho_{\text{eff}} R, \rho_{\text{eff}} \gamma\right) . \tag{4.83}$$

Then the ratio of the flux $F_1$, measured in the interval $(\gamma_0, \gamma_1)$, to the flux $F_0$, measured in the interval $(0, \gamma_0)$ depends on the two cloud characteristics only: $\rho_{\text{eff}}$ and the scattering coefficient $\sigma^D$:

$$\frac{F_1}{F_0} = \frac{F(\gamma_1) - F(\gamma_0)}{F(\gamma_0)} = \frac{\gamma_1 \, \mathrm{f}\left(\rho_{\text{eff}} R, \rho_{\text{eff}} \gamma_1\right)}{\gamma_0 \, \mathrm{f}\left(\rho_{\text{eff}} R, \rho_{\text{eff}} \gamma_0\right)} - 1 . \tag{4.84}$$

All other quantities are known values defined by the experiment geometry.

In that case, if the scattering coefficient is known from some other measurements, e.g., through the depth dependence of a lidar return, then the ratio of fluxes $\varsigma = F_1/F_0$ is a function of the single variable $\rho_{\text{eff}}$, which can be easily retrieved by measuring the value of $\varsigma$.

The dependence of multiply scattered signal on $\rho_{\text{eff}}$ is strongest at small angles: the smaller the angle $\theta$, the stronger the dependence. However, at $\theta < R/(z+H)$ the significant portion of the measured signal is the singly scattered light, which contains no information about $\rho_{\text{eff}}$. Therefore, the largest amount

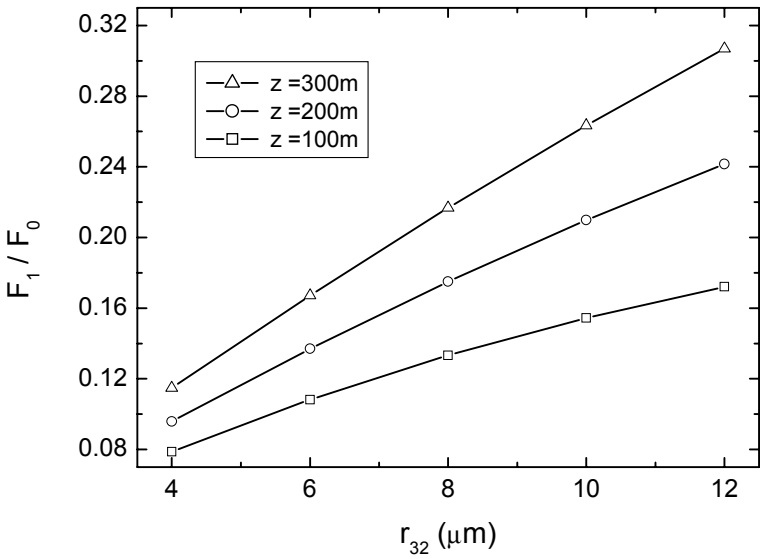



**Fig. 4.4.** The ratio $F_1/F_0$ vs. the effective droplets radius.

of information will be carried by a signal measured with an annular receiver at the angular range $(\gamma_0, \gamma_1)$ satisfying the conditions:

$$\gamma_1 > \gamma_0 > \frac{R}{z+H}\,, \tag{4.85}$$

$$\gamma_1 \frac{z+H}{R} - 1 < 1\,. \tag{4.86}$$

Figure 4.4 presents an example of the dependence of $F_1/F_0$ on the effective radius of cloud droplets $r_{32}$ for clouds with the particle size distribution (4.80). The calculations were carried out using Eq. (4.54) with no additional approximations. It is evident from Fig. 4.4 that the ratio $F_1/F_0$ is quite sensitive to the radius of particles, and this sensitivity increases with sounding depth. As it follows from the results presented in the plot, the relative error of retrieving $r_{32}$ has the same order of magnitude as the relative error of measuring the flux ratio $F_1/F_0$.

### 4.5.2 Use of double scattering for retrieving the volume concentration of scatterers

The double-scattering approximation was used by many authors to simulate lidar return (see, for example, Eloranta, 1972; Samokhvalov, 1979). We will consider its small-angle modification and show that for the range of its validity (small FOVs and not very large optical depths), it can be used to retrieve information about a light scattering medium without the use of additional *a priori* data.

Let us consider a homogeneous cloud. In this case the expression for the flux (4.81) takes the form:

$$F(\gamma) = W_0\sigma_R(\pi)\frac{R\,\gamma}{z+H}\int\frac{2\pi\,\mathrm{d}\nu}{\nu}\,J_1\left(\rho_{\mathrm{eff}}R\,\nu\right)J_1\left(\rho_{\mathrm{eff}}(z+H)\gamma\,\nu\right) \times \exp\left(-2\tau+2\sigma^D\int_0^z\chi(\nu\xi)\,\mathrm{d}\xi\right), \tag{4.87}$$

where $\tau$ is the optical depth defined as:

$$\tau=\frac{\varepsilon_{\mathrm{eff}}z}{2}. \tag{4.88}$$

Passing to the dimensionless variables $p$ and $\eta$ and making the change in Eq. (4.87):

$$\nu=\frac{p}{z},\quad \xi=\frac{z}{p}\eta, \tag{4.89}$$

we get:

$$F(\gamma) = W_0\sigma_R(\pi)\frac{R\,\gamma}{z+H}\int\frac{2\pi\,\mathrm{d}p}{p}\,J_1\left(\rho_{\mathrm{eff}}\frac{R}{z}p\right)J_1\left(\rho_{\mathrm{eff}}\frac{z+H}{z}\gamma p\right) \times \exp\left(-2\tau+2\frac{\sigma^D z}{p}\int_0^p\chi(\eta)\,\mathrm{d}\eta\right). \tag{4.90}$$

The double-scattering approximation in the Raman case means that one event of Raman and one event of elastic scattering are taken into account, i.e., one should take the term of the first order of $\sigma^D z$ in Eq. (4.90):

$$F(\gamma) = W_0\sigma_R(\pi)\frac{R\,\gamma}{z+H}\int\frac{2\pi\,\mathrm{d}p}{p}\,J_1\left(\rho_{\mathrm{eff}}\frac{R}{z}p\right)J_1\left(\rho_{\mathrm{eff}}\frac{z+H}{z}\gamma p\right) \times\mathrm{e}^{-2\tau}\left(1+2\frac{\sigma^D z}{p}\int_0^p\chi(\eta)\,\mathrm{d}\eta\right). \tag{4.91}$$

The unity in the parentheses corresponds to single scattering and the second term corresponds to double scattering, with one forward elastic and one backward Raman scattering event.

The integral (4.91) can be calculated approximately at

$$\rho_{\mathrm{eff}}\frac{z+H}{z}\gamma\ll 1. \tag{4.92}$$

This corresponds to the light field close to the light source. In that case the integral (4.91) is defined by the asymptotic behaviour of the function to be integrated at $p\gg 1$, and the upper limit can be put equal to infinity. This integral can be calculated exactly. To do so, we use Eq. (4.75), according to

which the function $\chi(\eta)$ is the Hankel transform of the function $\Phi(x)$. Then, it follows:

$$\int_0^p \chi(\eta)\,\mathrm{d}\eta \;\approx\; \int_0^\infty \chi(\eta)\,\mathrm{d}\eta = \frac{1}{2}\int_0^\infty \Phi(x)\,\mathrm{d}x = \frac{1}{2\rho_{\text{eff}}}\int_0^\infty P_{\text{eff}}^D(\theta)\,\mathrm{d}\theta$$
$$= \frac{1}{2\rho_{\text{eff}}}\,\frac{16}{3\pi}\,\frac{2\pi}{\lambda_{\text{eff}}}\,\frac{\langle r^3\rangle}{\langle r^2\rangle} = \frac{1}{2\rho_{\text{eff}}}\,\frac{16}{3\pi}\rho_{\text{eff}} = \frac{8}{3\pi} \tag{4.93}$$

(when integrating over $\theta$, Eqs (4.59) and (4.70) and also the Parseval equality have been used).

Substituting (4.93) into (4.91), we get:

$$F(\gamma) \;=\; W_0\sigma_R(\pi)\frac{R\gamma}{z+H}$$
$$\times\int \frac{2\pi\,\mathrm{d}p}{p}\,J_1\left(\rho_{\text{eff}}\frac{R}{z}p\right)J_1\left(\rho_{\text{eff}}\frac{z+H}{z}\gamma p\right)\,\mathrm{e}^{-2\tau}\left(1+\frac{16}{3\pi}\frac{\sigma^D z}{p}\right). \tag{4.94}$$

Using conditions (4.85), (4.86), and (4.92), and taking the first order of the value $\gamma(z+H)/R-1$, we get an approximate expression:

$$F(\gamma) = W_0\sigma_R(\pi)\,\mathrm{e}^{-2\tau}\frac{\pi R^2}{(z+H)^2}\left(1+\frac{64\,\sigma^D\rho_{\text{eff}}}{3\pi^2}\left((z+H)\gamma-R/3\right)\right). \tag{4.95}$$

If the first receiver (circular) measures the signal $F_0$ in the interval from 0 to $\gamma_0$ and the second receiver (annular) measures the signal $F_1$ in the interval from $\gamma_0$ to $\gamma_1$, then the ratio $F_1/F_0$ is equal to:

$$\frac{F_1}{F_0} = \frac{(z+H)\,(\gamma_1-\gamma_0)}{\dfrac{3\pi^2}{64\,\sigma^D\rho_{\text{eff}}}+(z+H)\gamma_0-R/3}\,. \tag{4.96}$$

From this equation the product $\sigma^D$ by $\rho_{\text{eff}}$ can be easily retrieved:

$$\sigma^D\rho_{\text{eff}} = \frac{3\pi^2}{64}\,\frac{F_1/F_0}{c_1-c_2\,F_1/F_0}\,, \tag{4.97}$$

where the coefficients $c_1$ and $c_2$ are equal to:

$$\begin{aligned} c_1 &= (z+H)\,(\gamma_1-\gamma_0),\\ c_2 &= (z+H)\gamma_0-R/3. \end{aligned} \tag{4.98}$$

Within the approximation of Fraunhofer diffraction, the scattering cross-section equals the cross-sectional area of a scattering particle. Therefore, the scattering coefficient is:

$$\sigma^D = \langle\pi r^2\rangle\,C_N = \langle\pi r^2\rangle\,\frac{C_V}{\langle 4\pi r^3/3\rangle} = \frac{3}{4}\frac{C_V}{r_{32}}\,, \tag{4.99}$$

where $C_N$ is the numeric particle concentration and $C_V$ is the volume particle concentration.

It should be noted that the diffraction scattering coefficient $\sigma^D$ used here is approximately twice less than the total Mie scattering coefficient.

For the product of $\sigma^D$ by $\rho_{\text{eff}}$, we get:

$$\sigma^D \rho_{\text{eff}} = \frac{3}{4}\frac{C_V}{r_{32}}\frac{2\pi r_{32}}{\lambda_{\text{eff}}} = \frac{3\pi}{2}\frac{C_V}{\lambda_{\text{eff}}}\,. \tag{4.100}$$

This means that through the ratio $F_1/F_0$ one can retrieve the volume concentration of droplets using the following simple formula:

$$C_V = \lambda_{\text{eff}}\frac{\pi}{32}\,\frac{F_1/F_0}{c_1 - c_2\,F_1/F_0}\,. \tag{4.101}$$

Table 4.1 represents an example of such a retrieval. The five types of clouds with the droplets distribution (4.80) and different effective sizes were used for modelling. The scattering coefficient for all the cloud types was taken of $0.01\,\text{m}^{-1}$, the sounding depth is 100 m. The retrieved and true volume concentrations (and the relative error of retrieval, as well) are shown in Table 4.1. As can be seen, the retrieval error is not greater than 20%, but it increases with sounding depth, when the double-scattering approximation becomes invalid. The applicability of the formula (4.101) is bounded by small optical depths (about unity), as well as by the conditions (4.86) and (4.92) imposed on the geometry of an experiment. However, the use of formula (4.101) requires no additional knowledge about the properties of the medium and it can be used as a reference point in other algorithms.

**Table 4.1.** Example of the droplets volume concentration retrieval from Eq. (4.97)

| Effective radius (μm) | True concentration (ppm) | Retrieved concentration (ppm) | Relative error (%) |
|---|---|---|---|
| 4 | 0.02462 | 0.02924 | 18.7 |
| 6 | 0.03764 | 0.04177 | 11.0 |
| 8 | 0.05072 | 0.05324 | 5.0 |
| 10 | 0.06383 | 0.06350 | −0.5 |
| 12 | 0.07698 | 0.07256 | −5.7 |

### 4.5.3 The algorithm of simultaneous retrieval of the scattering coefficient and the effective droplet size

Within the single scattering approximation, the scattering coefficient can be found through the logarithmic derivative:

$$\varepsilon^0(z) = -\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}z}\ln\left[F_R(z)\,(z+H)^2\right]\,, \tag{4.102}$$

However, if the single scattering albedo and the phase function are known, the scattering coefficient can also be found in the case of significant influence of

multiple scattering. It can be done within the following iterative scheme: the starting scattering profile is found through Eq. (4.102), then the contribution of multiple scattering is found and the scattering profile is corrected with regard to multiple scattering. The formula for this iterative scheme is:

$$\varepsilon^{k+1}(z) = \varepsilon^k(z) - \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}z}\ln\left[\frac{F_R(z)}{F_R\left(\varepsilon^k(z)\right)}\right], \tag{4.103}$$

where $F_R(z)$ denotes the measured signal and $F_R\left(\varepsilon^k(z)\right)$ denotes the signal, calculated with account for multiple scattering with the extinction profile $\varepsilon^k(z)$ (the extinction coefficient is equal to the scattering coefficient for clouds in the visible range).

Therefore, Raman lidar sounding of warm clouds with an MFOV receiver makes possible the simultaneous retrieval of both the extinction coefficient profile and the effective size of droplets.

Let us consider the abovementioned measurement scheme. As was shown, if the scattering profile is known (the single scattering albedo for clouds is equal to unity in the visible), one can easily retrieve the effective radius of droplets through the ratio of two signals, the first being determined mainly by single scattering ($F_0$) and the second being determined exclusively by multiple scattering ($F_1$). This means that the extinction profile retrieval should be the first step. It can be done within the single scattering approximation, using the signal $F_0$:

$$\varepsilon^0(z) = -\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}z}\ln\left[F_0(z)\,(z+H)^2\right]. \tag{4.104}$$

Then, using the ratio $F_1/F_0$, the effective radius of droplets is retrieved. If the retrieval accuracy is insufficient, knowing the particles size (i.e., the phase function), one can correct the extinction profile with regard to multiple scattering:

$$\varepsilon^{k+1}(z) = \varepsilon^k(z) - \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}z}\ln\left[\frac{F_0(z)}{F_0\left(\varepsilon^k(z), r_{32}^k\right)}\right], \tag{4.105}$$

where $\varepsilon^k(z)$ denotes the $k$th iteration of the extinction profile, $r_{32}^k$ is the $k$th iteration of the effective radius, $F_0(z)$ is the 'true' signal, measured with FOV $\gamma_0$, and $F_0\left(\varepsilon^k(z), r_{32}^k\right)$ is the signal, simulated with the profile $\varepsilon^k(z)$ and the effective radius $r_{32}^k$.

New value of $r_{32}^{k+1}$ is found through the ratio $F_1/F_0$ with new profile $\varepsilon^{k+1}(z)$. The iterations go on until the required accuracy is achieved.

The example of such retrieval is presented in Fig. 4.5 and in Table 4.2. Only 0th (starting) and 11th iterations are plotted in Fig. 4.5. The first six iterations are shown in Table 4.2. It is seen that the retrieval within the single scattering approximation can underestimate the value of the extinction coefficient by two times, whereas after ten iterations the retrieved profile in fact replicates the true one. The derived value of $r_{32}$ is fairly accurate at the first iteration (overestimated of about 10%) and converges rapidly to the true value.

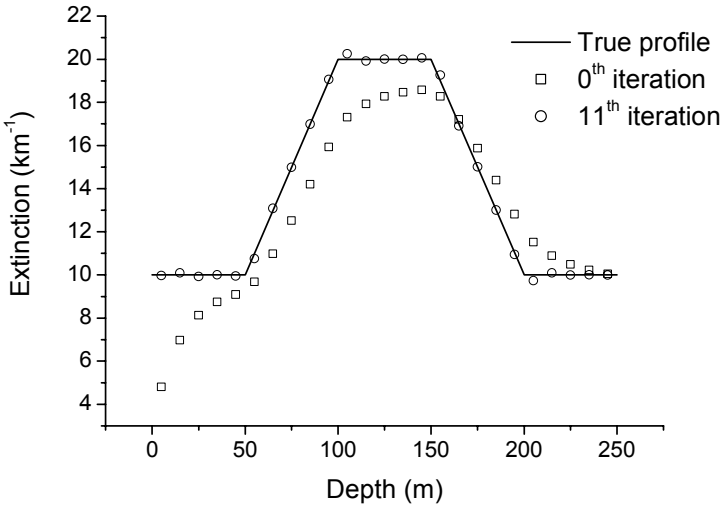



**Fig. 4.5.** Example of the extinction profile retrieval.

**Table 4.2.** Example of the effective droplets radius retrieval

| Iteration number | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Effective radius (μm) | 6.57 | 5.84 | 6.02 | 6.02 | 6.01 | 6.00 |

## 4.6 Conclusion

This chapter has given a review of the main methods of Raman lidar sounding. It has displayed the advantages of and the additional possibilities in the remote sensing of the environment due to the usage of Raman lidars. It was emphasized that most of the methods of processing Raman lidar data rely on the lidar equation formulated within the single scattering approximation. The analytical theory of the Raman lidar return with regard to multiple scattering has been described in this chapter in great depth. The theory is based on the developed earlier approach of small-angle quasi-single scattering approximation, used to describe multiple scattering in usual (elastic) lidar sounding. The investigation of the angular structure of the Raman lidar return made it possible to relate the angular characteristics of multiple scattering to the microphysical characteristics of a sounding medium and to suggest the method of their retrieval by measuring the Raman lidar return with a multiple-field-of-view receiver, parting multiple scattering. In particular, the method of measuring the extinction coefficient and also the effective radius of droplets in warm clouds was described. The method could have an advantage over the similar method when using elastic scattering, because of the simple angular behaviour of the Raman backscattering phase function.

**Acknowledgment**

The author is grateful to Eleonora P. Zege for her support at all stages of the work and to Vera Lozanova for her help in preparing this chapter.

## References

Ansmann, A., M. Reibesell, and C. Weitkamp, 1990: Measurement of atmospheric aerosol extinction profiles with Raman lidar, *Opt. Lett.*, **15**, 746–748.

Ansmann, A., M. Riebesell, U. Wandinger, C. Weitkamp, E. Voss, W. Lahmann, and W. Michaelis, 1992a: Combined Raman elastic-backscatter lidar for vertical profiling of moisture, aerosol extinction, backscatter, and lidar ratio, *Appl. Phys. B*, **55**, 18–28.

Ansmann, A., U. Wandinger, M. Riebesell, and C. Weitkamp, 1992b: Independent measurement of extinction and backscatter profiles in cirrus clouds by using a combined Raman elastic-backscatter lidar, *Appl. Opt.*, **31**, 7113–7131.

Bissonnette, L., and D. Hutt, 1990: Multiple scattering lidar, *Appl. Opt.*, **29**, 5045–5046.

Bruscaglioni, P., M. Gai, and A. Ismaelli, 1999: Molecular lidar and Mie multiple scattering, MUSCLE 10 (International Workshop on Multiple Scattering Lidar Experiments): Proceedings, Florence, Italy, 19-22 April 1999, 206-211.

Cohen, A., M. Kleiman, and J. Cooney, 1978: Lidar measurements of rotational Raman and double scattering, *Appl. Opt.*, **17**, 1905–1910.

Cooney, J., 1972: Measurements of atmospheric temperature profiles by Raman backscatter, *Appl. Meteorol.*, **11**, 108–112.

Deirmendjian, D., 1969: *Electromagnetic Scattering on Spherical Polydispersions*, Elsevier, New York.

Egert, S., A. Cohen, M. Kleiman, and N. Ben-Yosef, 1983: Instantaneous integrated Raman scattering, *Appl. Opt.*, **22**, 1592–1597.

Eloranta, E., 1972: Calculation of doubly scattered lidar returns, Ph.D. dissertation, University of Wisconsin, Madison.

Fiocco, L., Z. Smullin, 1963: Detection of scattering layers in the upper atmosphere (60–140 km) by optical radar, *Nature*, **199**, 1275–1276.

Girolamo, P. D., A. Behrendt, and V. Wulfmeyer, 2006: Spaceborne profiling of atmospheric temperature and particle extinction with pure rotational Raman lidar and of relative humidity in combination with differential absorption lidar : performance simulations, *Appl. Opt.*, **45**, 2474–2494.

Grund C., and E. Eloranta, 1991: University of Wisconsin high spectral resolution lidar, *Opt. Eng.*, **30**, 6–12.

Katsev, I. , E. Zege, A. Prikhach, and I. Polonsky, 1997: Efficient technique to determine backscattered light power for various atmospheric and oceanic sounding and imaging systems, *JOSA A*, **14**, 1338–1346.

Klyshko, D. N., and V. V. Fadeev, 1978: Remote detecting the water impurity by means of laser spectroscopy calibrated by Raman scattering, *Reports of Academy of Sciences of the USSR*, **238**, 320–323 (in Russian).

Kokhanovsky, A. A., and E. P. Zege, 1997: Parametrization of local optical characteristics of cloudy media, *Izv. Atmos. and Ocean. Phys.*, **33**, 190–198.

Lee, Z. P., K. L. Carder, S. K. Hawes, R. G. Steward, T. G. Peacock, and C. O. Davis, 1994: Model for the interpretation of hyperspectral remote-sensing reflectance, *Appl. Opt.*, **33**, 5721–5732.

Malinka, A., and E. Zege, 2001: Analytical modeling of the Raman lidar return including multiple scattering, *Current Problems in Optics of Natural Waters (ONW-2001)*: Proceedings of International Conference, St. Petersburg, 25–29 September 2001, 176–181.

Malinka, A., and E. Zege, 2003a: Analytical modeling of the Raman lidar return from clouds and ocean including multiple scattering, *MUSCLE-12, Proc. of SPIE*, **5059**, 153–159.

Malinka, A., and E. Zege, 2003b: Analytical modeling of the Raman lidar return, including multiple scattering, *Appl. Opt.*, **24**, 1075–1081.

Malinka, A., and E. Zege, 2003c: Using multiple scattering in Raman lidar sounding of warm clouds, *6th International Symposium on Tropospheric Profiling: Needs and Technologies* (ISTP-2003): Proceedings, Leipzig, Saxony, Germany, 14–20 September 2003, 297–299.

Mattis, I., A. Ansmann, D. Althausen, V. Jaenisch, U. Wandinger, D. Mueller, Y. Arshinov, S. Bobrovnikov, and I. Serikov, 2002: Relative-humidity profiling in the troposphere with a Raman lidar, *Appl. Opt.*, **41**, 6451–6462.

McLean J. W., J. D. Freeman, and R.E. Walker, 1998: Beam Spread Function with Time Dispersion, *Appl. Opt.*, **37**, 4701–4711.

Mobley, C., B. Gentili, H. Gordon, Z. Sin, G. Kattautar, H. Morel, P. Reinersman, K. Stamnes, and R. Stavn, 1993: Comparison of numerical models for computing underwater light fields, *App. Opt.*, **32**, 7484–7503.

Mueller, D., U. Wandinger, and A. Ansmann, 1999a: Microphysical particle parameters from extinction and backscatter lidar data by inversion with regularization: theory, *Appl. Opt.*, **38**, 2346–2357.

Mueller, D., U. Wandinger, and A. Ansmann, 1999b: Microphysical particle parameters from extinction and backscatter lidar data by inversion with regularization: simulation, *Appl. Opt.*, **38**, 2358–2367.

Mueller, D., U. Wandinger, and A. Ansmann, 2000: Microphysical particle parameters from extinction and backscatter lidar data by inversion with regularization: experiment, *Appl. Opt.*, **39**, 1879–1892.

Muscari, G., M. Cacciani, G. Giuliani, and G. Fiocco, 1996: An imaging lidar for multiple scattering measurements, Advances in Atmospheric Remote Sensing with Lidar: selected papers of the 18th International Laser Radar Conference (ILRC), Berlin, 22–26 July 1996, Springer, Berlin, 103–106.

Piironen, P., E. Eloranta, 1994: Demonstration of a high-spectral-resolution lidar based on an iodine absorption filter, *Opt. Lett.*, **19**, 234–236.

Polonsky, I. N., E. P. Zege, I. L. Katsev, 2001: Lidar sounding of warm clouds and determination of their microstructure parameters, *Izv. Atmos. and Ocean Phys.*, **37**, 624–632.

Reichardt, J., M. Hess, A. Macke, 2000: Lidar inelastic multiple-scattering parameters of cirrus particle ensembles determined with geometrical-optics crystal phase functions, *Appl. Opt.*, **39**, 1895–1910.

Reichardt, J., U. Wandinger, M. Serwazi, and C. Weitkamp, 1996: Combined Raman lidar for aerosol, ozone, and moisture measurements, *Opt. Eng.*, **5**, 1457–1465.

Roy, G., L. Bissonnette, C. Bastille, and G. Vallee, 1997: Estimation of cloud droplet-size density distribution from multiple-field-of-view lidar returns, *Opt. Eng.*, **36**, 3404–3415.

Samokhvalov, I.V., 1979: Double scattering approximation of the lidar equation for inhomogeneous atmosphere, *Opt. Lett.*, **4**, 12–14.

Sherlock, V., A. Hauchecorne, and J. Lenoble, 1999a: Methodology for the independent calibration of Raman backscatter water-vapour lidar system, *Appl. Opt.*, **38**, 5816–5837.

Sherlock, V., A. Garnier, A. Hauchecorne, and P. Keckhut, 1999b: Implementation and validation of a Raman lidar measurement of middle and upper tropospheric water vapour, *Appl. Opt.*, **38**, 5838–5850.

Shipley, S. et al., 1983: High spectral resolution lidar to measure optical scattering properties of atmospheric aerosols. 1. Theory and instrumentation, *Appl. Opt.*, **22**, 3716–3724.

Tomasi, F., M. R. Perrone, and M. L. Protopapa, 2001: Monitoring $O_3$ with Solar-Blind Raman Lidars, *Appl. Opt.*, **40**, 1314–1320.

van de Hulst, H. C., 1957: Light Scattering by Small Particles, Chapman & Hall, London.

Wandinger, U., 1998: Multiple-scattering influence on extinction and backscatter-coefficient measurement with Raman and high-spectral-resolution lidars, *Appl. Opt.*, **37**, 417–427.

Wandinger, U., A. Ansmann, J. Reichardt, and T. Deshler, 1995: Determination of stratospheric-aerosol microphysical properties from independent extinction and backscattering measurements with Raman lidar, *Appl. Opt.*, **34**, 8315–8329.

Weinman, J., and S. Shipley, 1972: Effects of multiple scattering on laser pulses transmitted through clouds, *J. of Geophys. Res.*, **77**, 7123–7128.

Whiteman, D., and S. Melfi, 1999: Cloud liquid water, mean droplet radius, and number density measurements using a Raman lidar, *J. Geophys. Res. D*, **134**, 31411–31419.

Zege, E., A. Ivanov, and I. Katsev, 1991: Image Transfer Through a Scattering Medium, Springer-Verlag, Berlin, Heidelberg.

Zege, E., I. Katsev, and I. Polonsky, 1995: Analytical solution to LIDAR return signals from clouds with regards to multiple scattering, *Appl. Phys. B*, **60**, 345–354.

Zege, E. et al., 2003a: Analytical and computer modeling of the ocean lidar performance, *MUSCLE-12, Proc. of SPIE*, **5059**, 189–199.

Zege E., I. Katsev, A. Prikhach, and G. Ludbrook, 2001: Computer simulation with regard to pulse stretching for oceanic lidar return, *ONW-2001 (Current Problems in Optics of Natural Waters)*: Proceedings of the International Conference, St. Petersburg, Russia, 25–29 September 2001, 255–262.

Part II

# Inverse Problems

# 5 Linearization of vector radiative transfer by means of the forward-adjoint perturbation theory and its use in atmospheric remote sensing

Otto P. Hasekamp and Jochen Landgraf

## 5.1 Introduction

Aerosols directly affect the Earth's climate by scattering and absorption of radiation, and indirectly by changing the microphysical properties of clouds. The total effect of aerosols on climate is very uncertain, both in magnitude and even in sign, representing one of the largest uncertainties in climate research. In order to improve our understanding of the effect of aerosols on climate, global measurements are needed of a number of aerosol properties such as size of particles, their refractive index and aerosol optical thickness. The only way to obtain these parameters at a global scale is by means of satellite remote sensing.

Information on aerosol properties is contained in the spectral and angular behavior of the total intensity and the polarization properties of backscattered sunlight. Most satellite instruments that are used for aerosol retrieval only measure the intensity spectrum of backscattered light. Among these instruments are the Advanced Very High Resolution Radiometer (AVHRR), the Moderate Resolution Imaging Spectroradiometer (MODIS), the Total Ozone Mapping Spectrometer (TOMS), the Global Ozone Monitoring Experiment (GOME), the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY), the Medium Resolution Imaging Sensor (MERIS), and the Ozone Monitoring Instrument (OMI). Retrieval algorithms for instruments of this type allow the choice between a number of standard aerosol models (a combination of size distribution and refractive index), where the model that agrees best with the measured spectrum is used to determine the aerosol optical thickness. On the one hand these intensity -only retrievals do not provide enough information to answer the relevant questions in climate research, and on the other hand the retrieved optical thickness depends critically on the choice of the aerosol model. The information content with respect to aerosol properties is significantly larger for multiple-viewing-angle intensity measurements as performed by the Multi-angle Imaging Spectro-Radiometer (MISR), for multiple-viewing-angle intensity and polarization measurements as performed by the Polarization and Anisotropy of Reflectances for Atmospheric Sciences Coupled with Observations from a Lidar instrument (PARASOL), and for single-viewing-angle intensity and polar-

ization measurements, as performed by GOME-2. These instruments contain enough information to do more than just distinguishing between a number of aerosol models. Therefore, the 'classical' retrieval approach described above is not sufficient for these more advanced instruments. Instead, a retrieval approach is required that makes full use of the information content of the measurement.

Recently a new approach to the retrieval of aerosol properties has been developed (Hasekamp and Landgraf, 2005a,b). Instead of assuming a number of standard aerosol models, the developed method aims to retrieve microphysical aerosol properties corresponding to a bi-modal aerosol size distribution. The retrieval of these aerosol parameters from satellite measurements requires a forward model $\mathbf{F}$ that describes how the measured data depend on the aerosol parameters, viz.

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) + \mathbf{e}_y. \tag{5.1}$$

Here $\mathbf{y}$ is the measurement vector containing the measured data, e.g. intensity and/or polarization measurements at different wavelengths and/or different viewing angles, and $\mathbf{e}_y$ is the corresponding error vector. $\mathbf{x}$ is the state vector containing the aerosol parameters to be retrieved. The forward model consists of two parts. The first part relates the physical aerosol properties (size distribution, refractive index) to their optical properties (scattering and extinction coefficients, phase matrix). This relation can be described by Mie theory for spherical particles (van der Hulst, 1957) or alternative theories for particles of other shapes (Wiscombe and Grams, 1988; Koepke and Hess, 1988; Mishchenko and Travis, 1994; Mishchenko et al., 1995). The second part of the forward model is an atmospheric radiative transfer model that simulates the intensity vector at the top of the atmosphere for given optical input parameters. Since the forward model $\mathbf{F}$ is non-linear in the microphysical aerosol parameters contained in $\mathbf{x}$, the inversion of Eq. (5.1) has to be performed iteratively. Hereto, the forward model $\mathbf{F}$ in Eq. (5.1) is replaced by its linear approximation in each iteration step $n$

$$\mathbf{y} \approx \mathbf{F}(\mathbf{x}_\mathrm{n}) + \mathbf{K}\,(\mathbf{x} - \mathbf{x}_n) + \mathbf{e}_y, \tag{5.2}$$

where $\mathbf{x}_\mathrm{n}$ is the state vector for the iteration step under consideration and $\mathbf{K}$ is the Jacobian matrix containing the derivatives of the forward model with respect to the elements of $\mathbf{x}$, where element $K_{ij}$ of $\mathbf{K}$ is defined by:

$$K_{ij} = \frac{\partial F_i}{\partial x_j}(\mathbf{x}_\mathrm{n}). \tag{5.3}$$

The inversion of Eq. (5.2) can be performed analytically. Once $\mathbf{x}_n$ is close enough to the true state vector $\mathbf{x}$, the Jacobian matrix $\mathbf{K}$ can be used to calculate the mapping of the measurement errors $\mathbf{e}_y$ to errors on the retrieved aerosol parameters $\mathbf{e}_x$. Thus, the Jacobian matrix $\mathbf{K}$ plays an important role in the retrieval process, both for finding an appropriate solution of the inversion problem and for a solid error analysis. Therefore, in the most general case of aerosol retrieval from intensity and polarization measurements, a linearized vector radiative transfer model is needed that simulates the intensity vector at the top of the model at-

mosphere and additionally calculates the derivatives of the Stokes parameters with respect to the aerosol properties to be retrieved.

For scalar radiative transfer, a general linearization approach was proposed by Marchuk (1964) who employed the forward-adjoint perturbation theory approach, known from neutron transport theory (see, for example, Bell and Classtone (1970)), to atmospheric scalar radiative transfer. This approach has been used by, for example, Ustinov (1991), Rozanov et al. (1998) and Landgraf et al. (2002) for the linearization of scalar radiative transfer with respect to atmospheric absorption properties, by Landgraf et al. (2001) for the linearization with respect to surface properties, and by, for example, Ustinov (1992) and Sendra and Box (2000) for the linearization with respect to atmospheric scattering properties. Another approach for the linearization of scalar radiative transfer has been followed by Spurr et al. (2001), who developed an analytical linearization with respect to absorption and scattering properties for the discrete ordinate method of scalar radiative transfer (Chandrasekhar, 1960; Stamnes et al., 1988). For plane-parallel vector radiative transfer, an analytical linearization with respect to atmospheric absorption properties has been developed by Hasekamp and Landgraf (2002), who extended the forward-adjoint perturbation theory to include polarization.

An analytical linearization of vector radiative transfer with respect to atmospheric scattering properties has recently been achieved by Hasekamp and Landgraf (2005a). They combined the linearization of radiative transfer with a linearization of Mie theory to obtain a radiative transfer model that provides the requested derivatives of the Stokes parameters at the top of the atmosphere with respect to microphysical aerosol properties. Based on this linearized vector radiative transfer model, Hasekamp and Landgraf (2005b) developed a novel approach to the retrieval of microphysical aerosol properties from multi-spectral single-viewing-angle measurements of intensity and polarization. This retrieval approach uses the Phillips–Tikhonov regularization method for the analytical inversion of the linearized radiative transfer model. A powerful feature of this approach is that it quantifies the information content of the measurement as part of the retrieval procedure and extracts the available information. The aim of this chapter is to review this retrieval approach based on linearized radiative transfer and Phillips–Tikhonov regularization. The first part of this chapter (section 5.2–5.5) is devoted to the linearization of vector radiative transfer with respect to microphysical aerosol properties. The second part of the chapter (section 5.6 and 5.7) discusses the application of the linearized vector radiative transfer model in a retrieval scheme using Phillips–Tikhonov regularization.

## 5.2 Radiative transfer model

The radiance and state of polarization of light at a given wavelength can be described by an intensity vector $\mathbf{I}$ which has the Stokes parameters as its components (see, for example, Chandrasekhar (1960)):

$$\mathbf{I} = [I, Q, U, V]^T , \tag{5.4}$$

where $T$ indicates the transposed vector, and the Stokes parameters are defined with respect to a certain reference plane. The angular dependence of single scattering of polarized light can be described by means of the scattering phase matrix $\mathbf{P}$. We will restrict ourselves to scattering phase matrices of the form

$$\mathbf{P}(\theta) = \begin{pmatrix} p_1(\theta) & p_5(\theta) & 0 & 0 \\ p_5(\theta) & p_2(\theta) & 0 & 0 \\ 0 & 0 & p_3(\theta) & p_6(\theta) \\ 0 & 0 & -p_6(\theta) & p_4(\theta) \end{pmatrix}. \tag{5.5}$$

where $p_1, p_2, \ldots, p_6$ are certain functions of scattering angle $\theta$ and the scattering plane is the plane of reference. This type of scattering matrix is valid for (see, for example, van de Hulst (1957)) (i) scattering by an assembly of randomly oriented particles each having a plane of symmetry, (ii) scattering by an assembly containing particles and their mirror particles in equal numbers and with random orientations, (iii) Rayleigh scattering with or without depolarization effects.

To discuss the single scattering properties of aerosol particles we will use the scattering plane as the plane of reference. However, for the atmospheric radiative transfer calculations in this chapter we will use the local meridian plane, defined as the plane going through the direction of propagation and the vertical direction, as reference plane.

### 5.2.1 Radiative transfer equation in operator form

We consider a plane-parallel, macroscopically isotropic atmosphere bounded below by a reflecting surface. Furthermore, we ignore inelastic scattering and thermal emission. The equation of transfer for polarized light is now given in its forward formulation by

$$\hat{\mathbf{L}}\,\mathbf{I} = \mathbf{S}, \tag{5.6}$$

where the transport operator

$$\begin{aligned} \hat{\mathbf{L}} \;=\; & \int_{4\pi} \mathrm{d}\tilde{\Omega} \Bigg\{ \left[ \mu \frac{\partial}{\partial z} + K_{\mathrm{ext}}(z) \right] \delta(\mathbf{\Omega} - \tilde{\mathbf{\Omega}})\, \mathbf{E} - \\ & \frac{K_{\mathrm{sca}}(z)}{4\pi}\, \mathbf{Z}(z, \tilde{\mathbf{\Omega}}, \mathbf{\Omega}) - \delta(z)\, \Theta(\mu) |\mu|\, \mathbf{R}_s(\tilde{\mathbf{\Omega}}, \mathbf{\Omega})\, \Theta(-\tilde{\mu}) |\tilde{\mu}| \Bigg\}, \end{aligned} \tag{5.7}$$

is adopted from scalar radiative transfer (Marchuk, 1964; Box et al., 1988; Ustinov, 2001; Landgraf et al., 2002). Here, $z$ describes altitude, the direction $\mathbf{\Omega}$ is given by $(\mu, \varphi)$ where $\varphi$ is the azimuthal angle measured clockwise when looking downward and $\mu$ is the cosine of the zenith angle ($\mu < 0$ for downward directions and $\mu > 0$ for upward directions). Furthermore, $d\Omega = d\mu\, d\varphi$, $\mathbf{E}$ is the $4 \times 4$ unity matrix, $K_{\mathrm{ext}}$ and $K_{\mathrm{sca}}$ represent the extinction and scattering coefficients, respectively, $\Theta$ represents the Heaviside step function, and $\delta$ is the Dirac-delta function with $\delta(\mathbf{\Omega} - \tilde{\mathbf{\Omega}}) = \delta(\mu - \tilde{\mu})\delta(\varphi - \tilde{\varphi})$. The first term of the radiative transfer operator describes the extinction of light, whereas the second

term represents scattering of light from direction $\tilde{\mathbf{\Omega}}$ to $\mathbf{\Omega}$ with the phase matrix $\mathbf{Z}(z, \tilde{\mathbf{\Omega}}, \mathbf{\Omega})$, defined with respect to the local meridian plane. The last term on the right-hand side of Eq. (5.7) describes the surface reflection at the lower boundary of the atmosphere with reflection matrix $\mathbf{R}_s$.

The right-hand side of Eq. (5.6) provides the source of light and can either be a volume source inside the atmosphere or a surface source chosen to reproduce the incident flux conditions at the boundaries of the atmosphere, or some combination of the two. In the UV and visible part of the spectrum the radiation source $\mathbf{S}$ is determined by the unpolarized sunlight that illuminates the top of the Earth atmosphere:

$$\mathbf{S}(z, \mathbf{\Omega}) = \mu_o \delta(z - z_{\text{top}}) \delta(\mathbf{\Omega} - \mathbf{\Omega}_o) \mathbf{F}_o. \tag{5.8}$$

Here, $z_{\text{top}}$ is the height of the model atmosphere, $\mathbf{\Omega}_o = (-\mu_o, \varphi_o)$ describes the geometry of the incoming solar beam (we define $\mu_o > 0$), and $\mathbf{F}_o$ is given by

$$\mathbf{F}_o = [F_o, 0, 0, 0]^T, \tag{5.9}$$

where $F_o$ is the solar flux per unit area perpendicular to the direction of the solar beam. Because the reflection of light at the ground surface is already included in the radiative transfer operator (5.7) and the incoming solar beam is represented by the radiation source of Eq. (5.8), the intensity vector $\mathbf{I}$ is subject to homogeneous boundary conditions:

$$\begin{aligned} \mathbf{I}(z_{\text{top}}, \mathbf{\Omega}) &= [0, 0, 0, 0]^T \qquad \text{for } \mu < 0, \\ \mathbf{I}(0, \mathbf{\Omega}) &= [0, 0, 0, 0]^T \qquad \text{for } \mu > 0. \end{aligned} \tag{5.10}$$

In conjunction with these boundary conditions, the radiation source $\mathbf{S}$ can be interpreted as located at a vanishingly small distance below the upper boundary. Similarly, the surface reflection takes place at a vanishingly small distance above the lower boundary (see, for example, Morse and Feshbach (1953)).

In order to handle the integration over azimuth angle in Eq. (5.6) we use a decomposition of the radiative transfer equation into corresponding equations per Fourier component (Hovenier and van der Mee, 1983; de Haan et al., 1987):

$$\hat{\mathbf{L}}^m \, \mathbf{I}^{\pm m} = \mathbf{S}^{\pm m}, \tag{5.11}$$

with

$$\begin{aligned} \hat{\mathbf{L}}^m = \int_{-1}^{1} d\tilde{\mu} \Bigg\{ & \left[ \mu \frac{\partial}{\partial z} + K_{\text{ext}}(z) \right] \delta(\mu - \tilde{\mu}) \, \mathbf{E} - \\ & \frac{K_{\text{sca}}(z)}{2} \, \mathbf{Z}^m(z, \tilde{\mu}, \mu) - \delta(z) \, \Theta(\mu) |\mu| \, \mathbf{R}_s^m(\tilde{\mu}, \mu) \, \Theta(-\tilde{\mu}) |\tilde{\mu}| \Bigg\}. \end{aligned} \tag{5.12}$$

The corresponding Fourier expansion of the intensity vector is given by

$$\mathbf{I}(z,\mathbf{\Omega}) = \sum_{m=0}^{\infty}(2-\delta_{m0})\left[\mathbf{B}^{+m}(\varphi_o-\varphi)\mathbf{I}^{+m}(z,\mu)+\mathbf{B}^{-m}(\varphi_o-\varphi)\mathbf{I}^{-m}(z,\mu)\right], \tag{5.13}$$

where $\delta_{m0}$ is the Kronecker delta, and

$$\mathbf{B}^{+m}(\varphi) = \mathrm{diag}[\cos m\varphi, \cos m\varphi, \sin m\varphi, \sin m\varphi], \tag{5.14}$$
$$\mathbf{B}^{-m}(\varphi) = \mathrm{diag}[-\sin m\varphi, -\sin m\varphi, \cos m\varphi, \cos m\varphi]. \tag{5.15}$$

The Fourier coefficients of the intensity vector are given by

$$\mathbf{I}^{+m}(z,\mu) = \frac{1}{2\pi}\int_0^{2\pi} d\varphi\ \mathbf{B}^{+m}(\varphi_o-\varphi)\ \mathbf{I}(z,\mathbf{\Omega}),$$
$$\mathbf{I}^{-m}(z,\mu) = \frac{1}{2\pi}\int_0^{2\pi} d\varphi\ \mathbf{B}^{-m}(\varphi_o-\varphi)\ \mathbf{I}(z,\mathbf{\Omega}). \tag{5.16}$$

Similarly, a Fourier expansion of the radiation source $\mathbf{S}(z,\mathbf{\Omega})$ is obtained, with Fourier coefficients

$$\mathbf{S}^{+m}(z,\mu) = \frac{1}{2\pi}\,\mu_o\,\delta(z-z_{top})\,\delta(\mu_o+\mu)\,\mathbf{F}_o,$$
$$\mathbf{S}^{-m}(z,\mu) = [0,0,0,0]^T. \tag{5.17}$$

From Eqs (5.10) and (5.17) it follows that $\mathbf{I}^{-m}=0$, so the Fourier expansion of the intensity vector contains terms of $\mathbf{I}^{+m}$ only.

The Fourier expansion of the phase matrix is given by

$$\mathbf{Z}(z,\tilde{\mathbf{\Omega}},\mathbf{\Omega}) = \frac{1}{2}\sum_{m=0}^{\infty}(2-\delta_{m0})\big[\mathbf{B}^{+m}(\tilde{\varphi}-\varphi)\mathbf{Z}^m(z,\tilde{\mu},\mu)(\mathbf{E}+\mathbf{\Lambda}) + \mathbf{B}^{-m}(\tilde{\varphi}-\varphi)\mathbf{Z}^m(z,\tilde{\mu},\mu)(\mathbf{E}-\mathbf{\Lambda})\big], \tag{5.18}$$

where

$$\mathbf{\Lambda} = \mathrm{diag}\,[1,1,-1,-1]\,. \tag{5.19}$$

The $m$th Fourier coefficient of the phase matrix can be calculated by

$$\mathbf{Z}^m(z,\tilde{\mu},\mu) = (-1)^m\sum_{l=m}^{L}\mathbf{P}^l_{\ m}(-\mu)\ \mathbf{S}^l(z)\ \mathbf{P}^l_{\ m}(-\tilde{\mu}), \tag{5.20}$$

where $L$ is a suitable truncation index (Ustinov, 1988) and $\mathbf{P}^l_{\ m}$ is the generalized spherical function matrix given by

$$\mathbf{P}^m_l(\mu) = \begin{pmatrix} P^l_{m0}(\mu) & 0 & 0 & 0 \\ 0 & P^l_{m+}(\mu) & P^l_{m-}(\mu) & 0 \\ 0 & P^l_{m-}(\mu) & P^l_{m+}(\mu) & 0 \\ 0 & 0 & 0 & P^l_{m0}(\mu) \end{pmatrix}, \tag{5.21}$$

where

$$P^l_{m\pm} = \frac{1}{2}\left(P^l_{m,-2} \pm P^l_{m,2}\right), \tag{5.22}$$

and $P^l_{mn}(\mu)$ are the generalized spherical functions (Gel'fand et al., 1963), which were introduced in atmospheric radiative transfer by Kuščer and Ribarič (1959). $\mathbf{S}^l$ is the expansion coefficient matrix having the form

$$\mathbf{S}^l = \begin{pmatrix} \alpha^l_1 & \alpha^l_5 & 0 & 0 \\ \alpha^l_5 & \alpha^l_2 & 0 & 0 \\ 0 & 0 & \alpha^l_3 & \alpha^l_6 \\ 0 & 0 & -\alpha^l_6 & \alpha^l_4 \end{pmatrix}, \tag{5.23}$$

where the expansion coefficients follow from the scattering phase matrix $\mathbf{P}$ in Eq. (5.5) (see, for example, de Rooij and van der Stap (1984)):

$$\alpha^l_1 = \frac{2l+1}{2}\int_{-1}^{1} P^l_{0,0}(\cos\theta) p_1(\theta)\, d(\cos\theta), \tag{5.24}$$

$$\alpha^l_2 + \alpha^l_3 = -\frac{2l+1}{2}\sqrt{\frac{(l-2)!}{(l+2)!}}\int_{-1}^{1} P^l_{2,2}(\cos\theta)(p_2(\theta)+p_3(\theta))\,\mathrm{d}(\cos\theta), \tag{5.25}$$

$$\alpha^l_2 - \alpha^l_3 = -\frac{2l+1}{2}\sqrt{\frac{(l-2)!}{(l+2)!}}\int_{-1}^{1} P^l_{2,-2}(\cos\theta)(p_2(\theta)+p_3(\theta))\,\mathrm{d}(\cos\theta), \tag{5.26}$$

$$\alpha^l_4 = \frac{2l+1}{2}\int_{-1}^{1} P^l_{0,0}(\cos\theta) p_4(\theta)\,\mathrm{d}(\cos\theta), \tag{5.27}$$

$$\alpha^l_5 = \frac{2l+1}{2}\int_{-1}^{1} P^l_{0,2}(\cos\theta) p_5(\theta)\,\mathrm{d}(\cos\theta), \tag{5.28}$$

$$\alpha^l_6 = \frac{2l+1}{2}\int_{-1}^{1} P^l_{0,2}(\cos\theta) p_6(\theta)\,\mathrm{d}(\cos\theta). \tag{5.29}$$

In this chapter we assume that the surface reflection matrix obeys the same symmetry relations as the scattering phase matrix (Hovenier, 1969) and thus can also be expanded in a Fourier series. The Fourier coefficients $\mathbf{R}^m_s$ of the surface reflection matrix $\mathbf{R}_s$ are given by:

$$\mathbf{R}^m_s(\tilde{\mu},\mu) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}(\tilde{\varphi}-\varphi)\left[\mathbf{B}^{+m}(\tilde{\varphi}-\varphi) + \mathbf{B}^{-m}(\tilde{\varphi}-\varphi)\right]\, \mathbf{R}_s(\tilde{\mathbf{\Omega}},\mathbf{\Omega}). \tag{5.30}$$

In order to obtain the Fourier coefficients $\mathbf{I}^m$ from Eq. (5.11) for a vertically inhomogeneous atmosphere, the model atmosphere has to be divided in a number of homogeneous layers, where each layer is characterized by a height-independent scattering coefficient, extinction coefficient, and scattering matrix. Several numerical models exist to solve the corresponding radiative transfer problem. We will use the Gauss–Seidel model described by Landgraf et al. (2002) and Hasekamp and Landgraf (2002).

## 5.3 Mie scattering calculations

The optical input parameters of the radiative transfer equation (5.6) are the extinction and scattering coefficients and the phase matrix $\mathbf{Z}$ in the form of expansion coefficient matrices $\mathbf{S}^l$. These parameters are determined by scattering and absorption by aerosols, scattering by air molecules, and absorption by atmospheric gases, and are obtained from these different components:

$$K_{\text{ext}} = K^a_{\text{ext}} + K^r_{\text{ext}} + K^g_{\text{ext}}, \tag{5.31}$$

$$K_{\text{sca}} = K^a_{\text{sca}} + K^r_{\text{sca}}, \tag{5.32}$$

$$\alpha = \frac{K^a_{\text{sca}}\,\alpha^a}{K_{\text{sca}}} + \frac{K^r_{\text{sca}}\,\alpha^r}{K_{\text{sca}}}, \tag{5.33}$$

where the superscript $a$ denotes aerosol, the superscript $r$ denotes Rayleigh scattering, the superscript $g$ denotes gas absorption, and we omitted the sub- and superscripts of the expansion coefficients $\alpha^l_j$.

The optical properties of aerosols depend on the size, shape, and type of aerosols. In this chapter we restrict ourselves to spherical aerosols which means that the optical properties of aerosols can be calculated using Mie theory. Here we will summarize the most important formulas needed for Mie calculations (see, for example, de Rooij and van der Stap (1984)). A complete overview of Mie scattering theory is given by van de Hulst (1957).

In order to calculate the elements of the Mie scattering phase matrix $\mathbf{P}$ in Eq. (5.5), we first consider the transformation matrix $\boldsymbol{\mathcal{F}}$ (van de Hulst, 1957) which is defined as

$$\boldsymbol{\mathcal{F}} = \frac{k^2\,C_{\text{sca}}}{4\pi}\,\mathbf{P}, \tag{5.34}$$

with elements $f_1, f_2, \ldots, f_6$, analogous to the elements $p_1, p_2, \ldots, p_6$ in Eq. (5.5). In Eq. (5.34) $C_{\text{sca}}$ is the scattering cross-section, and $k = 2\pi/\lambda$, where $\lambda$ denotes wavelength. For a single sphere of radius $r$ the elements of the transformation matrix $\boldsymbol{\mathcal{F}}$ are given by

$$f_1 = \frac{1}{2}\,(S_1 S_1^* + S_2 S_2^*)\,, \tag{5.35}$$

$$f_2 = f_1, \tag{5.36}$$

$$f_3 = \frac{1}{2}\,(S_1 S_2^* + S_2 S_1^*)\,, \tag{5.37}$$

$$f_4 = f_3, \tag{5.38}$$

$$f_5 = \frac{1}{2}\,(S_1 S_1^* - S_2 S_2^*)\,, \tag{5.39}$$

$$f_6 = \frac{i}{2}\,(S_1 S_2^* - S_2 S_1^*)\,, \tag{5.40}$$

where we omitted the dependence on scattering angle $\theta$ and particle radius $r$. In Eqs (5.35)–(5.40), $S_1$ and $S_2$ are the elements of the two-by-two scattering amplitude matrix relating the electric field vector (containing the component

parallel and the component perpendicular to the scattering plane) of the scattered beam to that of the incoming beam, and the asterisk denotes the complex conjugate. The functions $S_1$ and $S_2$ are given by

$$S_1(\theta) = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} \left(a_n \pi_n(\theta) + b_n \tau_n(\theta)\right), \tag{5.41}$$

$$S_2(\theta) = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} \left(b_n \pi_n(\theta) + a_n \tau_n(\theta)\right), \tag{5.42}$$

where $\pi_n$ and $\tau_n$ are functions of only scattering angle and are expressed in associated Legendre functions as

$$\pi_n(\theta) = \frac{1}{\sin\theta} P_n^1(\cos\theta), \tag{5.43}$$

$$\tau_n(\theta) = \frac{d}{d\theta} P_n^1(\cos\theta). \tag{5.44}$$

The most substantial part of the Mie calculations is the computation of the Mie coefficients $a_n$ and $b_n$ in Eqs (5.41) and (5.42) which are functions of the particle's complex refractive index $m = m_r + im_i$ and the size parameter $kr$. A numerical procedure for calculating $a_n$ and $b_n$ is given by Rooij and van der Stap (1984), and is summarized in Appendix A of this chapter. The scattering and extinction cross-sections, $C_{\text{sca}}$ and $C_{\text{ext}}$, of a single sphere with radius $r$ can also be calculated using the coefficients $a_n$ and $b_n$:

$$C_{\text{sca}}(r) = \frac{2\pi}{k^2} \sum_{n=1}^{\infty} (2n+1)[|a_n|^2 + |b_n|^2], \tag{5.45}$$

$$C_{\text{ext}}(r) = \frac{2\pi}{k^2} \sum_{n=1}^{\infty} (2n+1)\text{Re}(a_n + b_n). \tag{5.46}$$

Equations (5.35)–(5.42), (5.45), and (5.46) provide expressions for the scattering matrix and absorption and extinction cross-sections for a single sphere. In nature, a distribution of particles with different sizes is normally encountered. Under the assumption of independent scattering (see, for example, Hansen and Travis (1974)) the scattering and extinction cross-section for a size distribution are given by

$$\bar{C}_{\text{sca}} = \int_0^{\infty} C_{\text{sca}}(r)\, n(r)\, dr, \tag{5.47}$$

$$\bar{C}_{\text{ext}} = \int_0^{\infty} C_{\text{ext}}(r)\, n(r)\, dr, \tag{5.48}$$

where $n(r)$ is the aerosol size distribution normalized to unity (e.g. a lognormal distribution; see Appendix B).

Similarly, element $f_j$ of the transformation matrix $\boldsymbol{\mathcal{F}}$ (5.34) for a size distribution is given by

$$\bar{f}_j = \int_0^\infty f_j(r)\, n(r)\, \mathrm{d}r. \tag{5.49}$$

In this chapter the integrations over size distribution are approximated by a sum over different size bins. Here, for a function $g(r)$ the integral over size distribution is approximated by

$$\bar{g} \approx \sum_{i=1}^{N} g(r_i)\, n_i\, \Delta r_i, \tag{5.50}$$

where $r_i$ is the middle of size interval $i$, $n_i = n(r_i)$ and $\Delta r_i$ is the width of size interval $i$.

The elements $p_j$ of the scattering phase matrix $\mathbf{P}$ of an ensemble with given size distribution can be obtained from $\bar{f}_j$ and $\bar{C}_{\mathrm{sca}}$ via

$$p_j = \frac{4\pi}{k^2}\, \frac{\bar{f}_j}{\bar{C}_{\mathrm{sca}}}. \tag{5.51}$$

The expansion coefficients $\alpha_i^j$ can now be calculated from the elements of the scattering phase matrix $\mathbf{P}$ using Eqs (5.24)–(5.29). Here, the integrals over $\cos\theta$ can be calculated analytically (Domke, 1975) or numerically (de Rooij and van der Stap, 1984). The latter approach will be adopted for the calculations in this chapter. Furthermore, the aerosol extinction and scattering coefficients $K^a_{\mathrm{ext}}$ and $K^a_{\mathrm{sca}}$ for each homogeneous layer are obtained by multiplying the corresponding cross-section by the layer integrated aerosol number concentration in that layer.

## 5.4 Linearization of the forward model

As follows from the previous section, the relevant optical properties of spherical homogeneous aerosol particles can be obtained from the aerosol size distribution, the aerosol number concentration, and the aerosol refractive index. Often, the size distribution consists of different modes, where each mode contains particles of the same refractive index. Thus, in the most general case the elements of the state vector $\mathbf{x}$ for aerosol retrieval in Eq. (5.1) are for each homogeneous layer and mode of the size distribution, the real part of the refractive index $m_r$, the imaginary part of the refractive index $m_i$, the elements $n_i$ of the discretized aerosol size distribution, and the aerosol number concentration.

The elements of the forward model vector $\mathbf{F}$ are in the most general case modeled values of light intensity and polarization at the top of the model atmosphere (at different wavelengths and in different viewing directions of the satellite). We will use the symbol $E_i$ to refer to the modeled value of the $i$th Stokes parameter at the top of the atmosphere, at a given wavelength and viewing direction of the satellite measurement. $E_i$ is called a radiative effect. The radiative effect $E_i$

can be extracted from the intensity vector field $\mathbf{I}$ (i.e. the solution of the radiative transfer equation (5.6)) with a suitable response vector function $\mathbf{R}_i$, via the inner product (see, for example, Marchuk (1964)):

$$E_i = \langle \mathbf{R}_i, \mathbf{I} \rangle. \tag{5.52}$$

Here, the inner product of two arbitrary vector functions $\mathbf{a}$ and $\mathbf{b}$ is defined by

$$\langle \mathbf{a}, \mathbf{b} \rangle = \int \mathrm{d}z \int \mathrm{d}\Omega \, \mathbf{a}^T(z, \mathbf{\Omega}) \mathbf{b}(z, \mathbf{\Omega}), \tag{5.53}$$

with integration over full solid angle and altitude range of the model atmosphere. The response functions $\mathbf{R}_i$ are given by

$$\mathbf{R}_i(z, \mathbf{\Omega}) = \delta(z - z_{\mathrm{top}}) \delta(\mathbf{\Omega} - \mathbf{\Omega}_v) \mathbf{e}_i, \tag{5.54}$$

where $\mathbf{e}_i$ is the unity vector in the direction of the $i$th component of the intensity vector, and $\mathbf{\Omega}_v = (\mu_v, \varphi_v)$ denotes the viewing direction of the instrument. In this context the response function formalism may seem somewhat awkward, but it is essential for a proper presentation of the adjoint formulation of radiative transfer, which will be described in section 5.4.1.

The requested derivatives of the elements of the forward model vector $\mathbf{F}$ in Eq. (5.2) can be expressed by the corresponding derivatives of the radiative effects $E_i$. Thus, the derivatives that we need to calculate are the derivatives $\partial E_i / \partial x_k$. These derivatives can be written as

$$\frac{\partial E_i}{\partial x_k} = \sum_{j=1}^{6} \sum_{l=0}^{L} \frac{\partial E_i}{\partial \alpha_j^l} \frac{\partial \alpha_j^l}{\partial x_k} + \frac{\partial E_i}{\partial K_{\mathrm{ext}}} \frac{\partial K_{\mathrm{ext}}}{\partial x_k} + \frac{\partial E_i}{\partial K_{\mathrm{sca}}} \frac{\partial K_{\mathrm{sca}}}{\partial x_k}. \tag{5.55}$$

Thus, the linearization corresponds to the calculation of two types of derivatives: (i) the derivatives $\partial E_i / \partial \alpha_i^j$, $\partial E_i / \partial K_{\mathrm{ext}}$, $\partial E_i / \partial K_{\mathrm{sca}}$ and (ii) the derivatives $\partial \alpha_j^l / \partial x_k$ $\partial K_{\mathrm{ext}} / \partial x_k$, and $\partial K_{\mathrm{sca}} / \partial x_k$. In the following we present an analytical approach to calculate these derivatives.

### 5.4.1 Linearization of radiative transfer

#### 5.4.1.1 Forward-adjoint perturbation theory

For the linearization of radiative transfer with respect to the optical input parameters of the radiative transfer equation we will employ the forward-adjoint perturbation theory. Here, the adjoint formulation of radiative transfer is of essential importance. The transport operator adjoint to $\hat{\mathbf{L}}$, which is called $\hat{\mathbf{L}}^\dagger$, is defined by requiring that (see, for example, Marchuk (1964); Box et al. (1988))

$$\langle \mathbf{I}_2, \hat{\mathbf{L}} \mathbf{I}_1 \rangle = \langle \hat{\mathbf{L}}^\dagger \mathbf{I}_2, \mathbf{I}_1 \rangle \tag{5.56}$$

for arbitrary vector functions $\mathbf{I}_1$ and $\mathbf{I}_2$. The adjoint vector field $\mathbf{I}^\dagger$ is the solution of the adjoint transport equation

$$\hat{\mathbf{L}}^{\dagger}\mathbf{I}^{\dagger} = \mathbf{S}^{\dagger} \tag{5.57}$$

with any suitable adjoint source $\mathbf{S}^{\dagger}$. The adjoint operator $\hat{\mathbf{L}}^{\dagger}$ is given by (Marchuk, 1964; Carter et al., 1978)

$$\begin{aligned}\hat{\mathbf{L}}^{\dagger} \;=\; \int_{4\pi} \mathrm{d}\tilde{\Omega} \Bigg\{ \Bigg[ -\mu\frac{\partial}{\partial z} + K_{\mathrm{ext}}(z) \Bigg] \delta(\mathbf{\Omega}-\tilde{\mathbf{\Omega}})\,\mathbf{E} \\ -\frac{K_{\mathrm{sca}}(z)}{4\pi}\mathbf{Z}^{T}(z,\mathbf{\Omega},\tilde{\mathbf{\Omega}}) - \delta(z)\Theta(-\mu)|\mu|\mathbf{R}_s^{T}(\mathbf{\Omega},\tilde{\mathbf{\Omega}})\;\Theta(\tilde{\mu})|\tilde{\mu}| \Bigg\}. \end{aligned} \tag{5.58}$$

The inclusion of the surface reflection term (last term on the right-hand side) is discussed by Ustinov (2001) and Landgraf et al. (2002). We see that compared to the forward operator $\hat{\mathbf{L}}$ the adjoint operator $\hat{\mathbf{L}}^{\dagger}$ has a different sign in the first term, and the phase matrix $\mathbf{Z}(z,\tilde{\mathbf{\Omega}},\mathbf{\Omega})$ and the surface reflection matrix $\mathbf{R}_s(\tilde{\mathbf{\Omega}},\mathbf{\Omega})$ are replaced by $\mathbf{Z}^{T}(z,\mathbf{\Omega},\tilde{\mathbf{\Omega}})$, and $\mathbf{R}_s^{T}(\mathbf{\Omega},\tilde{\mathbf{\Omega}})$, respectively. The adjoint vector field $\mathbf{I}^{\dagger}$ has to fulfill the boundary conditions (Box et al., 1988)

$$\begin{aligned}\mathbf{I}^{\dagger}(z_{\mathrm{top}},\mathbf{\Omega}) &= [0,0,0,0]^{T} \qquad \text{for } \mu > 0, \\ \mathbf{I}^{\dagger}(0,\mathbf{\Omega}) &= [0,0,0,0]^{T} \qquad \text{for } \mu < 0. \end{aligned} \tag{5.59}$$

The forward radiative transfer equation (5.6) and the adjoint transport equation (5.57) do not describe two independent problems. The solutions $\mathbf{I}$ and $\mathbf{I}^{\dagger}$ are linked by the relation

$$\langle \mathbf{S}^{\dagger}, \mathbf{I} \rangle = \langle \mathbf{I}^{\dagger}, \mathbf{S} \rangle, \tag{5.60}$$

which can be derived in a straightforward fashion using Eqs (5.6), (5.56), and (5.57). We now take the response vector function $\mathbf{R}_i$ of Eq. (5.54) as the adjoint source $\mathbf{S}^{\dagger}$. In this particular case, the left-hand side of Eq. (5.60) represents the definition of the radiative effect $E_i$ (see Eq. (5.52)). Thus we see from Eq. (5.60) that there are two ways of computing the radiative effect $E_i$. The first is the forward approach: solve the radiative transfer equation (5.6) and take the inner product of the response function $\mathbf{R}_i$ with the solution $\mathbf{I}$. The second is the adjoint approach: solve the adjoint transport equation (5.57) for the adjoint source $\mathbf{S}^{\dagger} = \mathbf{R}_i$ and take the inner product of its solution $\mathbf{I}^{\dagger}$ with the radiation source $\mathbf{S}$. Now also the physical meaning of the adjoint field becomes clear. Namely, the value of the adjoint field at a given altitude $z_s$ and in a certain direction $\mathbf{\Omega}_s$ gives the effect of a point source $\delta(z-z_s, \mathbf{\Omega}-\mathbf{\Omega}_s)$ on the radiative effect $E_i$. In other words, the adjoint field gives us the importance of a radiation source anywhere in the atmosphere for the radiative effect $E_i$ (Lewins, 1965). Thus, if the adjoint vector field is known the radiative effect $E_i$ can be calculated for any radiation source via Eq. (5.60).

Let us consider an atmosphere with a set of optical parameters ($\alpha_j^l$, $K_{\mathrm{ext}}$ and $K_{\mathrm{sca}}$) contained in the vector $\mathbf{g}_o$. We call this atmosphere the unperturbed atmosphere. We denote the corresponding vector intensity field by $\mathbf{I}_o$, and the adjoint field corresponding to the adjoint source $\mathbf{R}_i$ by $\mathbf{I}_o^{\dagger}(\mathbf{R}_i)$. We also consider a perturbed atmosphere with a vector of optical parameters $\mathbf{g} = \mathbf{g}_o + \Delta\mathbf{g}$, where

the optical parameters are perturbed in one layer of the model atmosphere. The radiative effect $E_i$ for the perturbed atmosphere is given by (Marchuk, 1964)

$$E_i(\mathbf{g}) = E_i(\mathbf{g}_o) - \langle \mathbf{I}_o^\dagger(\mathbf{R}_i), \Delta\hat{\mathbf{L}}\, \mathbf{I}_o \rangle + \mathcal{O}(\Delta\mathbf{g}^2), \tag{5.61}$$

where $\mathcal{O}(\Delta\mathbf{g}^2)$ denotes second and higher order terms. The change $\Delta\hat{\mathbf{L}}$ in the radiative transfer operator $\hat{\mathbf{L}}$ caused by the perturbation $\Delta\mathbf{g}$ can be written as:

$$\Delta\hat{\mathbf{L}} = \sum_{k=1}^{K} \Delta g_k\, \Delta\hat{\mathbf{L}}_k, \tag{5.62}$$

where $\Delta\hat{\mathbf{L}}_k$ is the the change in $\Delta\hat{\mathbf{L}}$ per unit in parameter $g_k$, and $K$ is the total number of optical parameters. The explicit form of $\Delta\hat{\mathbf{L}}_k$ follows from the definition of the transport operator $\hat{\mathbf{L}}$ (5.7). Substitution of Eq. (5.62) in Eq. (5.61) and comparison with a Taylor expansion yields the requested derivatives of the radiative effect $E_i$ with respect to the optical parameters $g_k$:

$$\frac{\partial E_i}{\partial g_k} = -\langle \mathbf{I}_o^\dagger(\mathbf{R}_i), \Delta\hat{\mathbf{L}}_k\, \mathbf{I}_o \rangle. \tag{5.63}$$

So, in order to calculate the requested derivative the intensity vector field $\mathbf{I}_o$ is required as well as the adjoint fields $\mathbf{I}_o^\dagger$ for the adjoint sources $\mathbf{R}_i$ with $i = 1, \ldots, 4$.

#### 5.4.1.2 Transformation to pseudo-forward problem

The adjoint field can be calculated with the same radiative transfer model as the forward intensity field, because the adjoint transport equation (5.57) may be transformed to a pseudo-forward problem. For this purpose we consider the vector function

$$\mathbf{\Psi}(z, \mathbf{\Omega}) = \mathbf{I}^\dagger(z, -\mathbf{\Omega}). \tag{5.64}$$

With substitution of Eq. (5.64) in Eq. (5.57), and with the symmetry relation of the scattering phase matrix (Hovenier, 1969)

$$\mathbf{Z}^T(z, -\mathbf{\Omega}, -\tilde{\mathbf{\Omega}}) = \mathbf{Q}\mathbf{Z}(z, \tilde{\mathbf{\Omega}}, \mathbf{\Omega})\mathbf{Q}, \tag{5.65}$$

with

$$\mathbf{Q} = \mathrm{diag}[1, 1, 1, -1], \tag{5.66}$$

and a similar relation for the surface reflection matrix $\mathbf{R}_s$, the adjoint transport equation transforms to a pseudo-forward equation

$$\hat{\mathbf{L}}_{\mathbf{\Psi}}\mathbf{\Psi} = \mathbf{S}_\Psi, \tag{5.67}$$

where

$$\mathbf{S}_\Psi(z, \mathbf{\Omega}) = \mathbf{R}_i(z, -\mathbf{\Omega}). \tag{5.68}$$

Here, the transport operator $\hat{\mathbf{L}}_{\mathbf{\Psi}}$ is the same as $\hat{\mathbf{L}}$ defined in Eq. (5.7), except that $\mathbf{Z}(z, \tilde{\mathbf{\Omega}}, \mathbf{\Omega})$ is replaced by $\mathbf{Q}\mathbf{Z}(z, \tilde{\mathbf{\Omega}}, \mathbf{\Omega})\mathbf{Q}$ and $\mathbf{R}_s(z, \tilde{\mathbf{\Omega}}, \mathbf{\Omega})$ is replaced by

$\mathbf{QR}_s(z,\tilde{\mathbf{\Omega}},\mathbf{\Omega})\mathbf{Q}$. According to Eqs (5.64) and (5.59), $\mathbf{\Psi}$ has to fulfill the same boundary conditions as $\mathbf{I}$ in Eq. (5.10).

For the pseudo-forward problem a Fourier expansion can be performed as described in section 5.2. However, here the Fourier coefficients $\mathbf{S}_{\mathbf{\Psi}}^{\pm m}$ of the pseudo-forward source $\mathbf{S}_\Psi(z,\mathbf{\Omega}) = \mathbf{R}_i(z,-\mathbf{\Omega})$ depend on the index $i$, indicating the radiative effect $E_i$ under consideration. For $i = 1,2$, i.e. for the calculation of the derivatives of $I$ and $Q$, we obtain

$$\begin{aligned}\mathbf{S}_{\mathbf{\Psi}}^{+m}(z,\mu) &= \frac{1}{2\pi}\,\delta(z-z_{top})\,\delta(\mu+\mu_v)\,\mathbf{e}_i,\\ \mathbf{S}_{\mathbf{\Psi}}^{-m}(z,\mu) &= [0,0,0,0]^T.\end{aligned} \tag{5.69}$$

Hence, for the corresponding pseudo-forward problems we obtain a Fourier expansion of $\mathbf{\Psi}$ containing terms of $\mathbf{\Psi}^{+m}$ only. For $i = 3,4$, i.e. for the calculation of the derivatives of $U$ and $V$, we obtain

$$\begin{aligned}\mathbf{S}_{\mathbf{\Psi}}^{+m}(z,\mu) &= [0,0,0,0]^T,\\ \mathbf{S}_{\mathbf{\Psi}}^{-m}(z,\mu) &= \frac{1}{2\pi}\,\delta(z-z_{top})\,\delta(\mu+\mu_v)\,\mathbf{e}_i.\end{aligned} \tag{5.70}$$

Hence, for the corresponding pseudo-forward problems we obtain a Fourier expansion of $\mathbf{\Psi}$ containing terms of $\mathbf{\Psi}^{-m}$ only.

#### 5.4.1.3 Calculation of the derivatives

In the following we will work out the expressions for the derivatives with respect to the expansion coefficients $\alpha_j^l$, the scattering coefficient $K_{\text{sca}}$, and the extinction coefficient $K_{\text{ext}}$. Hereto, we write instead of Eq. (5.62)

$$\Delta\hat{\mathbf{L}} = \sum_{j=1}^{6}\sum_{l=0}^{L}\Delta\alpha_j^l\;\Delta\hat{\mathbf{L}}_j^l \;+\; \Delta K_{\text{sca}}\Delta\hat{\mathbf{L}}_{\text{sca}} \;+\; \Delta K_{\text{ext}}\Delta\hat{\mathbf{L}}_{\text{ext}}, \tag{5.71}$$

where

$$\Delta\hat{\mathbf{L}}_j^l = \frac{\beta^s}{4\pi}\int_{4\pi}\mathrm{d}\tilde{\Omega}\frac{\partial\mathbf{Z}(z,\tilde{\mathbf{\Omega}},\mathbf{\Omega})}{\partial\alpha_j^l}, \tag{5.72}$$

$$\Delta\hat{\mathbf{L}}_{\text{sca}} = \frac{1}{4\pi}\int_{4\pi}\mathrm{d}\tilde{\Omega}\mathbf{Z}(z,\tilde{\mathbf{\Omega}},\mathbf{\Omega}), \tag{5.73}$$

$$\Delta\hat{\mathbf{L}}_{\text{ext}} = \int_{4\pi}\mathrm{d}\tilde{\Omega}\;\delta(\mathbf{\Omega}-\tilde{\mathbf{\Omega}})\;\mathbf{E}. \tag{5.74}$$

In order to obtain expressions for the derivatives with respect to $\alpha_j^l$, $K_{\text{sca}}$, and $K_{\text{ext}}$, we substitute $\Delta\hat{\mathbf{L}}_j^l$, $\Delta\hat{\mathbf{L}}_{\text{sca}}$, and $\Delta\hat{\mathbf{L}}_{\text{ext}}$ in Eq. (5.63), respectively. Additionally, we use the Fourier expansion of $\mathbf{I}$, $\mathbf{\Psi}$, and $\mathbf{Z}$, and evaluate the integrals over azimuth angle. We then obtain expressions in the form of cosine- and sine expansions which have a similar form for the three types of derivatives. For the

radiative effects $E_i$ with $i = 1, 2$ (i.e. corresponding to Stokes parameters $I$ and $Q$, respectively), the derivatives are given by a cosine expansion:

$$\frac{\partial E_i}{\partial g_k} = -\sum_{m=0}^{\infty} (2 - \delta_{m0}) \cos m(\phi_v - \phi_0) \, K_i^{+m}(g_k). \tag{5.75}$$

For the radiative effects $E_i$ with $i = 3, 4$ (i.e. corresponding to Stokes parameters $U$ and $V$, respectively) the derivatives are given by a sines expansion:

$$\frac{\partial E_i}{\partial g_k} = -\sum_{m=0}^{\infty} (2 - \delta_{m0}) \sin m(\phi_v - \phi_0) \, K_i^{-m}(g_k). \tag{5.76}$$

The specific integral kernels for $\alpha_j^l$, $K_{\text{sca}}$, and $K_{\text{ext}}$ are determined by $\Delta\hat{\mathbf{L}}_j^l$, $\Delta\hat{\mathbf{L}}_{\text{sca}}$, and $\Delta\hat{\mathbf{L}}_{\text{ext}}$, respectively. For the derivative of the radiative effect $E_i$ with respect to the expansion coefficient $\alpha_j^l$, we obtain for the integral kernel

$$\begin{aligned} K_i^{\pm m}(\alpha_j^l) \;=\; & \frac{\pi}{4} \int_{z_{\text{bot}}}^{z_{\text{top}}} \mathrm{d}z \, K_{\text{sca}}(z) \\ & \int_{-1}^{1} \int_{-1}^{1} \mathrm{d}\mu \, \mathrm{d}\tilde{\mu} \, \mathbf{\Psi}_o^{\pm mT}(\mathbf{R}_i, z, -\mu) \, \boldsymbol{\Lambda} \, \dot{\mathbf{Z}}^m(z, \tilde{\mu}, \mu) \, \mathbf{I}_o^{+m}(z, \tilde{\mu}), \end{aligned} \tag{5.77}$$

where $\boldsymbol{\Lambda} = \text{diag}\,[1, 1, -1, -1]$, and the derivative $\dot{\mathbf{Z}}^m = \partial \mathbf{Z}^{\mathbf{m}} / \partial \alpha_j^l$ is given by

$$\dot{\mathbf{Z}}^m(z, \tilde{\mu}, \mu) = (-1)^m \, \mathbf{P}_m^l(-\mu) \, \mathbf{H}_j \, \mathbf{P}_m^l(-\tilde{\mu}). \tag{5.78}$$

Here, the matrix $\mathbf{H}_j$ has the same structure as the expansion coefficient matrix (5.23) and is given by

$$\mathbf{H}_j = \begin{pmatrix} \delta_{j1} & \delta_{j5} & 0 & 0 \\ \delta_{j5} & \delta_{j2} & 0 & 0 \\ 0 & 0 & \delta_{j3} & \delta_{j6} \\ 0 & 0 & -\delta_{j6} & \delta_{j4} \end{pmatrix}, \tag{5.79}$$

where $\delta$ is the Kronecker delta.

The integral kernel corresponding to the derivative of the radiative effect $E_i$ with respect to $K_{\text{sca}}$ is given by:

$$K_i^{\pm m}(K_{\text{sca}}) = \frac{\pi}{4} \int_{z_{\text{bot}}}^{z_{\text{top}}} \mathrm{d}z \int_{-1}^{1} \int_{-1}^{1} \mathrm{d}\mu \, \mathrm{d}\tilde{\mu} \, \mathbf{\Psi}_o^{\pm mT}(\mathbf{R}_i, z, -\mu) \, \boldsymbol{\Lambda} \, \mathbf{Z}^m(z, \tilde{\mu}, \mu) \, \mathbf{I}_o^{+m}(z, \tilde{\mu}), \tag{5.80}$$

and the integral kernel corresponding to the derivative of the radiative effect $E_i$ with respect to $K_{\text{ext}}$ has the following form:

$$K_i^{\pm m}(K_{\text{ext}}) = 2\pi \int_{z_{\text{bot}}}^{z_{\text{top}}} \mathrm{d}z \int_{-1}^{1} \mathrm{d}\mu \, \mathbf{\Psi}_o^{\pm mT}(\mathbf{R}_i, z, -\mu) \, \boldsymbol{\Lambda} \, \mathbf{I}_o^{+m}(z, \mu). \tag{5.81}$$

Equations (5.75), (5.76), together with (5.77)–(5.81) provide analytical expressions for the derivatives of the radiative effect $E_i$ with respect to the expansion coefficients $\alpha_j^l$, the scattering coefficient $K_{\rm sca}$, and the extinction coefficient $K_{\rm ext}$, respectively. Thus, to calculate these derivatives one needs to solve the forward radiative transfer equation (5.6) and the adjoint transport equation (5.57) for the sources $\mathbf{S}_{\mathbf{\Psi}}(z,\mathbf{\Omega}) = \mathbf{R}_i(z,-\mathbf{\Omega})$, with $i = 1, 4$. These fields can be determined by any vector radiative transfer model that calculates the internal radiation in the atmosphere, such as the doubling and adding model of Stammes et al. (1989), the discrete ordinate model VDISORT of Schulz et al. (1999) and the Gauss–Seidel model of Hasekamp and Landgraf (2002). The latter model is used for all numerical simulations in this chapter. The corresponding integral kernels of equations (5.77)–(5.81) are worked out by Landgraf et al. (2004) for this model.

### 5.4.2 Linearization of Mie theory

The derivatives of the optical input parameters of the radiative transfer equation with respect to the different elements $x_k$ of the state vector can be found from the corresponding derivatives of the optical aerosol parameters (see Eqs (5.31)–(5.33)):

$$\frac{\partial K_{\rm ext}}{\partial x_k} = \frac{\partial K_{\rm ext}^a}{\partial x_k}, \tag{5.82}$$

$$\frac{\partial K_{\rm sca}}{\partial x_k} = \frac{\partial K_{\rm sca}^a}{\partial x_k}, \tag{5.83}$$

$$\frac{\partial \alpha}{\partial x_k} = \frac{K_{\rm sca}^a}{K_{\rm sca}}\frac{\partial \alpha^a}{\partial x_k} + \frac{\alpha^a}{K_{\rm sca}}\frac{\partial K_{\rm sca}^a}{\partial x_k} - \frac{K_{\rm sca}^a\,\alpha^a}{(K_{\rm sca})^2}\frac{\partial K_{\rm sca}^a}{\partial x_k} - \frac{K_{\rm sca}^r\,\alpha^r}{(K_{\rm sca})^2}\frac{\partial K_{\rm sca}^a}{\partial x_k}, \tag{5.84}$$

where we omitted the indices for the expansion coefficients $\alpha_j^l$. In this subsection we will derive expressions for the requested derivatives $\partial\alpha^a/\partial x_k$, $\partial K_{\rm ext}^a/\partial x_k$, and $\partial K_{\rm sca}^a/\partial x_k$. For notational convenience, we will omit the superscript $a$ for the expansion coefficients in the remainder of this subsection.

In order to calculate the derivatives of the expansion coefficients $\alpha_j^l$ with respect to the real and imaginary part of the refractive index, $m_r$ and $m_i$, respectively, we first need to calculate the corresponding derivatives of the elements $f_j$ of the transformation matrix $\boldsymbol{\mathcal{F}}$ in Eq. (5.34). These derivatives are expressed via the derivatives of $S_1$ and $S_2$ (see Eqs (5.35)–(5.40)):

$$[f_1]' = \frac{1}{2}\big(S_1[S_1^*]' + [S_1]'S_1^* + S_2[S_2^*]' + [S_2]'S_2^*\big), \tag{5.85}$$

$$[f_2]' = [f_1]', \tag{5.86}$$

$$[f_3]' = \frac{1}{2}\big(S_1[S_2^*]' + [S_1]'S_2^* + S_2[S_1^*]' + [S_2]'S_1^*\big), \tag{5.87}$$

$$[f_4]' = [f_3]', \tag{5.88}$$

$$[f_5]' = \frac{1}{2}\big(S_1[S_1^*]' + [S_1]'S_1^* - S_2[S_2^*]' + [S_2]'S_2^*\big), \tag{5.89}$$

$$[f_6]' = \frac{i}{2}\big(S_1[S_2^*]' + [S_1]'S_2^* - S_2[S_1^*]' + [S_2]'S_1^*\big). \tag{5.90}$$

Here, the prime denotes the derivative with respect to either $m_r$ or $m_i$. The derivatives of $S_1$ and $S_2$ are given by

$$[S_1]' = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} \left([a_n]'\pi_n + [b_n]'\tau_n\right), \tag{5.91}$$

$$[S_2]' = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} \left([b_n]'\pi_n + [a_n]'\tau_n\right). \tag{5.92}$$

The derivatives of the scattering and extinction coefficients with respect to $m_r$ and $m_i$ follow from the corresponding derivatives of the scattering and extinction cross-sections. These derivatives also depend on the derivatives of $a_n$ and $b_n$:

$$[C_{\text{sca}}]' = \frac{4\pi}{k^2}\sum_{n=1}^{\infty}(2n+1)\big(a_n[a_n^*]' + [a_n]'a_n^* + b_n[b_n^*]' + [b_n]'b_n^*\big), \tag{5.93}$$

$$[C_{\text{ext}}]' = \frac{2\pi}{k^2}\sum_{n=1}^{\infty}(2n+1)\text{Re}([a_n]' + [b_n]'). \tag{5.94}$$

Thus, all derivatives with respect to $m_r$ and $m_i$ depend on the corresponding derivatives of the Mie coefficients $a_n$ and $b_n$, for which analytical expressions are given in Appendix A.

The derivatives of $f_j$, $C_{\text{sca}}$, and $C_{\text{ext}}$ with respect to $m_r$ and $m_i$ given above all correspond to a single sphere with a given radius. The derivatives for an ensemble of particles with a given size distribution can be easily obtained via integration over the size distribution as in Eq. (5.50). The derivatives of elements $p_j$ of the scattering phase matrix (5.5) for the size distribution can then be calculated using the derivatives of $\bar{f}_j$ and $\bar{C}_{\text{sca}}$, viz.

$$[p_j]' = \frac{4\pi}{k^2}\left(\frac{[\bar{f}_j]'}{\bar{C}_{\text{sca}}} - \frac{\bar{f}_j[C_{\text{sca}}]'}{\bar{C}_{\text{sca}}^2}\right). \tag{5.95}$$

The derivatives of the expansion coefficients $\alpha_j^l$ can be calculated from Eqs (5.24)–(5.29), replacing $\alpha_i^j$ and $p_j$ by their corresponding derivatives.

The other type of derivatives that are needed are the derivatives with respect to the elements $n_i$ of the discretized size distribution (see Eq. (5.50)). These derivatives can be calculated in a straightforward manner. The derivative of an averaged element $\bar{f}_j$ of the transformation matrix $\boldsymbol{\mathcal{F}}$ with respect to $n_i$ is given by:

$$\frac{\partial \bar{f}_j}{\partial n_i} = f_j(r_i)\,\Delta r_i, \tag{5.96}$$

and similar expressions hold for the derivatives of $\bar{C}_{\rm sca}$ and $\bar{C}_{\rm ext}$ with respect to $n_i$. Using the derivatives of $\bar{f}_j$ and $\bar{C}_{\rm sca}$ with respect to $n_i$, the derivatives of the elements $p_j$ of the scattering phase matrix (5.5) with respect to $n_i$ be calculated using Eq. (5.95). The corresponding derivatives of the aerosol expansion coefficients $\alpha_i^j$ can be calculated from Eqs (5.24)–(5.29), replacing $\alpha_i^j$ and $p_j$ by their corresponding derivatives.

For aerosol retrieval it is often useful to describe the size distribution by a limited number of parameters, for example the effective radius $r_{\rm eff}$ and the effective variance $v_{\rm eff}$ (see Appendix B) of a prescribed size distribution (e.g. a lognormal distribution). For such retrievals one needs to calculate the derivatives with respect to these parameters. For an averaged parameter $\bar{g}$ this derivative is given by

$$\frac{\partial \bar{g}}{\partial r_{\rm eff}} = \sum_{i=1}^{N} \frac{\partial \bar{n}_i}{\partial r_{\rm eff}} \frac{\partial \bar{g}}{\partial \bar{n}_i}, \tag{5.97}$$

and a similar expression holds for the derivative with respect to $v_{\rm eff}$.

## 5.5 Numerical implementation and results

The linearization approach described in section 5.4 has been implemented in the Gauss-Seidel vector radiative transfer model described by Landgraf et al. (2002) and Hasekamp and Landgraf (2002), combined with the Mie scattering algorithm of de Rooij and van der Stap (1984). Expressions for the integral kernels of section 5.4.1 can be found in the paper of Landgraf et al. (2004) for our Gauss–Seidel radiative transfer model.

All radiative transfer calculations in this section were performed for a model atmosphere that includes Rayleigh scattering and scattering and absorption by homogeneous spherical aerosol particles. All aerosols were homogeneously distributed over the lowest 2 km of the atmosphere. We used a bi-modal lognormal aerosol size distribution, with a mode containing small particles referred to as the small mode and a mode containing large particles referred to as the large mode. For this model atmosphere the aerosols are characterized by ten parameters, i.e. five per mode of the size distribution. These parameters are: per mode the effective radius $r_{\rm eff}$, the effective variance $v_{\rm eff}$, the column integrated aerosol number concentration $N$, and the real and imaginary part of the refractive index $m$.

Figures 5.1 and 5.2 show the derivatives of Stokes parameters $I$ and $Q$ at the top of the atmosphere with respect to the logarithm of the effective radius, real refractive index, and aerosol loading of the two size modes, as a function of viewing zenith angle for a solar zenith angle of $40^o$, for a wavelength of 350 nm and 800 nm, respectively. The relative azimuth angle $\phi_o - \phi_v = 180^o$ for negative viewing zenith angles and $\phi_o - \phi_v = 0^o$ for positive viewing zenith angles. For these geometries the Stokes parameters $U$ and $V$ are equal to zero, so Stokes parameter $Q$ fully describes the polarization of the backscattered light at the top of the atmosphere. The derivative with respect to the logarithm of a given

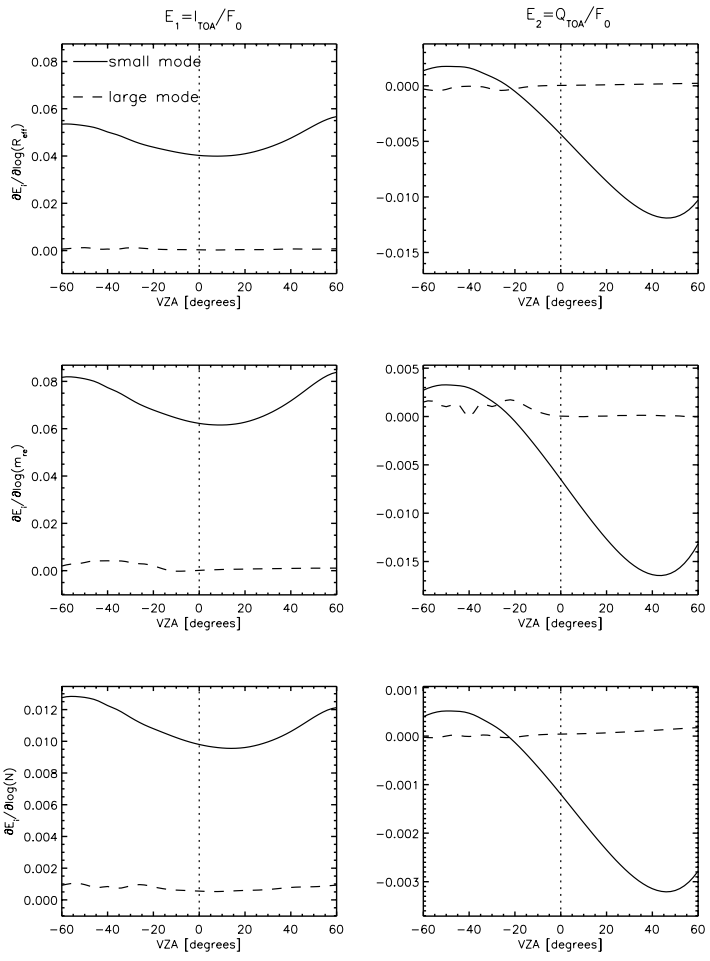



**Fig. 5.1.** Derivatives of the Stokes parameters $I$ (left panels) and $Q$ (right panels) at the top of the atmosphere (TOA), with respect to the logarithm of: (upper panels) $r_{\text{eff}}$, (middle panels) $m_r$, and (lower panels) the column integrated aerosol number concentration N. The solid lines correspond to parameters of the small mode and the dashed lines correspond to parameters of the large mode. The derivatives are shown as a function of viewing zenith angle (VZA), where $VZA < 0$ refer to the relative azimuth angle $\phi_o - \phi_v = 180^o$, and $VZA > 0$ refer to $\phi_o - \phi_v = 0^o$. The solar zenith angle is $40^o$, and the calculation is performed for a wavelength of 350 nm. The solar and viewing zenith angles are defined as the smallest angle between the zenith direction and the solar and viewing direction, respectively. The range $-60^o$ to $60^o$ of viewing zenith angles used in this figure corresponds to a horizon-to-horizon scan from a satellite at approximately 800 km. The internal radiation field was discretized in 16 Gaussian streams. A bimodal aerosol size distribution was used with $r_{\text{eff}} = 0.05$ for the small mode, $r_{\text{eff}} = 0.75$ for the large mode, $v_{\text{eff}} = 0.2$ for both modes, $m_r = 1.45$, $m_i = -0.0045$. The optical thickness at 550 nm is 0.15 with equal contribution from the small and the large mode. The model atmosphere is bounded below by a black surface.

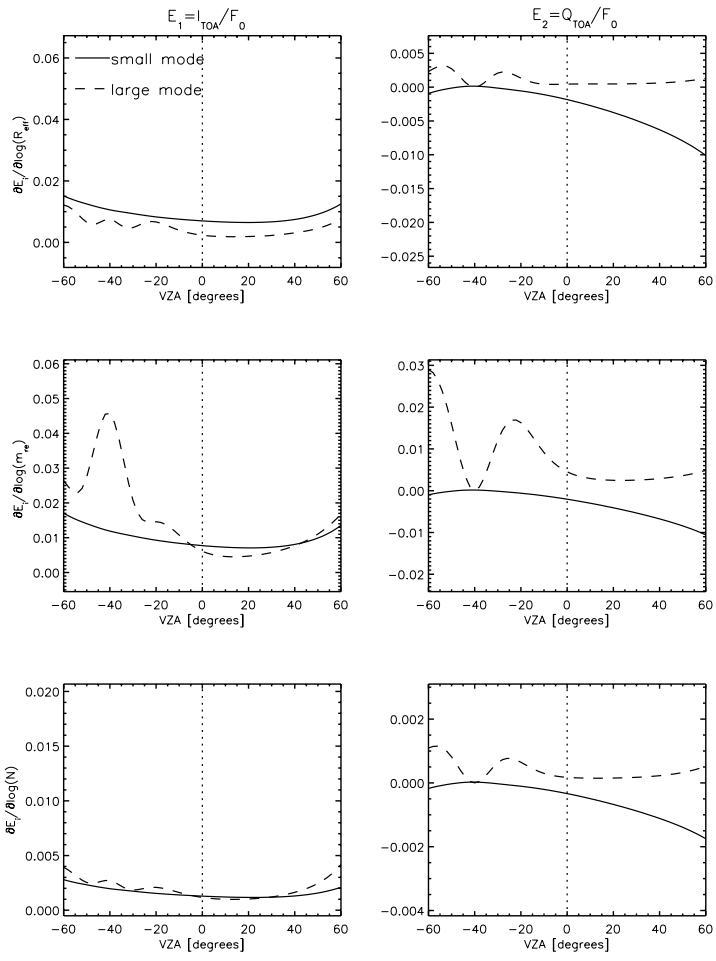



**Fig. 5.2.** Same as Fig. 5.1 but for a wavelength of 800 nm.

aerosol parameter is a measure for the sensitivity of $I$ and $Q$ to a relative change in this aerosol parameter, which is a convenient quantity in order to compare the sensitivities to the different parameters.

The angular dependence of the derivatives is caused by the angular dependence of the following effects: (i) The derivatives of the relevant elements of the aerosol scattering phase matrix with respect to the different aerosol parameters. (ii) The light path inside the aerosol layer, which increases with viewing zenith angle. This causes an increase in sensitivity up to a certain viewing angle because an increasing fraction of the light is scattered by aerosol particles. However, if the viewing angle becomes too large this effect causes a decrease in sensitivity because of increasing extinction within the aerosol layer along the line of sight. (iii) Multiple scattering effects, which in general smear out the angular effects of the aerosol scattering phase matrix.

From Fig. 5.1 it follows that at 350 nm the sensitivities of both $I$ and $Q$ with respect to the parameters of the small mode are much larger than the corresponding sensitivities with respect to the parameters of the large mode, which are in general negligibly small at 350 nm. The angular dependence of the derivatives of the intensity in Fig. 5.1 is relatively weak which is caused by the weak angular dependence of the corresponding derivatives of element (1,1) of the scattering phase matrix (except for the forward scattering direction, which is not shown in Fig. 5.1). The derivatives with respect to the different aerosol parameters of element (2,1) of the scattering phase matrix have a larger angular dependence, which is the most dominant effect in the right panels of Fig. 5.1. Also multiple scattering effects can be seen here, because both the Rayleigh and aerosol scattering optical thickness are relatively large at 350 nm. For example, in the backward scattering direction element (2,1) of the scattering phase matrix is zero independent of the aerosol properties, i.e. it is insensitive with respect to aerosol properties. However, the sensitivity of Stokes parameter $Q$ is not zero in the backward single scattering direction (viewing angle $= -40°$), because the sensitivity is influenced by aerosol scattering in all directions via multiple scattering.

At 800 nm (Fig. 5.2) the derivatives of both $I$ and $Q$ with respect to the parameters of the large mode, are significantly larger than at 350 nm, while the derivatives with respect to the parameters if the small mode are much smaller. The angular dependence of the derivatives of the phase matrix plays the most important role at 800 nm, especially for the derivatives of Stokes parameter $Q$. Here, the strong angular variation in the derivatives of $Q$ around the single scattering backward direction ($-40°$) is also present in the sensitivity of the (2,1)-element of the aerosol scattering phase matrix. A similar, but weaker effect can be seen in the derivatives of the intensity with respect to parameters of the large mode. Another effect that can be seen in Fig. 5.2 is the slight increase in sensitivity towards larger (absolute values of) viewing zenith angle. This increase in sensitivity is caused by an enhanced light path inside the aerosol layer.

## 5.6 Retrieval method

In this section we will discuss how the linearized vector radiative transfer model can be incorporated in a retrieval algorithm for aerosol properties over the ocean. Here, we assume that the aerosol size distribution can be described by a bi-modal lognormal function, where each mode is characterized by the effective radius $r_{\text{eff}}$, the effective variance $v_{\text{eff}}$ (see Appendix B) and the column integrated aerosol number concentration $N$. In what follows we use the superscripts $l$ and $s$ to refer to the small and large mode of the size distribution, respectively. Additionally, the complex refractive index $m = m_r + im_i$ is needed to characterize aerosols. Furthermore, we assume an altitude distribution with a constant aerosol density $\rho_o$ in the lowest layer with height $z_b$ of the atmosphere. Above that layer the

aerosol density decreases with the fourth power in pressure $p$ till a certain height $z_t$ above which we assume no aerosols are located, i.e.

$$\begin{aligned} \rho(z) &= \rho_o && \text{for } z < z_b, \\ \rho(z) &= \rho_o(p(z)/p(z_b))^4 && \text{for } z_b < z < z_t, \\ \rho(z) &= 0 && \text{for } z > z_t. \end{aligned} \tag{5.98}$$

In the retrievals described here nine unknown aerosol parameters are considered. These are the effective radius $r_{\text{eff}}$ of the small and large mode, the effective variance $v_{\text{eff}}$ of the small and large mode, the column integrated aerosol number concentration $N$ of the small and large mode, the real and imaginary part of the refractive index, and the height $z_b$ of the layer where the bulk of the aerosols is located. Here, we assume that the wavelength dependence of the refractive index is known. For all retrieval simulations in this paper the model atmosphere is bounded below by a rough ocean surface (see Appendix B). Here, the oceanic pigment concentration is included as an additional parameter to be retrieved in addition to the nine aerosol parameters.

### 5.6.1 Inversion of linearized forward model

In this subsection we consider the inversion of the linearized forward model (5.2), assuming that the state vector $\mathbf{x}_n$ of the iteration step under consideration is close enough to the true state vector so that the linear approximation is valid. In this case the inversion of Eq. (5.2) provides the solution of our retrieval problem. Rearranging terms in Eq. (5.2) we obtain

$$\tilde{\mathbf{y}} = \mathbf{K}\,\mathbf{x} + \mathbf{e}_y, \tag{5.99}$$

with $\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{F}(\mathbf{x}_\text{n}) + \mathbf{K}\mathbf{x}_\text{n}$. Here, $\mathbf{y}$ is the measurement vector, $\mathbf{F}(\mathbf{x}_\text{n})$ is the forward model vector for state vector $\mathbf{x}_\text{n}$, $\mathbf{K}$ is the Jacobian matrix defined by Eq. (5.3), and $\mathbf{e}_y$ is the measurement error vector.

For most types of satellite instruments the inversion of Eq. (5.99) represents an ill-posed problem. This means that many combinations of the state vector parameters fit the measurement almost equally well. As a result, the least-squares solution $\hat{\mathbf{x}}_\text{lsq}$ to our retrieval problem, viz.

$$\hat{\mathbf{x}}_\text{lsq} = \min_{\mathbf{x}} ||\mathbf{S}_y^{-\frac{1}{2}}(\mathbf{K}\mathbf{x} - \tilde{\mathbf{y}})||^2, \tag{5.100}$$

is overwhelmed by noise. In order to reduce the effect of noise, we use the Phillips–Tikhonov regularization method (Phillips, 1962; Tikhonov, 1963), which introduces a side constraint in addition to the minimization of the least-squares norm. As a side constraint we choose for our application the minimization of a weighted norm of the state vector, viz.

$$\hat{\mathbf{x}}_\text{reg} = \min_{\mathbf{x}} \left( ||\mathbf{S}_y^{-\frac{1}{2}}(\mathbf{K}\mathbf{x} - \tilde{\mathbf{y}})||^2 + \gamma\, ||\mathbf{\Gamma}\mathbf{x}||^2 \right), \tag{5.101}$$

where $\mathbf{\Gamma}$ is a diagonal matrix that contains weighting factors for the different state vector elements in the side constraint, and the regularization parameter $\gamma$ balances the two minimizations in Eq. (5.101). For each iteration step the solution $\hat{\mathbf{x}}_{\mathrm{reg,n+1}}$ in Eq. (5.101) can be written as a matrix equation:

$$\hat{\mathbf{x}}_{\mathrm{reg}} = \mathbf{D}\,\tilde{\mathbf{y}}, \tag{5.102}$$

where $\mathbf{D}$ is the contribution matrix defined by

$$\mathbf{D} = \left(\mathbf{K}^T\,\mathbf{S}_y^{-1}\,\mathbf{K} + \gamma\mathbf{\Gamma}\right)^{-1}\,\mathbf{K}^T\,\mathbf{S}_y^{-1}, \tag{5.103}$$

where the superscript $T$ denotes the transposed matrix.

The rationale of minimizing the norm of the state vector as a side constraint in Eq. (5.101) is to reduce the effect of measurement noise on the solution. Since the norm of the state vector is a quantity that is very sensitive to noise contributions, these contributions are reduced using Eq. (5.101) instead of Eq. (5.100). Clearly, a good choice of $\gamma$ is of crucial importance for the Phillips–Tikhonov solution. If $\gamma$ is chosen too large, the noise contribution will be low but the least squares norm deviates significantly from its minimum value, i.e. the fit between forward model and measurement is poor. On the other hand, if $\gamma$ is chosen too small the measurement is fitted well but the solution norm is high, i.e. the solution is overwhelmed by noise. Thus, $\gamma$ should be chosen such that the two minimizations are well balanced. Such a value for $\gamma$ can be found from the L-curve (Hansen and O'Leary, 1993). The L-curve is a parametric plot of the weighted least-squares norm $||\mathbf{S}_y^{-\frac{1}{2}}(\mathbf{K}(\mathbf{x}) - \tilde{\mathbf{y}})||$ and the weighted solution norm $||\mathbf{\Gamma x}||$, with a characteristic L-shaped corner. The corner of the L-curve corresponds to the optimum value of the regularization parameter. A numerical stable and efficient method for determining the corner of the L-curve is given by Hansen (1992), who defines the corner of the L-curve as the point with maximum curvature, where the curvature is calculated analytically. Visual inspection of the L-curves of our retrievals showed that in all cases the method of Hansen (1992) provided a value for the regularization parameter that corresponds to the 'true' corner of the L-curve. An example of an L-curve with the corresponding curvature is given in Fig. 5.3 for aerosol retrieval from synthetic GOME-2 measurements of intensity and polarization.

Due to the inclusion of the side constraint, the state vector $\hat{\mathbf{x}}_{\mathrm{reg}}$ retrieved using Eq. (5.101) does not represent an estimate of the true state vector $\mathbf{x}_{\mathrm{true}}$, but its elements represent weighted averages of the elements of $\mathbf{x}_{\mathrm{true}}$. The relation between $\hat{\mathbf{x}}_{\mathrm{reg}}$ and $\mathbf{x}_{\mathrm{true}}$ is expressed by the averaging kernel $\mathbf{A}$ (Rodgers, 2000), viz.

$$\hat{\mathbf{x}}_{\mathrm{reg}} = \mathbf{A}\mathbf{x}_{\mathrm{true}} + \mathbf{e}_x. \tag{5.104}$$

Here, $\mathbf{e}_x$ represents the error in the state vector caused by measurement errors, and the averaging kernel is given by

$$\mathbf{A} = \frac{\partial\hat{\mathbf{x}}}{\partial\mathbf{x}_{\mathrm{true}}} = \frac{\partial\hat{\mathbf{x}}}{\partial\mathbf{x}_{\mathrm{true}}} = \left(\mathbf{K}^T\,\mathbf{S}_y^{-1}\,\mathbf{K} + \gamma\mathbf{\Gamma}\right)^{-1}\,\mathbf{K}^T\,\mathbf{S}_y^{-1}\,\mathbf{K} = \mathbf{D}\,\mathbf{K}. \tag{5.105}$$

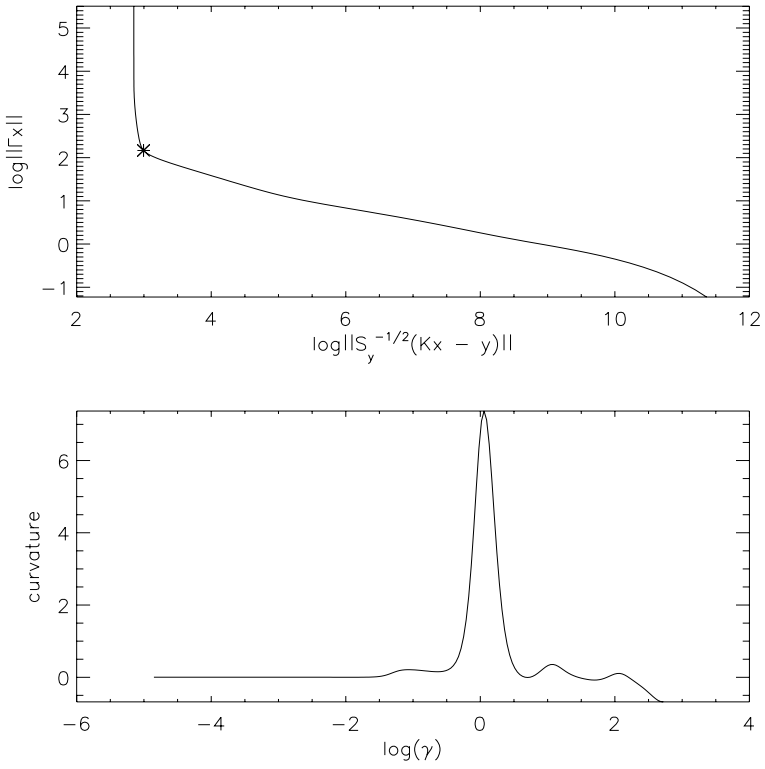



**Fig. 5.3.** L-curve (upper panel) and corresponding curvature (lower panel) for aerosol retrieval from synthetic GOME-2 measurements of intensity and polarization.

The matrix $\mathbf{A}$ is strongly related to the information content of the measurement $\mathbf{y}$, i.e. the closer $\mathbf{A}$ is to the unity matrix, the higher the information content. From the matrix $\mathbf{A}$ the Degrees of Freedom for Signal (DFS) can be derived (Rodgers, 2000), which indicates the number of independent pieces of information that is retrieved from the measurement:

$$\text{DFS} = \text{trace}\,(\mathbf{A}). \tag{5.106}$$

If $\mathbf{x}_{\text{true}}$ would have represented a discretization of a continuous function, then the weighted averages contained in $\hat{\mathbf{x}}_{\text{reg}}$ Eq. (5.104) would have a clear physical meaning, i.e. an estimate of $\mathbf{x}_{\text{true}}$ at a reduced resolution. However, for our application the elements of $\hat{\mathbf{x}}_{\text{reg}}$ represent weighted averages of different aerosol parameters, which have a limited physical meaning. Therefore, we include information from an *a priori* state vector $\mathbf{x}_a$ in the solution to make it a meaningful estimate of the $\mathbf{x}_{\text{true}}$. Hereto, we add the term $(\mathbf{I}-\mathbf{A})\mathbf{x}_a$ to $\hat{\mathbf{x}}_{\text{reg}}$ in order to obtain the final retrieval product $\hat{\mathbf{x}}$, viz.

$$\begin{aligned}\hat{\mathbf{x}} &= \hat{\mathbf{x}}_{\text{reg}} + (\mathbf{I}-\mathbf{A})\mathbf{x}_a,\\ &= \mathbf{A}\mathbf{x}_{\text{true}} + (\mathbf{I}-\mathbf{A})\mathbf{x}_a + \mathbf{e}_x.\end{aligned} \tag{5.107}$$

Thus, in Eq. (5.107) $\hat{\mathbf{x}}_{\text{reg}}$ represents an estimate of $\mathbf{A}\mathbf{x}_{\text{true}}$ and $(\mathbf{I}-\mathbf{A})\mathbf{x}_a$ represents an estimate of the part $(\mathbf{I}-\mathbf{A})\mathbf{x}_{\text{true}}$ of the true state vector that cannot be retrieved from the measurement. Here, the dependence of a retrieved element $\hat{x}_i$ of the state vector on its corresponding *a priori* value $x_{a,i}$ is given by

$$\frac{\partial \hat{x}_i}{\partial x_{a,i}} = 1 - a_{ii}, \tag{5.108}$$

where $a_{ii}$ is element (i,i) of $\mathbf{A}$. An equation similar to Eq. (5.107) has been used by Rodgers and Connor (2003) to represent retrieval results with respect to a different *a priori* state vector than had been used in the retrieval. The reason that we first solve the minimization problem (5.101) and later add *a priori* information in Eq. (5.107), instead of directly including $\mathbf{x}_a$ in the side constraint of Eq. (5.101), is that in our approach the amount of information extracted from the measurement is independent of the *a priori* state vector $\mathbf{x}_a$. So, this approach is especially suited for characterizing the information content of satellite measurements.

The weighting factors in the matrix $\mathbf{\Gamma}$ are defined relative to the values of the corresponding state vector element for the iteration step under consideration. This makes the vector $\mathbf{\Gamma x}$ dimensionless. From Eq. (5.101) it can be seen that if the weighting factor for a certain parameter decreases while the other weighting factors are kept constant, more information about this parameter is obtained from the measurement. This means that the parameters with small relative weight are less dependent on the *a priori* information added in Eq. (5.107). So, if for certain state vector elements less reliable *a priori* information is available than for others, the relative weighting factors corresponding to these parameters should be set to small values. In this way the dependence on *a priori* information for the state vector elements with small relative weight becomes smaller while for the other parameters the dependence on *a priori* assumptions becomes larger, compared to the situation where all parameters have unity relative weight. For our application it may be expected that no reliable *a priori* information will be available for the aerosol columns of both modes, because these two parameters are highly variable. Therefore, the relative weighting factors corresponding to these two parameters are set to a very low value $\epsilon$ while the other factors get a unity relative weight. We found that for $\epsilon = 1 \times 10^{-8}$ the retrieved aerosol columns for both modes are virtually independent of their *a priori* values.

From Eq. (5.107) it is clear that the retrieved state vector $\hat{\mathbf{x}}$ is affected by errors in the *a priori* state vector $\mathbf{x}_a$. The error on $\hat{\mathbf{x}}$ caused by an error on $\mathbf{x}_a$ is called the regularization error (called smoothing error by Rodgers (2000)). The regularization error covariance matrix $\mathbf{S}_r$ is given by

$$\mathbf{S}_r = (\mathbf{I}-\mathbf{A})\,\mathbf{S}_a\,(\mathbf{I}-\mathbf{A})^T, \tag{5.109}$$

where $\mathbf{S}_a$ is the *a priori* covariance matrix. Ideally, $\mathbf{S}_a$ is calculated from an ensemble of states that also include the retrieved state (Rodgers and Connor, 2003). However, for the application of aerosol satellite remote sensing $\mathbf{S}_a$ is in general

not known, which makes it difficult to calculate $\mathbf{S}_r$ for individual retrievals. However, an estimate for the upper boundary of the regularization error can be obtained by calculating $\mathbf{S}_r$ from Eq. (5.109) by assuming an *a priori* covariance matrix representing maximum values for the errors on the elements of $\mathbf{x}_a$. In order to estimate the maximum errors on the elements of $\mathbf{x}_a$ we used the 17 tropospheric aerosol models of Torres et al. (2001). For these 17 aerosol models we calculated the mean value and considered the maximum difference between the mean value and the actual value as *a priori* error. This resulted in the following *a priori* standard deviations $\sigma_a$ for the different parameters: $\sigma_a(r^s_{\text{eff}}) = 0.05\ \mu\text{m}, \sigma_a(v^s_{\text{eff}}) = 0.23, \sigma_a(r^l_{\text{eff}}) = 1.29\ \mu\text{m}, \sigma_a(v^l_{\text{eff}}) = 0.22, \sigma_a(m_r) = 0.065$, and $\sigma_a(m_i) = 0.01$. For $z_b$ and $C_{\text{pig}}$ we assumed *a priori* errors of 100%.

The contribution matrix $\mathbf{D}$ plays an important role for calculating the error propagation from measurement $\mathbf{y}$ to state vector $\mathbf{x}$. Assuming that the forward model is linear within the range of the errors, the effect of a random measurement error on the state vector is called retrieval noise. The retrieval noise covariance matrix $\mathbf{S}_x$ is given by

$$\mathbf{S}_x = \mathbf{D}\ \mathbf{S}_y\ \mathbf{D}^T. \tag{5.110}$$

Systematic state vector errors $\Delta\mathbf{x}$ due to systematic measurement errors $\Delta\mathbf{y}$ can also be evaluated using the contribution matrix:

$$\Delta\mathbf{x} = \mathbf{D}\ \Delta\mathbf{y}, \tag{5.111}$$

and a similar expression holds for systematic forward model errors $\Delta\mathbf{F}$, but with $\Delta\mathbf{y}$ replaced by $-\Delta\mathbf{F}$. Of course, the systematic errors in measurement and forward model are not known, because otherwise they would have been corrected for. However, examples of systematic state vector errors can be calculated for some reasonable scenarios of systematic measurement and forward model errors.

For estimating direct radiative forcing by aerosols, aerosol optical properties such as (spectral) optical thickness and single scattering albedo are very important. These optical properties can be derived from the microphysical aerosol parameters contained in the state vector $\mathbf{x}$. The standard deviation $\sigma_\tau$ on the optical thickness can be obtained from the retrieval noise covariance matrix $\mathbf{S}_x$ via

$$\sigma_\tau = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} s_{i,j}\ \frac{\partial\tau}{\partial x_i}\ \frac{\partial\tau}{\partial x_j}}, \tag{5.112}$$

where $s_{i,j}$ denotes element (i,j) of $\mathbf{S}_x$. The effect of the regularization error covariance matrix can be obtained in the same way. Systematic errors $\Delta\tau$ on the aerosol optical thickness $\tau$ are given by

$$\Delta\tau = \sum_{i=1}^{N} \Delta x_i\ \frac{\partial\tau}{\partial x_i}. \tag{5.113}$$

Expressions similar to Eqs (5.112) and (5.113) hold for the aerosol single scattering albedo $\omega$.

For completeness we would like to note that if in Eq. (5.101) the *a priori* state vector is included in the side constraint with $\gamma = 1$, and the inverse of the *a priori* covariance matrix is used for $\mathbf{\Gamma}$, Eq. (5.101) would be identical to the optimal estimation solution (Rodgers, 1976). The optimal estimation solution represents the maximum likelihood solution given the measurement, *a priori* information, and the corresponding covariance matrices. As stated above for the application of satellite aerosol remote sensing the *a priori* covariance matrix is not known with useful accuracy. If an *ad hoc* matrix is used that has been constructed as a rough estimate of the *a priori* covariance matrix, the optimal estimation solution reduces to the Phillips–Tikhonov solution, but with an arbitrarily chosen value for the regularization parameter (i.e. $\gamma = 1$) which is not necessarily close to the corner of the L-curve. In the Phillips–Tikhonov minimization (5.101), the matrix $\mathbf{\Gamma}$ is also an *ad hoc* matrix, but its absolute values do not affect the solution, because the side constraint is weighted by the regularization parameter, for which a suitable value is found using the L-curve. Therefore, in this chapter we prefer the use of Phillips–Tikhonov regularization in combination with the L-curve criterion for choosing the regularization parameter, instead of the use of the optimal estimation method with an *ad hoc* matrix as a rough estimate of the *a priori* covariance matrix.

### 5.6.2 Levenberg–Marquardt iteration

In general, the inversion of Eq. (5.1) represents a highly non-linear problem. Therefore, if the first guess state vector $\mathbf{x}_o$ is too far from the true state vector the linear approximation in Eq. (5.99) may be poor. In that case, the inversion of the linearized forward model in Eq. (5.99) may result in a new state vector that yields a higher $\chi^2$ difference between forward model $\mathbf{F}(\mathbf{x}_{n+1})$ and measurement $\mathbf{y}$ than the first guess state vector $\mathbf{x}_o$, i.e. a step has been taken in the wrong direction. In order to prevent the inversion from taking a large step away from the minimum $\chi^2$, we use the Levenberg–Marquardt method (Levenberg, 1944; Marquardt, 1963), which minimizes the step size between two iteration steps in addition to minimizing the difference between linearized forward model and measurement. Thus, for the first few iteration steps we replace our original cost function (5.101) by the Levenberg–Marquardt cost function, given by

$$\hat{\mathbf{x}}_{n+1}(\nu) = \min_{\mathbf{x}} \left( ||(\mathbf{K}\mathbf{x} - \tilde{\mathbf{y}})||^2 + \nu \, ||\mathbf{\Gamma}(\mathbf{x} - \mathbf{x}_n)||^2 \right), \tag{5.114}$$

where the subscripts $n$ and $n+1$ denote the current and next iteration step, respectively, $\nu$ is a parameter that controls the step size, and $\mathbf{\Gamma}$ is the same weighting factor matrix as in Eq. (5.101). Ideally, $\nu$ should be chosen such that the step taken yields an optimal improvement in $\chi^2$ between $\mathbf{F}(\mathbf{x}_{n+1})$ and $\mathbf{y}$. However, in order to find that value of $\nu$, it is necessary to evaluate $\chi^2$ for a large number of trial values of $\nu$, which requires a lot of computation time because for each trial $\nu$ a new forward model calculation is required. A more efficient application of Eq. (5.114) was proposed by Press et al. (1992). They start with a certain first guess value for $\nu$ and evaluate for that $\nu$ the $\chi^2$ between

$\mathbf{F}(\mathbf{x}_{n+1})$ and $\mathbf{y}$. If this $\chi^2$ is smaller than the $\chi^2$ of the previous iteration step, they proceed the iteration with $\mathbf{x}_{n+1}$ and decrease $\nu$ by a certain factor. If $\chi^2$ is larger than the $\chi^2$ of the previous iteration step, they proceed the iteration with $\mathbf{x}_n$ and increase $\nu$ by a certain factor. For our aerosol retrieval problem also this more efficient method still requires a large number of iteration steps (often more than 40 iteration steps are needed). Therefore, we developed a method to speed up the iteration process considerably. Here, for each iteration step we solve Eq. (5.114) for many different (say 50) values of $\nu$. Each $\nu$ results in a different retrieved state vector $\mathbf{x}_{n+1}$. For all these retrieved state vectors we use an approximate forward model to evaluate (an approximation of) the $\chi^2$ difference between forward model $\mathbf{F}(\mathbf{x}_{n+1})$ and measurement $\mathbf{y}$. This approximate forward model makes use of the fact that the forward model is more linear as a function of the discretized size distribution than as a function of $r_{\text{eff}}$ and $v_{\text{eff}}$, viz.

$$\mathbf{F}_{\text{appr}}(\mathbf{x}) = \mathbf{F}(\mathbf{x}_n) + \sum_{j=1}^{2} \sum_{i=1}^{N} \frac{\partial \mathbf{F}}{\partial n_i} \Delta n_i + \frac{\partial \mathbf{F}}{\partial \mathbf{x}_{\text{rest}}} \Delta \mathbf{x}_{\text{rest}}. \tag{5.115}$$

Here, $n_i$ is the value of the size distribution in size bin $i$, $\Delta n_i$ is a change in $n_i$ caused by a change in $r_{\text{eff}}$ and $v_{\text{eff}}$, and $\mathbf{x}_{\text{rest}}$ is the part of the state vector not including $r_{\text{eff}}$ and $v_{\text{eff}}$. Furthermore, the summation over $j$ represents the summation over the two modes of the size distribution and the summation over $i$ from 1 to $N$ represents the summation over all size bins. The forward model $\mathbf{F}_{\text{appr}}$ of Eq. (5.115) is a better approximation of the true forward model $\mathbf{F}$ than the linear approximation of Eq. (5.2), because the $\Delta n_i$ in Eq. (5.115) are calculated using the non-linear expression for the size distribution $n_i(r_{\text{eff}}, v_{\text{eff}})$. Using, the approximate forward model (5.115) we efficiently estimate the value $\nu_{\text{opt}}$ which yields optimal improvement in $\chi^2$, and use the corresponding state vector $\mathbf{x}_{n+1}(\nu_{\text{opt}})$ to proceed the iteration.

We follow the iteration process described above until $\chi^2$ becomes smaller than a certain threshold. For values of $\chi^2$ below this threshold we assume that the problem has become sufficiently linear for us to be able to replace Eq. (5.114) by the original cost function Eq. (5.101). In the final iteration step, this minimization yields the final solution of our retrieval problem with the corresponding regularization and error analysis.

## 5.7 Application to GOME-2

In this section we apply the retrieval concept presented in the previous subsections to synthetic measurements of intensity and polarization of the Polarization Measuring Device (PMD) of GOME-2. The first of three GOME-2 satellite instruments has been launched in October 2006 on the Eumetsat's Metop satellite. In total, a GOME-2 measurement series will be performed till 2020. Here, we will first demonstrate that the proposed retrieval concept is well suited to solve the non-linear aerosol retrieval problem. Furthermore, we will demonstrate that

a linear error mapping procedure, as used in Eq. (5.110), allows a sound error analysis for the application of aerosol retrieval. Finally, we will present an analysis of the information content of GOME-2 measurements of intensity and polarization, including a comparison with the information content of intensity -only retrievals from the GOME-2 PMD.

### 5.7.1 GOME-2 measurements

The PMD of GOME-2 measures the 312–800 nm spectral range using 200 detector pixels with a spectral resolution of 2.8 nm at 312 nm and about 40 nm at 800 nm. The components of the intensity polarized parallel and perpendicular to the optical plane are measured simultaneously. These components are denoted by $I_i^l$ and $I_i^r$, respectively, for detector pixel $i$. The design of the PMD was driven by the optical identity of the l- and r-channels. Given the need of a lightweight instrument, a prism spectrometer was chosen instead of a more complex grating solution. The measurement $I_i^l$ is simulated by

$$I_i^l = \int_0^\infty \mathrm{d}\lambda \; S_i(\lambda) \; I^l(\lambda), \tag{5.116}$$

where the integration over wavelength $\lambda$ describes the effect of a Gaussian spectral response function $S(\lambda)$, where $I^l(\lambda)$ denotes the $l$-component of the intensity at the entrance of the instrument. The measurement $I_i^r$ is simulated in the same manner as in Eq. (5.116). From the measured intensities $I_i^l$ and $I_i^r$ the Stokes parameters $I_i$ and $Q_i$ can be obtained, viz.,

$$I_i = I_i^l + I_i^r, \tag{5.117}$$

$$Q_i = I_i^l - I_i^r. \tag{5.118}$$

The simulated values of $I_i$ and $Q_i$ are superimposed by a random Gaussian noise. Here, we calculated the contributions of photon-shot-noise and detector-noise using the transmission properties of GOME-2, known from the on-ground calibration. However, we believe that considering only these two error sources an unrealistically positive retrieval diagnose will be obtained, because the measurement possibly also contains other errors which may introduce a random-like structure, such as errors due to spatial aliasing, unknown spectral features introduced by the diffuser plate, and spectral calibration errors. Also the forward model may contain random-like errors, for example due to errors in accounting for molecular absorption, the description of underlight, the assumed distribution of surface slopes of the oceanic waves, and the prediction of whitecap coverage from wind speed. Therefore, in addition to the contributions of photon-shot-noise and detector-noise, a noise floor of 1% is added to the simulations of $I_i$ and $Q_i$, to account for such errors.

Due to limitations in the GOME-2 data rate the information of the 200 detector pixels has to be co-added onboard to form 15 programmable bands. The expected wavelength ranges of these bands are denoted in Table 5.1. For

**Table 5.1.** Probable PMD band selection for GOME-2 (Hasekamp et al., 2004a). The wavelength range is indicated by the center wavelengths of the first and last PMD pixel, respectively, not including the Full Width at Half Maximum (FWHM) of the slit function, which is shown in the last column

| Band no. | Wavelength range (nm) | No. of pixels | FWHM (nm) |
|---|---|---|---|
| 1 | spare | – | – |
| 2 | 311.8–314.3 | 5 | 3.1 |
| 3 | 316.9–318.8 | 4 | 3.3 |
| 4 | 321.5–329.3 | 12 | 3.5 |
| 5 | 330.8–334.6 | 6 | 3.8 |
| 6 | 336.2–340.0 | 24 | 4.8 |
| 7 | 361.0–377.9 | 20 | 4.8 |
| 8 | 380.3–383.9 | 4 | 6.1 |
| 9 | 399.8–428.0 | 19 | 7.8 |
| 10 | 435.1–493.5 | 23 | 10.2 |
| 11 | 495.5–552.4 | 23 | 12.5 |
| 12 | 567.9–598.0 | 5 | 25.2 |
| 13 | 600.0–660.0 | 11 | 30.0 |
| 14 | 743.1–766.6 | 3 | 38.5 |
| 15 | 783.6–792.4 | 2 | 43.9 |

aerosol retrieval it is anticipated that band 6–15 will be used. The intensity $I_{\text{pmd}}$ for a given PMD band is given by

$$I_{\text{pmd}} = \sum_{i=1}^{N} I_i, \tag{5.119}$$

where the summation in Eq. (5.119) describes the co-adding over a number of $N$ detector pixels. The Stokes parameter $Q_{\text{pmd}}$ is obtained in the same manner. The standard deviation $\sigma_{\text{pmd}}$ for the Gaussian error on $I_{\text{pmd}}$ or $Q_{\text{pmd}}$ is given by

$$(\sigma_{\text{pmd}})^2 = \frac{1}{N}\left(\sum_{i=1}^{N} (\sigma_i)^2\right), \tag{5.120}$$

where $\sigma_i$ is the standard deviation for a detector pixel.

In this study we consider retrievals using intensity as well as polarization measurements and retrievals using only intensity measurements. For the latter retrievals we also use the PMD spectral bands of Table 5.1. In this comparison we want to avoid differences in information content due to the fact that adding an extra set of intensity measurements improves the signal-to-noise ratio by a factor $\sqrt{2}$. Therefore, for the calculation of the measurement error covariance matrix for the intensity -only retrievals we used the signal-to-noise ratio corresponding to a double set of intensity measurements.

### 5.7.2 Retrieval results

The retrieval procedure described in this chapter is tested on a set of 100 synthetic GOME-2 PMD measurements of intensity and polarization, created for randomly chosen aerosol parameters within a specified range. The chosen ranges for the different parameters were: 0.1–0.2 $\mu m$ for $r_{\text{eff}}$ of the small mode, 0.65–3.40 $\mu m$ for $r_{\text{eff}}$ of the large mode, 0.16–0.65 for $v_{\text{eff}}$ of the small mode, 0.5–0.9 for $v_{\text{eff}}$ of the large mode, 1.4–1.6 for $m_r$, $5 \cdot 10^{-7}$–0.02 for $|m_i|$, and aerosol columns for both modes corresponding to an optical thickness at 550 nm in the range 0.05–0.5. The oceanic pigment concentration ranged from 0.5–2 mg/m$^3$, whereas the height $z_b$ of the layer where the bulk of the aerosols is located (see Eq. 5.98) was kept fixed at 2 km.

In all cases the iteration converged to a stable solution and also the $\chi^2$ difference between forward model and measurement was close to 1 in all cases. Figure 5.4 shows the retrieved optical thickness (i.e. derived from the retrieved parameters) versus the true optical thickness at 550 nm. It can be seen that the retrieved optical thickness corresponds well to the true optical thickness. In general, the agreement is within 5%. This example indicates that the implemented retrieval approach is suited for aerosol retrieval.

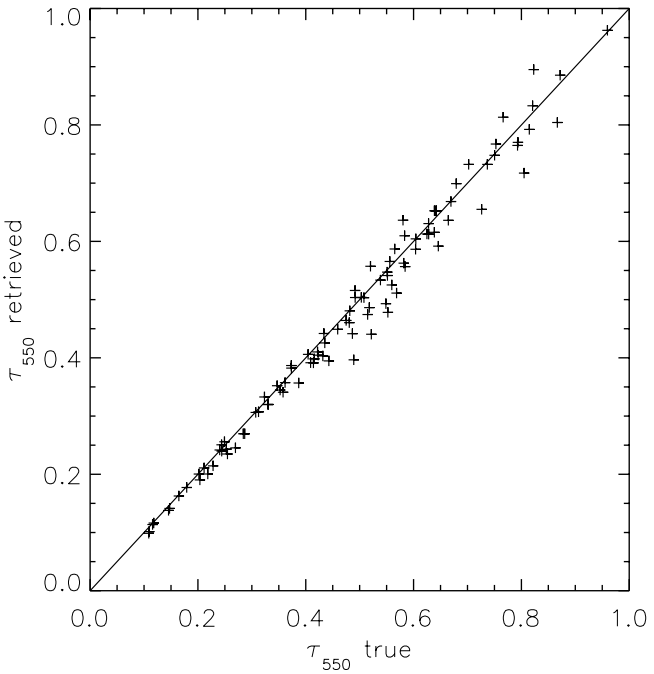



**Fig. 5.4.** Retrieved optical thickness versus true optical thickness at 550 nm for 100 synthetic retrievals.

An error analysis based on a linearized forward model, as is performed in Eq. (5.110), is only valid if the forward model is (approximately) linear within the error range. In order to test if a sound error analysis can be performed using a linearized forward model, we investigate whether the differences between the retrieved state vector and the true state vector are consistent with the retrieval noise covariance matrix calculated by Eq. (5.110). In order to exclude the effect of the *a priori* state vector from the comparison we replace $\mathbf{x}_a$ by $\mathbf{x}_{\text{true}}$ in the term $(\mathbf{I}-\mathbf{A})\mathbf{x}_a$ in Eq. (5.107) for the final iteration step. If the differences between the retrieved state vector and the true state vector are consistent with the retrieval noise covariance matrix (5.110), then the distribution of $(x_i^r - x_i^t)/\sigma_i$, where $x_i^r$ is the $i$th element of the retrieved state vector, $x_i^t$ is the corresponding true value, and $\sigma_i$ the standard deviation that follows from Eq. (5.110), is given by the standard Gaussian distribution

$$f(y) = \frac{\exp(-y^2/2)}{\sqrt{2\pi}}. \tag{5.121}$$

For the retrievals on the 100 synthetic measurements described above, the corresponding distribution is shown in Fig. 5.5. The distribution shown contains the values $(x_i^r - x_i^t)/\sigma_i$ for all state vector elements. From Fig. 5.5 it follows that the distribution of the retrieved aerosol parameters reproduces the standard Gaussian distribution well. This demonstrates that the linear approximation of Eq. (5.110) is valid for the calculation of the retrieval noise covariance matrix. So, a linear error mapping procedure can be used to investigate the retrieval capabilities of a given instrument concept, without doing a full iterative retrieval. The linearized radiative transfer model described in this chapter is a powerful tool for this purpose.

### 5.7.3 Information content

We investigated the information content of GOME-2 measurements using linearized radiative transfer calculations for the two aerosol types in Table 5.2, where the optical thickness at 550 nm $\tau_{550} = 0.3$. Here, aerosol type A corresponds to biomass-burning aerosols and type B corresponds to oceanic aerosols. Figure 5.6 shows the DFS as a function of viewing zenith angle (VZA) for a solar zenith angle (SZA) of 40° and a relative azimuth angle of 60° (positive VZA) or −120° (negative VZA), for retrievals using intensity as well as polarization measurements (left panel) and for retrievals using only intensity measurements (right panel). From this figure it follows that for retrievals using intensity as well as polarization measurements the DFS is in the range 6–8 which is 1–4 degrees higher than for retrievals using only intensity measurements. So, the use of polarization measurements significantly improves the information content.

In order to interpret by which parameters the DFS is mainly determined, we show in Fig. 5.7 the derivatives of the retrieved parameters with respect to their *a priori* values for the biomass-burning aerosol type. Here, if for a parameter $\partial x/\partial x_a = 1$, this parameter is fully determined by its *a priori* value, whereas if $\partial x/\partial x_a = 0$ the parameter is not influenced by its *a priori* value at all. Here, it

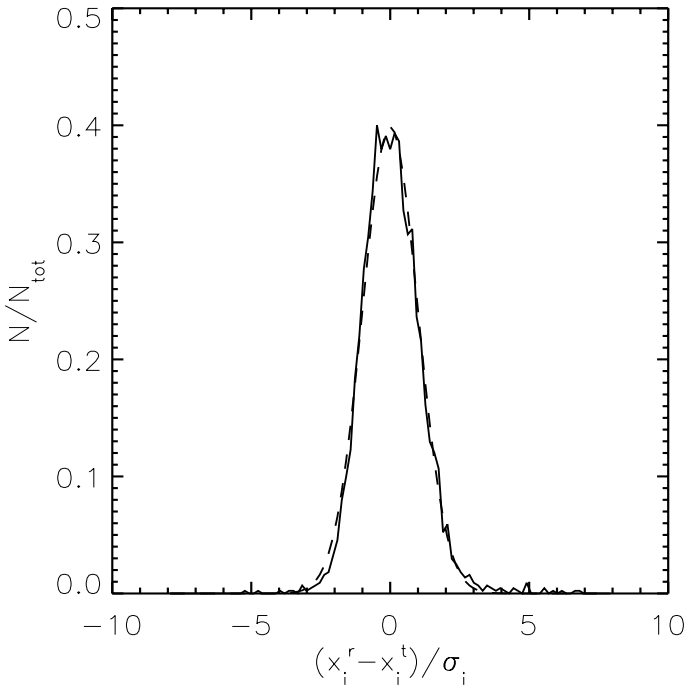



**Fig. 5.5.** Distribution of $(x_i^r - x_i^t)/\sigma_i$ for the 100 retrievals on synthetic GOME-2 measurements (solid line). $N/N_{\rm tot}$ indicates the number of points in a certain size bin normalized to the total number of points. The standard Gaussian distribution of Eq. (5.121) is given by the dashed line. The distribution contains 101 bins between $-8$ and 8.

**Table 5.2.** Aerosol types used to create synthetic measurements of intensity and polarization. The aerosol types are adopted from Torres et al. (2001). Type A corresponds to biomass-burning aerosols and type B corresponds to oceanic aerosols. See Appendix B for definitions of $r_{\rm eff}$ and $v_{\rm eff}$. $f_l$ denotes the fraction of large mode particles. Concerning the aerosol altitude distribution of Eq. (5.98), $z_b = 2$ km and $z_t = 10$ km

| Type | $r^s_{\rm eff}$ | $v^s_{\rm eff}$ | $r^l_{\rm eff}$ | $v^l_{\rm eff}$ | $f_l$ | $\tau^l_{550}/\tau^{\rm tot}_{550}$ | $m_r$ | $m_i$ | $\tau_{350}$ | $\tau_{550}$ | $\omega_{350}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 0.119 | 0.174 | 2.671 | 0.704 | $2.05 \cdot 10^{-4}$ | 0.087 | 1.50 | $-0.02$ | 0.657 | 0.300 | 0.892 |
| B | 0.105 | 0.651 | 0.840 | 0.651 | $1.53 \cdot 10^{-2}$ | 0.851 | 1.40 | $-5 \cdot 10^{-8}$ | 0.336 | 0.300 | 1.000 |

is important to note that due to our choice of the matrix $\mathbf{\Gamma}$ in Eq. (5.101) we force the aerosol columns of both modes to be fully independent of their *a priori* values, i.e. $\partial x/\partial x_a = 0$ for these parameters. Therefore, these parameters are not included in Fig. 5.7. It follows from Fig. 5.7 that the polarization measurements mostly add information on the effective variance of the small mode, the imagi-

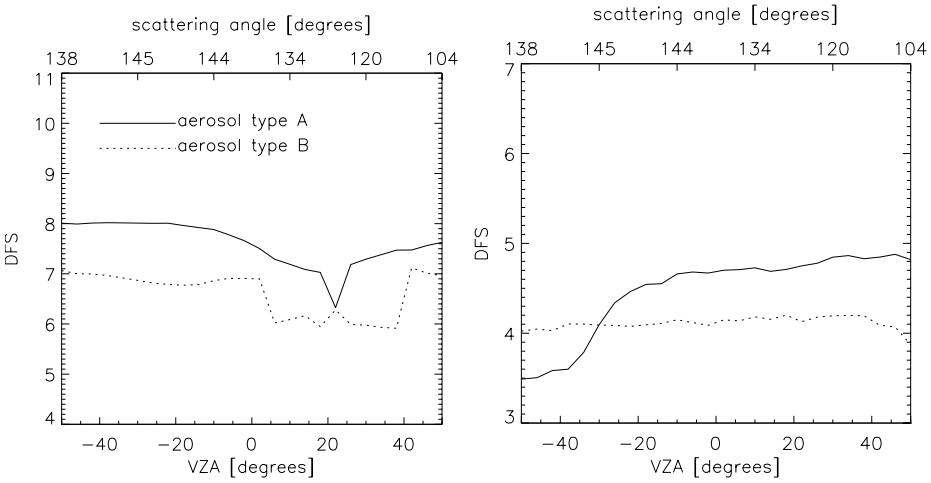



**Fig. 5.6.** Degrees of Freedom for Signal (DFS) for retrievals from GOME-2 PMD intensity and polarization measurements (left panel) and for retrievals using only intensity measurements (right panel). The two aerosol types of Table 5.2 are used. The DFS is shown as a function of viewing zenith angle (VZA) for a solar zenith angle of $40^o$ and a relative azimuth angle $\varphi_o - \varphi_v = 60^o$ for positive VZA and $\varphi_o - \varphi_v = -120^o$ for negative VZA. An oceanic pigment concentration of 1 mg/m$^3$ was used for the simulations.

nary part of the refractive index, and the height $z_b$ of the layer where the bulk of the aerosols is located. Furthermore, significant additional information can be retrieved on the effective radius of the small mode and the oceanic pigment concentration. Both retrievals contain little information on the size distribution parameters of the large mode. This can be explained by the fact that for the biomass-burning aerosol type the contribution of the large mode to the total optical thickness is relatively small. In contrast, the oceanic aerosol type (not shown) is dominated by the large mode. Therefore, for this aerosol type more information is available on the effective radius of the large mode, while the effective radius of small mode depends stronger on *a priori* information. However, the effective variance of both modes strongly depends on *a priori* for the oceanic aerosol type.

From Fig. 5.7 we conclude that the use of polarization measurements makes it possible to retrieve information on aerosol size and refractive index that cannot be retrieved using only intensity measurements. This can be explained by the characteristic sensitivity of polarization properties of light to aerosol microphysical properties, as shown, for example, by Hansen and Travis (1974). Furthermore, as follows from Fig. 5.7, polarization measurements allow the retrieval of information on aerosol height. This information mainly comes from measurements at wavelengths below about 450 nm, where the Rayleigh scattering optical thickness is relatively large. Since most Rayleigh scattering takes place low in the atmosphere, the Rayleigh scattering signal is more strongly attenuated if

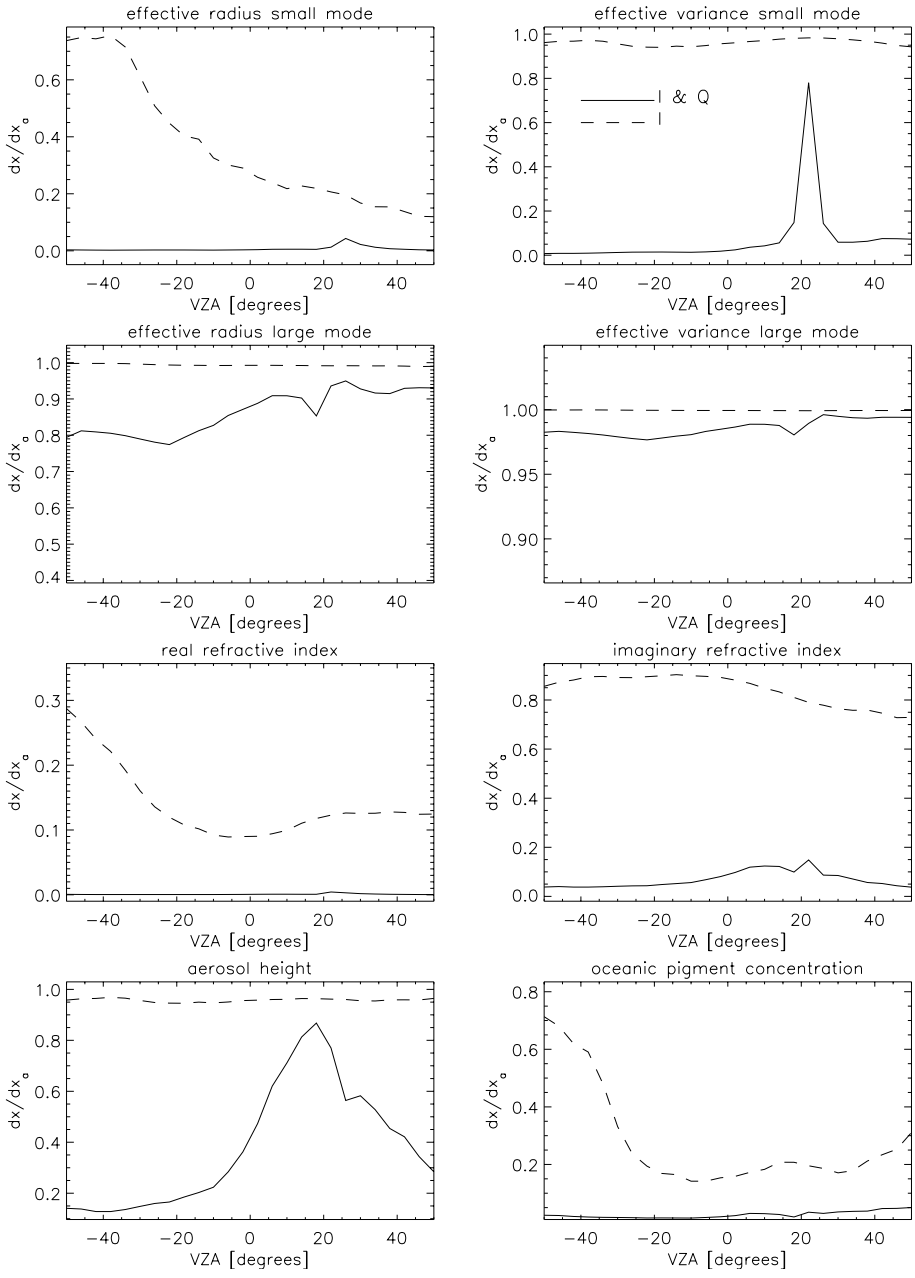



**Fig. 5.7.** Derivatives of retrieved values with respect to their *a priori* values for the biomass-burning aerosol type (A), as a function of viewing zenith angle. Other angles as in Fig. 5.6.

the aerosols are located higher in the atmosphere. So, the degree of polarization of the backscattered light becomes lower for increasing aerosol height, because Rayleigh scattering generally causes a higher degree of polarization than aerosol scattering. For the oceanic aerosol type, there is significantly less information available on aerosol height (not shown) then for the biomass-burning aerosol type, because for the oceanic aerosol type the aerosol optical thickness at short wavelengths is much smaller than for biomass-burning aerosols. In addition to the aerosol parameters, the oceanic pigment concentration also can be retrieved using GOME-2 intensity and polarization measurements. This is due to the characteristic spectral signature of oceanic pigment.

Figure 5.8 shows for aerosol type A the total retrieval error (retrieval noise and regularization error) on the optical thickness at 350 nm and 550 nm, respectively, for retrievals using intensity and polarization measurements and retrievals using only intensity measurements. It can be seen that for the retrievals using intensity and polarization measurements, the optical thickness error shows a distinct maximum around a VZA of 20°. The reason for this is that at these geometries the sensitivity of Stokes parameter $Q$ to atmospheric properties is rather low, which means that here the retrievals rely for a large part on intensity measurements. This strong dependence on viewing geometry demonstrates that the aerosol information retrieved using single-viewing-angle polarization measurements is for some geometries less useful than for other geometries. These geometries are well defined and the corresponding aerosol retrieval products should be labeled as less reliable. Away from this maximum, the optical thickness error is around 0.025 (3.7%) at 350 nm and around 0.017 (5.7%) at 550 nm. The optical thickness errors for retrievals using only intensity measurements are

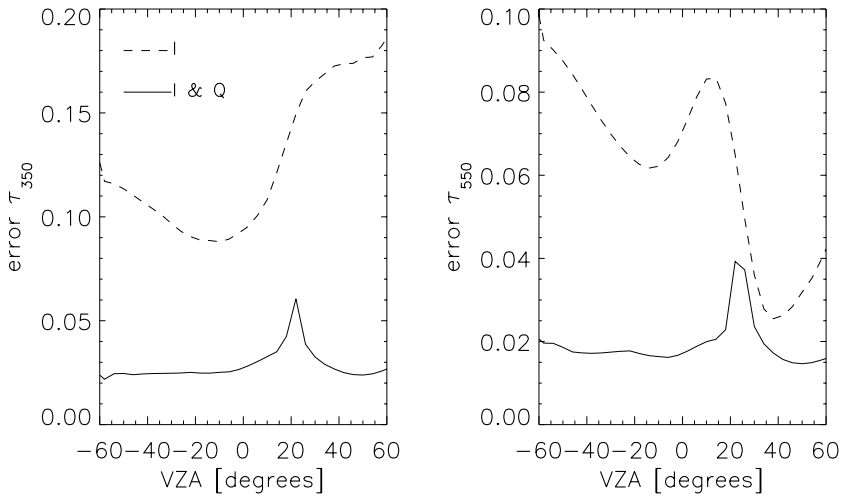



**Fig. 5.8.** Total retrieval error on the retrieved optical thickness at 350 nm (left panel) and 550 nm (right panel), as a function of viewing zenith angle. Other angles as in Fig. 5.6.

a factor 2–7 higher. The increase in total optical thickness error mainly comes from the regularization error, which means that optical thickness retrievals from only intensity measurements are very sensitive to *a priori* information on aerosol size distribution and refractive index. So, the additional information on aerosol size distribution and refractive index (see Fig. 5.7) that can be retrieved including polarization measurements, is not only important information on its own, but is also essential if a reliable optical thickness retrieval is to be obtained.

Figure 5.9 shows the total retrieval error on the aerosol single scattering albedo $\omega_{350}$ at 350 nm, for retrievals using intensity as well as polarization measurements, and retrievals using only intensity measurements. Here, we use a wavelength of 350 nm because information about aerosol single scattering albedo mainly comes from shorter wavelengths due to interaction with Rayleigh scattering (Torres et al., 1998). For retrievals using intensity as well as polarization measurements, the total error on the single scattering albedo is mostly below 0.015. For intensity -only retrievals the total error on $\omega_{350}$ is about a factor 2–4 larger than for retrievals using polarization measurements. So, the retrieval of single scattering albedo also benefits significantly from including polarization measurements.

To summarize, multi-wavelength single-viewing-angle measurements of intensity as well as polarization in the range 340–800 nm contain valuable information on aerosol size, refractive index, spectral optical thickness, and UV-single scattering albedo. These aerosol characteristics are of essential importance for climate research. Using only intensity measurements in the same spectral range, significantly less information on microphysical aerosol properties can be retrieved,

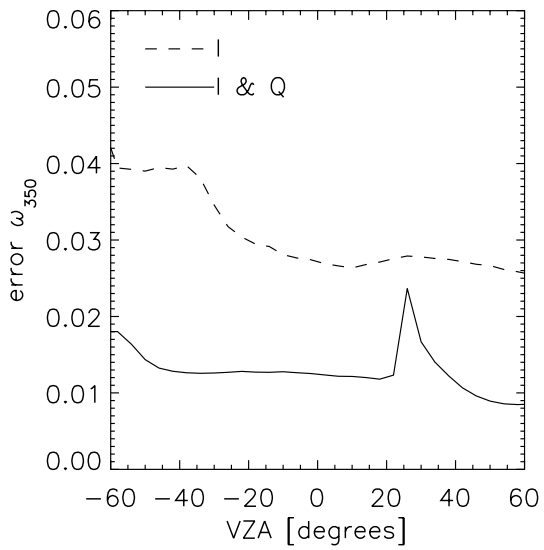



**Fig. 5.9.** Total retrieval error on the retrieved single scattering albedo at 350 nm as a function of viewing zenith angle. Other angles as in Fig. 5.6.

leading to (much) larger errors on the corresponding retrieved optical thickness and single scattering albedo. These conclusions have also been tested for other aerosol types, and similar results were obtained. In addition, multi-viewing-angle measurements will provide information on the aerosol phase matrix which in turn will provide additional constraints on microphysical aerosol properties and on surface reflectance properties.

## 5.8 Conclusions

The analytical linearization of vector radiative transfer with respect to physical aerosol properties and its use in satellite remote sensing have been reviewed. The linearization consists of two steps. The first step is the calculation of the derivatives of the four Stokes parameters at the top of the atmosphere with respect to scattering coefficient, absorption coefficient, and the expansion coefficients of the scattering phase matrix. These derivatives are calculated analytically employing the forward-adjoint perturbation theory. Here, general expressions are presented that can be applied for the linearization of any vector radiative transfer model that calculates the internal radiation field in the model atmosphere. The second step is the calculation of the derivatives of the scattering coefficient, absorption coefficient, and the expansion coefficients of the scattering phase matrix, with respect to the real and imaginary part of the refractive index, and parameters describing the size distribution (e.g. effective radius, effective variance). These derivatives are analytically calculated following Mie theory. The developed linearization approach has been implemented in a Gauss–Seidel vector radiative transfer model. The linearized radiative transfer model has been incorporated in a retrieval algorithm based on the Phillips–Tikhonov regularization method in combination with the Levenberg–Marquardt iterative method. This retrieval algorithm aims to retrieve microphysical aerosol parameters corresponding to a bi-modal aerosol size distribution. Additionally, the oceanic pigment concentration and information on aerosol height are retrieved from the measurement. We used synthetic GOME-2 measurements of intensity and polarization to demonstrate that the developed iterative retrieval approach based on linearized radiative transfer is well suited to solve the non-linear aerosol retrieval problem. Furthermore, we demonstrated that a linear error mapping procedure can be used to perform a solid error analysis, without doing a full iterative retrieval.

Finally, we presented and overview of the information content of GOME-2 measurements of intensity and polarization. Here, we considered the retrieval of nine aerosol parameters corresponding to a bi-modal aerosol size distribution: the column integrated aerosol number concentration of both modes, the effective radius of both modes, the effective variance of both modes, the real- and imaginary part of the refractive index, and the height of the layer where the bulk of the aerosols is located. In addition to the nine aerosol parameters we also considered the oceanic pigment concentration as an unknown parameter. It is demonstrated that for this retrieval setup the DFS is in the range 6–8. Here, the aerosol loading of both modes, the effective radius of at least one mode, the

real and imaginary part of the refractive index, the height of the layer where the bulk of the aerosols is located, and the oceanic pigment concentration can for most viewing geometries be retrieved from the measurement with negligible dependence on *a priori* information.

For retrievals that only use intensity measurements the DFS is significantly less than for retrievals also using polarization measurements, namely in the range 3.5–5. For these retrievals no significant information on aerosol imaginary refractive index, effective variance, and aerosol height can be retrieved. Furthermore, the information on effective radius, real part of the refractive index, and oceanic pigment concentration is much more affected by *a priori* information than retrievals that include polarization measurements.

To conclude, the results of this chapter demonstrate that a linearized radiative transfer model as presented here provides a powerful tool for efficiently solving the aerosol retrieval problem, and additionally for a solid error analysis. Using this tool, we showed that multi-wavelength single-viewing-angle measurements of intensity as well as polarization in the range 340–800 nm contain valuable information on aerosol size, refractive index, spectral optical thickness, and UV-single scattering albedo. Using only intensity measurements in the same spectral range significantly less information on microphysical aerosol properties can be retrieved, leading to (much) larger errors in the corresponding retrieved optical thickness and single scattering albedo. The retrievals can be further improved using multiple-viewing-angle measurements and highly spectrally resolved measurements in absorption bands of well mixed atmospheric gases, such as oxygen.

## Appendix A: The Mie coefficients and their derivatives

The Mie coefficients $a_n$ and $b_n$ are calculated using the method of de Rooij and van der Stap (1984). Here, we will summarize the relevant formulas and for further details we refer to the corresponding paper. Furthermore, we give expressions for the derivatives of $a_n$ and $b_n$ with respect to the real and imaginary part of the refractive index, used in section 5.4.2. The Mie coefficients are given by (see, for example, Deirmendjian (1969)):

$$a_n = \frac{(D_n(z)/m + n/x)\ \Psi_n(x) - \Psi_{n-1}(x)}{(D_n(z)/m + n/x)\ \zeta_n(x) - \zeta_{n-1}(x)}, \tag{5.122}$$

$$b_n = \frac{(mD_n(z) + n/x)\ \Psi_n(x) - \Psi_{n-1}(x)}{(mD_n(z) + n/x)\ \zeta_n(x) - \zeta_{n-1}(x)}, \tag{5.123}$$

where $m = m_r + im_i$ is the complex refractive index, $x$ is the size parameter $2\pi r/\lambda$, and $z = mx$. Furthermore,

$$\Psi_n(x) = xj_n(x), \tag{5.124}$$

$$\zeta_n(x) = \Psi_n(x) + i\chi_n(x), \tag{5.125}$$

with

$$\chi_n(x) = -xy_n(x), \tag{5.126}$$

where $j_n(x)$ and $y_n(x)$ are the spherical Bessel functions of the first and second kind, respectively. $D_n(z)$ is the only function that depends on refractive index and is given by

$$D_n(z) = \frac{d}{dz} \ln \Psi_n(z) = -\frac{n}{z} \frac{\Psi_{n-1}(z)}{\Psi_n(z)}. \tag{5.127}$$

The functions $\Psi_n(x)$ and $\chi_n(x)$ and $D_n(z)$ are all calculated using recurrence relations. Here, $\chi_n(x)$ is calculated by upward recursion using the recurrence relation

$$\chi_{n+1}(x) = \frac{2n+1}{x} \chi_n(x) - \chi_{n-1}(x), \tag{5.128}$$

with initial functions

$$\chi_{-1}(x) = \sin x, \qquad \chi_0(x) = \cos x. \tag{5.129}$$

$\Psi_n(x)$ is calculated using downward recursion:

$$\Psi_n(x) = r_n(x)\ \Psi_{n-1}(x), \tag{5.130}$$

where

$$r_n(x) = \left[\frac{2n+1}{x} - r_{n+1}(x)\right]^{-1}. \tag{5.131}$$

The recursion is started at $n = N_1(x)$ where

$$N_1(x) = x + 4.05x^{1/3} + 60, \tag{5.132}$$

and $r_{N_1}(x) = 0$ (de Rooij and van der Stap, 1984).

$D_n(z)$ is calculated using the following downward recursion relation:

$$D_n(z) = \frac{n+1}{z} - \left(D_{n+1}(z) + \frac{n+1}{z}\right)^{-1}, \tag{5.133}$$

where the recursion is started at $n = N_2(z)$ with

$$N_2(z) = z + 4.05z^{1/3} + 10, \tag{5.134}$$

and $D_{N_2}(z) = 0$ (de Rooij and van der Stap, 1984).

The derivatives of $a_n$ and $b_n$ with respect to the real and imaginary part of the the refractive index are given by

$$[a_n]' = \frac{\left([D_n(z)]'/m - D_n(z)/m^2\right)(\Psi_{n-1}\zeta_n - \Psi_n\zeta_{n-1})}{[(D_n(z)/m + n/x)\ \zeta_n(x) - \zeta_{n-1}(x)]^2}, \tag{5.135}$$

$$[b_n]' = \frac{m[D_n(z)]'\ (\Psi_{n-1}\zeta_n - \Psi_n\zeta_{n-1})}{[(mD_n(z) + n/x)\ \zeta_n(x) - \zeta_{n-1}(x)]^2}, \tag{5.136}$$

where the prime indicates the derivative with respect to either $m_r$ or $im_i$. Here, it is important to note that in section 5.4.2 we use the derivatives with respect

to $m_i$ which follow directly from the here given derivatives with respect to $im_i$. The derivative $[D_n(z)]'$ is found by backward recursion via

$$[D_n(z)]' = \frac{-x(n+1)}{z^2} + \left([D_{n+1}(z)]' + \frac{-x(n+1)}{z^2}\right) \left(D_{n+1}(z) + \frac{n+1}{z}\right)^{-2}, \tag{5.137}$$

starting the recursion at $N_2$ with $[D_{N_2}(z)]' = 0$.

## Appendix B: Aerosol and ocean properties

### B.1 Aerosol size distribution

For all simulations in this chapter we assume that the aerosol size distribution is bi-modal, where the size distribution $n$ for each mode is given by a lognormal function

$$n(r) = \frac{1}{\sqrt{2\pi}\,\sigma_g\,r} \exp\left[-(\ln r - \ln r_g)^2/(2\sigma_g^2)\right], \tag{5.138}$$

where $r$ describes particle radius,

$$\ln r_g = \int_0^\infty \ln r\; n(r)\,\mathrm{d}r, \tag{5.139}$$

and

$$\sigma_g^2 = \int_0^\infty (\ln r - \ln r_g)^2\; n(r)\,\mathrm{d}r. \tag{5.140}$$

As shown by Hansen and Travis (1974) it is useful to characterize (a mode of) the size distribution by the effective radius $r_{\text{eff}}$ and effective variance $v_{\text{eff}}$, because these parameters are relatively independent from the actual shape of the distribution. Here,

$$r_{\text{eff}} = \frac{1}{G} \int_0^\infty r\pi r^2 n(r)\,\mathrm{d}r, \tag{5.141}$$

and

$$v_{\text{eff}} = \frac{1}{G r_{\text{eff}}^2} \int_0^\infty (r - r_{\text{eff}})^2 \pi r^2 n(r)\,\mathrm{d}r, \tag{5.142}$$

where $G$ is the geometrical cross-section. We use the superscripts $l$ and $s$ to refer to the small and large mode of the size distribution, respectively.

### B.2 Ocean description

For the retrieval simulations in this chapter, the lower boundary of the model atmosphere is characterized by the reflection matrix of the ocean. The ocean reflection can be described by three contributions (see, for example, Chowdhary (1999) and references therin): (1) Fresnel reflection on the oceanic waves.

This contribution is mainly determined by the wind speed $W$. (2) Scattering inside the ocean body called underlight. In this chapter we restrict ourselves to the open ocean (so called 'case-1 waters' (Morel and Prieur, 1977)) for which the reflection due to underwater scattering is predominantly determined by the concentration of phytoplankton and its derivative products, referred to as the oceanic pigment concentration $C_{\text{pig}}$. (3) Reflection by oceanic foam, which depends on the foam albedo $A_{\text{fm}}$ (see, for example, Koepke (1984), Frouin et al. (1996) and Kokhanovsky (2004)) and the fraction of the ground pixel that is covered by foam, which depends on the wind speed. So, the total ocean reflection depends mainly on the wind speed, the oceanic pigment concentration, and the foam albedo.

For the simulations in this chapter, the Fresnel reflection on the waves is calculated using the method of Mishchenko and Travis (1997), assuming the wind speed dependent distribution of surface slopes proposed by Cox and Munk (1954). Here, we used a windspeed $W = 7$ m/s throughout this chapter. For the foam albedo $A_{fm}$ we assume a fixed value of 0.2, which is close to the value proposed by Koepke (1984) for the visible spectral range. For the wind speed dependent fraction $l_{fm}$ of the ground pixel that is covered by foam we use $l_{fm} = 2.95 \times 10^{-6}\, W^{3.52}$ (Monahan and O'Muircheartaigh, 1980).

The underlight contribution is described using a Lambertian albedo that depends on the oceanic pigment concentration, using the dependence given by Morel (1988) and Morel and Gentili (1993) (an improved model has been published by Morel and Maritorena (2001)), in combination with the data of Smith and Baker (1981). So, bi-directional effects and polarization are neglected using this simplified description of underlight. Since the underlight contribution is largest below 500 nm, the errors in the underlight contribution also will be largest for these wavelengths. However, the effect of errors in the ocean description on the intensity vector at the top of the atmosphere will be relatively small, since the atmospheric contribution to the intensity vector at the top of the atmosphere is much larger than the oceanic contribution at these wavelengths. The neglect of bi-directional effects can cause an error in the underlight contribution of roughly 20% directly above the ocean surface (Morel and Gentili, 1993; Chowdhary, 1999) but is in general smaller than 1.5% at the top of the atmosphere for wavelengths below 500 nm. The neglect of polarization in the underlight contribution causes for some geometries maximum errors of 1–2% in Stokes parameter $Q$ at the top of the atmosphere for realistic ocean models (Chowdhary, 1999). We expect that the simplified description of the underlight contribution does not significantly affect the sensitivity study results of this chapter. However, for aerosol retrieval from real measurements it is worthwhile to consider a more advanced ocean description (Chowdhary et al., 2005). Also for the retrieval of aerosol properties over coastal waters a more advanced ocean description should be considered.

**Acknowledgements**

This research was supported by the Dutch User Support Programme 2001-2005 (USP) under project number EO-069 and an Ozone Monitoring SAF Visiting Scientist Grant.

## References

Bell, G. I. and Glasstone, S., 1970. *Nuclear Reactor Theory*. Van Nostrand Reinhold Company, New York.

Box, M., Gerstl, S., and Simmer, C., 1988. Application of the adjoint formulation to the calculation of atmospheric radiative effects. *Beitr. Phys. Atmos.*, **61**, 303–311.

Carter, L. L., Horak, H. G., and Sandford, M. T., 1978. An adjoint Monte Carlo treatment of the equation of radiative transfer for polarized light. *J. Comp. Phys.*, **26**, 119–138.

Chandrasekhar, S., 1960. *Radiative transfer*. Dover Publications, Inc., New York.

Chowdhary, J., 1999. *Multiple scattering of polarized light in atmosphere-ocean systems: Application to sensitivity analyses of aerosol polarimetry*. Ph.D. thesis, Columbia University.

Chowdhary, J., Cairns, B., Mishchenko, M., Hobbs, P., Cota, G., Redemann, J., Rutledge, K., Holben, B., and Russel, E., 2005. Retrieval of aerosol scattering and absorption properties from photo-polarimetric observations over the ocean during the CLAMS experiment. *J. Atmos. Sci.*, **62**(4), 1093–1117.

Cox, C. and Munk, W., 1954. Statistics of the sea surface derived from sun glitter. *J. Mar. Res.*, **13**, 198–227.

de Haan, J., Bosma, P., and Hovenier, J., 1987. The adding method for multiple scattering calculations of polarized light. *Astron. Astrophys.*, **181**, 371–391.

Deirmendjian, D., 1969. *Electromagnetic Scattering on Spherical Polydispersions*. Elsevier, New York.

de Rooij, W. A. and van der Stap, C. C. A. H., 1984. Expansion of Mie scattering matrices in generalized spherical functions. *Astron. Astrophys.*, **131**, 237–248.

Domke, H., 1975. Fourier expansion of the phase matrix for Mie scattering. *Z. Meteorologie*, **25**, 357.

Frouin, R., Schwindling, M., and Deschamps, P., 1996. Spectral reflectance of sea foam in the visible and near-infrared: In situ measurements and remote sensing implications. *J. Geophys. Res.*, **101**, 14 361–14 371.

Gel'fand, I., Minlos, R., and Shapiro, Z., 1963. *Representations of the Rotation and Lorentz Groups and their Applications*. Pergamon Press, Oxford.

Hansen, J. E. and Travis, L. D., 1974. Light scattering in planetary atmospheres. *Space Sci. Rev.*, **16**, 527–610.

Hansen, P., 1992. Analysis of discrete ill posed problems by means of the L-curve. *SIAM Rev.*, **34**, 561–580.

Hansen, P. and O'Leary, D., 1993. The use of the L-curve in the regularization of discrete ill posed problems. *SIAM J. Sci. Comput.*, **14**, 1487–1503.

Hasekamp, O. and Landgraf, J., 2002. A linearized vector radiative transfer model for atmospheric trace gas retrieval. *J. Quant. Spectrosc. Radiat. Transfer*, **75**, 221–238.

Hasekamp, O., Landgraf, J., Hartmann, W., and Aben, I., 2004. Proposal for GOME-2 PMD wavelength band selection and the effect on aerosol retrieval. Techn. Rep. SRON-EOS/RP/04-002, SRON, Utrecht, The Netherlands.

Hasekamp, O. P. and Landgraf, J., 2005a. Linearization of vector radiative transfer with respect to aerosol properties and its use in satellite remote sensing. *Journal of Geophysical Research (Atmospheres)*, **110**, 4203.

Hasekamp, O. P. and Landgraf, J., 2005b. Retrieval of aerosol properties over the ocean from multispectral single-viewing-angle measurements of intensity and polarization: Retrieval approach, information content, and sensitivity study. *Journal of Geophysical Research (Atmospheres)*, **110**, 20 207.

Hovenier, J. W., 1969. Symmetry relationships for scattering of polarized light in a slab of randomly oriented particles. *J. Atmos. Sci.*, **26**, 488–499.

Hovenier, J. W. and van der Mee, C. V. M., 1983. Fundamental relationships relevant to the transfer of polarized light in a scattering atmosphere. *Astron. Astrophys.*, **128**, 1–16.

Koepke, P., 1984. Effective reflectance of oceanic whitecaps. *Appl. Opt.*, **23**, 1816–1824.

Koepke, P. and Hess, M., 1988. Scattering functions of tropospheric aerosols: The effect of nonspherical particles. *Appl. Opt.*, **27**, 2422–2430.

Kokhanovsky, A. A., 2004. Spectral reflectance of whitecaps. *Journal of Geophysical Research (Oceans)*, **109**, 5021.

Kuščer, I. and Ribarič, M., 1959. Matrix formalism in the theory of diffusion of light. *Opt. Acta*, **6**, 42–51.

Landgraf, J., Hasekamp, O., and Trautmann, T., 2002. Linearization of radiative transfer with respect to surface properties. *J. Quant. Spectrosc. Radiat. Transfer*, **72**, 327–339.

Landgraf, J., Hasekamp, O., Trautmann, T., and Box, M., 2001. A linearized radiative transfer model for ozone profile retrieval using the analytical forward-adjoint perturbation theory. *J. Geophys. Res.*, **106**, 27,291–27,306.

Landgraf, J., Hasekamp, O., van Deelen, R., and Aben, I., 2004. Rotational Raman scattering of polarized light in the Earth atmosphere: A vector radiative transfer model using the radiative transfer perturbation theory approach. *J. Quant. Spectrosc. Radiat. Transfer*, **87**, 399–433

Levenberg, K., 1944. A method for the solution of certain problems in least squares. *Quart. Appl. Math.*, **2**, 164–168.

Lewins, J., 1965. *Importance, the Adjoint Function*. Pergamon Press, Oxford, England.

Marchuk, G., 1964. Equation for the value of information from weather satellites and formulation of inverse problem s. *Cosmic Res.*, **2**, 394–409.

Marquardt, D., 1963. An algorithm for least squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, **11**, 431–441.

Mishchenko, M. and Travis, L., 1994. Light scattering by polydispersions of randomly oriented spheroids with sizes comparable to wavelengths of observation. *Appl. Opt.*, **33**, 7206–7225.

Mishchenko, M. I., Lacis, A. A., Carlson, B., and Travis, L., 1995. Nonsphericity of dust-like tropospheric aerosols: Implications for aerosol remote sensing and climate modeling. *Geophys. Res. Lett.*, **22**, 1077–1980.

Mishchenko, M. I. and Travis, L. D., 1997. Satellite retrieval of aerosol properties over the ocean using polarization as well as intensity of reflected sunlight. *J. Geophys. Res.*, **102**, 16,989–17,013.

Monahan, E. and O'Muircheartaigh, I., 1980. Optical power law description of oceanic whitecap coverage dependence on windspeed. *J. Phys. Oceanogr.*, **10**, 2094.

Morel, A., 1988. Optical modeling of the upper ocean in relation to its biogenous matter content (case I waters). *J. Geophys. Res.*, **93**, 10 749–10 768.

Morel, A. and Gentili, B., 1993. Diffuse reflectance of oceanic waters. II. Bidirectional effects. *Appl. Opt.*, **32**, 6864–6879.

Morel, A. and Maritorena, S., 2001. Bio-optical properties of oceanic waters: A reappraisal. *J. Geophys. Res.*, **106**, 7163–7180.

Morel, A. and Prieur, L., 1977. Ananlysis of variations in ocean color. *Limnol. Oceanogr.*, **19**, 591–600.

Morse, P. M. and Feshbach, H., 1953. *Methods of Theoretical Physics*. McGraw-Hill, New York.

Phillips, P., 1962. A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.*, **9**, 84–97.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B., 1992. *Numerical Recipes in FORTRAN, the Art of Scientific Computing*. Cambridge University Press, Cambridge.

Rodgers, C., 1976. Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Rev. Geophys.*, **14**, 609–624.

Rodgers, C., 2000. *Inverse Methods for Atmospheric Sounding: Theory and Practice*. World Sc., River Edge, NJ.

Rodgers, C. D. and Connor, B. J., 2003. Intercomparison of remote sounding instruments. *Journal of Geophysical Research (Atmospheres)*, **108**, 13.

Rozanov, V., Kurosu, T., and Burrows, J., 1998. Retrieval of atmospheric constituents in the UV-visible: A new quasi-analytical approach for the calculation of weighting functions. *J. Quant. Spectrosc. Radiat. Transfer*, **60**, 277–299.

Schulz, F. M., Stamnes, K., and Weng, F., 1999. VDISORT: An improved and generalized discrete ordinate method for polarized (vector) radiative transfer. *J. Quant. Spectrosc. Radiat. Transfer*, **61**, 105–122.

Sendra, C. and Box, M., 2000. Retrieval of the phase function and scattering optical thickness of aerosols: A radiative transfer perturbation theory application. *J. Quant. Spectrosc. Radiat. Transfer*, **64**, 499–515.

Smith, R. and Baker, K., 1981. Optical properties of the clearest natural waters. *Appl. Opt.*, **20**, 177–184.

Spurr, R., Kurosu, T., and Chance, K., 2001. A linearized discrete ordinate radiative transfer model for atmospheric remote-sensing retrieval. *J. Quant. Spectrosc. Radiat. Transfer*, **68**, 689–735.

Stammes, P., de Haan, J., and Hovenier, J., 1989. The polarized internal radiation field of a planetary atmosphere. *Astron. Astrophys.*, **225**, 239–259.

Stamnes, K., Tsay, S.-C., Wiscombe, W., and Jayaweera, K., 1988. Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media. *Appl. Optics*, **27**, 2502–2509.

Tikhonov, A., 1963. On the solution of incorrectly stated problems and a method of regularization. *Dokl. Akad. Nauk SSSR*, **151**, 501–504.

Torres, O., Bhartia, P. K., Herman, J. R., Ahmad, Z., and Gleason, J., 1998. Derivation of aerosol properties from satellite measurements of backscattered ultraviolet radiation: Theoretical basis. *J. Geophys. Res.*, **103**, 17 099–17 110.

Torres, O., Decae, R., Veefkind, P., and de Leeuw, G., 2001. OMI aerosol retrieval algorithm. *ATBD-OMI-03*, pages 47–69.

Ustinov, E. A., 1988. Methods of spherical harmonics: Application to the transfer of polarized radiation in a vertically non-uniform planetary atmosphere. Mathematical apparatus. *Cosmic Res.*, **26**, 473.

Ustinov, E. A., 1991. Inverse problem of photometric observation of solar radiation reflected by an optically dense planetary atmosphere. Mathematical methods and weighting functions of linearized inverse problem. *Cosmic Res.*, **29**, 519–532.

Ustinov, E. A., 1992. Inverse problem of the photometry of solar radiation reflected by an optically thick planetary atmosphere. 3. Remote sensing of minor gaseous constituents and an atmospheric aerosol. *Cosmic Res.*, **30**, 170–181.

Ustinov, E. A., 2001. Adjoint sensitivity analysis of radiative transfer equation: Temperature and gas mixing ratio weighting functions for remote sensing of scattering atmospheres in thermal IR. *J. Quant. Spectrosc. Radiat. Transfer*, **68**, 195–211.

van de Hulst, H. C., 1957. *Light Scattering by Small Particles*. John Wiley and Sons, New York.

Wiscombe, W. and Grams, G., 1988. Scattering from nonspherical Chebyshev particles, 2, Means of angular scattering patterns. *Appl. Opt.*, **27**, 2405–2421.

# 6 Derivatives of the radiation field and their application to the solution of inverse problems

V. V. Rozanov, A. V. Rozanov, A. A. Kokhanovsky

## 6.1 Introduction

Spectral distribution of the solar radiation traveling through the Earth's atmosphere contain an important information about numerous atmospheric and surface parameters. This information can be gained from the measured spectra employing so-called inverse theory and the problem to be solved thereby is usually referred to as the inverse problem. The first step to be done in the solution of any inverse problem is to formulate a model usually referred to as the forward model which will allow us to simulate the measured quantity assuming all relevant atmospheric and surface parameters to be known. Generally, in the case of the scattered, reflected, or transmitted solar radiance measured in the ultraviolet, visible, or near-infrared spectral range by means of satellite, airborne, or ground-based instruments, the corresponding forward model is nonlinear, i.e., there is no linear relationship between measured values of intensity and atmospheric parameters. However, the theoretical basis of the inverse problem solution is well investigated in the case of linear inverse problems only [19]. Thus, to make use of the existing numerical methods the forward model has to be linearized, i.e., a linear relationship between intensity of radiation and the atmospheric parameters has to be obtained. This can formally be done considering the intensity as a function or functional of the corresponding parameters and expanding it in the Taylor series with respect to the variations of the desired parameters. Then, in the linear approximation the partial or the variational derivatives of the intensity, usually referred to as the weighting functions (WFs), provide a linear relationship between the variation of the intensity and variations of parameters. Unfortunately, analytical expressions for the weighting functions can be derived in a few simple cases only whereas numerical calculations employing their mathematical definition in most practical situations are very computational-time-expensive.

In the previous issue of *Light Scattering Reviews* [24] we have presented a general approach to derive partial and variational derivatives based on the joint solution of the direct radiative transfer equation and the adjoint radiative transfer equation. Employing this approach we have derived the general expressions

for the partial and variational derivatives of the Stokes vector with respect to the main optical parameters such as the extinction coefficient and the single scattering albedo. However, most optical parameters depend not only on the atmospheric parameters (e.g., on concentrations of absorbing gases, pressure, temperature, aerosol and cloud particles number density) but also on the wavelength. Therefore, in multispectral inverse problems these are the atmospheric parameters which are commonly used as quantities to be retrieved rather than the optical parameters. For this reason we will derive here analytical expressions for WFs for the different atmospheric parameters which can be directly used to solve practical inverse problems. In particular, we will obtain WFs for the aerosol particles number density, for the atmospheric pressure and temperature, for the liquid water content and the effective radius of cloud droplets, as well as for geometrical cloud parameters such as cloud top height and geometrical cloud thickness. Furthermore, we will demonstrate that WFs for all considered atmospheric parameters can be obtained as a linear combination of WFs for main optical parameters.

Taking into account that for the solution of most inverse problems the scalar form of the radiative transfer equation is sufficient, we can simplify our treatment considering only variations of the intensity of radiation and neglecting other components of the Stokes vector, i.e., polarization effects. All expressions for WFs obtained here for the scalar case are implemented in the software package SCIATRAN 2.0 [22] and its successor version SCIATRAN 2.1 (see www.iup.uni-bremen.de/sciatran for further information). All numerical examples demonstrated throughout this chapter were completed employing the software package SCIATRAN 2.1.

The layout of the chapter is as follows. In section 6.2 we consider a relationship between the partial and variational derivatives of the intensity of radiation and the weighting functions. Section 6.3 gives the review the basics of the direct and adjoint radiative transfer equations which are used in the following sections for the derivation of WFs. In section 6.4 we derive general expressions for WFs. Section 6.5 presents expressions for WFs of the absorption and scattering coefficients. In section 6.6 we extend the obtained expressions to the case of a mixture of scattering and absorbing components. Numerical examples of WFs for the aerosol and cloud scattering coefficients are shown in section 6.7. In section 6.8 we derive WFs for the pressure and temperature and show some numerical examples in the ultraviolet and near infrared spectral ranges. Section 6.9 is aimed to the derivation of WFs for cloud parameters such as particle number density, liquid water content, and effective radius of water droplets or ice crystals. Some illustrations of these WFs are presented subsequently in section 6.10. In section 6.11 we introduce a modification for the weighting functions for the effective radius of water droples or ice crystals making them more suitable for the solution of the corresponding inverse problems. WFs for the geometrical cloud parameters such as cloud top and bottom height are presented in section 6.12. The derivation of WFs for main optical parameters such as the extinction coefficient and the single scattering albedo is presented in the Appendix.

## 6.2 Derivatives of the intensity and weighting functions

The intensity of the radiation field in a vertically inhomogeneous medium depends not only on the optical depth and the viewing direction but also on the vertical distribution of such parameters as, for example, the extinction coefficient and the single scattering albedo. This set of parameters depending on a vertical coordinate will be referred to as parameter-functions. In the case of the reflecting and emitting underlying surface the intensity depends also on the surface albedo, surface emissivity, etc. which will be referred to as scalar parameters. From the mathematical point of view the intensity can be considered as a functional of parameter-functions and a function of scalar parameters. A linear relationship between variations of these parameters and the variation of the intensity can be obtained employing the Taylor series expansion of the intensity. Restricting ourselves at this point with one parameter-function and one scalar parameter only which will be denoted as $p(\tau)$ and $c$, respectively, we have in the linear approximation:

$$I'(\tau,\Omega) = I(\tau,\Omega) + \frac{\partial I(\tau,\Omega)}{\partial c}\,\Delta c + \int_0^{\tau_0} \frac{\delta I(\tau,\Omega)}{\delta p(\tau')}\,\delta p(\tau')\,\mathrm{d}\tau'\,. \tag{6.1}$$

Here, $\tau \in [0,\tau_0]$ is the optical depth, $\tau_0$ is the optical thickness of the medium, $\Omega := \{\mu,\phi\}$ describes the set of variables $\mu \in [-1,1]$ and $\phi \in [0,2\pi]$, $\mu$ and $\phi$ are the cosine of the polar angle and the azimuthal angle, respectively, the perturbed intensity, $I'(\tau,\Omega)$, corresponds to perturbed parameters $c' = c + \Delta c$ and $p'(\tau) = p(\tau) + \delta p(\tau)$,

$$\frac{\partial I(\tau,\Omega)}{\partial c} = \lim_{\Delta c\to 0} \frac{I(\tau,\Omega;p(\tau),c') - I(\tau,\Omega;p(\tau),c)}{\Delta c} \tag{6.2}$$

is the partial derivative of the intensity with respect to the scalar parameter $c$ and

$$\frac{\delta I(\tau,\Omega)}{\delta p(\tau')} = \lim_{\Delta\tau'\to 0} \frac{I(\tau,\Omega;p'(\tau),c) - I(\tau,\Omega;p(\tau),c)}{\int_{(\Delta\tau')}\delta p(x)\,\mathrm{d}x} \tag{6.3}$$

is the variational derivative of the intensity with respect to the parameter-function $p(\tau)$ at the optical depth $\tau' \in [0,\tau_0]$. The integration of the perturbation in the denominator of this expression is carried out over a small range $\Delta\tau'$ around the point $\tau'$ where $\delta p(x)$ is non-zero. A complete discussion of functionals and variational derivatives is presented among others by Volterra [33]. The partial derivative at a given optical depth $\tau$ and viewing direction $\Omega$ given by (6.2) will be considered as a function of the wavelength, whereas the variational derivative given by (6.3) as a function of the vertical coordinate $\tau'$ and the wavelength. The explicit notation of the dependence of all relevant functions on the wavelength will be omitted.

Introducing the variation of the intensity as

$$\delta I(\tau,\Omega) = I'(\tau,\Omega) - I(\tau,\Omega)\,, \tag{6.4}$$

we rewrite (6.1) in the following form:

$$\delta I(\tau,\Omega) = \mathcal{V}_c(\tau,\Omega)\Delta c + \int_0^{\tau_0} \mathcal{V}_p(\tau';\tau,\Omega)\delta p(\tau')\,\mathrm{d}\tau'\,, \tag{6.5}$$

where the variational and partial derivatives of the intensity are denoted as $\mathcal{V}_p(\tau';\tau,\Omega)$ and $\mathcal{V}_c(\tau,\Omega)$, respectively.

Equation (6.5) provides a linear relationship between variations of the atmospheric and surface parameters and the variation of the intensity at a given optical depth $\tau$ and viewing direction $\Omega$. Functions $\mathcal{V}_p(\tau';\tau,\Omega)$ and $\mathcal{V}_c(\tau,\Omega)$ describe the contribution of the variation of a certain parameter to the variation of the intensity. These functions which are usually referred to as the weighting functions (WFs) can be defined for both absolute and relative variations of parameters. Introducing the relative variation of a scalar parameter $c$ as $v_c = \Delta c/c$ and the relative variation of a parameter-function $p(\tau)$ as $v_p(\tau) = \delta p(\tau)/p(\tau)$, Eq. (6.5) can be rewritten as follows:

$$\delta I(\tau,\Omega) = \mathcal{W}_c(\tau,\Omega)v_c + \int_0^{\tau_0} \mathcal{W}_p(\tau';\tau,\Omega)v_p(\tau')\,\mathrm{d}\tau'\,. \tag{6.6}$$

The relationship between weighting functions for the relative and absolute variations of parameters is obvious:

$$\mathcal{W}_c(\tau,\Omega) = c\mathcal{V}_c(\tau,\Omega)\,, \quad \mathcal{W}_p(\tau';\tau,\Omega) = p(\tau')\mathcal{V}_p(\tau';\tau,\Omega)\,. \tag{6.7}$$

The obtained linear relationship between the variation of the intensity and variations of parameters can be directly used to solve various inverse problems. To demonstrate this let us consider the simplest case of one scalar parameter to be retrieved. The perturbed intensity, $I'(\tau,\Omega)$, should be considered in this case as the measured value at the optical depth $\tau$ and viewing direction $\Omega$ corresponding to an unknown value of the scalar parameter, $c'$, whereas the intensity, $I(\tau,\Omega)$, is calculated employing an appropriate radiative transfer model for a known value of scalar parameter, $c$. Taking into account that in the case under consideration (6.5) contains only one unknown parameter, $\Delta c$, we obtain the following estimation for this scalar parameter:

$$\Delta c = \frac{\delta I(\tau,\Omega)}{\mathcal{V}_c(\tau,\Omega)} \quad \longrightarrow \quad c' = c + \frac{\delta I(\tau,\Omega)}{\mathcal{V}_c(\tau,\Omega)}\,, \tag{6.8}$$

where, according to its definition, $\Delta c = c' - c$.

Thus, the considered example demonstrates that the solution of a certain inverse problem requires not only the measured and simulated values of intensities but also an appropriate weighting function. However, a derivation of the weighting functions employing their definition according to Eqs (6.2) and (6.3) is possible in a few simplest cases only. Therefore, in the following sections we present an approach to derive weighting functions which allow us to avoid using Eqs (6.2) and (6.3). In particular, we consider how the weighting functions can be obtained employing the joint solution of the linearized radiative transfer equation (RTE) and the corresponding adjoint RTE.

## 6.3 Basic formulation of the direct and adjoint radiative transfer equations in the operator form

The derivation of the weighting functions presented below requires the formulation of the direct and adjoint radiative transfer equations in the so-called generalized form [25]. This representation as well as the adjoint approach used to derive analytical expressions for WFs will be briefly discussed below.

We start from the standard RTE for a plane-parallel scattering, absorbing, and emitting atmosphere illuminated by the incident solar radiation with the zenith angle $\theta_0$. The irradiance flux is considered to be equal to $\pi\mu_0$ at the top of the medium, where $\mu_0 = \cos\theta_0$. The direct RTE can be written in the following form (see e.g. [4, 11, 12, 27] for derivation):

$$\mu \frac{\mathrm{d}I(\tau,\Omega)}{\mathrm{d}\tau} + I(\tau,\Omega) = J(\tau,\Omega) + (1-\omega(\tau))B(\tau)\,, \tag{6.9}$$

and appropriate boundary conditions are given by:

$$I(0,\Omega) \;=\; \pi\delta(\Omega-\Omega_0),\;\; \mu>0\,, \tag{6.10}$$

$$I(\tau_0,\Omega) \;=\; \frac{A}{\pi}\int_{\Omega_+}\rho(\Omega,\Omega')I(\tau_0,\Omega')\mu'\,\mathrm{d}\Omega' + \epsilon B(T_s),\;\; \mu<0\,, \tag{6.11}$$

where $J(\tau,\Omega)$ is the multiple scattering source function:

$$J(\tau,\Omega) = \frac{\omega(\tau)}{4\pi}\int_{4\pi} p(\tau,\Omega,\Omega')I(\tau,\Omega')\,\mathrm{d}\Omega'\,, \tag{6.12}$$

$\tau \in [0,\tau_0]$ is the optical depth, $\tau_0$ is the optical thickness of the medium, $\omega(\tau) \in [0,1]$ is the single scattering albedo, $\mu \in [-1,1]$ is the cosine of the polar angle as measured from the positive $\tau$-axis (i.e., negative values of $\mu$ correspond to the light propagated upwards), $\phi \in [0,2\pi]$ is the azimuthal angle, the variable $\Omega := \{\mu,\phi\}$ describes a set of variables $\mu \in [-1,1]$ and $\phi \in [0,2\pi]$, the variable $\Omega_+ := \{\mu,\phi\}$ describes the set of variables $\mu \in [0,1]$ and $\phi \in [0,2\pi]$, $\delta(\Omega-\Omega_0) = \delta(\mu-\mu_0)\delta(\phi-\phi_0)$ is the Dirac $\delta$-function, $B(\tau)$ is the Planck function depending on the kinetic temperature of the medium, $A$ is the spherical albedo of the underlying surface, $\rho(\Omega,\Omega')$ is a function determining the angular reflection properties of the boundary surface, $\epsilon$ is the surface emissivity, $B(T_s)$ is the Planck function at the surface temperature $T_s$, $p(\tau,\Omega,\Omega')$ is the phase function.

It is convenient to rewrite (6.9) and the corresponding boundary conditions given by (6.10) and (6.11) in the operator form. Let us define a linear differential-integral operator, $L_e$, which comprises all operations on the intensity $I(\tau,\Omega)$ in (6.9) as follows:

$$L_e = \mu\frac{d}{d\tau} + 1 - \frac{\omega(\tau)}{4\pi}\int_{4\pi}\mathrm{d}\Omega' p(\tau,\Omega,\Omega')\otimes\,, \tag{6.13}$$

where symbol $\otimes$ is used to denote an integral operator rather than a finite integral. The operator $L_e$ is referred to as the direct radiative transfer operator. The radiative transfer equation is now written in the following operator form:

$$L_e\, I = S_e\,, \tag{6.14}$$

where

$$S_e \equiv S_e(\tau, \Omega) = [1 - \omega(\tau)]B(\tau)\,. \tag{6.15}$$

Although in the considered RTE the internal emission source function, $S_e(\tau, \Omega)$, is isotropic the argument $\Omega$ will be retained throughout this chapter for generality. To rewrite (6.10) and (6.11) in the operator form as well, we define two linear integral operators $L_t$ and $L_b$ as follows:

$$L_t \;=\; \int_0^{\tau_0} \mathrm{d}\tau \delta(\tau) \otimes\,, \tag{6.16}$$

$$L_b \;=\; \int_0^{\tau_0} \mathrm{d}\tau \delta(\tau - \tau_0) \left( \otimes - \frac{A}{\pi} \int_{4\pi} \mathrm{d}\Omega' \lambda(\mu') \rho(\Omega, \Omega') \otimes \right)\,. \tag{6.17}$$

Here, $\delta(\tau)$ and $\delta(\tau - \tau_0)$ are the Dirac $\delta$-functions and $\lambda(\mu)$ is an auxiliary function introduced following [30] as $\lambda(\mu) = \mu\Theta(\mu)$, where $\Theta(\mu)$ is the Heaviside step-function over $\mu \in [-1, 1]$ given by

$$\Theta(\mu) = \begin{cases} 1, & \mu > 0 \\ 0, & \mu < 0 \end{cases}\,. \tag{6.18}$$

Operators $L_t$ and $L_b$ operate analogically to the operator $L_e$ on the intensity $I(\tau, \Omega)$ and have the same domain. Thus, the operator form of the direct RTE alone with the boundary conditions is written as follows:

$$L_e I \;=\; S_e\,, \tag{6.19}$$

$$L_t I \;=\; S_t(\Omega),\ \mu > 0\,, \tag{6.20}$$

$$L_b I \;=\; S_b(\Omega),\ \mu < 0\,, \tag{6.21}$$

where according to (6.10) and (6.11)

$$S_t(\Omega) \;=\; \pi\delta(\Omega - \Omega_0)\,, \tag{6.22}$$

$$S_b(\Omega) \;=\; \epsilon B(T_s)\,. \tag{6.23}$$

The formulated direct RTE can be used to simulate the radiation field in a plane-parallel atmosphere in a wide spectral range from the ultraviolet to the thermal infrared. The polarization can be easily accounted for (see e.g. [6]).

### 6.3.1 Generalized form of the direct radiative transfer equation

The operator representation of the direct RTE and the corresponding boundary conditions formulated above describe a specific boundary value problem consisting of three independent operator equations. This boundary value problem can also be rewritten in the form of a single operator equation. Such representation is called the generalized form of the direct RTE. The generalized form of the direct RTE has been suggested by Ustinov [30] and its rigorous derivation has

been given by Rozanov and Rozanov [25] in the scalar case and by Rozanov [24] in the vector case (i.e., including polarization).

Following [25] the generalized form of the direct RTE is written as

$$LI = S\,. \tag{6.24}$$

Here, the generalized form of the direct radiative transfer operator, $L$, and the right-hand side, $S(\tau,\Omega)$, are determined as follows:

$$L = L_e + \psi_t(\tau,\mu)L_t + \psi_b(\tau,-\mu)L_b\,, \tag{6.25}$$

$$S(\tau,\Omega) = S_e(\tau,\Omega) + \psi_t(\tau,\mu)S_t(\Omega) + \psi_b(\tau,-\mu)S_b(\Omega)\,, \tag{6.26}$$

where auxiliary functions $\psi_t(\tau,\mu)$ and $\psi_b(\tau,-\mu)$ are given by

$$\psi_t(\tau,\mu) = \mu\delta(\tau)\Theta(\mu)\,, \tag{6.27}$$

$$\psi_b(\tau,-\mu) = -\mu\delta(\tau-\tau_0)\Theta(-\mu)\,. \tag{6.28}$$

The derived radiative transfer equation is equivalent to Eqs (6.19)–(6.21) but incorporate all operations with respect to the intensity at the boundaries, i.e., at this point, the boundary conditions are already included in the radiative transfer equation and do not need to be specified separately. This equation is the desired generalized form of the direct RTE containing all operations on the intensity field including boundary conditions.

### 6.3.2 Adjoint radiative transfer operator

Let $B$ be a linear operator operating on a function $I(\tau,\Omega)$. Then according to its definition [9], the corresponding adjoint operator $B^*$ has to satisfy the following identity:

$$(I^*, BI) = (B^*I^*, I)\,, \tag{6.29}$$

where the notation ( , ) denotes the scalar product in the appropriate functional space and $I^*(\tau,\Omega)$ is an arbitrary function which belongs to the domain of the operator $B^*$. Throughout this chapter we will assume that the scalar product of two arbitrary functions $f(\tau,\Omega)$ and $g(\tau,\Omega)$ is defined as follows:

$$(f,g) = \int_0^{\tau_0}\int_{4\pi} f(\tau,\Omega)g(\tau,\Omega)\,\mathrm{d}\tau\,\mathrm{d}\Omega\,. \tag{6.30}$$

Since the derivation of the adjoint radiative transfer operator using different approaches has already been described elsewhere [1,14,18,25,30], we will present just the final result here. Following [25], we write the expression for the adjoint radiative transfer operator in the following form:

$$L^* = L_e^* + \psi_t(\tau,-\mu)L_t^* + \psi_b(\tau,\mu)L_b^*\,, \tag{6.31}$$

where operators $L_e^*$, $L_t^*$ and $L_b^*$ are

$$L_e^* = -\mu\frac{d}{d\tau} + 1 - \frac{\omega(\tau)}{4\pi}\int_{4\pi} \mathrm{d}\Omega' p(\tau, \Omega', \Omega)\otimes\,, \tag{6.32}$$

$$L_t^* = \int_0^{\tau_0} \mathrm{d}\tau\delta(\tau)\otimes\,, \tag{6.33}$$

$$L_b^* = \int_0^{\tau_0} \mathrm{d}\tau\delta(\tau-\tau_0)\left[\otimes - \frac{A}{\pi}\int_{4\pi} \mathrm{d}\Omega'\rho(\Omega', \Omega)\lambda(-\mu')\otimes\right]\,. \tag{6.34}$$

Equation (6.31) is the desired generalized form of the adjoint radiative transfer operator. The adjoint operator, $L^*$, contains analogically to the direct operator, $L$, all operations on the adjoint intensity including boundary conditions.

### 6.3.3 Adjoint approach and the adjoint radiative transfer equation

In the previous subsection the generalized form of the adjoint radiative transfer operator has been presented. Here we will demonstrate how the adjoint radiative transfer equation can be formulated. We start from the generalized form of the direct RTE according to (6.24):

$$LI = S\,, \tag{6.35}$$

where $S$ is given by (6.26). Assume we need to calculate a functional, say $\Phi$, of the intensity $I$

$$\Phi = (W, I)\,, \tag{6.36}$$

where $W$ is an arbitrary function of variables $\tau$ and $\Omega$. There are two ways to solve this problem. One is to find the solution $I$ of the direct RTE and apply (6.36) to calculate $\Phi$ (forward approach) and the other (adjoint approach) is to calculate the scalar product of both sides of (6.35) and an arbitrary function $I^*$ and rewrite the left-hand side of the resulting equation using the definition of the adjoint operator, Eq. (6.29), as follows:

$$(I^*, LI) = (L^*I^*, I) = (I^*, S)\,. \tag{6.37}$$

Then, we obtain

$$(I^*, S) = (L^*I^*, I)\,. \tag{6.38}$$

If we require $I^*$ to be the solution of the following adjoint equation:

$$L^*I^* = W\,, \tag{6.39}$$

equation (6.38) results in

$$(I^*, S) = (W, I) = \Phi\,. \tag{6.40}$$

Thus, the functional $\Phi$ can be found also as the scalar product of the solution of the adjoint RTE, $I^*$, and the right-hand side of the direct RTE written in the generalized form, $S$.

Similar to the forward boundary value problem defined by (6.35), Eq. (6.39) describes a boundary value problem for the adjoint intensity written in the generalized form. Operator $L^*$ given by (6.31) includes the boundary conditions operators similar to the operator $L$ given by (6.25). Thereby, similar to the right-hand side of the direct RTE, the right-hand side of the adjoint RTE given by (6.39) has to contain the boundary conditions as well. Thus, it should be possible to rewrite $W(\tau, \mu)$ in the following form:

$$W(\tau, \Omega) = W_e(\tau, \Omega) + \psi_t(\tau, -\mu) W_t(\Omega) + \psi_b(\tau, \mu) W_b(\Omega)\,, \tag{6.41}$$

where subscripts 'e', 't', and 'b' denote 'equation', upper ('top'), and lower ('bottom') boundary conditions, respectively. Substituting (6.41) into (6.39) we obtain the generalized form of the adjoint radiative transfer equation as follows:

$$L^* I^* = W_e + \psi_t(\tau, -\mu) W_t(\Omega) + \psi_b(\tau, \mu) W_b(\Omega)\,. \tag{6.42}$$

This equation can be treated in analogy to the generalized form of the direct RTE as a sum of three independent operator equations, namely, operator form of the adjoint RTE and boundary condition equations written in the operator form premultiplied by appropriate functions. Therefore, similar to the boundary value problem for the direct RTE, Eq. (6.42) can be separated into three equations as follows:

$$L_e^* I^* = W_e(\tau, \Omega)\,, \tag{6.43}$$

$$L_t^* I^* = W_t(\Omega)\,,\ \mu < 0\,, \tag{6.44}$$

$$L_b^* I^* = W_b(\Omega)\,,\ \mu > 0\,. \tag{6.45}$$

Equations (6.43)–(6.45) are referred to as the operator representation of the adjoint RTE. Substituting in these equations operators $L_t^*$, $L_b^*$ and $L_e^*$ as given by (6.33), (6.34) and (6.32), respectively, the standard form of the adjoint RTE can be formulated:

$$-\mu \frac{\mathrm{d}I^*(\tau, \Omega)}{\mathrm{d}\tau} + I^*(\tau, \Omega) = \frac{\omega}{4\pi} \int_{4\pi} p(\tau, \Omega', \Omega) I^*(\tau, \Omega')\, \mathrm{d}\Omega' + W_e(\tau, \Omega)\,, \tag{6.46}$$

$$I^*(0, \Omega) = W_t(\Omega),\ \mu < 0\,, \tag{6.47}$$

$$I^*(\tau_0, \mu) = W_b(\Omega) - \frac{A}{\pi} \int_{\Omega_-} \rho(\Omega', \Omega) I^*(\tau_0, \Omega') \mu'\, \mathrm{d}\Omega',\ \mu > 0\,. \tag{6.48}$$

We note that in contrast to the forward intensity the boundary conditions for the adjoint intensity are given for the outgoing adjoint radiation.

As clearly seen the adjoint and the direct RTEs are closely related to each other. Moreover, as demonstrated by many authors [1, 2, 15–17, 25] the solution of the adjoint RTE given by (6.46) can be found employing standard numerical methods developed for direct RTEs. Thereby, in fact, we do not need to develop special methods to solve the adjoint RTE.

## 6.4 General expressions for weighting functions

For the solution of most practical inverse problems a linear relationship between the measured functional of the radiative field and the atmospheric parameters is required. To obtain such a relationship let us formulate the measured functional as follows:

$$\Phi(\tau_v, \Omega_v) = (R, I) \equiv \int_0^{\tau_0} \int_{4\pi} R(\tau_v, \Omega_v; \tau, \Omega) I(\tau, \Omega) \, \mathrm{d}\tau \, \mathrm{d}\Omega \,, \tag{6.49}$$

where $R(\tau_v, \Omega_v; \tau, \Omega)$ is an instrument response function and a short notation for the scalar product as ( , ) is used. The specific form of the response function is of minor importance for our theoretical consideration. Therefore, following [30], we introduce the instrument response function appropriate to the instrument with an infinitesimally small field of view placed inside or at a boundary of a medium as follows:

$$R(\tau_v, \Omega_v; \tau, \Omega) = \delta(\tau - \tau_v)\delta(\Omega - \Omega_v) \,, \tag{6.50}$$

where $\delta(\Omega - \Omega_v) = \delta(\mu - \mu_v)\delta(\phi - \phi_v)$. The measured value, $\Phi$, given by (6.49) is in this case the intensity of the radiation field at the optical depth $\tau = \tau_v$ in the direction $\Omega_v \equiv [\mu_v, \phi_v]$ characterized by the cosine of the zenith angle, $\mu_v$, and azimuthal angle, $\phi_v$.

Taking into account that the response function is independent of both atmospheric and surface parameters, the variation of the measured functional $\delta\Phi$ can be expressed as follows:

$$\delta\Phi(\tau_v, \Omega_v) = \Phi'(\tau_v, \Omega_v) - \Phi(\tau_v, \Omega_v) = (R, \delta I) \,, \tag{6.51}$$

where $\Phi'(\tau_v, \Omega_v)$ corresponds to the set of perturbed parameters. To express the scalar product in the right-hand side of (6.51) via the variation of the atmospheric and surface parameters we employ the adjoint approach. As discussed in section 6.3.3 the scalar product of the known function, $R$, and the function, $\delta I$, can be expressed as the scalar product of the adjoint intensity, $I^*$, and the right-hand side of the RTE describing the variation of the intensity, $\delta I$. To derive the corresponding RTE we rewrite following Marchuk [14] the direct RTE given by (6.24) for the perturbed values of the operator $L'$, intensity $I'$, and right-hand side $S'$ :

$$L'I' = S' \,. \tag{6.52}$$

Here, the primed source function and the primed operator correspond in analogy to the primed intensity to the set of perturbed parameters. Substituting into this equation $I' = I + \delta I$, $L' = L + \delta L$ and $S' = S + \delta S$, we obtain

$$(L + \delta L)(I + \delta I) = S + \delta S \implies L\delta I + \delta L I + \delta L \delta I = \delta S \,. \tag{6.53}$$

In the linear approximation we can neglect the term containing the product $\delta L \delta I$. Therefore, the RTE describing the variation of intensity can be formulated in the following form:

$$L\delta I = \delta S - \delta L I \,. \tag{6.54}$$

Varying the operator $L$ and the function $S$ given by (6.25) and (6.26), respectively, and taking into account that the operator $L_t$ and the function $S_t$ as given by (6.16) and (6.22), respectively, are independent of the atmospheric and surface parameters, and therefore $\delta S_t = 0$ and $\delta L_t = 0$, we obtain the following expression for the right-hand side of (6.54):

$$\delta S - \delta L I = \sum_{p=1}^{P} \Big[ (\delta_p S_e - \delta_p L_e I) + \psi_b(\tau, -\mu)(\delta_p S_b - \delta_p L_b I) \Big] , \tag{6.55}$$

where $P$ is the full number of parameters and $\delta_p$ denotes that the variation of the corresponding function or operator is caused by the variation of the parameter $p$ only.

Variations of functions $S_e$ and $S_b$ as well as of operators $L_e$ and $L_b$ in the right-hand side of (6.55) can easily be expressed via the variation of the corresponding parameter $p_p(\tau)$ expanding them in the Taylor series with respect to this parameter and restricting with the linear term relative to $\delta p_p(\tau)$. For example, for the function $S_e$ we obtain

$$\delta_p S_e(\tau, \Omega) = S'_e(\tau, \Omega) - S_e(\tau, \Omega) = \frac{\partial S_e(\tau, \Omega)}{\partial p_p(\tau)} \delta p_p(\tau) , \tag{6.56}$$

where $S'_e(\tau, \Omega)$ is the source function for the perturbed parameter $p'_p(\tau)$ and $\partial S_e(\tau, \Omega)/\partial p_p(\tau)$ is the partial derivative of the function $S_e(\tau, \Omega)$ with respect to the parameter $p_p(\tau)$.

Employing Eq. (6.56) and introducing the relative variation of the corresponding parameter as $v_p(\tau) = \delta p_p(\tau)/p_p(\tau)$, the terms in the right-hand side of (6.55) can be rewritten in the linear approximation as follows:

$$\delta_p S_e - \delta_p L_e I = \left[ \frac{\partial S_e(\tau, \Omega)}{\partial p_p(\tau)} - \frac{\partial L_e}{\partial p_p(\tau)} I \right] \delta p_p(\tau) = v_p(\tau) Y_p(\tau, \Omega) , \tag{6.57}$$

$$\delta_p S_b - \delta_p L_b I = \left[ \frac{\partial S_b(\Omega)}{\partial p_p(\tau)} - \frac{\partial L_b}{\partial p_p(\tau)} I \right] \delta p_p(\tau) = v_p(\tau) G_p(\Omega) , \tag{6.58}$$

where functions $Y_p(\tau, \Omega)$ and $G_p(\Omega)$ contain partial derivatives of the appropriate functions and operators:

$$Y_p(\tau, \Omega) = p_p(\tau) \left[ \frac{\partial S_e(\tau, \Omega)}{\partial p_p(\tau)} - \frac{\partial L_e}{\partial p_p(\tau)} I \right] , \tag{6.59}$$

$$G_p(\Omega) = p_p(\tau) \left[ \frac{\partial S_b(\Omega)}{\partial p_p(\tau)} - \frac{\partial L_b}{\partial p_p(\tau)} I \right] . \tag{6.60}$$

Introducing an auxiliary function $\Psi_p(\tau, \Omega)$ as

$$\Psi_p(\tau, \Omega) = Y_p(\tau, \Omega) + \psi_b(\tau, -\mu) G_p(\Omega) \tag{6.61}$$

and substituting it into the right-hand side of (6.55), we obtain

$$\delta S - \delta LI = \sum_{p=1}^{P} v_p(\tau)\Psi_p(\tau,\Omega)\,. \tag{6.62}$$

We note that for a scalar parameter we only need to replace $v_p(\tau)$ by $v_c$, where $v_c = \Delta c/c$ is the relative variation of this parameter. Therefore, this case does not need to be considered separately. Substituting further (6.62) into the right-hand side of (6.54), we have

$$L\delta I = \sum_{p=1}^{P} v_p(\tau)\Psi_p(\tau,\Omega)\,. \tag{6.63}$$

Equation (6.63) provides a linear relationship between the variation of the intensity and the relative variations of the atmospheric and surface parameters. This equation will be referred to as the generalized form of the linearized direct radiative transfer equation (LRTE).

Multiplying both sides of the LRTE given by (6.63) by an arbitrary function $I^*(\tau,\Omega)$ and using the definition of the adjoint operator as given by (6.29), we obtain

$$\left(I^*, L\delta I\right) = \left(L^*I^*, \delta I\right) = \left(I^*, \sum_{p=1}^{P} v_p(\tau)\Psi_p\right). \tag{6.64}$$

Let us require now that the function $I^*(\tau,\Omega)$ is the solution of the following linear operator equation:

$$L^*I^* = R\,, \tag{6.65}$$

where $R$ is the response function given by (6.50). Then, substituting the response function $R$ instead of $L^*I^*$ into (6.64), we obtain the expression for the variation of the measured functional $\delta\Phi(\tau_v,\Omega_v) = (R,\delta I)$ as the scalar product of the adjoint intensity, $I^*$, and the right-hand side of LRTE given by (6.63):

$$\delta\Phi(\tau_v,\Omega_v) = (R,\delta I) = \left(I^*, \sum_{p=1}^{P} v_p(\tau)\Psi_p\right). \tag{6.66}$$

Thus, we have obtained the desired linear relationship between the variation of the measured functional and variations of the atmospheric and surface parameters. We note that the adjoint intensity, $I^*$, as a solution of (6.65) is a function of variables $\tau$ and $\Omega$ and depends parametrically on variables $\tau_v$ and $\Omega_v$ describing the position and direction of observation, i.e., $I^* \equiv I^*(\tau,\Omega;\tau_v,\Omega_v)$. However, for simplicity the explicit notation of the dependence of the adjoint intensity on the observation position, $\tau_v$, will be omitted in the following discussion.

Introducing a short notation for the integral of the product of two arbitrary functions $f(\tau,\Omega)$ and $g(\tau,\Omega)$ over $\Omega$ as

$$\int_{4\pi} f(\tau,\Omega)g(\tau,\Omega)\,\mathrm{d}\Omega \equiv \langle fg\rangle\,, \tag{6.67}$$

and taking into account that the scalar product is defined according to (6.30), the variation of the measured functional $\delta\Phi$ as given by (6.66) can be rewritten as follows:

$$\delta\Phi(\tau_v, \Omega_v) = \sum_{p=1}^{P} \int_0^{\tau_0} \langle \Psi_p I^* \rangle \, v_p(\tau) \, \mathrm{d}\tau \,. \tag{6.68}$$

Thus, in the case of one scalar parameter, $c$, and one parameter-function, $p(\tau)$, we obtain

$$\delta\Phi(\tau_v, \Omega_v) = v_c \int_0^{\tau_0} \langle \Psi_c I^* \rangle \, \mathrm{d}\tau + \int_0^{\tau_0} \langle \Psi_p I^* \rangle \, v_p(\tau) \, \mathrm{d}\tau \,. \tag{6.69}$$

Comparing this equation to (6.6), we conclude that the weighting functions for a scalar parameter and for a parameter-function can be written as

$$\mathcal{W}_c(\tau_v, \Omega_v) = \int_0^{\tau_0} \left\langle \Psi_c I^* \right\rangle \mathrm{d}\tau \equiv \int_0^{\tau_0} \int_{4\pi} \Psi_c(\tau, \Omega) I^*(\tau, \Omega; \Omega_v) \, \mathrm{d}\Omega \, \mathrm{d}\tau \tag{6.70}$$

and

$$\mathcal{W}_p(\tau; \tau_v, \Omega_v) = \left\langle \Psi_p I^* \right\rangle \equiv \int_{4\pi} \Psi_p(\tau, \Omega) I^*(\tau, \Omega; \Omega_v) \, \mathrm{d}\Omega \,, \tag{6.71}$$

respectively. Here, functions $\Psi_c(\tau, \Omega)$ and $\Psi_p(\tau, \Omega)$ are given by (6.61) and the adjoint intensity, $I^*(\tau, \Omega; \Omega_v)$, is the solution of the adjoint RTE given by (6.65).

The derived WFs for the parameter-functions provide a linear relationship between the variation of the intensity at a given position and direction and the variation of a certain optical parameter at any point inside the medium. Equatioins (6.70) and (6.71) signify the importance of the adjoint intensity, $I^*(\tau, \Omega)$, for the calculations of the weighting functions and, therefore, for the solution of inverse problems.

Although Eqs (6.59)–(6.61) can be used to obtain analytical expressions for the auxiliary functions $\Psi_p(\tau, \Omega)$ and, thus, for the corresponding weighting functions, for any desired parameter, in practice only the analytical expressions for the weighting functions with respect to the directly involved in the formulated RTE parameters are required whereas WFs for all other parameters can be obtained as a linear combination of the weighting functions for these main parameters. In the case under consideration main parameter-functions are the extinction coefficient, single scattering albedo and kinetic temperature of the medium, and the scalar parameters are the surface albedo, surface emissivity and surface temperature, whereas secondary parameter-functions whose weighting functions can be obtained as a linear combination of the main parameter WFs mentioned above are, for example, the scattering and absorption coefficients. For this reason, only the analytical expressions for auxiliary functions corresponding to the main parameters are presented in the Appendix.

## 6.5 Weighting functions for absorption and scattering coefficients

The expressions for the extinction coefficient and the single scattering albedo weighting functions derived in the Appendix (see Eqs (6.240) and (6.244), respectively) allow us to formulate the weighting functions for the scattering and absorption coefficients as well. Indeed, assuming that both extinction coefficient and single scattering albedo are varying simultaneously, we can write

$$\delta\phi(\tau;\tau_v,\Omega_v) = \mathcal{W}_e(\tau;\tau_v,\Omega_v)v_e(\tau) + \mathcal{W}_\omega(\tau;\tau_v,\Omega_v)v_\omega(\tau)\,, \tag{6.72}$$

where $\delta\phi(\tau;\tau_v,\Omega_v)\,\mathrm{d}\tau$ can be considered as the contribution of the extinction coefficient and single scattering albedo variations within an infinitesimal layer positioned at the optical depth $\tau$, into the variation of the measured functional, $\delta\Phi(\tau_v,\Omega_v)$, i.e.,

$$\delta\Phi(\tau_v,\Omega_v) = \int_0^{\tau_0} \delta\phi(\tau;\tau_v,\Omega_v)\,\mathrm{d}\tau\,. \tag{6.73}$$

The relative variations $v_e(\tau)$ and $v_\omega(\tau)$ can be caused by variations of the absorption and/or scattering coefficients, namely

$$v_e(\tau) = v_s(\tau)\omega(\tau) + v_a(\tau)[1-\omega(\tau)]\,, \tag{6.74}$$

$$v_\omega(\tau) = [v_s(\tau) - v_a(\tau)][1-\omega(\tau)]\,, \tag{6.75}$$

where $\omega(\tau)$ is the single scattering albedo and the relative variations of the scattering and absorption coefficients ($\sigma_s(\tau)$ and $\sigma_a(\tau)$, respectively) are given by $v_s(\tau) = \delta\sigma_s(\tau)/\sigma_s(\tau)$ and $v_a(\tau) = \delta\sigma_a(\tau)/\sigma_a(\tau)$, respectively. Substituting $v_e(\tau)$ and $v_\omega(\tau)$ according to (6.74) and (6.75) into (6.72) and introducing functions $\mathcal{W}_a$ and $\mathcal{W}_s$ as follows:

$$\mathcal{W}_a(\tau;\tau_v,\Omega_v) = \big[\mathcal{W}_e(\tau;\tau_v,\Omega_v) - \mathcal{W}_\omega(\tau;\tau_v,\Omega_v)\big]\big[1-\omega(\tau)\big]\,, \tag{6.76}$$

$$\mathcal{W}_s(\tau;\tau_v,\Omega_v) = \mathcal{W}_e(\tau;\tau_v,\Omega_v)\omega(\tau) + \mathcal{W}_\omega(\tau;\tau_v,\Omega_v)[1-\omega(\tau)]\,, \tag{6.77}$$

we have

$$\delta\phi(\varpi_v,\tau) = \mathcal{W}_a(\tau;\tau_v,\Omega_v)v_a(\tau) + \mathcal{W}_s(\tau;\tau_v,\Omega_v)v_s(\tau)\,. \tag{6.78}$$

Thus, functions $\mathcal{W}_a(\tau;\tau_v,\Omega_v)$ and $\mathcal{W}_s(\tau;\tau_v,\Omega_v)$ defined by (6.76) and (6.77) are the weighting functions for the absorption and scattering coefficients, respectively. As pointed out in the previous section, these WFs are expressed as a linear combination of the corresponding WFs for the extinction coefficient and the single scattering albedo.

## 6.6 Weighting functions for a mixture of scattering and absorbing components

In the previous sections we have derived the expressions for the weighting functions appropriate to the direct RTE formulated for the extinction coefficient and the single scattering albedo as two independent variables. In this section we extend the derived expressions for the case of a mixture of the scattering and absorbing components. This is of great importance for the Earth's atmosphere, where the solar radiation can be scattered by the molecules, aerosol particles, or cloud droplets and it be absorbed by various gases. The scattering and the extinction coefficients are defined in this case as follows:

$$\sigma_s(\tau) = \sum_{i=1}^{N_s} s_i(\tau)\,, \tag{6.79}$$

$$\sigma_e(\tau) = \sigma_s(\tau) + \sum_{k=1}^{N_a} a_k(\tau)\,, \tag{6.80}$$

where $s_i(\tau)$ is the scattering coefficient for the $i$th component, $N_s$ is the number of the scattering components, $a_k(\tau)$ is the $k$th absorption coefficient, and $N_a$ is the number of absorbing components including the absorption by aerosol, gases and clouds. The probability of the photon scattering on the $i$th component can now be defined as follows:

$$\omega_i(\tau) = \frac{s_i(\tau)}{\sigma_e(\tau)}\,. \tag{6.81}$$

This probability will be referred to as the partial single scattering albedo. Taking into account that each sort of the scattering components has its phase function, the radiative transfer operator, $L_e$, given by (6.13) should be rewritten as follows:

$$L_e = \mu\frac{\mathrm{d}}{\mathrm{d}\tau} + 1 - \sum_{i=1}^{N_s} \frac{\omega_i(\tau)}{4\pi} \int_{4\pi} \mathrm{d}\Omega' p_i(\tau, \Omega, \Omega')\otimes\,, \tag{6.82}$$

where $p_i(\tau, \Omega, \Omega')$ is the phase function appropriate to the $i$th scattering component. Assuming that the scattering processes on the different components are independent, we do not need any additional modification to describe the radiative transfer in the medium consisting of a mixture of the scattering and absorbing particles.

To modify the derived expressions for WFs, we consider the extinction coefficient and the partial single scattering albedos to be main parameter-functions. Although the expression for the extinction coefficient WF remains the same as given by (6.240), the multiple scattering source function, $J(\tau, \Omega)$, consists now of a sum of the multiple scattering source functions for all scattering components:

$$J(\tau, \Omega) = \sum_{i=1}^{N_s} J_i(\tau, \Omega) = \sum_{i=1}^{N_s} \frac{\omega_i(\tau)}{4\pi} \int_{4\pi} p_i(\tau, \Omega, \Omega') I(\tau, \Omega')\,\mathrm{d}\Omega'\,. \tag{6.83}$$

Thus, the expression for the extinction coefficient WF can be written as follows:

$$\mathcal{W}_e(\tau;\tau_v,\Omega_v) = \int_{4\pi} I^*(\tau,\Omega;\Omega_v)\Big[J(\tau,\Omega) + S_e(\tau,\Omega) - I(\tau,\Omega)\Big]\,\mathrm{d}\Omega\,, \quad (6.84)$$

where $J(\tau,\Omega)$ is given by (6.83).

The WF for the partial single scattering albedo, $\omega_i(\tau)$, can be derived in a way analogical to the derivation of the WF for the single scattering albedo in the case of one scattering component. Considering $\omega_i(\tau)$ as an independent variable, the result can be written in the form of Eq. (6.244) as follows:

$$\mathcal{W}_{\omega_i}(\tau;\tau_v,\Omega_v) = \int_{4\pi} I^*(\tau,\Omega;\Omega_v)\Big[J_i(\tau,\Omega) - B(\tau)\omega_i(\tau)\Big]\,\mathrm{d}\Omega\,. \quad (6.85)$$

Let us now derive WF for the absorption coefficient and the scattering coefficient of the $k$th component. Assuming that the extinction coefficient and all partial single scattering albedos are varying, we can write

$$\delta\phi(\tau;\tau_v,\Omega_v) = \mathcal{W}_e(\tau;\tau_v,\Omega_v)v_e(\tau) + \sum_{i=1}^{N_s}\mathcal{W}_{\omega_i}(\tau;\tau_v,\Omega_v)v_{\omega_i}(\tau)\,, \quad (6.86)$$

where $\delta\phi(\tau;\tau_v,\Omega_v)\,\mathrm{d}\tau$ is the contribution of the extinction coefficient and the partial single scattering albedo variations within an infinitesimal layer positioned at the optical depth $\tau$, into the variation of the measured functional, $\delta\Phi(\tau_v,\Omega_v)$, as given by (6.73), and $v_{\omega_i}(\tau)$ is the relative variation of the $i$th partial single scattering albedo, i.e., $v_{\omega_i}(\tau) = \delta\omega_i(\tau)/\omega_i(\tau)$. Varying the extinction coefficient, $\sigma_e(\tau)$, and the partial single scattering albedo, $\omega_i(\tau)$, with respect to the $k$th absorption coefficient, we obtain

$$v_e(\tau) = \frac{\delta_k\sigma_e(\tau)}{\sigma_e(\tau)} = \frac{a_k(\tau)}{\sigma_e(\tau)}v_{a_k}(\tau)\,, \quad (6.87)$$

$$v_{\omega_i}(\tau) = \frac{1}{\omega_i(\tau)}\delta_k\left[\frac{s_i(\tau)}{\sigma_e(\tau)}\right] = -\frac{a_k(\tau)}{\sigma_e(\tau)}v_{a_k}(\tau)\,, \quad (6.88)$$

where $v_{a_k}(\tau) = \delta a_k(\tau)/a_k(\tau)$. Substituting (6.87) and (6.88) into (6.86), we have

$$\delta_k\phi(\tau;\tau_v,\Omega_v) = \left[\mathcal{W}_e(\tau;\tau_v,\Omega_v) - \sum_{i=1}^{N_s}\mathcal{W}_{\omega_i}(\tau;\tau_v,\Omega_v)\right]\frac{a_k(\tau)}{\sigma_e(\tau)}v_{a_k}(\tau)\,. \quad (6.89)$$

Introducing the WF for the single scattering albedo, $\omega(\tau) = \sum\omega_i(\tau)$, as follows:

$$\mathcal{W}_\omega(\tau;\tau_v,\Omega_v) = \sum_{i=1}^{N_s}\mathcal{W}_{\omega_i}(\tau;\tau_v,\Omega_v)\,, \quad (6.90)$$

and taking into account (6.85) for the partial single scattering albedo WF, we have

$$\mathcal{W}_\omega(\tau;\tau_v,\Omega_v) = \int_{4\pi} I^*(\tau,\Omega;\Omega_v)\Big[J(\tau,\Omega) - B(\tau)\omega(\tau)\Big]\,\mathrm{d}\Omega\,, \tag{6.91}$$

where the multiple scattering source function, $J(\tau,\Omega)$, is given by (6.83) and $\omega(\tau)$ is the single scattering albedo. Thus, the expression for the $k$th absorption coefficient WF can be written in the following form:

$$\mathcal{W}_{a_k}(\tau;\tau_v,\Omega_v) = \frac{a_k(\tau)}{\sigma_e(\tau)}\Big[\mathcal{W}_e(\tau;\tau_v,\Omega_v) - \mathcal{W}_\omega(\tau;\tau_v,\Omega_v)\Big]\,. \tag{6.92}$$

In an analogous way we can derive the expression for the $k$th scattering coefficient WF. Indeed, varying now the extinction coefficient, $\sigma_e(\tau)$, and the partial single scattering albedo, $\omega_i(\tau)$, with respect to the $k$th scattering coefficient, we obtain

$$v_e(\tau) = \omega_k(\tau)v_{s_k}(\tau)\,, \tag{6.93}$$

$$v_{\omega_i}(\tau) = -\omega_k(\tau)v_{s_k}(\tau),\ i \neq k\,, \tag{6.94}$$

$$v_{\omega_k}(\tau) = v_{s_k}(\tau) - \omega_k(\tau)v_{s_k}(\tau),\ i = k\,, \tag{6.95}$$

where $v_{s_k}(\tau) = \delta s_k(\tau)/s_k(\tau)$.

Substituting (6.93)–(6.95) into (6.86), we have

$$\begin{aligned}\delta_k\phi(\tau;\tau_v,\Omega_v) \;&=\; \left[\mathcal{W}_e(\tau;\tau_v,\Omega_v) - \sum_{i=1}^{N_s}\mathcal{W}_{\omega_i}(\tau;\tau_v,\Omega_v)\right]\omega_k(\tau)v_{s_k}(\tau)\\ &+\; \mathcal{W}_{\omega_k}(\tau;\tau_v,\Omega_v)v_{s_k}(\tau)\,.\end{aligned} \tag{6.96}$$

Thus, the expression for the $k$th scattering coefficient WF can be written in the following form:

$$\begin{aligned}\mathcal{W}_{s_k}(\tau;\tau_v,\Omega_v) \;&=\; \omega_k(\tau)\Big[\mathcal{W}_e(\tau;\tau_v,\Omega_v) - \mathcal{W}_\omega(\tau;\tau_v,\Omega_v)\Big]\\ &+\; \mathcal{W}_{\omega_k}(\tau;\tau_v,\Omega_v)\,.\end{aligned} \tag{6.97}$$

Here, functions $\mathcal{W}_e$, $\mathcal{W}_{\omega_k}$, and $\mathcal{W}_\omega$ are given by (6.84), (6.85), and (6.91), respectively. Thus, we can see that also for a mixture of scattering and absorbing components the weighting functions for the scattering and absorption coefficients of each particle type can be expressed as a linear combination of WF derived for the main parameter-functions.

## 6.7 Examples of weighting functions for the aerosol and cloud parameters

In this section we compare the weighting functions for the absorption and scattering coefficients of the aerosol particles and for the scattering coefficient of the water droplets calculated according to the analytical expressions given by (6.92) and (6.97) to WFs obtained employing the numerical perturbation approach.

We will assume here that the measured functional, $\Phi(\tau_v, \Omega_v)$, is the intensity of the reflected radiation at the top of atmosphere (TOA) in the nadir direction. As pointed out in section 6.2, the WF for the absolute variation of a certain atmospheric parameter corresponds to the variational derivative of the intensity with respect to this parameter. Therefore, the WF can be calculated directly according to the definition of the variational derivative given by (6.3). Employing the finite-difference approximation, this expression can be rewritten as follows:

$$\mathcal{V}_p(\tau_i; \tau_v, \Omega_v) \approx \frac{I[\tau_v, \Omega_v; p(\tau) + \Delta p(\tau_i)] - I[\tau_v, \Omega_v; p(\tau)]}{\Delta p(\tau_i)}\,, \tag{6.98}$$

where $I[\tau_v, \Omega_v; p(\tau)]$ and $I[\tau_v, \Omega_v; p(\tau) + \Delta p(\tau_i)]$ are the unperturbed and perturbed intensities at the optical depth $\tau_v$ in direction $\Omega_v$ and $\Delta p(\tau_i)$ is the perturbation of the parameter $p(\tau)$ at the altitude level $\tau_i$. Equation (6.98) provides the finite-difference approximation for the variational derivative and, therefore, for the weighting function at an altitude level with the optical depth $\tau_i$. Assuming that the entire atmosphere is divided into $N-1$ layers, WF can be obtained for all altitude levels successively employing (6.98) for $i = 1, \ldots, N$. Thus, WFs for all discrete levels can be obtained running the radiative transfer model once for the unperturbed value of the desired parameter and subsequently $N$ times to account for parameter perturbations at all discrete altitude levels:

$$p_i(\tau) = p(\tau) + \Delta p(\tau_i),\ i = 1, \ldots, N\,. \tag{6.99}$$

Weighting functions calculated according to the described approach will be referred to as the numerical weighting functions.

It is worth noting that numerical WFs calculated according to (6.98) cannot be directly compared to the analytical WFs resulting from, for example, Eq. (6.97). The main reason for this is as follows. Any numerical method of solving the radiative transfer equation assumes a certain distribution of optical parameters within each discrete layer in the atmosphere and this distribution is based on the values specified at the boundaries of layers, i.e., at internal discretization levels. For this reason, a perturbation of the parameter $p(\tau)$ at an altitude level corresponding to the optical depth $\tau_i$ affects the vertical distribution of this parameter in both upper and lower layers having their boundary at this altitude, i.e., effectively the vertical distribution of the parameter is perturbed within two layers. Thus, the numerical WFs describe the variation of the intensity caused by the variation of the optical parameter within two altitude layers having a finite geometrical and optical thicknesses, whereas the analytical WFs describe the intensity variation caused by the variation of the parameter in an infinitesimally thin altitude layer. Therefore, to compare the numerical and analytical WFs we introduce the layer-integrated analytical WF corresponding to the variation of the parameter $p(\tau)$ at the discrete level $\tau_i$ as follows:

$$\mathcal{W}_p(\tau_i; \tau_v, \Omega_v) = \frac{1}{2}\int_{\tau_{i-1}}^{\tau_{i+1}} \mathcal{W}_p(\tau; \tau_v, \Omega_v)\,\mathrm{d}\tau\,, \tag{6.100}$$

where the integration is performed over two neighboring layers bordered by the altitude level $\tau_i$.

Although the computation of the numerical WFs employing (6.98) is very time-consuming, they are often needed to prove the correctness of the derivation and of the numerical implementation of the corresponding analytical expressions.

### 6.7.1 Weighting functions for the aerosol scattering coefficient and aerosol particles number density

At first we compare the numerical and the layer-integrated analytical WFs for the aerosol scattering coefficient in the $O_2$-A absorption band (spectral range 758–773 nm). The vertical profiles of the aerosol extinction coefficient and of the aerosol single scattering albedo used in calculations are presented in Fig. 6.1. The profiles are shown at 758 nm wavelength. The phase function of the aerosol particles was reperesented by the Heney–Greenstein phase function with a constant asymmetry parameter of 0.67.

According to the definition of the layer-integrated WFs the variation of the aerosol scattering coefficient, $\Delta s_a(\tau_i)$, at the level $\tau_i$ causes the variation of the intensity at the optical depth $\tau_v$ in the direction $\Omega_v$ which can be expressed as follows:

$$\Delta_i I(\tau_v, \Omega_v) = \mathcal{W}_{s_a}(\tau_i; \tau_v, \Omega_v) v_{s_a}(\tau_i)\,. \tag{6.101}$$

Here, $\mathcal{W}_{s_a}(\tau_i; \tau_v, \Omega_v)$ is the layer-integrated WF for the aerosol scattering coefficient and $v_{s_a}(\tau_i)$ is its relative variation. Dividing both sides of this equation

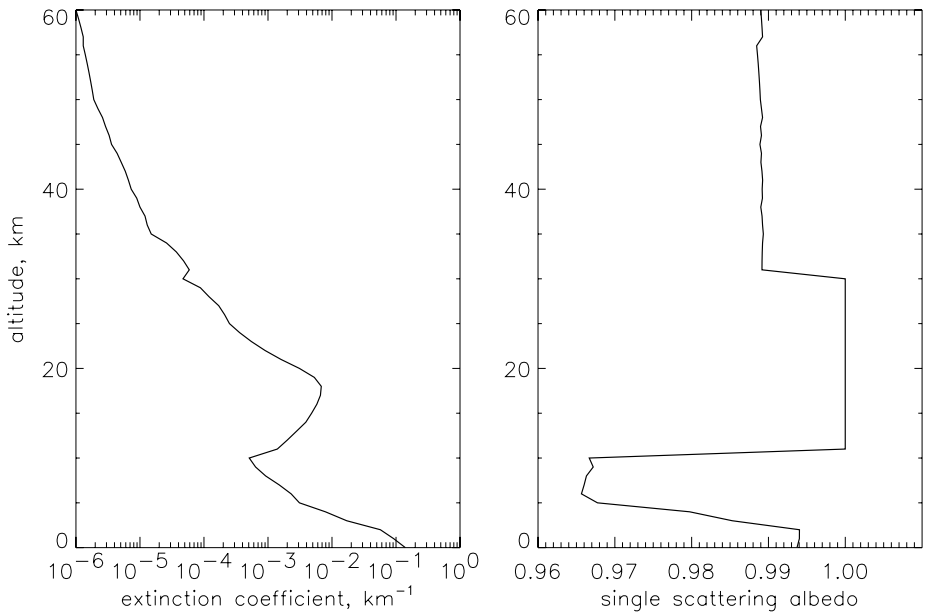



**Fig. 6.1.** Vertical profiles of the aerosol extinction coefficient (left panel) and of the aerosol single scattering albedo (right panel) used in the comparison of the weighting functions.

by the intensity $I(\tau_v, \Omega_v)$, we obtain

$$\frac{\Delta_i I(\tau_v, \Omega_v)}{I(\tau_v, \Omega_v)} = \frac{\mathcal{W}_{s_a}(\tau_i; \tau_v, \Omega_v)}{I(\tau_v, \Omega_v)} v_{s_a}(\tau_i) = \mathcal{R}_{s_a}(\tau_i; \tau_v, \Omega_v) v_{s_a}(\tau_i)\,. \tag{6.102}$$

Thus, assuming 1% variation of the aerosol scattering coefficient, $v_{s_a}(\tau_i)$, within a layer with a geometrical thickness $\Delta_i z = z_{i+1} - z_{i-1}$, the percentage variation of the intensity becomes numerically equivalent to the normalized layer-integrated WF, $\mathcal{R}_{s_a}(\tau_i; \tau_v, \Omega_v)$.

Comparisons of the numerical and the normalized layer-integrated analytical WFs for the aerosol scattering coefficient in the monochromatic case at selected wavelengths within the $O_2$-A absorption band are shown in Fig. 6.2. The numerical WFs were calculated assuming the relative variation of the aerosol scattering coefficient to be 0.01% at each atmospheric level. The entire atmosphere between 0 and 60 km was divided into 60 layers. As seen from the plot the shape and the maximum value of WFs are strongly dependent on the gaseous absorption. At very week absorption ($\tau_g = 0.006$ in Fig. 6.2) the maximum of WF is located near the surface and the maximum value is about 0.004%. An enhanced gaseous absorption causes the second maximum in the altitude region of the stratospheric aerosol layer (∼18 km, see Fig. 6.1) to appear. Further increase in the optical

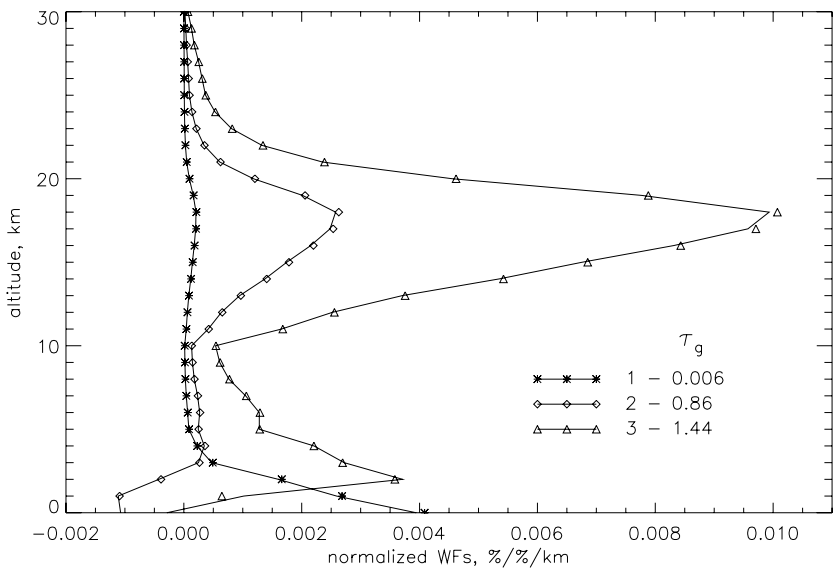



**Fig. 6.2.** Comparison of the numerical (symbols) and the normalized layer-integrated analytical WFs (solid line) for the aerosol scattering coefficient at selected wavelengths within the $O_2$-A absorption band: 1, 758.2 nm; 2, 763.8 nm; 3, 760.055 nm. The optical thicknesses of gaseous absorbers ($\tau_g$) at these wavelengths is shown in the legend. The weighting functions were calculated for the monochromatic reflected intensity at TOA observed in the nadir viewing geometry at a solar zenith angle of 45°. The surface albedo was set to 0.3.

thickness of the gaseous absorber to 1.44 raises the lower maximum of the WF and increases the value of the upper maximum to about 0.01%.

Obtained results demonstrate that spectral measurements of the reflected intensity within the absorptions bands of atmospheric gases, e.g., $O_2$-A band, contain certian information about the vertical distribution of the aerosol optical parameters. To illustrate this we consider more realistic case of measurements of the reflected solar radiance at TOA in the entire $O_2$-A absorption band with a finite spectral resolution. The instrument slit function is approximated by the boxcar function with a full width of 0.05 nm. The corresponding spectrum of the reflected solar radiance observed at TOA in nadir viewing geometry is shown in Fig. 6.3. The intensity were simulated assuming the incident solar flux at TOA to be equal $\pi$. This intensity is closely related to the reflection function, $R$, defined as follows:

$$R = \frac{\pi I}{\mu_0 F}\,, \tag{6.103}$$

where $I$ is the intensity of radiation, $F$ is the incident solar flux, and $\mu_0$ is the cosine of the solar zenith angle. If the incident solar flux, $F$, is assumed to be equal $\pi$ as mentioned above, the following relationship between the reflection function and the intensity is obtained:

$$R = \frac{I}{\mu_0}\,. \tag{6.104}$$

The variations of the aerosol scattering and absorption coefficients are caused in particular by the variation of the aerosol particles number density, $N_a(\tau)$.

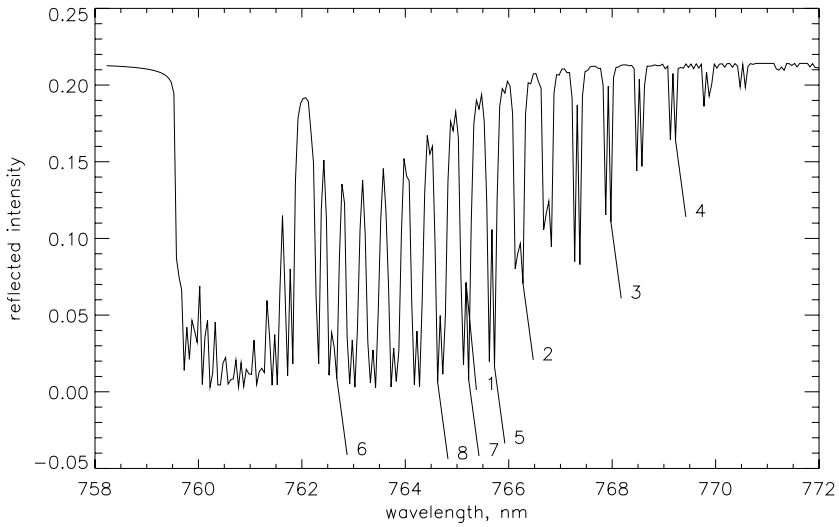



**Fig. 6.3.** Reflected solar radiation in the $O_2$-A absorption band spectral range observed at TOA in the nadir viewing geometry at a solar zenith angle of 45°. The surface albedo was set to 0.3. The wavelengths selected for presentation of the weighting functions in Fig. 6.4 are marked by the numbers.

Their relative variations can be expressed via the variation of $N_a(\tau)$ as follows:

$$\frac{\delta s_a(\tau)}{s_a(\tau)} = \frac{\delta N_a(\tau)}{N_a(\tau)} \quad \text{and} \quad \frac{\delta a_a(\tau)}{a_a(\tau)} = \frac{\delta N_a(\tau)}{N_a(\tau)}\,. \tag{6.105}$$

Since a variation of the aerosol particles number density leads to variations of both aerosol scattering and absorption coefficients, the corresponding weighting function has to be given by a linear combination of the aerosol scattering and absorption coefficient WFs, namely, taken into account (6.105), by their sum:

$$\mathcal{W}_{N_a}(\tau;\tau_v,\Omega_v) = \mathcal{W}_{s_a}(\tau;\tau_v,\Omega_v) + \mathcal{W}_{a_a}(\tau;\tau_v,\Omega_v)\,. \tag{6.106}$$

Figure 6.4 shows the normalized layer-integrated WFs for the aerosol particle number density at selected wavelengths within the $O_2$-A absorption band. As can be clearly seen, the shape and the maximum value of WFs are strongly dependend on the optical thickness of the gaseous absorber. For example, in a case of a weak gaseous absorption, $\tau_g \lesssim 6$, (left panel in Fig. 6.4) the maxima of the WFs are located in the lower troposphere and 1% variation in the particle number density within the boundary layer causes the variation of the reflected intensity in the range from $-0.0015\%$ to $+0.002\%$ depending on $\tau_g$. In the spectral channels characterized by a stronger gaseous absorption, $\tau_g \gtrsim 9$, (right panel in Fig. 6.4) WFs have their maxima in the altitude region of the

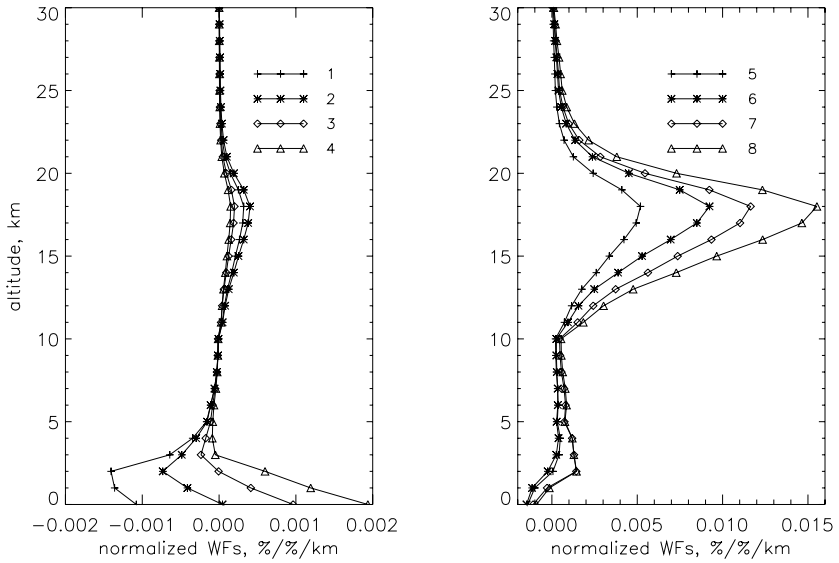



**Fig. 6.4.** Normalized layer-integrated WFs for the aerosol particle number density at selected wavelengths within the $O_2$-A absorption band. The optical thicknesses of gaseous absorbers at these wavelengths is: 1, 0.5; 2, 5.5; 3, 1.1; 4, 0.2; 5, 9.1; 6, 28.5; 7, 14.2; 8, 17.5. The weighting functions were calculated for the reflected intensity at TOA observed in the nadir viewing geometry at a solar zenith angle of 45°. The surface albedo was set to 0.3.

stratospheric aerosol layer (∼18 km). The sensitivity of the reflected radiance to the variation of the stratospheric aerosol particles concentration is much higher than in the case of the boundary layer aerosol. For example, 1% variation of the aerosol number density at the altitude of about 18 km causes a relative variation of the reflected intensity at the TOA of about 0.015% in a spectral channel with $\tau_g = 17.5$.

A quick look analysis of the vertical behavior of the aerosol number density WFs shows that their maxima are located at 0, 2, or 18 km altitudes depending on the optical thickness of the gaseous absorber. Therefore, we can conclude that the reflected solar radiation measured in the $O_2$-A absorption band contains information about three to four parameters characterizing the vertical distribution of the aerosol particle number density. This conclusion agrees very well with results of extensive analyses of the information content of spectral measurements of the reflected solar radiation in the $O_2$-A absorption band with respect to the vertical distribution of aerosol optical properties reported in [20, 29].

### 6.7.2 Weighting functions for the cloud scattering coefficient

Let us further discuss WFs for the cloud scattering coefficient. As an example, we consider the reflected radiance at TOA in the $O_2$-A absorption band spectral range. Figure 6.5 shows vertical profiles of the scattering coefficient and the single scattering albedo within a water cloud located between 3 and 4 km altitude. The profiles are presented at two selected wavelengths for both vertically homogeneous and inhomogeneous water clouds. Here, the single scattering albedo of cloud, $\omega_c(z)$, is defined as follows:

$$\omega_c(z) = \frac{s_c(z)}{\sigma_e(z)}\,, \tag{6.107}$$

where $s_c(z)$ is the cloud scattering coefficient and $\sigma_e(z)$ is the extinction coefficient comprizing all absorption coefficients of atmospheric gases, Rayleigh scattering coefficient, and aerosol and cloud extinction coefficients. Therefore, the cloud single scattering albedo depends on the altitude even in the case of a vertically homogeneous cloud.

Now let us introduce the normalized weighting functions for cloud scattering coefficient, i.e., the weighting functions describing the relative variation of the intensity. Similarly to (6.102) we obtain for the relative variation of the intensity:

$$\frac{\Delta_i I(\tau_v, \Omega_v)}{I(\tau_v, \Omega_v)} = \frac{\mathcal{W}_{s_c}(\tau_i; \tau_v, \Omega_v)}{I(\tau_v, \Omega_v) s_c(\tau_i)} \Delta s_c(\tau_i) = \mathcal{R}_{s_c}(\tau_i; \tau_v, \Omega_v) \Delta s_c(\tau_i)\,. \tag{6.108}$$

Unlike the aerosol WFs, the normalized WFs, $\mathcal{R}_{s_c}(\tau_i; \tau_v, \Omega_v)$, introduced here describe the relative variation of the intensity caused by an absolute, $\Delta s_c(\tau_i)$, rather than a relative, $\Delta s_c(\tau_i)/s_c(\tau_i)$, variation of the cloud scattering coefficient. Thus, the introduced normalized WF, $\mathcal{R}_{s_c}(\tau_i; \tau_v, \Omega_v)$, is numerically equivalent to the relative variation of the intensity observed at the optical depth $\tau_v$ and

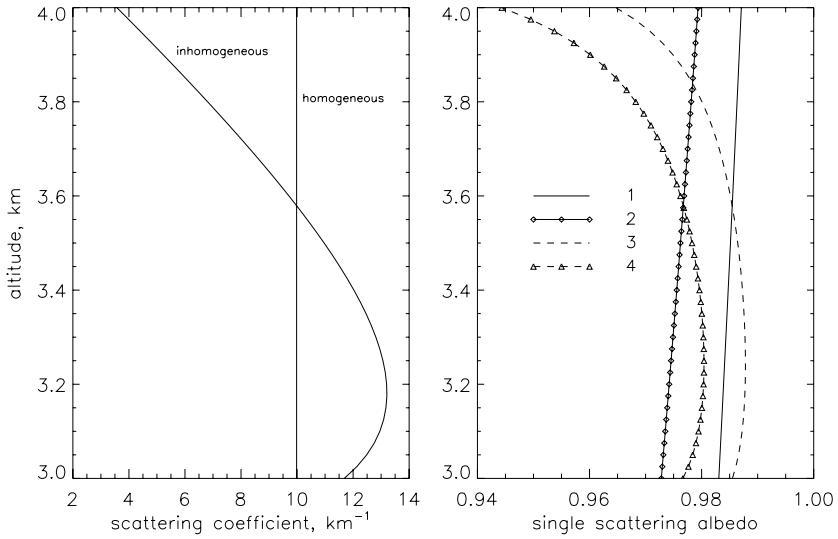



**Fig. 6.5.** Vertical profiles of the cloud scattering coefficient (left panel) and the cloud single scattering albedo (right panel) within a water cloud. The single scattering albedo is shown with solid lines (cases 1 and 2) for a homogeneous and with dashed lines (cases 3 and 4) for an inhomogeneous water cloud. The curves with and without symbols represent the single scattering albedo at 760.055 and 763.775 nm, respectively. The optical tickness of the gaseous absorbers at these wavelengths is 1.44 and 0.86, respectively.

direction $\Omega_v$ caused by $1\,\mathrm{km}^{-1}$ variation of the cloud scattering coefficient at the level $\tau_i$.

Comparisons of the numerical and normalized layer-integrated analytical WFs for the cloud scattering coefficient at two selected wavelengths for a vertically homogeneous and inhomogeneous cloud are shown in Fig. 6.6 for the monochromatic case. The layer-integrated WFs were obtained according to (6.100) assuming the geometrical thickness of integration layers (i.e., $\tau_{i+1} - \tau_{i-1}$) to be 50 m. The numerical WFs were calculated employing the numerical perturbation approach assuming 0.01% variation of the cloud scattering coefficient. As can be seen from the plot an enhancement of the gaseous absorption inscreases maximum values of WFs. Although the altitudinal behavior of WFs in homogeneous and inhomogeneous clouds is similar, their maxima are located closer to the cloud top height in the case of a homogeneous cloud.

To illustrate the influence of the gaseous absorption on the WFs of the cloud scattering coefficient we consider the reflected radiation within the entire $O_2$-A absorption band. The spectral distribution of the reflected solar radiation observed at TOA in the nadir viewing geometry is shown in Fig. 6.7. The calculations were performed for a finite spectral resolution assuming the instrument slit function to be the boxcar function with a full width of 0.05 nm. To investigate the dependence of WFs on the optical thickness of gaseous absorbers we have selected five wavelengths marked in Fig. 6.7 by the numbers. The optical

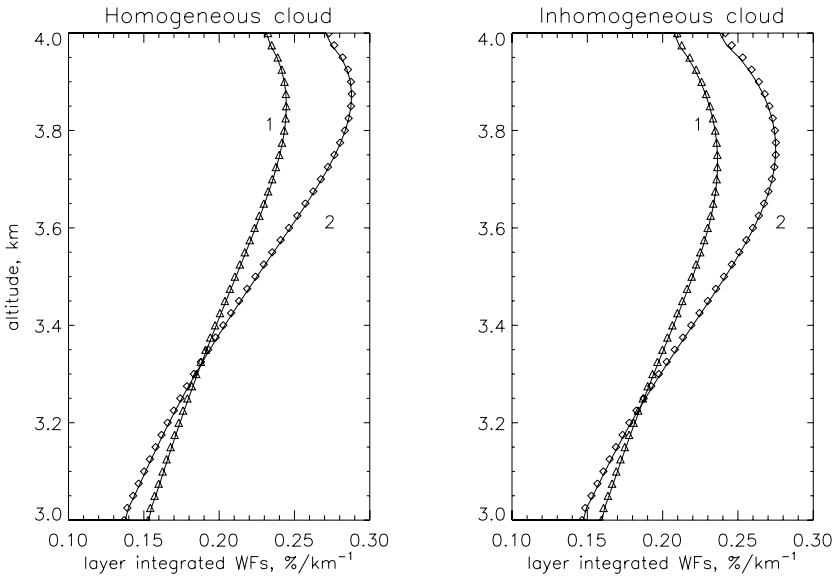



**Fig. 6.6.** Comparisons of the numerical (symbols) and normalized layer-integrated (solid line) WFs (layer geometrical thickness 50 m) for the cloud scattering coefficient at two selected wavelengths: 1, 763.775 nm ($\tau_g = 0.86$); 2, 760.055 nm ($\tau_g = 1.44$). The comparisons were performed for a solar zenith angle of 45°, surface albedo of 0.3, and cloud optical thickness of 10.

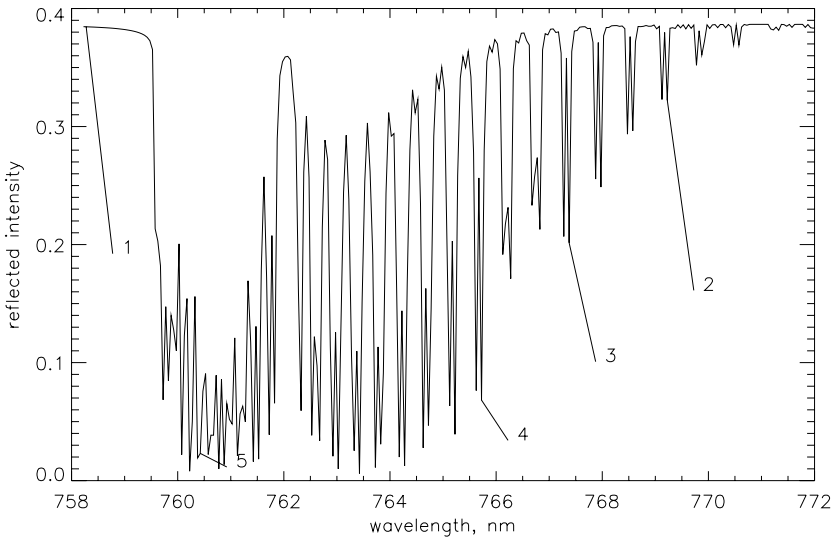



**Fig. 6.7.** Reflected solar radiation in the $O_2$-A absorption band spectral range in the presence of a homogeneous water cloud calculated for the same conditions as in Fig. 6.6. The wavelengths selected for presentation of the weighting functions in Fig. 6.8 are marked by the numbers.

thickness of gaseous absorbers ranges from 0.01 to 22 at these wavelengths. Figure 6.8 shows the normalized layer-integrated WFs at the selected wavelengths. At wavelengths where the gaseous absorption is weak, WFs are almost constant within the cloud demonstrating that no information on the vertical distribution of the cloud scattering coefficient within the cloud can be obtained. However, since the variation of the scattering coefficient at any altitude level within the cloud causes nearly the same variation of the reflected intensity, information on the optical thickness of the cloud independent of its vertical structure can be retrieved from measurements at these wavelengths. An enhancement in the optical thickness of the gaseous absorbers leads to an increased sensitivity of the reflected solar radiation to variations of the scattering coefficient in the upper part of a cloud. Thus, as can be seen from Fig. 6.8, for example, for the optical thickness of gaseous absorber of 9.11 a variation of the scattering coefficient of $1\,\mathrm{km}^{-1}$ near the cloud top height causes a variation of the intensity of about 0.24% whereas the same variation of the scattering coefficient near the cloud bottom results in just 0.15% variation of the intensity.

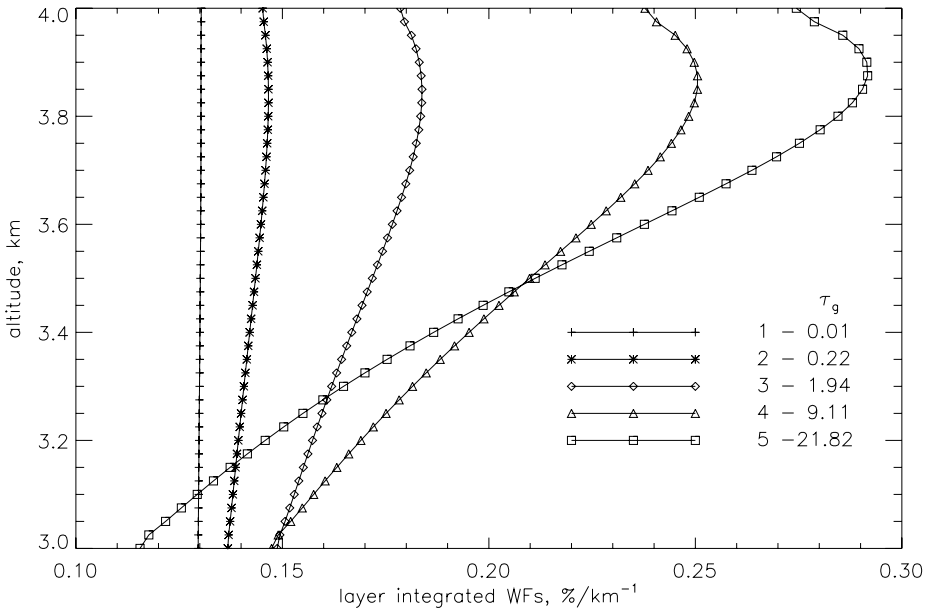



**Fig. 6.8.** Normalized layer-integrated WFs (layer geometrical thickness 50 m) for the cloud scattering coefficient at the wavelengths selected according to Fig. 6.7 corresponding to different values of optical thickness of gaseous absorber, $\tau_g$.

## 6.8 Weighting functions for temperature and pressure

### 6.8.1 Theory

Variations in the vertical distributions of the temperature and pressure cause variations of the Rayleigh scattering coefficient as well as of the cross-sections of atmospheric gases. The strongest dependence on the temperature and pressure takes place in the near-infrared spectral range for the absorption cross-sections of atmospheric gases and in the ultraviolet spectral range for the Rayleigh scattering coefficient. In this section we will demonstrate that appropriate expressions for the temperature and pressure WFs can be obtained as a linear combination of WFs for the Rayleigh scattering coefficient and absorption coefficients of atmospheric gases employing the expressions for the scattering and absorption coefficients WFs derived in section 6.6. We will not consider here the variation of the Planck function caused by the variation in the vertical distributions of the temperature which is of great importance in the thermal spectral region only (see [30–32]).

Let us assume that the variation of the measured functional is caused by the variation of the gaseous absorption coefficient, $a_k(\tau)$, only. Then we can rewrite (6.6) as follows:

$$\delta\Phi(\tau_v, \Omega_v) = \int_0^{\tau_0} \mathcal{W}_{a_k}(\tau; \tau_v, \Omega_v) v_{a_k}(\tau)\, \mathrm{d}\tau\,, \tag{6.109}$$

where $\mathcal{W}_{a_k}(\tau; \tau_v, \Omega_v)$ given by (6.92) and $v_{a_k}(\tau)$ are the weighting function and the relative variation of the $k$th gaseous absorption coefficient, respectively. Assuming further that the concentration of the absorbing gas does not vary, we have

$$\delta a_k(\tau) = n_k(\tau)\delta\sigma_k(\tau)\,, \tag{6.110}$$

where $n_k(\tau)$ is the number density profile of $k$th absorbing gas and the variation of the absorption cross-section, $\delta\sigma_k(\tau)$, is caused by variations of the temperature and pressure. Employing the Taylor series expansion of the absorption cross-section as a function of the temperature, $T(\tau)$, and the pressure, $P(\tau)$, we obtain in the linear approximation:

$$\delta\sigma_k(\tau) = \frac{\partial\sigma_k(\tau)}{\partial T(\tau)}\delta T(\tau) + \frac{\partial\sigma_k(\tau)}{\partial P(\tau)}\delta P(\tau)\,, \tag{6.111}$$

where $\partial\sigma_k(\tau)/\partial T(\tau)$ and $\partial\sigma_k(\tau)/\partial P(\tau)$ are the partial derivatives of the cross-section with respect to temperature and pressure, respectively. The partial derivatives can be calculated analytically or numerically. Substituting $\delta a_k(\tau)$ given by (6.110) into (6.109) and taking into account (6.111), we obtain the following linear relationship between the variation of the measured functional and variations of $T(\tau)$ and $P(\tau)$:

$$\delta\Phi(\tau_v, \Omega_v) = \int_0^{\tau_0} \frac{\mathcal{W}_{a_k}(\tau; \tau_v, \Omega_v)}{\sigma_k(\tau)} \left[\frac{\partial\sigma_k(\tau)}{\partial T(\tau)}\delta T(\tau) + \frac{\partial\sigma_k(\tau)}{\partial P(\tau)}\delta P(\tau)\right] \mathrm{d}\tau\,. \tag{6.112}$$

Thus, WFs for the temperature and pressure profiles are obtained as follows:

$$\mathcal{W}_T^a(\tau;\tau_v,\Omega_v) = \sum_{k=1}^{K} \frac{\partial \ln \sigma_k(\tau)}{\partial T(\tau)} \mathcal{W}_{a_k}(\tau;\tau_v,\Omega_v)\,, \tag{6.113}$$

$$\mathcal{W}_P^a(\tau;\tau_v,\Omega_v) = \sum_{k=1}^{K} \frac{\partial \ln \sigma_k(\tau)}{\partial P(\tau)} \mathcal{W}_{a_k}(\tau;\tau_v,\Omega_v)\,, \tag{6.114}$$

where $K$ is the full number of absorbing atmospheric gases. The superscript 'a' is introduced to emphasize that the corresponding WFs describe the variation of the measured functional caused by variations of the gaseous absorption coefficients.

In the ultraviolet and visible spectral ranges the variations of the temperature and pressure cause not only variations of the absorption cross-sections but also the variation of the Rayleigh scattering coefficient which is strongly dependent on the air number density and, therefore, on the temperature and pressure. To account for the contribution of the Rayleigh scattering coefficient variation into the variation of the measured functional we rewrite (6.109) in the following form:

$$\delta\Phi(\tau_v,\Omega_v) = \int_0^{\tau_0} \mathcal{W}_{s_k}(\tau;\tau_v,\Omega_v) v_{s_k}(\tau)\,\mathrm{d}\tau\,, \tag{6.115}$$

where we have assumed that the variation of the measured functional is caused only by the variation of the Rayleigh scattering coefficient whose relative variation is represented by $v_{s_k}(\tau)$. Employing the Taylor series expansion of the Rayleigh scattering coefficient as a function of the temperature and pressure, we obtain in the linear approximation:

$$\delta s_k(\tau) = \frac{\partial s_k(\tau)}{\partial T(\tau)} \delta T(\tau) + \frac{\partial s_k(\tau)}{\partial P(\tau)} \delta P(\tau)\,. \tag{6.116}$$

Similarly to (6.113) and (6.114) the contribution of the Rayleigh scattering coefficient to the temperature and pressure WFs can be obtained as follows:

$$\mathcal{W}_T^r(\tau;\tau_v,\Omega_v) = \frac{\partial \ln s_k(\tau)}{\partial T(\tau)} \mathcal{W}_{s_k}(\tau;\tau_v,\Omega_v)\,, \tag{6.117}$$

$$\mathcal{W}_P^r(\tau;\tau_v,\Omega_v) = \frac{\partial \ln s_k(\tau)}{\partial P(\tau)} \mathcal{W}_{s_k}(\tau;\tau_v,\Omega_v)\,. \tag{6.118}$$

Here, the superscript 'r' is introduced to emphasize that the corresponding WFs describe the variation of the measured functional caused by the variation of the Rayleigh scattering coefficient.

Thus, accounting for contributions of both the gaseous absorption and the Rayleigh scattering, the temperature and pressure WFs are obtained as

$$\mathcal{W}_T(\tau;\tau_v,\Omega_v) = \mathcal{W}_T^a(\tau;\tau_v,\Omega_v) + \mathcal{W}_T^r(\tau;\tau_v,\Omega_v)\,, \tag{6.119}$$

$$\mathcal{W}_P(\tau;\tau_v,\Omega_v) = \mathcal{W}_P^a(\tau;\tau_v,\Omega_v) + \mathcal{W}_P^r(\tau;\tau_v,\Omega_v)\,. \tag{6.120}$$

Derived WFs describe the contribution of variations of the temperature and pressure at a given optical depth into the variation of the measured functional. Assuming that the absolute variation of the temperature and the relative variation of the pressure can be considered to be independent of the altitude, the integrated WFs can be introduced as follows:

$$\mathcal{W}_T(\tau_v, \Omega_v) = \int_0^{\tau_0} \mathcal{W}_T(\tau; \tau_v, \Omega_v)\, \mathrm{d}\tau\,, \tag{6.121}$$

$$\mathcal{W}_P(\tau_v, \Omega_v) = \int_0^{\tau_0} \mathcal{W}_P(\tau; \tau_v, \Omega_v) P(\tau)\, \mathrm{d}\tau\,. \tag{6.122}$$

### 6.8.2 Examples

Using the integrated WFs, the variation of the measured functional can be rewritten in the following form:

$$\delta\Phi(\tau_v, \Omega_v) = \mathcal{W}_T(\tau_v, \Omega_v)\Delta T + \mathcal{W}_P(\tau_v, \Omega_v)\frac{\Delta P}{P}\,, \tag{6.123}$$

where $\Delta T$ and $\Delta P/P$ are the absolute variation of the temperature and the relative variation of the pressure, respectively, considered to be constant in the entire atmosphere. Dividing both sides of (6.123) by the measured functional, we introduce the normalized WFs for the temperature and pressure as follows:

$$\frac{\delta\Phi(\tau_v, \Omega_v)}{\Phi(\tau_v, \Omega_v)} = \mathcal{R}_T(\tau_v, \Omega_v)\Delta T + \mathcal{R}_P(\tau_v, \Omega_v)\frac{\Delta P}{P}\,, \tag{6.124}$$

where

$$\mathcal{R}_T(\tau_v, \Omega_v) = \frac{\mathcal{W}_T(\tau_v, \Omega_v)}{\Phi(\tau_v, \Omega_v)}\,,\ \mathcal{R}_P(\tau_v, \Omega_v) = \frac{\mathcal{W}_P(\tau_v, \Omega_v)}{\Phi(\tau_v, \Omega_v)}\,. \tag{6.125}$$

Spectral dependence of $\mathcal{R}_T(\tau_v, \Omega_v)$ and $\mathcal{R}_P(\tau_v, \Omega_v)$ functions appropriate to observations of the reflected solar radiance in the nadir viewing geometry is illustrated in Fig. 6.9 for the $CO_2$ absorption band spectral range, where only the gaseous absorption is substantial, and in Fig. 6.10 for 320–330 nm spectral range, where contributions of both the Rayleigh scattering and the ozone absorption are significant. The numerical WFs for the temperature and pressure shown in Figs 6.9 and 6.10 were calculated using the numerical perturbation approach according to (6.98) assuming a relative variation of the pressure of 0.01% and an absolute variation of the temperature of 0.1 K at each atmospheric level. As can be seen from Fig. 6.9 the variation of the reflected intensity within the $CO_2$ absorption band ranges from 0% to about −0.055% then caused by 1% variation of the pressure and from −0.055% to about 0.045% then caused by 1K variation in the temperature. Moreover, spectral dependencies of these variations are very different. Contributions of other gaseous absorbers as well as of the Rayleigh

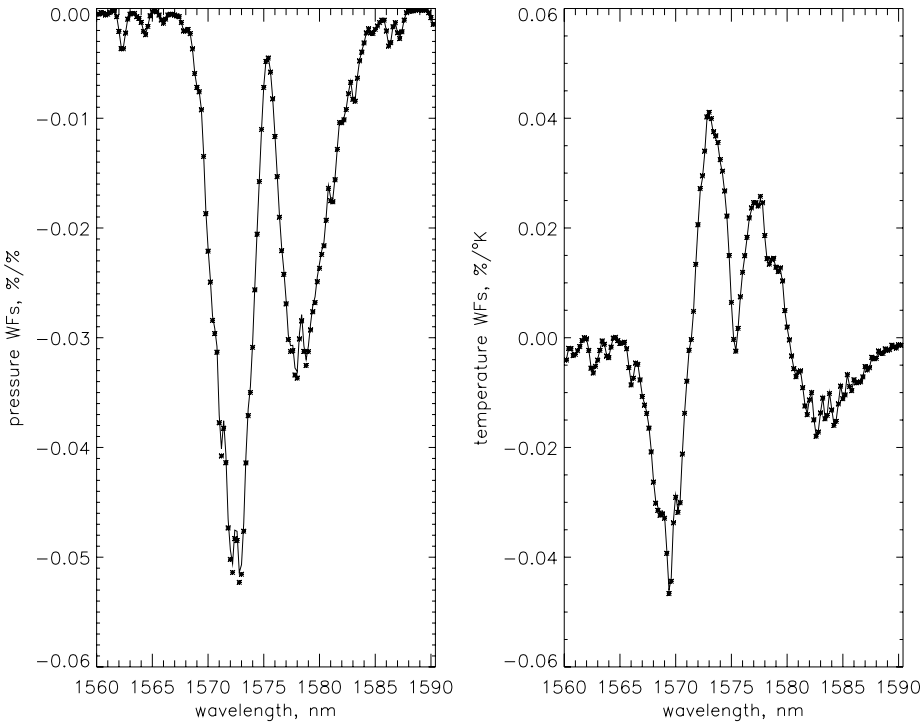



**Fig. 6.9.** Numerical (symbols) and normalized integrated analytical (solid line) WFs for relative variations of the pressure (left panel) and absolute variations of the temperature (right panel). Calculations were performed for the reflected intensity observed at TOA in the nadir viewing geometry at a solar zenith angle of 40°. The surface albedo was set to zero.

scattering can be neglected in this spectral range. This is, however, not the case in 320–330 nm spectral range shown in Fig. 6.10. Here, due to the strong contribution of the Rayleigh scattering, the relative variations of the intensity are about an order of magnitude larger than in the $CO_2$ absorption band spectral range. For example, 1% variation of the pressure causes 0.7% variation of the reflected intensity, whereas 1 K variation in the temperature results in −0.35% variation of the intensity.

Even if the relative variations of the reflected intensity cased by variations of atmospheric trace gas cross-sections are quite small, as occurs in the $CO_2$ absorption band spectral range shown in Fig. 6.9, they can result in noticeable additional errors in the retrieved trace gas number densities if corresponding parameters are neglected. For known uncertainties in the pressure and temperature these errors can be estimated employing the obtained WFs.

Relative variations of the reflected intensity in $CO_2$ absorption band spectral range due to 1% variation of $CO_2$ number density, 1% variation of the pressure, and 1 K variation of the temperature are shown in Fig. 6.11. We note that relative

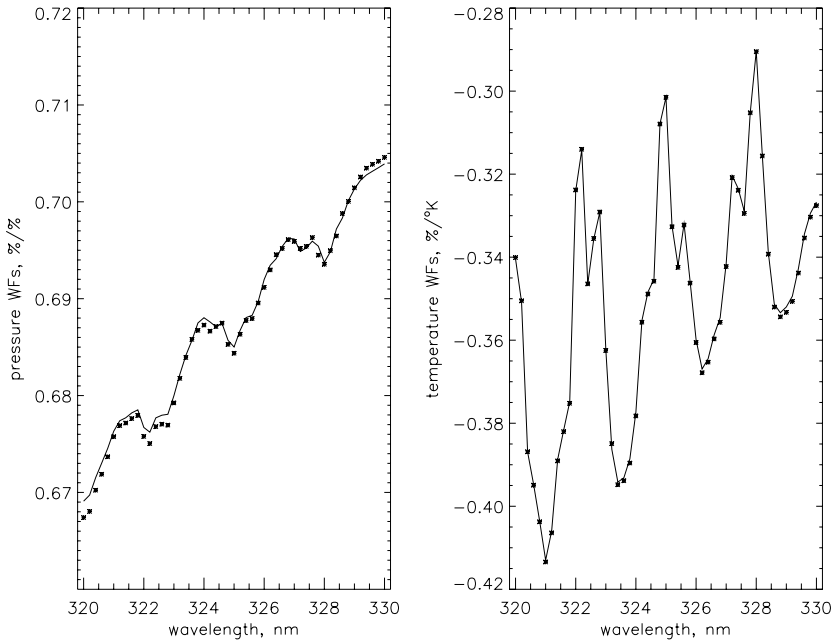



**Fig. 6.10.** Same as Fig. 6.9 but for 320–330 nm spectral range.

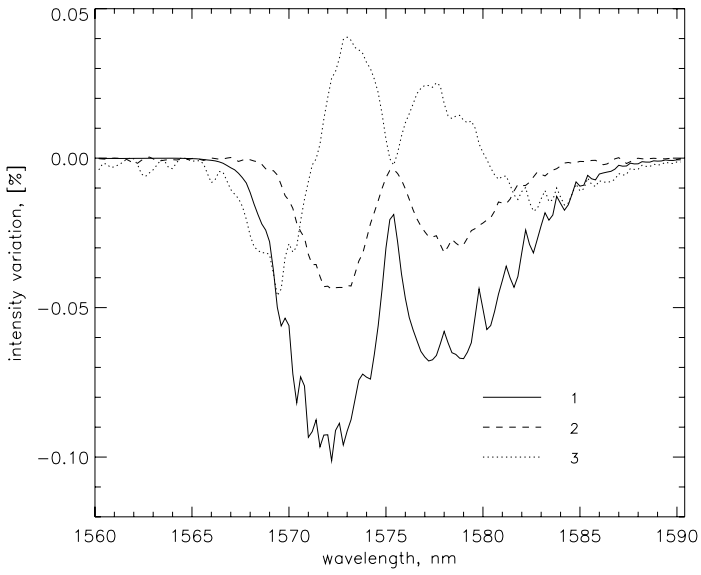



**Fig. 6.11.** Relative variations of the reflected intensity in $CO_2$ absorption band spectral range due to 1% variation of $CO_2$ number density (1), 1% variation of the pressure (2), and 1K variation of the temperature (3). Calculations were performed for the reflected intensity observed at TOA in the nadir viewing geometry at a solar zenith angle of 40°. The surface albedo was set to zero.

variations of the absorber number density and of the corresponding absorption coefficient result in the same relative variation of the reflected intensity (see [24] for details). As can be seen from Fig. 6.11, the relative variations of the observed intensity due to variations of the pressure and $CO_2$ number density have very similar spectral dependence, whereas the response to 1% variation of $CO_2$ number density is by a factor ∼2.5 stronger as compared to 1% variation of the pressure. Thus, this simple comparison allows us to conclude that 1% retrieval accuracy for $CO_2$ number density can only be achieved if the uncertainty of the atmospheric pressure does not exceed 2.5%. In the same manner, the influence of the temperature and other atmospheric parameters on the retrieval accuracy of $CO_2$ number density can be evaluated as well. Essential results of a quantitative analysis using the weighting function approach in the considered spectral range are reported in [3].

## 6.9 Weighting functions for particle number concentration and effective radius of droplets

### 6.9.1 Cloud parameters

The main optical parameters of a cloud are the scattering coefficient, the absorption coefficient, and the phase function. In the case of a vertically inhomogeneous cloud these parameters are usually defined within the cloud as functions of the altitude. Taking into account that the same cloud can be located at different altitudes in the atmosphere, it is more convenient to employ a dimensionless variable to describe vertical profiles of cloud parameters. As such a variable we introduce

$$x = \frac{z_t - z}{z_t - z_b}\,, \tag{6.126}$$

where $z_t$ and $z_b$ are the cloud top height (CTH) and cloud bottom height (CBH), respectively. The variable $x$ is dimensionless and ranges from 0 at CTH to 1 at CBH. The vertical profiles of cloud optical parameters can be defined using two different approaches. The first one is to define the shape of the cloud extinction coefficient, $s_h(x)$, within the cloud as a function of the dimensionless variable $x$. Then assuming the water droplet or ice crystal single scattering albedo, $\omega_k$, to be constant within the cloud, we obtain the following expressions for the cloud scattering, $s_k(z)$, and absorption, $a_k(z)$, coefficients:

$$s_k(z) = c\ \omega_k s_h(z)\,, \tag{6.127}$$

$$a_k(z) = c\ (1 - \omega_k) s_h(z)\,, \tag{6.128}$$

where $c$ is the scaling parameter obtained requiring the optical thickness of the cloud to be equal a pre-defined value of the optical thickness, $\tau_c$, which is considered to be an input optical integral parameter. Introducing the optical thickness, $\tau_{s_h}$, corresponding to the shape of the extinction coefficient, $s_h(z)$, as follows:

$$\tau_{s_h} = \int_{z_b}^{z_t} s_h(z)\,\mathrm{d}z\,, \tag{6.129}$$

we obtain for the scaling factor:

$$c = \frac{\tau_c}{\tau_{s_h}}\,. \tag{6.130}$$

Thus, according to this approach the input parameters for the cloud are: single scattering albedo, cloud optical thickness, and shape of the extinction coefficient. The phase function is assumed to be constant within the cloud. This approach will be used to define optical characteristics of water or ice clouds.

The second approach is based on the parameterization of the local optical characteristics of the cloud such as the extinction coefficient and the single scattering albedo. The extinction coefficients of water droplets, $e_1(z)$, and ice crystals, $e_2(z)$, can be represented in the form of the following analytical expressions [8]:

$$e_1(z) = \frac{3l_1(z)}{2r_1(z)\rho_1}K_e(z)\,, \tag{6.131}$$

$$K_e(z) = 1 + Ax_1^{-2/3}(z) - B\left[1 - e^{Cx_1^{-2/3}(z)}\right]\,, \tag{6.132}$$

$$e_2(z) = \frac{3l_2(z)}{2r_2(z)\rho_2}\,, \tag{6.133}$$

where subscripts '1' and '2' correspond to water droplets and ice crystals, respectively; $l_1(z)$ and $l_2(z)$ are the liquid water content (LWC) and ice water content (IWC), respectively; $r_1(z)$ and $r_2(z)$ are the effective radii of particles defined as

$$r_k(z) = \frac{3\bar{V}_k(z)}{\bar{G}_k(z)}\,, \quad k = 1, 2\,, \tag{6.134}$$

where $\bar{V}_k(z)$ and $\bar{G}_k(z)$ are the average volume and the average surface area of particles, respectively; $x_k(z) = 2\pi r_k(z)/\lambda$; $\lambda$ is the wavelength; $\rho_1$ and $\rho_2$ are the densities of the water and ice, respectively. Constants $A$, $B$, and $C$ are calculated employing the Mie theory: $A = 1.1$, $B = 1.7\ 10^{-6}$ and $C = 56.3$ [8].

The absorption properties of the water and ice particles are defined by the parameterization of the water droplets absorption coefficient, $a_1(z)$, and ice crystals single scattering albedo, $\omega_2(z)$, as follows [8]:

$$a_1(z) = \frac{4\pi\chi_1}{\lambda\rho_1}l_1(z)K_a(z)\,, \tag{6.135}$$

$$K_a(z) = 1.23\Big[1 - 2.6x_1(z)\chi_1\Big]\Big[1 + 0.34\left(1 - \mathrm{e}^{-8\lambda/r_1(z)}\right)\Big]\,, \tag{6.136}$$

$$\omega_2(z) = 1 - 0.47K_\omega(z)\,, \tag{6.137}$$

$$K_\omega(z) = 1 - e^{-2\eta x_2(z)\chi_2}\,. \tag{6.138}$$

Here, $\chi_k$ represents the imaginary part of the refractive index $m_k = n_k - i\chi_k$ of water ($k = 1$) and ice ($k = 2$), the parameter $\eta$ depends on the assumed shape of ice crystals. It is equal to 3.6 for fractal particles used in this work.

Summing up all obtained results, the scattering coefficient of the cloud will be defined as

$$s_1(z) = \frac{l_1(z)}{\rho_1}\left[\frac{3K_e(z)}{2r_1(z)} - \frac{4\pi\chi_1}{\lambda}K_a(z)\right], \tag{6.139}$$

$$s_2(z) = \frac{3}{2}\frac{l_2(z)}{r_2(z)\rho_2}\Big[1 - 0.47K_\omega(z)\Big], \tag{6.140}$$

and the absorption coefficient as

$$a_1(z) = \frac{l_1(z)}{\rho_1}\frac{4\pi\chi_1}{\lambda}K_a(z), \tag{6.141}$$

$$a_2(z) = \frac{3}{2}\frac{l_2(z)}{r_2(z)\rho_2}0.47K_\omega(z). \tag{6.142}$$

Thus, according to the second approach the input parameters for the cloud are:

- the vertical profiles of the liquid water and/or ice content;
- the vertical profiles of the effective radius of water droplets and/or ice crystals;
- the imaginary part of the refractive index for water and/or ice.

The phase function is assumed to be constant within the cloud. This approach will be used to define the local optical properties of water, ice or mixed clouds.

The accuracy of the proposed approximations has been studied in [8]. Comparisons of the obtained approximative formulae to the results of exact calculations show that the relative errors of the extinction coefficient approximation are below 1% for $\lambda < 2.5\,\mu$m and effective radii of water droplets greater than $4\,\mu$m. The accuracy of the approximative expression for the absorption coefficient is better than 10% for $\lambda < 2.5\,\mu$m and the effective radii of water droplets between 4 and $16\,\mu$m. In the spectral range, where the absorption of the liquid water is weaker ($\lambda < 1.8\,\mu$m), the accuracy of the absorption coefficient approximation is better than 5% for the same range of the effective radius.

The integral properties of the cloud such as cloud optical thickness (COT), liquid water path ($LWP$), and ice water path ($IWP$) are obtained as follows:

$$\tau_{c_k} = \int_{z_b}^{z_t} e_k(z)\,\mathrm{d}z\,,\ k = 1, 2\,, \tag{6.143}$$

$$LWP = \int_{z_b}^{z_t} l_1(z)\,\mathrm{d}z\,, \quad IWP = \int_{z_b}^{z_t} l_2(z)\,\mathrm{d}z\,, \tag{6.144}$$

where $e_k(z)$ is the extinction coefficient of water ($k = 1$) or ice ($k = 2$). Thus, according to this approach, COT, $LWP$, and $IWP$ are not considered as input parameters.

In the parameterization of the local cloud optical parameters given by (6.139)–(6.142) the liquid or ice water content and effective radius are considered as two independent variables. However, at least for a water cloud, there is a known relationship between LWC and the radius of water droplets. Assuming that the size distribution of water droplets is given by the function $f(r, z)$ normalized as follows:

$$\int_0^\infty f(r, z)\,\mathrm{d}r = 1\,, \tag{6.145}$$

where $r$ is the radius of water droplets, LWC at a given altitude $z$ can be expressed as

$$l_1(z) = \frac{4\pi\rho_1}{3} N(z) \int_0^\infty r^3 f(r, z)\,\mathrm{d}r\,, \tag{6.146}$$

where $\rho_1$ is the density of the liquid water and $N(z)$ is the particle number concentration of droplets. Introducing the third moment of the size distribution function as follows:

$$\langle r^3(z)\rangle = \int_0^\infty r^3 f(r, z)\,\mathrm{d}r \tag{6.147}$$

and substituting it into (6.146), we obtain

$$l_1(z) = \frac{4\pi\rho_1}{3} N(z)\langle r^3(z)\rangle\,. \tag{6.148}$$

In most cases, the measured particle size distribution function, $f(r, z)$, can be approximated by the gamma distribution (see [7] and references therein):

$$f(r, z) = Br^\mu\,\mathrm{e}^{-\mu[r/r_0(z)]}\,, \tag{6.149}$$

where,

$$B = \frac{\mu^{\mu+1}}{\Gamma(\mu+1) r_0^{\mu+1}(z)} \tag{6.150}$$

is the normalization constant and $\Gamma(\mu+1)$ is the gamma function. The parameters $\mu$ and $r_0(z)$ characterize the width and the maximum position of the particle size distribution function, respectively. Now let us assume that only parameter $r_0(z)$ depends on the altitude $z$. In this case the following expressions for $\langle r^3(z)\rangle$ and the effective radius, $r_1(z)$, can be obtained:

$$\langle r^3(z)\rangle = \left(\frac{r_0(z)}{\mu}\right)^3 \frac{\Gamma(4+\mu)}{\Gamma(1+\mu)}\,, \tag{6.151}$$

$$r_1(z) = \frac{r_0(z)}{\mu}\,(3+\mu)\,. \tag{6.152}$$

Taking into account properties of the gamma function [13], we obtain the following relationship between the third moment and the effective radius:

$$\langle r^3(z)\rangle = r_1^3(z)\frac{(\mu+2)(\mu+1)}{(\mu+3)^2}\,. \tag{6.153}$$

Substituting obtained expression into (6.148) it follows:

$$l_1(z) = \frac{4\pi\rho_1}{3}\frac{(\mu+2)(\mu+1)}{(\mu+3)^2}N(z)r_1^3(z)\,. \tag{6.154}$$

Thus, for a water cloud, instead of the parameterization via dependent parameters $l_1(z)$ and $r_1(z)$ we have obtained parameterization via an independent pair $N(z)$ and $r_1(z)$. This relationship is often used by the retrieval of water-cloud droplet effective radius (see [5] and references therein).

### 6.9.2 Weighting functions

In this subsection we derive the weighting functions for the particle number concentration, $N(z)$, liquid water content, $l_1(z)$, and the effective radius of water droplets, $r_1(z)$. We start from expressions for the weighting functions for the cloud absorption and scattering coefficients given by (6.92) and (6.97), respectively. Let us assume that the variation of the measured functional, $\delta\Phi$, is caused by the variation of the cloud scattering, $s_1(\tau)$, and absorption, $a_1(\tau)$, coefficients of the water droplets only. Then according to (6.68) we have

$$\delta\Phi(\tau_v,\Omega_v) = \int_0^{\tau_0}\left[\mathcal{W}_{s_1}(\tau;\tau_v,\Omega_v)v_{s_1}(\tau) + \mathcal{W}_{a_1}(\tau;\tau_v,\Omega_v)v_{a_1}(\tau)\right]\mathrm{d}\tau\,, \tag{6.155}$$

where $v_{s_1}(\tau) = \delta s_1(\tau)/s_1(\tau)$ and $v_{a_1}(\tau) = \delta a_1(\tau)/a_1(\tau)$ are the relative variations of the scattering and absorption coefficients of water droplets, respectively. The variation of the scattering and absorption coefficients given by (6.139) and (6.141), respectively, can be caused by the variations of the particle number concentration and the effective radius. Varying the expression for the scattering coefficient given by (6.139) with respect to $N(z)$ and $r_1(z)$, we obtain

$$\delta s_1(z) \;=\; F_s(z)\frac{\partial l_1(z)}{\partial N(z)}\delta N(z) + \left[F_s(z)\frac{\partial l_1(z)}{\partial r_1(z)} + l_1(z)\frac{\partial F_s(z)}{\partial r_1(z)}\right]\delta r_1(z)\,, \tag{6.156}$$

where the function $F_s(z)$ is defined as follows:

$$F_s(z) = \frac{1}{\rho_1}\left[\frac{3}{2}\frac{K_e(z)}{r_1(z)} - \frac{4\pi\chi_1}{\lambda}K_a(z)\right]\,. \tag{6.157}$$

Equation (6.156) was obtained taking into account that according to (6.154) the liquid water content is a function of $N(z)$ and $r_1(z)$. If we consider the liquid water content as an independent variable, then varying the expression for the scattering coefficient given by (6.139) with respect to $l_1(z)$ and $r_1(z)$, we obtain

$$\delta s_1(z) \;=\; F_s(z)\delta l_1(z) + l_1(z)\frac{\partial F_s(z)}{\partial r_1(z)}\delta r_1(z)\,. \tag{6.158}$$

Taking into account that $\partial l_1(z)/\partial N(z) = l_1(z)/N(z)$ and $\partial l_1(z)/\partial r_1(z) = 3l_1(z)/r_1(z)$ as well as $F_s(z) = s_1(z)/l_1(z)$, we rewrite (6.156) and (6.158) in the following form:

$$\delta s_1(z) = s_1(z)\frac{\delta N(z)}{N(z)} + 3s_1(z)\frac{\delta r_1(z)}{r_1(z)} + l_1(z)\frac{\partial F_s(z)}{\partial r_1(z)}\delta r_1(z)\,. \quad (6.159)$$

$$\delta s_1(z) = s_1(z)\frac{\delta l_1(z)}{l_1(z)} + l_1(z)\frac{\partial F_s(z)}{\partial r_1(z)}\delta r_1(z)\,. \quad (6.160)$$

Dividing both sides of this equation by $s_1(z)$, the relative variation of the scattering coefficient can now be obtained via the variation of the particle number concentration or of the liquid water content and of the effective radius of water droplets as follows:

$$v_{s_1}(z) = v_N(z) + \left(3 + \frac{r_1(z)}{F_s(z)}\frac{\partial F_s(z)}{\partial r_1(z)}\right) v_r(z)\,, \quad (6.161)$$

$$v_{s_1}(z) = v_{l_1}(z) + \frac{r_1(z)}{F_s(z)}\frac{\partial F_s(z)}{\partial r_1(z)} v_r(z)\,, \quad (6.162)$$

where $v_N(z)$, $v_{l_1}(z)$, and $v_r(z)$ are the relative variations of the particle number concentration, liquid water content and the effective radius, respectively. Taking into account that

$$\frac{r_1(z)}{F_s(z)}\frac{\partial F_s(z)}{\partial r_1(z)} = \frac{\partial \ln F_s(z)}{\partial \ln r_1(z)}\,, \quad (6.163)$$

Equations (6.161) and (6.162) can be rewritten as follows:

$$v_{s_1}(z) = v_N(z) + \left(3 + \frac{\partial \ln F_s(z)}{\partial \ln r_1(z)}\right) v_r(z)\,, \quad (6.164)$$

$$v_{s_1}(z) = v_{l_1}(z) + \frac{\partial \ln F_s(z)}{\partial \ln r_1(z)} v_r(z)\,. \quad (6.165)$$

The expression for the variation of the absorption coefficient, $\delta a_1(z)$, can be derived in a way analogous to the derivation of $\delta s_1(z)$. Introducing the function $F_a(z)$ as follows:

$$F_a(z) = \frac{4\pi\chi_1}{\lambda\rho_1} K_a(z)\,, \quad (6.166)$$

we obtain

$$v_{a_1}(z) = v_N(z) + \left(3 + \frac{\partial \ln F_a(z)}{\partial \ln r_1(z)}\right) v_r(z)\,, \quad (6.167)$$

$$v_{a_1}(z) = v_{l_1}(z) + \frac{\partial \ln F_a(z)}{\partial \ln r_1(z)} v_r(z)\,. \quad (6.168)$$

The weighting functions for relative variations of the particle number concentration and effective radius of water droplets can now be obtained substituting $v_{s_1}(z)$, and $v_{a_1}(z)$ given by (6.164) and (6.167), respectively, into (6.155):

$$\mathcal{W}_N(\tau;\tau_v,\Omega_v) = \mathcal{W}_{s_1}(\tau;\tau_v,\Omega_v) + \mathcal{W}_{a_1}(\tau;\tau_v,\Omega_v)\,, \tag{6.169}$$

$$\mathcal{W}_r^N(\tau;\tau_v,\Omega_v) = \tilde{F}_s(z)\mathcal{W}_{s_1}(\tau;\tau_v,\Omega_v) + \tilde{F}_a(z)\mathcal{W}_{a_1}(\tau;\tau_v,\Omega_v)\,, \tag{6.170}$$

where

$$\tilde{F}_s(z) = 3 + \frac{\partial \ln F_s(z)}{\partial \ln r_1(z)}\,, \quad \tilde{F}_a(z) = 3 + \frac{\partial \ln F_a(z)}{\partial \ln r_1(z)}\,, \tag{6.171}$$

and the superscript '$N$' of the WF for the effective radius denotes that this weighting function corresponds to the pair $(N,r)$, i.e., the particle number concentration and effective radius are considered as independent parameters.

The weighting functions for relative variations of the liquid water content and effective radius of water droplets can be obtained in an analogous way substituting $v_{s_1}(z)$, and $v_{a_1}(z)$ given by (6.165) and (6.168), respectively, into (6.155):

$$\mathcal{W}_{l_1}(\tau;\tau_v,\Omega_v) = \mathcal{W}_{s_1}(\tau;\tau_v,\Omega_v) + \mathcal{W}_{a_1}(\tau;\tau_v,\Omega_v)\,, \tag{6.172}$$

$$\mathcal{W}_r^{LWC}(\tau;\tau_v,\Omega_v) = \frac{\partial \ln F_s(z)}{\partial \ln r_1(z)}\mathcal{W}_{s_1}(\tau;\tau_v,\Omega_v) + \frac{\partial \ln F_a(z)}{\partial \ln r_1(z)}\mathcal{W}_{a_1}(\tau;\tau_v,\Omega_v)\,, \tag{6.173}$$

where the superscript '$LWC$' of the WF for the effective radius denotes that this weighting function corresponds to the pair (LWC,$r$). Comparing expressions for WFs obtained for the pairs $(N,r)$ and (LWC,$r$), we see that WFs for the relative variation of the particle number concentration and the relative variation of the liquid water content are identical. The relationship between $\mathcal{W}_r^{LWC}$ and $\mathcal{W}_r^N$ obtained for pairs (LWC,$r$) and $(N,r)$, respectively, can be easily derived:

$$\mathcal{W}_r^N(\tau;\tau_v,\Omega_v) = \mathcal{W}_r^{LWC}(\tau;\tau_v,\Omega_v) + 3\,\mathcal{W}_N(\tau;\tau_v,\Omega_v)\,. \tag{6.174}$$

Therefore, we obtain the linear relationship between the variation of the measured functional, $\delta\Phi(\tau_v,\Omega_v)$, and relative variations of the particle number concentration and effective radius of water droplets as follows:

$$\delta\Phi(\tau_v,\Omega_v) = \int_0^{\tau_0} \left[\mathcal{W}_N(\tau;\tau_v,\Omega_v)v_N(\tau) + \mathcal{W}_r^N(\tau;\tau_v,\Omega_v)v_r(\tau)\right] \mathrm{d}\tau\,. \tag{6.175}$$

Introducing as a vertical coordinate the altitude $z$ instead of the optical depth, $\tau$, and taking into account that variations of cloud parameters are non-zero only within the cloud, we can rewrite (6.175) in the following form:

$$\delta\Phi(z_v,\Omega_v) = \int_{z_b}^{z_t} \left[\mathcal{W}_N(z;z_v,\Omega_v)v_N(z) + \mathcal{W}_r^N(z;z_v,\Omega_v)v_r(z)\right] \sigma_e(z)\,\mathrm{d}z\,. \tag{6.176}$$

This equation can be used to retrieve vertical profiles of the particle number concentration and the effective radius of water droplets from measurements of the transmitted or reflected solar radiation in the range of the liquid water absorption or both within and outside the absorption bands of gaseous components such as $O_2$, $O_4$, or $CO_2$.

## 6.10 Examples of weighting functions for particle number concentration, liquid water content, and effective radius of water droplets

In this section we consider selected examples of WFs for the particle number concentration, liquid water content and effective radius of water droplets calculated using expressions obtained in the previous section. The normalized layer-integrated WFs for relative variations of the effective radius of water droplets and the liquid water content corresponding to the pair (LWC,$r$) are shown in Fig. 6.12 at four selected wavelengths. All calculations were performed for a vertically homogeneous water cloud having an optical thickness of 10. The wavelengths were selected as follows: one spectral point within the $O_2$-A absorption band characterized by a strong gaseous absorption (760.425 nm), two spectral points within absorption bands of the liquid water (1.6 μm and 2.2 μm), and one point in a spectral range where no absorption features are present (758.2 nm). WFs shown here are defined as:

$$\mathcal{R}_p(\tau_i;\tau_v,\Omega_v) = \frac{\mathcal{W}_p(\tau_i;\tau_v,\Omega_v)}{I(\tau_v,\Omega_v)}\,, \tag{6.177}$$

where subscript $p$ is equal to '$r$' or to '$l_1$' for the effective radius of water droplets and LWC, respectively. The function $\mathcal{R}_p(\tau_i;\tau_v,\Omega_v)$ equals to the percentage variation of the reflected intensity observed at TOA in the nadir viewing geometry caused by a 1% variation of corresponding parameters at a given position within a layer of 50 m geometrical thickness.

As can be seen from Fig. 6.12, WFs for LWC are always positive whereas WFs for the effective radius of water droplets are always negative. This can be easily explained considering the analytical expression for the water cloud scattering coefficient given by (6.139). Indeed, the scattering coefficient is proportional to LWC and inversely proportional to the effective radius of water droplets. Therefore, an enhancement of the liquid water content leads to an enhancement of the scattering coefficient and, thus, of the cloud optical thickness which results, in turn, in an increase in the reflected intensity. This is why WFs for the liquid water content are positive. These WFs show very week dependence on the altitude at all considered wavelengths with exception of the spectral point at a strong gaseous absorption (760.425 nm) where the sensitivity of the reflected radiation to the variation of LWC near the cloud top is as large as twice the sensitivity to the variation in the bottom cloud layers. Unlike LWC weighting functions, WFs for the effective radius of water droplets show a significant dependence on the altitude for all wavelengths where the gaseous or liquid water absorption occurs whereas there is absolutely no altitude dependence of WF at spectral points where no absorption signatures are present (758.2 nm). The maxima of WFs are located near the cloud top. For example, as seen from the left panel in Fig. 6.12, WF for the effective radius of water droplets at 2.2 μm wavelength in the layers near the cloud top height is as large as twice the WF in the bottom cloud layers.

In the previous section we have derived the expression for the effective radius WF introducing as an independent variable the particle number concentration

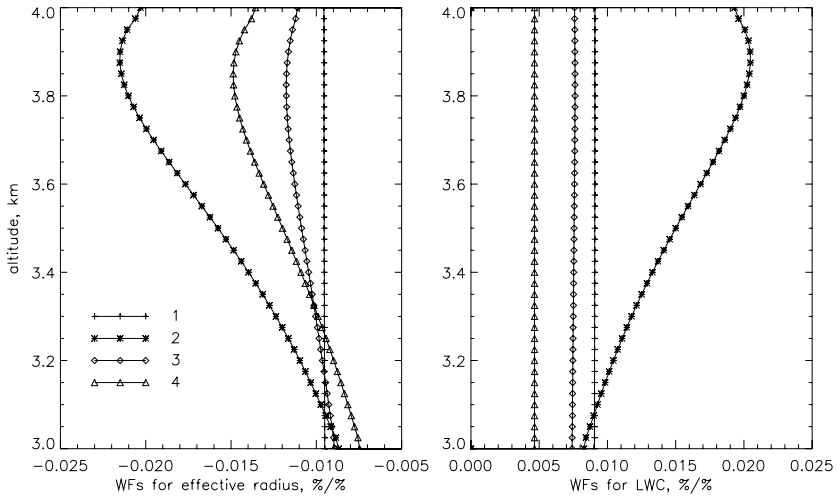



**Fig. 6.12.** Normalized layer-integrated WFs (layer geometrical thickness 50 m) for the relative variation of the effective radius of water droplets (left panel) and liquid water content (right panel) corresponding to the pair (LWC,r). Calculations were performed for the reflected solar radiance observed at TOA in nadir viewing geometry at a solar zenith angle of 45° at the following wavelengths: 1, 758.2 nm; 2, 760.425 nm; 3, 1.6 μm; 4, 2.2 μm. A vertically homogeneous water cloud with an optical thickness of 10 was assumed. The surface albedo was set to 0.3.

instead of the liquid water content. The corresponding WFs for the pair $(N, r)$ are shown in Fig. 6.13. As clearly seen, WFs change the sign if $(N, r)$ instead of (LWC,$r$) representation is used. This fact can be explained taking into account that LWC expressed via the particle number concentration and the effective radius of water droplets according to (6.154) is proportional to $r^3$. Therefore, in the spectral range where liquid water absorption is weak the scattering coefficient given by (6.139) is proportional to $r^2$. Thus, an increase of the effective radius of water droplets under assumption that the particle number concentration remains the same leads to an increase in the scattering coefficient and, therefore, in the cloud optical thickness which results, in turn, in an increase in the reflected radiation.

According to its definition, the weighting function provides a linear relationship between the variation of the corresponding parameter and the variation of the reflected or transmitted intensity. However, the intensity of the radiation depends nonlinearly on such cloud parameters as the particle number concentration and the effective radius of water droplets. The accuracy of the linear approximation depends on the magnitude of the parameter variation as well as on the observation geometry, on the solar zenith angle, on the cloud optical thickness, on the absorption of light within the cloud, etc. To investigate the accuracy of the linear approximation we assume that relative variations of all parameters are constant within the cloud. Under this assumption the relative

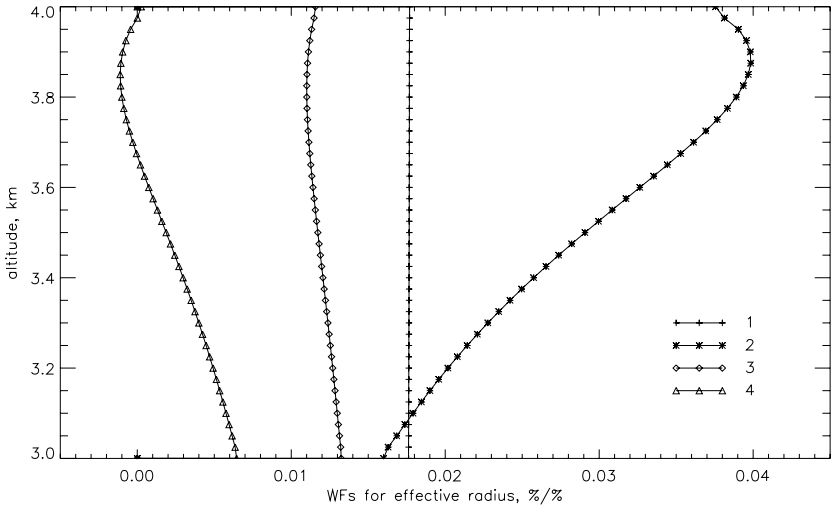



**Fig. 6.13.** Normalized layer-integrated WFs (layer geometrical thickness 50 m) for the relative variation of the effective radius of water droplets corresponding to the pair (N,r). Calculations were performed for the same conditions as in Fig. 6.12.

variation of the intensity can be expressed as follows:

$$\frac{\Delta I(\tau_v, \Omega_v)}{I(\tau_v, \Omega_v)} = \sum_{i=1}^{N_l} \mathcal{R}_p(\tau_i; \tau_v, \Omega_v) v_p \,, \tag{6.178}$$

where $N_l$ is the number of layers within the cloud and subscript 'p' corresponds to the effective radius or to the particle number concentration. The relative variations of the reflected intensity corresponding to relative variations of the effective radius, $v_r$, and particle number concentration, $v_N$, in the range of 5–20% are shown in Fig. 6.14 and Fig. 6.15, respectively. The numerical values for intensity variations are calculated according to the following expression:

$$\frac{\Delta I(\tau_v, \Omega_v)}{I(\tau_v, \Omega_v; p)} = \frac{I(\tau_v, \Omega_v; p + \Delta p) - I(\tau_v, \Omega_v; p)}{I(\tau_v, \Omega_v; p)} \,, \tag{6.179}$$

where $I(\tau_v, \Omega_v; p)$ and $I(\tau_v, \Omega_v; p + \Delta p)$ are solutions of the radiative transfer equation corresponding to the parameters p and p + $\Delta p$, respectively.

The comparison of results presented in Figs 6.14 and 6.15 shows that the variations of the intensity calculated employing the linear approximation are systematically higher than the exact values. It follows as well that the linear approximation works better for variations of the particle number concentration. This is due to the fact that at least the water cloud scattering coefficient is the linear function of the particle number concentration whereas its dependence on the effective radius of water droplets is quadratic. The effect of the nonlinearity is especially pronounced in the spectral range where the liquid water absorption is strong. For example, as can be seen from Fig. 6.14, at 2.2 µm 20% variation

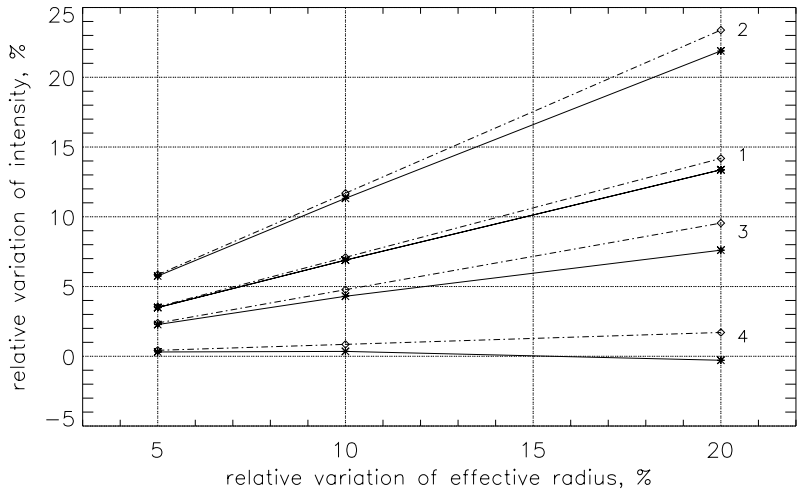



**Fig. 6.14.** Comparison of the relative variations of the reflected intensity caused by variations of the effective radius of water droplets. Solid line represents the results of numerical calculations and dash-dotted line corresponds to the linear approximation. The calculations were performed for a solar zenith angle of 45° and a surface albedo of 0.3 at the following wavelengths: 1, 758.2 nm; 2, 760.425 nm; 3, 1.6 μm; 4, 2.2 μm.

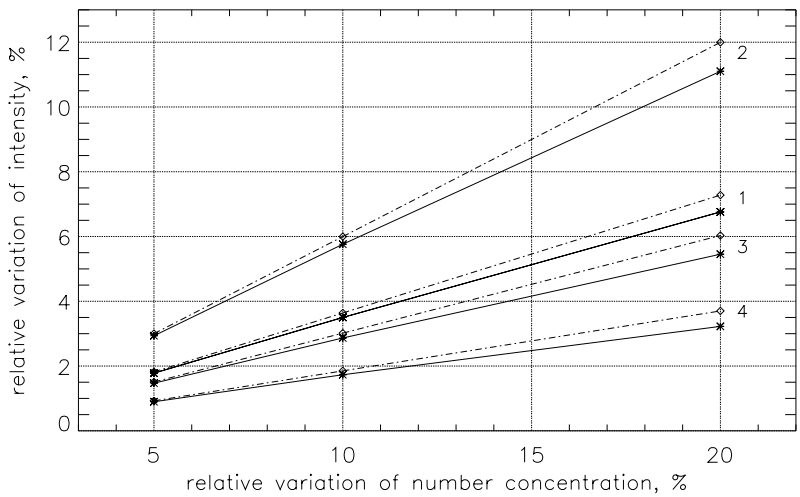



**Fig. 6.15.** Comparison of the relative variations of the reflected intensity caused by variations of the particle number concentration of water droplets. Solid line represents the results of numerical calculations and dash-dotted line corresponds to the linear approximation. The calculations were performed for a solar zenith angle of 45° and a surface albedo of 0.3 at the following wavelengths: 1, 758.2 nm; 2, 760.425 nm; 3, 1.6 μm; 4, 2.2 μm.

of the effective radius of water droplets results in the linear approximation in a variation of the intensity of about 1.8% whereas the exact value is even negative.

## 6.11 Application to the retrieval of the effective radius of water droplets

The weighting functions $\mathcal{W}_N(z; z_v, \Omega_v)$ and $\mathcal{W}_r^N(z; z_v, \Omega_v)$ can be used for the retrieval of the vertical profiles of $N(z)$ and $r(z)$ within a cloud. However, these functions considered as functions of the wavelength are highly correlated and a simultaneous retrieval of both parameters, i.e., $N(z)$ and $r_1(z)$ is a very complicated matter. Therefore, it is reasonable to assume that the optical thickness of the cloud, $\tau_c$, can be estimated using the measured reflected radiation in the spectral range, where the gaseous and liquid water absorption is weak [7]. Let us use the retrieved optical thickness to estimate the particle number concentration. Obviously, we can not retrieve the profile of $N(z)$ since in the spectral ranges with weak gaseous and liquid water absorption the WF for the particle number concentration is almost independent of the altitude within the cloud (see Fig. 6.12). Therefore, we will assume that the shape of the particle number concentration is known *a priori* and only the scaling factor needs to be estimated. Namely, the unknown profile of the particle number concentration, $N'(z)$, will be approximated as follows:

$$N'(z) = CN(z)\,, \tag{6.180}$$

where $N(z)$ is the known shape of the particle number concentration and $C$ is the scaling factor. Using (6.143) for the optical thickness of the cloud, we obtain the estimation for the scaling factor $C$ as follows:

$$\hat{C} = \frac{\tau_c}{\int\limits_{z_b}^{z_t} e_1'(z)\,\mathrm{d}z}\,, \tag{6.181}$$

where $\tau_c$ is the estimated value of the cloud optical thickness. The extinction coefficient of water droplets, $e_1'(z)$, corresponds to the known shape of the particle number concentration $N(z)$ and an unknown profile of the effective radius $r_1'(z)$, which should be estimated as well. To obtain $\hat{C}$ according to (6.181), we take into account that in the linear approximation the extinction coefficient, $e_1'(z)$, corresponding to an unknown effective radius $r_1'(z)$ can be represented as follows:

$$e_1'(z) = e_1(z) + \frac{\partial e_1(z)}{\partial r_1(z)}\delta r_1(z)\,, \tag{6.182}$$

where the extinction coefficient, $e_1(z)$, corresponds to the known profile of the effective radius, $r_1(z)$, and $\delta r_1(z) = r_1'(z) - r_1(z)$. Substituting $e_1'(z)$ as given by (6.182) into (6.181), we obtain

$$\hat{C} = \frac{\tau_c}{\int\limits_{z_b}^{z_t} \Big[ e_1(z) + \Delta e_1(z) \Big] \, \mathrm{d}z} \,. \tag{6.183}$$

Here, the variation of the extinction coefficient, $\Delta e_1(z)$, is caused by the variation of the effective radius, $\delta r_1(z)$, i.e.,

$$\Delta e_1(z) = \frac{\partial e_1(z)}{\partial r_1(z)} \delta r_1(z) \,. \tag{6.184}$$

Assuming that $\Delta e_1(z)$ is small compared to $e_1(z)$, it follows in the linear approximation:

$$\hat{C} = \frac{\tau_c}{\int\limits_{z_b}^{z_t} e_1(z) \, \mathrm{d}z} - \frac{\tau_c}{\Big[ \int_{z_b}^{z_t} e_1(z) \, \mathrm{d}z \Big]^2} \int_{z_b}^{z_t} \Delta e_1(z) \, \mathrm{d}z \,. \tag{6.185}$$

The first term in this equation corresponds to the estimation of the scaling factor obtained ignoring the variation of the effective radius. Introducing the following definition for this scaling factor:

$$C[r_1(z)] = \frac{\tau_c}{\int\limits_{z_b}^{z_t} e_1(z) \, \mathrm{d}z} \,, \tag{6.186}$$

and substituting it into (6.185), we obtain

$$\hat{C} = C[r_1(z)] - \frac{C[r_1(z)]}{\int_{z_b}^{z_t} e_1(z) \, \mathrm{d}z} \int_{z_b}^{z_t} \frac{\partial e_1(z)}{\partial r_1(z)} \delta r_1(z) \, \mathrm{d}z \,. \tag{6.187}$$

Introducing for the simplification the function $E_1(z)$ as

$$E_1(z) = -\frac{1}{\int_{z_b}^{z_t} e_1(z) \, \mathrm{d}z} \frac{\partial e_1(z)}{\partial r_1(z)} r_1(z) \,, \tag{6.188}$$

Eq. (6.187) can be rewritten in the following form:

$$\hat{C} = C[r_1(z)] + C[r_1(z)] \int_{z_b}^{z_t} E_1(z) v_r(z) \, \mathrm{d}z \,, \tag{6.189}$$

where $v_r(z)$ is the relative variation of the effective radius of water droplets. Substituting now $\hat{C}$ given by (6.189) into (6.180) instead of $C$ and introducing the zero-order estimation of the particle number concentration obtained ignoring the variation of the effective radius as

$$\bar{N}(z) = C[r_1(z)] N(z) \,, \tag{6.190}$$

we obtain an estimation for the particle number concentration accounting for the variation of the effective radius as follows:

$$N'(z) = \bar{N}(z) + \bar{N}(z) \int_{z_b}^{z_t} E_1(z) v_r(z)\, \mathrm{d}z\,. \tag{6.191}$$

This equation provides the following relationship between the relative variation of the particle number concentration around $\bar{N}(z)$ and the relative variation of the effective radius of water droplets:

$$v_N(z) = \frac{N'(z) - \bar{N}(z)}{\bar{N}(z)} = \int_{z_b}^{z_t} E_1(z) v_r(z)\, \mathrm{d}z\,. \tag{6.192}$$

The variation of the measured functional, $\delta\Phi(z_v, \Omega_v)$, in the spectral range containing absorption structures of the liquid water can be written now as follows:

$$\begin{aligned} \delta\Phi(z_v, \Omega_v) &= \Phi'(z_v, \Omega_v) - \Phi(z_v, \Omega_v) \\ &= \int_{z_b}^{z_t} \left[ \mathcal{W}_N(z; z_v, \Omega_v) v_N(z) + \mathcal{W}_r^N(z; z_v, \Omega_v) v_r(z) \right] \sigma_e(z)\, \mathrm{d}z\,, \end{aligned} \tag{6.193}$$

where $\Phi'(z_v, \Omega_v)$ is the measured value and $\Phi(z_v, \Omega_v)$ is calculated for the particle number concentration $\bar{N}(z)$ and the effective radius $r_1(z)$. Substituting further $v_N(z)$ given by (6.192) into (6.193), we obtain the expression for the variation of the measured functional, $\delta\Phi(z_v, \Omega_v)$, containing only the effective radius of water droplets as unknown parameter:

$$\begin{aligned} \delta\Phi(z_v, \Omega_v) &= \int_{z_b}^{z_t} \left[ \mathcal{W}_N(z; z_v, \Omega_v) \int_{z_b}^{z_t} E_1(z') v_r(z') dz' \right] \sigma_e(z)\, \mathrm{d}z \\ &+ \int_{z_b}^{z_t} \mathcal{W}_r^N(z; z_v, \Omega_v) v_r(z) \sigma_e(z)\, \mathrm{d}z\,. \end{aligned} \tag{6.194}$$

Introducing the specific weighting function for the effective radius in the following form:

$$\begin{aligned} \mathcal{W}_r^{\tau_c}(z; z_v, \Omega_v) &= \mathcal{W}_r^N(z; z_v, \Omega_v) \sigma_e(z) \\ &+ E_1(z) \int_{z_b}^{z_t} \mathcal{W}_N(z; z_v, \Omega_v) \sigma_e(z)\, \mathrm{d}z\,, \end{aligned} \tag{6.195}$$

we obtain

$$\delta\Phi(z_v, \Omega_v) = \int_{z_b}^{z_t} \mathcal{W}_r^{\tau_c}(z; z_v, \Omega_v) v_r(z)\, \mathrm{d}z\,. \tag{6.196}$$

This equation can be employed to estimate the vertical profile of the effective radius of water droplets using the measured radiation in the spectral range containing liquid water absorption bands, e.g., around 1.6, 2.2 or 3.7 µm.

## 6.12 Weighting functions for cloud geometrical parameters

In this section we derive the analytical expressions for geometrical cloud parameters such as the cloud top and cloud bottom height defining the position of a cloud layer in the atmosphere.

### 6.12.1 Theory

Let us assume that the variation of the measured functional, $\delta\Phi$, is caused by the variation of the cloud scattering, $s_k(\tau)$, and absorption, $a_k(\tau)$, coefficients only. Then according to (6.68) we have

$$\delta\Phi(\tau_v,\Omega_v)=\sum_{k=1}^{2}\int_0^{\tau_0}\left[\mathcal{W}_{s_k}(\tau;\tau_v,\Omega_v)v_{s_k}(\tau)+\mathcal{W}_{a_k}(\tau;\tau_v,\Omega_v)v_{a_k}(\tau)\right]\mathrm{d}\tau\,, \quad (6.197)$$

where $v_{s_k}(\tau)=\delta s_k(\tau)/s_k(\tau)$ and $v_{a_k}(\tau)=\delta a_k(\tau)/a_k(\tau)$ are the relative variations of the cloud scattering and absorption coefficients, respectively, and subscript $k=1$ corresponds to water droplets and $k=2$ to ice crystals. The expressions for WFs $\mathcal{W}_{a_k}(\tau;\tau_v,\Omega_v)$ and $\mathcal{W}_{s_k}(\tau;\tau_v,\Omega_v)$ are given by (6.92) and (6.97), respectively. To introduce the cloud geometrical parameters we rewrite at first (6.197) in the form containing the absolute variation of the cloud absorption and scattering coefficients as well as the integration over the altitude $z$ instead of the optical depth $\tau$:

$$\begin{aligned}\delta\Phi(z_v,\Omega_v) &= \sum_{k=1}^{2}\int_0^{H}\left[\mathcal{W}_{s_k}(z;\tau_v,\Omega_v)\frac{\delta s_k(z)}{s_k(z)}\right.\\ &+ \left.\mathcal{W}_{a_k}(z;\tau_v,\Omega_v)\frac{\delta a_k(z)}{a_k(z)}\right]\sigma_e(z)\,\mathrm{d}z\,, \quad (6.198)\end{aligned}$$

where $H$ is the top of atmosphere altitude. Taking into account that the optical parameters of the cloud are non-zero within the cloud layer only, we can represent them in the following form:

$$p_k(z)=\Theta(z_t-z)\Theta(z-z_b)p_k(z)\,, \quad (6.199)$$

which allows us to define the cloud parameters in the entire atmosphere. Here, $z_t$ and $z_b$ are altitudes of the cloud top and bottom, respectively, function $p_k(z)$ is assumed to be the scattering or the absorption coefficient of the cloud, and functions $\Theta(z_t-z)$ and $\Theta(z-z_b)$ are the Heaviside step-functions given by (6.18).

Employing (6.199), variations of the cloud optical parameters caused by variations of the cloud top and bottom heights can be expressed in the linear approximation as follows:

$$\begin{aligned}\delta p_k(z) &= \left[\frac{\mathrm{d}\Theta(z_t-z)}{\mathrm{d}z_t}p_k(z)+\Theta(z_t-z)\frac{\mathrm{d}p_k(z)}{\mathrm{d}z_t}\right]\Theta(z-z_b)\Delta z_t\\ &+ \left[\frac{\mathrm{d}\Theta(z-z_b)}{\mathrm{d}z_b}p_k(z)+\Theta(z-z_b)\frac{\mathrm{d}p_k(z)}{\mathrm{d}z_b}\right]\Theta(z_t-z)\Delta z_b\,. \quad (6.200)\end{aligned}$$

Here, $dp_k(z)/dz_t$ and $dp_k(z)/dz_b$ are derivatives of the cloud parameters with respect to the cloud top and bottom height, respectively, $\Delta z_t=z_t'-z_t$ and $\Delta z_b=z_b'-z_b$ are the variations of CTH and CBH. Derivatives of the Heaviside

step-functions with respect to the cloud top height, $z_t$, and the cloud bottom height, $z_b$, can be obtained analytically [9]:

$$\frac{\mathrm{d}\Theta(z_t - z)}{\mathrm{d}z_t} = \delta(z_t - z)\,, \tag{6.201}$$

$$\frac{\mathrm{d}\Theta(z - z_b)}{\mathrm{d}z_b} = -\delta(z_b - z)\,, \tag{6.202}$$

where $\delta(z_t - z)$ and $\delta(z_b - z)$ are the Dirac $\delta$-functions. Substituting these expressions into (6.200), dividing both sides of this equation by $p_k(z)$ and introducing for the simplification the following functions:

$$t_{p_k}(z) = \left[\delta(z_t - z) + \Theta(z_t - z)\frac{1}{p_k(z)}\frac{\mathrm{d}p_k(z)}{\mathrm{d}z_t}\right]\Theta(z - z_b)\,, \tag{6.203}$$

$$b_{p_k}(z) = \left[-\delta(z_b - z) + \Theta(z - z_b)\frac{1}{p_k(z)}\frac{\mathrm{d}p_k(z)}{\mathrm{d}z_b}\right]\Theta(z_t - z)\,, \tag{6.204}$$

we obtain

$$\frac{\delta p_k(z)}{p_k(z)} = t_{p_k}(z)\Delta z_t + b_{p_k}(z)\Delta z_b\,. \tag{6.205}$$

Substituting further these expressions into (6.198), we obtain the variation of the measured functional as a linear function with respect to the variations of CTH and CBH:

$$\delta\Phi(z_v, \Omega_v) \;=\; \int_0^H \Big[\mathcal{W}_{z_t}(z; \tau_v, \Omega_v)\Delta z_t + \mathcal{W}_{z_b}(z; \tau_v, \Omega_v)\Delta z_b\Big]\sigma_e(z)\,\mathrm{d}z\,, \tag{6.206}$$

where functions $\mathcal{W}_{z_t}(z; \tau_v, \Omega_v)$ and $\mathcal{W}_{z_b}(z; \tau_v, \Omega_v)$ are defined as follows:

$$\mathcal{W}_{z_t}(z; \tau_v, \Omega_v) = \sum_{k=1}^{2}\Big[t_{s_k}(z)\mathcal{W}_{s_k}(z; \tau_v, \Omega_v) + t_{a_k}(z)\mathcal{W}_{a_k}(z; \tau_v, \Omega_v)\Big]\,, \tag{6.207}$$

$$\mathcal{W}_{z_b}(z; \tau_v, \Omega_v) = \sum_{k=1}^{2}\Big[b_{s_k}(z)\mathcal{W}_{s_k}(z; \tau_v, \Omega_v) + b_{a_k}(z)\mathcal{W}_{a_k}(z; \tau_v, \Omega_v)\Big]\,, \tag{6.208}$$

and functions $t_{s_k}(z)$, $t_{a_k}(z)$ and $b_{s_k}(z)$, $b_{a_k}(z)$ are given by (6.203) and (6.204) for $p_k(z) = s_k(z)$ and $p_k(z) = a_k(z)$, respectively. Thus, we obtain the following expressions for CTH and CBH weighting functions:

$$\mathcal{W}_{z_t}(z_v, \Omega_v) = \int_0^H \mathcal{W}_{z_t}(z; \tau_v, \Omega_v)\sigma_e(z)\,\mathrm{d}z\,, \tag{6.209}$$

$$\mathcal{W}_{z_b}(z_v, \Omega_v) = \int_0^H \mathcal{W}_{z_b}(z; \tau_v, \Omega_v)\sigma_e(z)\,\mathrm{d}z\,. \tag{6.210}$$

The derived expressions can be used to calculate the weighting functions after the expressions for functions $t_{p_k}(z)$ and $b_{p_k}(z)$ containing derivatives of the cloud

optical parameters with respect to CTH and CBH having been obtained. For this reason we will first derive a general expression for derivatives of the cloud optical parameters with respect to CTH assuming that the actual value of a certain parameter is obtained scaling the known shape profile:

$$p_k(z) = c_k p_h(z)\,, \tag{6.211}$$

where $p_h(z)$ is the shape profile of the corresponding parameter and $c_k$ is a scaling factor. The scaling factor is obtained to match the corresponding integral parameter such as the optical thickness or liquid water path of a cloud.

The derivative $\partial p_k(z)/\partial z_t$ can be obtained using (6.211) as follows:

$$\frac{\partial p_k(z)}{\partial z_t} = \frac{\partial}{\partial z_t}\Big[c_k p_h(z)\Big] = \left[\frac{\partial c_k}{\partial z_t} p_h(z) + c_k \frac{\partial p_h(z)}{\partial z_t}\right]. \tag{6.212}$$

Assuming that the scaling factor is chosen to match the optical thickness of the cloud, the derivative $\partial c_k/\partial z_t$ can be found as

$$\frac{\partial c_k}{\partial z_t} = \frac{\partial}{\partial z_t}\Big[\frac{\tau_c}{\tau_{p_h}}\Big] = -\frac{\tau_c}{\tau_{p_h}^2}\frac{\partial \tau_{p_h}}{\partial z_t}\,, \tag{6.213}$$

where we have taken into account that $\tau_c$ is a given constant value of the cloud optical thickness. To find the analytical expression for the derivative $\partial\tau_{p_h}/\partial z_t$ we use the following relationship:

$$\int_0^1 p_h(x)\,\mathrm{d}x \equiv \frac{1}{L}\int_{z_b}^{z_t} p_h(z)\,\mathrm{d}z = \frac{\tau_{p_h}}{L}\,, \tag{6.214}$$

where $x$ is the dimensionless altitude given by (6.126), and $L = z_t - z_b$ is the geometrical thickness of the cloud. Differentiating (6.214) with respect to the parameter $z_t$, we obtain

$$-\frac{1}{L^2}\tau_{p_h} + \frac{1}{L}\frac{\partial \tau_{p_h}}{\partial z_t} = 0 \quad\Longrightarrow\quad \frac{\partial \tau_{p_h}}{\partial z_t} = \frac{\tau_{p_h}}{L}\,. \tag{6.215}$$

Substituting the obtained result into (6.213), we have

$$\frac{\partial c_k}{\partial z_t} = -\frac{\tau_c}{\tau_{p_h} L}\,. \tag{6.216}$$

For the derivative $\partial p_h(z)/\partial z_t$ we obtain

$$\frac{\partial p_h(z)}{\partial z_t} = \frac{\partial p_h(z)}{dz}\,\frac{\partial z}{\partial x}\,\frac{\partial x}{\partial z_t} = -\frac{\partial p_h(z)}{\partial z}(1-x)\,. \tag{6.217}$$

Taking into account that the dimensionless altitude $x$ is given by (6.126) and substituting it into the previous equation, we obtain

$$\frac{\partial p_h(z)}{\partial z_t} = -\frac{\partial p_h(z)}{\partial z}\frac{z - z_b}{L}\,. \tag{6.218}$$

Substituting (6.216) and (6.218) into (6.212), it follows that

$$\frac{\partial p_k(z)}{\partial z_t} = \left[ -\frac{\tau_c}{\tau_{p_h} L} p_h(z) - c_k \frac{\partial p_h(z)}{\partial z} \frac{z - z_b}{L} \right] . \tag{6.219}$$

Taking into account (6.211), we obtain the final expression for the derivative of the optical parameters with respect to CTH as follows:

$$\frac{\partial p_k(z)}{\partial z_t} = -\frac{1}{L} \left[ p_k(z) + \frac{\partial p_k(z)}{\partial z} (z - z_b) \right] . \tag{6.220}$$

Employing the same approach, the derivative of the optical parameters with respect to CBH is obtained in the following form:

$$\frac{\partial p_k(z)}{\partial z_b} = \frac{1}{L} \left[ p_k(z) - \frac{\partial p_k(z)}{\partial z} (z_t - z) \right] . \tag{6.221}$$

The first term in the derived expressions results from the dependence of the scaling factor on the cloud geometrical parameters and the second contains the derivative of the corresponding parameter with respect to the altitude. Substituting (6.220) and (6.221) into (6.203) and (6.204), we obtain the final expressions for functions $t_{p_k}(z)$ and $b_{p_k}(z)$ in the following form:

$$\begin{aligned} t_{p_k}(z) &= \left\{ \delta(z_t - z) - \frac{\Theta(z_t - z)}{L} \left[ 1 + \frac{\partial \ln p_k(z)}{\partial z} (z - z_b) \right] \right\} \\ &\times \Theta(z - z_b) , \end{aligned} \tag{6.222}$$

$$\begin{aligned} b_{p_k}(z) &= \left\{ -\delta(z_b - z) + \frac{\Theta(z - z_b)}{L} \left[ 1 - \frac{\partial \ln p_k(z)}{\partial z} (z_t - z) \right] \right\} \\ &\times \Theta(z_t - z) . \end{aligned} \tag{6.223}$$

In some special cases the obtained expressions can be simplified. For example, for a vertically homogeneous cloud derivatives of the cloud scattering and absorption coefficients with respect to the altitude are zero. Thus, we obtain the following expressions for functions $t_{p_k}(z)$ and $b_{p_k}(z)$:

$$t_{p_k}(z) = \left[ \delta(z_t - z) - \frac{\Theta(z_t - z)}{L} \right] \Theta(z - z_b) , \tag{6.224}$$

$$b_{p_k}(z) = \left[ -\delta(z_b - z) + \frac{\Theta(z - z_b)}{L} \right] \Theta(z_t - z) . \tag{6.225}$$

Further simplification can be obtained if the optical parameters in an vertically homogeneous cloud are not rescaled for varying geometrical parameters. In this case we have

$$t_{p_k}(z) = \delta(z_t - z) \Theta(z - z_b) , \tag{6.226}$$

$$b_{p_k}(z) = -\delta(z_b - z) \Theta(z_t - z) , \tag{6.227}$$

and the weighting functions for CTH and CBH given by (6.209) and (6.210) can be rewritten as follows:

$$\mathcal{W}_{z_t}(z_v,\Omega_v)=\sum_{k=1}^{2}\Big[\mathcal{W}_{s_k}(z_t;\tau_v,\Omega_v)+\mathcal{W}_{a_k}(z_t;\tau_v,\Omega_v)\Big]\sigma_e(z_t)\,, \tag{6.228}$$

$$\mathcal{W}_{z_b}(z_v,\Omega_v)=-\sum_{k=1}^{2}\Big[\mathcal{W}_{s_k}(z_b;\tau_v,\Omega_v)+\mathcal{W}_{a_k}(z_b;\tau_v,\Omega_v)\Big]\sigma_e(z_b)\,. \tag{6.229}$$

Having defined the weighting functions for CTH and CBH, we can easily obtain WFs for other parameters characterizing the cloud geometry. In particular, the WF for the geometrical thickness of the cloud is

$$\mathcal{W}_{gt}(z_v,\Omega_v)=\mathcal{W}_{z_t}(z_v,\Omega_v)-\mathcal{W}_{z_b}(z_v,\Omega_v)\,. \tag{6.230}$$

Assuming that the cloud geometrical thickness is constant, we can obtain the WF characterizing the position of the cloud in the atmosphere as well. In this case the shift of a cloud can be described by simultaneous variations of the same magnitude of both CTH and CBH. Therefore, WF for the position of a cloud can be written as follows:

$$\mathcal{W}_{sh}(z_v,\Omega_v)=\mathcal{W}_{z_t}(z_v,\Omega_v)+\mathcal{W}_{z_b}(z_v,\Omega_v)\,. \tag{6.231}$$

### 6.12.2 Numerical results

In this subsection we consider examples of WFs for different cloud geometrical parameters corresponding to vertically homogeneous and inhomogeneous water clouds as well as to mixed clouds consisting of water droplets and ice crystals. At first we compare WFs for CTH and CBH obtained employing the analytical expressions derived in the previous subsection to results of the numerical calculations. For this purpose we consider the most general case of a vertically inhomogeneous mixed cloud consisting of water droplets and ice crystals. Corresponding vertical profiles of the liquid and ice water contents are shown in Fig. 6.16. The vertical profiles of the effective radius of water droplets and ice crystals were assumed to change linearly within the cloud. At the top of the cloud the effective radii of water droplets and ice crystals were set to 6 µm and 70 µm, respectively, and at the bottom of the cloud to 2 µm and 120 µm, respectively. The scattering and absorption coefficients of water droplets and ice crystals were calculated according to the analytical expressions given by (6.139)–(6.142). The numerical WFs were obtained using the numerical perturbation approach as follows:

$$\mathcal{W}_{z_t}(z_v,\Omega_v)=\frac{I(0,\Omega_v,z_t+\Delta z_t)-I(0,\Omega_v,z_t)}{\Delta z_t}\,, \tag{6.232}$$

where $I(0,\Omega_v,z_t)$ and $I(0,\Omega_v,z_t+\Delta z_t)$ are the reflected intensities at the top of the atmosphere corresponding to the cloud top height $z_t$ and $z_t+\Delta z_t$, respectively, and $\Delta z_t$ is the perturbation of the cloud top height which was set to 0.01 km.

The simulated spectrum of the reflected solar radiation observed at TOA in the nadir viewing geometry in the presence of a homogeneous water cloud

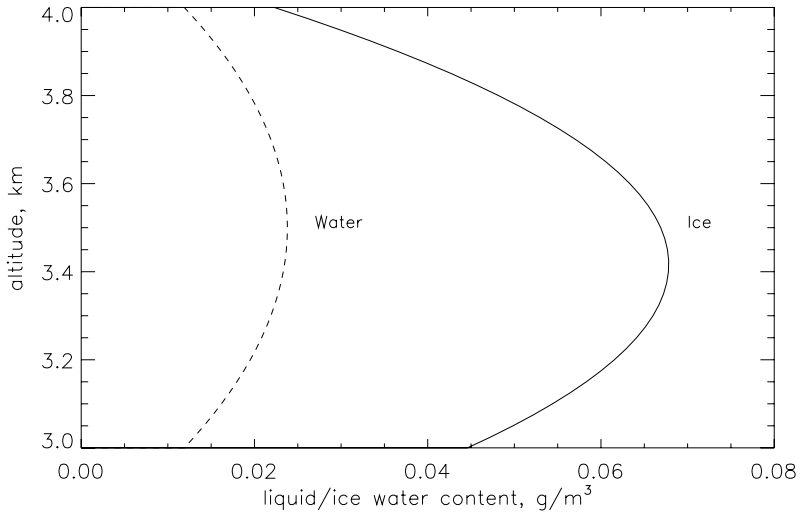



**Fig. 6.16.** Vertical profiles of the liquid and ice water contents for a mixed cloud having an optical thickness of 10.

having an optical thickness of 10 is shown in Fig. 6.17. The calculations were performed for a finite spectral resolution assuming the instrument slit function to be the boxcar function with a full width of 0.1 nm. Relative differences between the reflected intensities corresponding to a homogeneous and an inhomogeneous water cloud as well as to a homogeneous and a mixed cloud are shown in the lower panel of Fig. 6.17. As can clearly be seen, at spectral points where the gaseous absorption is week the difference between reflected intensities corresponding to a vertically homogeneous and a vertically inhomogeneous water cloud is very small whereas, due to a difference in the phase functions, the reflected intensity in a presence of a mixed cloud is approximately 2% higher than for a homogeneous water cloud of the same optical thickness.

Similarly to the parameters discussed in previous sections let us introduce the normalized WFs for CTH and CBH which are numerically equivalent to relative variations of the observed intensity caused by 1 km variations of CTH or CBH, respectively. For example, the normalized weighting function for CTH is introduced as

$$\frac{\Delta I(\tau_v, \Omega_v)}{I(\tau_v, \Omega_v)} = \frac{\mathcal{W}_{z_t}(z_v, \Omega_v)}{I(\tau_v, \Omega_v)} \Delta z_t = \mathcal{R}_{z_t}(z_v, \Omega_v) \Delta z_t \,. \tag{6.233}$$

The normalized weighting functions for CTH and CBH in the spectral range of the $O_2$-A absorption band obtained employing analytical expressions, Eqs (6.209) and (6.210), as well as calculated using the numerical perturbation approach as given by (6.232) are shown in Fig. 6.18. Good agreement between the numerical and analytical weighting functions confirms the correctness of the derived general analytical expressions. Although the overall wavelength behavior of both weighting functions is similar, it is clearly seen that the fine spectral structure is

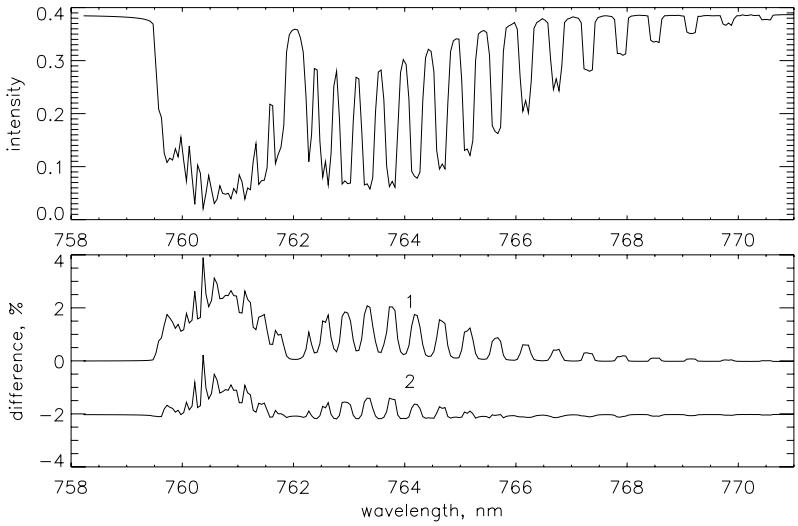



**Fig. 6.17.** Upper panel: reflected intensity at TOA for a homogeneous water cloud. Lower panel: relative differences between the reflected intensities for: 1, homogeneous and inhomogeneous water cloud, 2, homogeneous and mixed cloud ($\tau_{\text{water}} = 9$, $\tau_{\text{ice}} = 1$). Calculations were performed for a solar zenith angle of 45° and a surface albedo of 0.3.

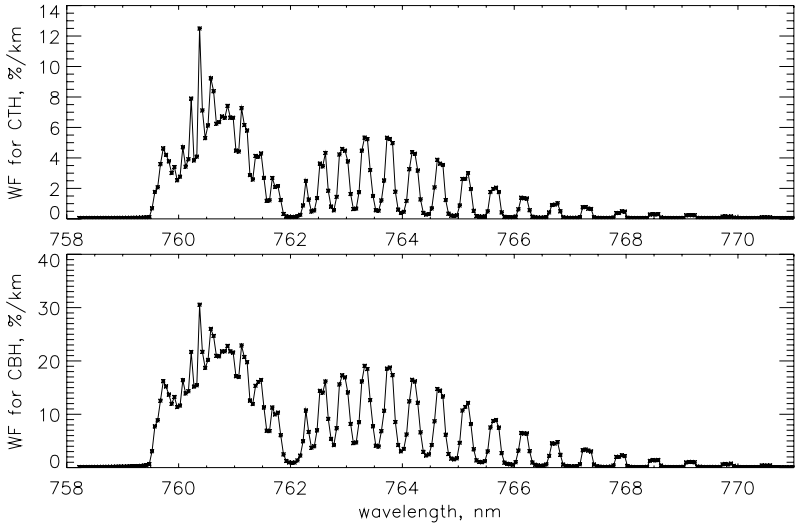



**Fig. 6.18.** Comparison of the numerical (symbols) and the normalized layer integrated analytical (solid line) WFs for the cloud top (upper panel) and cloud bottom height (lower panel) in a presence of a mixed cloud of optical thickness 10. The calculations were performed for a solar zenith angle of 45° and a surface albedo of 0.3.

significantly different, i.e, the spectral structures of the reflected intensity variations caused by variations of CTH and CBH are different and, thus, these parameters can be considered to be independent. As seen from the plot, CTH and CBH weighting functions are very small in spectral intervals where $O_2$ absorption is weak which is in accordance with the well known fact that in a non-absorbing atmosphere the reflection of the cloud is independent of its altitude and geometrical thickness. Since both weighting functions are always positive, an increase of CTH or CBH under the assumption of constant optical thickness leads to an enhancement in the reflected intensity. In particular, as follows from Fig. 6.18, the relative variation of the reflected intensity due to a 1 km shift of CTH or CBH reaches about 13% and 30%, respectively. The reflected intensity is most sensitive to variations of CTH and CBH in spectral ranges where the gaseous absorption is strong.

In the most general case of a vertically inhomogeneous mixed cloud the weighting functions for CTH and CBH contain derivatives of vertical profiles of the scattering and absorption coefficients with respect to the altitude (see Eqs (6.222) and (6.223) for the auxiliary functions $t_{p_k}(z)$ and $b_{p_k}(z)$, respectively). Therefore, WFs depend on the vertical structure of the cloud, i.e., on the vertical distribution of the scattering and absorption coefficients within the cloud. Nevertheless, due to a lack of information on the cloud structure, retrievals of the cloud top height from real measurements are often performed assuming the cloud to be vertically homogeneous. The validity of this approximation can be investigated analyzing the influence of the cloud vertical inhomogeneity on the weighting functions. This was done comparing CTH WF for a vertically homogeneous water cloud to that for vertically inhomogeneous water and mixed clouds. If the geometrical thickness of the cloud is known, the inverse problem can be simplified because the variation of the reflected intensity is resulted only from a variation of the vertical position of the cloud, i.e., of the cloud altitude, and the corresponding WF is given by (6.231). The cloud altitude weighting functions describing the simplified inverse problem were also compared in the similar manner as CTH WFs. The comparison was performed for the cloud optical thickness of 10. For mixed clouds the optical thicknesses of water droplets and ice crystals were selected to be 9 and 1, respectively.

Figure 6.19 illustrates the sensitivity of the weighting functions for the cloud top height (curves 1 and 2) and the cloud altitude (curves 3 and 4) to the vertical inhomogeneity of a cloud. The ratios of WFs for a vertically homogeneous water cloud to WFs for an inhomogeneous water cloud are shown by dashed lines and the ratios of WFs for a vertically homogeneous water cloud to WFs for an inhomogeneous mixed cloud are represented by solid lines. As can clearly be seen, in the considered observation geometry CTH WF for a vertically homogeneous cloud is much larger than that for both inhomogeneous water clouds and mixed clouds, whereas in the case of a constant geometrical thickness the influence of the vertical cloud inhomogeneity is substantially weaker. Therefore, an estimation of the cloud top height under the assumption of the vertical homogeneity of a cloud is more accurate in the case of a known geometrical thickness of the cloud.

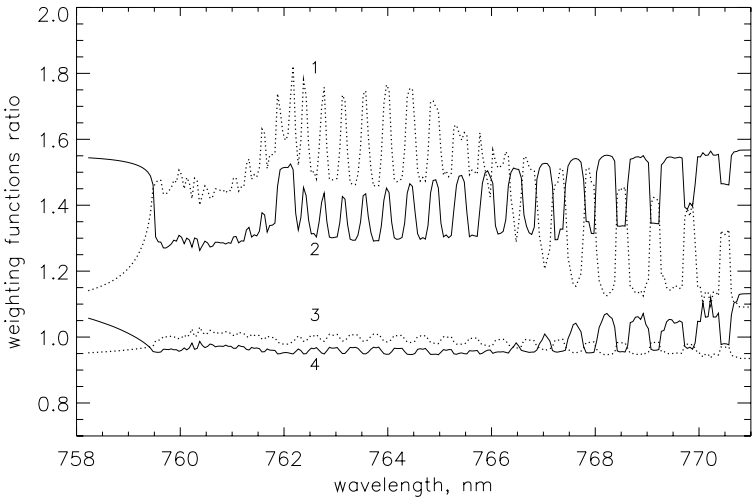



**Fig. 6.19.** Comparison of the weighting functions for the cloud top height (curves 1 and 2) and the cloud altitude (curves 3 and 4). Dashed lines: ratios of WFs for a vertically homogeneous water cloud to WFs for a inhomogeneous water cloud. Solid lines: ratios of WFs for a vertically homogeneous water cloud to WFs for a inhomogeneous mixed cloud.

A variation of the reflected intensity in a cloudy atmosphere can be caused not only by variations of the cloud altitude, cloud top or bottom height, or vertical distributions of cloud parameters, but also by variations of the pressure and temperature. In spectral regions where the gaseous absorption is not negligible, variations of the pressure and temperature in the atmosphere cause variations of the absorption cross-sections of atmospheric gases and, thus, alter the absorption of the solar radiation. Fig. 6.20 illustrates contributions into the relative variation of the observed intensity due to 1% variation of the pressure, 1 K variation of the temperature, and 100 m variation of the cloud top height. As clearly seen, all plotted variations are of the same magnitude providing an estimation for the error propagation in the case of uncertain atmospheric parameters.

Concluding the discussion of WFs for cloud geometrical parameters let us consider the cloud top height weighting functions for a two-layered cloud system. The cloud system was assumed to comprise an ice crystal cloud with a cloud top height of 8 km in the upper layer and a water cloud with a cloud top height of 4 km in the lower layer. Both cloud layers were assumed to have a geometrical thickness of 1 km. Corresponding WFs in the $O_2$-A absorption band are shown in Fig. 6.21 for different optical thicknesses of cloud layers. As can be seen from the plot, the spectral structure and magnitude of CTH WFs for the upper and lower cloud layers are strongly dependent on the distribution of the optical thickness between the cloud layers. For example, if the optical thickness of the upper layer is much smaller than that of the lower layer, 1 km variation of the cloud top height of the lower layer results in a relative variation of the reflected intensity of

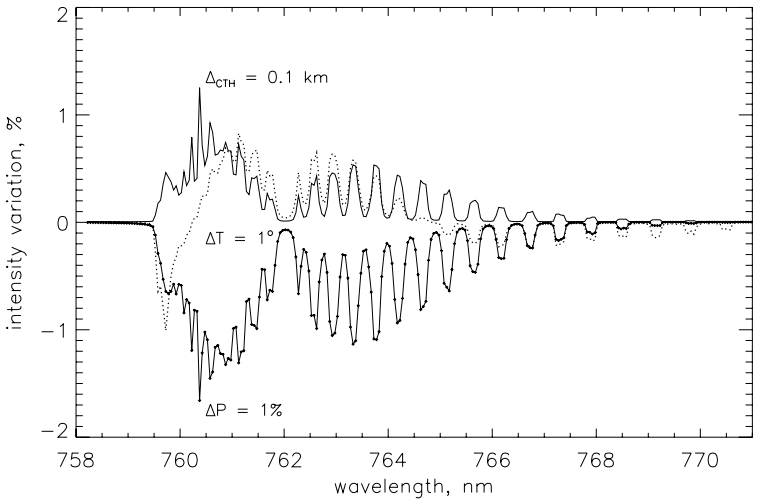



**Fig. 6.20.** Relative variations of the reflected intensity in a presence of a mixed cloud due to 1% variation of the pressure, 1 K variation of the temperature, and 100 m variation of the cloud top height.

about 11% at wavelengths around 760.4 nm (upper panel in Fig. 6.21), whereas, the relative variation of the reflected intensity caused by the same variation of CTH of the upper layer is about a factor of 5 smaller. An increase in the optical thickness of the upper cloud layer for a constant total optical thickness leads to an increase in CTH WF of the upper cloud layer and a decrease in CTH WF of the lower cloud layer (lower panel in Fig. 6.21). Thus, in the case of a two-layered cloud system the reflected solar radiance observed within a gaseous absorption band spectral range contains an information about the cloud top heights of both cloud layers. However, the retrieval of these parameters requires independent information about optical thicknesses of the upper and lower cloud layers.

## 6.13 Conclusion

Generally, to find a solution of an inverse problem, the weighting functions of the atmospheric parameters of interest are required. In this chapter we have extensively discussed the weighting functions not only for common atmospheric parameters, such as atmospheric trace gas number densities, pressure, temperature, and aerosol particle number density, but also for various cloud parameters, such as geometrical thickness, cloud top and bottom height, liquid water content and effective radius of water droplets and ice crystals. All obtained expressions for the weighting functions are implemented in the software package SCIATRAN 2.1 [26] and verified against the numerical perturbation technique. SCIATRAN 2.1 is freely available for non-commercial use at *www.iup.physik.uni-bremen.de/sciatran*. A brief description of the mathematical background of the

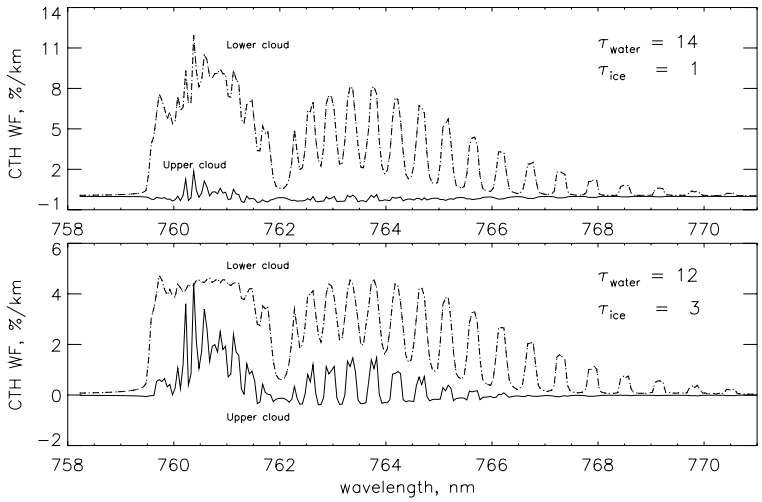



**Fig. 6.21.** Cloud top height weighting functions for a two-layered cloud system. The calculations were performed for the reflected intensity observed at TOA at a solar zenith angle of 45° and a surface albedo of 0.3. The cloud system was assumed to comprise an ice crystal cloud with a cloud top height of 8 km in the upper layer and a water cloud with a cloud top height of 4 km in the lower layer. Both cloud layers were assumed to have a geometrical thickness of 1 km. The total optical thickness of the cloud system was set to 15.

linearized radiative transfer equation, adjoint radiative transfer equation, and adjoint approach, as well as detailed derivation of the expressions for the weighting functions presented in this chapter, is aimed to facilitate the usage of the software package SCIATRAN 2.1 to solve various practical inverse problems.

The fact that, due to linearization errors, most of inverse problems need to be solved iteratively requiring an update of the forward intensity and weighting functions at each iterative step, was accounted for in the SCIATRAN 2.1 software package coupling the retrieval block with the radiative transfer model. The retrieval block, which was originally developed to retrieve vertical distributions or column amounts of atmospheric species, incorporates various inversion techniques such as the optimal estimation [19], Tikhonov regularization [28], and information operator approach [10], and can be adapted by a user to solve various inverse problems arising in remote sensing of the Earth's atmosphere. One of the retrieval block extensions already implemented in the inversion procedure is a retrieval of the cloud top height using satellite measurements of the backscattered solar radiation in the oxygen absorption A-band [21]. Some examples of successful applications of the SCIATRAN software package to the retrieval of vertical profiles of $NO_2$ and BrO as well as vertical columns of $CH_4$ and $CO_2$ using SCIAMACHY measurements are presented in [3, 23].

**Acknowledgments**

This work has been funded in parts by the German Ministry of Education and Research BMBF (grant 07UFE12/8) and the German Aerospace Center DLR (grant 50EE0027).

## Appendix A: Derivation of weighting functions for main parameters

*Extinction coefficient.* For a variation of the extinction coefficient, we have: $p(\tau) = \sigma_e(\tau)$. Functions $Y_e$ and $G_e$ are given by Eqs (6.59) and (6.60). In the considered case they can be written as follows:

$$Y_e(\tau,\Omega) = \sigma_e(\tau)\left[\frac{\partial S_e(\tau,\Omega)}{\partial\sigma_e(\tau)} - \frac{\partial L_e}{\partial\sigma_e(\tau)}I\right] = -\sigma_e(\tau)\frac{\partial L_e}{\partial\sigma_e(\tau)}I\,, \quad (6.234)$$

$$G_e(\Omega) = \sigma_e(\tau)\left[\frac{\partial S_b(\Omega)}{\partial\sigma_e(\tau)} - \frac{\partial L_b}{\partial\sigma_e(\tau)}I\right] = 0\,, \quad (6.235)$$

where we have taken into account that functions $S_e$ and $S_b$ as well as the operator $L_b$ as given by Eqs (6.15), (6.23) and (6.17), respectively, are independent of the extinction coefficient. We note that the extinction coefficient and the single scattering albedo are considered to be independent parameters. Substituting Eqs (6.234) and (6.235) into Eq. (6.61), we obtain

$$\Psi_e(\tau,\Omega) = -\sigma_e(\tau)\frac{\partial L_e}{\partial\sigma_e(\tau)}I\,. \quad (6.236)$$

The partial derivative of the operator $L_e$ as given by Eq. (6.13) with respect to the extinction coefficient $\sigma_e(\tau)$ can be obtained as

$$\frac{\partial L_e}{\partial\sigma_e(\tau)} = \frac{\partial}{\partial\sigma_e(\tau)}\left[-\frac{\mu}{\sigma_e(z)}\frac{\mathrm{d}}{\mathrm{d}z}\right] = \frac{\mu}{\sigma_e^2(\tau)}\frac{\mathrm{d}}{\mathrm{d}z} = -\frac{\mu}{\sigma_e(\tau)}\frac{\mathrm{d}}{\mathrm{d}\tau}\,, \quad (6.237)$$

where we have used that $\mathrm{d}\tau = -\sigma_e(\tau)\,\mathrm{d}z$. Substituting Eq. (6.237) into Eq. (6.236), we obtain

$$\Psi_e(\tau,\Omega) = \mu\frac{\mathrm{d}I}{\mathrm{d}\tau}\,. \quad (6.238)$$

Substituting into this equation $\mu\,\mathrm{d}I/\mathrm{d}\tau$ as given by Eq. (6.9), the final expression for the auxiliary function $\Psi_e(\tau,\Omega)$ is written as follows:

$$\Psi_e(\tau,\Omega) = -I(\tau,\Omega) + J(\tau,\Omega) + S_e(\tau,\Omega)\,, \quad (6.239)$$

where $J(\tau,\Omega)$ and $S_e(\tau,\Omega)$ are given by Eqs (6.12) and (6.15), respectively. The final expression for the extinction coefficient WF follows after substitution the function $\Psi_e(\tau,\Omega)$ into Eq. (6.71):

$$\mathcal{W}_e(\tau;\tau_v,\Omega_v) = \int_{4\pi} I^*(\tau,\Omega;\Omega_v)\Big[J(\tau,\Omega) + S_e(\tau,\Omega) - I(\tau,\Omega)\Big]\,\mathrm{d}\Omega\,. \quad (6.240)$$

*Single scattering albedo.* Employing the same approach for a variation of the single scattering albedo, we obtain

$$\Psi_\omega(\tau,\Omega) = \omega(\tau)\left[\frac{\partial S_e(\tau,\Omega)}{\partial\omega(\tau)} - \frac{\partial L_e}{\partial\omega(\tau)}I\right], \tag{6.241}$$

where we have taken into account that the function $S_b$ and the operator $L_b$ as given by Eqs (6.23) and (6.17), respectively, are independent of the single scattering albedo and, therefore, $G_\omega(\Omega) = 0$. Taking into account Eqs (6.13) and (6.15), partial derivatives of the function $S_e$ and of the operator $L_e$ with respect to the single scattering albedo can be written as follows:

$$\frac{\partial S_e}{\partial\omega(\tau)} = -B(\tau)\,, \qquad \frac{\partial L_e}{\partial\omega(\tau)} = -\frac{1}{4\pi}\int_{4\pi} \mathrm{d}\Omega' p(\tau,\Omega,\Omega')\otimes\,. \tag{6.242}$$

Substituting these expressions into Eg. (6.241), we obtain the final expression for the auxiliary function $\Psi_\omega(\tau,\Omega)$:

$$\Psi_\omega(\tau,\Omega) = J(\tau,\Omega) - B(\tau)\omega(\tau)\,. \tag{6.243}$$

The final expression for the extinction coefficient WF follows after substitution the function $\Psi_\omega(\tau,\Omega)$ into Eq. (6.71):

$$\mathcal{W}_\omega(\tau;\tau_v,\Omega_v) = \int_{4\pi} I^*(\tau,\Omega;\Omega_v)\Big[J(\tau,\Omega) - B(\tau)\omega(\tau)\Big]\,\mathrm{d}\Omega\,, \tag{6.244}$$

*Surface albedo.* A variation of the surface albedo, A, causes only a variation of the lower boundary condition operator, $L_b$, as given by Eq. (6.17). Therefore, using Eqs (6.59)–(6.61), the expression for the function $\Psi_A(\tau,\Omega)$ can be written as follows:

$$\Psi_A(\tau,\Omega) = -A\psi_b(\tau,-\mu)\frac{\partial L_b}{\partial A}I\,. \tag{6.245}$$

Considering the operator $L_b$ given by Eq. (6.17) as a function of the surface albedo, the partial derivative of the $L_b$ with respect to $A$ can be found in the following form:

$$\frac{\partial L_b}{\partial A} = -\frac{1}{\pi}\int_0^{\tau_0}\mathrm{d}\tau\delta(\tau-\tau_0)\int\limits_{4\pi}\mathrm{d}\Omega'\lambda(\mu')\rho(\Omega,\Omega')\otimes\,. \tag{6.246}$$

Applying this operator to the intensity $I(\tau,\Omega)$ and substituting the result into Eq. (6.245), we obtain

$$\Psi_A(\tau,\Omega) = \psi_b(\tau,-\mu)\frac{A}{\pi}\int\limits_{\Omega_+}\rho(\Omega,\Omega')I(\tau_0,\Omega')\mu'\,\mathrm{d}\Omega'\,. \tag{6.247}$$

Having defined the auxiliary function for the surface albedo, the corresponding WF can be obtained substituting it into Eq. (6.70):

$$\mathcal{W}_A(\tau_v,\Omega_v)=\frac{A}{\pi}\int_0^{\tau_0}\int_{4\pi} I^*(\tau,\Omega;\Omega_v)\psi_b(\tau,-\mu)F(\Omega,\tau_0)\,\mathrm{d}\Omega\,\mathrm{d}\tau\,, \tag{6.248}$$

where the function

$$F(\Omega,\tau_0)=\int\limits_{\Omega_+}\rho(\Omega,\Omega')I(\tau_0,\Omega')\mu'\mathrm{d}\Omega' \tag{6.249}$$

is introduced for the simplification reason. The final expression for the surface albedo WF follows after substitution the function $\psi_b(\tau,-\mu)$ given by Eq. (6.28) into Eq. (6.248). Taking into account properties of the Dirac $\delta$-function and the Heaviside step-function, we obtain

$$\mathcal{W}_A(\tau_v,\Omega_v)=-\frac{A}{\pi}\int_{\Omega_-} I^*(\tau_0,\Omega;\Omega_v)F(\Omega,\tau_0)\mu\,\mathrm{d}\Omega\,, \tag{6.250}$$

*Surface emissivity.* A variation of the surface emissivity, $\epsilon$, causes only a variation of the function $S_b(\Omega)$ given by Eq. (6.23). Therefore, taking into account Eqs (6.59)–(6.61), the expression for the function $\Psi_\epsilon(\tau,\Omega)$ can be written as follows:

$$\Psi_\epsilon(\tau,\Omega)=\epsilon\psi_b(\tau,-\mu)\frac{\partial S_b(\Omega)}{\partial\epsilon}\,. \tag{6.251}$$

The derivative of the function $S_b$ with respect to the surface emissivity is

$$\frac{\partial S_b(\Omega)}{\partial\epsilon}=B(T_s)\,. \tag{6.252}$$

Substituting Eq. (6.252) into Eq. (6.251), we obtain

$$\Psi_\epsilon(\tau,\Omega)=\psi_b(\tau,-\mu)\epsilon B(T_s)\,. \tag{6.253}$$

Having defined the auxiliary function for the surface emissivity, the corresponding WF can be obtained substituting it into general expression for the scalar-parameter WF given by Eq. (6.70):

$$\mathcal{W}_\epsilon(\tau_v,\Omega_v)=\epsilon B(T_s)\int_0^{\tau_0}\int_{4\pi} I^*(\tau,\Omega;\Omega_v)\psi_b(\tau,-\mu)\,\mathrm{d}\Omega\,\mathrm{d}\tau\,. \tag{6.254}$$

The final expression for the surface emissivity WF can be obtain in a way analogous to the surface albedo. Substituting the function $\psi_b(\tau,-\mu)$ given by Eq. (6.28) into Eq. (6.254), we have

$$\mathcal{W}_\epsilon(\tau_v,\Omega_v)=-\epsilon B(T_s)\int_{\Omega_-} I^*(\tau_0,\Omega;\Omega_v)\mu\,\mathrm{d}\Omega\,. \tag{6.255}$$

## References


1. G. I. Bell, S. Glasstone: *Nuclear reactor theory* (Van Nostrand, New York, Reinholt 1970)
2. M. A. Box, S. A. W. Gerstl, C. Simmer: Application of the adjoint formulation to the calculation of atmospheric radiative effects, *Beitr. Phys. Atmosph.*, **61**, 303–311 (1988)
3. M. Buchwitz, R. de Beek, J. P. Burrows, H. Bowensmann, T. Warneke, J. Notholt, J. F. Meirink, A. P. H. Goede, P. Bergamaschi, S. Koerner, M. Heimann, and A. Schulz: Atmospheric methane and carbon dioxide from SCIAMACHY satellite data: initial comparison with chemistry and transport models, *Atmos. Chem. Phys.*, **5**, 941–962 (2005)
4. S. Chandrasekhar: *Radiative transfer* (Oxford University Press, London, 1950)
5. Fu-Lung Chang and Zhanqing Li: Retrieving vertical profiles of water-cloud droplet effective radius: Algorithm modification and preliminary application, *J. Geophys. Res.*, **108**, D24, 4763, doi:10.1029/2003JD003906, (2003)
6. J. W. Hovenier, C. Van der Mee, H. Domke: *Transfer of polarized light in planetary atmospheres. Basic concepts and practical methods* (Kluwer Academic Publishers, Dordrecht Boston London, 2004)
7. Alexander A. Kokhanovsky: *Cloud optics* (Springer, Netherlands, 2003)
8. A. A. Kokhanovsky: Local optical characteristics of mixed clouds: simple parameterizations, *Atmos. Res.*, in press (2007)
9. G. A. Korn, T. M. Korn: *Mathematical handbook for scientists and engineers* (McGraw-Hill Book Company, New York San Francisco Toronto London Sydney, 1968)
10. V. P. Kozlov: Design of experiments related to the inverse problem of mathematical physics. In: *Mathematical theory of experiment design*, ed. by C. M. Ermakov, pp. 216–246, Nauka, Moscow, 1983, in Russian.
11. K. N. Liou: *An introduction to atmospheric radiation* (Academic Press, New York, 1980)
12. J. Lenoble: *Atmospheric radiative transfer* (A. Deepak Publishing, Hampton, Virginia USA, 1993)
13. Y. L. Luke: *Mathematical functions and their approximations* (Academic Press Inc., New York San Francisco London, 1975)
14. G. I. Marchuk: Equation for the value of information from weather satellites and formulation of inverse problems, *Cosmic Res.*, (2), 394–409 (1964)
15. Q. Min, L. E. Harrison: An adjoint formulation of the radiative transfer method, *J. Geophys. Res.*, **101**, D1, 1635–1640 (1996)
16. I. N. Polonsky, M. A. Box: General perturbation technique for the calculation of radiative effects in scattering and absorbing media, *J. Opt. Soc. of America*, **19**, Issue 11, page 2281 (2002)
17. I. N. Polonsky, M. A. Box, A. B. Davis: Radiative transfer through inhomogeneous turbid media: implementation of the adjoint perturbation approach at the first order, *J. Quant. Spectrosc. Radiat. Transfer*, **78**, 85–98 (2003)
18. G. C. Pomraning: *Linear kinetic theory and particle transport in stochastic mixtures* (World Scientific Publishing, Singapore New Jersey London Hong Kong, 1991)
19. C. D. Rodgers: *Inverse methods for atmospheric sounding: theory and practice* (World Scientific, Singapore, 2000)
20. V. V. Rozanov, Yu. M. Timofeev: Retrieval of vertical profiles of atmospheric aerosol optical characteristics using satellite measurements of outgoing radiation

in 0.76 micrometers absorption band, *Izvestiya AN. Physics of Atmosphere and Ocean*, **30**, N6, 779–785 (1994)
21. V. V. Rozanov and A. A. Kokhanovsky: Semi-analytical cloud retrieval algorithm as applied to the cloud top altitude and the cloud geometrical thickness determination from top of atmosphere reflectance measurement in the oxygen absorption bands, *J. Geophys. Res.*, **109**, D05202, 10.1029/2003JD004104, (2004)
22. A. Rozanov, V. Rozanov, M. Buchwitz, A. Kokhanovsky, J. P. Burrows: SCIATRAN 2.0 – A new radiative transfer model for geophysical applications in the 175–2400 nm spectral region, *Adv. Space Res.*, doi:10.1016/j.asr.2005.03.012
23. A. Rozanov, H. Bovensmann, A. Bracher, S. Hrechanyy, V. Rozanov, M. Sinnhuber, F. Stroh, J. P. Burrows: $NO_2$ and BrO vertical profile retrieval from SCIAMACHY limb measurements: Sensitivity studies, *Adv. Space Res.*, **36**, 846–854 (2005)
24. V. V. Rozanov: Adjoint radiative transfer equation and inverse problems. In: *Light Scattering Reviews*, ed. by A. A. Kokhanovsky. (Springer, Berlin, 2006)
25. V. V. Rozanov, A. V. Rozanov: Generalized form of the direct and adjoint radiative transfer equations. *J. Quant. Spectrosc. Radiat. Transfer*, **104**, 155–170 (2007).
26. A. Rozanov, V. Rozanov, J. P. Burrows: Software package SCIATRAN 2.1 – New developments in the radiative transfer modeling and the retrieval technique. Paper presented at COSPAR 2006, to appear in *Adv. Space Res.*, 2007
27. G. E. Thomas, K. Stamnes: *Radiative transfer in the atmosphere and ocean* (Cambridge University Press, Cambridge, 1999)
28. A. N. Tikhonov, V. Ya. Arsenin: *Solutions of ill-posed problems* (Halsted Press, Washington, DC 1977 )
29. Yu. M. Timofeyev, A. V. Vasilyev, V. V. Rozanov: Information content of the spectral measurements of the 0.76 µm $O_2$ outgoing radiation with respect to the vertical aerosol optical properties, *Adv. Space Res.*, **16**, No. 10, 1091–1094 (1995)
30. E. A. Ustinov: Adjoint sensitivity analysis of radiative transfer equation: temperature and gas mixing ratio weighting functions for remote sensing of scattering atmospheres in thermal IR, *J. Quant. Spectrosc. Radiat. Transfer*, **68**, 195–211 (2001)
31. E. A. Ustinov: Adjoint sensitivity analysis of radiative transfer equation: 2. Applications to retrievals of temperature in scattering atmospheres in thermal IR. *J. Quant. Spectrosc. Radiat. Transfer*, **73**, 29–40 (2002)
32. E.A. Ustinov: Atmospheric weighting functions and surface partial derivatives for remote sensing of scattering planetary atmospheres in thermal spectral region: general adjoint approach, *J. Quant. Spectrosc. Radiat. Transfer*, **92**, 351–371 (2005)
33. V. Volterra: *Theory of functionals and of integral and integro-differential equations* (Dover, New York, 1959)

Part III

# Numerical Techniques

# 7 Studies of light scattering by complex particles using the null-field method with discrete sources

Thomas Wriedt

## 7.1 Introduction

Light scattering has an increasing importance in modern technologies. Examples are characterization of particles in natural or technical environments, surface characterization, biomedical sensing and nanotechnology. As a consequence, the development of accurate and fast methods devoted to the numerical simulation of electromagnetic and light scattering has become of fundamental importance.

There is a long interest in light scattering computations for nonspherical natural or artificial particles. To perform scattering computations the T-matrix method is considered of advantage because in the T-matrix all information on the polarization scattering effects is included. Thus from a precomputed T-matrix a scattering problem under slightly different conditions of incident wave orientation or scattering angles can quite easily be computed. There are also efficient ways to compute orientation-averaged scattering quantities from a precomputed T-matrix.

Although there has been much development in the T-matrix method over the last two decades there still have been problems with scattering computations for some types of particles. These include arbitrarily shaped nonaxisymmetric particles, particles having a large aspect ratio such as finite fibres or flat discs, and compound particles consisting of regions of different refractive indices. Other problems have been with chiral or optically anisotropic particles.

In the recent development of the Null-Field Method with Discrete Sources (NFM-DS) all of the method problems mentioned have been solved. In the standard T-matrix method a single system of spherical vector wave functions is used for internal field expansions. In the NFM-DS different kinds of discrete sources having different positions can be used for field expansion and this helps to overcome the stability problems with the standard method.

In this chapter the development of the Null-Field Method with Discrete Sources will be reviewed and some exemplary scattering results will be presented to demonstrate the capabilities of the concept. Recent developments will be mentioned.

First we give a short review of the state of the art of the Discrete Sources Method as well as the T-Matrix Method. Next we will introduce the Null-Field method with Discrete Sources as an extension of these two methods.

## 7.2 Discrete Sources Method

The Discrete Sources Method (DSM) [1] and related methods are widely used techniques for the numerical solution of elliptic boundary value problems including electromagnetic scattering.

The main idea of the DSM consists of approximating the solution of the problem by a linear combination of discrete sources. This discrete sources are the fundamental solution of the differential equation of the problem. The introduction of the Discrete Sources Method is generally attributed to Kupradze and Aleksidze [2].

Since that time the method has been applied in various fields such as acoustics, elasticity theory, electromagnetism, fluid dynamics, geophysics and solid mechanics. For a full theoretical outline of the method we refer the interested reader to the book by Doicu, Eremin and Wriedt [3]. An excellent review of the DSM and related methods for elliptic boundary value problems over recent decades has been given by Fairweather and Karageorghis [4]. Fairweather, Karageorghis and Martin [5] also surveyed the DSM applications in scattering and radiation problems. An edited volume covering different variants of this method has been published by Wriedt [6].

The DSM method is able to solve the problem of scattering from arbitrary shaped scatterers. Using point matching or point collocation of the boundary condition on the surface of the scatterer the original problem is reduced to determining the unknown coefficients of the discrete sources by solving a linear system of equations. The coefficients can also be obtained by matching the fields at the boundaries of the regions using a least squares fit of the boundary data. In this way the scattered field can be expressed in terms of a complete set of discrete sources.

There are various other names used for similar kind of concepts such as Charge Simulation Method [7], Yasuura Method [8], Multiple Multipole Program [9], Method of Auxiliary Sources [10], Discrete Singularity Method [11], Fictitious Sources Method [12], Method of Fictitious Sources [13], Method of Fundamental Solutions [14], and Generalized Multipole Technique [15].

The advantage of the Discrete Sources Method is that it provides a reduction in the size of the linear system that has to be solved and thus leads to a reduction in the computation time and memory storage.

The representation of electromagnetic fields by the use of discrete sources placed apart from the surface of the scatterer helps to simulate scattering by complex particles which cannot be solved using the standard T-matrix method. These particles include elongated scatterers, flat scatterers and concave scatterers. Compared to the surface integral method it does not have the problems encountered with singularities of the kernels. The method includes the possibility

of checking the accuracy of the computational results by means of a boundary matching error. The method also allows free choice of the kind and the location of the discrete sources, but if optimization of the coordinates of the discrete sources is included in the computational algorithms this would lead to a time consuming nonlinear least-squares minimization procedure. Another advantage of the DSM over other methods is that it does not require an elaborate discretization of the surface of the scatterer as in the Boundary Element Method (BEM); also integrations over the particle boundary surface, as needed in the T-matix method, are avoided.

## 7.3 T-matrix method

The T-matrix method is a widely used method for obtaining numerical solutions to electromagnetic scattering problems. The T-matrix method found a wide range of applications because a corresponding FORTRAN program for a conducting scatterer had already been published in the early 1970s [16] which was later extended by others to the dielectric case. The T-matrix method is also called the Null-Field Method (NFM) or Extended Boundary Condition Method (EBCM). It is based on a series of papers by Waterman [17]. An early collection of conference papers on this method was edited by Varadan and Varadan [18]. With this method the incident, transmitted and scattered field is expanded into a series of spherical vector wavefunctions as shown for the scattered field:

$$\mathbf{E}_s(\mathbf{x}) = \sum_{\nu=1}^{\infty} f_\nu \mathbf{M}_\nu^3(k_s\mathbf{x}) + g_\nu \mathbf{N}_\nu^3(k_s\mathbf{x}) \,, \tag{7.1}$$

$$\begin{bmatrix} f_\nu \\ g_\nu \end{bmatrix} = \mathbf{T} \begin{bmatrix} a_\nu^0 \\ b_\nu^0 \end{bmatrix} . \tag{7.2}$$

In this equation $\mathbf{N}_\nu^3$ and $\mathbf{N}_\nu^3$ represent the rediating spherical vector wavefunctions. The expansion coefficients of the scattered field $f_\nu$, $g_\nu$ are related to the coefficients of the incident field $a_\nu^0, b_\nu^0$ by the T-matrix (transition matrix).

The elements of the T-matrix are obtained by numerical integration. For an arbitrarily shaped particle lacking rotational symmetry a surface integral has to be computed. As this is computationally expensive, most implementations of the method are restricted to axisymmetric scatterers. In this case line integrals have to be computed. Nevertheless, there are some papers in which the T-matrix method has been applied to arbitrarily shaped scatterers. Early scattering computation for nonaxisymmetric scatterers using the T-matrix method have been done by P. W. Barber in his Ph.D. thesis [19] and by Schneider and Peden [20], both presenting results for ellipsoids. Wriedt and Doicu [21] presented computational examples with results of scattering by a dielectric cube of size parameter 2. A 3D variant of the T-matrix method has also been developed by Laitinen and Lumme [22] and by Kahnert et al. [23] both presenting scattering results for rounded cubes and another implementation was published by Havemann and Baran [24] giving results for hexagonal ice crystals.

A review of the status of the T-matrix approach up to 1996 has been published by Mishchenko, Travis and Mackowski [25]. A more recent review can be found in the book by Mishchenko, Hovenier and Travis [26]. The T-matrix method is also subject of books by Mishchenko et al. [27] and Borghese et al. [28]. A database of the literature on the T-matrix method has recently been compiled by Mishchenko et al. [29].

The notion of a T-matrix of a single scatterer makes it possible to solve problems of scattering from an arbitrary number of homogeneous objects also in the vicinity of a plane surface through the use of the T-matrix formalism. This method fully takes into account the interaction between the objects from multiple scattering and can deal with a large number of scattering particles. This feature, computing the T-matrix for a group of scatterers from the T-matrix of each constituent, makes the method very powerful and is considered its main advantages over other methods.

It has been found that the numerical performance of the T-matrix method is strongly dependent on the shape of the scatterer. It tends to degrade as the shape deviates from a sphere. An efficient approach for overcoming the numerical-instability problem in computing the T-matrix for highly nonspherical particles is the Null-Field Method with Discrete Sources, which is introduced in the next section.

## 7.4 Null-Field method with Discrete Sources

The Null-Field Method with Discrete Sources (NFM-DS) was originally developed to solve the stability problems in the standard T-matrix method with elongated and flat particles.

In this section we would like to outline the basics of the Null-field Method with Discrete Sources [3]. Let us consider a three-dimensional space $D$ consisting of the union of a closed surface $S$, its interior $D_i$ and its exterior $D_s$. We denote by $k_t$ the wave number in the domain $D_t$, where $k_t = k\sqrt{\varepsilon_t \mu_t}$, $k = \omega/c$, $t = s, i$ and $\varepsilon_t$ s the permeability, $\mu_t$ is the permitivity.

The transmission boundary-value problem can be formulated as follows**.** Let $\mathbf{E}_0, \mathbf{H}_0$ be an entire solution to the Maxwell equations representing an incident electromagnetic field. Find the vector fields, $\mathbf{E}_s, \mathbf{H}_s \in C^1(D_s) \cap C(\overline{D}_s)$ and $\mathbf{E}_i, \mathbf{H}_i \in C^1(D_i) \cap C(\overline{D}_i)$ satisfying the Maxwell's equations

$$\begin{aligned} \nabla \times \mathbf{E}_t &= jk\mu_t \mathbf{H}_t \,, \\ \nabla \times \mathbf{H}_t &= -jk\varepsilon_t \mathbf{E}_t \,, \end{aligned} \tag{7.3}$$

in $D_t$, where $t = s, i$; $j = \sqrt{(-1)}$ and two boundary conditions:

$$\begin{aligned} \mathbf{n} \times \mathbf{E}_i - \mathbf{n} \times \mathbf{E}_s &= \mathbf{n} \times \mathbf{E}_0 \,, \\ \mathbf{n} \times \mathbf{H}_i - \mathbf{n} \times \mathbf{H}_s &= \mathbf{n} \times \mathbf{H}_0 \,, \end{aligned} \tag{7.4}$$

on $S$, where $\mathbf{n}$ is the outward directing normal to the boundary. In addition, the scattered fields $\mathbf{E}_s, \mathbf{H}_s$ must satisfy the Silver–Müller radiation condition uniformly for all directions $\mathbf{x}/x$.

For solving the transmission boundary-value problem in the framework of the NFM-DS the scattering object is replaced by a set of surface current densities $\mathbf{e}$ and $\mathbf{h}$, so that in the exterior domain the sources and fields are exactly the same as those existing in the original scattering problem. The entire analysis can conveniently be broken down into the following three steps:

(I) A set of integral equations for the surface current densities $\mathbf{e}$ and $\mathbf{h}$ is derived for a variety of discrete sources. Physically, the set of integral equations in question guarantees the null-field condition within $D_i$. It is noted that localized and distributed vector spherical functions, magnetic and electric dipoles or vector Mie-potentials can be used as discrete sources. Essentially, the NFM-DS consists in the projection relations:

$$\begin{aligned}
\int_S \left[ (\mathbf{e}-\mathbf{e}_0)\cdot\boldsymbol{\Psi}_\nu^3 + j\sqrt{\frac{\mu_s}{\varepsilon_s}}(\mathbf{h}-\mathbf{h}_0)\cdot\boldsymbol{\Phi}_\nu^3 \right] \mathrm{d}S &= 0 \\
\int_S \left[ (\mathbf{e}-\mathbf{e}_0)\cdot\boldsymbol{\Phi}_\nu^3 + j\sqrt{\frac{\mu_s}{\varepsilon_s}}(\mathbf{h}-\mathbf{h}_0)\cdot\boldsymbol{\Psi}_\nu^3 \right] \mathrm{d}S &= 0, \ \ \nu = 1,2,\ldots
\end{aligned} \tag{7.5}$$

where $\mathbf{e}_0 = \mathbf{n}\times\mathbf{E}_0$ and $\mathbf{h}_0 = \mathbf{n}\times\mathbf{H}_0$ are the tangential components of the incident electric and magnetic fields. The set $\left\{\boldsymbol{\Psi}_\nu^3, \boldsymbol{\Phi}_\nu^3\right\}_{\nu=1,2,\ldots}$ consists of radiating solutions to Maxwell equations and depends on the system of discrete sources which is used for imposing the null-field condition. Actually, this set together with the set of regular solutions to Maxwell equations $\left\{\boldsymbol{\Psi}_\nu^1, \boldsymbol{\Phi}_\nu^1\right\}_{\nu=1,2,\ldots}$ stands for

– localized vector spherical functions $\left\{\mathbf{M}_{mn}^{1,3}, \mathbf{N}_{mn}^{1,3}\right\}_{m\in\mathbb{Z},n\geq\max(1,|m|)}$,

$$\begin{aligned}
\mathbf{M}_{mn}^{1,3}(k\mathbf{x}) &= \sqrt{D_{mn}} z_n(kr)\left[ jm\frac{P_n^{|m|}(\cos\theta)}{\sin\theta}\mathbf{e}_\theta - \frac{\mathrm{d}P_n^{|m|}(\cos\theta)}{\mathrm{d}\theta}\mathbf{e}_\varphi \right] e^{jm\varphi}, \\
\mathbf{N}_{mn}^{1,3}(k\mathbf{x}) &= \sqrt{D_{mn}}\left\{ n(n+1)\frac{z_n(kr)}{kr}P_n^{|m|}(\cos\theta)\,\mathrm{e}^{jm\varphi}\mathbf{e}_r \right. \\
&\quad \left. + \frac{[krz_n(kr)]'}{kr}\left[ \frac{\mathrm{d}P_n^{|m|}(\cos\theta)}{\mathrm{d}\theta}\mathbf{e}_\theta + jm\frac{P_n^{|m|}(\cos\theta)}{\sin\theta}\mathbf{e}_\varphi \right] \right\} \mathrm{e}^{jm\varphi},
\end{aligned} \tag{7.6}$$

where $\left(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\varphi\right)$ are the unit vectors in spherical coordinates $(r,\theta,\varphi)$ , $z_n$ designates the spherical Bessel functions $j_n$ or the spherical Hankel functions of the first kind $h_n^1$, $P_n^{|m|}$ denotes the associated Legendre polynomial of order $n$ and $m$, and $D_{mn}$ is a normalization constant given by

$$D_{mn} = \frac{2n+1}{4n(n+1)}\cdot\frac{(n-|m|)!}{(n+|m|)!}, \tag{7.7}$$

– distributed vector spherical functions $\left\{\mathcal{M}_{mn}^{1,3}, \mathcal{N}_{mn}^{1,3}\right\}_{m\in\mathbb{Z},n=1,2,\ldots}$:

$$\begin{aligned}
\mathcal{M}_{mn}^{1,3}(k\mathbf{x}) &= \mathbf{M}_{m,|m|+l}^{1,3}\left(k(\mathbf{x}-z_n\mathbf{e}_3)\right), \ \mathbf{x}\in\mathbb{R}^3 - \{z_n\mathbf{e}_3\}_{n=1}^\infty, \\
\mathcal{N}_{mn}^{1,3}(k\mathbf{x}) &= \mathbf{N}_{m,|m|+l}^{1,3}\left(k(\mathbf{x}-z_n\mathbf{e}_3)\right), \ \mathbf{x}\in\mathbb{R}^3 - \{z_n\mathbf{e}_3\}_{n=1}^\infty,
\end{aligned} \tag{7.8}$$

where $m \in \mathbb{Z}^0$, $n = 1, 2, \ldots$; $l = 1$ if $m = 0$ and $l = 0$ if $m \neq 0$, and $\{z_n\}_{n=1}^{\infty}$ is a set of points located on a segment $\Gamma_z$ of the $z$-axis,

– magnetic and electric dipoles $\left\{\mathcal{M}_{ni}^{1,3}, \mathcal{N}_{ni}^{1,3}\right\}_{n=1,2,\ldots,i=1,2}$ :

$$\begin{aligned} \mathcal{M}_{ni}^{1,3}(k\mathbf{x}) &= \mathbf{m}(\mathbf{x}_n^{\pm}, \mathbf{x}, \tau_{ni}^{\pm}),\ \mathbf{x} \in \mathbb{R}^3 - \{\mathbf{x}_n^{\pm}\}_{n=1}^{\infty}\ , \\ \mathcal{N}_{ni}^{1,3}(k\mathbf{x}) &= \mathbf{n}(\mathbf{x}_n^{\pm}, \mathbf{x}, \tau_{ni}^{\pm}),\ \mathbf{x} \in \mathbb{R}^3 - \{\mathbf{x}_n^{\pm}\}_{n=1}^{\infty}\ , \end{aligned} \tag{7.9}$$

where $n = 1, 2, \ldots$; $i = 1, 2, \ldots$; $\tau_{n1}$ and $\tau_{n2}$ are two tangential linear independent unit vectors at the point $\mathbf{x}_n$,

$$\mathbf{m}(\mathbf{x}, \mathbf{y}, \mathbf{a}) = \frac{1}{k^2}\mathbf{a}(\mathbf{x}) \times \nabla_{\mathbf{y}} g(\mathbf{x}, \mathbf{y}, k),\ \mathbf{n}(\mathbf{x}, \mathbf{y}, \mathbf{a}) = \frac{1}{k}\nabla_{\mathbf{y}} \times \mathbf{m}(\mathbf{x}, \mathbf{y}, \mathbf{a}),\ \mathbf{x} \neq \mathbf{y}\,, \tag{7.10}$$

and the sequence $\{\mathbf{x}_n^-\}_{n=1}^{\infty}$ is dense on a smooth surface $S^-$ enclosed in $D_i$, while the sequence $\{\mathbf{x}_n^+\}_{n=1}^{\infty}$ is dense on a smooth surface $S^+$ enclosing $D_i$, or finally for the set of

– vector Mie-potentials $\left\{\mathcal{M}_n^{1,3}, \mathcal{N}_n^{1,3}\right\}_{n=1,2,\ldots}$ :

$$\begin{aligned} \mathcal{M}_n^{1,3}(k\mathbf{x}) &= \tfrac{1}{k}\nabla\varphi_n^{\pm}(\mathbf{x}) \times \mathbf{x},\ \mathbf{x} \in \mathbb{R}^3 - \{\mathbf{x}_n^{\pm}\}_{n=1}^{\infty}\ , \\ \mathcal{N}_n^{1,3}(k\mathbf{x}) &= \tfrac{1}{k}\nabla \times \mathcal{M}_n^{1,3}(k\mathbf{x}),\ \mathbf{x} \in \mathbb{R}^3 - \{\mathbf{x}_n^{\pm}\}_{n=1}^{\infty}\ , \end{aligned} \tag{7.11}$$

where the Green functions

$$\varphi_n^{\pm}(\mathbf{x}) = g(\mathbf{x}_n^{\pm}, \mathbf{x}, k),\ n = 1, 2, \ldots$$

have singularities $\{\mathbf{x}_n^-\}_{n=1}^{\infty}$and $\{\mathbf{x}_n^+\}_{n=1}^{\infty}$ distributed on the auxiliary surfaces $S^-$ and $S^+$, respectively. By convention, when we refer to the null-field equations (7.5) we implicitly refer to all equivalent forms of these equations.

(II) The surface current densities are approximated by fields of discrete sources. In this context let $\mathbf{e}$ and $\mathbf{h}$ solve the null-field equations (7.5) and assume that the system $\left\{\mathbf{n} \times \mathbf{\Psi}_\mu^1, \mathbf{n} \times \mathbf{\Phi}_\mu^1\right\}_{\mu=1}^{\infty}$ forms a Schauder basis in $\mathcal{L}_{\tan}^2(S)$. Then there exists a sequence $\{a_\mu, b_\mu\}_{\mu=1}^{\infty}$ such that

$$\begin{aligned} \mathbf{e}(\mathbf{y}) &= \sum_{\mu=1}^{\infty} a_\mu \mathbf{n} \times \mathbf{\Psi}_\mu^1(k_i\mathbf{y}) + b_\mu \mathbf{n} \times \mathbf{\Phi}_\mu^1(k_i\mathbf{y})\,,\ \mathbf{y} \in S\,, \\ \mathbf{h}(\mathbf{y}) &= -j\sqrt{\tfrac{\varepsilon_i}{\mu_i}} \sum_{\mu=1}^{\infty} a_\mu \mathbf{n} \times \mathbf{\Phi}_\mu^1(k_i\mathbf{y}) + b_\mu \mathbf{n} \times \mathbf{\Psi}_\mu^1(k_i\mathbf{y})\,,\ \mathbf{y} \in S\,. \end{aligned} \tag{7.12}$$

We recall that a system $\{\psi_i\}_{i=1}^{\infty}$ is called a Schauder basis of a Banach space $X$ if any element $u \in X$ can be uniquely represented as $u = \sum_{i=1}^{\infty} \alpha_i \psi_i$, where the convergence of the series is in the norm of $X$. It is noted that in the case of localized vector spherical functions the notion of Schauder basis is closely connected with the Rayleigh hypothesis. This hypothesis says that the series representation of the scattered field in terms of radiating localized vector spherical functions, which uniformly converges outside the circumscribing sphere, also converges on $S$.

(III) Once the surface current densities are determined the scattered field outside the circumscribing sphere is obtained by using the Stratton–Chu representation theorem. We get the series representation

$$\mathbf{E}_s(\mathbf{x}) = \sum_{\nu=1}^{\infty} f_\nu \mathbf{M}_\nu^3(k_s\mathbf{x}) + g_\nu \mathbf{N}_\nu^3(k_s\mathbf{x}) \,, \tag{7.13}$$

where

$$\begin{aligned} f_\nu &= \tfrac{jk_s^2}{\pi} \int_S \left[ \mathbf{e}(\mathbf{y}) \cdot \mathbf{N}_{\overline{\nu}}^1(k_s\mathbf{y}) + j\sqrt{\frac{\mu_s}{\varepsilon_s}} \mathbf{h}(\mathbf{y}) \cdot \mathbf{M}_{\overline{\nu}}^1(k_s\mathbf{y}) \right] \mathrm{d}S(\mathbf{y}) \,, \\ g_\nu &= \tfrac{jk_s^2}{\pi} \int_S \left[ \mathbf{e}(\mathbf{y}) \cdot \mathbf{M}_{\overline{\nu}}^1(k_s\mathbf{y}) + j\sqrt{\frac{\mu_s}{\varepsilon_s}} \mathbf{h}(\mathbf{y}) \cdot \mathbf{N}_{\overline{\nu}}^1(k_s\mathbf{y}) \right] \mathrm{d}S(\mathbf{y}) \,. \end{aligned} \tag{7.14}$$

Here, $\overline{\nu}$ is a complex index incorporating $-m$ and $n$, i.e. $\overline{\nu} = (-m, n)$.

### 7.4.1 T-matrix computation

Now, for deriving the T-matrix, let us assume that the incident field can be expressed inside a finite region containing $S$ as a series of regular vector spherical functions

$$\begin{aligned} \mathbf{E}_0(\mathbf{x}) &= \sum_{\nu=1}^{\infty} a_\nu^0 \mathbf{M}_\nu^1(k_s\mathbf{x}) + b_\nu^0 \mathbf{N}_\nu^1(k_s\mathbf{x}) \,, \\ \mathbf{H}_0(\mathbf{x}) &= -j\sqrt{\frac{\varepsilon_s}{\mu_s}} \sum_{\nu=1}^{\infty} a_\nu^0 \mathbf{N}_\nu^1(k_s\mathbf{x}) + b_\nu^0 \mathbf{M}_\nu^1(k_s\mathbf{x}) \,. \end{aligned} \tag{7.15}$$

Then, using (7.5)–(7.15) we see that the relation between the scattered and the incident field coefficients is linear and is given by a transition matrix $\mathbf{T}$ as follows

$$\begin{bmatrix} f_\nu \\ g_\nu \end{bmatrix} = \mathbf{T} \begin{bmatrix} a_\nu^0 \\ b_\nu^0 \end{bmatrix} . \tag{7.16}$$

Here

$$\mathbf{T} = \mathbf{B}\mathbf{A}^{-1}\mathbf{A}_0 \,, \tag{7.17}$$

where $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{A}_0$ are block matrices written in general as

$$\mathbf{X} = \begin{bmatrix} X_{\nu\mu}^{11} & X_{\nu\mu}^{12} \\ X_{\nu\mu}^{21} & X_{\nu\mu}^{22} \end{bmatrix} , \; \nu, \mu = 1, 2, \ldots , \tag{7.18}$$

with $\mathbf{X}$ standing for $\mathbf{A}, \mathbf{B}$ and $\mathbf{A}_0$. Explicit expressions for the elements of these matrices are given by Doicu, Eremin and Wriedt [3].

It is noted that the exact, infinite T-matrix is independent of the expansion systems used on $S$. However, the approximate truncated matrix, computed according to

$$\mathbf{T}_N = \mathbf{B}_N \mathbf{A}_N^{-1} \mathbf{A}_{0N} \tag{7.19}$$

does contain such a dependence.

Energy characteristics in the far field are computed from the far-field pattern $\mathbf{E}_{s0}^{N}$ for an unit amplitude incident electric field for p- or s-polarization. The angle-dependent intensity function plotted in the simulation section is the normalized differential scattering cross-section (DSCS)

$$\frac{\sigma_d}{\pi a^2} = \frac{\left|k_s \mathbf{E}_{s0}^{N}\right|^2}{\pi \left|k_s a\right|^2}\,, \tag{7.20}$$

where $a$ is a characteristic dimension of the particle.

### 7.4.2 Orientation averaged scattering

To numerically compute orientation averaged scattering three integrals with respect to the three Euler angles $\alpha, \beta, \gamma$ have to be computed. Thus the value of interest $f(\alpha, \beta, \gamma)$ is integrated over all directions and polarization of the incident plane wave. The numerical procedure used to do this is based on a step wise procedure

$$\begin{aligned} &\int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} f(\alpha,\beta,\gamma) \sin\beta \,\mathrm{d}\alpha \,\mathrm{d}\beta \,\mathrm{d}\gamma \\ &\quad\approx\; 4\pi^3 \sum_{n_\alpha=1}^{N_\alpha} \sum_{n_\beta=1}^{N_\beta} \sum_{n\gamma=1}^{N_\gamma} f(\alpha,\beta,\gamma) \sin(n_\beta \pi / N_\beta) \frac{n_\alpha n_\beta n_\gamma}{N_\alpha N_\beta N_\gamma}\,. \end{aligned} \tag{7.21}$$

The triple integral is converted to three summations. Angle $\alpha$ is digitized for $N_\alpha$ steps in the range $(0, 2\pi)$, angle $\beta$ is digitized for $N_\beta$ steps in the range $(0, \pi)$, and angle $\gamma$ is digitized for $N_\gamma$ steps in the range $(0, 2\pi)$.

### 7.4.3 Computation of surface integrals

In the NFM-DS method we use a polyhedral representation of the particle shape model of interest. This means that a particle shape is modeled by a closed surface formed by many planar triangles. Such shapes bounded by polygons are commonly used in computer graphics and there are different methods available to construct a polyhedral representation of implicit or parametric surfaces. In our case the representation by a triangular patch model should allow firstly a correct calculation of surface integrals and secondly a graphical visualization of the scattering particle.

Different methods are available to create a geometric surface mesh of a particle. For example the free *HyperFun* polygonizer software may be used for surface mesh generation [30]. It generates VRML output of a triangular patch model for implicit surfaces.

In the standard method to compute surface integrals a parametric equation is used and thus an equivalent integral in polar coordinates has to be evaluated. Thus partial derivatives of $x, y, z$ with respect to the parametrization (in our case parameters $\vartheta, \varphi$) are needed which may not be available analytically. If the

partial derivatives are not available, a numerical method via finite differences may be used.

We use an alternative approach based on a modified centroid quadrature that does not use the partial derivatives. This modified centroid quadrature has been proposed and investigated by Georg and Tauch [31]. The surface integrals to be computed are approximated by

$$\int_S f \, \mathrm{d}S \approx \sum_i f(v_{i,c}) \; area[v_{i,1}, v_{i,2}, v_{i,3}] \, . \tag{7.22}$$

Here, $v_{i,1}, v_{i,2}, v_{i,3}$ are the vertices spanning a triangle and point $v_{i,c}$ denotes the centre of mass of the triangle $[v_{i,1}, v_{i,2}, v_{i,3}]$ given by

$$v_{i,c} = \frac{1}{3} \sum_{j=1}^{3} v_{i,j} \, . \tag{7.23}$$

Thus the integral over each triangle is approximated by multiplying the value of the integrand at the centroid by the triangle area.

## 7.5 Scattering by complex particles

In this section we would like to present some exemplary scattering results for complex particles using the NFM-DS. We will show computational results and in some cases validation results obtained using other computational programs. More information on the method of validation of the developed programs will be given in the next section on validation.

### 7.5.1 Fibres

Field expansion using discrete sources helps very much when computing scattering by very elongated scatterers such as finite rotational symmetric fibres. With such type of scatterers the discrete sources are positioned on the axis of symmetry of the particle. As an example of an elongated particle we are considering a long circular cylinder which is rounded at the bottom and the top. The shape of this fibre is presented in Fig. 7.1.

In polar coordinates this fibre shape is described by the following equations:

$$\begin{aligned}
r &= a\cos\theta \pm \sqrt{b^2 - a^2\sin^2\theta} \quad \text{for} \quad 0 \le \theta \le \arctan\left(\frac{b}{a}\right) \\
r &= \frac{b}{\sin\theta} \quad \text{for} \quad \arctan\left(\frac{b}{a}\right) < \theta < \arctan\left(\frac{-a}{b}\right) \\
r &= -a\cos\theta \pm \sqrt{b^2 - a^2\sin^2\theta} \quad \text{for} \quad \arctan\left(\frac{-a}{b}\right) \le \theta \le \pi
\end{aligned}$$

So the total length of the fibre is $2(a + b)$ and its diameter is $2b$.

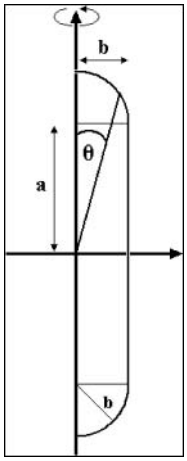



**Fig. 7.1.** Geometry of a fibre.

Light scattering computations of fibre-like particles are of interest for various scientific branches: astrophysics, atmospheric science, optical particle characterization – the latter especially in connection with airborne fibrous particles like mineral, glass or asbestos fibres, which are considered to cause serious health hazards. Here high aspect ratios are of special interest and so it is required that a light scattering simulation algorithm can handle them. Figure 7.2 presents the differential scattering cross-section (DSCS) of such a fibre and demonstrates

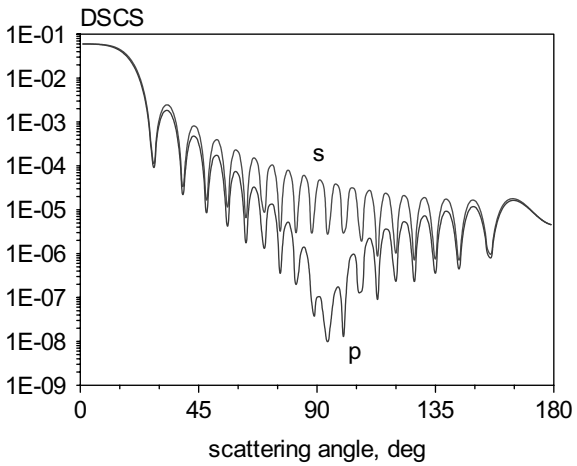



**Fig. 7.2.** Differential scattering cross-section of a fibre particle with length $2(a+b) = 6\,\mu\text{m}$ and diameter $2b = 0.12\,\mu\text{m}$. Incident wavelength is $\lambda = 632.8\,\text{nm}$, refractive index $n = 1.5$.

the capabilities of using discrete sources in the T-matrix method for fibre-like particles.

The total length of the fibre is 6 µm and the diameter is 0.12 µm which leads to an aspect ratio of 50:1, which is an already challenging value. The incident wavelength is 632.8$nm$ and the refractive index is 1.5. The plane wave is incident onto the long side of the fibre. Convergent computational results were achieved using 50 discrete sources and 3,000 surface points to compute the line integrals.

Further computational results have been published by Pulbere and Wriedt [32].

### 7.5.2 Flat plates

Application of discrete sources also helps in computing scattering by oblate particles which are very flat. In this case the discrete sources have to be arranged in the complex plane [33]. As an exemplary particle we use a flat circular disc which is rounded at its edges. The geometry of such disc is shown in Fig. 7.3.

This particle shape can be described in polar coordinates by the following equations:

$$
\begin{aligned}
r &= \frac{a}{\cos\theta} \quad \text{for} \quad 0 \le \theta \le \arctan\left(\frac{b}{a}\right) , \\
r &= a\cos\theta \pm \sqrt{b^2 - a^2\sin^2\theta} \quad \text{for} \quad \arctan\left(\frac{b}{a}\right) < \theta < \arctan\left(\frac{-a}{b}\right) , \\
r &= -\frac{a}{\cos\theta} \quad \text{for} \quad \arctan\left(\frac{-a}{b}\right) \le \theta \le \pi .
\end{aligned}
$$

For this shape we also would like to present scattering results. As an example we computed scattering by a flat circular disc having a total diameter of 6 µm and a thickness of 0.06 µm. This results in an extremely high aspect ratio of 100 : 1. Figure 7.4 presents the corresponding scattering pattern. The incident wavelength is 632.8$nm$ and refractive index is 1.5. The plane wave is incident onto the flat side of the disc. To achieve this convergent computational result 36 discrete sources were needed and the number of integration points needed is 5,000. Further scattering patterns of such flat particles have been published by Hellmers et al. [34].

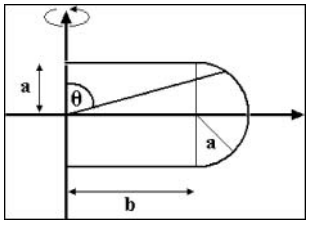



**Fig. 7.3.** Geometry of an oblate circular disc.

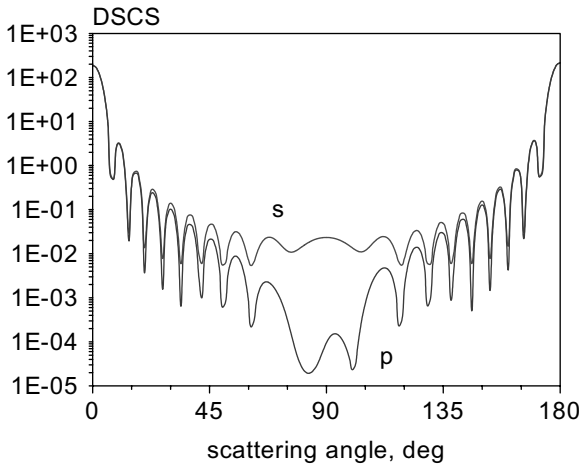



**Fig. 7.4.** Differential scattering cross-section of an oblate circular disc with a radius of $2(a+b) = 6\,\mu\text{m}$ and a thickness of $2a = 0.06\,\mu\text{m}$. Incident wavelength $\lambda = 632.8\,\text{nm}$, refractive index $n = 1.5$.

### 7.5.3 Cassini ovals

The particle shapes investigated so far have been convex in its shape. With NFM-DS it is also possible to calculate scattering by concave shapes, which has till now been considered hardly possible using a standard T-matrix algorithm. To demonstrate this we make use of Cassini ovals. This kind of curves was introduced by Giovanni Domenico Cassini (1625–1712), also known as Jean-Dominique Cassini, in 1680. These curves are characterized in such a way that the product of the distance of two fixed focal points is constant (while for a normal ellipse the sum of the distance of two fixed focal points is constant).

The Cassini ovals have the Cartesian equation:

$$\left[(x-a)^2 + y^2\right]\left[(x+a)^2 + y^2\right] = b^4\,.$$

This leads to:

$$y = \pm\left(-a^2 - x^2 \pm \left(4x^2a^2 + b^4\right)^{\frac{1}{2}}\ \right)^{\frac{1}{2}}\,. \tag{7.24}$$

The corresponding expression in polar coordinates is

$$r = \left(a^2 - 2a^2\sin(\theta)^2 + \left(-4a^4\sin(\theta)^2 + 4a^4\sin(\theta)^4 + b^4\right)^{\frac{1}{2}}\ \right)^{\frac{1}{2}}\,.$$

The Cassini shape therefore depends on the relation $b/a$ (Fig. 7.5).

If $a < b$ the curve is an oval loop, for $a = b$ the result is a lemniscate (like the $\infty$-symbol) and for $a > b$ the curve consists of two separate loops. If $a$ is chosen slightly smaller than $b$ one gets a concave, bone-like shape. The concavity on both sides will get deeper the closer $a$ gets to $b$. By rotating this two-dimensional

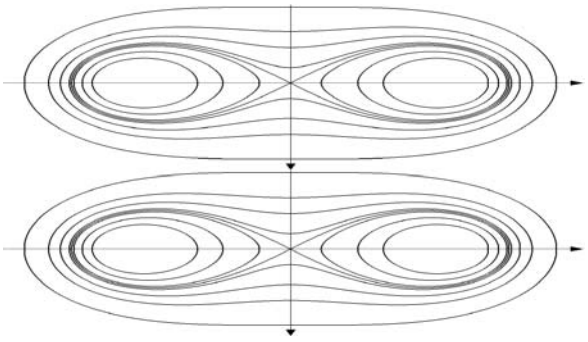


**Fig. 7.5.** Cassini ovals for a given $a$ and varying values of $b$.

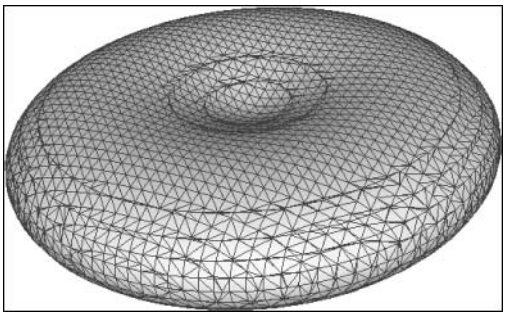


**Fig. 7.6.** 3D shape of a Cassini oval.

curve around the vertical axis we get a three-dimensional particle of the shape of an oblate disc with a concavity on its top and bottom (Fig. 7.6. To get more flexibility and to manipulate the thickness directly a factor $c$ can be introduced as the first term in equation (7.24).

Figure 7.7 shows the light scattering diagram for a Cassini oval based particle with $a = 1.1$, $b = 1.125$ and $c = 0.66$, which gives a total diameter of 3.15 µm and an aspect ratio of approximately 4:1. The incident wavelength is 632.8 nm and refractive index is 1.5. The plane wave is incident onto the flat side of the Cassini oval. For the computation 28 discrete sources positioned in complex plane were used and 1,000 points were needed for integration. Further computational results for concave particles have been published by Hellmers et al. [35] and Wriedt et al. [36].

### 7.5.4 Anisotropic particles

Colour pigments are commonly anisotropic in their refractive index. To be able to investigate scattering by such type of colour pigments the NFM-DS has been extended to compute scattering by uniaxial anisotropic particles [37]. The next scattering diagram (Fig. 7.8) presents scattering by an anisotropic sphere of di-

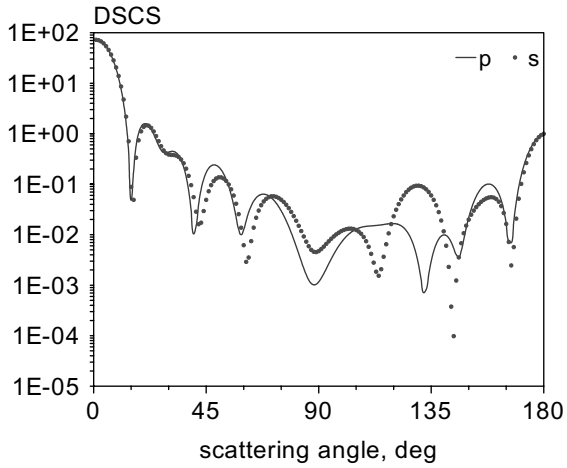



**Fig. 7.7.** Differential scattering cross-section of a Cassini oval with $a = 1.1$, $b = 1.125$ and $c = 0.66$. Incident wavelength is $\lambda = 632.8\,\text{nm}$, refractive index is $n = 1.5$.

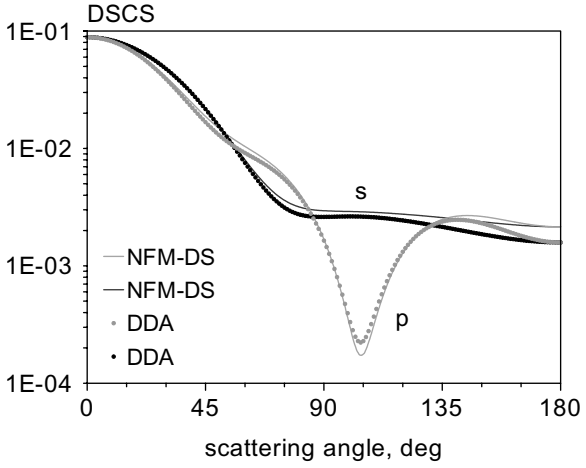



**Fig. 7.8.** Differential scattering cross section of an an unisotropic sphere ($d = 400\,\text{nm}$, $n_x = n_y = 2.5 + j0.725$, $n_z = 4.0 + j1.45$, $\lambda = 498\,\text{nm}$).

ameter $400nm$ and refractive indices of $n_x = n_y = 2.5 + j0.725$, $n_z = 4.0 + j1.45$ alongside scattering results obtained from the DDSCAT program for validation. The plane electromagnetic wave is incident along the $z$-axis and the incident wavelength in this case is 498 nm. The figure demonstrates close agreement between the results of both programs. The related NFM-DS computer programs are also capable computing scattering by rotational symmetric uniaxial anisotropic particles as well as uniaxial anisotropic particles of arbitrary shape without rotational symmetry.

### 7.5.5 Arbitrarily shaped 3D particles

The NFM-DS sources can be used to compute light scattering by arbitrarily shaped 3D particles which lack rotational symmetry. Scattering results have been computed for cubes [21], ellipsoids [3], [38], superellipsoids [39], rough particles [39], and hexagonal prisms [40]. In each case the shape of the scattering particle has to be triangulated into a suitable surface patch model to compute the surface integrals. There are various tools in computational graphics available to handle this problem.

To perform a convergence check versus the number of integration points some flexibility in surface triangulation is needed. Thus methods to increase or reduce the number of surface patches are of great help. To increase the number of surface triangles, the divide by three and the divide by four schemes implemented in the a free DOS program Triangles by David Sharp [41] is suitable. To reduce the number of triangles the SIM Rational Reducer program by Systems in Motion AS [42] can be used.

As an example, scattering by a rounded hexagonal prism has been computed and the results have been compared to results obtained from MMP and CST Microwave Studio. The dimension of the rounded hexagonal prism are $l = 2\,\mu\text{m}$ (rectangular face to rectangular face), $d = 1.15471\,\mu\text{m}$ (hexagonal face to hexagonal face), the refractive index used is $n = 1.5$ and the wavelength of the incident plane wave is 628.319 nm. The plane wave is incident along the $z$-axis and the scattering results are plotted in the $y, z$ scattering plane. The three-dimensional shape of the scattering hexagonal prism is plotted in Fig. 7.9. As can be seen from the scattering plots in Figs 7.10 and 7.11, there is almost perfect correspondence between the results of the different programs for p- as well as for s-polarization.

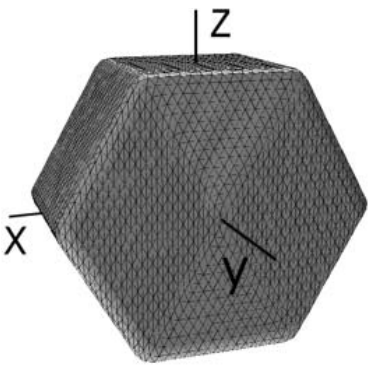



**Fig. 7.9.** Rounded hexagonal prism $l = 2\,\mu\text{m}$, $d = 1.15471\,\mu\text{m}$.

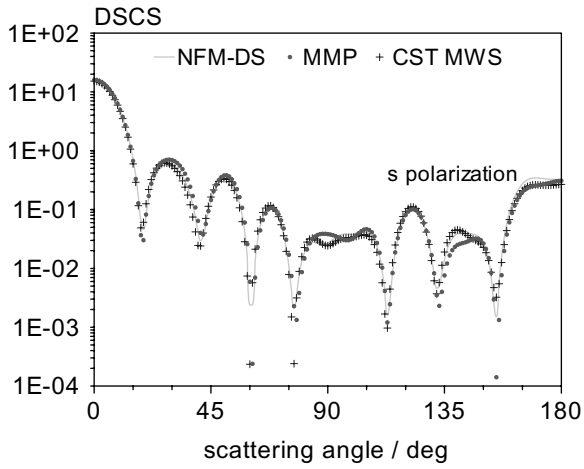



**Fig. 7.10.** Differential scattering cross-section of a rounded hexagonal prism $l = 2\,\mu\text{m}$, $d = 1.15471\,\mu\text{m}$, $n = 1.5$, $\lambda = 628.319\,\text{nm}$.

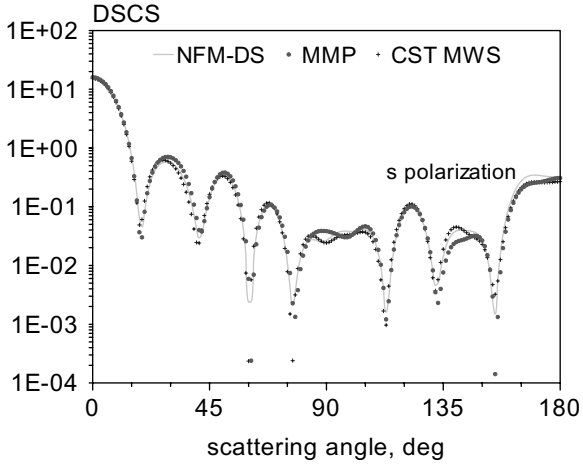



**Fig. 7.11.** Differential scattering cross-section of a rounded hexagonal prism $l = 2\,\mu\text{m}$, $d = 1.15471\,\mu\text{m}$, $n = 1.5$, $\lambda = 628.319\,\text{nm}$.

### 7.5.6 Agglomerates

Characterization of the size and the structure of agglomerates is also needed for many applications such as determination of soot in vehicle diesel exhaust or of aircraft exhaust soot in the atmosphere. For this angular distributions of light scattered by the particles are commonly measured and used for particle characterization. To compute scattering by such type of aggregates particle geometry data have to by generated using a cluster–cluster aggregation algorithm [43]. Figure 7.12 includes the geometry of an exemplary aggregate. The diameter of

**Fig. 7.12.** Figure of cluster cluster aggregate with parameters diameter of primary particle $d = 30\,\text{nm}$, fractal dimension $D_f = 1.3$, radius of gyration $r_g = 4.5462\,\mu\text{m}$, 130 primary spheres.

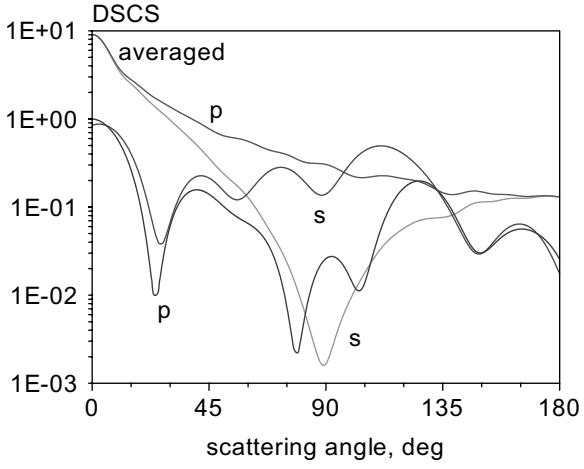



**Fig. 7.13.** Single scattering and orientation averaged DSCS of the cluster of Fig. 7.12 incident wavelength $\lambda = 514\,\text{nm}$, soot refractive index $n = 1.57 + j0.56$.

primary particles is $d = 30\,\text{nm}$, the fractal dimension is $D_f = 1.3$, and the radius of gyration is $r_g = 4.5462\,\mu\text{m}$. There are in total 130 primary spheres in the aggregate. Figure 7.13 shows the differential scattering cross-section of this soot aggregate alongside orientation-averaged scattering. Orientation averaging leads to a damping of oscillations which are present in the single scattering diagram which resembles Rayleigh scattering.

### 7.5.7 Inclusions

Commonly with particles having inclusions, the T-matrix method is restricted to spherical inclusions. Using a multiple scattering approach we extended the NFM-DS to handle scattering by off-center nonspherical inclusions. A full description of the theory and additional computational results are available in Doicu and Wriedt [44] and Schuh and Wriedt [45].

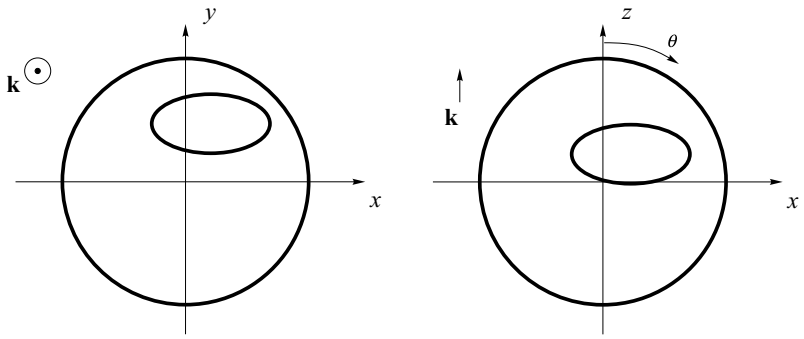



**Fig. 7.14.** Geometry of sphere with spheroidal inclusion, sphere size parameter $kr = 10$, spheroid size parameters $ka = 5$, $kb = 2.5$ placed at $kx_i = 2$, $ky_i = 4$, $kz_i = 2$; orientation Euler angles $\alpha = 0°$, $\beta = 90°$.

We consider a spherical particle with a spheroidal inclusion. The geometry of the scattering problem is shown in Fig. 7.14. The size parameter of the sphere is $kr = 10$, the refractive index is $n = 1.334$. The size parameters of the spheroid are $ka = 5$, $kb = 2.5$ placed at $kx_i = 2$, $ky_i = 4$, $kz_i = 2$; orientation Euler angles of the inclusion are $\alpha = 0°$, $\beta = 90°$. The refractive index of the inclusion is $n = 1.6$. In Figs 7.15 and 7.16, we plot the differential scattering cross-section computed with DFM-NS and the multiple multipole method for validation. There is close agreement between the two computational results.

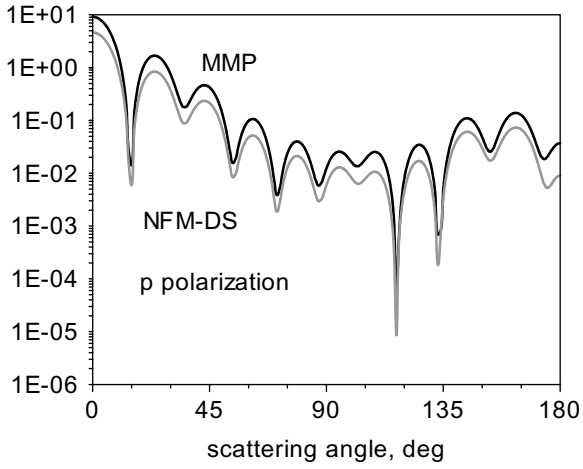



**Fig. 7.15.** DSCS of a sphere with a prolate spheroid inclusion with the scattering geometry of Fig. 7.14.

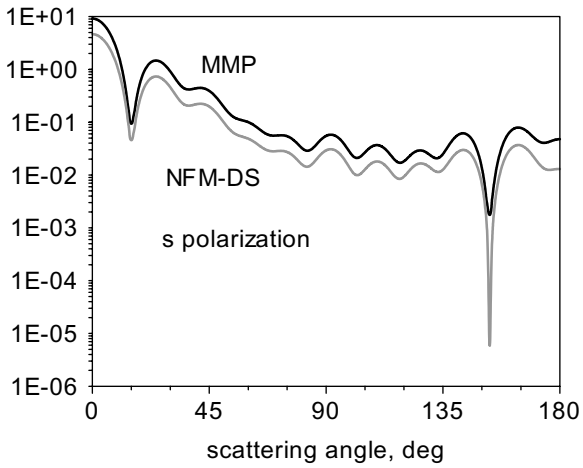



**Fig. 7.16.** DSCS of a sphere with a prolate spheroid inclusion with the scattering geometry of Fig. 7.14.

### 7.5.8 Particles on surfaces

Optical characterization of defects such as particles, bumps or pits on a silicon wafer surface is of great importance in semiconductor manufacturing. As semiconductor device dimensions become smaller, there is a need for optical wafer surface scanning systems to detect the size and composition of microcontaminations to sizes as low as 0.1 µm or even smaller. To expand the current detection ability an efficient mathematical model and computer simulation technique is needed. There are some appoximate approaches to computing scattering by a dielectric particle on a plane dielectric surface ([28] pp. 192).

But as these approaches are no longer sufficient for a modern design of particle surface scanners the NFM-DS has been extended to handle the particle surface scattering problem in an exact way. The theory is fully described in a book contribution by Doicu and Wriedt [46].

As an exemplary computational result, we present the scattering plot in Fig. 7.17 for a 1.09 µm diameter polystyrene sphere on a plane silicon surface with the following parameters: sphere diameter $d = 1.09\,\mu\text{m}$, polystyrene refractive index $n = 1.64$, silicon refractive index $n = 4.90 + i3.84$, incident wavelength $\lambda = 308\,\text{nm}$. The plane wave is incident normal to the plane surface.

Plasmon resonance phenomena, that is the local amplification of light by nanoscale silver or gold noble particles, have potential applications for biosensors, bio-labels and nano-optical devices. Plasmon resonances of small noble metal spheres can be detected as peaks in the measured light scattering spectra. Transmission dark-field microscopy is a technique where only the particles scatter light into the direction of the microscope objective. Such a measuring device can visualize very small particles as coloured discs. The surface plasmon resonance frequency from a nonspherical particle or a particle aggregate

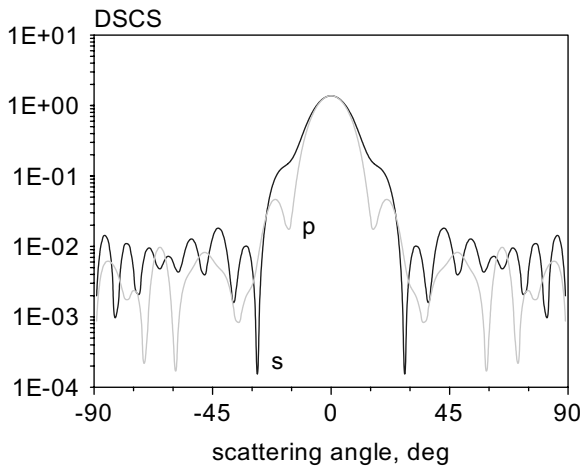



**Fig. 7.17.** DSCR of a sphere (diameter $d = 1.09\,\mu\text{m}$, $n = 1.64$), on a plane silicon surface with $n = 4.90 + j3.84$ and an incident wavelength $\lambda = 308\,\text{nm}$.

is different compared to a single spherical particle. With this effect, measuring techniques which use white light illumination are capable of differentiating between aggregated particles and a single particle because of their different color. Even when bioreceptor molecules attached to a gold or silver sphere detect its biomolecular counterpart, the resonance frequency is shifted. In the following we give some simulation examples of particles on or near a plane surface. We compute intensities at different scattering angles over the visible spectrum of wavelengths of small particles with diameter $d = 80\,\text{nm}$. The intensities will be detector-integrated over a range of $\phi_{NA} = 25°$ which corresponds to a numerical aperture of $NA = n\sin(NA)$ of the objective lens. The particles consist of silver. The wavelength-dependent refractive indices are interpolated values from Johnson et al. [47]. The numerical aperture depends on the medium surrounding of the particle. The incident beam angle with respect to the normal is $\phi_0 = 30°$ and the surrounding medium is water ($n = 1.333$).

From the computational results of the detector-integrated DSCS printed in Fig. 7.18 we see that the frequency and the form of the resonance peak very much depends on the distance of the silver particle from the plane silver surface. With a higher distance it shifts to a shorter wavelength and it also becomes broader.

## 7.6 Validation

An important step in development of the NFM-DS theory and the related computer programs is validation by comparing to results obtained from other programs. Although in scattering research measurement results were applied for validation of theory and corresponding programs [48] we did not consider this

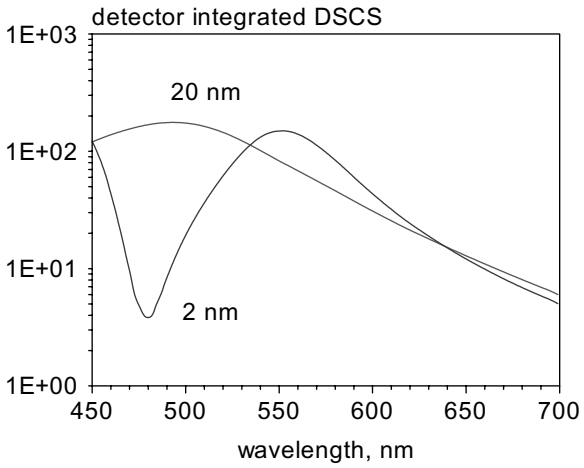



**Fig. 7.18.** Detector integrated DSCS of a silver sphere ($d = 80\,\mathrm{nm}$) on a silver plane surface with different heights of the particle.

as a suitable approach in our program development. In light scattering optics, it is much more difficult to obtain accurate measurements suitable for validation. Therefore we prefer to use other computation results for validation. Alongside validation it is, of course, important to care for convergence with respect to the number and order of discrete sources used in the simulation and with respect to the number of integration points used in computing the relevant surface integrals. The number of integration points proved to be much more stable than the number and order of discrete sources. In some cases we even found some plateau in the number and order of discrete sources where we obtained convergent scattering results. Increasing the number or order of discrete sources beyond this plateau again led to nonconvergence, which can make a convergence check somewhat tricky.

For program validation three different approaches have been used. In the first method, two different implementations of NFM-DS have been used for validation. For example, the program for the composite scatterers was compared to results of the multiple scattering program. In this case the composite particle consisted of three parts with the center part free space such that the problem could also be treated as a two-particle multiple scattering problem [40].

In the second approach we developed the Discrete Sources Method (DSM) together with the research group of Yuri Eremin, Lomonosov Moscow State University, for the same scattering problems. One may almost speak about co-evolution in program development. Examples are the flat disc [35], Cassini ovals [34], the long fibre [49] and total internal reflection microscopy (TIRM) [50].

In the third method we used other freely available programs for validation. In this we focused on three methods. These are Multiple Multipole Program (MMP) by Bomholt and Hafner [51], Discrete Dipole Approximation (DDA) [52] implemented in the FORTRAN program DDSCAT [53], and in various implementations of Finite Different Time Domain (FDTD) and the Volume Integral

Equation Method (VIEM) [54] and the related Finite Integration Technique FIT [55]. FIT is implemented in the commercial CST Microwave Studio program [56]. We also had a look at other commercial computational electromagnetics programs but found the CST Microwave Studio program to be the most suitable. With the other programs tested we hardly could input particle shape data or the program was restricted to small particle sizes compared to the incident wavelength. The reason for this is that with most programs not only the surface of a scatterer but the full volume of the scattering particle has to be discretized, which leads to a high demand in computer resources.

## 7.7 Applications

In this section we briefly mention the broad range of applications the computational programs based on NFM-DS have found in solving practical technological applications in recent years.

Phase Doppler Anemometry (PDA) was extended to size spheroidal particles [57,58]. The NFM-DS was used to develop an optical instrument for mineral and asbestos fibre characterization [32].

Extensive light scattering computations for aggregated particles helped to characterize soot particles in the flame of a Bunsen burner [43].

There is an ongoing project for characterization of red blood cells using light scattering [35]. Within a collaborative project total internal reflection microscopy (TIRM) for measurement of nanoforces acting on a colloidal particle near a plane surface has been developed [50]. Sensors for surface particle or surface defect characterization have been developed based on the NFM-DS.

## 7.8 Conclusion

In this chapter an overview of the progress in developing the NFM-DS has been given. Some open problems are still left. These include bi-anisotropic particles, bi-anisotropic host media, chiral media, nonaxisymmetric compound particles and optimal deposition of discrete sources and faces for surface integration.

Most of the FORTRAN programs developed within this project have been published on CD with a monograph on the Null-Field Method with Discrete Sources [40]. For further applications not covered by this review, such as chiral particles, layered particles [59], composite particles [60], Gaussian laser beam scattering [61], evanescent wave scattering [62] and dipole scattering the interested reader is referred to this book or to the cited papers.

**Acknowledgments**

I would like to thank Adrian Doicu for constant cooperation and Jens Hellmers, Roman Schuh, Norbert Riefler and Elena Eremina for providing computational results for this review. This work was supported by Deutsche Forschungsgemeinschaft (DFG).

## 7.9 Symbols and abbreviations

| | |
|---|---|
| $(a_\nu^0, b_\nu^0)$ | expansion coefficients of the incident field |
| $(a_\nu^0, b_\nu^0)$ | expansion coefficients of the incident field |
| $D_{mn}$ | normalization constant |
| $\mathbf{E}_0, \mathbf{E}_s$ | incident and scattered fields |
| $(\mathbf{e}, \mathbf{h})$ | surface current densities |
| $(f_\nu^0, g_\nu^0)$ | expansion coefficients of the scattered field |
| $k$ | wavenumber |
| $\{\mathbf{M}_{mn}^{1,3}, \mathbf{N}_{mn}^{1,3}\}$ | localized vector spherical functions |
| $\{\mathcal{M}_{mn}^{1,3}, \mathcal{N}_{mn}^{1,3}\}$ | distributed vector spherical functions |
| $\{\mathcal{M}_{ni}^{1,3}, \mathcal{N}_{ni}^{1,3}\}$ | magnetic and electric dipoles |
| $\{\mathcal{M}_{n}^{1,3}, \mathcal{N}_{n}^{1,3}\}$ | vector Mie-potentials |
| $n$ | refractive index |
| $S$ | particle surface |
| $[\mathbf{T}]$ | transition matrix |
| $(x, y, z)$ | Cartesian coordinate |
| $\mathbf{x}$ | position vector |
| $\alpha_n^{\pm}(\mathbf{x})$ | Green function |
| $\alpha, \beta, \gamma$ | Euler angles |
| $(\vartheta, \varphi)$ | angular coordinates |
| $\epsilon$ | permittivity |
| $\lambda_0$ | wavelength in vacuum |
| $\sigma_d/\pi a^2$ | normalized differential scattering cross-section (DSCS) |
| $\mu$ | permeability |
| $v_i$ | vertex points on particle surface |

## References


1. Y. A. Eremin and A. G. Sveshnikov: A computer technology for the discrete source method in scattering problems. *Computational Mathematics and Modeling*, **14** (2003) 1, 16–25.
2. V. D. Kupradze and M. A. Aleksidze: The method of functional equations for the approximate solution of certain boundary value problems. *USSR Comput. Math. Math. Phys.*, **4** (1964) 4, 82–126.
3. A. Doicu, Y. A. Eremin, T. Wriedt: *Acoustic and Electromagnetic Scattering Analysis Using Discrete Sources*. Academic Press, San Diego 2000.
4. G. Fairweather and A. Karageorghis: The method of fundamental solutions for elliptic boundary value problems. *Adv. Comput. Math.*, **9** (1998), 69–95.
5. G. Fairweather, A. Karageorghis, P. A. Martin: The method of fundamental solutions for scattering and radiation problems. *Engineering Analysis with Boundary Elements*, **27** (2003), 759–769.
6. T. Wriedt (Ed.): *Generalized Multipole Techniques for Electromagnetic and Light Scattering*. Elsevier, Amsterdam 1999.
7. M. Katsurada, A mathematical study of the charge simulation method II, *J. Fac. Sci.*, Univ. of Tokyo, Sect. 1A, Math. **36** (1989), 135–162.

8. M. Kawano, H. Ikuno, M. Nishimoto: Numerical analysis of 3-D scattering problems using the Yasuura method. *IEICE Trans. Electron.* **E79-C** (1996), 1358–1363.
9. Ch. Hafner, *Post-modern Electromagnetics Using Intelligent MaXwell Solvers*. John Wiley & Sons, Chichester 1999.
10. D. I. Kaklamani and H. T. Anastassiu: Aspects of the method of auxiliary sources in computational electromagnetics. *IEEE Antennas and Propagation Magazine* **44** (2002) 3, 48–64.
11. M. Nishimura, S. Takamatsu, and H. Shigesawa: A numerical analysis of electromagnetic scattering of perfect conducting cylinders by means of discrete singularity method improved by optimization process. *Electronics and Communications in Japan*, **67-B** (1984) 5, 552–558.
12. Y. Leviatan, Z. Baharav, E. Heyman: Analysis of electromagnetic scattering using arrays of fictitious sources. *IEEE Trans. Antennas Propagat.*, **AP-43** (1995) 10, 1091–1098.
13. D. Maystre, M. Saillard, G. Tayeb: Special methods of wave diffraction. in P. Sabatier and E.R. Pike (Eds): *Scattering*. Academic Press, London 2001.
14. M. A. Golberg and C. S. Chen: The method of fundamental solutions for potential, Helmholtz and diffusion problems. In M. A. Golberg (Ed.): *Boundary Integral Methods and Mathematical Aspects*. WIT Press/Computational Mechanics Publications, Boston 1999, 103–176.
15. A. C. Ludwig: The generalized multipole technique. *Comput. Phys. Commun.* **68** (1991), 306–314.
16. P. C. Waterman: Numerical solution of electromagnetic scattering problems. In R. Mittra: *Computer techniques for electromagnetics*. Pergamon Press, New York 1973, 97–157.
17. P. C. Waterman: Matrix formulation of electromagnetic scattering. *Proc. IEEE* **53** (1965), 803–812.
18. V. K. Varadan and V. V. Varadan (Ed): *Recent Developments in Classical Wave Scattering: Focus on the T-matrix*. Pergamon Press, Oxford 1980.
19. P. W. Barber: Differential scattering of electromagnetic waves by homogeneous isotropic dielectric bodies. Ph.D. thesis, University of California, Los Angeles 1973.
20. J. B. Schneider and I. C. Peden: Differential cross section of a dielectric ellipsoid by the T-matrix extended boundary condition method. *IEEE Trans. Antennas Propagat.* **AP 36** (1978), 1317–1321.
21. T. Wriedt and A. Doicu: Formulation of the extended boundary condition method for three-dimensional scattering using the method of discrete sources. *Journal of Modern Optics* **45** (1998) 1, 199–213.
22. H. Laitinen and K. Lumme: T-matrix method for general star-shaped particles: first results. *J. Quant. Spectrosc. Radiat. Transfer*, **60** (1998), 325–334.
23. F. M. Kahnert, J. J. Stamnes, K. Stamnes: Application of the extended boundary condition method to particles with sharp edges: a comparison of two surface integration approaches. *Applied Optics*, **40** (2001), 3101–3109.
24. S. Havemann and A. J. Baran: Extension of T-matrix to scattering of electromagnetic plane waves by non-axisymmetric dielectric particles: application to hexagonal ice cylinders. *J. Quant. Spectrosc. Radiat. Transfer*, **70** (2001), 139–158.
25. M. I. Mishchenko, L. D. Travis, D. W. Mackowski: T-matrix computations of light scattering by nonspherical particles: a review. *J Quant Spectrosc. Radiat. Transfer*, **55** (1996), 535–575.
26. M. I. Mishchenko, J. W. Hovenier, L. D. Travis (Eds): *Light Scattering by Nonspherical Particles*. Academic Press, San Diego 2000.

27. M. I. Mishchenko, L. D. Travis, A. A. Lacis: *Scattering, Absorption, and Emission of Light by Small Particles*. Cambridge University Press, Cambridge 2002.
28. F. Borghese, P. Dentl, R. Saija: *Scattering from Model Nonspherical Particles. Theory and Applications to Environmental Physics*. Springer Verlag, Berlin 2003.
29. M. I. Mishchenko, G. Videen, V. A. Babenko, N. G. Khlebtsov, T. Wriedt: T-matrix theory of electromagnetic scattering by particles and its applications: a comprehensive reference database. *J. Quant. Spectrosc. Radiat. Transfer*, **88** (2004), 357–406.
30. HyperFun Project, Language and Software Tools for F-rep Geometric Modeling. www.hyperfun.org.
31. K. Georg and J. Tausch: Some error estimares for the numerical approximation of surface integrals. *Math. Comp.*, **62** (1994), 755–763.
32. S. Pulbere, T. Wriedt: Light scattering by cylindrical fibers with high aspect ratio using the null-field method with discrete sources *Part. Part. Syst. Charact.*, **21** (2004), 213–218.
33. A. Doicu and T. Wriedt: Extended boundary condition method with multipole sources located in the complex plane. *Optics Commun.*, **139** (1997), 85–91.
34. J. Hellmers, T. Wriedt, A. Doicu: Light scattering simulation by oblate disc spheres using the null field method with discrete sources located in the complex plane. *Journal of Modern Optics*, **53** (2006) 3, 267–282.
35. J. Hellmers, E. Eremina, T. Wriedt: Simulation of light scattering by biconcave Cassini ovals using the nullfield method with discrete sources. *Journal of Optics A*, **8** (2006), 1–9.
36. T. Wriedt, J. Hellmers, E. Eremina, R. Schuh: Light scattering by single erythrocyte: Comparison of different methods. *J. Quant. Spectrosc. Radiat. Transfer*, **100** (2006), 444–456.
37. A. Doicu: Null-field method to electromagnetic scattering from uniaxial anisotropic particles. *Optics Commun.*, **218** (2003) 1–3, 11–17.
38. J. Hellmers and T. Wriedt: Influence of particle shape models on T-matrix light scattering simulation. *J. Quant. Spectrosc. Radiat. Transfer*, **89** (2004), 97–110.
39. T. Wriedt: Using the T-matrix method for light scattering computations by non-axisymmetric particles: Superellipsoids and realistically shaped particles. *Part. Part. Syst. Charact.*, **19** (2002) 4, 256–268.
40. A. Doicu, T. Wriedt, Yuri Eremin: *Light Scattering by Systems of Particles. Null-Field Method with Discrete Sources – Theory and Programs*. Springer Verlag, Berlin, Heidelberg, New York 2006.
41. Triangles, a triangle mesh construction utility: http://www.geocities.com/Athens/Academy/8764/triangles.htm.
42. Systems in Motion AS, www.sim.no.
43. N. Riefler, S. di Stasio, T. Wriedt: Structural analysis of clusters using configurational and orientational averaging in light scattering analysis. *J. Quant. Spectrosc. Radiat. Transfer*, **89** (2004), 323–342.
44. A. Doicu and T. Wriedt: T-matrix method for electromagnetic scattering from scatteres with complex structure. *J. Quant. Spectrosc. Radiat. Transfer*, **70** (2001), 663–673.
45. R. Schuh and T. Wriedt: Computer programs for light scattering by particles with inclusions. *J. Quant. Spectrosc. Radiat. Transfer*, **70** (2001), 715–723.
46. T. Wriedt and A. Doicu: T-matrix method for light scattering from particles on or near an infinite surface. in F. Moreno and F. González (Eds): *Light Scattering from Microstructures*. Springer Verlag, Berlin 2000, 113–132.

47. P. B. Johnson and R. W. Christy: Optical constants of noble metals. *Phys. Rev. B*, **6** (1972), 4370–4379.
48. A. Hizal and A. Marincic: New rigorous formulation of electromagnetic scattering from perfectly conducting bodies of arbitrary shape. *Proc. IEE*, **117** (1970) 8, 1639-1647.
49. E. Eremina, Y. Eremin, T. Wriedt: Review of light scattering by fiber particles with a high aspect ratio. *Recent Res. Devel. Optics*, **3** (2003), 297–318.
50. L. Helden, E. Eremina, N. Riefler, Ch. Hertlein, C. Bechinger, Y. Erimin, T. Wriedt: Single particle evanescent light scattering simulations for total internal reflection microscopy. *Applied Optics*, submitted.
51. Ch. Hafner and L. Bomholt: *The 3D Electromagnetic Wave Simulator, 3D MMP Software and User's Guide*. Wiley, Chichester 1993.
52. B. T. Draine and P. J. Flatau: Discrete-dipole approximation for scattering calculations. *J. Opt. Soc. Am. A*, **11** (1994), 1491–1499.
53. The Discrete Dipole Approximation for Scattering and Absorption of Light by Irregular Particles. www.astro.princeton.edu/∼draine/DDSCAT.html.
54. T. Wriedt and U. Comberg: Comparison of computational scattering methods. *J. Quant. Spectrosc. Radiat. Transfer*, **60** (1998) 3, 411–423.
55. T. Weiland: A discretization method for the solution of Maxwell's equation for six-components fields. *Electron. Commun. AEÜ*, **31** (1977), 116–120.
56. Computer Simulation Technology GmbH, www.cst.de.
57. A. Doicu, T. Wriedt, K. Bauckhage: Light scattering by homogeneous axisymmetrical particles for PDA calculations to measure both axes of spheroidal particles. *Part. Part. Syst. Charact.*, **14** (1997), 3–11.
58. A. Doicu, J. Köser, T. Wriedt, K. Bauckhage: Light scattering simulation and measurement of monodisperse spheroids using a Phase Doppler Anemometer. *Part. Part. Syst. Charac.*, **15** (1999), 257–262.
59. A. Doicu and T. Wriedt: Null-field method with discrete sources to electromagnetic scattering from layered scatterers. *Comput. Phys. Commun.*, **138** (2001), 136–142.
60. A. Doicu and T. Wriedt: Null-field method with discrete sources to electromagnetic scattering from composite scatterers. *Optics Commun.*, **190** (2001), 13–17.
61. A. Doicu and T. Wriedt: Formulations of the extended boundary condition method for incident Gaussian beams using multiple multipole expansions. *Journal of Modern Optics*, **44** (1997) 4, 785–801.
62. A. Doicu, Yu. Eremin, T. Wriedt: Scattering of evanescent waves by a particle on or near a plane surface. *Comput. Phys. Commun.*, **134** (2001), 1–10.

# 8 Radiative transfer in horizontally and vertically inhomogeneous turbid media

O. V. Nikolaeva, L. P. Bass, T. A. Germogenova, V. S. Kuznetsov,
A. A. Kokhanovsky

## 8.1 Introduction

Radiative transfer through turbid media is usually modeled on the basis of the stationary radiative transfer equation (RTE). As a rule, in addition various approximations of the radiative transfer equation, such as the spherical harmonics equations or small angle approximations, are used. The spherical harmonics equations are relevant for transport problems in optically thick and weakly heterogeneous media, whereas small angle approximation works well for radiation transfer problems in media characterized by the phase functions peaked in the forward scattering direction.

Modern methods of atmospheric research require multiply scattered radiation field intensity calculations at large discrete spatial-angular arrays for detailed modeling of light scattering media. Besides, radiation field calculations in the wide range of wavelengths are necessary to estimate solar radiation influence on weather and climate variations. In some cases, however, the information on the scattered radiation field for few wavelengths is sufficient.

So we shall consider the radiative transfer equation for the fixed wavelength value, rather than one for a spectral line. RTE for the radiance $\psi(\vec{r},\vec{\Omega})$ can be written in the form

$$\hat{L}\psi = B(\vec{r},\vec{\Omega})\,, \quad B(\vec{r},\vec{\Omega}) = \hat{S}\psi(\vec{r},\vec{\Omega}) + F(\vec{r},\vec{\Omega})\,, \tag{8.1}$$

where

$$\vec{L}\psi = \vec{\Omega}\cdot\vec{\nabla}\psi(\vec{r},\vec{\Omega}) + K_{\rm ext}(\vec{r})\psi(\vec{r},\vec{\Omega}) \tag{8.2}$$

is a differential transport operator,

$$\hat{S}\psi = K_{\rm ext}(\vec{r})\,\bar{\omega}_0(\vec{r})\int_{-1}^{+1}\mathrm{d}\gamma\int_0^{2\pi}\mathrm{d}\varphi\,p(\vec{r},\vec{\Omega}\vec{\Omega}')\,\psi(\vec{r},\vec{\Omega}') \tag{8.3}$$

is an integral operator of scattering, $\vec{\Omega}$ is the direction of propagation, defined by azimuth angle $\varphi$ and value $\gamma = \cos\theta$, where $\theta$ is polar angle (see Fig. 8.1).

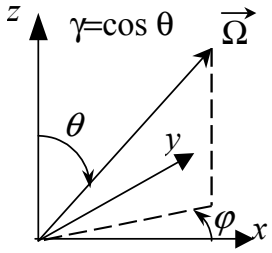



**Fig. 8.1.** Vector $\vec{\Omega}$.

The extinction coefficient $K_{\rm ext}(\vec{r})$, the single scattering albedo $\bar{\omega}_0(\vec{r})$ and the scattering phase function $p(\vec{r},\chi)$ are suggested to be positive. The following normalization condition for the scattering phase function $p(\vec{r},\chi)$ is used:

$$\int_{-1}^{+1}\int_{0}^{2\pi} {\rm d}\gamma\,{\rm d}\varphi\, p(\vec{r},\vec{\Omega}\cdot\vec{\Omega}') = 1\,, \tag{8.4}$$

where $\chi = \vec{\Omega}\cdot\vec{\Omega}'$ is the inner product of vectors $\vec{\Omega}(\gamma,\varphi)$ and $\vec{\Omega}'(\gamma',\varphi')$,

$$\vec{\Omega}\cdot\vec{\Omega}' = \gamma\,\gamma' + \sqrt{1-\gamma^2}\,\sqrt{1-(\gamma')^2}\,\cos(\varphi-\varphi')\,. \tag{8.5}$$

The function $F(\vec{r},\vec{\Omega})$ defines solar and heat radiation sources. For formulation of a transport problem in the atmospheric region $G$ with the boundary $\Gamma$, some boundary conditions on $\Gamma$ should be also defined:

$$\psi(\vec{r},\vec{\Omega}) = \overline{\psi}(\vec{r},\vec{\Omega}) = A(\vec{r})\,\widehat{R}\,\psi(\vec{r},\vec{\Omega}) + \psi_0(\vec{r},\vec{\Omega})\,, \quad \text{at } \vec{r}\in\Gamma,\ \vec{\Omega}\cdot\vec{n}(\vec{r}) < 0\,. \tag{8.6}$$

Here $\vec{n}(\vec{r})$ is the external normal at the point $\vec{r}$ of the boundary surface $\Gamma$ and the function $\overline{\psi}(\vec{r},\vec{\Omega})$ defines the radiation intensity entering the region G, $\psi_0(\vec{r},\vec{\Omega})$ is radiation intensity of the source on the boundary surface, $A(\vec{r}) \in [0,1]$ is the reflection albedo. The operator $\widehat{R}$, defining the radiation reflection for the surface $\Gamma$, can be written as

$$\widehat{R}\,\psi(\vec{r},\vec{\Omega}) = \int_{\vec{\Omega}'\cdot\vec{n}>0} {\rm Re}\,(\vec{r},\vec{\Omega},\vec{\Omega}')\psi(\vec{r},\vec{\Omega}')\,{\rm d}\vec{\Omega}'\,,$$

${\rm Re}\,(\vec{r},\vec{\Omega},\vec{\Omega}')$ is the bi-directional surface reflectance, normalized by equality

$$\widehat{R}\,[1] = \int_{\vec{\Omega}'\cdot\vec{n}>0} {\rm Re}\,(\vec{r},\vec{\Omega},\vec{\Omega}')\,{\rm d}\vec{\Omega}' = 1\,.$$

In particular, the operator

$$\hat{R}\,\psi(\vec{r},\vec{\Omega}) = \frac{1}{\int_{(\vec{\Omega}'\cdot\vec{n})>0}(\vec{\Omega}'\cdot\vec{n})\,{\rm d}\vec{\Omega}'}\int_{(\vec{\Omega}'\cdot\vec{n})>0}(\vec{\Omega}'\cdot\vec{n})\,\psi(\vec{r},\vec{\Omega}')\,{\rm d}\vec{\Omega}'$$

corresponds to the well known Lambert reflection law, the operator $\hat{R}\,\psi(\vec{r},\vec{\Omega}) = \psi(\vec{r},\vec{\Omega}^*)$ defines mirror reflection processes (here the function $\vec{\Omega}^*(\vec{r},\vec{\Omega})$ defines

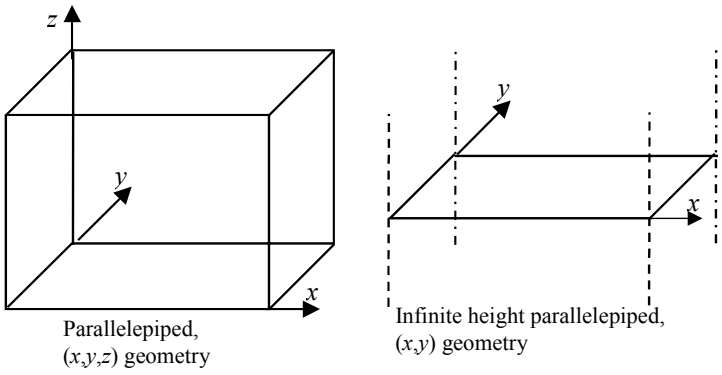



**Fig. 8.2.** Calculation regions with Cartesian coordinate systems.

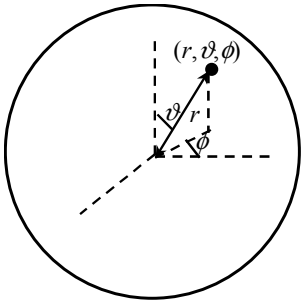



**Fig. 8.3.** Calculation region with the spherical coordinates system.

vector $\vec{\Omega}^*$, symmetrical to vector $\vec{\Omega}$ relatively the plane perpendicular to surface $\Gamma$ at the point $\vec{r}$), whereas the operator $\hat{R}\,\psi(\vec{r},\vec{\Omega}) = \psi(\vec{r},-\vec{\Omega})$ corresponds to the returned scattering law, relevant for irregular surfaces.

A parallelepiped is used as a calculation region G in many cases, and so three coordinates $\{x, y, z\}$ define the vector $\vec{r}$ introduced above. Obviously, it is the case of $(x, y, z)$-geometry, depicted in Fig. 8.2.

More rarely the spherical coordinates $\vec{r} = \{r, \vartheta, \varphi\}$ are used (see Fig. 8.3). There exist situations, when the solution $\psi(\vec{r},\vec{\Omega})$ is independent on one of the variables. Then the corresponding variable is vanished (see Fig. 8.2).

Numerical algorithms for transport equation solving are frequently based either on the stochastic Monte–Carlo (MC) method (Marchuk et al., 1980) or the deterministic Discrete Ordinates Method (DOM) (Chandrasekhar, 1950; Bass et al., 1986). These two methods usually complement each other, each possessing some advantages and shortcomings.

The MC method can easily allow us to take into account the complicated spatial structure of calculation regions. However, it provides solutions of the numerical transport problems only for a small number of radiation detectors. But, as is clear, the smaller detector array size in phase space $\{\vec{r}, \vec{\Omega}\}$ the less the accuracy of calculation of the functional, corresponding to the detector.

Obviously, the smaller the detector size the fewer photons it receives. So, in optically thick media (thick clouds or twilight atmosphere), a large number of histories is necessary. A more complicated way of region nonhomogeneity accounting is used in the DOM method compared to the MC method. However, DOM provides the solution in the whole spatially-angular grid located inside the calculation region, independently of the region size. Test calculations confirm high DOM accuracy even in the case when the solution undergoes variations over 18–20 orders. On the other hand, due to the necessity of radiation field calculation in the whole region, DOM is sometimes proved more computationally expensive as compared to the MC method.

The slab model approximation is relevant for radiation transfer problems in horizontally homogeneous regions (such as a cloud or clear sky). If slab optical thickness is more than 5–8 optical lengths, asymptotic methods may be used for reflected and transferred light intensity calculations. For complicated radiation transport problems, various decomposition techniques have been developed, consisting of combinations of analytical, asymptotic and simple numerical methods. The development of multiprocessor computers enables us to increase DOM codes efficiency. Parallel calculation algorithms enable us to decrease DOM calculation time and so to come close to MC, despite the MC itself parallelizing. Besides, the nonhomogeneity approximation accuracy via DOM is currently almost as high as that via the MC method.

Below we shall consider discrete ordinates methods for the transport equation, namely: the specification of transport equation parameters (section 8.2); the construction of angular (section 8.3) and spatial (section 8.6) grids; the scattering integral $\hat{S}$ (section 8.4) and the differential operator $\hat{L}$ (section 8.6) approximations; and the separation of diffused and direct light (section 8.5). In correspondence with the introduced classification (section 8.6) various grids schemes will be considered (sections 8.7–8.10), methods of grid equations will be presented (section 8.11) and parallel calculation organization will be introduced (section 8.12). Code implementations of various DOM versions for atmosphere optics problems will be given in sections 8.13 and 8.14.

## 8.2 Description of the calculation region

We present results of radiation field calculations in a cloudy atmosphere on the set of pixels. Each pixel size (on the $(x, y)$ plane) corresponds to the required calculation accuracy, vertical atmosphere characteristics (over $z$-axis) being considered as known. An example of the representation of atmosphere by $3.6 \times 10^5$ pixels is given by Cahalan et al. (2005).

In each pixel, the transfer equation coefficient behavior over the height ($z$-axis) is suggested to be known (via the direct problem solution). The spatial grid mesh in the calculation region cannot be larger than the corresponding pixel size.

Proper approximations for cross-sections of aerosol and molecular light scattering processes and light absorption by various gases are used in transport problems. Corresponding coefficients in the transport equation are then obtained as

weighted mean values over aerosol, molecular, gaseous and cloudy contributions. For example, for a cloud consisting of water ($w$) and ice ($i$), transport equation coefficients are defined by formulas

$$K_{\text{ext}} = K_{ext,w} + K_{ext,i}\,, \quad \omega_0 = \frac{\omega_{0,w} K_{ext,w} + \omega_{0,i}\, K_{ext,i}}{K_{ext,w} + K_{ext,i}}\,,$$

$$p = \frac{\omega_{0,w} K_{ext,w} p_w + \omega_{0,i} K_{ext,i} p_i}{\omega_{0,w} K_{ext,w} + \omega_{0,i} K_{ext,i}}\,.$$

These coefficients are usually approximated by constants either inside the pixel or inside the spatial grid mesh.

## 8.3 Discrete ordinates method and a angular quadratures

The discrete ordinates method is based on the transfer from the continuous angular dependence of the transport equation solution to the discrete one, that is on the introduction of the angular quadrature over variables $\theta$ and $\varphi$ on the unit sphere $\{-1 < \gamma < 1,\ 0 < \varphi < 2\pi\}$.

Let us consider such a quadrature for the hemisphere $\{-1 < \gamma < 1, 0 < \varphi < \pi\}$. At first, the interval $-1 < \gamma < 1$ is divided into subintervals

$$\begin{aligned} -1 &= \gamma_{-L-1/2} < \cdots < \gamma_{-\ell-1/2} < \ldots \gamma_{-1/2} \\ &= 0 = \gamma_{1/2} < \cdots < \gamma_{\ell+1/2} < \ldots \gamma_{L+1/2} = 1 \end{aligned}$$

with nodes $\gamma_\ell$, $\ell = 1, \ldots, L$, being chosen inside each subinterval:

$$\gamma_\ell \in [\gamma_{\ell-1/2}, \quad \gamma_{\ell+1/2}]\,, \quad \gamma_{-\ell} \in [\gamma_{-\ell-1/2},\ \gamma_{-\ell+1/2}]\,, \quad \Delta\gamma_\ell = \left|\gamma_{\ell+1/2} - \gamma_{\ell-1/2}\right|\,.$$

A similar subdivision of interval $0 < \varphi < \pi$ is introduced for each node $\gamma_\ell$:

$$0 = \varphi_{\ell,1/2} < \cdots < \varphi_{\ell,m+1/2} < \cdots < \varphi_{\ell,M_\ell+1/2} = \pi\,,$$

$$\Delta\varphi_{\ell,m} = \varphi_{\ell\,,m+1/2} - \varphi_{\ell,m-1/2}\,, \quad \varphi_{\ell,m} \in (\varphi_{\ell\,,m-1/2},\ \varphi_{\ell,m+1/2})\,.$$

Thus, the hemisphere is decomposed into the set of fragments, each point $\{\gamma_\ell, \varphi_{\ell,m}\}$ being chosen as a quadrature node, with the value of a hemisphere element $w_{\ell,m} = \Delta\gamma_\ell\, \Delta\varphi_{\ell,m}$ being equal to corresponding quadrature weight. Similarly, the quadrature can be introduced for another hemisphere.

The quadrature is called a *rectangular* one if the number $M_\ell$ of subintervals over $\varphi$ is the same for each $\gamma$-layer. Concentration of nodes near the unit sphere poles represents the main shortcoming of the rectangular quadrature. There is not this shortcoming in triangular quadratures: the closer node $\gamma_\ell$ is to the sphere pole the less the corresponding parameter $M_\ell$.

Two important requirements should be satisfied for the quadrature construction. The first requirement is that quadrature nodes should be distributed in a maximally uniform manner over the unit sphere. The second one is that the quadrature should provide an exact calculation of the following integrals:

$$\sum_{\ell=1}^{2L} \gamma_\ell^{k_1}\, \Delta\gamma_\ell = \int_{-1}^{1} \mathrm{d}\gamma\, \gamma^{k_1}\,, \tag{8.7a}$$

$$\sum_{\ell=1}^{2L} \Delta\gamma_\ell \sum_{m=1}^{M_\ell} \xi_{\ell,m}^{k_2} \Delta\varphi_{\ell,m} = \int_{-1}^{1} \mathrm{d}\gamma \int_0^{\pi} \mathrm{d}\varphi \xi^{k_2}(\gamma,\varphi)\,, \tag{8.7b}$$

$$\sum_{\ell=1}^{2L} \Delta\gamma_\ell \sum_{m=1}^{M_\ell} \eta_{\ell,m}^{k_3} \Delta\varphi_{\ell,m} = \int_{-1}^{1} \mathrm{d}\gamma \int_0^{\pi} \mathrm{d}\varphi \eta^{k_3}(\gamma,\varphi)\,, \tag{8.7c}$$

where the set of values $k_1$, $k_2$, $k_3$ depends on quadrature type. Here the values $\xi$, $\eta$, $\gamma$ are projections of transport vector $\vec{\Omega}$ on coordinate axes $x$, $y$ and $z$, correspondingly, see Fig. 8.1:

$$\xi_{\ell,m} = \xi(\gamma_\ell, \varphi_{\ell,m}),\ \eta_{\ell,m} = \eta(\gamma_\ell, \varphi_{\ell,m}),\ \xi = \sin\theta\cos\varphi,\ \eta = \sin\theta\sin\varphi,\ \gamma = \cos\theta\,.$$

The scattering integral in the transfer equation can be calculated with a high accuracy, if Eqs (8.7a)–(8.7c) are satisfied.

Two examples of the triangular angular quadrature with $M_\ell = 2(L - |\ell| + 1)$ are depicted in Fig. 8.4. Here the darker shading is prescribed for the larger sphere fragment.

Although the distribution of *LQ-quadrature* nodes over the sphere is not uniform (see Fig. 8.4), the quadrature is symmetrical with respect to rotation by 90°. Besides, it satisfies the conditions (8.7) at $k = 0, 1, \ldots, 2L$. Unfortunately, these useful quadratures exist only for $L \le 10$, because equations, defining weights and nodes of the LQ-quadarature, are not resolved at $L > 10$.

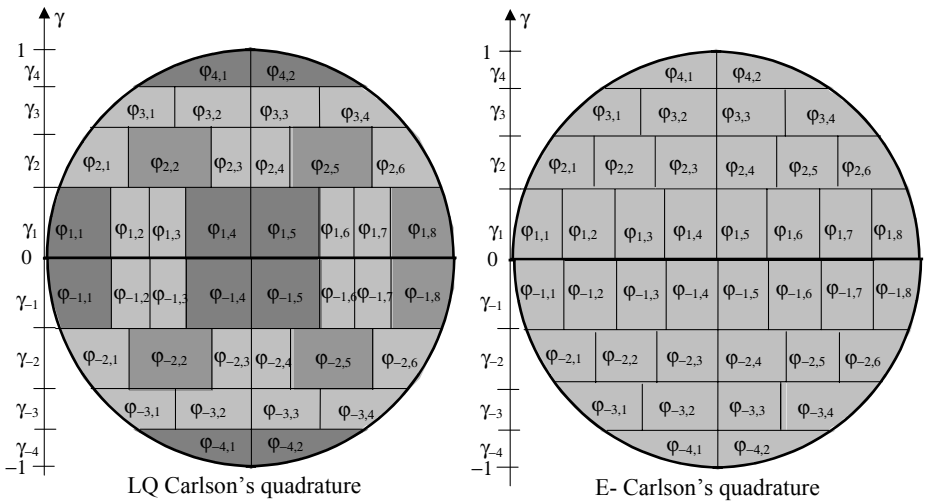



**Fig. 8.4.** Examples of angular quadratures for hemisphere $\{-1 < \gamma < 1,\ 0 < \varphi < \pi\}$.

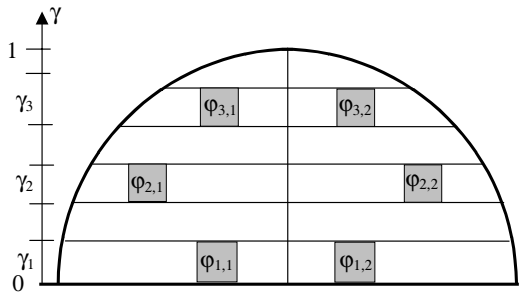



**Fig. 8.5.** Six nodes of one set for DCT quadrature.

*E-quadrature* (Carlson, 1976) exists for arbitrary $L$. All squares $w_{\ell,m}$ of the quadrature are equal to value $\pi/(L\,(L+1))$, see Fig. 8.4. Such severe constraint does not provide great freedom for quadrature node choice, because Eqs (8.7a)–8.7c) are exactly satisfied only for $k = 0, 1, 2$ in the case of E-quadrature.

*DCT (double circle triangle) quadratures* (Koch et al., 1995) are generalization of Carlson's ones. DCT grids consist of sets, each of which contains six nodes arranged on one quadrant; see an example in Fig. 8.5. Weights of all six nodes must be identical. This ensures the quadrature symmetry. But weights of different sets may be chosen differently that provide free parameters to increase the quadrature accuracy in the sense of Eq. (8.7).

Equation (8.7a) for $k = 0, 1, \ldots, 4L-1$ is satisfied in the case of *Gauss's grid* over $\gamma$. If one uses a uniform $\varphi$-grid with centered nodes for each $\gamma$-layer (i.e. at $\varphi_{\ell,m} = (\varphi_{\ell,m+1/2} + \varphi_{\ell,m-1/2})/2$), then conditions (8.7b) and (8.7c) are approximately satisfied (the accuracy being sufficiently high for a large number of quadrature nodes). The conditions (8.7b) and (8.7c) are exactly fulfilled for the Gaussian $\varphi$-grid. However, in this case the node distribution is inhomogeneous. It should be noted that another interpolation formulas can be used instead of those based on Gaussian grid. For example, one can use either *Radau quadrature* (where $\gamma_{-L} = -1$, and the condition (8.7a) is fulfilled for $k = 0, 1, \ldots, 4L-2$) or *Lobatto quadrature* (where $\gamma_{-L} = -1$, $\gamma_L = 1$, and the condition (8.7a) is satisfied for $k = 0, 1, \ldots, 4L-3$).

The quadratures with weights and nodes that permit exact integration of polynomials up to the highest possible order and are invariant to the desired rotation group have features of both Gauss's and Carlson's grids (Lebedev, 1976). The weights and nodes of these quadratures are defined by non-linear algebraic systems.

The significant shortcoming of all mentioned quadratures is that the grid condensing cannot be realized uniformly as $L$ increases. That is to say, each fragment of a coarser grid is not formed by proper combination of fragments of a finer grid. A *Special T-quadrature* (Aussourd, 2003) is constructed, where this problem is solved: each portion of a coarser grid is divided into four identical portions of a finer grid (an example is shown in Fig. 8.6).

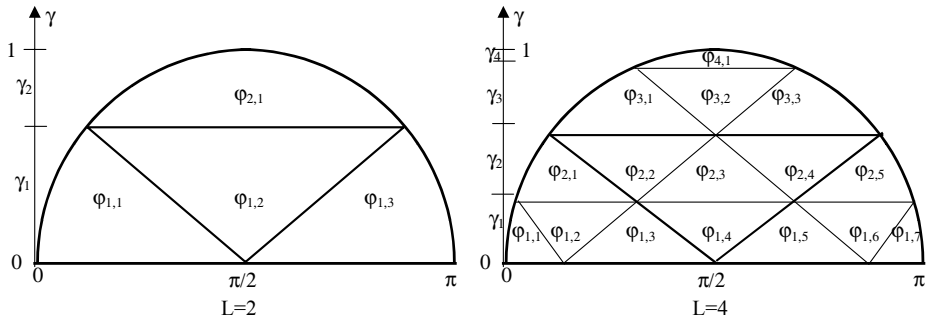



**Fig. 8.6.** T-quadrature for $\{0 < \gamma < 1,\ 0 < \varphi < \pi\}$.

The mentioned feature of the T-quadrature provides the capability to condense (refine) the angular grid in that solid angle, where the transport equation solution significantly depends on angular variables. Additional nodes are then simply added to the initial quadrature grid, whereas using other quadratures one has to correct the neighboring nodes in the process of grid condensing to keep Eqs (8.7) fulfillment for a finer grid (Longoni and Haghighat, 2001).

For problems with the forward peaked phase functions, common for atmospheric optics problems, special kinds of quadratures were constructed, for which quadrature nodes are concentrated near sphere poles (Sanchez and McCormic, 2004).

## 8.4 Scattering integral representation

After the introduction of angular quadratures in DOM, the solution $\psi(\vec{r}, \gamma, \varphi)$ is replaced by a collection of functions $\psi(\vec{r}, \gamma_\ell, \varphi_{\ell,m})$. Two types of the scattering integral $\hat{S}\psi$ representations exist. The first type is widely used in neutron physics and atmospheric optics problems. Here the phase function $p(\vec{r}, \vec{\Omega}\cdot\vec{\Omega}')$ is presented via expansion into a finite sum on the Legendre polynomials,

$$p(\vec{r}, \vec{\Omega}\cdot\vec{\Omega}') = \frac{1}{4\pi}\sum_{\nu=1}^{N(\vec{r})}(2\nu+1)\,\omega_\nu(\vec{r})\,P_\nu(\vec{\Omega}\cdot\vec{\Omega}')\,, \tag{8.8}$$

that are orthogonal in the interval $(-1,\,1)$:

$$\int_{-1}^{1} P_n(\chi)\,P_\nu(\chi)\,\mathrm{d}\chi = 2\,\delta_{n,\nu}/(2n+1)\,,$$

where $\delta_{n,\nu}$ is the Kronecker delta symbol. The expansion coefficients are defined by formulas

$$\omega_\nu(\vec{r}) = 2\,\pi\int_{-1}^{1} p(\vec{r}, \chi)\,P_\nu(\chi)\,\mathrm{d}\chi\,, \quad \omega_0(\vec{r}) \equiv 1\,. \tag{8.9}$$

The expression (8.5) and Legendre polynomials adding theorem (Ryzik and Gradstein, 1972)

$$P_\nu(\vec{\Omega}\,\vec{\Omega}') = P_\nu^0(\gamma)\,P_\nu^0(\gamma') + 2\sum_{\mu=1}^{\nu}\frac{(\nu-\mu)!}{(\nu+\mu)!}P_\nu^\mu(\gamma)\,P_\nu^\mu(\gamma')\,\cos\mu\,(\varphi-\varphi')\,,$$

where

$$P_\nu^\mu(\gamma) = (-1)^\mu\,(1-\gamma^2)^{\mu/2}\,\frac{\mathrm{d}^\mu}{\mathrm{d}\gamma^\mu}\,P_\nu(\gamma)$$

are the associated Legendre functions, allow us to represent the scattering integral $\hat{S}\psi$ by the following sum

$$\begin{aligned}\hat{S}\psi \;=\; & \frac{1}{4\pi}K_{\mathrm{ext}}(\vec{r})\,\bar{\omega}_0(\vec{r})\sum_{\nu=0}^{N(\vec{r})}(2\nu+1)\,\omega_\nu(\vec{r})\left\{P_\nu^0(\gamma)\,M_{\nu,0}^c(\vec{r})\right.\\ & \left.+\,2\sum_{\mu=1}^{\nu}\frac{(\nu-\mu)!}{(\nu+\mu)!}P_\nu^\mu(\gamma)\cdot\left[\cos\mu\varphi\,M_{\nu,\mu}^c(\vec{r})+\sin\mu\varphi\,M_{\nu,\mu}^s(\vec{r})\right]\right\}, \quad (8.10)\end{aligned}$$

where angular moments $M_{\nu,\mu}^c(\vec{r})$ and $M_{\nu,\mu}^s(\vec{r})$ of the solution are determined as

$$M_{\nu,\mu}^c(\vec{r}) = \int_{-1}^{1}\mathrm{d}\gamma\int_0^{2\pi}P_\nu^\mu(\gamma)\,\cos\mu\varphi\,\psi(\vec{r},\gamma,\varphi)\,\mathrm{d}\varphi\,,$$

$$M_{\nu,\mu}^s(\vec{r}) = \int_{-1}^{1}\mathrm{d}\gamma\int_0^{2\pi}P_\nu^\mu(\gamma)\,\sin\mu\varphi\,\psi(\vec{r},\gamma,\varphi)\,\mathrm{d}\varphi\,.$$

After replacing all integrals by corresponding quadrature sums

$$\begin{aligned}M_{\nu,\mu}^c(\vec{r}) \;&\simeq\; \sum_\ell\sum_m P_\nu^\mu(\gamma_\ell)\,\cos\mu\varphi_{\ell,m}\,\psi(\vec{r},\gamma_\ell,\varphi_{\ell,m})\,,\\ M_{\nu,\mu}^s(\vec{r}) \;&\simeq\; \sum_\ell\sum_m P_\nu^\mu(\gamma_\ell)\,\sin\mu\varphi_{\ell,m}\,\psi(\vec{r},\gamma_\ell,\varphi_{\ell,m})\,, \quad (8.11)\end{aligned}$$

we obtain the discrete representation of the scattering integral.

Utilizing representation (8.10), one can store a set of angular moments (8.11) and values $P_\nu^\mu(\gamma_\ell)\,\cos\mu\varphi_{\ell,m}$ and $P_\nu^\mu(\gamma_\ell)\,\sin\mu\varphi_{\ell,m}$ in computer memory, rather than solution values at grid nodes.

The second type of the integral $\hat{S}\psi$ representation consists in the direct replacement of $\hat{S}\psi$ by quadrature sums:

$$\hat{S}\psi(\vec{r},\gamma_\ell,\varphi_{\ell,m}) \simeq K_{\mathrm{ext}}\bar{\omega}_0(\vec{r})\sum_{\ell'}\sum_{m'}H_{\ell,\ell',m,m'}(\vec{r})\,\psi(\vec{r},\gamma_{\ell'},\varphi_{\ell',m'})\,w_{\ell',m'}\,,$$

$$H_{\ell,\ell',m,m'}(\vec{r}) \approx p\left(\vec{r},\gamma_\ell\,\gamma_{\ell'}+\sqrt{1-\gamma_\ell^2}\,\sqrt{1-(\gamma_{\ell'})^2}\cos(\varphi_{\ell,m}-\varphi_{\ell',m'})\right),$$

and normalization conditions (see (8.4)) must be kept

$$\sum_{\ell'} \sum_{m'} H_{\ell,\ell',m,m'}(\vec{r})\, w_{\ell',m'} = 1\,.$$

Now the values $\psi(\vec{r}, \gamma_\ell, \varphi_{\ell,m})$ themselves and the elements of the scattering matrix $H$ should be stored in computer memory during calculations. It is noteworthy that the number $\tilde{N}(\vec{r})$ of angular moments (8.11), which form the solution accordingly to formula (8.10), is equal to $[N(\vec{r})]^2$, where $N(\vec{r})$ is the number of moments used for the phase function representation (see Eq. (8.8)). If the solution is an even function of the azimuth angle $\varphi$, the moments $M^s_{\nu,\mu}(\vec{r})$ vanish, and so only $\tilde{N}(\vec{r}) = N(\vec{r})\,[N(\vec{r}) + 1]/2$ moments are used in calculations.

Thus, while using the first method, one should store $\tilde{N}(\vec{r})\,(1 + 2M)$ values (angular moments and auxiliary values $P^\mu_\nu(\gamma_\ell)\,\cos\mu\varphi_{\ell,m}$, $P^\mu_\nu(\gamma_\ell)\,\sin\mu\varphi_{\ell,m}$) for each spatial mesh $\vec{r}$, whereas in the case of using the second method, exactly $M + M^2$ values (solution results and scattering matrix elements) should be stored (here $M$ is full number of quadrature nodes). So, as one can see, the first method requires less computer memory than the second one, if $\tilde{N}(\vec{r}) < M\,(M + 1)/(2M + 1)$. The last inequality is valid in the case of reactor-shielding problems, where $N(\vec{r}) \le 5$, $M \le 80$. The phase functions are of more complex type in problems of radiation transfer through the terrestrial atmosphere, and therefore one should use the sums of a great number of Legendre polynomials to represent them. Hence the second method can be more economic because it uses a smaller number of quadrature nodes. Note also, that matrix elements $H_{\ell,\ell',m,m'}(\vec{r})$ may be calculated during the main computation procedure, rather than stored in computer memory. It will result in a decrease in the necessary memory volume, but an increase in the calculation time.

It is just the first method that is applied in the majority of codes using the scattering integral calculation. The associated Legendre functions $P^\mu_\nu(\gamma)$ are usually calculated via the recurrent formulas (Ryzik and Gradstein, 1972):

$$P^\mu_{\nu+1}(\gamma) = \frac{2\nu + 1}{\nu - \mu + 1}\,\gamma\,P^\mu_\nu(\gamma) - \frac{\nu + \mu}{\nu - \mu + 1}\,P^\mu_{\nu-1}(\gamma)\,, \tag{8.12}$$

where

$$\begin{aligned} P^\nu_\nu(\gamma) &= \left(1 - \gamma^2\right)^{\nu/2}\,(-1)^\nu\,\frac{(2\nu)!}{2^\nu \nu!}\,, \\ P^{\nu-1}_\nu(\gamma) &= \gamma\,\left(1 - \gamma^2\right)^{(\nu-1)/2}\,(-1)^{\nu-1}\,\frac{(2\nu - 1)!}{2^{\nu-1}\,(\nu - 1)!}\,. \end{aligned}$$

In atmospheric optics transport problems, one has often to deal with highly forward-peaked phase functions, defined via values at the nodes of a some grid over the scattering angle $\chi$. The calculation of coefficients $\omega_\nu(\vec{r})$ in the expansion (8.8) represents a complicated problem for such phase functions at $\nu > 60$, because of two facts:

(a) the Legendre polynomials of high order, contained in the integrals (8.9), are quickly oscillating functions, especially in the vicinity of points $\chi = \pm 1$,

(b) the Legendre polynomial values are incorrectly calculated near $\chi = \pm 1$ (see Eq. (8.12)). A widely used code, presented at the site http://rts.kiam.ru/verval/sfxp.htm, is based on a special approach for effective $\omega_\nu(\vec{r})$ calculations with a high accuracy.

An adequate representation of highly forward-peaked phase functions requires calculations with a large number of polynomials (up to some hundreds and even thousands). For this reason, it is helpful to decompose a highly-peaked phase function into the sum of a singular component (corresponding to the phase function peak) and a regular component. The expansion (8.8) can be applied to the regular component only, the singular component being approximated by the delta-function (Wiscombe, 1977; Landesman and Morel, 1989).

## 8.5 The general solution

As was mentioned in the introduction, function $F(\vec{r}, \vec{\Omega})$ in Eq. (8.1) describes radiation sources, usually representing either solar or thermal radiation. In the first case function $F(\vec{r}, \vec{\Omega})$ is singular with respect to angular variables:

$$F(\vec{r}, \vec{\Omega}) = F_0 \, \delta(\vec{\Omega} - \vec{\Omega}_0) \, \delta(f_0(\vec{r})) \,,$$

where $\delta$ is the Dirac delta-function, the vector $\vec{\Omega}_0$ defines the solar radiation direction, $F_0$ is the solar radiation flux and the function $f_0(\vec{r})$ determines the surface through which radiation penetrates into the region under consideration.

The solution can be presented as the sum of two functions

$$\psi(\vec{r}, \vec{\Omega}) = \psi_n(\vec{r}, \vec{\Omega}) + \psi_s(\vec{r}, \vec{\Omega}) \,,$$

function-terms being the solutions to the following problems

$$\hat{L}\psi_n(\vec{r}, \vec{\Omega}) = F(\vec{r}, \vec{\Omega}) \,, \quad \psi_n(\vec{r}, \vec{\Omega}) = 0 \,, \quad \text{as } \vec{r} \in \Gamma, \ \vec{\Omega} \cdot \vec{n}(\vec{r}) < 0 \tag{8.13}$$

and

$$\hat{L}\psi_s(\vec{r}, \vec{\Omega}) = \hat{S}\psi_s(\vec{r}, \vec{\Omega}) + \hat{S}\psi_n(\vec{r}, \vec{\Omega}) \,,$$

$$\psi_s(\vec{r}, \vec{\Omega}) = A(\vec{r}) \, \widehat{R} \, \psi_s(\vec{r}, \vec{\Omega}) + A(\vec{r}) \, \widehat{R} \, \psi_n(\vec{r}, \vec{\Omega}) + \varphi_0(\vec{r}, \vec{\Omega}) \quad \text{as } \vec{r} \in \Gamma, \vec{\Omega} \cdot \vec{n}(\vec{r}) < 0 \,.$$

Here the function $\psi_n(\vec{r}, \vec{\Omega})$ corresponds to the unscattered component of radiation, whereas the function $\psi_s(\vec{r}, \vec{\Omega})$ represents the scattered component. The problem (8.13) for the unscattered component can be easily solved analytically, whereas the scattered component $\psi_s(\vec{r}, \vec{\Omega})$ has to be obtained numerically with the help a grid scheme.

## 8.6 Approximation of differential operator $\hat{L}$

When grid approximations of operator $\hat{L}$ are performed, the function $B(\vec{r}, \vec{\Omega})$ in Eq. (8.1) and the boundary source $\overline{\psi}(\vec{r}, \vec{\Omega})$ are assumed to be known (further indices $\ell$, $m$ in all formulas are omitted). At the first step, the spatial grid itself is specified. For example, in the case of $(x, y, z)$ geometry we have

$$\begin{aligned} &x_{1/2} < \ldots < x_{i+1/2} < \ldots < x_{I-1/2} < x_{I+1/2}\,, \\ &y_{1/2} < \ldots < y_{j+1/2} < \ldots < y_{J-1/2} < y_{J+1/2}\,, \\ &z_{1/2} < \ldots < z_{k+1/2} < \ldots < z_{K-1/2} < z_{K+1/2}\,. \end{aligned} \tag{8.14}$$

Here $x_{i\pm1/2}, y_{j\pm1/2}, z_{k\pm1/2}$ are grid mesh bounds,

$$\Delta x_i = x_{i+1/2} - x_{i-1/2}\,, \quad \Delta y_i = y_{j+1/2} - y_{j-1/2}\,, \quad \Delta z_k = z_{k+1/2} - z_{k-1/2}$$

are mesh sizes,

$$x_i = (x_{i+1/2} + x_{i-1/2})/2\,, \;\; y_j = (y_{j+1/2} + y_{j-1/2})/2\,, \;\; z_k = (z_{k+1/2} + z_{k-1/2})/2$$

are mesh centers, and integer $I$, $J$, $K$ are numbers of meshes.

Each grid mesh is a parallelepiped. Algorithms, based on the grids with meshes of tetrahedral or even arbitrary polyhedral forms, have been developed, aimed at accuracy increasing in the representation of discontinuity surfaces for the coefficient of the transport equation (Morel and Larsen, 1990; Castriani and Adams, 1995; Grove and Pevey, 1995). Further, we do not consider the problems related to a design of such complicated grids and corresponding calculations. We consider only regular grids of a type given by Eq. (8.14).

*Spatial moments of the solution* rather than solution values themselves appear as calculation values at grid nodes. (Remember that *spatial moments* are integrals of the solution over a mesh and its bounds, integration being fulfilled with different weight functions.) For example, for grid (8.14) the *zero spatial moments*, arising in the case when the weight function is equal to 1, are defined by the formulas

$$\begin{aligned} \psi_{i,j,k} &= \frac{1}{\Delta x_i \Delta y_j \Delta z_k} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathrm{d}x \int_{y_{j-1/2}}^{y_{j+1/2}} \mathrm{d}y \int_{z_{k-1/2}}^{z_{k+1/2}} \mathrm{d}z \psi(x, y, z)\,, \\ \psi_{i\pm1/2,j,k} &= \frac{1}{\Delta y_j\, \Delta z_k} \int_{y_{j-1/2}}^{y_{j+1/2}} \mathrm{d}y \int_{z_{k-1/2}}^{z_{k+1/2}} \mathrm{d}z \psi(x_{i\pm1/2}, y, z)\,, \\ \psi_{i,j\pm1/2,k} &= \frac{1}{\Delta x_i\, \Delta z_k} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathrm{d}x \int_{z_{k-1/2}}^{z_{k+1/2}} \mathrm{d}z\, \psi(x, y_{j\pm1/2}, z)\,, \\ \psi_{i,j,k\pm1/2} &= \frac{1}{\Delta x_i\, \Delta y_j} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathrm{d}x \int_{y_{j-1/2}}^{y_{j+1/2}} \mathrm{d}y\, \psi(x, y, z_{k\pm1/2})\,. \end{aligned} \tag{8.15}$$

These values may be interpreted as average values of the solution over a mesh, including its bounds. The averages of the solution are expected to be more stable characteristics of the solution, than solution values themselves at grid nodes.

For moments of a higher order, the polynomials of a corresponding order appear as weight functions. For example, the moments of the first order can be defined by relations

$$\begin{aligned}
\psi^{1,x}_{i,j,k} &= \frac{3}{\Delta x_i \Delta y_j \Delta z_k} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathrm{d}x \bar{P}_1(x) \int_{y_{j-1/2}}^{y_{j+1/2}} \mathrm{d}y \int_{z_{k-1/2}}^{z_{k+1/2}} \mathrm{d}z \psi(x,y,z)\,, \\
\psi^{1,y}_{i,j,k} &= \frac{3}{\Delta x_i \Delta y_j \Delta z_k} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathrm{d}x \int_{y_{j-1/2}}^{y_{j+1/2}} \mathrm{d}y \bar{P}_1(y) \int_{z_{k-1/2}}^{z_{k+1/2}} \mathrm{d}z \psi(x,y,z)\,, \\
\psi^{1,z}_{i,j,k} &= \frac{3}{\Delta x_i \Delta y_j \Delta z_k} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathrm{d}x \int_{y_{j-1/2}}^{y_{j+1/2}} \mathrm{d}y \int_{z_{k-1/2}}^{z_{k+1/2}} \mathrm{d}z \bar{P}_1(z) \psi(x,y,z)\,, \qquad (8.16)
\end{aligned}$$

where

$$\bar{P}_1(x) = \frac{x - x_i}{\Delta x_i/2}\,, \quad \bar{P}_1(y) = \frac{y - y_j}{\Delta y_j/2}\,, \quad \text{and} \;\; \bar{P}_1(z) = \frac{z - z_k}{\Delta z_k/2}$$

are the first Legendre polynomial to meshes $[x_{i-1/2},\ x_{i+1/2}]$, $[y_{j-1/2},\ y_{j+1/2}]$ and $[z_{k-1/2},\ z_{k+1/2}]$, correspondingly. They enable us to expand the solution in the spatial mesh $[x_{i-1/2},\ x_{i+1/2}] \times [y_{j-1/2},\ y_{j+1/2}] \times [z_{k-1/2},\ z_{k+1/2}]$ in following manner

$$\begin{aligned}
&\psi(x,y,z) = \psi_{i,j,k} + \bar{P}_1(x)\,\psi^{1,x}_{i,j,k} + \bar{P}_1(y)\,\psi^{1,y}_{i,j,k} + \bar{P}_1(z)\,\psi^{1,z}_{i,j,k} + O(\Delta^2)\,, \\
&\Delta = \max\{\Delta x_i,\ \Delta y_j,\ \Delta z_k\}\,.
\end{aligned}$$

This presentation leads to the nodal grid scheme that is more accurate than schemes based on moments (8.15) and the assumption

$$\psi(x,y,z_k) = \psi_{i,j,k} + O(\Delta)\,.$$

### 8.6.1 Properties of DOM grid schemes

The *grid scheme accuracy* is of importance in scheme characteristics analysis. For some schemes considered below the following estimations can be obtained:

$$\left\| \vec{\psi}^{\,\mathrm{exact}} - \vec{\psi} \right\| \le C h^q \quad \text{as} \quad h \to 0,\ q = 1, 2, \ldots\,, \qquad (8.17)$$

where the vector $\vec{\psi}$ is formed by grid values of the solution moments, the vector $\vec{\psi}^{\,\mathrm{exact}}$ consists of exact solution moments at corresponding meshes, $h$ is the largest mesh 'diameter' in terms of the optical length, $C$ is a constant, which is independent on $h$. Usually the constant $C$ is proportional to the greatest absolute value of some derivative from the exact solution $\psi(\vec{r}, \vec{\Omega})$. The norm, figuring in Eq. (8.17), is either the uniform one (being defined by the maximum of the absolute value of the difference $(\vec{\psi}^{\,\mathrm{exact}} - \vec{\psi})$) or the mean square one. The estimation (8.17) means, that the grid solution converges to the exact solution under spatial grid refining. Under condition (8.17) the scheme is said to have $q$th order of accuracy on the set of smooth solutions.

Surely, the estimations (8.17) are not valid in those sub-regions, where the exact solution has no necessary derivatives. In particular, the exact solution is

not smooth near the surfaces, where coefficients of transport equation are discontinuous (Germogenova, 1985). Besides, spatial grid dimensions are naturally bounded by current computer capabilities, and so calculations cannot be always performed on the basis of grids that are as fine as may be desired. Therefore, it is impossible to be restricted to the accuracy estimations of type (8.17). Qualitative properties of the grid solutions on arbitrary grids are of a significant value. In particular, a quite important moment is that a grid solution should satisfy some *balance relation*, that is a result of the transport equation integration over a spatial mesh and over all directions. It guarantees the conservation of a photon number in the computation process. Significant errors in the grid solution are possible in the case when the balance relation is not satisfied.

Besides, as the computational practice demonstrates, reliable results can be obtained only with the help of schemes, that guarantee the preservation of main qualitative features of exact solutions. The non-negativity of the solution under non-negative sources and the monotonicity of the solution along any characteristics under monotonic effective source $B(\vec{r},\vec{\Omega})/K_{\rm ext}(\vec{r})$ belong to strictly established characteristic features of exact solutions (Bass et al., 1986). The computational schemes, ensuring these essential features of grid solutions, are called *positive* and *monotonic schemes* correspondingly. Calculations with non-positive schemes can result in non-positive grid solutions that should be considered as non-physical (because the function $\psi(\vec{r},\vec{\Omega})$ defines the positive radiance). In addition, grid solutions obtained by means of non-monotonic schemes can have significant non-physical oscillations.

Finally, an additional feature of grid schemes for the transport equation should be noted: the higher the scheme order, the lower the degree of positiveness and monotonicity of the grid solution.

### 8.6.2 Classification of grid schemes

At the step of the grid approximation construction, the left-hand side of the transport equation (see Eq. (8.1)) can be written in two forms. In the first form it is suggested, that the term $\vec{\Omega}\cdot\vec{\nabla}\psi$ in Eq. (8.2) is the derivative in the direction $\vec{\Omega}$:

$$\vec{\Omega}\cdot\vec{\nabla}\psi = \frac{\partial\psi}{\partial\vec{\Omega}}\,.$$

Then the solution $\psi(\vec{r},\vec{\Omega})$ is defined by the formula

$$\begin{aligned}\psi(\vec{r},\vec{\Omega}) \;=\; & \overline{\psi}(\vec{r}^{\,*},\vec{\Omega})\,\exp\left(-\int_0^{|\vec{r}-\vec{r}^{\,*}|} K_{\rm ext}(\vec{r}^{\,*}+\xi\,\vec{\Omega})\,{\rm d}\xi\right)\\ & +\int_0^{|\vec{r}-\vec{r}^{\,*}|} B(\vec{r}^{\,*}+\xi\,\vec{\Omega},\vec{\Omega})\exp\left(-\int_\xi^{|\vec{r}-\vec{r}^{\,*}|} K_{\rm ext}(\vec{r}^{\,*}+\xi'\,\vec{\Omega})\,{\rm d}\xi'\right)\,{\rm d}\xi\,, \quad (8.18)\end{aligned}$$

If one considers the term $\vec{\Omega}\cdot\vec{\nabla}\psi$ as an inner product of two vectors in $(x, y, z)$ geometry, the following expression for $\hat{L}$ can be written:

$$\hat{L}\psi = \gamma\frac{\partial\psi}{\partial z} + \eta\frac{\partial\psi}{\partial y} + \xi\frac{\partial\psi}{\partial x} + K_{\text{ext}}\,\psi\,. \tag{8.19}$$

The method of grid scheme construction based on Eq. (8.18) is known as the characteristics method (see Fig. 8.7) because integration in (8.18) is carried out either along the whole characteristics of the operator $\hat{L}$, defined by Eq. (8.1) (segment AB in Fig. 8.8) or along the characteristics segment, located inside a single mesh (segment A′B′ in Fig. 8.8). In the first case we obtain *long characteristics schemes* (section 8.7). In the second case we come to *short characteristics schemes* (section 8.8). Construction of *integro-interpolational schemes*

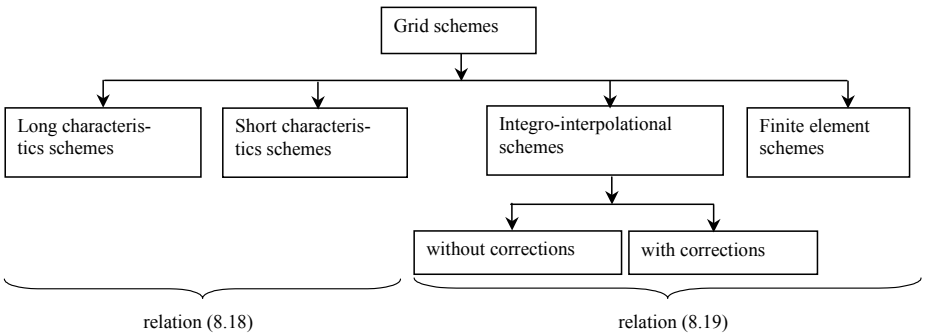



**Fig. 8.7.** Classification of schemes.

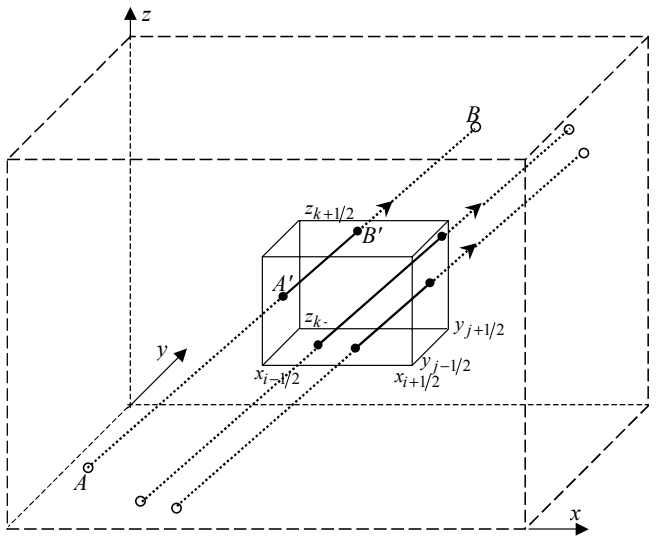



**Fig. 8.8.** Transfer of long characteristics through the whole calculation region (line AB) and one grid mesh (segment A′B′).

(section 8.9) (Samarsky, 1989) and *finite element schemes* (section 8.10) are based on Eq. (8.19). The grid scheme classification used in this chapter (that is close to widely accepted one), is depicted in Fig. 8.7.

Integro-interpolational schemes may be divided into schemes *with corrections* (section 8.9.2) and schemes without corrections (section 8.9.1). In the case of schemes with corrections, each mesh is calculated first according to a non-positive and non-monotonous scheme of a high order of accuracy. If the grid solution is negative or possesses non-physical oscillations, the mesh is recalculated via a lower-order accuracy scheme, which guarantees the solution positivity and, possibly, oscillation smoothing.

If scheme equations contain spatial moments of higher order (besides zero moments), the scheme is called *a nodal one* (section 8.9.3). Nodal schemes were developed for each class, presented in Fig. 8.7. It is suggested that such schemes may include grids with large meshes. Utilizing nodal schemes leads to an increase in the calculation scope, the calculation accuracy and the complexity being simultaneously increased. On the other hand nodal schemes provide accurate calculations with coarser grids compared to lower-order accuracy schemes. This feature leads to reduction of calculation time.

It is worth noting, that the characteristics schemes are often used in atmosphere optics problems (see section 8.7). Recently the code RADUGA-5.1(P) (Nikolaeva et al., 2005a,b), based on integro-interpolational schemes (section 8.9), has been successfully applied to the solution of the abovementioned problems (section 8.13.2).

## 8.7 Long characteristics schemes

Long characteristics schemes were initially developed for the case of one-dimensional spherical geometry (Vladimirov, 1958). They have since been extended to problems with Cartesian $(x, y, z)$ geometry (Suslov, 1988; Postma and Viujic, 1999; Evans, 1998). Very informative monographs (Sushkevich et al., 1990; Sushkevich, 2005) contain, in particular, algorithms of long and short characteristics in Cartesian and spherical geometries.

The calculation starts with the introduction of a spatial grid and definition of a set of characteristics for Eq. (8.1) (see segment AB in Fig. 8.8, where all characteristics are depicted by arrows). Several characteristics pass in each direction $\vec{\Omega}$ can intersect a spatial mesh (see segment A′B′ in Fig. 8.8), where intersection points of characteristics and mesh edges are denoted by spots.

Using spatial moments of the function $B(\vec{r}, \vec{\Omega})$ in a mesh, we calculate the values of this function at each point of each characteristic inside the mesh. The values of the solution $\psi(\vec{r}, \vec{\Omega})$ at points of characteristics can be found using Eq. (8.18). To decrease the calculation time, the exponents in Eq. (8.18) can be replaced by fractional-polynomial approximations (so called Padé approximations) (Marchuk and Lebedev, 1981), which have the form $\exp(-h) \simeq (2-h)/(2+h)$ or the form $\exp(-h) \simeq 1/(1+h)$.

The zero spatial moment (8.15) of the solution in each mesh can be found as the average of the solution over all characteristics in the same direction $\vec{\Omega}$, intersecting the mesh (see Fig. 8.8). Obviously, the obtained grid solution generally does not satisfy the balance relation. For this the special normalizing coefficients must be introduced.

The long characteristics schemes are in some sense equivalent to the MC methods. Just like the MC schemes, they always generate positive solutions under positive sources. Besides, the schemes are capable of taking into account the whole region with inhomogeneities accurately. On the other hand, the characteristics method demands calculations for a large number of directions $\vec{\Omega}$ for those problems, where the MC method has to use many histories, and so the long characteristics method is a very time-consuming one.

## 8.8 Short characteristics schemes

In short characteristics schemes, the edges of each mesh are divided into 'entering' (through them radiation enters the mesh) and 'outgoing' (through them radiation leaves the mesh). For example in $(x, y, z)$ geometry, if values $\gamma$ and $\varphi$ which define the transport direction $\vec{\Omega}$, belong to the octant $\{0 < \gamma < 1, 0 < \varphi < \pi/2\}$, then radiation enters the calculation region via the boundaries $x = x_{1/2}$, $y = y_{1/2}$ and $z = z_{1/2}$ (see Fig. 8.9). Therefore, 'entering' edges are $x = x_{i-1/2}$, $y = y_{j-1/2}$, $z = z_{k-1/2}$, and 'outgoing' ones are $x = x_{i+1/2}$, $y = y_{j+1/2}$, $z = z_{k+1/2}$ (see Fig. 8.9).

The grid scheme solution at 'entering' edges is approximated by some continuous functions. Further under these assumptions the transport equation solution is defined via Eq. (8.18) both inside the mesh and at its 'outgoing' edges. At last, the solution moments are calculated both inside the mesh and at the 'outgoing' edges.

According to the *SC (Step Characteristics) scheme* (Lathrop, 1969) the solution at each 'entering' edge is represented by a constant, which is equal to the corresponding value of the zero spatial moment (see Eq. (8.15)). Equations for the grid solution calculation can be written in the following form:

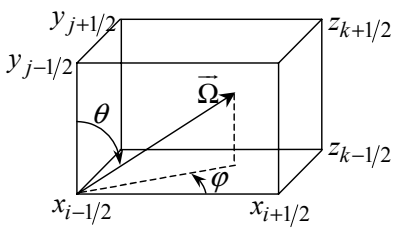



**Fig. 8.9.** Mesh in $(x, y, z)$ geometry.

$$\begin{aligned}
\psi_{i+1/2,j,k} &= a_{x,x}\,\psi_{i-1/2,j,k} + a_{y,x}\,\psi_{i,j-1/2,k} + a_{z,x}\,\psi_{i,j,k-1/2} + a_{B,x}\,B_{i,j,k}\,,\\
\psi_{i,j+1/2,k} &= a_{x,y}\,\psi_{i-1/2,j,k} + a_{y,y}\,\psi_{i,j-1/2,k} + a_{z,y}\,\psi_{i,j,k-1/2} + a_{B,y}\,B_{i,j,k}\,,\\
\psi_{i,j,k+1/2} &= a_{x,z}\,\psi_{i-1/2,j,k} + a_{y,z}\,\psi_{i,j-1/2,k} + a_{z,z}\,\psi_{i,j,k-1/2} + a_{B,z}\,B_{i,j,k}\,,\\
\psi_{i,j,k} &= a_{x,0}\,\psi_{i-1/2,j,k} + a_{y,0}\,\psi_{i,j-1/2,k} + a_{z,0}\,\psi_{i,j,k-1/2} + a_{B,0}\,B_{i,j,k}\,.
\end{aligned} \tag{8.20}$$

where $B_{i,j,k}$ is the zero spatial moment of $B(x,y,z)$ in the mesh:

$$B_{i,j,k} = \frac{1}{\Delta x_i \Delta y_j \Delta z_k} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathrm{d}x \int_{y_{j-1/2}}^{y_{j+1/2}} \mathrm{d}y \int_{z_{k-1/2}}^{z_{k+1/2}} \mathrm{d}z B(x,y,z)\,, \tag{8.21}$$

and coefficients $a_{t,h}$, $t = x,y,z,B$, $h = x,y,z,0$ depend on the transport direction $\vec{\Omega}\,\{\gamma,\varphi\}$, the mesh size $\Delta x_i$, $\Delta y_j$, $\Delta z_k$, and the extinction coefficient

$$K_{i,j,k} = K_{\mathrm{ext}}(x_i, y_j, z_k)\,. \tag{8.22}$$

Here it is assumed that all coefficients of the transport equation in each mesh are constant.

Equation (8.20) allows us to find the zero moment $\psi_{i,j,k}$ at the mesh and the moments $\psi_{i+1/2,j,k}$, $\psi_{i,j+1/2,k}$, $\psi_{i,j,k+1/2}$ at the 'outgoing' edges, if the zero moment $B_{i,j,k}$ and $\psi_{i-1/2,j,k}$, $\psi_{i,j-1/2,k}$, $\psi_{i,j,k-1/2}$ at the 'entering' edges are known.

Hence, for each fixed transport direction $\vec{\Omega}\ \{\gamma,\varphi\}$, the grid solution can be defined recursively. Let the values $\gamma$ and $\varphi$ belong to the octant $\{0 < \gamma < 1, 0 < \varphi < \pi/2\}$. Then the calculation begins from the mesh $[x_{1/2},\ x_{3/2}] \times [y_{1/2},\ y_{3/2}] \times [z_{1/2},\ z_{3/2}]$. The radiation intensity values at 'entering' edges, i.e. values $\psi_{1/2,1,1}$, $\psi_{1,1/2,1}$, $\psi_{1,1,1/2}$, are known from the boundary conditions (8.6). Eq. (8.20) permits us to find the values of the solution at 'outgoing' edges, i.e. the values $\psi_{3/2,1,1}$, $\psi_{1,3/2,1}$, $\psi_{1,1,3/2}$. They can be considered as moments at 'entering' edges for meshes $[x_{3/2},\ x_{5/2}]\times[y_{1/2},\ y_{3/2}]\times[z_{1/2},\ z_{3/2}]$, $[x_{1/2},\ x_{3/2}]\times[y_{3/2},\ y_{5/2}]\times[z_{1/2},\ z_{3/2}]$ and $[x_{1/2},\ x_{3/2}]\times[y_{1/2},\ y_{3/2}]\times[z_{3/2},\ z_{5/2}]$, correspondingly. So we can calculate all the grid solution values via sorting out successively all the meshes according to the increase of indices.

For each fixed transport direction $\vec{\Omega}$ different sequences of mesh calculations may be used. Two possible ways for mesh sorting out in a single $z$-layer are shown in Fig. 8.10. Note that the calculation sequence for $\vec{\Omega}$ directions is indifferent in the case of Cartesian spatial geometry.

We should also note that, according to formula (8.20), the solution moment at an 'outgoing' edge is the linear combination of right-side moments and solution moments at 'entering' edges. For example, the coefficient $a_{x,x}$ defines, what fraction of radiation, entering the mesh via the edge $x = x_{i-1/2}$, reaches the edge $x = x_{i+1/2}$.

Eq. (8.20) is illustrated in Fig. 8.11. All characteristics are depicted by arrows, and the shaded square $S_{x,x}$ corresponds to the share of edge $x = x_{i+1/2}$, illuminated by radiation, entering the mesh via the edge $x = x_{i-1/2}$. Square $S_{x,z}$ is a similar share of edge $z = z_{k+1/2}$.

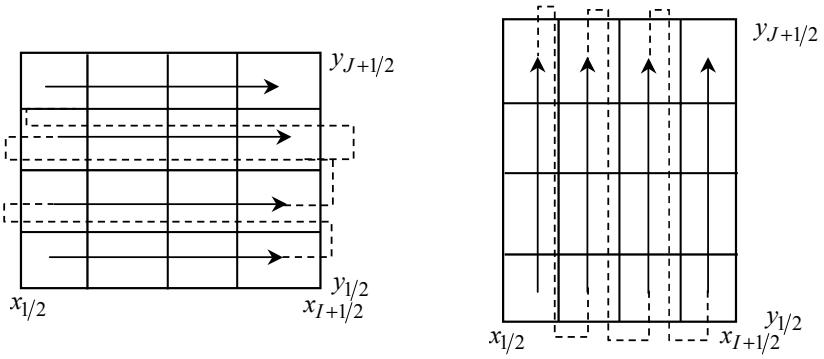



**Fig. 8.10.** Sequences of mesh calculation for the layer $(z_{k-1/2},\ z_{k+1/2})$.

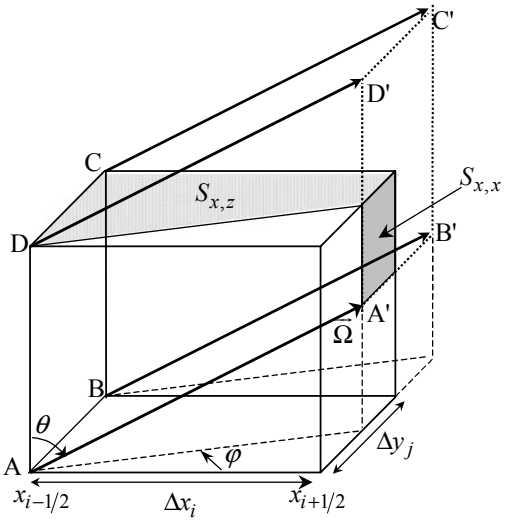



**Fig. 8.11.** Characteristics (↗) of the transport equation in a mesh.

The corresponding to pair $x = x_{i+1/2}$ and $x = x_{i-1/2}$ coefficient $A_{x,x}$ of the exact solution $\psi^{exact}(x, y, z)$ can be defined by the equality

$$A_{x,x} = U/V\,,$$

where

$$\begin{aligned} U &= \frac{1}{S_{x,x}} \int_{S_{x,x}} \mathrm{d}y\,\mathrm{d}z\,\psi^{\mathrm{exact}}(x_{i+1/2}, y, z)\,, \\ V &= \frac{1}{\Delta y_j\,\Delta z_k} \int_{y_{j-1/2}}^{y_{j+1/2}} \mathrm{d}y \int_{z_{k-1/2}}^{z_{k+1/2}} \mathrm{d}z\psi^{\mathrm{exact}}(x_{i-1/2}, y, z)\,. \end{aligned}$$

Similarly, the remaining coefficients $A_{t,h}$ can be defined.

Some of coefficients $A_{t,h}$ are equal to zero. It means that radiation from the corresponding 'entering' edge does not reach the corresponding 'outgoing' edge. For example, if $tg\varphi < \Delta y_j/\Delta x_i$, then radiation from the edge $y = y_{j-1/2}$ does not reach the edge $y = y_{j+1/2}$ and $A_{y,y} = 0$ (see Fig. 8.11).

Assuming, that the exact solution $\psi^{\text{exact}}(x, y, z)$ is a constant at each mesh edge, we can define the coefficients of SC scheme as coefficients of the exact solution:

$$a_{t,h} = A_{t,h}|_{\psi^{\text{exact}}(x,y,z)=\text{const}} ,$$

see above. In particular, if the coefficient $A_{t,h} = 0$, then the corresponding coefficient of the SC scheme is also equal to zero.

The features of the SC scheme discussed above permit us to calculate both smooth solutions and discontinuous ones in a correct way. This is confirmed by the calculation practice as well. At the same time, the scheme is of first-order accuracy only (in the uniform norm) (Nikolaeva, 2004), i.e. the parameter $q$ in the estimation (8.17) is equal to 1.

To increase the scheme accuracy, it is necessary to exploit more precise, possibly positive, approximations of the solution at 'entering' mesh edges. Constructing such approximations, we have to use solution moments of some adjacent meshes. The piecewise constant function (Mathews, 1999) and the exponent of linear function (Castriani and Adams, 1995) represent examples of positive approximations.

More accurate representations may be created in the frames of nodal schemes, where both zero moments and the ones of higher order are included. For example, using moments

$$\psi^{1,y}_{i+1/2,j,k} = \frac{3}{\Delta y_j\, \Delta z_k} \int_{y_{j-1/2}}^{y_{j+1/2}} \mathrm{d}y \bar{P}_1(y) \int_{z_{k-1/2}}^{z_{k+1/2}} \mathrm{d}z\, \psi(x_{i+1/2}, y, z) ,$$

$$\psi^{1,z}_{i+1/2,j,k} = \frac{3}{\Delta y_j\, \Delta z_k} \int_{y_{j-1/2}}^{y_{j+1/2}} \mathrm{d}y \int_{z_{k-1/2}}^{z_{k+1/2}} \mathrm{d}z \bar{P}_1(z)\, \psi(x_{i+1/2}, y, z) ,$$

compare with (8.16), we can construct the following linear approximation

$$\psi(x_{i+1/2}, y, z) \simeq \psi_{i+1/2,j,k} + \bar{P}_1(y)\, \psi^{1,y}_{i+1/2,j,k} + \bar{P}_1(z)\, \psi^{1,z}_{i+1/2,j,k} . \qquad (8.23)$$

The zero and the first moments of this expansion are equal to corresponding moments of the exact solution. Similarly, approximations with polynomials of higher order can be constructed.

As these approximations are not always positive, the corresponding grid schemes will not be positive as well (Azmy, 1992; Elsawi et al., 2003). The positive nodal scheme of short characteristics was successfully constructed based on long characteristics method applied to each spatial mesh (Santandrea and Sanchez, 2002).

It should be noted that all the considered schemes have no corrections.

## 8.9 Integro-interpolational schemes

### 8.9.1 Zero spatial moments schemes without corrections

The construction of integro-interpolational schemes in a mesh is always based on a *balance relation*, which can be obtained as a result of action of the integral operator

$$\frac{1}{\Delta x_i \Delta y_j \Delta z_k} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathrm{d}x \int_{y_{j-1/2}}^{y_{j+1/2}} \mathrm{d}y \int_{z_{k-1/2}}^{z_{k+1/2}} \mathrm{d}z$$

on Eqs (8.1), (8.19). In $(x, y, z)$ geometry, the balance relation has a form

$$\begin{aligned}(\psi_{i+1/2,j,k} - \psi_{i-1/2,j,k})\,\xi/\Delta x_i + (\psi_{i,j+1/2,k} - \psi_{i,j-1/2,k})\,\eta/\Delta y_j \\ +(\psi_{i,j,k+1/2} - \psi_{i,j,k-1/2})\,\gamma/\Delta z_k + K_{i,j,k}\,\psi_{i,j,k} = B_{i,j,k}\,. \qquad (8.24)\end{aligned}$$

Here values $\psi_{i,j,k}$, $\psi_{i\pm1/2,j,k}$, $\psi_{i,j\pm1/2,k}$, $\psi_{i,j,k\pm1/2}$, $B_{i,j,k}$, $\sigma_{t,i,j,k}$ are defined by the expressions (8.15), (8.21) and (8.22).

To uniquely determine values $\psi_{i,j,k}$, $\psi_{i+1/2,j,k}$, $\psi_{i,j+1/2,k}$, $\psi_{i,j,k+1/2}$, the values $\psi_{i-1/2,j,k}$, $\psi_{i,j-1/2,k}$, $\psi_{i,j,k-1/2}$, $B_{i,j,k}$ being known, we join the *additional equations* to the exact balance relation (8.24). The *WDD (Weighted Diamond Difference) scheme* is often used in calculations. The relations for radiation directions from the octant $\{0 < \gamma < 1,\ 0 < \varphi < \pi/2\}$, being written in the form

$$\begin{aligned}\psi_{i,j,k} &= (\psi_{i+1/2,j,k} + p_{x,i,j,k}\,\psi_{i-1/2,j,k})/(1 + p_{x,i,j,k})\,, \\ \psi_{i,j,k} &= (\psi_{i,j+1/2,k} + p_{y,i,j,k}\,\psi_{i,j-1/2,k})/(1 + p_{y,i,j,k})\,, \\ \psi_{i,j,k} &= (\psi_{i,j,k+1/2} + p_{z,i,j,k}\,\psi_{i,j,k-1/2})/(1 + p_{z,i,j,k})\,, \qquad (8.25)\end{aligned}$$

are used in WDD as additional equations. Here the weight parameters $p_{x,i,j,k}$, $p_{y,i,j,k}$, $p_{z,i,j,k}$ belong to the segment [0,1].

Traditionally, before solving the equations (8.24) and (8.25) in arbitrary mesh, the first moment $\psi_{i,j,k}$ is defined. After that the moments at 'outgoing' edges are calculated by means of Eqs (8.25). The approach permits to reduce the number of arithmetic operations. Further the sequential mesh sorting, which was outlined above for the SC scheme, is applied.

Features of a given WDD scheme depend on its weight parameters. Two WDD schemes are often used: the *DD (Diamond Difference) scheme* and the *St (Step) scheme*. Their weights are defined by the equalities $p_{x,i,j,k} = p_{y,i,j,k} = p_{z,i,j,k} = 1$ and $p_{x,i,j,k} = p_{y,i,j,k} = p_{z,i,j,k} = 0$, correspondingly. As has been shown (Madsen, 1975) only the DD scheme among all WDD schemes possesses second-order accuracy in mean square norm. The WDD schemes with other weights are of the first-order accuracy.

From the first glance, the obtained estimations indicate, that it is only worth using the DD scheme for transport problems. However, as calculation practice shows, really the DD solutions turn out to be negative or contain non-physical oscillations.

We would like to explain the reason for the mentioned effects. For this purpose, we explicitly rewrite the calculation formulas of DD scheme for cubic grid ($\Delta = \Delta x_i = \Delta y_j = \Delta z_k$). For the mesh with indices $i, j, k$ we have the following expressions:

$$
\begin{aligned}
&\psi_{i,j,k} \\
&= \frac{\xi\,\psi_{i-1/2,j,k} + \eta\;\psi_{i,j-1/2,k} + \gamma\,\psi_{i,j,k-1/2} + \Delta\,B_{i,j,k}/2}{\gamma + \xi + \eta + K_{i,j,k}\,\Delta/2}\,, \\
&\psi_{i+1/2,j,k} \\
&= \frac{(\xi - \gamma - \eta - K_{i,j,k}\Delta/2)\,\psi_{i-1/2,j,k} + 2\eta\,\psi_{i,j-1/2,k} + 2\gamma\,\psi_{i,j,k-1/2} + \Delta\,B_{i,j,k}}{\gamma + \xi + \eta + K_{i,j,k}\,\Delta/2}\,, \\
&\psi_{i,j+1/2,k} \\
&= \frac{2\xi\,\psi_{i-1/2,j,k} + (\eta - \gamma - \xi - K_{i,j,k}\Delta/2)\;\psi_{i,j-1/2,k} + 2\gamma\,\psi_{i,j,k-1/2} + \Delta\,B_{i,j,k}}{\gamma + \xi + \eta + K_{i,j,k}\,\Delta/2}\,, \\
&\psi_{i,j,k+1/2} \\
&= \frac{2\xi\,\psi_{i-1/2,j,k} + 2\eta\,\psi_{i,j-1/2,k} + (\gamma - \xi - \eta - K_{i,j,k}\Delta/2)\,\psi_{i,j,k-1/2} + \Delta\,B_{i,j,k}}{\gamma + \xi + \eta + K_{i,j,k}\,\Delta/2}\,.
\end{aligned}
\tag{8.26}
$$

As one can verify, these equations are satisfied identically in the case of the linear solution

$$\psi(x, y, z) = A + B\,x + C\,y + D\,z\,,$$

where $A$, $B$, $C$, $D$ are arbitrary constants. It confirms the fact, that DD scheme is of second-order accuracy for smooth solutions. In the addition, zero solution moment $\psi_{i,j,k}$ is always positive for positive $B_{i,j,k}$ and moments on 'entering' edges. Nevertheless, moments on 'outgoing' edges may be negative, if $\left|\psi_{i-1/2,j,k} - \psi_{i,j-1/2,k}\right| \gg 1$ or $\left|\psi_{i-1/2,j,k} - \psi_{i,j,k-1/2}\right| \gg 1$ or $\left|\psi_{i,j-1/2,k} - \psi_{i,j-1/2,k}\right| \gg 1$, i.e. if the grid solution decreases too quickly (see Fig. 8.12).

Let $\psi_{i,j-1/2,k} = \psi_{i,j,k-1/2} = \psi_{i+1,j-1/2,k} = \psi_{i+1,j,k-1/2} = B_{i,j,k} = 0$. Then Eq. (8.26) can be written in form

$$\psi_{i+1/2,j,k} = \rho\,\psi_{i-1/2,j,k}\,,\;\; \psi_{i+3/2,j,k} = \rho^2\,\psi_{i-1/2,j,k}\,,\;\; \rho = \frac{\xi - \gamma - \eta - K_{i,j,k}\Delta/2}{\gamma + \xi + \eta + K_{i,j,k}\Delta/2}\,.$$

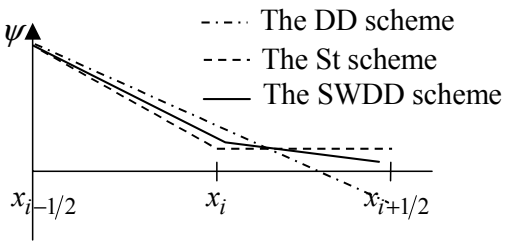



**Fig. 8.12.** WDD schemes solution in a mesh.

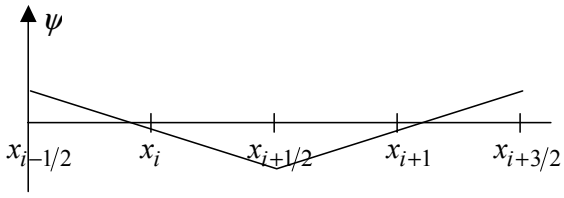



**Fig. 8.13.** DD scheme solution for a pair of meshes.

As $\xi - \gamma - \eta - \sigma_{t,i,j,k}\Delta/2 < 0$, the coefficient $\rho$ is negative and non-physical oscillations in DD solutions arise (see Fig. 8.13).

Therefore, the additional relations in the DD scheme are too coarse to calculate quickly varying functions.

Calculation formulas of the St scheme with a cubic mesh can be written in the form

$$\begin{aligned}\psi_{i+1/2,j,k} &= \psi_{i,j,k+1/2} = \psi_{i,j,k+1/2} = \psi_{i,j,k} \\ &= \frac{\xi\,\psi_{i-1/2,j,k} + \eta\,\psi_{i,j-1/2,k} + \gamma\psi_{i,j,k-1/2} + \Delta B_{i,j,k}}{\gamma + \xi + \eta + K_{i,j,k}\Delta}\,.\end{aligned}$$

These equations are converted into identities only if the solution is constant. It is the consequence of the fact, that St scheme is of first-order accuracy for smooth solutions. Thus, although this scheme is positive (see Fig. 8.12) it is of low accuracy.

So we have a non-positive and non-monotonic DD scheme of second-order accuracy and a positive and monotonic St scheme of first-order accuracy. This fact confirms the known theorem by S. K. Godunov: *it is impossible to construct monotonic linear discretizations for hyperbolic systems of equations that would be of second- or of higher-order accuracy, if one is restricted by zero solution moments only.*

For this reason schemes of two different classes have been developed. They are

- positive and monotonic schemes of a higher-order accuracy as compared to the St scheme,
- non-positive and non-monotonic schemes, which more rarely generate negative and non-physical oscillatory grid solutions for practical problems as compared to the DD scheme.

The first class includes WDD schemes with a more flexible algorithm with respect to the definition of weights. One such algorithm leads to the *SWDD (Special WDD) scheme* (Bass and Nikolaeva, 1997; Nikolaeva, 2004), its weights are calculated using the following equations:

$$a_{x,x}(p_x, p_y, p_z) = \bar{a}_{x,x}\,, \quad a_{y,y}(p_x, p_y, p_z) = \bar{a}_{y,y}\,, \quad a_{z,z}(p_x, p_y, p_z) = \bar{a}_{z,z}\,.$$

Here the values $a_{x,x}, a_{y,y}, a_{z,z}$ and $\bar{a}_{x,x}, \bar{a}_{y,y}, \bar{a}_{z,z}$ are coefficients of Eq. (8.20) for the SWDD scheme and for the SC scheme respectively. As it turned out, the following assertions are valid:

- All coefficients of the SWDD scheme are non-negative.
- Each non-diagonal coefficient ($a_{t,h}$ for $t \neq h$) of the SWDD scheme is close to the corresponding coefficient of SC scheme, if the spatial grid is a sufficiently fine one.
- If any coefficient of the SC scheme is equal to zero, then the corresponding coefficient of the SWDD scheme is equal to zero as well.

In other words, the features of integro-interpolational SWDD scheme are similar to those of the SC scheme of short characteristics, and so the SWDD scheme also can be used for calculation of discontinuous solutions (Nikolaeva, 2004). It should be noted, however, that the SWDD scheme is slightly more accurate than the SC one. In addition, the SWDD scheme, being convenient for the implementation (the code RADUGA-5.1(P), see section 8.13.2), has demonstrated its advantages in atmospheric optics and radiation protection problems.

It should mentioned, that the SWDD scheme weights in any mesh depend on the radiation transport direction $\vec{\Omega}$ (that is, on $\gamma$ and $\varphi$, mesh sizes $\Delta x_i$, $\Delta y_j$, $\Delta z_k$, and on the value of the extinction coefficient in the mesh $K_{i,j,k} = K_{\text{ext}}(x_i, y_j, z_k)$). Remember, that the grid solution inside a mesh is a positive piecewise linear function, see Fig. 8.12, whereas DD solution is a linear function, which can be negative (see Fig. 8.12).

The *MDSN*$_1$ *scheme* (Nikolaeva, 2004) is positive and monotonous one as well. Coefficients of the MDSN$_1$ scheme converge to coefficients of the SWDD scheme under grid condensing. But insofar as the grid is coarse, these two schemes are different.

Improved additional relations as compared to the DD scheme are used in schemes of the second class. Sometimes the additional relations include moments of adjacent meshes (Morel and Larsen, 1990; Adams, 1991). In other cases the relations are obtained with the help of the analytical solution of the transport equation (8.1), averaged over all spatial variables except some single variables (Azmy, 1988a; Zhougsheng et al., 1994). The *EC (Exponential) scheme* with non-linear additional relations (Barbucci and Pasquantonio, 1977) is worthly of special mention. It is the scheme with relations of the kind

$$\psi_{i,j,k} = \sqrt{\psi_{i+1/2,j,k}\ \psi_{i-1/2,j,k}} = \sqrt{\psi_{i,j+1/2,k}\ \psi_{i,j-1/2,k}} = \sqrt{\psi_{i,j,k+1/2}\ \psi_{i,j,k-1/2}}\,.$$

Schemes based on the quasi-stationary derivatives principle (Suslov and Pevey, 1997) are also of special interest. Differentiating the transport equation (8.1) with respect to each spatial variable and equating all second derivatives to zero (here the solution is suggested to be slowly varying), one can find approximate expressions for the first derivatives of $B(\vec{r}, \vec{\Omega})$. Using these expressions, more accurate representations for both $B(\vec{r}, \vec{\Omega})$ and the solution can be constructed.

### 8.9.2 Zero spatial moment schemes with corrections

The impossibility of constructing a positive and monotonous scheme of second-order accuracy forces us to use schemes with corrections. As a result *WDD schemes with corrections* have been actively developed. In these schemes, first a mesh is calculated via a DD scheme and then, if the grid solution turnes out to be a negative one or even oscillatory, the mesh is recalculated via a WDD scheme of first-order accuracy.

The earliest of the schemes is the *DD/St scheme*, where the mesh recalculation is carried out via the St scheme, if at least one of 'outgoing' edges provided a negative moment of the solution. However, correction of this kind is too rough, because the weight values undergo jumps inside a mesh. Thus, these schemes are positive, but non-monotonous ones.

More flexible corrections are exploited in the positive and non-monotonous *θ WDD scheme* (Rhoades and Engle, 1977; Petrovic and Haghighat, 1996). Finally, an algorithm of oscillation smoothing is included in the grid equations of the *AWDD (Adaptive WDD) scheme* (Carlson, 1976; Voloschenko and Germogenova, 1994). Here the user can choose the degree of oscillation smoothing via the specification of the parameter for monotonization of the solution. In *MDSN scheme* corrections the numerical algorithm relies on the condition of monotonicity of the grid solution (Voloschenko, 1981).

In multidimensional problems, both MDSN and AWDD schemes require the iterative construction of scheme equations in each mesh to ensure grid solution monotonocity for all variables. Usually one needs to use one or two iterations. Under many iterations, MDSN equations tend to equations of the monotonous $MDSN_1$ *scheme* (Nikolaeva, 2004). In slab geometry MDSN and MDSN1 schemes are equivalent.

Finally, in some cases corrections are carried out via an *MDSN scheme* in selected meshes, adjoint to the surfaces, where transport equation coefficients undergo discontinuities. Here it is presupposed that oscillations arise mainly in such meshes. Corrections are performed just in the mesh set. Numerical results show that monotonization techniques only slightly smooth oscillations.

### 8.9.3 Nodal schemes

To complete our review of modern numerical schemes for the solution of the transport equation, we describe the so-called the nodal integro-interpolational schemes. Spatial moments of zero- and first-order (and sometimes higher orders) are used in these schemes for the construction of grid equations.

Several balance relations are used in the scheme construction. Each of the balance relations is obtained via integration of the transport equation (8.1) over a spatial mesh with an appropriate weight function. In some nodal schemes, the additional relations are based on the non-positive polynomial (Walters, 1982, 1986; Badruzzaman, 1985; Azmy, 1988b; Warin, 1996; Voloschenko, 1997; Zmijarevic, 1999; Takeda and Yamamoto, 2001), the piecewise polynomial (Voronkov and Sychugova, 1997) or the exponential (Ullo et al., 1982) representations of a

grid solution in a mesh. Although these schemes have a high order of accuracy (up to four), usually they are neither positive nor monotonous.

Positive nodal integro-interpolational schemes are of lower-order accuracy (second to third) than non-positive ones. They are based on the approximation of the solution via piecewise constant (Mathews and Minor, 1991), piecewise linear (Mathews and Minor, 1993) or exponential linear (Walters et al., 1995; Wareing, 1997) functions.

Nodal schemes with corrections are also actively developed. They combine the preliminary grid calculation via non-positive and non-monotonous scheme of high order and its consequent recalculation via a positive and a partially monotonous scheme of a lower-order accuracy (Voloschenko, 1997; Shwetsov, 1997).

## 8.10 Finite element schemes

At preliminary step of the construction of finite elements schemes, a specific basis of functions $\{u_g(\vec{r},\vec{\Omega})\}$, $g = 1,\ldots,G$ is chosen. The solution in each mesh is represented by a linear combination:

$$\psi(\vec{r},\vec{\Omega}) \simeq \tilde{\psi}(\vec{r},\vec{\Omega}) = \sum_{g=1}^{G} u_g(\vec{r},\vec{\Omega})\, T_g\,.$$

Coefficients $T_g$ are defined by equating mesh solution moments at mesh edges to corresponding moments of $\tilde{\psi}(\vec{r},\vec{\Omega})$ representation. Further the residual $R(\vec{r},\vec{\Omega})$ for the function $\tilde{\psi}(\vec{r},\vec{\Omega})$ via Eq. (8.1) is calculated:

$$R(\vec{r},\vec{\Omega}) = \hat{L}\tilde{\psi}(\vec{r},\vec{\Omega}) - B(\vec{r},\vec{\Omega})\,.$$

After that the scheme equations are defined via the integration of the equality $R(\vec{r},\vec{\Omega}) = 0$ over the mesh. Also some weight functions, which in the general case do not coincide with basis functions $u_g(\vec{r},\vec{\Omega})$, are used. The chosen functions $u_g(\vec{r},\vec{\Omega})$, defining grid solution form, are named finite elements.

Using only zero spatial moments and bilinear combinations of functions $u_g(\vec{r},\vec{\Omega})$, one can successfully construct an effective calculation scheme (Morel et al., 1993).

Using moments of high orders, nodal schemes of finite elements can be created (Hennart and del Valle, 1997; del Valle and Alonso, 2001).

## 8.11 The solution of the grid equation

The previous sections were aimed at the construction of the grid equation, approximating both the differential operator $\widehat{L}$ and the integral operator $\widehat{S}$ in the transport equation (8.1). As a result we find a system of linear algebraic equations

$$\widehat{L}^{\text{grid}}\vec{\psi} = \widehat{S}^{\text{grid}}\,\vec{\psi} + \vec{Q}\,, \tag{8.27}$$

where the vector $\vec{\psi}$ is formed by grid solution values in all meshes of spatially-angular grid, and grid operators $\widehat{L}^{\text{grid}}$ and $\widehat{S}^{\text{grid}}$ correspond to the differential operator $\widehat{L}$ and the integral operator $\widehat{S}$. In section 8.7 an explicit method, based on operator $\widehat{L}^{\text{grid}}$ inversion, was presented. The system (8.27) is solved using the *simple iteration method*:

$$\vec{\psi}^{it+1} = (\widehat{L}^{\text{grid}})^{-1}[\widehat{S}^{\text{grid}}\,\vec{\psi}^{it} + \vec{Q}]\,. \tag{8.28}$$

Iterations are stopped when the relative variation of a scalar flux (i.e. solution averaged over angle variables) inside all spatial meshes is less then a value of $\varepsilon$ ($\varepsilon$ being specified by a user). Usually the values $\varepsilon$ inside the interval (0.0001–0.001) are used in calculations.

Insofar as the operator $\widehat{S}^{\text{grid}}$ describes light scattering processes, each new iteration adds to a grid solution contribution, provided photons undergo the next order of scattering. Therefore, the method (8.28) is known as *successive-orders-of-scattering method (SOSM)*.

If the size of the calculation region, measured in optical lengths, is large and absorption is weak (the typical situation for atmospheric optic problems in the visible and UV range), then on the average a photon undergoes a great number of scatterings, and so the iterative process (8.28) converges very slowly.

The Seidel scheme of Eq. (8.27) solving has proved to be more effective. It consists in obtaining solution values inside a separate octant at each iteration and subsequently utilizing them for Eq. (8.27) right-hand side calculation for other octants at the same iteration. This method is slightly more effective in solving Eq. (8.27) as compared to SOSM. Two-layer iterative schemes have proved to be even more effective. Their equations can be written in the form:

$$\widehat{L}^{\text{grid}}\vec{\psi}^{it+1/2} = \widehat{S}^{\text{grid}}\,\vec{\psi}^{it} + \vec{Q}\,, \quad \vec{\psi}^{it+1} = \vec{\psi}^{it+1/2} + \vec{\delta}^{it+1/2} \tag{8.29}$$

The equations for corrections $\vec{\delta}^{it+1/2}$ are constructed based on the fact that the equation for a grid solution error $\vec{E}^{it+1/2} = \vec{\psi} - \vec{\psi}^{it+1/2}$ satisfies the following relation:

$$\widehat{L}^{\text{grid}}\,\vec{E}^{it+1/2} = \widehat{S}^{\text{grid}}\,\vec{E}^{it+1/2} + \widehat{S}^{grid}(\vec{\psi}^{it} - \vec{\psi}^{it+1/2})\,.$$

Hence, the correction $\vec{\delta}^{it+1/2}$ can be naturally defined via some approximations of the vector $\vec{E}^{it+1/2}$. In a wide class of methods, a correction $\vec{\delta}^{it+1/2}$ can be sought *in the low approximation over angular variables* (Adams and Larsen, 2002), vector $\vec{\delta}^{it+1/2}$ being a linear function or a constant in each octant with respect to angular variables. Similar methods were successfully exploited in radiation shielding problems, because they make it possible to decrease the number of iterations tens or hundreds of times. But they are proved to be inapplicable in atmospheric optics problems, where light is scattered mostly in the forward direction and the correction $\vec{\delta}^{it+1/2}$ cannot be approximated by a polynomial of a low order. For this case, *multigrid angular methods* were developed, where calculations are performed at a sequence of angular grids for each iteration (Morel and Manteuffel, 1991; Pautz and Morel, 1999). Correction to a finer grid so-

lution is obtained via a coarser grid calculation. After adding all the obtained corrections, a new iteration begins with the finest grid.

All listed methods are effective (the number of iterations decreases in order), if only the correction equations are *consistent with* initial grid (Eq. (8.27)), i.e. if correction equations are a consequence of Eq. (8.27). It should be noted that the consistent correction system is more complicated as compared to the inconsistent one.

An effective iterative scheme construction for problem (8.27) is rather difficult due to the fact that the operator $(\widehat{L}^{\text{grid}})^{-1}\widehat{S}^{\text{grid}}$ is not self-adjoint in practical problems. Among all methods developed for systems of type (8.27), only the methods by Krylov (Morel, 2005), which do not use the suggestion of the operator self-adjointness, are proved to be applicable here. In these methods, the corrections $\vec{\delta}^{it+1/2}$ can be presented by the linear combination of vectors $\vec{v}_\mu$:

$$\vec{\delta}^{it+1/2} = \sum_{\mu=1}^{\mathcal{M}} \vec{v}_\mu \, \beta_\mu \, .$$

The vectors $\vec{v}_\mu$ should be chosen from the space formed by residual vectors

$$\{\vec{r}_\mu, \mu = 1, \ldots, M\}\colon \vec{r}_1 = (\widehat{L}^{\text{grid}} - \widehat{S}^{\text{grid}})\vec{\psi}^{it+1/2} - \vec{Q}\,, \quad \vec{r}_\mu = (\widehat{L}^{\text{grid}} - \widehat{S}^{grid})\,\vec{r}_{\mu-1}\,.$$

The coefficients $\beta_\mu$ can be found from the condition of the residual minimum for equations (8.27) in mean square norm, i.e. from the minimum condition for the sum of squares of elements of the vector

$$\vec{Y} = (\widehat{L}^{\text{grid}} - \widehat{S}^{\text{grid}})\,(\vec{\psi}^{it+1/2} + \sum_{\mu=1}^{\mathcal{M}} \vec{v}_\mu \, \beta_\mu) - \vec{Q}\,.$$

As a result, coefficients $\beta_\mu$ are defined by a system of linear algebraic equations of the order $\mathcal{M} \times \mathcal{M}$.

In each version of the methods by Krylov the vectors $\vec{v}_\mu$ are constructed in own manner, based on vectors $\vec{r}_\mu$. The conjugate gradient method, the least squares technique, the quasi-minimal residual algorithm, the generalized minimal residual method and many other approaches use different ways to construct and also to store vectors $\vec{v}_\mu$. The number $\mathcal{M}$ of vectors is also different in different approaches.

The acceleration degree of methods by Krylov is increased when the number $\mathcal{M}$ of vectors $\vec{v}_\mu$ increases, but the calculation of vectors and scalar products is very time-consuming. If the number $\mathcal{M}$ is too large, the method is not effective.

## 8.12 Technique of transport equation solving by the parallel discrete ordinates method

The appearance of parallel-architecture computers has led to a new step in the development of numerical techniques. Obviously, the parallel algorithm construction should correspond to architecture particularities of concrete multiprocessor computers.

There are two main types of multiprocessor systems. In *weakly connected systems with distributed memory* each processor possesses its own disk and operative memory and is able to send data to other processors only via a common network. The information exchange speed between different processors here is lower than one between a single processor and its operative memory. Such kinds of multiprocessor systems can contain a lot of processors (more than a thousand).

In *strongly connected systems* all the processors have an extensive *shared (common) memory*. By means of the shared memory the processors perform the information exchange. The operation system should be capable of preventing errors, which may arise when two or more processors simultaneously deal with the same memory cell. The number of processors in the strongly connected systems is essentially smaller than in weakly connected ones.

The modern architecture is the so-called nested parallelism, where a distributed-memory machine (called a *cluster*) has embedded nodes with several (2, 4, 8) processors. In these machines two mechanisms are used for the communication: (a) shared memory when processors belong to the same node; (b) network communication when the processors belong to the different nodes.

The *effectiveness* of a parallel algorithm is defined by the formula

$$\mathrm{Eff}(\mathcal{N}) = T(1)/[\mathcal{N}\,T(\mathcal{N})]100\%\,, \tag{8.30}$$

where $\mathcal{N}$ is processor number, T(1) is the calculation time for a single processor, $T(\mathcal{N})$ is the calculation time for $\mathcal{N}$ processors. Obviously, the more uniform is the calculation distribution among all the processors the higher is parallel algorithm effectiveness, and so the less time it takes for a single processor to exchange information with other processors.

The parallelizing technique depends on the problem type. If the line-by-line calculation is carried out, then a high effectiveness of parallelizing can be achieved by means of the uniform distribution of problems for different wavelengths among all processors.

A parallel algorithm for the long characteristics method can be reduced to the proportional distribution of the calculation of different characteristics among the processors (Dahmani et al., 2003).

The parallel algorithm for solving transport problems for a single wavelength can be based on the calculation region decomposition. In the case of the angular decomposition (Azmy, 1988b) at one iteration all nodes of angular quadrature are distributed among different processors and the solution at each node is calculated by the corresponding processor. This kind of parallelizing implies that each single processor should transmit to previously chosen summarizing processor solution values at all the meshes to provide the consequent scattering integral calculation. After that each processor should get from the summarizing processor the scattering integral values at all spatial meshes. Similar algorithms should be realized on strongly connected computers in order to increase the high calculation effectiveness under a great number of information exchange operations.

If a spatial decomposition is used, the calculation region is divided into subregions, incoming radiation fluxes being supposed to be known at the sub-region

boundary. The fluxes are known either from boundary conditions or via neighboring sub-region calculations at a previous iteration (Hanebutte and Lewis, 1991; Haghighat and Azmy, 1991; Baker et al., 1995; Pautz, 2001).

In some cases the decomposition is realized both on spatial and on angular variables (Dorr and Salo, 1995; Sjoden and Haghighat, 1997; Nowak and Nemanic, 1999; Fischer and Azmy, 2003). Specially developed expert systems permit the estimation of time and memory expenses for each decomposition type (spatial, angular or spatially-angular) for each concrete calculation (Patchimpattapong and Haghighat, 2003).

Different calculation procedures for sub-regions are possible. For example, each sub-region can be computed by own processor at each iteration (Haghighat and Azmy, 1991). In the last case the algorithm includes two iterative processes: one for the calculation of the scattering integral and yet another one for the calculation of the solution values at sub-region boundaries. Naturally, with the number of sub-regions the necessary number of iterations increases as well. Besides, the obtained grid solution usually does not coincide with that obtained by a single processor at each iteration. Special parallel acceleration algorithms should be developed in such cases.

Accordingly to the 'red-black algorithm' (Hanebutte and Lewis, 1991; Sjoden and Haghighat, 1997), all the sub-regions are specified as 'red' or 'black'. At each iteration either only 'red' sub-regions or only 'black' ones participate in the calculation procedure (see Fig. 8.14). It permits to obtain more accurate values of entering fluxes at sub-region boundaries at each iteration.

According to the 'diagonal scheme' (Baker et al., 1995; Dorr and Salo, 1995), for any transport direction $\vec{\Omega}$, a special sequence of sub-region calculations, based on chosen angular quadrature, is used. For example, calculation begins with the left lower corner for a direction $\vec{\Omega}\{\gamma, \varphi\}$ at $\{0 < \gamma < 1, 0 < \varphi < \pi/2\}$ (see Fig. 8.1). The sequence is shown in Fig. 8.15. A number, marking each sub-region in Fig. 8.15, denotes the iteration number of sub-region participation in the common calculation procedure. In this case the solution, obtained by the parallel method, coincides with one obtained via a single processor method. Thus, it makes possible to use, in a parallel calculation, procedure acceleration algorithms which have been developed for single-processor calculations.

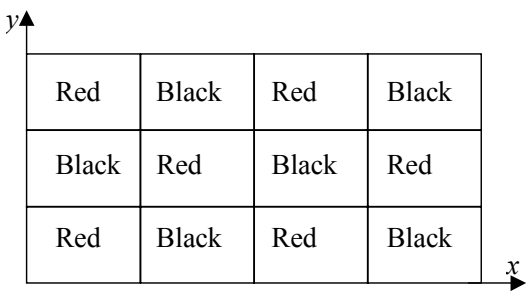



**Fig. 8.14.** Red-black algorithm scheme.

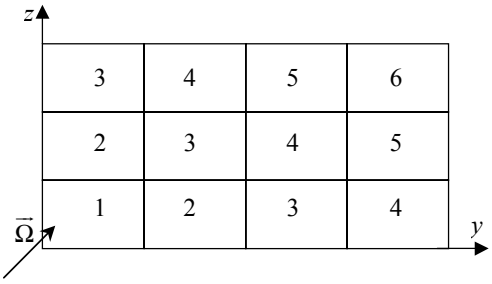



**Fig. 8.15.** Diagonal scheme.

## 8.13 Discrete ordinates codes

The project I3RC (http://i3rc.gsfc.nasa.gov) is of a great importance in 3D radiation transport (RT) codes verification. 'The goal of I3RC is to promote the improvement of algorithms that are used for all kinds of 3D RT processes in cloudy atmospheres. Activities include not only comparisons of results from state-of-the-art 3D RT codes, but also development of fast approximations that are more suitable for climate applications and community "open source" codes that distill the best current knowledge on how to treat the various interactions of ultraviolet, visible, and infrared photons with atmospheric constituents' (Cahalan et al., 2005).

For these purposes a number of test problems have been formulated for the verification of DOM and Monte Carlo algorithms. Participants in this project just created the codes providing calculation of the test problem with the required accuracy.

The main DOM code is SHDOM (Evans, 1998); recently RADUGAP-5.1(P) was included in the list of codes being used. Other results were obtained by Monte Carlo codes.

Now the project consists of three phases aimed at calculations of 1D, 2D and 3D problems accordingly. The authors suggest that 'the cloud and climate modeling community is further ahead of its remote sensing counterpart in incorporating the advances of 3D RT into its representation of radiative processes'.

### 8.13.1 SHDOM code

Currently the code SHDOM (Evans, 1998), belongs to widely used codes for the calculation of radiation transport in the terrestrial atmosphere (http://nit.colorado.edu/∼evans/shdom.html). Let us list its main attributes.

- The transport equation is solved in regions with Cartesian 1D, 2D and 3D coordinates (see Fig. 8.2).
- The radiation source is the Sun or thermal sources.
- Phase functions are represented by expansions on Legendre polynomials, the expansion order being arbitrary. To decrease the number of polynomials in the expansion, the delta-M method can be used (Wiscombe, 1977).

- Transport equation coefficients (the extinction coefficient, the single scattering albedo, the number of moments in phase function expansions) are defined by their values at spatial grid nodes.
- The boundary condition on the region external boundaries can be specified as periodicity condition or as boundary reflection law (in particular, the mirror reflection law and the Lambert reflection law are included).
- The transport equation for scattered radiation intensity is solved by the long characteristic method. Non-scattered radiation intensity is determined analytically.
- The spatial grid is defined by a user; it can be refined during the calculation procedure, if it is necessary for more accurate solution presentation.
- The simple iterative method with respect to scattering orders is used for solving grid equations; the iteration acceleration method, representing a version of the minimal residual method for a single-processor computer, is included.

The SHDOM is the numerical code for wide-range applications. Although a version of the long characteristics method is used in the code, it does not include the balance equation for a grid mesh. Besides, the acceleration method stability proof is absent. In some cases the iterative process does not converge.

### 8.13.2 The code RADUGA-5.1(P)

The code RADUGA-5.1(P) solves the transport equation in 2D and 3D regions under sufficiently common suggestions on source, phase functions and boundary conditions (Nikolaeva et al., 2005a,b). In particular, it may be used in such areas as atmospheric optics, radiation shielding problems, biomedicine, ray therapy, etc.

Below we outline the main characteristics of the RADUGA-5.1(P) code.

- The transport equation is solved in regions with Cartesian $(x, y, z)$, $(x, y)$ and cylindrical $(r, \vartheta, z)$, $(r, \vartheta)$, $(r, z)$ coordinates (see Fig. 8.2 and Fig. 8.16).

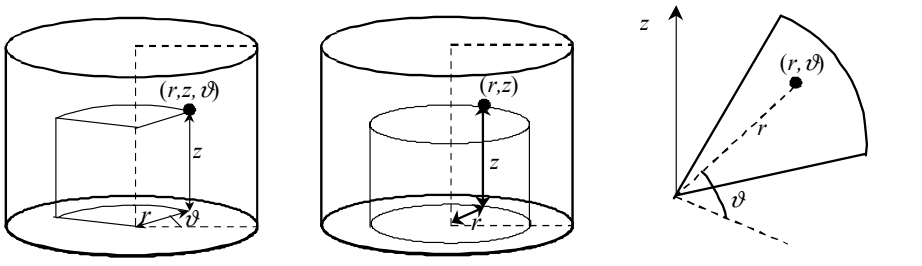



**Fig. 8.16.** Calculation regions with cylindrical coordinates.

- The radiation source may be given in the form of plane-parallel radiation flux (the Sun) or uniform radiation distribution in all directions (e.g., heated body).[1]
- Phase functions can be represented by expansions on Legendre polynomials, the expansion order being arbitrary, or by discrete-angular forms.
- Medium inhomogeneities are defined via surfaces of some model bodies (parallelepipeds, cylinders, spheres, cones, prisms and others). Transport equation coefficients (the extinction cross-section and the single scattering albedo) can be also defined at the nodes of a spatial grid.
- Periodic boundary conditions, the mirror reflection or the Lambert reflection law may be defined over any external boundary. The boundary condition at semi-infinity is also included.
- Angular quadratures of several widely used types (in particular, Carlson and Gauss quadratures) can be used for the scattering integral approximation; regular spatial grids are also used.
- Unscattered radiation intensity is calculated analytically. Grid schemes of first and second accuracy order of well known WDD family (DD, St, DD/St, AWDD, SWDD) are used for scattered radiation calculation.
- After spatial decomposition of a calculation region into sub-regions, the local spatial and angular grids may be introduced in each of them.
- Condensing of angular grids inside some solid angles is admissible.
- The simple iteration method with acceleration by the minimal residuals method is used for grid equations solving.
- The parallel algorithm is designed based on the international MPI standard. The spatial decomposition is used, each sub-region being calculated at each iteration.
- Calculation time is decreased while processor number $\mathcal{N}$ is increased, but the calculation time is not proportional to $\mathcal{N}$. So, the effectiveness of parallelizing is decreased as processor number increases (see (8.30)). In the case of a weakly connected computers with 20 processors the effectiveness is equal to 90%. If the number of processors is between 20 and 80, the effectiveness is near 65%. When the processor number is increased up to 200, the effectiveness is decreased up to 50%. The effectiveness reduction could be associated with increasing time of information exchange between processors.
- The code construction makes it possible to carry out calculations on weakly connected, on strongly connected and on clusters of multiprocessor computers, including personal computers.

Let us briefly outline the semi-infinity boundary condition, which is used in the case of a semi-infinite medium. An example is shown in Fig. 8.17, where both aerosol and cloud media are homogeneous and they possess a common vertical boundary.

[1] The code RADUGA-5.1(P) may be applied to solve problems with other sources. They are point isotropic, point anisotropic (in particular, source radiating in a cone of a small aperture) and ray sources.

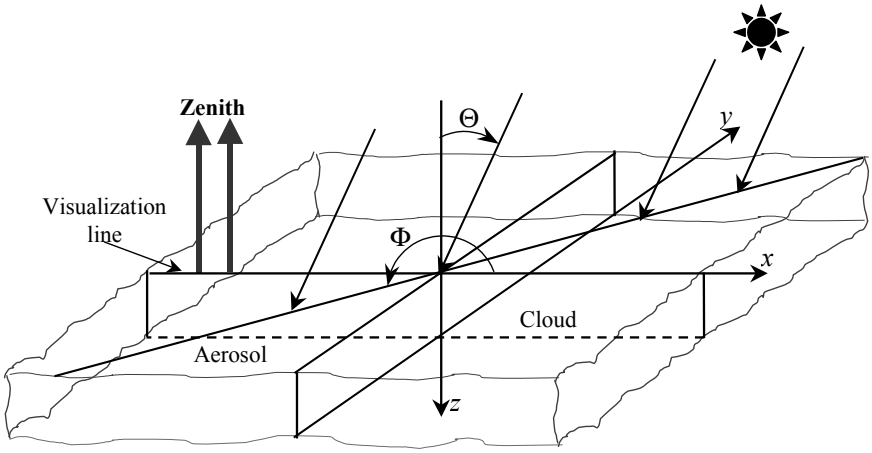



**Fig. 8.17.** Model problem calculation region.

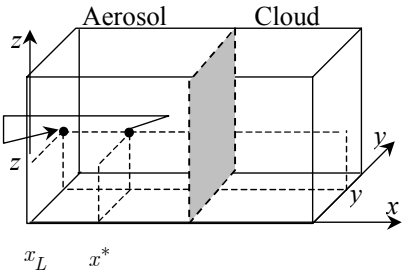



**Fig. 8.18.** Semi-infinity condition at the left boundary.

The semi-infinity condition with respect to variable $x$ at the left boundary $x = x_L$ can be expressed by the formula (see Fig. 8.18):

$$\psi(x_L, y, z, \vec{\Omega}) = \psi(x^*, y, z, \vec{\Omega}), \quad -1 < \gamma < 1,\ 0 < \varphi < \pi/2 \quad \text{or } 3\pi/2 < \varphi < 2\pi, \tag{8.31}$$

Equation (8.31) guarantees the solution independence on the variable $x$ far from the boundary of the media. Moreover, far from the boundary the solution can be approached by the constant, which can be defined from the corresponding 1D model.

Verification of the code RADUGA-5.1(P) was carried out, based on the model problem, depicted in Fig. 8.17. Aerosol scattering was modeled by the Heney–Greenstein phase function with asymmetry parameter $g = \omega_1/3 = 0.7$. The cloud C1 phase function (Kokhanovsky, 2006) was obtained based on Mie theory for the wavelength 412 nm. The asymmetry parameter $g$ is equal to 0.86. Layer height is equal to 4 km, cloud optical thickness is chosen to be 30, aerosol optical thickness is equal to 1.2. Photons enter the layer only via the top boundary $z = 0$ (see Fig. 8.17).

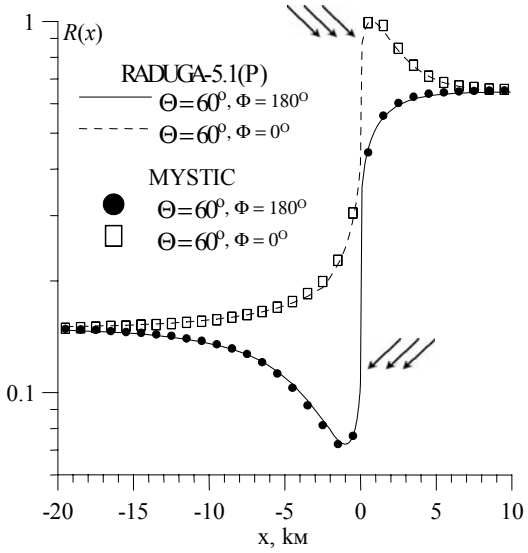



**Fig. 8.19.** Brightness coefficient in the model problem.

The brightness coefficients $R(x) = \pi\,\psi(x, 0, 0)/(F_0 \cos\Theta)$ for the radiation reflected in the zenith at the visualization line for two solar light directions $\Theta = 60°$, $\Phi = 0°$ and $\Theta = 60°$, $\Phi = 180°$ are depicted in Fig. 8.19. Here $F_0$ is intensity of incident radiation. Solar light direction is shown by arrows. Calculation results for the same problem, performed by the Monte Carlo method (code MYSTIC) (Mayer, 1999), are presented in Fig. 8.19 as well. The difference between results, obtained by two different methods, is less than 1% in the areas far from the boundary and less than 4.5% near the boundary, where the exact solution possesses great gradients.

The constant value of brightness far from the boundary can be obtained based on the slab model. The code ROZ-6.5 (Averin et al., 1991), was used for transport problem calculations in the optically thin aerosol slab, whereas asymptotic formulas were applied for radiation obtaining in optically thick cloud. The difference between results, obtained via RADUGA-5.1(P) and 1D results is less than 1%. This fact confirms, that the boundary condition (8.31) properly imitates the semi-infinite medium.

Two imitations of really observable atmospheric phenomena can be seen in Fig. 8.19. These are the *cloud shadow* (minimum of light brightness in aerosol slab near $\Phi = 180°$) and *brightening* (brightness maximum in aerosol slab near $\Phi = 0°$). These two effects are primarily caused by the direct light transfer through a scattering medium. They are relevant not only in the case of a simple model of a semi-infinite homogeneous cloud but also for more general broken cloud systems. They demonstrate the influence of the vertical medium boundary on scattered radiation brightness near this boundary and can be used to study radiative edge effects of cloud optics (Kokhanovsky, 2006).

## 8.14 Simplified discrete ordinates models

### 8.14.1 Accuracy estimation for simple 1D models

The solution of inverse problem s and radiance balance calculations in climate models are currently carried out mainly by means of simple 1D models, where radiation transfer in horizontal directions is not taken into account. The plane-parallel slab model (PP) and the independent column approximation (ICA) belong to the class of 1D models. In the frames of the PP model a cloud is considered as horizontally homogeneous, whereas in the ICA a cloudy model is composed of large homogeneous parallelepipeds (pixels). In the framework of ICA the parallelepipeds can also possess a stratified structure. Such an approximation is also called Independent Pixel Approximation (IPA).

In last decade a great number of papers have studied the features and accuracy of these 1D models under a variety of atmospheric parameters. In our opinion, the main conclusion was formulated by Marshak and Davis (2005): 'It is time to think of 3D theory as the golden standard in atmospheric radiative transfer rather than a perturbation of standard 1D theory'. However, we mention here some results obtained via 1D models to complete the picture.

The capabilities of the ICA model were analyzed in problems of short-wave radiation propagation through deep convective clouds (Giuseppe and Thompkins, 2003) by means of the code SHDOM (Evans, 1998). It has been shown, that energy imbalance is mainly caused by two opposing effects: side illumination and shadowing (see Fig. 8.19). The comparison between ICA and PP models shows that for deep convective clouds geometry-related effects can have a larger influence on radiative transfer calculations than the internal optical inhomogeneities. Similar results were obtained for ultraviolet light transfer problems as well (Scheirer and Macke, 2003). The shortcomings of these investigations are indicated by Giuseppe and Thompkins (2003). They are due to the simplified model of the cloud structure. Hence, the characteristic radiative biases, obtained in the frames of the 1D models, cannot be applied to most real situations. The numerical experiment example and the results obtained are depicted in Table 8.1 and Figs 8.20 and 8.21.

**Table 8.1.** Reflectance, transmittance, and absorptance for IPA and PP biases for two numerical experiments (the solar zenith angle (SZA) is equal to 0 (SZA0) and 60 (SZA60) degrees) (after Guiseppe and Thompkins, 2003)

| | IPA Bias SZA0 | | PP Bias SZA0 | | IPA Bias SZA60 | | PP Bias SZA60 | |
|---|---|---|---|---|---|---|---|---|
| | Abs. | Rel.% | Abs. | Rel.% | Abs. | Rel.% | Abs. | Rel.% |
| Reflectance | 0.016 | 16 | −0.023 | −19 | 0.013 | 8.6 | −0.0095 | −5.9 |
| Transmittance | 0.024 | 3.4 | 0.023 | 3.5 | −0.018 | −2.8 | 0.0082 | 1.3 |
| Absorptance | −0.040 | −23 | 0.0 | 0.0 | 0.0049 | 2.3 | 0.0013 | 0.63 |

Abs., absolute; Rel., relative.

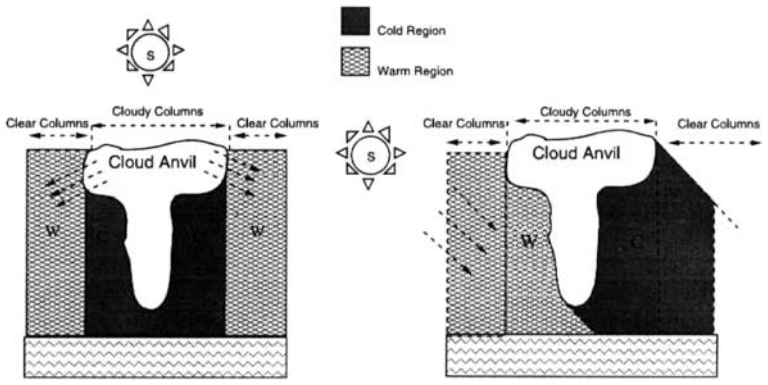



**Fig. 8.20.** Scheme of the redistribution of the radiation between clear and cloudy columns. The clear regions surrounding cloudy columns undergo an enhancement of fluxes with an associated warming (W, lighter shading) in the Sun-overhead case due to the 'spilling' of radiation from the cloud to the clear region and due to the horizontal transport of photons. The opposite happens when the Sun is set at degree zenith angle, with an increased role of shading (darker shading), while clear- sky heating rates on the sunny side of the cloud are enhanced (after Guiseppe and Thompkins, 2003).

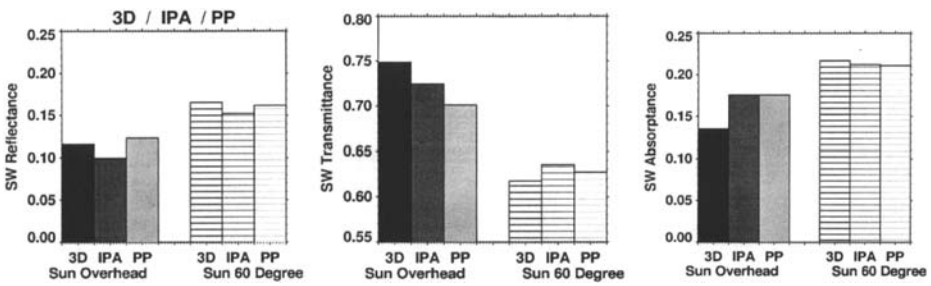



**Fig. 8.21.** Reflectance, transmittance, and absorptance for the two experiments SZA0 and SZA60. The calculation is performed for the whole domain using the full 3D radiative transfer, the independent pixel approximation (IPA), and plane-parallel (PP) methods (after Guiseppe and Thompkins, 2003).

Horizontal radiation transfer in clouds, modeled as inhomogeneous stratified columns, and its influence on cloud energy absorption was investigated by Titov (1998) for the case of lognormal particle distribution and the power law of energy spectrum of cloud depth. In addition, a realistic fractal cloud model was used. It was shown in the frames of the ICA model, that the contribution of the horizontal component to the full radiation field in clouds possesses the following features: (1) it is comparable with transmitted and absorbed radiation; (2) it is strongly dependent on cloud fractal structure and optical depth. The results allow us to conclude that in the majority of situations neglecting horizontal transport will lead to uncertainties in absorption estimations (anomalous absorption). Besides, not accounting for the horizontal transport will result in violation of the one-

to-one correspondence between optical and radiative pixel characteristics. Two possible ways of overcoming the difficulties mentioned are discussed in the paper.

The smoothness properties of radiation fields for stratified clouds, horizontal fluctuations of extinction, modeled via multiplicative cascades, were analyzed by Marshak and Davis (1995). As it turns out, the ICA model can be applied for sufficiently large pixels, whereas in the case of small pixels (i.e. pixel sizes less then 200–500 m) the ICA highly underestimates the liquid water variability in clouds.

A novel approach for 3D effect estimation and exclusion is proposed by Varnai and Marshak (2002). The method is based on satellite image analysis, that is to say, on the estimation of reflective cloudy brightness for visible and thermal infrared light. This technique makes it possible to decrease errors in the problem of cloud optical property reconstruction by means of ICA models, although the errors cannot be excluded completely. The necessity of 3D model introduction for processing reflected radiation fields, obtained in real measurements with high-resolution accuracy, has been also stressed by Marshak and Davis (1998).

Some final results of longstanding extended studies of 1D atmosphere models are presented in a detailed investigation by Barker et al. (2003). Here 1D and 3D model comparison was carried out in the frames of three test problems, typical for tropical atmosphere:

clear-sky – clouds and aerosol are absent, slab height is near 100 km.
CLOUD A – low cloudiness (a cloud is located between 3.5 and 4 km).
CLOUD B – high cloudiness (a cloud is located between 10.5 and 11 km).

The calculation accuracy of three parameters has been studied. These are

- $\alpha_p$ – top-of-atmosphere albedo,
- $\alpha_{\mathrm{atm}}$ – atmospheric absorption,
- $\alpha_{\mathrm{sfc}}$ – surface absorption.

3D calculations were performed by the Monte Carlo method by means of four different codes. Deviations in the obtained results were less than 2%.

Twenty-five codes carried out 1D calculations, the line-by-line model being realized in two of them. Obtained results essentially differ among themselves and strongly deviate from 3D results. In particular, 1D models underestimate the value $\alpha_{\mathrm{atm}}$ on the average by 15–25 $\mathrm{Wm}^{-2}$, independently of cloud presence, for the problem with the Sun at zenith. Errors in 1D calculations are usually about 10% and even more. An example of a top-of-atmosphere albedo calculation is presented in Fig. 8.22. These facts give the authors the opportunity to conclude that a new methodology of radiation field calculations in large-scale models is very necessary. New codes should be fast and operative.

Results obtained by Kokhanovsky (2003a, 2005) can be used in problems of algorithm development. A important case of 3D light scattering problems in a turbid medium layer is considered in these papers. It is shown that in the case of optically thick layer the statistical distributions for reflected, transmitted and absorbed radiation are related via simple analytical expressions with the statis-

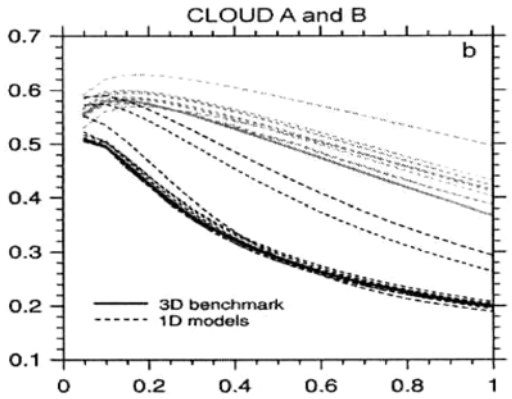



**Fig. 8.22.** Dashed lines represent broadband TOA albedos as a function of solar zenith angle cosines for the homogeneous CLOUD A (lighter lines) and CLOUD B (darker lines) predicted by all 1D codes assuming maximum/random overlap. Solid lines are corresponding values for one of the 3D MC codes (after Barker et al., 2003).

tical distribution of the optical thickness. These expressions may be applicable to both direct and inverse problem s of cloud optics.

### 8.14.2 Spherical atmosphere models

The 3D spherical coordinate system exactly corresponds to the terrestrial atmosphere geometry (see Fig. 8.23).

The transport equation in the spherical geometry is solved mainly by a Monte Carlo method. However, it is not always possible to use MC methods for atmospheric radiation field calculations (Ougolnikov, 1999) – because of large optical thickness of cloudy atmosphere. From this and other works we can see that grid spherical atmospheric models should be developed.

Currently algorithms, based on mesh schemes of the discrete ordinates method, are included in a small number of codes only. They use the charac-

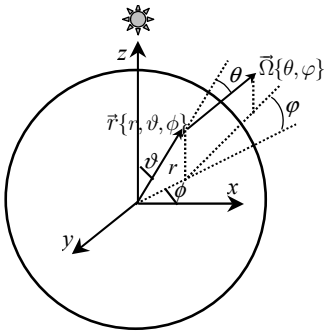



**Fig. 8.23.** 3D spherical coordinate system.

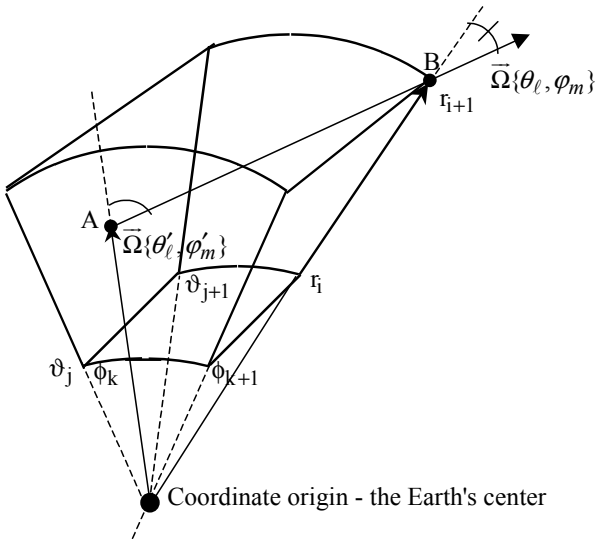



**Fig. 8.24.** Example of characteristics intersection scheme with a grid mesh in spherical coordinate system.

teristics method, where the transport equation is written in the integral form (see (8.18)). Integration in (8.18) is carried out along characteristics of the differential operator $\hat{L}$ (see Eq. (8.2)). An example of characteristics propagation through a curvilinear mesh in 3D spherical geometry is shown in Fig. 8.24, where

- AB – characteristics segment inside the mesh,
- $r_i$ – grid node radii,
- $\vartheta_j$ – grid node latitudes,
- $\phi_k$ – grid node longitudes,
- $\vec{\Omega}\,(\theta_\ell, \varphi_m)$ – grid node angles of viewing.

Figure 8.24 demonstrates the main difficulty arising in solving the transport equation in spherical geometry: its own local coordinate system for viewing angles $\{\theta, \varphi\}$ is being used at each spatial point $\{r, \vartheta, \phi\}$, and so viewing angles in a local coordinate system are changing, when the viewing point is moved along characteristics (see points B and A).

Because the right-hand side of transport equation (8.1) depends on $\psi(\vec{r}, \vec{\Omega})$ (that is, on the solution sought for), the iterative process on successive scatterings is usually applied to calculate $\psi(\vec{r}, \vec{\Omega})$.

An example of the algorithm of the transport equation being solved in 3D spherical geometry is presented in the work of Emde (2005). It is given in detail the next section.

So-called 'limb' observations play an important role in atmospheric investigations. Their scheme is presented in Fig. 8.25, where $\theta_T$ and $\phi_T$ are zenith and azimuth solar angles respectively (Griffioen and Oikarinen, 2000). Now these

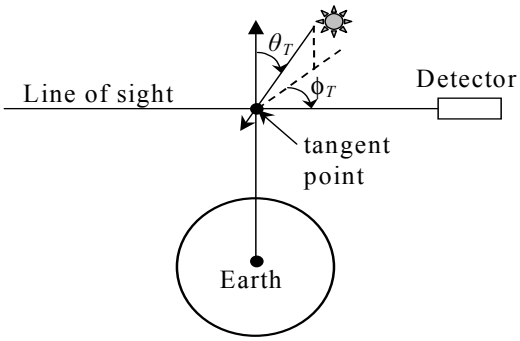



**Fig. 8.25.** 'Limb' observation scheme.

observations are performed by means of satellites for different spectral ranges. They permit

- detection of aerosol layers in the atmosphere,
- measurement of the temperature of the stratosphere, ozone concentration in the mesosphere and concentration of $NO_2$ in the stratosphere.

In the codes, which are intended for 'limb' observation treatments, the simplified combined models are used, because much time is required to solve the transport equation in 3D geometry.

These combined models are usually pseudo-spherical, because the scattering integral, defining multi-scattered radiation intensity, is calculated in the frames of the plane-parallel model. The examples are:

(a) model LIDORT (Linearized Discrete Ordinate Radiative Transfer) (Spurr, 2002);
(b) model LIMBTRAN (Griffioen and Oikarinen, 2000);
(c) model CDI (Combined Differential-Integral) (Rozanov et al., 2001), where light refraction is taking into account.

Besides, special spherical models are developed. Among them there is the model CDIPI (CDI Picard Iteration), based on characteristics method in combination with successive scattering order iterations. Here the initial approximation is calculated by means of the pseudo-spherical model CDI of Rozanov et al. (2000), refraction effects being taken into account.

The spherical model GSS (Gauss–Seidel Spherical) is intended for radiation intensity calculation in fixed desired radial directions (Herman et al., 1994). For the purpose a cone with a vertex in the Earth's center is constructed, the deviation angle between cone generatrix and the considered direction being equal to $\vartheta_0$ (see Fig. 8.26). A grid over cone height with nodes $R_0 + r_i$, $i = 1, 2, \ldots$, where $R_0$ is the Earth's radius, is introduced inside the cone. The transport equation solution inside the cone is calculated via the characteristics method in the frames of 1D spherical model. To determine the radiation intensity, entering the cone, it is assumed that the ratio of multiply scattered to unscattered light

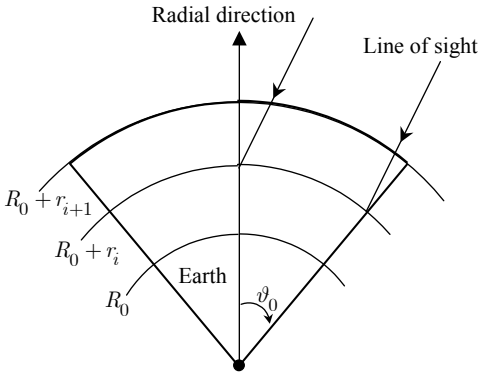



**Fig. 8.26.** Spatial grid and conical boundaries in the spherical model GSS.

intensity is constant over the atmospheric shell. A special iterative scheme has been constructed, taking into account both successive radiation scattering orders and boundary conditions. It provides a fast and accurate technique for finding the solution. Results of calculations obtained via both the MC method and other methods were used for model testing. The solutions obtained by the other grid methods were employed too. The interval of admissible angle $\vartheta_0$ variation was found. The angles are between 0.25° and 2°. Detailed comparison of reflected and transmitted light intensity for spherical and plane-parallel atmospheres for different solar and viewing angles has been carried out. Reflection from the Earth was not taken into account. Numerous illustrations were prepared for 50-km atmosphere, the Earth's radius being assumed to be equal to 6380 km. Figure 8.27 given by the authors clearly demonstrates the influence of the sphericity of the Earth's atmosphere.

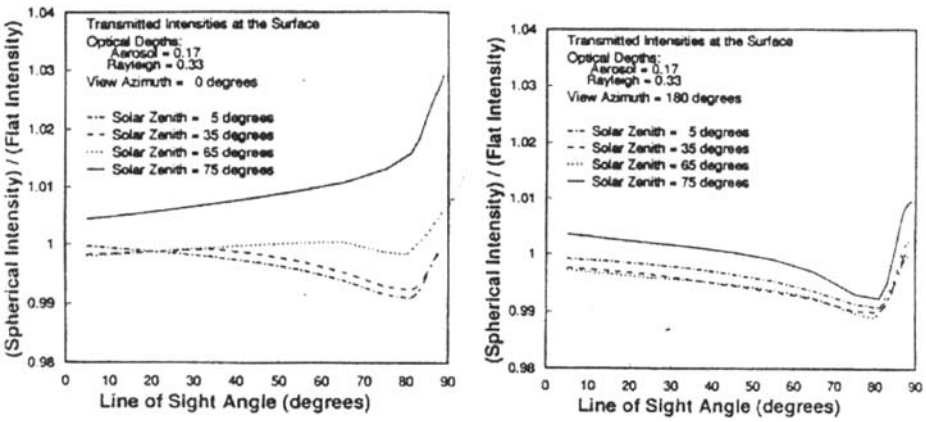



**Fig. 8.27.** Ratio of spherical- to flat-atmosphere transmitted intensities at the surface for a scattering atmosphere with a 0.50 optical depth. Left-view azimuth = 0 degrees, right −180 degrees (after Herman et al., 1994).

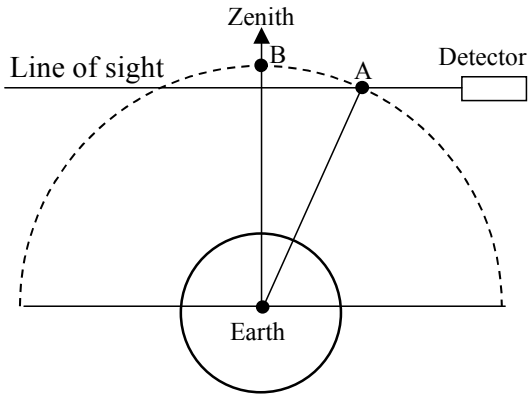



**Fig. 8.28.** Spherical model GSLS.

It should be noted that iterations over boundary conditions, being taken into account in the model, can be more easily performed both in $(x, y)$ and in $(x, y, z)$ geometries (see section 8.14). The boundary condition inclusion enables us to decrease the calculation time. However, a theoretical base for this inclusion is so far absent.

The spherical GSLS (Gauss–Seidel Limb Spherical) model of (Loughman et al., 2004) represents a more accurate and faster model than the GSS model (Herman et al., 1994, 1995). In the frames of the GSLS model, scattered radiation intensity in the line of sight is assumed to be equal to scattered radiation intensity at the point, which is characterized by the same coordinate $r$, but is located at the zenith radial direction. For example, the intensity value at the point A is equated to the intensity value at the point B (see Fig. 8.28). In addition polarization effects are also taken into account in the frames of GSLS model.

Finally, note that in some special cases the transport equation solution in 1D spherical geometry can be defined as the solution of the same equation in plane-parallel geometry under special boundary conditions (El-Wakil et al., 2001; Yildiz, 2002). This enables the use of analytical methods, developed for the plane-parallel transport problems, to obtain special solution classes for transport problems with spherical geometry. It can be important when one deals with inverse transport problems.

The paper by Loughman et al. (2004), is devoted to the investigation and comparison of six atmospheric models. In two of the models the MC method was used, whereas four models were based on DOM. These are the CDI, GSLS, LIMBTRAN and CDIPI models (see above). The detailed numerical analysis of these models has been carried out, including:

– comparison of results for test problems, obtained via different models;
– comparison of computational speeds;
– estimation of the accuracy of various techniques.

A solution qualified as an exact one if the calculation results obtained via two different MC codes provided a difference in the results smaller than 1–2%.

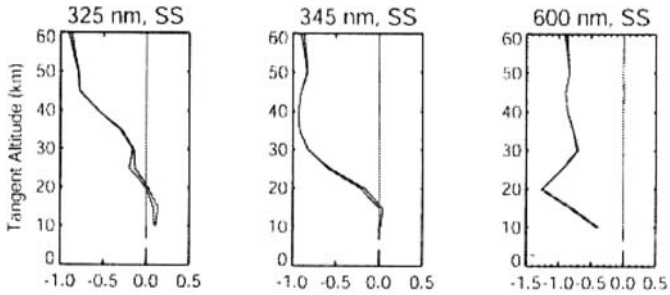



**Fig. 8.29.** Comparison between the vector RT models (left-hand two lines: GSLS and MC++ (Postylyakov, 2004)). Calculations were performed for an aerosol-free atmosphere, the underlying surface albedo equal to zero and for the same viewing geometry as in Figs. 8.21, 8.25. Right-hand vertical line corresponds to the calculations according to the scalar RT theory. The figures present percent differences for the exact SS (single scattering) source function from the case when the polarization is neglected (after Loughman et al., 2004).

As the comparison shows, the DOM codes mentioned provided low accuracy in situations where the optical path between calculation and detector points was great or when the zenith viewing angle changed significantly along the line of sight. The comparison of calculations according to different radiative transfer models is presented in Fig. 8.29.

### 8.14.3 DOM in problems with polarization

Light polarization arises mainly owing to single scattering of photons by water droplets, aerosol particles, ice crystals or air molecules. Multiple scattering processes lead to the decrease of the light polarization degree and increase the entropy of the radiation field (Kokhanovsky, 2003b; Hovenier and Domke, 2005; Mishchenko et al., 2006).

The retrieval of microscopic and macroscopic properties of clouds is based often on the measurements of the polarization characteristics of the scattered light. Hence, in many cases polarization effects should be taken into account in atmospheric optics problems. In this section we provide a brief description of some models on polarized radiation transfer that rely on the Vector Radiation Transport Equation (VRTE).

In these models four unknown functions, forming the so-called Stokes vector $S(I, Q, U, V)$, contain full information of light beam intensity, degree of polarization and polarization form.

1. In a paper of Rozanov and Kokhanovsky (2006) the algorithms of the VRTE solution based on the discrete-ordinates technique (DOT) (Siewert, 2000) is presented. A new code SCIAPOL_1.0 for the solution of the VRTE for the Stokes vector in a plane-parallel turbid slab, illuminated by the monodirected wide beam, is described. The phase matrix is presented in the form of a decomposition into series on associated Legendre functions.

The code can be applied to the solving of direct and inverse problem s of atmospheric optics, including analysis of the applicability of scalar approximation in complicated light transfer problems in media containing molecular atmosphere, aerosols and clouds. Some new approximations for the calculation of reflection function and light polarization degree under unpolarized light illumination conditions are proposed. Code capabilities have been demonstrated in a number of real physical problems. In particular, comparison of scalar and vector intensity approaches has been carried out.

2. The DOIT (Discrete Ordinate Iterative) algorithm of the polarized radiation field calculation in 1D and 3D spherical geometries is presented in the work of Emde (2005). It relies on the numerical solution of the transport equation for Stokes parameters by the characteristics method, the iterative successive-orders-of-scattering method being applied.

The main equations of the algorithm provide solution values corresponding to grid meshes so that the solution values can be immediately compared with satellite measurements in 'limb' geometry (zenith angle $\theta_{\ell\mathrm{imb}}$), in geometry 'down viewing' (angle $\theta_{\mathrm{down}}$) and in geometry 'up viewing' (angle $\theta_{\mathrm{up}}$) (see Fig. 8.30). Note, that scattered light intensity is much larger in the 'limb' geometry, than in the geometry 'down viewing' due to a larger optical path along the line of sight. So, the 'limb' measurements are more useful from the viewpoint of cloud optical property reconstruction.

The influence of particle size and orientation on light polarization was studied in the framework of these models. It was established that for clouds with horizontally oriented particles the polarization accounting significantly changes light intensity. Therefore, in many cases polarization effects should be taken into account, even if it is necessary to calculate only light intensity. The issue of calculation time is of importance, as usual. The time greatly increases as cloud size increases. For example, calculation of a thin cloud via a 3D model for the four-component Stokes vector with the help of a 3 Ghz Pentium 4 takes 50 min. The same calculation of the two-component Stokes vector takes 37 min. Complete calculation of a similar thick cloud takes 150 min.

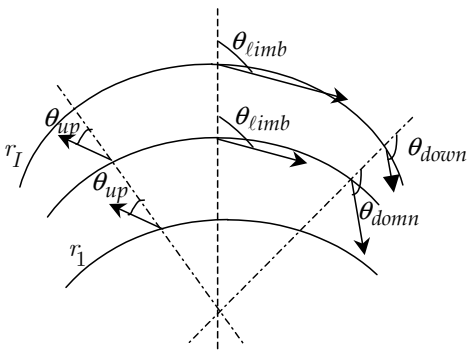



**Fig. 8.30.** Observation schemes.

The calculation via a 1D model requires less time. Full calculation of the thick cloud via a 1D model requires 30 sec. Therefore, 3D models cannot be so far used in operational or 'line-by-line' calculations.

3. The linearized MCC++ spherically symmetric 2D model is presented in the paper of Postylyakov (2004). It is intended for radiation field calculations at a large number of wavelengths by the Monte Carlo method. The technique is based on calculation of radiative functionals for radiation transfer problems with a single wavelength (that is the linearization point for the transport equation). These functionals and their derivatives are necessary for the solving of atmosphere optics inverse problem s. The MCC++ model takes into account aerosol and molecular scattering, gaseous and aerosol absorption and radiation reflection by a boundary surface according to the Lambert law.

## 8.15 Conclusion

Radiation propagation through a horizontally and vertically inhomogeneous atmosphere is usually described by the transport equation. To solve this equation either statistical algorithms (Monte Carlo), or grid methods, in particular the discrete ordinates methods, are used.

The Monte Carlo methods rely on direct modelling of random photon paths. They allow us to take into account the arbitrarily complicated structure of a calculation region. But they require large time expenditure for transport problems with optically thick media, where great numbers of trajectories should be calculated to guarantee high accuracy of results.

The chapter is aimed at the description of the discrete ordinates method, intended for multi-dimensional transport equation solving. The representation of the transport equation coefficients, grid constructions and grid scheme developments are considered in detail. In addition, long characteristics schemes that are frequently used for solving of atmospheric optics problems, and the other types of schemes, that are currently successfully exploited both in model radiation transfer problems and in applications, are also considered and discussed. Particular attention is focused on the accuracy of the schemes and their capabilities in the calculation of smooth solutions in homogeneous media and quickly varied solutions in strongly heterogeneous regions. In the latter case grid solutions can be distorted by non-physical oscillations of large amplitude, if only special approaches are not applied to reduce or eliminate the oscillations.

Different iterative algorithms for solving grid equations are considered. The popular successive-orders-of-scattering method converges fairly slowly (especially in the case of optically thick and weakly absorbing media). So, the main types of convergence acceleration method are described.

Methods of grid algorithm parallelizing are presented as well, and the conditions of their effectiveness are considered.

The main characteristics of two codes for solving 3D radiation transport problems in the atmosphere by the discrete ordinates method are described. These are SHDOM and RADUGA-5.1(P).

Simplified models are also presented, which permit the reducing of the scope of calculations for radiation transport problems that account for the 'limb' model of the atmosphere. The features of the simplified ICA model, which neglects horizontal radiation transfer, are discussed. In particular, the issue of the accuracy of medium optical characteristics retrieval via using a radiation transport model, neglecting horizontal radiation transfer, is also discussed.

The described grid methods, which were previously used mainly in different neutron transport problems, will be also useful in light scattering media optics, especially if the radiation field in turbid media having complex geometrical shapes is of interest.

## References

Adams, M.L., 1991: A new transport discretization scheme for arbitrary spatial mesh in (x, y) geometry, *Proc. of Int. Top. Meeting Advances Mathematics, Computations and Reactor Physics, Pittsburg, USA, April 28 – May 2, 1991*, Vol. 3, 13.2.2–1 – 13.2.2–9.

Adams, M.L., Larsen, E.W., 2002: Fast iterative methods for discrete-ordinates particle transport calculations, *Progress in Nuclear Energy*, **40**, 3–159.

Aussourd, Ch., 2003: A TN adaptive ray effect mitigation for styx 3D, *Proc. of Joint Int. Conf. on Supercomputing in Nuclear Applications, Paris, France, September 22–24*, 2003, CD.

Averin, A.V., Voloschenko, A.M., et al., 1991: The ROZ–6.4 one–dimensional discrete ordinates neutrons, gamma–rays and charged particles transport code, *Proc. of Int. Top. Meeting Advances Mathematics, Computations and Reactor Physics, Pittsburg, USA, April 28 – May 2, 1991*, Vol. 5, 30.3.5–1 – 30.3.5–4.

Azmy, Y.Y., 1988a: The weighted diamond-difference form of nodal transport methods, *Nucl. Sci. Eng.*, **98**, 29–40.

Azmy, Y.Y., 1988b: Multidimensional nodal transport methods for multi-instruction and multi-data distributed memory machine, *Trans. Am. Nucl. Soc.*, **56**, 292.

Azmy, Y.Y., 1992: Arbitrary high order characteristic method for solving the neutron transport equation, *Ann. Nucl. Energy*, **19**, 593–606.

Badruzzaman, A., 1985: An efficient algorithm for nodal transport solution in multidimensional geometry, *Nucl. Sci. Eng.*, **89**, 281–290.

Baker, R.S., Asano, C.A., Shirley, D.N., 1995: Implementation of the first order form of the 3D discrete ordinate equations on T3D, *Trans. Am. Nucl. Soc.*, **73**, 170.

Barbucci, P., Pasquantonio D.Di, 1977: Exponential supplementary equations for Sn methods: The one–dimensional case, *Nucl. Sci. Eng.*, **63**, 179–187.

Barker, H.W. et al., 2003: Assessing 1D atmospheric solar radiative transfer models: Interpretation and handling of unresolved clouds, *J. Climate*, **16**, 2676-2699.

Bass, L.P., Voloschenko, A.M., Germogenova, T.A., 1986: *Discrete ordinates techniques in problems of radiative transfer*, Keldysh Institute of Applied Mathematics, Moscow (in Russian).

Bass, L.P., Nikolaeva, O.V., 1997: Correct calculation of angular flux distributions in strongly heterogeneous media and voids, *Proc. of Joint Int. Conf. on Mathematical Methods and Supercomputing for Nuclear Applications, Saratoga Springs, October 5–9, 1997*, Vol. 2, 995–1004.

Cahalan, R.F., Oreopoulos, L., et al., 2005: I3RC. Bringing together the most advanced radiative transfer tools for cloudy atmospheres, *Bull. Amer. Meteor. Soc.*, September 2005, 1275–1293.

Carlson, B.G., 1976: A method of characteristics and other improvements in solutions methods for the transport equations, *Nucl. Sci. Eng.*, **61**, 408–425.

Castriani, Ch.L., Adams, M.L., 1995: A nonlinear corner balance spatial discretization for transport on arbitrary grids, *Proc. of Int. Conf. Mathematics and Computations, Reactor Physics, and Environmental Analyses, Portland, USA, April 30 – May 4, 1995*, Vol. 2, 916–927.

Chandrasekhar, S., 1950: *Radiative transfer*, Oxford University Press, Oxford.

Dahmani, M., Chavez-Guzman, J., Roy, R., Koclas, J., 2003: Data management in parallel transport calculation, *Proc. of Joint Int. Conf. on Supercomputing in Nuclear Applications, Paris, France, September, 22–24*, 2003, CD.

Dorr, M.R., Salo, E.M., 1995: Performance of a neutron transport code with full phase space decomposition on the Cray research T3D, *Proc. of Int. Conf. Mathematics and Computations, Reactor Physics, and Environmental Analyses, Portland, USA, April 30 – May 4, 1995*, Vol. 2, 1535–1544.

Elsawi, M.A., Abdurrahman, N.M., Azmy, Y.Y., 2003: Asymptotic analysis of the spatial weights of the arbitrarily high order transport method of the characteristic type, *Proc. of Int. Conf. Nuclear Mathematical and Computational Sciences, Tennessee, April 6–11*, 2003, CD.

El-Wakil, S.A., et al., 2001: Radiative transfer in a spherical medium, *J. Quant. Spectrosc. Radiat. Transfer*, **69**, 49-59.

Emde, C., 2005: A polarized discrete ordinate scattering model for radiative transfer simulations in spherical atmospheres with thermal source. PhD thesis, University of Bremen, Bremen.

Evans, K.F., 1998: The spherical harmonics discrete ordinate method for three–dimensional atmospheric radiative transfer, *J. Atmos. Sci.*, **55**, 429–446.

Fischer, J.W., Azmy, Y.Y., 2003: Parallel performance of the angular versus spatial domain decomposition for discrete ordinate method, *Proc. of Joint Int. Conf. on Supercomputing in Nuclear Applications, Paris, France, September, 22–24*, 2003, CD.

Germogenova, T.A., 1985: *Local properties of transport equation solutions*, Nauka, Moscow (in Russian).

Giuseppe, F.Di., Thompkins, A.M., 2003: Three-dimensional radiative transfer in tropical deep convective clouds, *J. Geophys. Res.*, **108**, 10.1029/2003JD003392.

Griffioen, E., Oikarinen, L., 2000: LIMBTRAN: A pseudo three-dimensional radiative transfer model for the limb-viewing imager OSIRIS on the ODIN satellite, *J. Geophys. Res.*, **105**, 29,717 – 29,730.

Grove R.E., Pevey R.E., 1995: A characterictics based multiple balance approach for SN on arbitrary polygonal grids, *Proc. of Int. Conf. Mathematics and Computations, Reactor Physics, and Environmental Analyses, Portland, USA, April 30 – May 4, 1995*, 928–937.

Haghighat, A., Azmy, Y.Y., 1991: Parallelization of a spherical Sn algorithm based on the spatial domain decomposition, *Proc. of Int. Top. Meeting Advances Mathematics, Computations and Reactor Physics, Pittsburg, USA, April 28 – May 2, 1991*, 1.1.3.-1. – 1.1.3.-12.

Hanebutte, U.R., Lewis, E.E., 1991: A discrete ordinate response matrix method for massively parallel computers, *Proc. of Int. Top. Meeting Advances Mathematics, Computations and Reactor Physics, Pittsburg, USA, April 28 – May 2, 1991*, 1.1.1-1. – 1.1.1.-10.

Hennart, L.P., del Valle, E., 1997: New nodal element schemes for the discrete ordinates transport equation, *Proc. of Joint Int. Conf. on Mathematical Methods and Supercomputing for Nuclear Applications, Saratoga Springs, October 5–9, 1997*, Vol. 1, 19–28.

Herman, B.M. et al., 1994: Numerical techniques for solving the radiative transfer equation for a spherical shell atmosphere, *Appl. Opt.*, **33**, 1760-1770.

Herman, B.M. et al., 1995: Comparison of the Gauss-Seidel spherical polarized radiative transfer code with other radiative transfer codes, *Appl. Opt.*, **34**, 4563–4572.

Hovenier, J., Domke, H., 2005: *Transfer of polarized light in planetary atmospheres. Basic concept and practical methods*, Springer, Berlin.

Koch, R., Krebs, W., Wittig, S., Viskanta, R., 1995: Discrete ordinates quadrature schemes for multidimensional radiative transfer, *J. Quant. Spectrosc. Radiat. Transfer*, **53**, 353–372.

Kokhanovsky, A.A., 2003a: The influence of horizontal inhomogeneity on radiative characteristics of clouds: an asymptotic case study, *IEEE Trans. Geosci. Remote Sensing.*, **41**, 817–825.

Kokhanovsky, A.A., 2003b: *Polarization optics of random media*, Springer, Berlin.

Kokhanovsky, A.A., 2005: Statistical properties of light reflected and transmitted by a thick horizontally inhomogeneous turbid layer, *J. Opt. Soc. Am. A.*, **22**, 2419–2423.

Kokhanovsky, A.A., 2006: *Cloud Optics*, Springer, Dordrecht.

Landesman, M., Morel, J.E., 1989: Angular Fokker–Planck decomposition and representation techniques, *Nucl. Sci. Eng.*, **103**, 1–11.

Lathrop, K.D., 1969: Spatial differencing of the transport equation: positivity vs. accuracy, *J. Comp. Phys.*, **4**, 475–490.

Lebedev, V.I., 1976. On quadratures on spheres. *Computational Mathematics and Mathematical Physics*, **16**, 293-307.

Longoni, G., Haghighat, A., 2001: Development of new quadrature sets with the 'ordinate splitting' technique, *Proc. of Int. Meeting on Mathematical Methods For Nuclear Application, Salt Lake City, Utah, September, 2001*, CD.

Loughman, R.P., Griffioen, E., Oikarinen, L., Postylyakov, O.V., Rozanov, A., Flittner, D.E., Rault, D.F., 2004: Comparison of radiative transfer models for limb-viewing scattering sunlight measurements, *J. Geophys. Res.*, **109**, 10.1029/2003JD003854.

Madsen, N.K., 1975: Spatial convergence properties of the diamond difference method in $x$, $y$ geometry, *Nucl. Sci. Eng.*, **12**, 164–176.

Marchuk, G.I., et al.,: 1980: *The Monte Carlo methods in atmospheric optics*, Springer, Berlin.

Marchuk, G.I., Lebedev, V.I., 1981: *Numerical methods in neutron transport theory*, Atomizdat, Moscow (in Russian).

Marshak, A., Davis, A., et al., 1995: Radiative smoothing in fractal clouds. *J. Geophys. Res.*, **100**, 26,247 – 26,261.

Marshak, A., Davis, A., et al., 1998: Radiative effects of sub-mean-free-path liquid water variability observed in stratiform clouds, *J. Geophys. Res.*, **103**, 19557-19567.

Marshak, A., Davis, A., 2005: *Three-dimensional radiative transfer in cloudy atmospheres, Springer*, Berlin.

Mathews, K.A., Minor, B.M., 1991: Step adaptive characteristic spatial quadrature in two-dimensional Cartesian coordinates, *Proc. of Int. Top. Meeting Advances Mathematics, Computations and Reactor Physics, Pittsburg, USA, April 28 – May 2, 1991*, Vol. 3, 13.2.4–1 – 13.2.4–12.

Mathews, K.A., Minor, B.M., 1993: Adaptive characteristic spatial quadratures for discrete ordinates neutral particle transport – the rectangular cell case, *Transport Theory Stat. Phys.*, **22**, 655–685.

Mathews, K.A., 1999: On the propagation of rays in discrete ordinates, *Nucl. Sci. Eng.*, **132**, 155–180.

Mayer, B., 1999: I3RC phase 1 results from the MYSTIC Monte Carlo model, *Proc. of I3RC Workshop, Tucson, Arizona, November 17–19, 1999*.

Mishchenko, M.I., Travis L.D., Lacis, A.A., 2006: *Multiple scattering of light by particles: radiative transfer and coherent backscattering*, Cambridge University Press, Cambridge.

Morel, J.E., Larsen, E.W., 1990: A multiple balance approach for differencing the Sn equations, *Nucl. Sci. Eng.*, **105**, 1–15.

Morel, J.E., Manteuffel, T.A., 1991: An angular multigrid acceleration technique for Sn equations with highly forward-peaked scattering, *Nucl. Sci. Eng.*, **107**, 330–342.

Morel, J.E., Dendy Jr, J.E., Wareing, T.A., 1993: Diffusion–accelerated solution of the two–dimensional Sn equations with bilinear-discontinuous difference, *Nucl. Sci. Eng.*, **115**, 304–319.

Morel, J.E., 2005: Basic Krylov methods with application to transport, *Proc. of Int. Conf. on Mathematics and Computation, Supercomptuting, Reactor Physics and Nuclear Biological Applications, Avignon, France, September 12–15, 2005*, CD.

Nikolaeva, O.V., 2004: Special grid approximations for the transport equation in strongly heterogeneous media with the $(x, y)$-geometry, *Computational Mathematics and Mathematical Physics*, **44**, 835–854.

Nikolaeva, O.V., Bass, L.P., Germogenova, T.A., Kokhanovsky, A.A., Kuznetsov, V.S., Mayer, B., 2005a: The influence of neighbouring clouds on the clear sky reflectance studied with the 3-D transport code RADUGA, *J. Quant. Spectrosc. Radiat. Transfer*, **94**, 405-424.

Nikolaeva, O.V., Germogenova, T.A., Bass, L.P., Kuznetsov, V.S., 2005b: The discrete ordinates methods development to the transport equation solving. The 3D code Raduga-5.1 and multiprocessors computers, *Proc. of 19th International Conference on Transport Theory, Budapest, Hungary, July 24–29, 2005*, 115–117.

Nowak, P.F., Nemanic, M.K., 1999: Radiation transport calculation on unstructured grids using spatially decomposed and threaded algorithm, *Proc. of Int. Conf. Mathematics and Computation, Reactor Physics and Environmental Analysis in Nuclear Application, Madrid, Spain, September 27–30, 1999*, Vol. 1, 379–390.

Ougolnikov, O.S., 1999: Twilight sky photometry and polarimetry: The problem of multiple scattering at the twilight time, *Cosmic Res.*, **37**, 159–165.

Patchimpattapong, A., Haghighat, A., 2003: Testing an expert system for selection of mesh and decomposition on parallel Sn method, *Proc. of Int. Conf. Nuclear Mathematical and Computational Sciences, Tennessee, April 6–11, 2003*, CD.

Pautz, S.D., Morel, J.E., 1999: An angular multigrid acceleration method for SN equations with highly forward-peaked scattering, *Proc. of Int. Conf. Mathematics and Computation, Reactor Physics and Environmental Analysis in Nuclear Application, Madrid, Spain, September 27–30, 1999*, Vol. 1, 647–656.

Pautz, S.D., 2001: An algorithm for parallel Sn sweep on unstructured mesh, *Proc. of Int. Meeting on Mathematical Methods For Nuclear Application, Salt Lake City, Utah, September, 2001*, CD.

Petrovic, B., Haghighat, A., 1996: New directional theta–weighed SN differencing scheme and its application to pressure vessel fluence calculations, *Proc. of*

*Radiation Protection and Shielding Topical Meeting, Falmouth, MA*, Vol. 1, 3–10.

Postma, T., Viujic, J., 1999: The method of characteristics in general geometry with anisotropic scattering, *Proc. of Int. Conf. Mathematics and Computation, Reactor Physics and Environmental Analysis in Nuclear Application, Madrid, Spain, September 27–30, 1999*, Vol. 2, 1215–1224.

Postylyakov, O.V., 2004: Linearized vector radiative transfer model MCC++ for spherical atmosphere, *J. Quant. Spectrosc. Radiat. Transfer*, **88**, 297–317.

Rhoades, W.A., Engle, W.W., 1977: A new weighted-diamond formulation for discrete ordinates calculations, *Trans. Am. Nucl. Soc.*, **27**, 776–777.

Rozanov, A., Rozanov, V., Burrows, J.P., 2000: Combined differential-integral approach for the radiation field computation in a spherical shell atmosphere: non-limb geometry, *J. Geophys. Res.*, **105**, 22937–22942.

Rozanov, A., Rozanov, V., Burrows, J.P., 2001: A numerical radiative transfer model for a spherical planetary atmosphere: combined differential-integral approach involving the Picard iterative approximation, *J. Quant. Spectrosc. Radiat. Transfer*, **69**, 491–512.

Rozanov V.V., Kokhanovsky, A.A., 2006: The solution of the vector radiative transfer equation using the discrete ordinates technique: Selected applications, *Atm. Res.*, **79**, 241–265.

Ryzik, I.M., Gradstein, I.S., 1972: *Tables of integrals, series and products*, Academy Press, New York.

Samarsky, A.A., 1989: *Theory of difference schemes*. Nauka, Moscow (in Russian).

Sanchez, R., McCormic, N.J., 2004: Discrete ordinates solutions for highly forward peaked scattering, *Nucl. Sci. Eng.*, **147**, 249–274.

Santandrea, S., Sanchez, R., 2002: Positive linear and nonlinear surface characteristic scheme for the neutron transport equation in unstructured geometries, *Proc. of Physor-2002*, CD.

Scheirer, R., Macke, A., 2003: Cloud inhomogeneity and broadband solar fluxes, *J. Geophys. Res.*, **108**, 1029/2002JD003321.

Shwetsov, A.V., 1997: Solving the x–y geometry transport equation by linear discontinuous scheme with consistent flux correction, *Proc. of Joint Int. Conf. on Mathematical Methods and Supercomputing for Nuclear Applications, Saratoga Springs, October 5–9, 1997*, Vol. 2, 1487–1496.

Sjoden, G.E., Haghighat, A., 1997: Pentran-A-3D cartesian parallel Sn code with angular, energy, and spatial decomposition, *Proc. of Joint Int. Conf. on Mathematical Methods and Supercomputing for Nuclear Applications, Saratoga Springs, October 5–9, 1997*, Vol. 1, 553-562.

Siewert, C.E., 2000: A discrete-ordinates solution for radiative-transfer models that include polarization effects, *J. Quant. Spectrosc. Radiat. Transfer*, **64**, 227–254.

Spurr, R.J.D., 2002: Simultaneous derivation of intensities and weighting functions in a general pseudo-spherical discrete ordinate radiative transfer treatment, *J. Quant. Spectrosc. Radiat. Transfer*, **75**, 129–175.

Sushkevich, T.A., Strelkov, S.A., Ioltuhovsky, A.A., 1990: *Characteristic method in atmospheric optics*, Nauka Publishers, Moscow (in Russian).

Sushkevich, T.A., 2005: *Mathematical models of radiation transfer*, Moscow (in Russian).

Suslov, I.R., 1988: Characteristics method in region with complicated geometry, *Atomic energy*, **65**, 57–58 (in Russian).

Suslov, I.R., Pevey, R.E., 1997: Application of the quasi-stationary derivatives principle in transport theory, *Trans. Am. Soc.*, **76**, 214.

Takeda, T., Yamamoto, T., 2001: Nodal transport methods for $XYZ$ and hexagonal-$Z$ geometry, *Transport Theory Stat. Phys.*, **30**, 401–420.

Titov, G.A., 1998: Radiative Horizontal transport and absorption in stratocumulus clouds, *J. Atmos. Sci.*, **55**, 2549–2560.

Ullo, J.J., Dorning, J.J., Dodds, H.L., Pevey, R.E., 1982: A comparison of nodal transport methods based on exponential and polynomial expansions, *Trans. Am. Nucl. Soc.*, **43**, 367–369.

del Valle, E., Alonso, G., 2001: Exponential weakly-discontinuous nodal scheme for the transport equation, *Trans. Am. Nucl. Soc.*, CD.

Varnai, T., Marshak, A., 2002: Observations of three-dimensional radiative effects that influence MODIS cloud optical thickness retrievals, *J. Atmos. Sci.*, **59**, 1607–1618.

Vladimirov, V.S., 1958: Numerical solution of kinetic equation for sphere, *Trans. 'Computing mathematics'*, **3**, 3–33 (in Russian).

Voloschenko, A.M., 1981: Transport equation solving by the DSn method in heterogeneous media. Part 2. One-dimensional spherical and cylindrical geometries, in collection *'Numerical solution of the transport equation for one-dimensional problems'*, Keldysh Institute of Applied Mathematics, 64–91 (in Russian).

Voloschenko, A.M., Germogenova, T.A., 1994: Numerical solution of the time-dependent transport equation with pulsed sources, *Transport Theory Stat. Phys.*, **23**, 845–869.

Voloschenko, A.M., 1997: Geometrical interpretation of family of weighted nodal schemes and adaptive positive approximations for transport equations, *Proc. of Joint Int. Conf. on Mathematical Methods and Supercomputing for Nuclear Applications, Saratoga Springs, New York, October 5–9, 1997*, Vol. 2, 1517–1526.

Voronkov, A.V., Sychugova, E.P., 1997: Analysis of a class of characteristic methods for solving the transport equation in x–y and x–y–z geometry, *Proc. of Joint Int. Conf. on Mathematical Methods and Supercomputing for Nuclear Applications, Saratoga Springs, New York, October 5–9, 1997*, Vol. 2, 985–994.

Walters, W.F., 1982: Recent development in nodal and characteristic method in transport theory, *Trans. Am. Nucl. Soc.*, **43**, 611–612.

Walters, W.F., 1986: Augmented weighted-diamond form of the linear-nodal scheme of Cartesian coordinate systems, *Nucl. Sci. Eng.*, **92**, 192–196.

Walters, W.F., Wareing, T.A., Marr, D.R., 1995: The nonlinear characteristic scheme for (x, y) geometry transport problems, *Proc. of Int. Conf. Mathematics and Computations, Reactor Physics, and Environmental Analyses, Portland, USA, April 30 – May 4, 1995*, 340–348.

Wareing, T.A., 1997: An exponential discontinuous scheme for discrete–ordinate calculation in Cartesian geometries, *Proc. of Joint Int. Conf. on Mathematical Methods and Supercomputing for Nuclear Applications, Saratoga Springs, October 5–9, 1997*, Vol. 2, 1257–1266.

Warin, X., 1996: Recent results about Sn nodal method in neutron transport. *Proc. of Int. Conf. 3-D Deterministic Radiation Transport Computer Programs: Features, Applications and Perspectives, Paris, France, December 2–3, 1996.* 325–337.

Wiscombe, W., 1977: The delta-M method: Rapid yet accurate radiative flux calculations for strongly asymmetric phase functions, *J. Atmos. Sci.*, **34**, 1408–1422.

Yildiz, C., 2002: The FN solution of the time-dependent neutron transport equation for a sphere with forward scattering, *J. Quant. Spectrosc. Radiat. Transfer*, **74**, 521–529.

Zhougsheng, Xie, Xuehua, Zhy, Honghun, Wu., 1994: Two–dimensional discrete nodal transport method for solving neutron transport problems in curvilinear coordinates, *Proc. of the Int. Conf. on Reactor Physics and Reactor Computations, Tel–Aviv, Israel, January 23–26, 1994*, Vol. 3, 688–694.

Zmijarevic, I., 1999: Multidimensional discrete ordinate nodal and characteristics methods for the Apollo2 code, *Proc. of Int. Conf. Mathematics and Computation, Reactor Physics and Environmental Analysis in Nuclear Application, Madrid, Spain, September 27–30, 1999*, Vol. 2, 1587–1597.

# Index

Printing: Mercedes-Druck, Berlin
Binding: Stein + Lehmann, Berlin