ENVIRONMENTAL SYSTEMS ANALYSIS

CENG 6652

Chapter 4

System Sensitivity and Uncertainty Analysis

Contents

1. Issues, Concerns, and Terminology

- The usefulness of any model is in part dependent on the accuracy and reliability of its **output data**.
- Output values are subject to imprecision because:
  - All models are abstractions of reality,
  - Precise input data are rarely available.
  - We simply cannot forecast the future with precision.
- The input data and modeling uncertainties can interact in various ways.
- This chapter focuses on ways of identifying, quantifying, and communicating the uncertainties in model outputs.

Issues, …

- Some input data uncertainties can be reduced by additional research and further data collection and analysis.
- Before spending money and time to gather and analyze additional data, it is reasonable to ask:
    - What improvement in estimates of system performance or what reduction in the uncertainty associated with those estimates would result, if all data and model uncertainties could be reduced (if not eliminated).
- Such information helps to determine how much one would be willing to "pay" to reduce model output uncertainty.
- **Obtaining additional information, however measured, should exceed the cost of obtaining it.**

Issues, …

- **Risk vs. Uncertainty**
  - The term risk is often reserved to describe situations for which probabilities are available to describe the likelihood of various possible events or outcomes. Often risk refers to these probabilities times the magnitude of the consequences of these events or outcomes.
  - If probabilities of various events or outcomes cannot be quantified, or if the events themselves are unpredictable, some would say the problem is then one of uncertainty.
- Uncertainty stems from:
  - Inadequate information
  - Incorrect assumptions, and
  - The variability of natural processes.

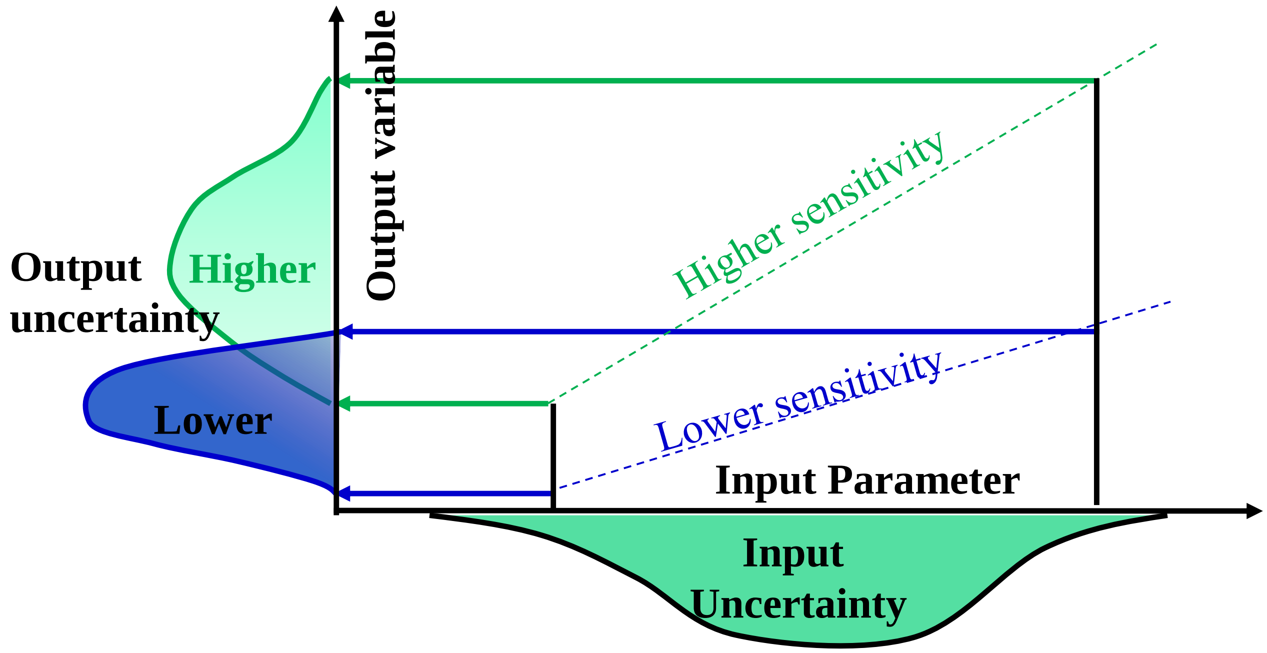
Simple approach to uncertainty

- Among ways to deal with uncertainty, the simplest approach is to replace each uncertain quantity either by its expected or average value or by some critical (e.g., “safest-case”) value and then proceed with a deterministic approach.
- Replacement of uncertain quantities by either expected or safest-case values can adversely affect the evaluation of project performance especially when important parameters are highly variable.
- When important quantities are uncertain, one should evaluate both the expected performance of a system and possible magnitude of system failures and their consequences.

Issues, …

- Errors and approximations in input data measurement, parameter values, model structure and model solution algorithms, are all sources of uncertainty.
- Uncertainty analyses employing probabilistic descriptions of model inputs can be used to derive probability distributions of model outputs and system performance indices.

Figure: the impact of both input data sensitivity and input data uncertainty on model output uncertainty.



Issues, …

- There is also **uncertainty with respect to human behavior** and reaction related to particular outcomes and their likelihoods, i.e., to their risks and uncertainties. These are not usually part of the models themselves.
- Social uncertainty may often be the most significant component of the total uncertainty associated with just how a water environment system will perform.
- Sensitivity analysis procedures explore and quantify the impact of possible changes (errors) in input data on predicted model outputs.
- Simple sensitivity analysis procedures can be used to illustrate the consequences of alternative assumptions about the future.

Issues, …

- An uncertainty analysis attempts to describe the entire set of possible outcomes, together with their associated probabilities of occurrence.
- A sensitivity analysis attempts to determine the relative change in model output values given modest changes in model input values. A sensitivity analysis thus measures the change in the model output in a localized region of the space of inputs.
- However, one can often use the same set of model runs for both uncertainty analyses and sensitivity analyses.
- It is possible to carry out a sensitivity analysis of the model around a current solution and then use it as part of a first-order uncertainty analysis.

2. Uncertainty Analyses (UA)

- Uncertainty involves the notion of randomness.
- If a value of a performance indicator or performance measure, or in fact **any variable** (like the chemical concentration or the depth of water at a particular location) varies and this variation over space and time **cannot be predicted with certainty, it is called a random variable**.
- One cannot say with certainty what the value of a random variable will be but only the likelihood or probability that it will be within some specified range of values.
- The probabilities of observing particular ranges of values of a random variable are described or defined by a probability distribution.
- Here we are assuming each student know, or can compute, or can estimate, this distribution.

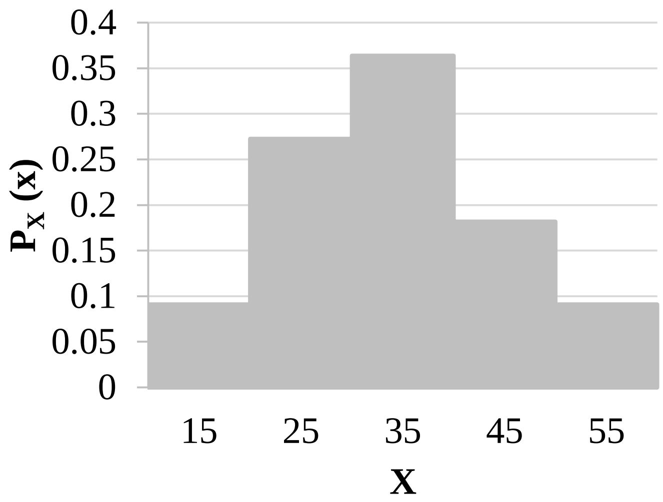
Probability distribution

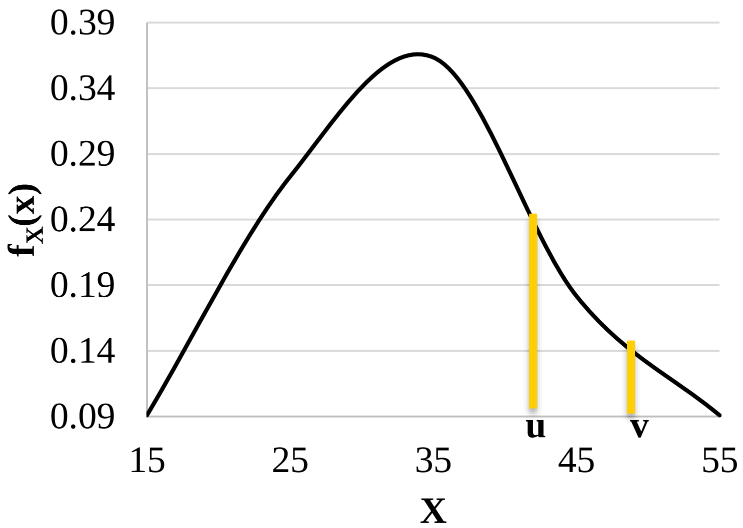
- A random variable x takes on a defined set of values with different probabilities.
  - For example, if you roll a die, the outcome is random (not fixed) and there are 6 possible outcomes, each of which occur with probability one-sixth.
- Roughly, probability is how frequently we expect different outcomes to occur if we repeat the experiment over and over (“frequentist” view)
- A probability function maps the possible values of x against their respective probabilities of occurrence, p(x)
- p(x) is a number from 0 to 1.0.
- The area under a probability function is always 1.

UA

- Suppose the random variable is X. For discrete values, the probability distribution of X can be expressed as a histogram, as shown below (see next slide).
- If the random variable is a continuous variable, the probability distribution of X can be expressed as a continuous distribution as shown as smooth line (see next slide).
- The shaded area under the density function for the continuous distribution is 1 as the sum of the probabilities for all possible outcomes must equal 1.
- The area between two values (u and v) of the random variable represents the probability that the observed value x of the random variable value X will be within that range of values.

UA

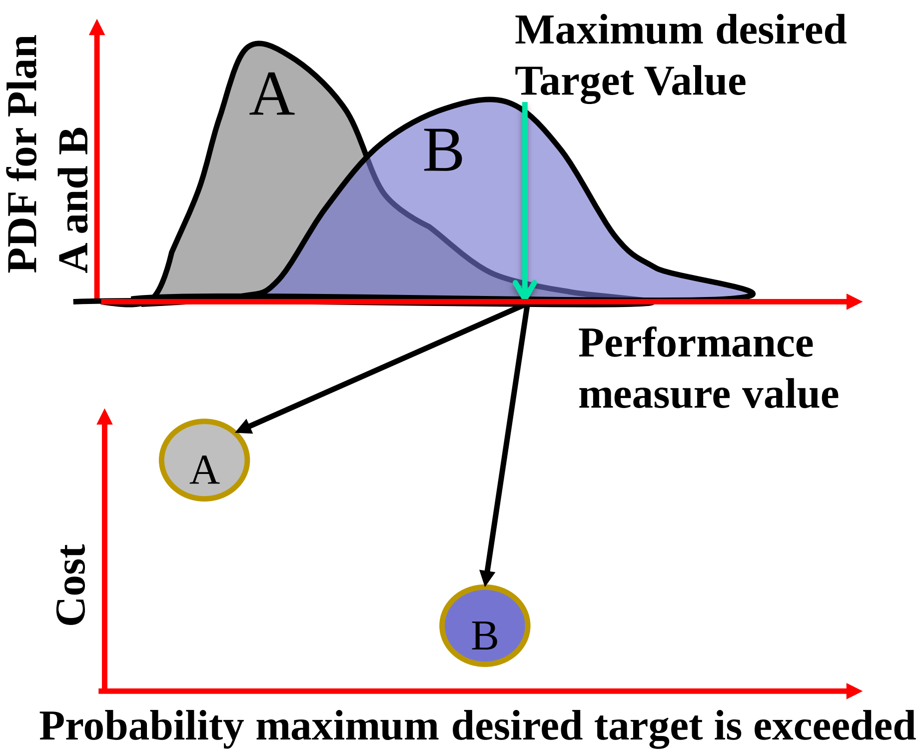




- The probability distribution, $P_X(x)$ shown in left is called a probability mass function.
- The probability distributions shown in the right, is called probability density functions (pdf) and are denoted by $f_X(x)$.
- The subscript X of P_X and f_X represents the random variable, and the variable x (on the horizontal axes) is some value of that random variable X.

UA

- UA involve a comprehensive identification of all sources of uncertainty that contribute to the joint Probability Distribution of each input or output variable.
- PDF for two alternative project plans, A and B, for a specified performance measure with the corresponding costs are shown.

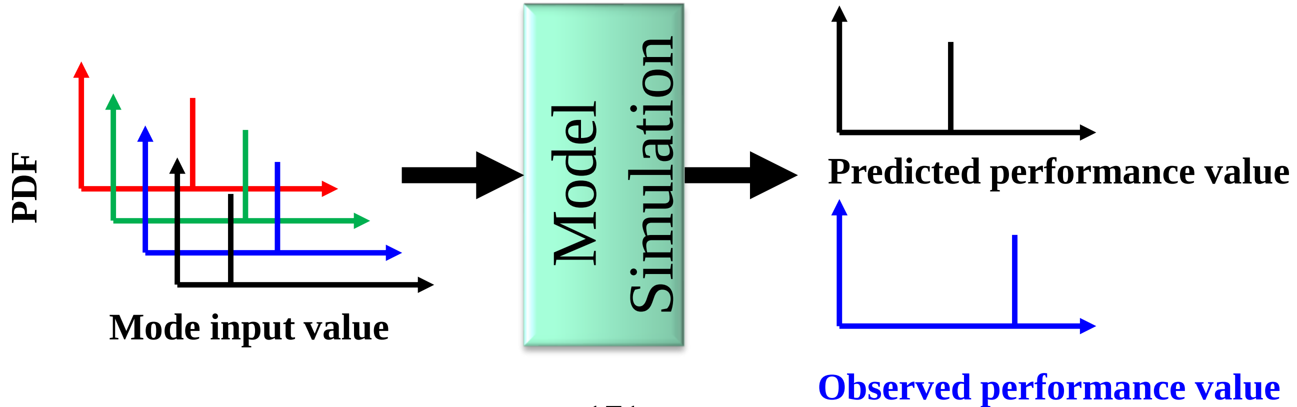


- The introduction of two performance criteria, cost and probability of exceeding a performance measure target (e.g., a polluant concentration standard) introduces a conflict where a tradeoff must be made.

In this illustration, we want the lowest cost (B is best) and the lowest probability of exceedance (A is best)

Model and Model Parameter Uncertainties

- Consider a situation, in which for a specific set of model inputs parameters (here 4), the model outputs differ from the observed values, and for those model inputs, the observed values are always the same. Here nothing randomly occurs.
  - The 4 input values or model structure needs to be changed.
- This is typically done in a model calibration process.
- The gap between the two (predicted and observed) values could result from imprecision in the measurement of observed data, imprecision in the model parameter values, the model structure, or the algorithm used to solve the model.

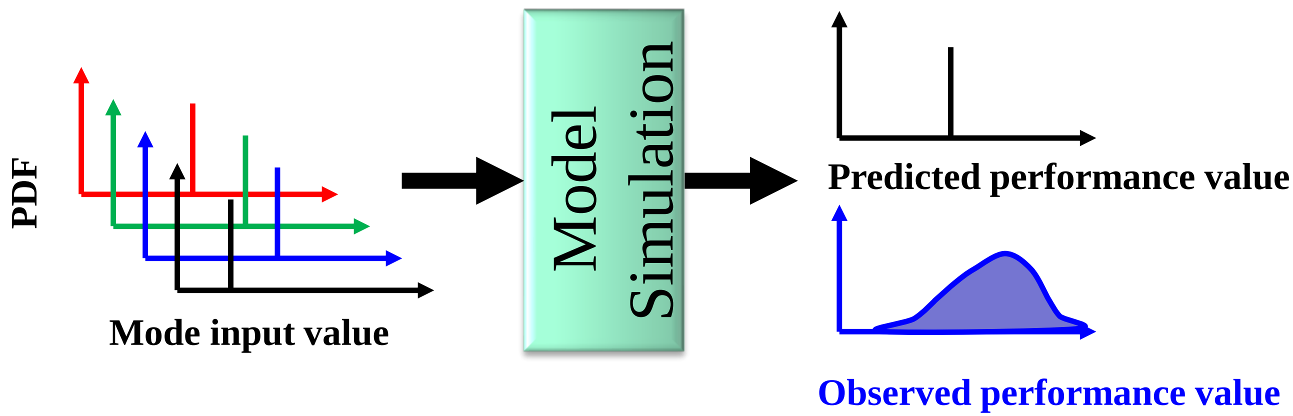


Model and Model Parameter Uncertainties

- If some of the inputs are random, i.e., not predictable, (e.g. when random outputs of one model are used as inputs into another model) they will yield random outputs. Here the model input and output values shall be described by PDFs.
- If the uncertainty in the output is due only to the uncertainty in the input, the situation is similar to that shown above slide.
- If the PDF of performance measure output values does not fit or is not identical to the PDF of observed performance measure values, then calibration of model parameter values or modification of model structure may be needed.
- If a model calibration or "identification" exercise finds the "best" values of the parameters to be outside reasonable ranges of values based on scientific knowledge, then the model structure or algorithm might be in error.

Model and Model Parameter Uncertainties

- Assuming the algorithms used to solve the models are correct and observed measurements of system performance vary for the same model inputs, as shown in Fig. below, it can be assumed that the model structure does not capture all the processes that are taking place and that impact the value of the performance measures.
- This is often the case when relatively simple and low-resolution models are used. However, even large and complex models can fail to include or adequately describe important phenomena.



Example

- Consider the prediction of a pollutant concentration at some site downstream of a pollutant discharge site.
- Given a streamflow Q (1000 m^3/day), the distance between the discharge site and the monitoring site, x (m), the pollutant decay rate constant k (day^{-1}), and the pollutant discharge w (kg/day),
- Use the following simplified model to predict the concentration of the pollutant C (g/m^3 = mg/l) at the downstream monitoring site $C = \frac{w}{Q} e^{-k\frac{x}{u}}$
- The velocity u (m/day) is a known function of the streamflow Q.
- Observed C may differ from the computed C even for the same inputs of *w*, *Q*, *k*, *x*, and *u*. This difference varies in different time periods. This apparent variability, is as illustrated in Figure upper slide.

Example

- This apparent variability, can be simulated using the same model but by assuming a distribution of values for the decay rate constant *k*.
- Alternatively the model structure can be modified to include the impact of streamflow temperature T on the prediction of C as $C = \frac{w}{Q} e^{-k\frac{x}{u}\theta^{T-20}}$
- Now there are two model parameters, k and the dimensionless temperature correction factor θ, and an additional model input, the streamflow temperature, T.
- It could be that the variation in streamflow temperature was the sole cause of the first equation's "uncertainty" and that the assumed parameter distribution of *k* was simply the result of the distribution of streamflow temperatures on the term $k\theta^{T-20}$.

Example

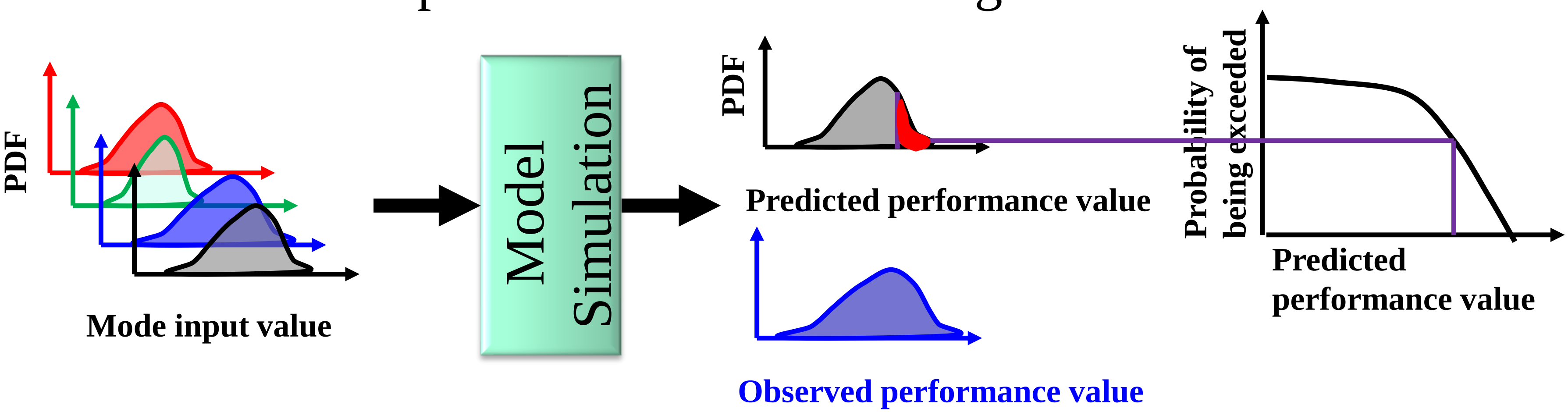
- If the output were still random given constant values of all the inputs, then another source of uncertainty exists.
- This uncertainty might be due to additional random loadings of the pollutant, possibly from nonpoint sources.
- Once again the model could be modified to include these additional loadings if they are knowable.
- If these additional loadings are not known, a new random parameter could be added to the input variable w or to the right hand side of the equations above that would attempt to capture the impact on C of these additional loadings.
- A potential problem, however, might be the likely correlation between those additional loadings and the streamflow Q.

Example

- NOTE
  - While adding model detail removed some “uncertainty” in the above example, increasing model complexity will not always eliminate or reduce uncertainty in model output.
  - Adding complexity is generally not a good idea when the increased complexity is based on processes whose parameters are difficult to measure, the right equations are not known at the scale of application, or the amount of data for calibration is small compared to the number of parameters.

What Uncertainty Analysis Can Provide

- An uncertainty analysis takes a set of randomly chosen input values (that can include parameter values), passes them through a model (or transfer function) to obtain the distributions (or statistical measures of the distributions) of the resulting outputs. The output distributions can be used to
  - Describe the range of potential outputs of the system at some probability level.
  - Estimate the probability that the output will exceed a specific threshold or performance measure target value.



What Uncertainty Analysis Can Provide

- Common uses for uncertainty analyses are to make general inferences, like:
  - Estimating the mean and standard deviation of the outputs.
  - Estimating the probability the performance measure will exceed a specific threshold.
  - Putting a reliability level on a function of the outputs, e.g., the range of function values that is likely to occur with some probability.
  - Describing the likelihood of different potential system outputs.
- Implicit in any uncertainty analysis are the assumptions that statistical distributions for the input values are correct and that the model is a sufficiently realistic description of the processes taking place in the system. Neither of these assumptions is likely to be entirely correct.

3. Sensitivity Analyses (SA)

- "Sensitivity analysis" is aimed at describing how much output values are affected by changes in model input values.
- It is the investigation of the importance of imprecision or uncertainty in inputs in a decision-making or modeling process.
- The exact character of sensitivity analysis depends upon the particular context and the questions of concern.
- Sensitivity studies can provide a general assessment of model precision when used to assess system performance for alternative scenarios, as well as detailed information addressing the relative significance of errors in various parameters.

SA

- Initial sensitivity analysis studies focus on two products:
  - Detailed results to guide research and assist model development efforts, and
  - Calculation of general descriptions of uncertainty associated with model predictions so that policy decisions can reflect both the predicted system performance and the precision of such predictions.
- Most sensitivity analysis approaches examine the effects of changes in a single parameter value or input variable assuming no changes in all the other inputs.
- Sensitivity analyses can be extended to examine the combined effects of multiple sources of error as well.

Sensitivity Coefficients

- These are one measure of sensitivity, defined as the derivative of a model output variable w.r.t an input variable or parameter.
- A sensitivity coefficient can be used to measure the magnitude of change in an output variable Q per unit change in the magnitude of an input parameter value P from its base value P_o.
- Let SI_{PQ} be the sensitivity index for an output variable Q with respect to a change ΔP in the value of the input variable P from its base value P_o. Noting that the value of the output Q(P) is a function of P, a sensitivity index could be defined as
  - $SI_{PQ} = [Q(P_o + \Delta P) - Q(P_o - \Delta P)]/(2\Delta P)$
- A dimensionless expression of sensitivity is the elasticity index, EI_{PQ} that measures the relative change in output Q for a relative change in input P. $EI_{PQ} = [P_o/Q(P_o)]SI_{PQ}$

Example

- The simple water quality model provided by Vollenweider's (1976) empirical relationship for the average phosphorus concentration in lakes is given by: $P = {(L/q)}\Big/\left[1+\sqrt{z/q}\right]$
- Where P is the phosphorus concentration, (mg/m^3),
- L is the annual phosphorus loading rate, (milligrams per square meter per year, mg/m^2 a),
- q is the annual hydraulic loading, (m/a or more exactly m^3/m^{2}a),
- z is the mean water depth, z (m).

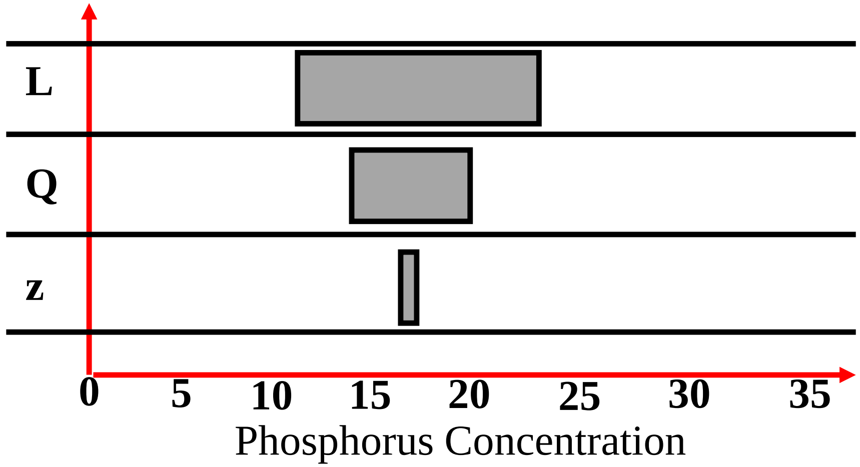
Example

- For Lake having L = 680 mg/m^{2}a; q = 10.6 m/a; and z = 84 m, yields P = 16.8 mg/m^3.
- Values of phosphorus concentrations less than 10 mg/m^3 are considered oligotrophic, whereas values greater than 20 mg/m^3 generally correspond to eutrophic conditions.
- Reasonable ranges reflecting possible errors in the three parameters yield the values in Table below.

Parameter	Parameter Value		Phosphorous Concentration		
	Low	High	Low	High	Range
L (mg/m^{2}a)	500	900	12.36	22.26	9.89
q (m/a)	8	13.5	14.51	20.05	5.54
Z (m)	81	87	16.61	17.04	0.43

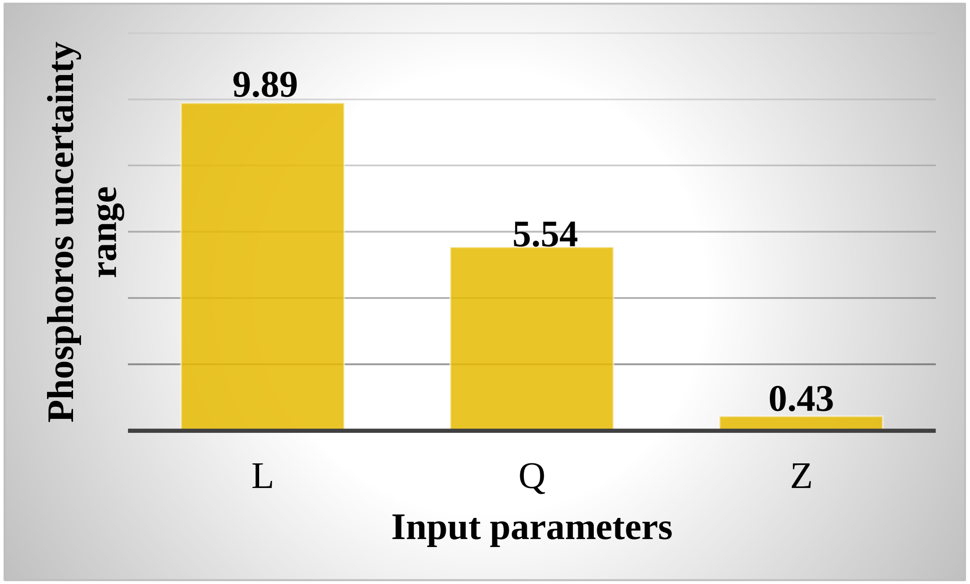
Tornado Diagram

- One may want to display these results so they can be readily visualized and understood. A tornado diagram (Eschenback 1992) would show the lower and upper values of P obtained from variation of each parameter, with the parameter with the widest limits displayed on top, and the parameter having smallest limits on the bottom.
- These error bars shown indicate there is substantial uncertainty associated with the phosphorus concentration P, primarily due to uncertainty in the loading rate L.



Pareto chart

- An alternative to tornado diagrams is a Pareto chart showing the width of the uncertainty range associated with each variable, ordered from largest to smallest.



Spider plot

- Another visual presentation is a spider plot showing the impact of uncertainty in each parameter on the variable in question, all on the same graph.
- It shows the particular functional response of the output to each parameter on a common scale, so one needs a common metric to represent changes in all of the parameters. Here we use percentage change from the nominal or best values.