Prepared for:

RWS-RIKZ

Uncertainty Communication

Report

December 2007

WL | delft hydraulics

Q3435

Prepared for:

RWS-RIKZ

Uncertainty Communication

Joost Beckers

Report

December 2007

Client:	RWS-RIKZ
Title:	Uncertainty Communication

Abstract:

This report is about uncertainty communication. Its goal is to give some helpful directions to anyone that needs to communicate about uncertainty, for example to get support for an uncertainty analysis in a model study or to explain the results to an audience of non-scientists. The report gives a glossary of theoretical ideas and practical guidelines for uncertainty communication in the process of policy making, including examples of verbal and graphical presentation. It can be used by statisticians, technicians and scientists that want to increase awareness of uncertainty, incorporate uncertainties in the policy making process or explain probabilistic findings to an audience.

References:	<begin hier>

Ver	Author		Date	Remarks	Review		Approved by	
1.0	J.V.L. Beckers		20-12-2007	final	A.H. Weerts		C.A. Bons	

Project number:	Q3435
Keywords:	uncertainty, probability, statistics, communication
Number of pages:	29
Classification:	None
Status:	Final

Contents

1 Introduction

The goal of this report is to give some helpful directions to anyone that needs to communicate about uncertainty. It can be used by statisticians, technicians and scientists that want to increase awareness of uncertainty, incorporate uncertainties in the policy making process or explain probabilistic findings to an audience. The report gives a glossary of theoretical work and some practical guidelines for uncertainty communication in the process of policy making, including examples of verbal and graphical presentation. It is not a detailed and comprehensive overview of the theoretical work on this subject. For further reading, the following references are recommended:

- Van der Sluijs et al, RIVM/MNP Guidance for Uncertainty Assessment and Communication: Detailed Guidance (2003)
- Morgan and Henrion, Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis (1992)
- Friedman et al, Communicating Uncertainty : Media Coverage of New and Controversial Science (1999)
- IPCC, Guidance Notes for Lead Authors of the IPCC Fourth Assessment Report on Addressing Uncertainties (2005)
- Patt and Schrag, Using specific language to describe risk and probability, Climatic Change 61, 17-30 (2003).
- Patt and Dessai, Communicating uncertainty: lessons learned and suggestions for climate change assessment, Comptes Rendu Geosciences 337, 425-441 (2004)

1.1 Outline

The outline of this document is as follows. Chapter 2 discusses the role of uncertainty communication in the context of policy making. The rest of the report deals with practical uncertainty communication techniques. Chapter 3 gives a glossary of statistical expressions, definitions and uncertainty terminology. Uncertainty represented by numbers is discussed in Chapter 4. Chapter 5 gives examples of how to present uncertainty in graphical representations. In Chapter 6 we draw some conclusions.

2 Uncertainty management

2.1 Modelling as a part of the process of policy making

Science is only one of many inputs in the process of policy making. Social, legal, financial and political aspects can be just as important as any technical point of view. Ideally, there is a good balance between scientific and political arguments. Using too much science in politics leads to technocracy at the cost of democracy. Too much politics in science reduces scientific quality and authority.

There are several theories about interaction between science and politics. In the classical, or *normal science* model, science is an objective, valid and reliable practice (Thomas Kuhn, 1962). Scientists believe that for every scientific question there is an ultimate conclusive answer, which they can find by doing research. Uncertainty can always be minimized by doing more research, even though this can take some time. An example of this view is found in the first scientific assessment of climate change by the IPCC: "We are confident that the uncertainties can be reduced by further research." (Houghton, 1990). In the normal science model, uncertainty and disagreement between scientists are not welcomed. At best, it indicates that scientific methods are not yet developed far enough to come to conclusive results. At worst it suggests that one of the experts is biased or corrupt.

The normal science model breaks down when science cannot give conclusive answers, even after extensive research. A way to solve this problem is by *evidence evaluation*: to bring the different perspectives together and use dialogue to build consensus. The available scientific evidence and arguments are discussed by multidisciplinary expert panels. The goal is to reach consensus on robust findings by dialogue and participation by all stakeholders. An example is the IPCC Third Assessment Report on global warming in 2001, which states that: "the balance of evidence suggests a discernible human influence on global climate", but on the other hand: "understanding of processes related to ice flow is limited and there is no consensus on their magnitude."

For some complex issues today, such as genetic modification of organisms, even the evidence evaluation strategy fails, because no experiments can be done, the system itself is unpredictable, or because there is more ethics than science involved. For this type of problems, the model of *post-normal science* (Funtowicz and Ravetz, 1990) was developed. This model accepts that science is always developing and will never give a 100% certainty. The best one can get is the current 'state of the art'. Politicians and scientists discuss both the quality of the available scientific evidence and the non-technical aspects and collaborate to make the best policy. Uncertainty and disagreement between scientists is, in this model, a normal phenomenon. One needs to weight the different opinions along with non-scientific arguments to come to the right decision.

2.2 How to address uncertainty

Ideally, a strategy is chosen how to deal with uncertainty before the actual research is started. The required (and obtainable) confidence levels are discussed beforehand and an uncertainty analysis is included as an integrated part of the assignment. This way, it can be avoided that any (unexpected) uncertainty is seen as a weakness or even failure of the model calculation.

The acceptable level of uncertainty depends on the application and must be seen in a socio-economic context. It should, therefore, be defined through a dialogue between the modeller and the water manager. An analysis of the key sources of uncertainty is crucial in order to focus the study on the elements that produce most information of relevance to the problem concerned.

The following is a guidance for the modeller to address uncertainty issues and increase the water manager's awareness of uncertainties in a proposed model study. The guidance is based on the SPIN sales method (Rackam, 1989) and gives four types of questions: Situation, Problem, Implication, and Need-payoff. The questions should be asked in this order.

1. Situation: Facts about the background of the water manager and the proposed model study.

What needs to be modelled and why?
What do you expect to learn from these modelling studies?
What do you expect to be the outcome?
What will you do with the results?

2. Problem: Questions about the water manager's difficulties or dissatisfactions.

Do you expect that the model calculation will be totally accurate and conclusive?
How accurate are the measurements? Is there a natural variability?
Do you expect any differences between the model study and future observations?
How large do you think these differences are for the current state-of-the art models?
What is the maximum acceptable difference between model result and measurement?
If we applied a safety margin for these differences, how large should it be?
Do you think that people in general believe that the model is 100% accurate?
What will the public think when we present only a single model result without a confidence interval?
Is a single model result without a confidence interval credible?

3. Implications: Questions about the consequences or effects of a water manager's problems.

What are the consequences of a discrepancy between model and measurement?
What will happen when people find out that there are discrepancies between model and observations?
Would the project will be delayed?
Would people take legal action?
Will you loose credibility?
Will this cost you time?

4. Need, pay-off: Questions about the value that the water manager perceives in a solution.

Would it help if we calculate a confidence interval and a safety margin for the model?
Do you think a model study with an uncertainty analysis would be more credible?
How much time would you save when the model study would be accepted immediately?
Do you see the value of some form of uncertainty analysis?
Would you like to see some examples of uncertainty studies?
Which of these would fit your situation?

2.3 Dealing with technical uncertainty

Once the modeller and water manager have agreed to analyze uncertainty in some way, there are several ways to deal with uncertainty in the process of policy making.

2.3.1 Safety margins

If a full probabilistic model is not available, but it is possible to make an estimate of the uncertainty, this can be used as a safety margin. An advantage of this approach is that it can be explained easily to a broad audience. The estimated uncertainty defines the safety margin, which is added to the model calculation and the result is used as a normative scenario.

The magnitude of the safety margin will, however, often be the subject of discussion between different stakeholders. This magnitude determines the probability of an outcome beyond the safety margin, most likely an event that should be avoided. Although this probability is never zero, a large safety margin will make it small enough to be acceptable. On the other hand, a larger safety margin will be more expensive.

2.3.2 Scenarios

If a probabilistic model is not available, but there are several deterministic models, these can be used as scenarios to give some indication of how the system may develop in the future. In climatology, for example, the various climate models give a range of scenarios. A scenario is not a forecast, it is only one of many plausible futures.

The use of scenario’s requires making assumptions that in most cases are not verifiable. It is impossible to assess the probability of the different scenarios. All scenarios are probable to some extent, but it is unknown how probable. The contingency plan should prepare for each scenario.

A pitfall when using multiple scenarios is to average over the scenarios to produce an overall ‘most probable’ outcome. By doing so it is assumed that all scenarios are equally probable and that they represent a probability distribution. However, this is not known. Moreover, the next step is usually to ignore uncertainty and consider only the single ‘most probable’ scenario.

2.3.3 Probabilistic decision support

Another method to deal with technical uncertainty is to incorporate the uncertainty information in the process of decision making. In a probabilistic cost benefit analysis (CBA) the uncertainty information is used to reach a balance between the risk and the cost of countermeasures for a particular application. In order to do this, detailed information about the uncertainty is required, i.e. the full probability density function (PDF) of the model results.

Consider a community that is liable to suffer a loss of 2 million euros when a flood occurs. This loss can be reduced to 500,000 euros by evacuating the area. However, the actual evacuation costs are 100,000 euros. The decision whether or not to evacuate should be based on the probability of the flood.
Suppose that a probabilistic forecasting model predicts that the probability of flooding is 10%. In case of evacuation, the expected cost are 100,000 + 500,000 * 10% = 150,000 euros. If the community decides not to evacuate, the expected cost are 2,000,000 * 10% = 200,000 euros. Based on this CBA, the correct decision would be to evacuate, even though there is a 90% probability that the flooding will not occur. In the above example, due to the large difference between loss of evacuated and non-evacuated, the critical probability for evacuation is 6.67%.

A disadvantage of this approach is that it focuses only on the technical aspects. Any social or ethical perspectives are very difficult to incorporate into the cost-benefit analysis. For example, the hypothetical cost of a human life is not widely accepted and probably will not in the near future.

2.4 Specification of uncertainty and probability

Whichever approach is used, it is very important that the requirements to a probabilistic study are clearly specified. This includes a definition of the quantities that will be studied. An *uncertainty* must refer to a continuous variable. A *probability* refers to an event that either occurs or does not occur (www.metoffice.gov.uk).

For example, a statement that there is "a 30% probability of elevated water levels in the Rhine" is meaningless because it is not clear whether it is for a specific place or just somewhere in the river, there is no time given and it is not stated how high the water level will be. Examples of well-defined probability forecasts could be:

- 30% probability of a discharge of more than 8000 m^3/s at Lobith between 12:00 and 18:00.
- 70% probability of a critical water level (waarschuwingspeil) for at least one location along the Rhine on Tuesday.
- 10% probability of flooding in Borgharen overnight.

It is generally easier to define events and verify them unambiguously for specific locations, but as the second example shows it is also possible to define probabilities for multiple locations. The third example illustrates how even quite a low probability can give a useful warning of a serious event likely to lead to significant disruption. Even though there is a 90% probability that the event will not occur, knowledge of the 10% risk enables users to be prepared for the worst rather than being caught out.

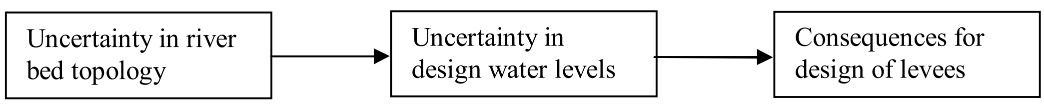
The use of examples can help in specifying the products of the probabilistic study. The next chapters will show some of these examples of various presentations of uncertainty and probability.

2.5 From uncertainty to decision making

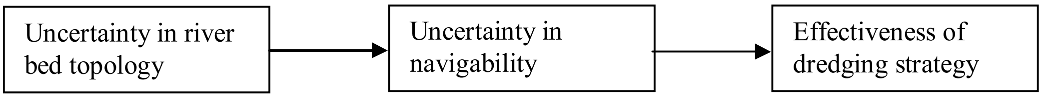
The direct results of an uncertainty analysis are often meaningless to a non-specialist. Even a bandwidth around a model result is perhaps appealing to people from a technical background, but its practical applicability is limited. A stronger connection should be made between uncertainty and its consequences and preferably expressed in terms of decision making. For example, uncertainty can lead to increased financial risk, which may change the outcome of a cost benefit analysis. Or an uncertainty can cause some system to fail a safety standard. These are consequences that are understood by water managers and decision makers.

It is impossible to give a general recipe for post-processing of uncertainty information, because the consequences of uncertainty depend on the context of the decision. The most practical way to understand this context is for researchers to interact with decision makers, ideally in an ongoing manner that allows a two-way flow of ideas and information. Below are some examples of translating uncertainty information to practical support for decision making. These examples might help to understand how this process works.

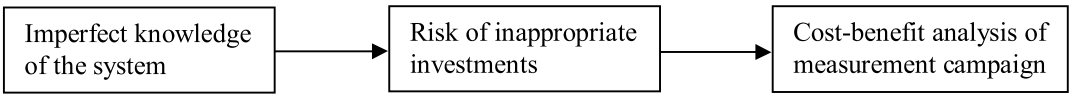
- Van Vuren (2005) performed a stochastic modelling study of river morphodynamics and made a link to flood risk management. The consequences of uncertainties in river morphology were translated to practical river management by calculating the effect on the design water level. To the flood risk manager, this design water level is the most important input parameter for the design of flood defense systems. An increase in design water level leads to an increase of the minimum crest level of levees along the river.



- In a second example, Van Vuren (2005) translated the uncertainty in morphology to river navigability. The river manager is responsible for a certain level of navigability (95% of the time) of the shipping route. Maintenance dredging is organized such that a minimum water depth along the entire route is guaranteed. Van Vuren investigated different dredging strategies in a probabilistic study and calculated the effect on the navigable time, taking into account uncertainties in the morphodynamical processes. This way, the water manager could assess the different dredging strategies and determine which was most cost effective.

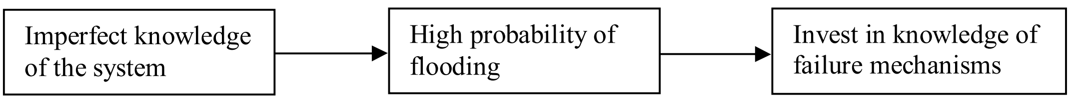


- Korving (2004) developed a probabilistic method to calculate the performance of sewer systems, taking into account uncertainties in knowledge of the sewer system dimensions, pump failure probabilities, etc. Korving used so-called Bayesian decision-making to extract as much information as possible from the available data, some of which was imperfect or unreliable. The costs of measurements is weighted against the risk of inappropriate reconstruction as a result of unreliable or insufficient measurements. The results were used to support the decision whether or not to start a costly measurement campaign, to support investments and maintenance.



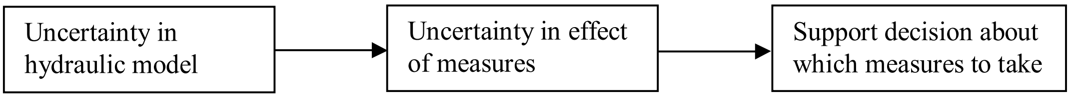
- The Hydraulic Boundary Conditions (HBC) for the North Sea coast are calculated probabilistically, accounting for variability and uncertainty in water levels and wave conditions. Part of the calculation, namely the failure mechanism of a levee is, however, fully deterministic and assumed free of error. This is justified by the argument that the result of the calculation is the HBC, not the probability of flooding.

Within the project VNK (Veiligheid Nederland in Kaart: DWW, 2005), the uncertainty associated with the failure mechanisms was taken into account. This led to an increase in the (expected) probability of flooding, mainly due to the lack of knowledge of certain failure mechanisms. If a failure mechanism is very uncertain, one has to accept a relatively large probability of failure, even at low water levels. The conclusion of the project was that, in order to reduce the probability of flooding, the uncertainty in these failure mechanisms should be reduced. This is far more cost-effective than upgrading the coastal defense system.



- The risk of flooding of areas along the Dutch major rivers can be counteracted by several measures, such as building buffers (flood planes) and enhancing the discharge capacity of the river. Several combinations of measures can be taken, each with different cost and effectiveness in lowering the water level. The preferred combination of measures is selected using these technical arguments and weighing them against socio-economical arguments.

The effectiveness of a measure is calculated using a hydraulic model of the river. In 2006, WL | Delft Hydraulics (WL, 2006) investigated the uncertainty associated with these model calculations. The results of the uncertainty analysis indicated that the outcome of some measures were more uncertain than others. This information can be used to support the decision about which measures should be taken. On the short-term, the results of the uncertainty analysis helped to prioritize further developments in the hydraulic model.



3 Glossary & definitions

Statisticians have a different perception of some statistical concepts, such as uncertainty, probability and risk than other people. For a clear communication between the two it is important for the scientist to understand these differences. Below is a glossary of terms that are often used in statistics and risk assessment, and some potential misconceptions. This list can help to enhance clear communication.

3.1 Accuracy and precision

The terms accuracy and precision are mostly used in the context of measurements. They can also be used for model results, however with some precautions.

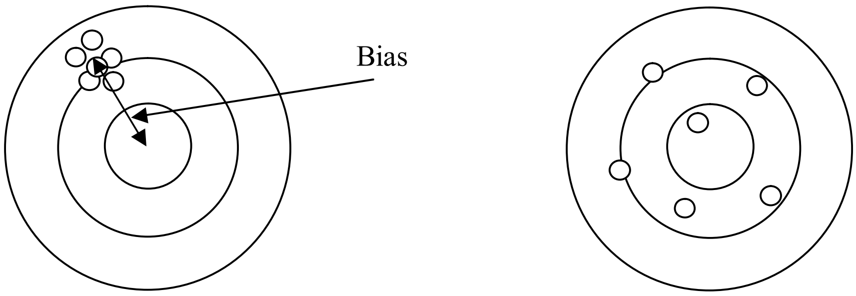
Precision is a measure of how well a result can be determined, without reference to a theoretical or true value. It is the degree of consistency and agreement among repeated experiments; also the reliability or reproducibility of the result. Precision can be measured quantitatively by the *variance* of the readings. The variance is obtained by calculating the squared difference of each result from the mean and taking the mean of all these squares. Statisticians actually prefer to err a bit on the safe side by calculating a slightly larger number than the mean of the squares. Instead they divide the sum of the squares by a number which is one less than the number of readings. For example in case of ten readings, divide by nine. The *standard deviation*, or *spread*, is the square root of the variance.

Accuracy is the correctness, or the closeness of agreement between a model result and a true or accepted value, often an observation. It is a qualitative term referring to whether there is agreement between the results of the model and reality. Because accuracy refers to a true value, the degree of accuracy can only be determined by using a measurement method that is more accurate than the model.

There are two definitions for the quantitative measure of accuracy:

- The systematic error, or bias. This is the difference between the mean of a very long series of readings and the true value.
- A combination of the systematic error and the statistical error. The statistical error, or variance, is related to the precision, so in this definition, accuracy includes precision.

Whenever accuracy is quantified, make sure that it is clear which definition is used.



High precision, low accuracy | Low precision, better accuracy

There is a subtle difference between precision and resolution. *Resolution* is the smallest step of interval that can be distinguished on the object or parameter being measured, or on the instrument that is used. The resolution of an instrument is, however, not necessarily the precision of the reading. The precision of the reading could suffer from other random errors, such as temperature fluctuations.

In deterministic models, all results of repeated runs on the same computer will be identical. The precision is then determined by the machine precision, typically the round off error in the last byte of the floating point representation. In stochastic models, the precision of the result can be determined by repeated runs with a different seed for the random number generator.

3.2 Probability and frequency

There are two interpretations of the term *probability*:

- The '*Frequentist*' view is that probabilities are a natural phenomenon based on relative frequencies of occurrence. For example, the probability of a flooding is something can be calculated from historical records of water levels. This approach relies on the availability of observations.
- The '*Bayesian*' view is that probabilities represent our degree of certainty about the truth of propositions (i.e. statements that can be true or false). In this approach, the probability of flooding is based on our knowledge of the system and aspects that are unknown enter as uncertainties.

In risk assessment a Bayesian view of probability is often assumed, because there are very few observations of catastrophic events. However, the frequentist interpretation can help to understand and assess probability. For instance, a probability of 1% for an event to occur annually, is understood as the event happening once every 100 years. It should be stressed though, that probability represents an *average* frequency. People tend to interpret the *average* frequency of a stochastic process as the fixed frequency of a periodic process. In reality, if the event has occurred last year, this does not reduce the probability for the event happening again this year. The probability of an event in a stochastic process is always the same, whether the event just occurred or not. Also, if an event has a 1% probability of occurring each year, then the probability of this event happening in the next 100 years is not 100%, but only 63%[1].

[1] You can check this by considering the annual probability of the event not happening. If the events are uncorrelated, they represent a Poisson process and their interarrival times are given by an exponential distribution

The probability of a continuous random variable taking a certain value can be described by a probability density function (PDF), sometimes called marginal probability. The PDF can be integrated to obtain the probability that the variable takes a value in a given interval. The cumulative distribution function (CDF) gives the probability that the variable is less than or equal to a value. The CDF is the integral of the PDF from minus infinity to this value and the PDF is the derivative of the CDF.

People tend to adjust their interpretation of probability according to the potential consequences. The perceived probability is either inflated to the point where one can feel good about taking precautions or the probability is diminished to zero, so one can feel okay about not taking precautions. A large scale catastrophe for which no precautions can be taken is perceived as less probable than a small scale event that is equally probable. On the other hand, the probability of an event with large positive consequences is perceived as more probable than actually true. The lottery is based on this principle. What is very hard for people to do is to say: This event is exceedingly unlikely and any precaution we take will almost certainly be wasted, but the consequences are so horrible that it is still cost-effective to take precautions.

Likelihood is the probability that an event that has already occurred would yield a specific outcome, usually an observation. The concept differs from that of a probability in that a probability refers to the occurrence of future events, while a likelihood refers to past events with known outcomes (from mathworld.wolfram.com).

3.3 Uncertainty and error

Uncertainty is the admitted or expected error in a quantitative statement. The error is the disagreement between the given value and a true or accepted reference value. Error exists in all experiments and can only be minimized. Errors can be random or fixed in value, frequency of occurrence, linear or non-linear in relation to the magnitude of the model result. Some errors skew the data in one direction, other times they just cause random scattering of the data.

Uncertainty can refer to lack of knowledge about a true state (Bayesian interpretation of probability) or natural variability, or randomness of a phenomenon (frequentist interpretation). Uncertainty can result from systematic errors (bias) and statistical errors (due to natural variability and lack or measurements). The total error encompasses both types. Unless the type of error is specified explicitly, people will assume that the total uncertainty is meant, i.e. all errors combined.

Without further specification, the uncertainty is given by the smallest digit in the result. For example, if we say that the water level is 312 cm, we implicitly claim that the true value is between 311.5 cm and 312.5 cm and that the uncertainty is smaller than 1 cm. If the uncertainty is actually larger than this we will quickly loose credibility, because people find out that we are wrong most of the time.

Uncertainty can be quantified by an uncertainty analysis, in which all uncertain aspects in the model calculation and the propagation of errors in the model result are analyzed. An uncertainty analysis is sometimes confused with a sensitivity analysis or risk analysis. The former is an analysis of the variation of the model results caused by a variation of each individual model parameter. A sensitivity analysis does not include any reference to uncertainty in the model parameters. A risk analysis takes into account the effect of all possible outcomes, usually expressed in terms of money.

Reported uncertainty often includes only technical uncertainty. Unreliability of the computing method and unknown processes that may influence the results are disregarded, simply because there is insufficient knowledge to quantify them. This can lead to very strange results, such as physical constants that change over time beyond their original confidence intervals (Taylor, 1969). The NUSAP (Funtowicz and Ravetz, 1990, vd Sluijs, 2003) initiative is an effort to take into account uncertainty due to unknown factors that cause systematic errors. NUSAP identifies four different types of uncertainty:

- Technical (inexactness)
- Methodological (unreliability)
- Epistemological (ignorance)
- Societal (limited social robustness)

NUSAP (Numeric, Unit, Spread, Assessment, Pedigree) introduces the qualitative parameters ‘assessment’ and ‘pedigree’ to measure the non-technical sources of uncertainty. The parameter ‘assessment’ expresses any qualitative judgments about the scientific findings, such as ‘optimistic’ or ‘pessimistic’. The ‘pedigree’ represents the mode of production of the data. For example, a finding from an externally reviewed research is considered of better pedigree than a study that was not reviewed. Both ‘assessment’ and ‘pedigree’ are measured by a quality index and reported along with the scientific finding and its traditional uncertainty.

Instead of ‘confidence interval’, people sometimes use the terms *spread* or *bandwidth*. Statisticians, however, mainly associate ‘spread’ with ‘statistical error’, derived from a series of repeated experiments, ignoring the systematic error. Bandwidth can be used as a synonym for confidence interval.

Uncertainty is sometimes confused with *tolerance*. However, a tolerance is a 100% guarantee that the true value lies within the given range. Uncertainty can be expressed as a 90% confidence interval, leaving a considerable 10% probability for the true value being outside the interval.

In statistics, *bias* simply means non-randomness: a biased result is a result with a systematic error, or a deviation from the true value that does not cancel out if the sample size is increased. For many people, bias is associated with cheating. Some additional explanation is therefore advisable if a model is said to be biased. An alternative is to use the term ‘systematic error’ (that can be corrected for).

To a general audience, an *error* is an indication that something is wrong and therefore the result cannot be used. An alternative is to use the term confidence interval. The confidence interval is defined by a lower bound and an upper bound. The probability that the true value lies between these bounds is called the confidence level. A confidence interval without a confidence level is meaningless. The Bayesian analogue to the confidence interval is called a credibility interval.

People tend to over- or underestimate the uncertainty in results of model calculations, depending on their relation with the production of results (McKenzie, 1990):

- People that were involved in the process of modelling and data retrieval often have a good sense of the uncertainty, at least if they are regularly confronted with measurement data and validation.
- People that are somewhat further from the production process (e.g. managers, clients) tend to overestimate the reliability of the results, because they would like the uncertainty to be smaller than it actually is.
- The public has a general distrust in technical studies, causing their estimated uncertainty larger than the true uncertainty.

The word 'uncertainty' implies the lack of something (certainty) and is sometimes seen as a sign of weakness of the underlying theory. This is in most cases unjust, because virtually all quantitative statements contain some degree of uncertainty. Alternatively, one can use a 'confidence interval' to express the same quantity in a more linguistically positive way.

3.4 Risk and risk analysis

In risk assessment, *risk* is the probability that some event will occur times the consequences, or the effect, if it does occur. Unfortunately, the rest of the world uses only the probability half of this definition. When people ask how big a risk is, they usually mean how likely the event is, not how bad. They define risk as the *probability* of the realization of an adverse event. People also use risk to mean uncertainty; a course of action may be called risky either because the probability of a bad outcome is high or because the probability is unknown. The term risk thus needs to be used with care.

Risk assessors discriminate between *individual risk* and *group risk*. Individual risk is the risk someone is personally exposed to by being in the wrong place at the wrong time. The individual risk can be location-specific, for an individual who is present at a particular location 24 hours per day, 365 days per year, or individual-specific, for someone who is present at different locations during different periods.

The group risk is the risk associated with some event or activity, that can affect a group of people. A group risk is often presented by a table, with different probabilities for events of increasing magnitude. For example, the group risk of a chemical plant can be presented as an annual probability of 1/100 for events that cause 10 casualties, and 1/1000 for 500 casualties. Society usually accepts a relatively high probability for small scale events rather than a rare but large scale catastrophe.

Scenarios are often used in risk analyses to envisage the possible outcome of an event tree. A *worst case scenario* can be used to get an impression of the maximum damage. In most cases, however, the literal worst case scenario is so unlikely that it is completely irrelevant. Any scenario can be made worse by adding a meteorite impact. The value of such a scenario is, of course, very limited. Instead it is better to talk about the *worst credible scenario*, which has a bad outcome and a reasonable probability of happening. What is reasonable depends on the situation. In general, a scenario is sensible if it is necessary to make a contingency plan for it.

The term *probabilistic* is associated with a calculation that explicitly takes into account the probability of multiple event trees or scenarios in an integrated fashion. In a *deterministic* calculation, a single scenario is selected, such as the most probable or the worst case, as representative for all possible scenarios and continue our calculation with this single scenario. The deterministic approach usually leads to more conservative results.

A calculation can be partly probabilistic and partly deterministic. For example, the probability of flooding can be calculated using a probabilistic approach, taking into account multiple combinations of water level and wave conditions, each with a probability of occurring. In contrast, the consequences (damage, casualties) of this flooding can be based on a deterministic calculation. The risk (probability times consequences) is then partly probabilistic and partly deterministic.

A *stochastic*, or random process, is a process that evolves with some indeterminacy. The outcome can be described by probability distributions. A *deterministic* process has only one outcome.

The *expected value* is a mathematical concept that differs from what is generally perceived as the *expected outcome*. It is the probability weighted average outcome, or the mean. For example, if you play a game where you roll a die and receive €10 for any number between 2 and 6, but lose €50 for throwing 1, the *expected value* of each throw is 10 * 5/6 - 50 * 1/6 = zero. However, most people would think of winning €10 as the *expected outcome*, with a risk of losing €50.

Probabilities for *rare events* are difficult to communicate to large audiences. A 1/10,000 annual probability of flooding in the Netherlands is such a small number that it is hard to envisage. Sometimes it helps to compare to a generally known event, such as a plane crash or winning the lottery. If people are willing to pay 20 euros for a probability of winning 1 million euros, why not invest 1 euro to prevent an equally probable loss of 100,000 euros? Another possibility is to express the annual probability in terms of the event happening during a lifetime, or a number of generations. People will often not accept even an extremely small possibility of a very disastrous event to happen. If an audience is asked what is the acceptable probability of flooding of their residence area, most will respond with ‘zero’.

Saying something is *safe* means that the probability of failure is zero. But there is always a remote probability of failure and nothing is risk-free, so nothing should ever be called safe. Instead, one can say that something is safer than a standard, or safer than something else. Likewise, *evidence* or *proof* in a statistical sense is based on probability. There is no such thing as 100% proof. We simply accept a (small) probability of being wrong.

3.5 Qualitative expressions

A *conservative* estimate of probability is an estimate that errs on the side of caution. For a conservative estimate we intentionally overestimate uncertain factors in order to be confident we are not underestimating them. However, to the public a conservative estimate is always a low estimate. When the media talk about a conservative estimate of how much damage an accident has caused they mean that the accident probably caused more damage than the estimate suggests. So to the public a conservative estimate of the size of a risk is an estimate that probably understates the risk — exactly the opposite of what the word means to professionals. Consider using alternatives such as optimistic/pessimistic, or robust, or make very clear what is meant by conservative.

In real life, something is positive if it's good news. But in statistics, a positive relationship means that when one variable goes up, so does the other. A positive conclusion about a study on climate change and the frequency of severe storms means that there is a correlation and that the frequency of storms will increase. Emotionally, however, this is a very *negative* finding, since the risk of flooding will increase.

To statisticians, *significant* means that a finding is unlikely to have occurred by chance: the probability of the outcome being the result of the stochastic nature or randomness of the process is smaller than a predefined threshold. For example, the outcome of an experiment is called *significant* if it supports a hypothesis with a probability that is larger than 95%. Statistical significance is not related to the magnitude of the effect.

To most other people, something is *significant* if it is important or relevant to you personally and this is very much related to the magnitude of the effect. For example, if a model predicts a 1 m sea level rise over the next 50 years, but this result falls within the uncertainty of the model, then a statistician would call it not significant. However, an audience would judge otherwise, depending on their fear of flooding. Therefore the term (in)significant should be used with care.

3.6 Probability-linguistics

Some audiences prefer uncertainty being expressed in words rather than numbers. For this reason scales have been developed that translate probability to vocabulary. Below are some examples:

- IPCC 5 point confidence level scale

Degree of confidence in being correct	Terminology
Less than 1 out of 10 chance	Very low confidence
About 2 out of 10 chance	Low confidence
About 5 out of 10 chance	Medium confidence
About 8 out of 10 chance	High confidence
At least 9 out of 10 chance of being correct	Very High confidence

- IPCC 7 point probability scale (Moss and Schneider, 2000)

Likelihood of the occurrence/ outcome	Terminology
< 1% probability	Exceptionally unlikely
<10% probability	Very unlikely
< 33% probability	Unlikely
33 to 66% probability	About as likely as not
> 66% probability	Likely
> 90% probability	Very likely
> 99% probability of occurrence	Virtually certain

- Weiss 12 point scale, often used for legal purposes.

Probability of being true	
0%	Impossible
0-1%	Hunch
1-10%	Reasonable suspicion
10-20%	Reasonable indication
20-33%	Reasonable belief
33-50%	Clear indication
50-67%	Preponderance of the evidence
67-80%	Substantial and credible
80-90%	Clear showing
90-99%	Clear and convincing
99-100%	Beyond a reasonable doubt
100%	Beyond any doubt

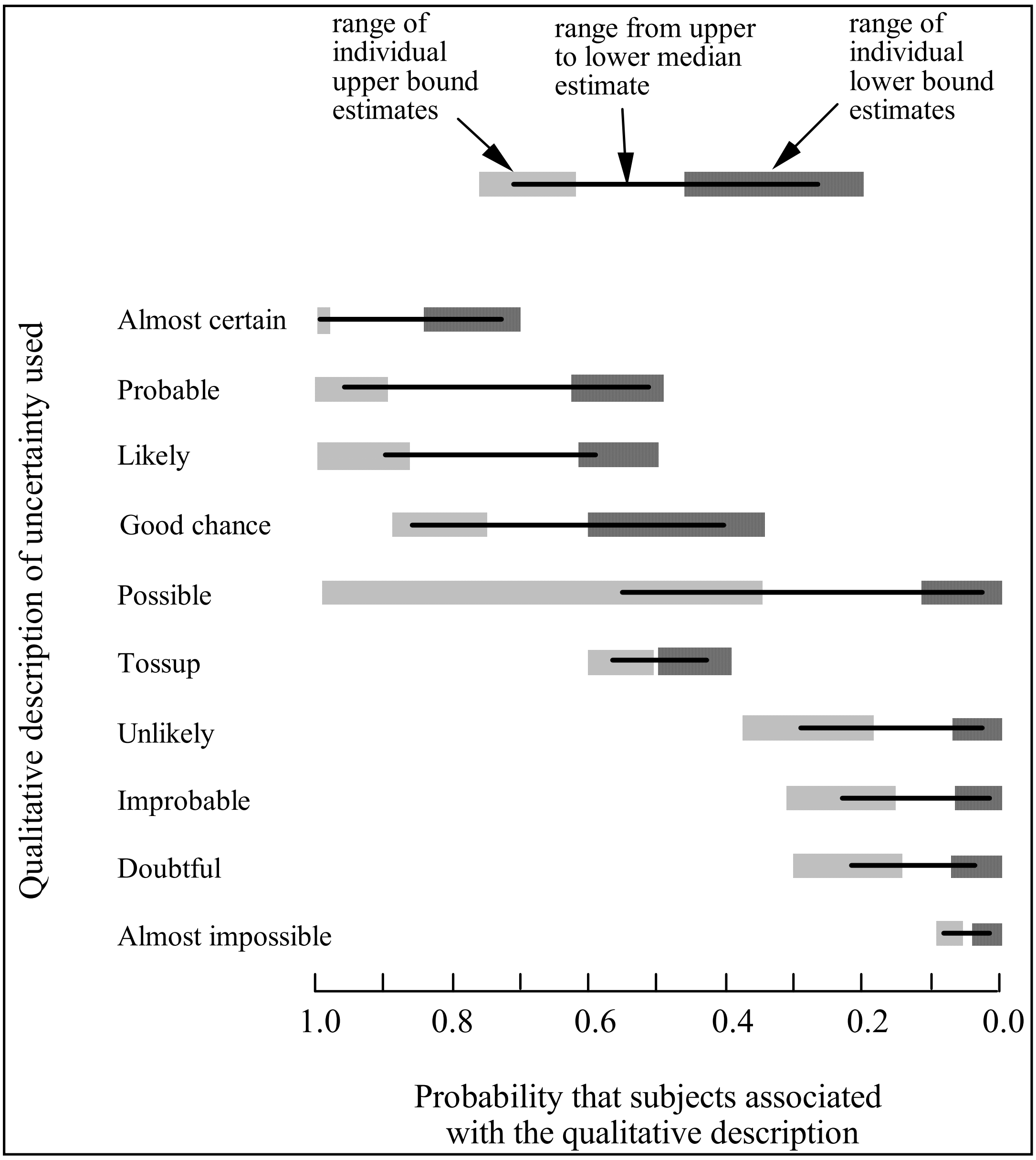
- Schinzer et al 13 point scale.

Probability	Linguistic expression
0%	Absolutely impossible
0-9%	Rarely
9-18%	Very unlikely
18-27%	Fairly unlikely
27-36%	Somewhat unlikely
36-45%	Uncertain
45-54%	Tossup
54-63%	Better than even
63-72%	Rather likely
72-81%	Quite likely
81-90%	Highly probable
90-100%	Almost certain
100%	Absolutely certain

Note the difference between the Schinzer and IPCC scales for the expression ‘very unlikely’.

The use of these scales has been criticized. The meaning of the words can differ between persons, as is shown in the figure below[2].

2. From http://www.geo.uu.se/workshop/vandersluijs.pdf


range of
individual
upper bound
estimates
range from upper
to lower median
estimate
range of
individual
lower bound
estimates
Qualitative description of uncertainty used
Almost certain
Probable
Likely
Good chance
Possible
Tossup
Unlikely
Improbable
Doubtful
Almost impossible
1.0
0.8
0.6
0.4
0.2
0.0
Probability that subjects associated
with the qualitative description

4 Representation of uncertainty by numbers

This chapter discusses various representations of uncertainty in terms of numbers. These include mean absolute error, root mean square error, confidence intervals and quantiles.

4.1 Absolute error

The uncertainty of a continuous variable is often defined by an absolute error. The absolute error is the expected discrepancy between a true value and some approximation to it. Some common notations are:

$x = 5.1 \pm 0.1$, $x=5.1(1)$, $x=5.1, \Delta x = 0.1$, $x=5.1, \delta x = 0.1$

All these expressions claim that the true value of the variable *x* probably lies in the interval [5.0, 5.2]. Alternatively, uncertainty can be defined by giving the relative error, usually written as a percentage:

$x = 5.1 \pm 2\%$

The problem with absolute and relative error is that it is usually not specified what is meant by it. There are two possibilities:

- The mean absolute error (MAE). This an *average of the absolute errors*, which would be measured by performing multiple experiments and if the true value of *x* was known.
- The standard deviation, or *root mean squared error* (RMSE), also to be determined from a series of experiments.

The MAE differs from the RMSE. In fact, the RMSE will always be larger or equal to the MAE.

In many cases it is assumed that the difference between a measured value and the true value is normally distributed, with the standard deviation of the distribution being the uncertainty. This can lead to false interpretations. For non-normal distributions, the 'most probable' value can differ from the ensemble mean. Also, confidence intervals and quantiles can be very different from a normally distributed PDF with the same standard deviation. Therefore, if there is reason to believe that the PDF deviates from a normal distribution, it is wise not to specify uncertainty just by giving a MAE or a standard deviation.

Moreover, for the normal distribution, the standard deviation corresponds to a confidence interval of only 68%. The MAE bounds (for the normal distribution) correspond to an even lower confidence interval of 56%. These probabilities are much lower than the usual interpretation of the uncertainty interval by the public: that it is almost impossible that the true value lies outside the bounds of the interval. It is therefore wise to clearly state that the given uncertainty is a mean absolute error, or a standard deviation, not a confidence interval.

4.2 Confidence interval

A way to avoid these misinterpretations is by means of a confidence interval (Bayesian term: credibility interval). For example: we are 95% confident that our quantity of interest lies between 4.9 and 5.3, where 95% is our confidence level and 4.9-5.3 is the confidence interval. This statement is much more specific than the absolute error and it is likely to be understood by a broad audience[3].

The confidence level thus represents our degree of belief that the true value lies within the confidence interval. The appropriate confidence level depends on the application and can be discussed with the customer before any calculation is made. Specification of the required level of confidence is a good way to bring the uncertainty issue under attention.

The engineering standard confidence interval is 95% or 2σ for a normal distribution (Zar 1984). This seemingly high 95% confidence level still implies a fairly large chance of 5% that the outcome lies outside the confidence interval. This is why higher confidence levels are customary in, for instance, legislative matters. After all, 5% erroneous convictions of innocent people is considerable.

For a normal distribution, confidence intervals are easily derived from the standard deviation. Commonly used confidence levels and their associated intervals in terms of standard deviation for the normal distribution are given in the table below.

confidence level	confidence interval (normal distribution)
68%	$\pm \sigma$
80%	$\pm 1.28\sigma$
90%	$\pm 1.64\sigma$
95%	$\pm 2\sigma$
99%	$\pm 2.58\sigma$
99.7%	$\pm 3\sigma$
99.9%	$\pm 3.29\sigma$

For non-symmetric distributions the confidence interval becomes asymmetric around the most probable value and the plus-or-minus type expression can no longer be used.

If it is not known whether the distribution is normal, one can always use Chebyshev's inequality:

At least 50% of the values are within 1.4 standard deviations from the mean.
At least 75% of the values are within 2 standard deviations from the mean.
At least 89% of the values are within 3 standard deviations from the mean.
At least 94% of the values are within 4 standard deviations from the mean.
At least 96% of the values are within 5 standard deviations from the mean.
At least 97% of the values are within 6 standard deviations from the mean.
At least 98% of the values are within 7 standard deviations from the mean.

[3] Although there is some philosophical controversy about the frequentionist interpretation of a confidence interval as a probabilistic statement. Stating that '90% of the time the value of a true quantity is within the confidence interval' is nonsense if that true quantity is not a stochastic variable.

In general, at least $(1 - 1/k^2)$ of the values are within k standard deviations from the mean. This holds for all probability distributions.

4.3 Quantile

A quantile can be used to represent our degree of belief that the true value of a quantity of interest lies below a certain value. Quantiles are, in fact, one-sided confidence intervals. Quantiles can be derived by inverting the cumulative distribution function (CDF). If the CDF is not known, it can be approximated by a series of ordered samples from a Monte Carlo simulation.

Some quantiles have special names:

The 100-quantiles are called percentiles.
The 10-quantiles are called deciles.
The 9-quantiles are called noniles, common in educational testing.
The 5-quantiles are called quintiles.
The 4-quantiles are called quartiles.
The 2-quantile is the median.

5 Graphical representation of uncertainty

Below is a glossary of some graphical representations of uncertainty. Numerous variations are possible. These examples can help to find a presentation technique that best suits the application.

5.1 Error bars and candle sticks

A well-known technique to indicate uncertainty is by error bars (see Figure 1). The common convention is to use one standard deviation. Error bars can be asymmetric in case of an asymmetric PDF.

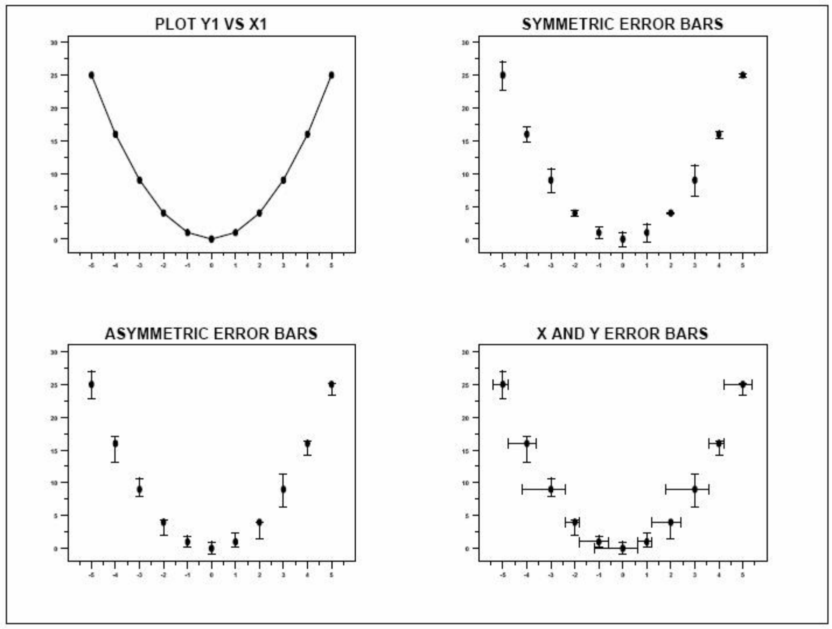


Figure 1: Error bars.

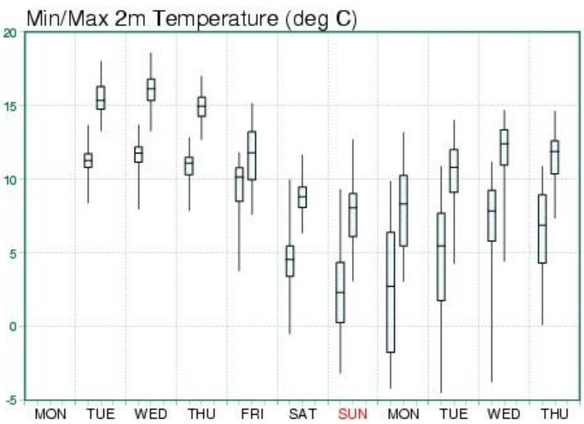


Figure 2: Candle sticks or box plot.

The box plot, box-and-whisker, or candlestick plot (Figure 2) is an extension to the common error bar, giving two confidence intervals. The original box plot, introduced by Tukey in the 1970's uses the lower quartile, median and upper quartile to define the box. The vertical lines (the "whiskers") denote the smallest and largest observations and extend to at most 1.5 times the box width from either ends of the box.

5.2 Point cloud

Uncertainty can be visualized by a cloud of points, which can be generated by stochastic sampling. A problem is that the uncertainty seems to increase if more sample points are used (see Figure 3). This can be resolved by adding a confidence interval (Figure 4). If the uncertainty differs between the points in the graph, this can be represented by increasing symbol size (Figure 5).

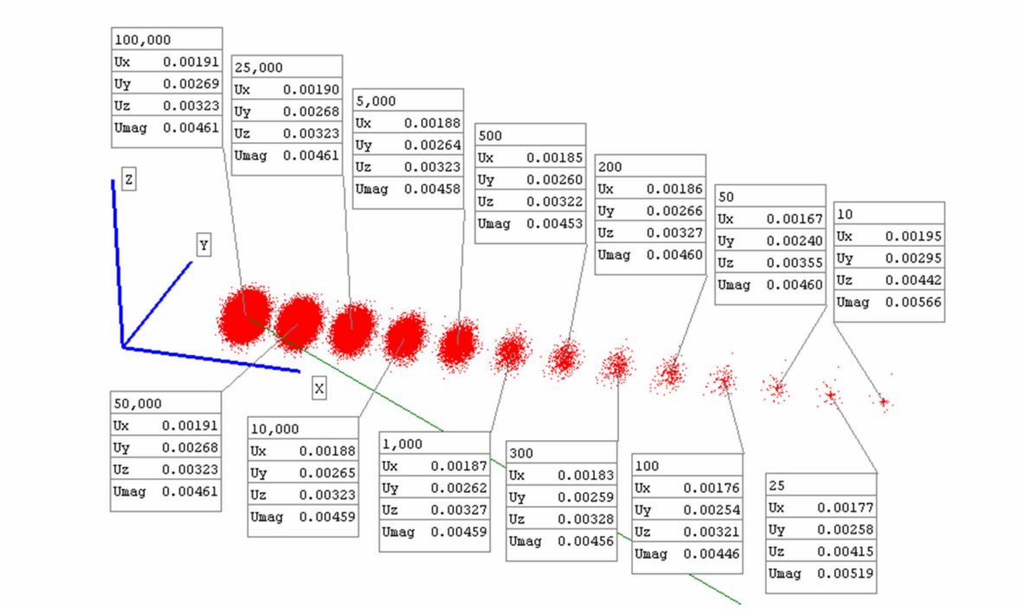


Figure 3: Point cloud.

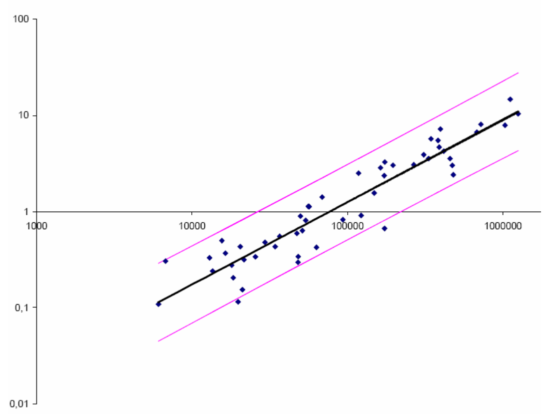


Figure 4: Point cloud with confidence interval.

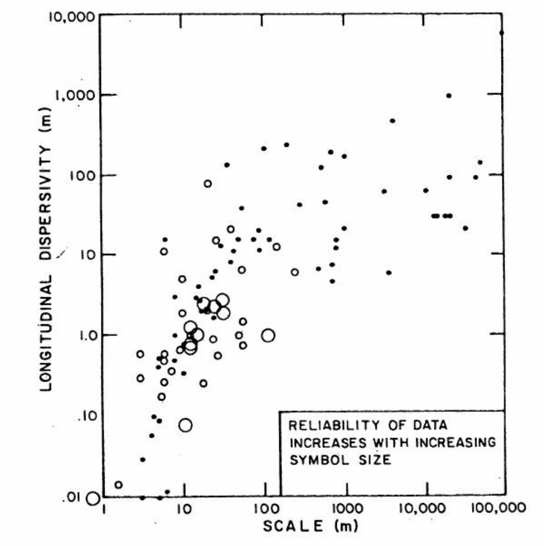


Figure 5: Uncertainty represented by circles.

5.3 PDF/CDF

If the full probability distribution function (PDF) is known, this function can be plotted or its integral, the cumulative distribution function (CDF). A PDF is best for indicating the relative probabilities of values and the shape of the distribution (e.g. skewness). A CDF is better for indicating fractiles and for displaying distributions from discrete sampling (see Figure 6).

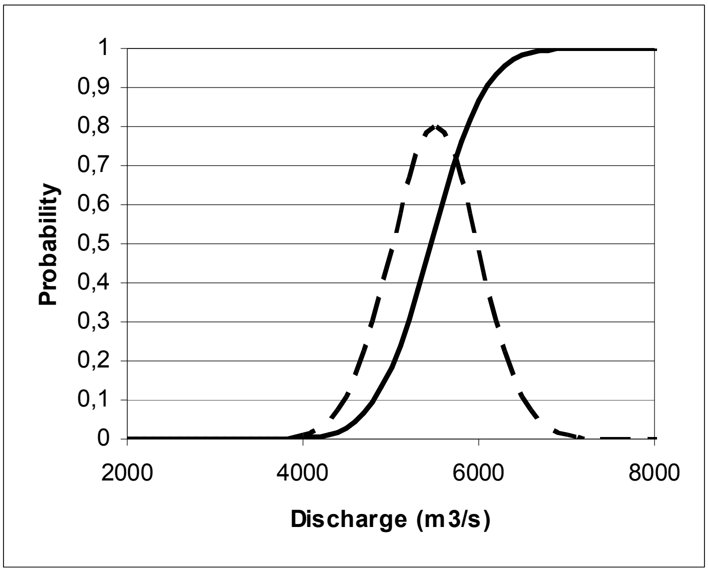


Figure 6 Probability distribution function (PDF) and cumulative distribution function (CDF).

If you decide to present a PDF or CDF, present both. Display them in the same figure or at least on the same page and with identical horizontal scales (Ibrek and Morgan, 1987).

5.4 Pie chart

Another way to visualize probabilities is by using a pie chart. The circle is a natural design element that represents 100% probability. The parts represent different events that can happen with smaller probabilities. The pie chart is most convenient for events that can be well-categorized. A very playful example of the use of pie charts is given in Figure 7, from the web site of the Australian bureau of meteorology[4]. A Java applet spins the circles to give the user a notion of randomness.

[4] http://www.bom.gov.au/lam/Students_Teachers/climprob/rainprbprim.shtml

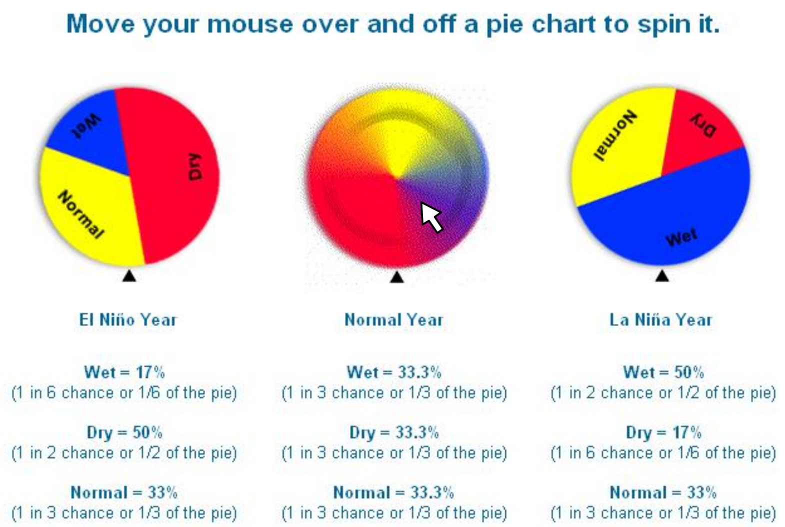


Figure 7: The probabilities of dry, wet and normal years explained by spinning pie charts.

5.5 Bar chart

The bar chart is a variation of the pie chart. A bar charts is more convenient for classes of events that should be visualized in an ordered way. It is also more suitable for comparing a series of probability distributions (see Figure 8).

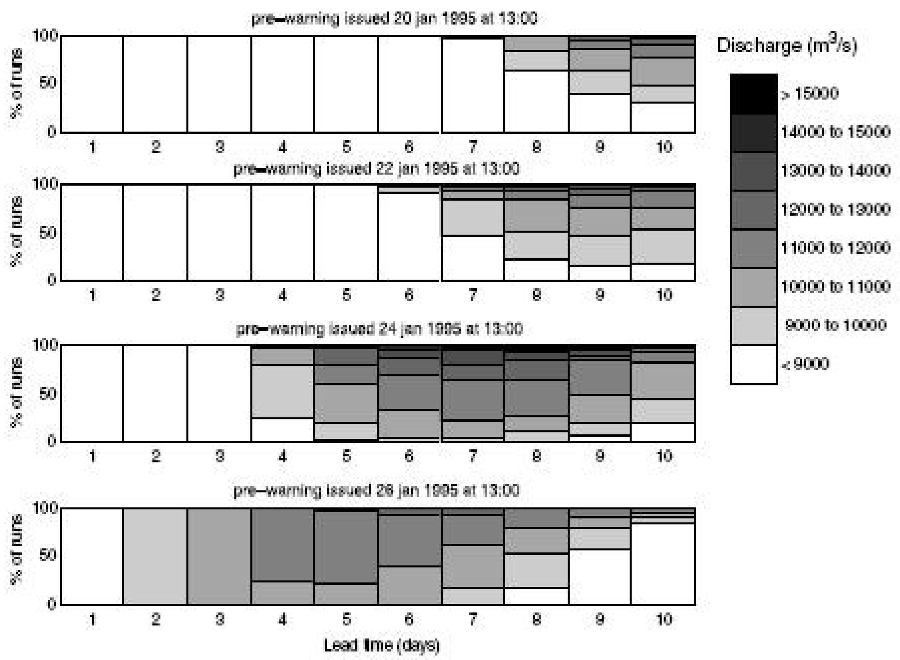


Figure 8: Bar charts representing the probability of discharge classes.

5.6 Time dependent data

An ensemble is, in fact, a time dependent point cloud. Most ensembles are generated from random perturbations of model parameters, boundary conditions or input parameters. They are designed to estimate probabilities by sampling the range of possible forecast outcomes. The probability of a particular event is estimated by counting the proportion of ensemble members which forecast that event to occur. Taking the first forecast example above, 30% would result when 15 out of 50 ensemble members predict more than 5 mm of rain to fall at the specified location in the defined period.

In practice this method does not always give reliable probabilities, because ensemble prediction forecasts are often not calibrated. Any physical process that is neglected in the model causes a deviation from the measurement. If a process is absent in the model, its influence cannot be captured by an ensemble run. The ensemble spread is then smaller than the spread of the absolute error and it is called *underdispersed*. Moreover, if the model is biased, so will the ensemble.

However, there is often a reasonable correlation between the ensemble spread and the true uncertainty. This property can be used by adjusting the ensemble spread every time step, using a predetermined correction factor. If the ensemble is also corrected for any bias, a calibrated probabilistic forecast is obtained, i.e. events that are predicted to have probability P happen a proportion P of the time.

Example of visualization of scenarios are shown in Figure 9, Figure 10 and Figure 11. This visualization is very similar to that of an ensemble and it is tempting to interpret the scenarios as such. Sometimes this is allowed, seeing the scenarios as a ‘poor man’s ensemble’, other times this may lead to false conclusions.

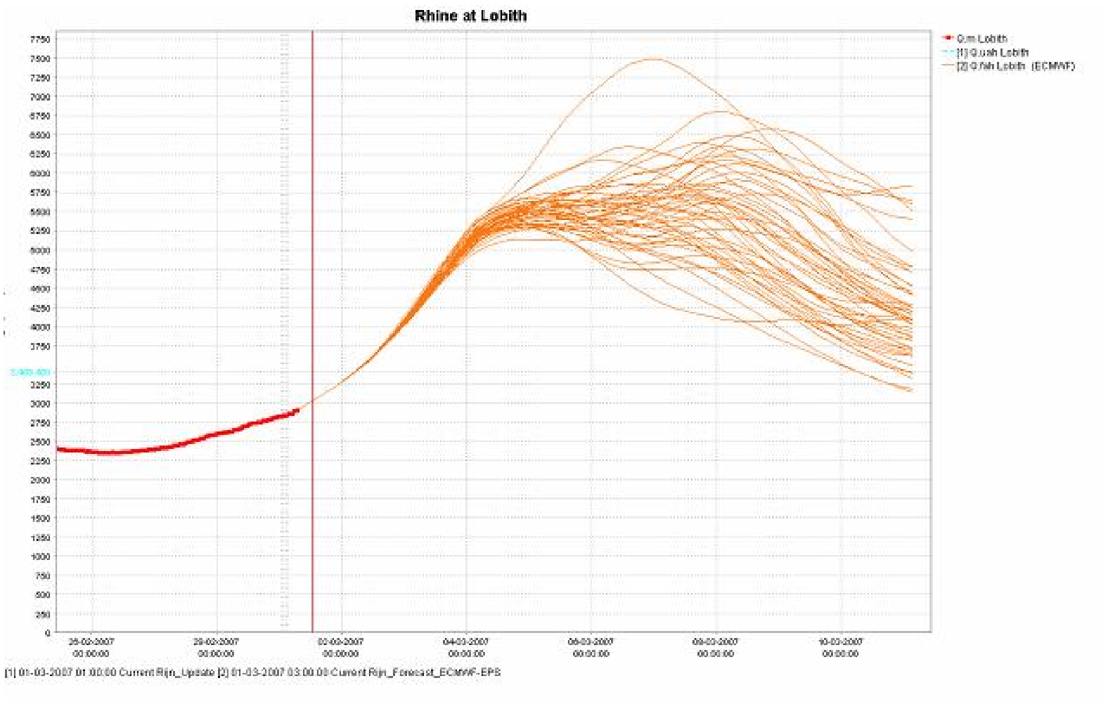


Figure 9: Ensemble visualization (Rhine discharge in FEWS NL).

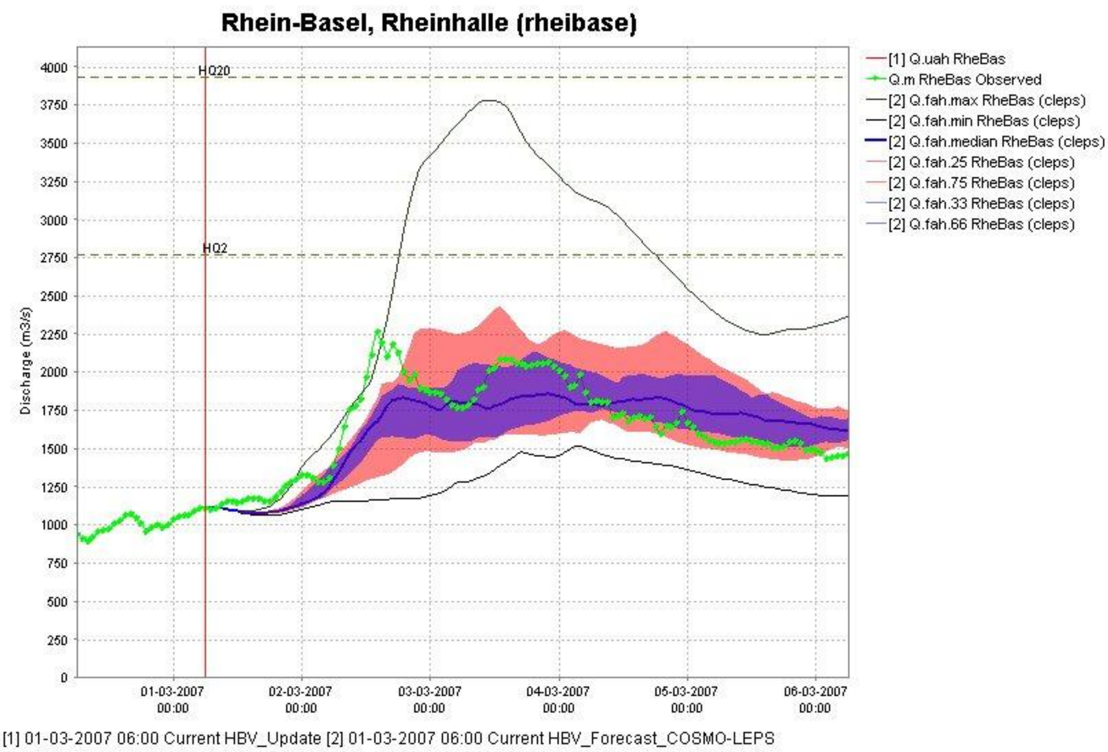


Figure 10: Shading is used to indicate confidence intervals in FEWS-FOEN (Switzerland).

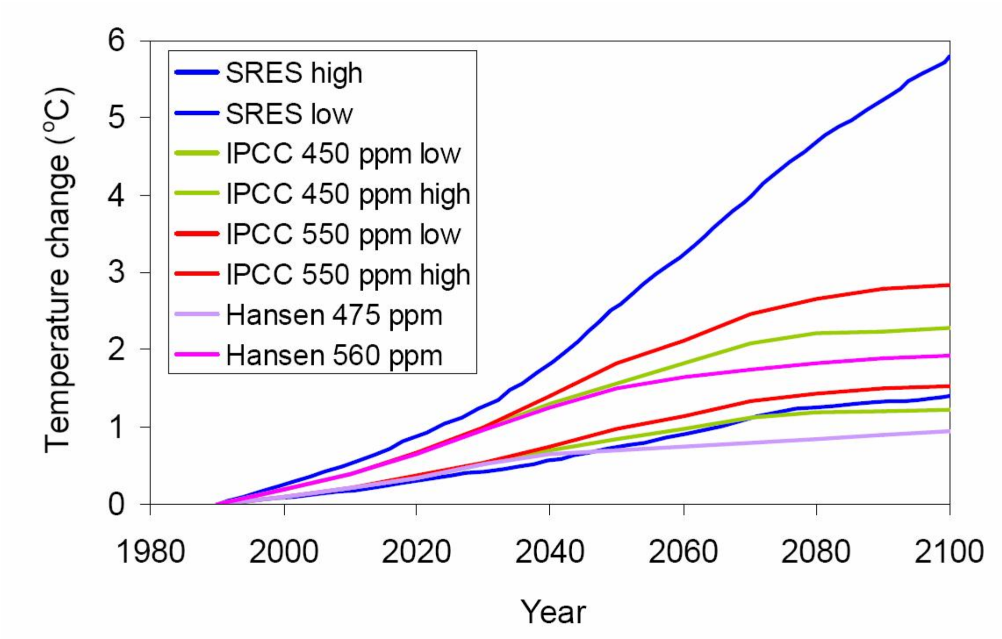


Figure 11: IPCC Climate Change Scenarios. Comparison of global warming based of various assumptions and models.

5.7 Multivariate uncertainty

In case of more than one variable and the uncertainty depending on the value of each of them, the visualization becomes more complex. Figure 12 shows four different representations of multivariate uncertainty.

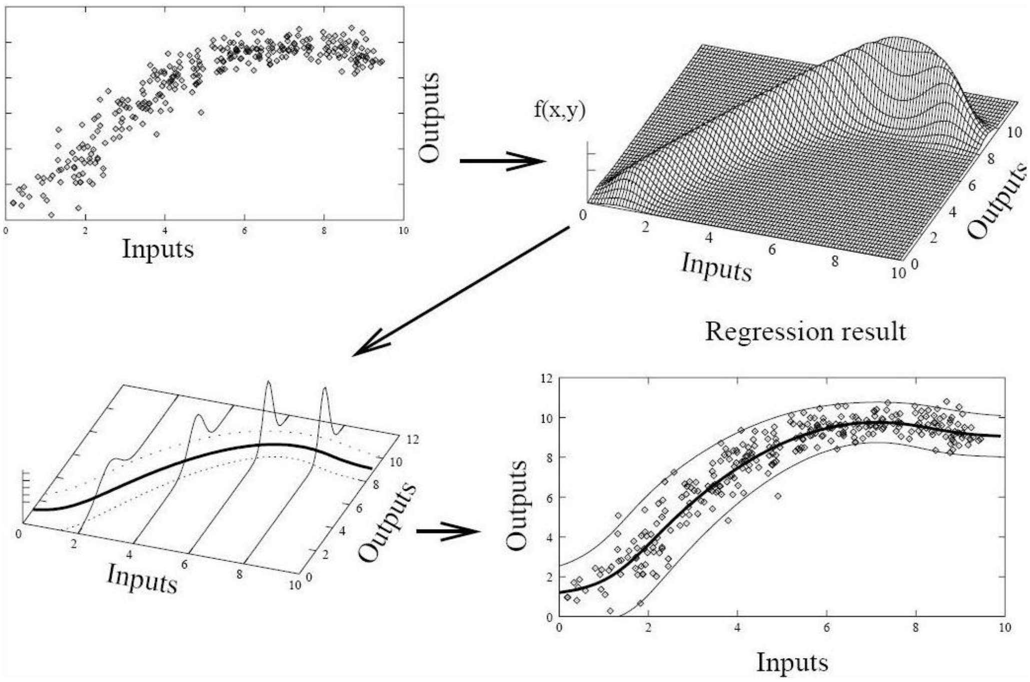


Figure 12: Visualization of multivariate uncertainty (from Torfs and Warmerdam, 2002).

Uncertainty in spatial (geographical) data can be represented in numerous ways:
1. A difference map between model and measurement
2. Two or more maps representing scenarios with specific parameter settings
3. Ensemble maps, several equally possible realizations with random parameter settings
4. Uncertainty represented by whiteness, certainty represented by bright colours.
5. Uncertainty plotted in a third dimension
6. Uncertainty plotted using a different plotting technique, such as isolines
7. Absolute or relative uncertainty map, percentile maps

For untrained audiences, the high-density information of maps 4-6 is often too much. A separate figure to display uncertainty usually works better. If a dynamic (animated) representation is possible this can also be an elegant way.

Figure 13 is an example of probabilistic weather information from the university of Washington (http://www.probcast.washington.edu). The geographical spreading of the probability of precipitation for the next 48 hours varies is represented by grey and colour scales.

Figure 14 shows the uncertainty in river bathymetry as a function of space (km along the river) and time. A darker colour represents a larger uncertainty.

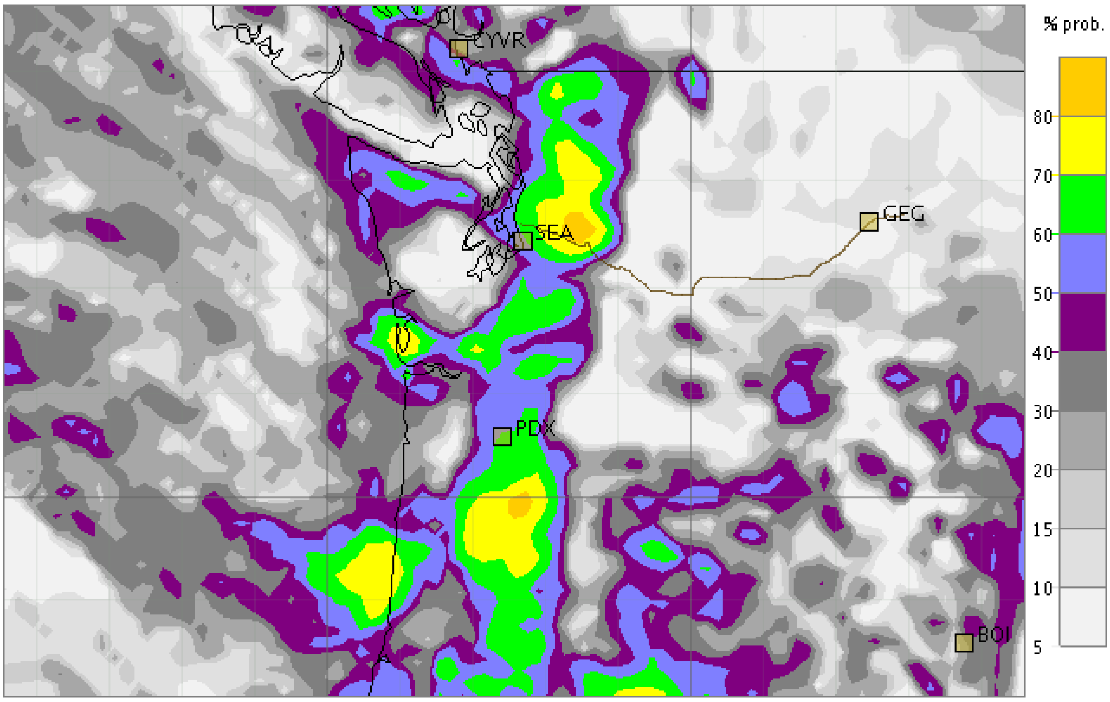


Figure 13: The probability of more than 0 mm of precipitation for the next 48 hrs from a probabilistic weather forecast (from: www.probcast.com).

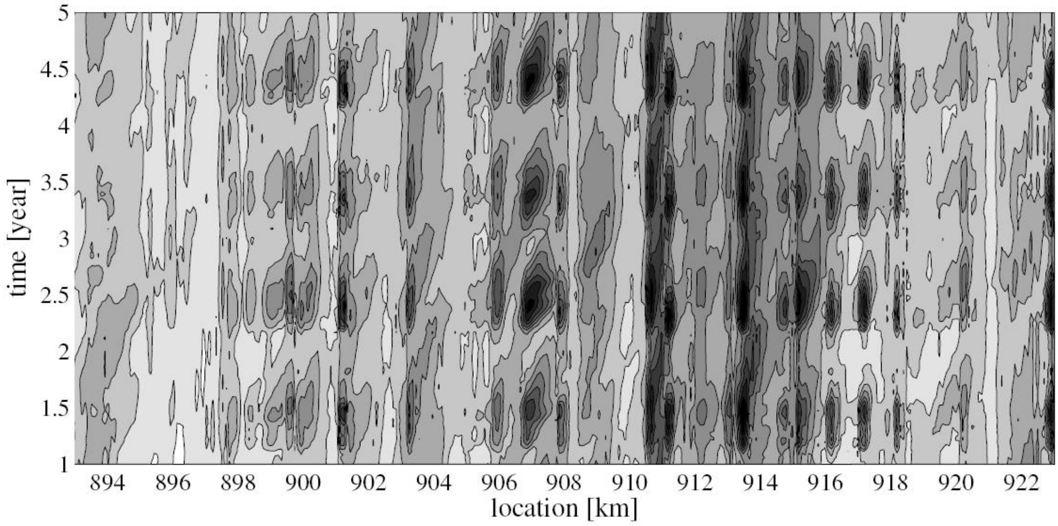


Figure 14: Uncertainty in river bathymetry as a function of location (x-axis) and time (y-axis). From: Van Vuren, 2005.

6 Conclusions

Communication about uncertainty can be done in many different ways, some of which have been discussed in this report. In chapter 2 the role of uncertainty communication in the context of policy making has been discussed. Science is considered one of many inputs in the process of policy making. Uncertainty information is of great value to this process, especially if a connection is made between the uncertainty and its consequences for decision making.

Chapters 3, 4 and 5 discussed practical uncertainty communication techniques, including a glossary of statistical expressions and uncertainty terminology and various representations of uncertainty by means of language, numbers and graphics. This material can help scientists and engineers to explain uncertainty to non-technical audiences.

7 References

Friedman, S.M., Dunwoody, S. and Rogers, C.L. (1999): *Communicating Uncertainty: Media Coverage of New and Controversial Science*, Lawrence Erlbaum Ass., Mahwah, NJ.

Funtowicz, S.O. and Ravetz, J. R. (1990): *Uncertainty and Quality in Science for Policy*, Kluwer Academic Publishers, Netherlands.

Houghton et al., (1990) in: *Climate Change, The IPCC Scientific Assessment*, Cambridge University Press, UK.

Ibrekk, H., and M.G. Morgan. 1987. *Graphical Communication of Uncertain Quantities to Nontechnical People*. Risk Analysis, Vol. 7, No. 4, pp. 519–529.

IPCC, 2001, *Climate Change 2001, IPCC Third Assessment Report*, Intergovernmental Panel on Climate Change, available from: http://www.ipcc.ch/pub/reports.htm.

IPCC, 2005: *Guidance Notes for Lead Authors of the IPCC Fourth Assessment Report on Addressing Uncertainties*, Intergovernmental Panel on Climate Change.

Kuhn, T.S. (1962): *The Structure of the Scientific Revolutions*, University of Chicago Press, Chicago.

Korving, J.L., 2004, *Sewer system; probabilistic modelling; uncertainty analysis,* Thesis, Faculty of Civil Engineering and Geosciences, TU Delft, Netherlands.

Morgan, M.G. and Henrion, M. (1992): *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.*, Cambridge University Press, UK.

Moss, R., and S. Schneider. 2000. Uncertainties, in Guidance Papers on the Cross Cutting Issues of the Third Assessment Report of the IPCC, edited by R. Pachauri, T. Taniguchi, and K. Tanaka, Intergovernmental Panel on Climate Change (IPCC), Geneva.

Patt, A. G. and Schrag, D. 2003. Using specific language to describe risk and probability. Climatic Change 61, 17-30 (2003).; and Patt, A. G. and S. Dessai. 2004. Communicating uncertainty: lessons learned and suggestions for climate change assessment. Comptes Rendu Geosciences 337, 425-441.

Rackam, N., 1989: *SPIN Selling,* McGraw-Hill, NJ.

Sluijs, van der J., Kloproggea, P., Risbeyb, J. and Ravetzc, J., 2003, *Towards a synthesis of qualitative and quantitative uncertainty assessment: Applications of the Numerical, Unit, Spread, Assessent, Pedidree (NUSAP) system.* Int. Workshop on Uncertainty, Sensitivity, and Parameter Estimation for Multimedia Environmental Modeling. Also: www.nusap.net

Sluijs, van der, J.P., *et al* (2003): RIVM/MNP Guidance for Uncertainty Assessment and Communication: Detailed Guidance"; Utrecht University and RIVM.

Taylor, B.N., Parker, W. H., and Langenberg, D. N. (1969): *Determination of e / h, Using Macroscopic Quantum Phase Coherence in Superconductors: Implications for Quantum Electrodynamics and the Fundamental Physical Constants*

Rev. Mod. Phys. 41, 375.

Torfs, P. and Warmerdam, P. (2002): *Rainfall-runoff models with Parzen regression,* Proceedings of the ERB and Northern European FRIEND Project 5 Conference, Demänovská dolina, Slovakia.

Tukey, J.W. (1977): *Exploratory Data Analysis*, Addison-Wesley, MA.

Vuren, van, B.G., 2005, *Stochastic modelling of river morphodynamics.* Ph.D Thesis, Fac. of Civil Engineering, Delft University of Technology, The Netherlands.

WL (2006): *Zicht op onzekerheden in de PKB Ruimte voor de Rivier*, WL | Delft Hydraulics report Q4198, June 2006.

Zar, J.H. (1984) *Biostatistical Analysis*. Prentice Hall International, NJ.

WL | Delft Hydraulics

Rotterdamseweg 185
postbus 177
2600 MH Delft
telefoon 015 285 85 85
telefax 015 285 85 82
e-mail info@wldelft.nl
internet www.wldelft.nl

Rotterdamseweg 185
p.o. box 177
2600 MH Delft
The Netherlands
telephone +31 15 285 85 85
telefax +31 15 285 85 82
e-mail info@wldelft.nl
internet www.wldelft.nl