On the lack of information in hydraulic engineering models

Karen A.H. Slijkhuis
Risk Analysis Department, Ministry of Public Works and Water Management, Utrecht, the Netherlands

Pieter H.A.J.M. van Gelder
Department of Civil Engineering, Delft University of Technology, Delft, the Netherlands

Han K. Vrijling
Department of Civil Engineering, Delft University of Technology, Delft, the Netherlands

Anton C.W.M. Vrouwenvelder
TNO Bouw, Delft, the Netherlands

ABSTRACT: This paper is concerned with the problems connected with the lack of information in hydraulic engineering models. The paper focusses on hydraulic engineering models that result in a probability of failure. The different types of uncertainty in such models will be discussed. Future research might give information and reduce some of the uncertainties. It will be discussed how to present the current and the uncertain future probability of failure. A simplified model which calculates the probability of failure of a dike will be used as an example.

1 INTRODUCTION

The present safety standards of the Dutch dikes and flood defences date to the report of the Delta-committee (1960) and are expressed as an exceedance frequency of the design water level. The Dutch Ministry of Public Works wants to change this policy to bring it in line with the approach in areas like planning and transport, where failure probabilities are given in a framework of acceptable risk. The transition from water level criteria towards flooding probabilities -and finally to a flood risk approach- requires a model to calculate the probability of a failure of a dike system with several boundary conditions:

1 The model should be widely accepted by the flood defence community and in some sense by the general public.

2 The results of the model should be sufficiently robust. The answers should not vary substantially with slight moderations of the input.

3 The results of the model should not actuate the decision to alter the dike systems. The dike systems are now in full compliance with the Delta-committee standard and are generally considered to be sufficiently safe.

In the past few years a model has been developed in order to determine the probability of failure of a dike system, including several failure mechanisms and their statistical dependencies. The mentioned boundary conditions to the model were approached by carrying out eight case studies of eight typical, slightly schematised polders in the Netherlands. In the case studies a comparison between the old and the new safety approach has been made, to show that the differences are minor. (condition 3) Subsequently these case studies are a good first step on the road to acception of the model. (condition 1).

A problem that had to be researched in accordance with the second condition was whether a future change in the uncertainties in the model could cause a significant change in the resulting probability of failure. To investigate this an uncertainty analysis of the model has been carried out. An uncertainty analysis can give insight into the contributions and the effects that various types of uncertainty have on the resulting probability of flooding.

The discussion of this paper will be the problems connected with an uncertainty analysis of a model with a probability of failure as a result.

The paper is organized as follows. First the different types of uncertainty will be discussed followed by the available options to present the results of an uncertainty analysis. A model for the effects of the uncertainty due to lack of information will be presented and applied on a case study. Finally the conclusions are given.

2 TYPES OF UNCERTAINTY

Uncertainties in decision and risk analysis can primarily be divided in two categories: uncertainties that stem from variability in known (or observable) populations and therefore represent randomness in samples (inherent uncertainty), and uncertainties that come from basic lack of knowledge of fundamental phenomena (epistemic uncertainty).

Inherent uncertainties represent randomness or the variations in nature. For example, even with a long history of data, one cannot predict the maximum water level that will occur in, f.i., the coming year at the North Sea. It is not possible to reduce inherent uncertainties.

Epistemic uncertainties are caused by lack of knowledge of all the causes and effects in physical systems, or by lack of sufficient data. It might be possible to obtain the type of the distribution, or the exact model of a physical system, when enough research could and would be done. Epistemic uncertainties may change as knowledge increases.

2.1 Five types of uncertainty

The inherent uncertainty and epistemic uncertainty can be subdivided in the following five types of uncertainty (Paté-Cornell 1996): inherent uncertainty in time and in space, parameter uncertainty and distribution type uncertainty (together also known as statistical uncertainty) and finally model uncertainty.

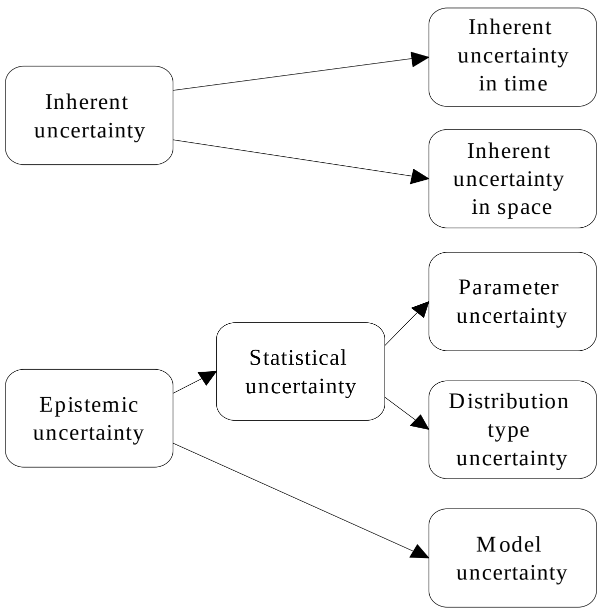

Figure 1: Types of uncertainty

i) Inherent uncertainty in time

Stochastic processes running in time such as the occurrence of water levels and wave heights are examples of the class of inherent uncertainty in time. Unlimited data will not reduce this uncertainty because the realizations of the process in the future stay uncertain.

ii) Inherent uncertainty in space

Random variables that represent the fluctuation in space, such as the dike height. Just as for inherent uncertainty in time it holds that unlimited data (e.g. if the height would be known every centimeter) will not reduce this uncertainty. There will always still be a fluctuation in space.

iii) Parameter uncertainty

This uncertainty occurs when the parameters of a distribution are determined with a limited number of data. The smaller the number of data, the larger the parameter uncertainty.

iv) Distribution type uncertainty

This type represents the uncertainty of the distribution type of the variable. It is for example not clear whether the occurrence of the water level of the North Sea is exponentially or Gumbel distributed or whether it has a completely different distribution.

Remark: a choice was made to divide statistical uncertainty into parameter- and distribution type uncertainty although it is not always possible to draw the line; in case of unknown parameters (because of lack of observations), the distribution type will be uncertain as well.

Later in this paper we will try to distinguish what type of uncertainty a stochastic parameter represents. Since parameter uncertainty and distribution type uncertainty can not be discerned, another practical -less scientific- division has been chosen. The statistical uncertainty is divided in two parts: 'statistical uncertainty of variations in time' and 'statistical uncertainty of variations in space'.

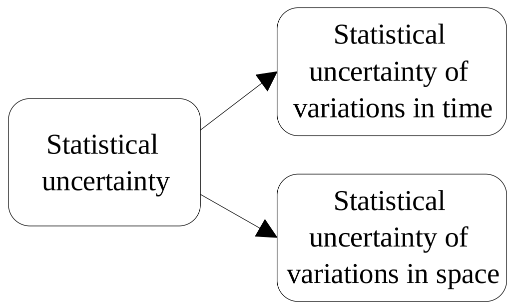

Figure 2: Statistical uncertainty

v) Statistical uncertainty of variations in time

When determining the probability distribution of random variable that represents the variation in time of a process (like the occurrence of a water level), there essentially is a problem of information scarcity. Records are usually too short to ensure reliable estimates of low-exceedance probability quantiles in many practical problems. The uncertainty caused by this shortage of information is the statistical uncertainty of variations in time. This uncertainty can theoretically be reduced by keeping record of the process for the coming centuries.

vi) Statistical uncertainty of variations in space

When determining the probability distribution of random variable that represents the variation in space of a process (like the fluctuation in the height of a dike), there essentially is a problem of shortage of measurements. It is usually too expensive to measure the height or width of a dike in great detail. This statistical uncertainty of variations in space can be reduced by taking more measurements (see also Vrijling and Van Gelder, 1998).

vii) Model uncertainty

Many of the engineering models that describe the natural phenomena like wind and waves are imperfect. They may be imperfect because the physical phenomena are not known (for example when regression models without underlying theory are used), or they can be imperfect because some variables of lesser importance are omitted in the engineering model for reasons of efficiency.

2.2 Reduction of uncertainty

Before was mentioned that inherent uncertainties represent randomness or the variations in nature. Inherent uncertainties cannot be reduced.

Epistemic uncertainties, on the other hand, are caused by lack of knowledge. Epistemic uncertainties may change as knowledge increases.

In general there are three ways to increase knowledge:

- Gathering data
- Research
- Expert-judgment

Data can be gathered by taking measurements or by keeping record of a process in time. Research can f.i. be done into the physical model of a phenomenon or into the better use of existing data. By using expert opinions it is possible to acquire the probability distributions of variables that are too expensive or practically impossible to measure.

The goal of all this research obviously is to reduce the uncertainty in the model. Nevertheless it is also thinkable that uncertainty will increase. Research might show that an originally flawless model actually contains a lot of uncertainties. Or after taking some measurements the variations of the dike height can be a lot larger. It is also thinkable that the average value of the variable will change because of the research that has been done.

The consequence is that the calculated probability of failure will be influenced by future research. In order to guarantee a stable and convincing flood defence policy after the transition, it is important to understand the extent of this effect.

In the next chapter will be discussed what influence the future reduction of uncertainty can have on the probability of failure and how to present this influence. It is assumed that in practice research will rarely increase uncertainty or change the average value of a variable.

3 THE EFFECT OF UNCERTAINTY DUE TO LACK OF INFORMATION

The problem of lack of information in hydraulic engineering models is studied in detail in this chapter. Data records are usually too short to ensure reliable estimates of low-exceedance probability quantiles in many practical problems (Noortwijk and Van Gelder, 1998).

The information about a random variable can be updated by the help of expert judgements. Cooke (1991) describes various methods to do so. Expert judgements can also be used to reduce the uncertainty of the quantiles. In order to include expert judgements in the quantile estimation of a certain quantity, the following approach is proposed in this paper. The approach is applied on a case study of quantile estimation of water levels.

Consider R the random variable describing the river level with exceedance probability p per year ($p \ll 1$).

Consider H the random variable describing the height of the dike modelled with a normal distribution with mean μ_H and standard deviation σ_H.

The effect of the value of information on the random variable R may be modelled by correcting its original mean value μ_R to its new value $\mu_R + v\sigma_I$ in which v is the standard normal distribution and σ_I is the standard deviation of the expert information. Furthermore the standard deviation of R will be reduced from $\sqrt{(\sigma_R^2+\sigma_I^2)}$ to σ_R under the influence of the expert. Summarized in a table:

Without Information			With Information	
	μ	σ	μ	σ
R	μ_R	$\sqrt{(\sigma_R^2+\sigma_I^2)}$	$\mu_R+ v\sigma_I$	σ_R
H	μ_H	σ_H	μ_H	σ_H
$\beta_{ni}=(\mu_R-\mu_H)/\sqrt{(\sigma_R^2+\sigma_I^2+\sigma_H^2)}$			$\beta_{wi}=(\mu_R+v\sigma_I-\mu_H)/\sqrt{(\sigma_R^2+\sigma_H^2)}$	

Table 1: The effect of information on the random variables R and H.

The exceedance probability or reliability index β_{wi} after including expert opinion can be seen as a random variable with a normal distribution with the following mean and standard deviation:

$$\beta_{wi}\sim N\left((\mu_R-\mu_H)/\sqrt{(\sigma_R^2+\sigma_H^2)},\ \sigma_I/\sqrt{(\sigma_R^2+\sigma_H^2)} \right)$$

Using the notation $\beta_m = (\mu_R-\mu_H)/\sqrt{(\sigma_R^2+\sigma_H^2)}$, and $\sigma_\beta = \sigma_I/\sqrt{(\sigma_R^2+\sigma_H^2)}$, β_{wi} can be written as:

$$\beta_{wi} = \beta_m + v\,\sigma_\beta$$

in which v is the standard normal distribution.

In order to determine the uncertainty in the probability of failure and its influence factors, a FORM analysis can be performed. The following reliability function is therefore considered:

$$Z = \beta_{wi} - u$$

In which u is a standard normal distribution (indepedent of v). Together with the expression for β_{wi} the reliability function can be seen as a function of the 2 standard normal distributions u and v:

$$Z = \beta_m + v\,\sigma_\beta - u$$

Because $\partial Z/\partial u = -1$ and $\partial Z/\partial v = \sigma_\beta$, the reliability index for this Z-function can be derived to:

$$\beta = \beta_m / \sqrt{(1+\sigma_\beta^2)}$$

which appears to be exactly the same as the reliability index without information: β_{ni} (the proof is given in the Appendix).

The α-factors are given by:

$$\alpha_u = 1/\sqrt{(1+\sigma_\beta^2)} \text{ and } \alpha_v =\sigma_\beta /\sqrt{(1+\sigma_\beta^2)}$$

α_u and α_v represent the influence of the uncertainties in u and v on the reliability index β respectively. In terms of β_m and σ_β this can also be written as:

$$\beta_m = \beta / \alpha_u$$

$$\sigma_\beta = \sqrt{(1 - \alpha_u^2)} / \alpha_u = \alpha_v / \alpha_u$$

The implications of the changes in β and the related flooding frequency will be discussed in chapter 5.

4 OPTIONS FOR PRESENTATION

As discussed in paragraph 2.2 it is possible that some uncertainties might be reduced in the future. This means that the future probability of failure will be smaller. One could say that those uncertainties add to the ‘uncertainty’ of the probability of failure.

The obvious problem is that it can not be predicted which uncertainties will be reduced in the future. There are several philosophies thinkable about what uncertainty will remain part of the future probability of failure.

It is important to realise that, for every philosophy, the difference between the ‘true’ probability of failure and the ‘uncertainty’ of the probability of failure, is an artificial difference. If all uncertainties are integrated this will result in the same, current probability of failure.

Let us consider four different philosophies (options) about the uncertainties that might be reduced and form the ‘uncertainty’ of the probability of flooding, and the uncertainties that cause the ‘true’ probability of flooding, as given in table 2:

	Option 1	Option 2	Option 3	Option 4
Inherent uncertainty (in time)	P_f	P_f	P_f	P_f
Inherent uncertainty (in space)	P_f	P_f	P_f	-
Statistical uncertainty (of variations in time)	P_f	P_f	-	-
Statistical uncertainty (of variations in space)	P_f	-	-	-
Model uncertainty	P_f	-	-	-

Table 2: Four different options

Option 1

No incertainties will be reduced. All uncertainties in the model will be integrated to determine the probability of failure. This probability of failure also represents the current, actual probability of failure.

Option 2

A practical division is made between the uncertainties that might be influenced by taking measurements or doing research and the remaining uncertainties. Statistical uncertainty of variations in time will be a part of the probability of failure. This uncertainty can only theoretically be reduced by keeping record of the underlying process for the coming centuries.

Option 3

A theoretical division is made between inherent uncertainty and epistemic uncertainty. The assumption is that through research all epistemic uncertainty disappears. This would result in a probability of failure that is solely caused by inherent uncertainties.

Option 4

Assumed is that only inherent uncertainty in time causes the probability of failure. This is also the philosohy used in the present safety standards.

5 EXAMPLE

A simplified model which calculates the probability of failure of a dike, will be used to demonstrate the use of the theory of chapter 3 and 4.

$$Z = m \cdot (h_d - h_w)$$

m = modelfactor [-] : m = N(1; 0.5)
h_d = dike height [m +NAP] : h_d =N(5; 0.25)
h_w = water level [m +NAP] : h_w =N(3.5; σ_{hw})
σ_{hw} = distribution parameter of h_w: σ_{hw}=N(0.5;0.05)

It is possible to identify the type(s) of uncertainty that every variable represents (Table 3).

m	model uncertainty
h_d	inherent and statistical uncertainty (space)
h_w	inherent uncertainty (time)
σ_{hw}	statistical uncertainty (time)

Table 3: The variables and their uncertainties

The FORM analysis of this reliability function results in a probability of failure of 2.7 x 10^{-4} and a reliability index of 3.46. The influence factors of the variables are presentated in Table 4.

	α_i^2
h_w	0.77
σ_{hw}	0.06
h_d	0.16
m	0.01

Table 4: The influence factors

In the next table 5 is shown per option of presentation and per parameter whether or not the uncertainty is part of the probability of failure. It is assumed that the uncertainty in the dike height represents exactly halfly the inherent uncertainty and halfly the statistical uncertainty.

	Option 1	Option 2	Option 3	Option 4
h_w	1	1	1	1
σ_{hw}	1	1	-	-
h_d	1	0.5	0.5	-
m	1	-	-	-

Table 5: The four options of the case study

Let p_i be the values in table 5 then with $\alpha_u^2 = \Sigma\, p_i.\ \alpha_i^2$ and with the equations given in chapter 3, it is quite simple to obtain the results per option as displayed in the tabel below. The results are also presented in the figure 3.

	Option 1	Option 2	Option 3	Option 4
α_u^2	1	0.91	0.85	0.77
α_v^2	0	0.09	0.15	0.23
β_m	3.46	3.62	3.75	3.94
$\sigma\beta$	0	0.31	0.42	0.55

Table 6: Results for the four options

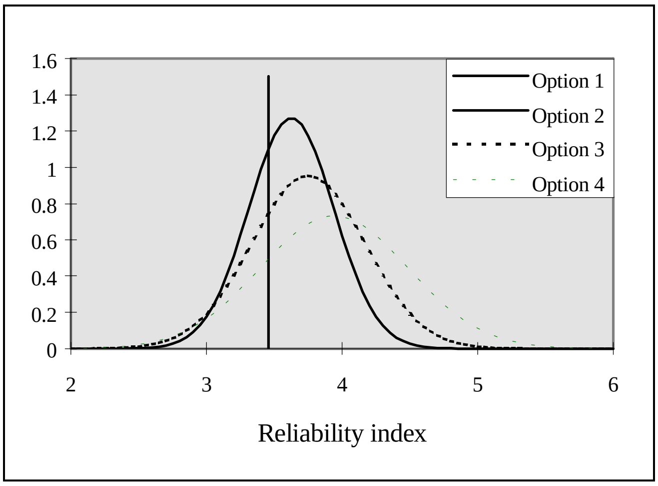

Figure 3: Probability distributions of β

In figure 3 the current probability of failure is shown in four different ways. Option 1 gives that probability with no uncertainty. With every next option, more uncertainty is removed from the 'true' probability of failure and added to the 'uncertainty' of the probability of failure.

It is clear that the more uncertainty is removed from the resulting probability of failure; the higher the mean and the larger the standard deviation of the distribution of the reliability index will be.

In table 4 can be seen that the uncertainty in the occurence of high water levels has the largest influence on the probability of failure. Since this uncertainty can not be reduced, this will always be part of the probability of failure. Therefore the mean of the distribution of the reliability index has not shifted considerably.

The model used for the present safety standards is deterministic, but for the random occurence of the water levels. This resembles option 4. The difference is that the result of the model is equal to the mean of the reliability index. This creates a seeming difference between the probability of flooding computed with the 'old' and the 'new' model (see chapter 1).

This causes a problem for the transition from the water level criteria towards flooding probabilities. It is still not decided how to deal with this seeming difference. There are three possible reactions:

1 Accept the difference and do nothing.

2 Heighten the dikes in order to lower the 'new' probability of flooding to the 'old' value.

3 Do research before the transition takes place in order to reduce some uncertainties and close the gap between the 'old' and the 'new' probability of flooding.

6 CONCLUSIONS

It can be concluded that the method described in this paper is a very practical and simple method to get insight in the problems concerned with the lack of information in hydraulic engineering models.

It shows that the more uncertainty is expected to be reduced in the future, the higher the mean and the larger the standard deviation of the distribution of the reliability index will be.

The method is demonstrated for a simple reliability function, but the advantage is that the method also works for a very complex reliability function with a large number of variables.

REFERENCES

Cooke, Roger M.. *Experts in uncertainty; opinion and subjective probability in science.* Publisher New York : Oxford University Press, 1991 ISBN 0-19-506465-8.

J.M. van Noortwijk, P.H.A.J.M. van Gelder, *Bayesian estimation of quantiles for the purpose of flood prevention*, ICCE 98: 26th International Conference on Coastal Engineering, 22 - 26 June 1998, Copenhagen, Denmark.

Vrijling, J.K. et.al.,*Framework for risk evaluation,* Hazardous Materials 43 ,1995.

Delta Committee, (1960) *Considerations on storm surges and tidal waves* (in Dutch), Rapport Deltacommissie, Deel 3, Staatsdrukkerij, Den Haag.

Jorissen, R.E., Vrijling, J.K., (1998), The transition to a risk-based concept for dike design in the Netherlands, ICCE, Copenhagen, Denmark.

Paté-Cornell, M.E. 1996. *Uncertainties in risk analysis; Six levels of treatment.* Reliability Engineering and System Safety. Nr. 54. Northern Ireland: Elsevier Science Limited.

Vrouwenvelder, A.C.W.M. 1997. *Project Onzekerheidsanalyse Mechanisme Overslag/ Overloop* (Project Uncertainty-analysis (in Dutch). Delft: TNO 97-CON-R0935.

J.K. Vrijling and P.H.A.J.M. van Gelder, *The effect of inherent uncertainty in time and space on the reliability of flood protection*, ESREL'98: European Safety and Reliability Conference 1998, pp.451-456, 16 - 19 June 1998, Trondheim, Norway.

K.A.H. Slijkhuis, M.P.C. Frijters, R.M. Cooke, A.C.W.M. Vrouwenvelder, *Probability of Flooding: An uncertainty analysis,* ESREL'98: European Safety and Reliability Conference 1998, pp.1419-1425, 16 - 19 June 1998, Trondheim, Norway.

APPENDIX

By definition:

$$\beta_m = (\mu_R-\mu_H)/\sqrt{(\sigma_R^2+\sigma_H^2)},$$

$$\sigma_\beta = \sigma_I/\sqrt{(\sigma_R^2+\sigma_H^2)},$$

So:

$$\beta = \beta_m / \sqrt{(1+\sigma_\beta^2)} = (\mu_R-\mu_H)/\sqrt{(\sigma_R^2+\sigma_H^2)}/\sqrt{(1+\sigma_\beta^2)}=$$

$$= (\mu_R-\mu_H)/\sqrt{(\sigma_R^2+\sigma_H^2)}/\sqrt{(1+ \sigma_I^2/(\sigma_R^2+\sigma_H^2))} =$$

$$= (\mu_R-\mu_H)/\sqrt{(\sigma_R^2+\sigma_I^2+\sigma_H^2)} = \beta_{ni} \qquad \square$$