

Bayesian decision analysis as a tool for defining monitoring needs in the field of effects of CSOs on receiving waters.

Hans Korving and François Clemens

Department of Environmental Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology, P.O.Box 5048, 2600 GA, Delft, the Netherlands. (E-mail: j.l.korving@citg.tudelft.nl; f.h.l.r.clemens@citg.tudelft.nl)

Abstract In recent years, decision analysis has become an important technique in many disciplines. It provides a methodology for rational decision-making allowing for uncertainties in the outcome of several possible actions to be undertaken. An example in urban drainage is the situation in which an engineer has to decide upon a major reconstruction of a system in order to prevent pollution of receiving waters due to combined sewer overflows (CSOs). This paper describes the possibilities of Bayesian decision-making in urban drainage. In particular, the utility of monitoring prior to deciding on the reconstruction of a sewer system to reduce CSO emissions is studied. Our concern is with deciding whether a price should be paid for new information and which source of information is the best choice given the uncertainties in the outcome. The influence of specific uncertainties (sewer system data and model parameters) on the probability of CSO volumes is shown to be significant. Using Bayes' rule, to combine prior information with new observations, reduces the risks linked with the planning of sewer system reconstructions.

Keywords Bayesian decision-making, CSO reduction, monitoring, sewer system, uncertainties.

Introduction

Since CSOs may cause deterioration of receiving water quality, it is generally accepted that their influence should be reduced. One obvious intervention to reduce effects of CSOs is to reconstruct a sewer system in such a way that CSO emissions diminish. Reconstruction of sewer systems however, demands major investments. The amount of money invested in sewer systems is large and will remain large in the future. As decisions on reconstruction investments are taken under substantial uncertainties, the effectiveness of investments in sewer systems may be questioned. For example, in the Netherlands a number of examples are known in which the reconstructions did not have the desired effect. The interventions turned out to be either too small, too large or even unnecessary. In current practice, decisions on investments in sewer systems meant to reduce CSO emissions need to be based on available measurement data that may be an incomplete and uncertain information source. So-called Bayesian decision-making provides opportunities to solve the problem of imperfect or unreliable information. The ability to extract as much information as possible from available data is typical of Bayesian decision-making. Due to this ability the risk related to planned investments is reduced.

The main aim of this paper is to show how to determine whether it is useful or not to carry out measurements prior to deciding on the reconstruction of a sewer system. For this purpose, the ratio of costs of measurements to reduced risk of inappropriate reconstruction as a result of the measurement outcomes is of importance. The technique of Bayesian decision-making is illustrated with a case study, in which the problem of choosing whether or not to start a measurement campaign to support decision-making on reconstructions of the sewer system is discussed.

Decision-making under uncertainty

Decisions on monitoring activities related to proposed interventions are made under uncertainties. Optimal decisions can be obtained using Bayesian decision theory. Such decisions are based on Bayes' theorem (Bayes, 1763) enabling the quantification of the probability of an event on the basis of a prior estimate of its probability and new observations. In other words, Bayes rule updates subjective beliefs on the occurrence of an event based on new data (Figure 1). The theory of Bayesian decision-making is described in more detail by many authors (e.g. Benjamin and Cornell (1970) and Pratt *et al.* (1995)). Bayesian decision theory is applied to several kinds of decision problems under uncertainty. In this article measurements prior to investments in a sewer system are studied.

Bayes' theorem, i.e. the conditional probability theorem, can be written as

$$\pi(\theta | x) = \frac{\ell(x | \theta)\pi(\theta)}{\int_{\Theta} \ell(x | \theta)\pi(\theta)d\theta} = \frac{\ell(x | \theta)\pi(\theta)}{\pi(x)}$$

in which,

$\pi(\theta | x)$ = posterior density of $\theta = (\theta_1, \dots, \theta_d)$ after observing data $x = (x_1, \dots, x_n)$,

$\ell(x | \theta)$ = likelihood function of observations $x = (x_1, \dots, x_n)$ when parameter $\theta = (\theta_1, \dots, \theta_d)$ is known,

$\pi(\theta)$ = prior density of $\theta = (\theta_1, \dots, \theta_d)$ before observing data $x = (x_1, \dots, x_n)$,

$\pi(x)$ = marginal density of observations $x = (x_1, \dots, x_n)$.

Using Bayes' theorem a prior distribution can be updated as soon as new observations are available. The more new observations are used, the smaller the parameter uncertainty in θ .

Although collecting data is possible prior to each decision on investments in a sewer system, the possibility to get more information about the system's behaviour is rarely used. Data collection prior to decisions on reconstructions can be used to prevent disinvestments; e.g. an unnecessary, too large or too small storm water settling tank. Monitoring reduces the risk on disinvestments. If monitoring costs are sufficiently small compared to the possible reduction of risk for disinvestments, monitoring prior to decisions on investments is worth the trouble. For this purpose Bayesian decision-making can be used (Korving, 2001). For example, Bayes' rule has been applied on decisions on the reduction of flood impacts (Van Noordwijk et al., 1997), analysis of extreme river discharges (Chbab et al., 2000) and risk-based design of civil structures (Van Gelder, 2000).

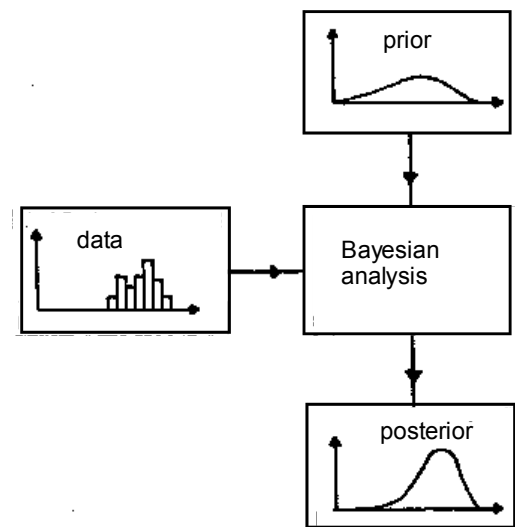


Figure 1 Bayesian analysis. The theorem of Bayes (1763) provides a solution to the problem of how to learn from new data

Decision tree

A decision-making process can be graphically displayed as a decision tree (Figure 2). The decision problem can be interpreted as a game between the decision maker and a fictitious character called "chance". The game comprises four moves. In move 1 the decision maker chooses a measurement from a set of possible measuring activities (e_0 (no measurement), e_1, e_2, \dots). Subsequently, "chance" reacts with a measuring result (z_1, z_2, \dots) in move 2. The decision maker has no influence on this outcome. In move 3 the decision maker chooses an action or an intervention from a set of possible actions or interventions A (a_1, a_2, \dots). In the final move "chance" chooses a state of nature ($\theta_1, \theta_2, \dots$). On this choice the decision maker has no influence also. The consequence of the sequence of choices by the decision maker and "chance" is put together in a so-called utility $u(e, z, a, \theta)$.

E = set of possible measuring activities (e_0 (no measurement), e_1, e_2, \dots).

Z = set of possible outcomes of a measurement activity, e_i (z_1, z_2, \dots).

A = set of possible interventions available to decision maker (a_1, a_2, \dots).

Θ = set of possible states of nature ($\theta_1, \theta_2, \dots$).

Loss or utility function

Decisions on interventions that reduce CSO emissions must be made under uncertainty and can be obtained by using decision theory. As discussed before, a decision problem is a problem in which the decision maker has to choose an action a from the set of all possible decisions A . Optimal decisions meet

the criterion of minimal expected loss, where the expectation is calculated with respect to the unknown parametric vector θ .

Let $L(\theta, a)$ be the loss when the decision maker chooses decision a and the state of nature turns out to be θ . The objective is to minimise $L(\theta, a)$. In the case of sewer system reconstructions related to CSO reduction this loss function consists of (1) (re)construction costs $c(\theta, a)$, (2) loss due to use of space by construction $s(\theta, a)$ and (3) damage due to overflows $o(\theta, a)$. Hence the loss can be written as,

$$L(\theta, a) = c(\theta, a) + s(\theta, a) + o(\theta, a).$$

Given the loss function the decision maker can best choose the decision with minimal expected loss. A loss function is also called utility function. One way to formulate a loss function is by capitalising all losses. For some losses it is difficult to translate them into terms of money, e.g. the environmental damage caused by CSOs. Moreover, determination of losses is subjective, which influences the whole decision-making process. This subject demands further research.

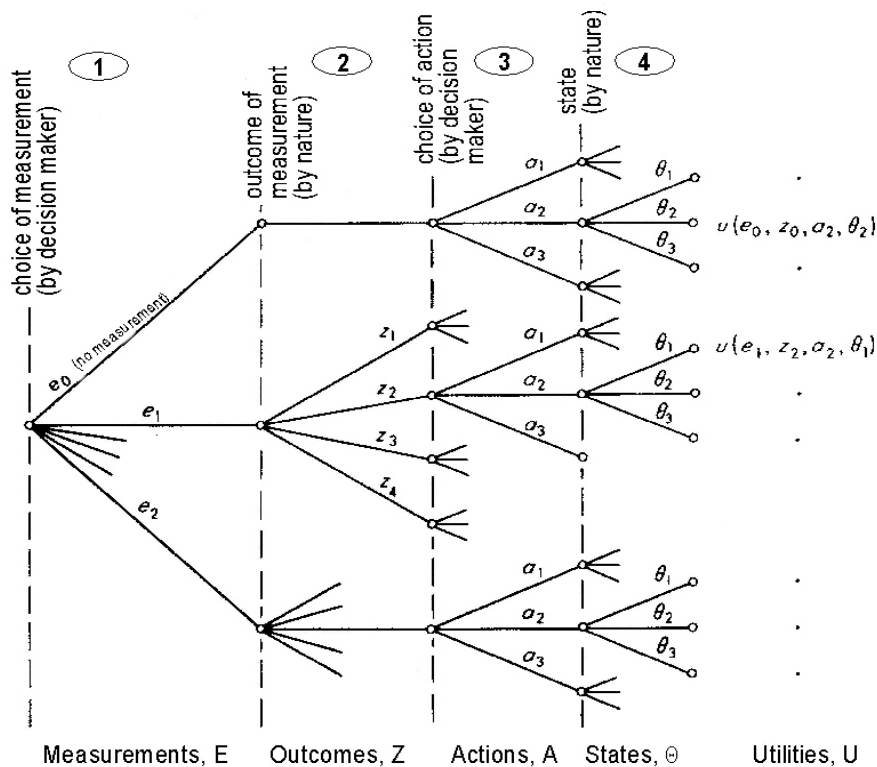


Figure 2 Decision tree (after: Benjamin and Cornell, 1970). Our concern is with deciding whether a price should be paid for new information and which source of information is the best choice.

Uncertainties

In general, the risks evolving from a decision are determined by the uncertain state of the world, θ . For CSO emissions the world is characterised by (random) variables including precipitation, dimensions of the sewer system (Clemens and Von der Heide, 1999) and resilience of the receiving water body.

The main objective of this paper is to find out if additional measuring results will reduce uncertainties in a decision-making process. By determining the joint probability density function of the above-mentioned random quantities the decision problem can be formulated. Uncertainties are either related to the randomness or variability in nature (inherent uncertainties) or related to the lack of knowledge about a (physical) system (epistemic uncertainties) (Van Gelder, 2000). Uncertainties influencing decisions on interventions in a sewer system comprise uncertainties in: (1) precipitation, (2) water levels in sewers, (3) exceedance of discharges, (4) dimensions of sewer system, (5) model parameters in hydrodynamic model, (6) prior information on precipitation and CSO volumes, (7) construction costs, (8) environmental damage caused by CSOs.

The relations of the uncertainties are shown in Figure 3. A hydrodynamic model is used to quantify the probability distribution of the design parameter (i.e. CSO volumes), since usually no series of observed system behaviour are available over a long enough period (several years). Modelling reduces the

uncertainties in the knowledge of the states of nature, the CSO volumes. If enough data on design parameters were available, modelling would not be necessary to this end.

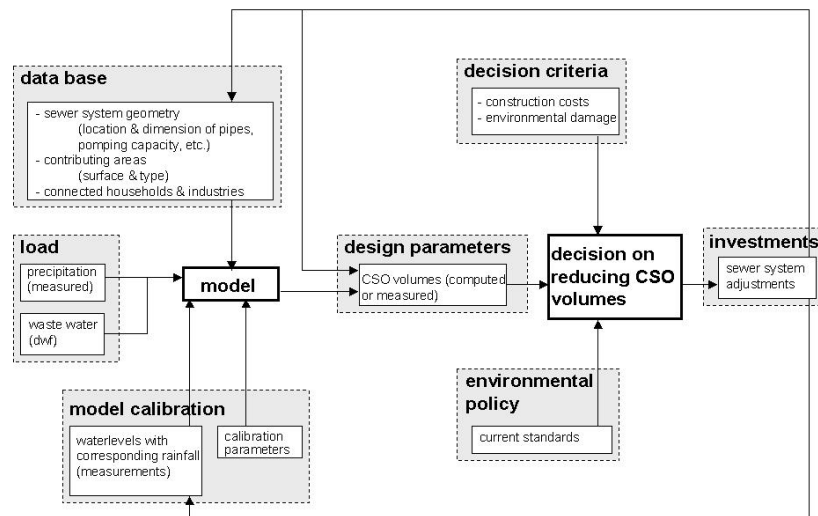


Figure 3 Uncertainties influencing decisions on interventions in a sewer system.

Application of Bayesian decision-making in CSO emission reduction

The Bayes' theorem can be used in the assistance of decision-making on reduction of CSO emissions by planning interventions in the sewer system (e.g. SST, RTC, etc.). In particular, the use of monitoring prior to decisions on interventions is studied. As an example a catchment area called 'De Hoven' is used. This catchment (12.69ha) is situated in the Netherlands on the banks of the river IJssel in the city of Deventer (Figure 4). The sewer system (865m³) is of the combined type and comprises one pumping station (119m³/h) transporting the sewage to a treatment plant and three CSO structures. As input of the computations a 10 year rainfall series (1955-1964) of KNMI is used.

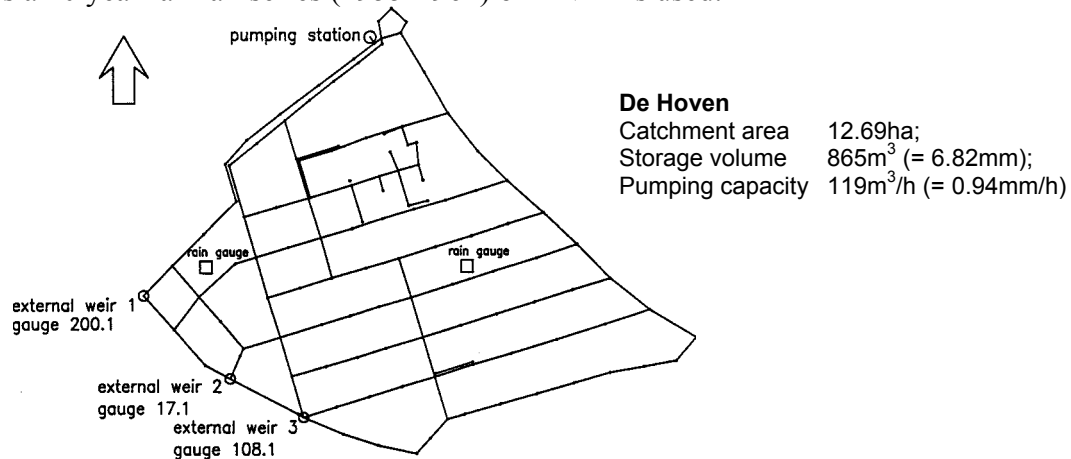


Figure 4 Layout sewer system 'De Hoven' (from: Clemens, 2001).

The case study 'De Hoven' is on deciding on the volume of a storage tank to be built in order to reduce CSO emissions, in particular on choosing to effect monitoring prior to the decision on interventions. The goal of the measuring activities is to reduce the risk on disinvestment related to building a storage facility with a certain volume. Alternative decisions can be weighed using Bayes' rule. Here three aspects of Bayesian decision-making are singled out. Firstly, the decision process is translated into a decision tree with discrete probabilities to clearly illustrate the process of decision-making. Secondly, uncertainties in the data on sewer system dimensions are examined by looking at their influence on the probability distribution of CSO volumes. Thirdly, uncertainties in the parameters of a calibrated hydrodynamic model are examined by performing Monte Carlo simulation based on the probability distribution of the residues

(differences between calibrated model and measuring results). The last step makes it possible to study the additional value of calibration measurements.

Decisions on monitoring prior to the planning of interventions translated to a simple decision tree

For the sake of simplicity in the explanation, the decision problem is simplified to a choice between building 1mm or 2mm additional storage volume in the sewer system. Furthermore, nature can take either one of the two states 1mm or 2mm required storage volume. The prior probabilities of possible states of nature are shown in Table 1.

In case of discrete probabilities the Bayes' rule becomes,

$$P[\theta_i | z_k] = \frac{P[z_k | \theta_i]P[\theta_i]}{\sum_j P[z_k | \theta_j]P[\theta_j]}$$

in which,

- $P[z_k | \theta_i]$ = probability of monitoring result z_k when state of nature θ_i is known,
- $P[\theta_i]$ = prior probability of θ_i before observing data z_k ,
- $P[\theta_i | z_k]$ = posterior probability of θ_i after observing data z_k ,
- $\sum_j P[z_k | \theta_j]P[\theta_j]$ = marginal density of observations z_k given all states of nature θ_j .

The utilities of the two alternative actions are summed up in Table 2. Since utilities are interpreted as risks, negative values are used. Table 2 shows that building too small a storage facility (still CSO emissions) is penalised heavier than building too large a facility (too large area used).

Table 2 Relative utilities (in units) of alternative actions (= $u(e,z,a,\theta_i)$)

		Decision	
		a ₁ : storage volume 1mm	a ₂ : storage volume 2mm
State of nature	θ ₁ : required volume 1mm	0	-10
	θ ₂ : required volume 2mm	-40	0

Table 3 Probability of monitoring results (= $P[z_k|\theta_j]$)

		State of nature	
		θ ₁ : required volume 1mm	θ ₂ : required volume 2mm
Outcome	z ₁ : monitoring result 1mm	0.6	0.3
	z ₂ : monitoring result 2mm	0.4	0.7

The decision maker may choose between either effecting specific monitoring activities or no monitoring at all. The relative costs of the monitoring amounts to -1.5 units and are added to the above-mentioned relative utilities. Table 3 summarises the probabilities of monitoring outcomes, including for example the accuracy of the measuring equipment.

The decision-making is summarised in a decision tree (Figure 5). The tree is analysed from right to left. The posterior probabilities are computed with,

$$P[\theta_i | z_k] = \frac{P[z_k | \theta_i]P[\theta_i]}{\sum_{j=1}^2 P[z_k | \theta_j]P[\theta_j]}$$

The posterior probabilities in the case study are computed with the probabilities from Table 1 and Table 3,

$$P[\theta_1 | z_1] = 0.88; \quad P[\theta_1 | z_2] = 0.67;$$

$$P[\theta_2 | z_1] = 0.12; \quad P[\theta_2 | z_2] = 0.33.$$

Afterwards, the expected utilities per alternative action are computed with,

$$E[u(e, z, a) | e, z] = \sum_i u(e, z, a, \theta_i) P[\theta_i | z],$$

and the action with minimal loss is chosen,

$$u^*(e, z) = \max_a [E[u(e, z, a) | e, z]].$$

The expected utilities per action are (using the probabilities computed above and the utilities in Table 2),

$$u^*(e_0) = -7.8; \quad u^*(e_1, z_1) = -8.3; \quad u^*(e_1, z_2) = -6.3.$$

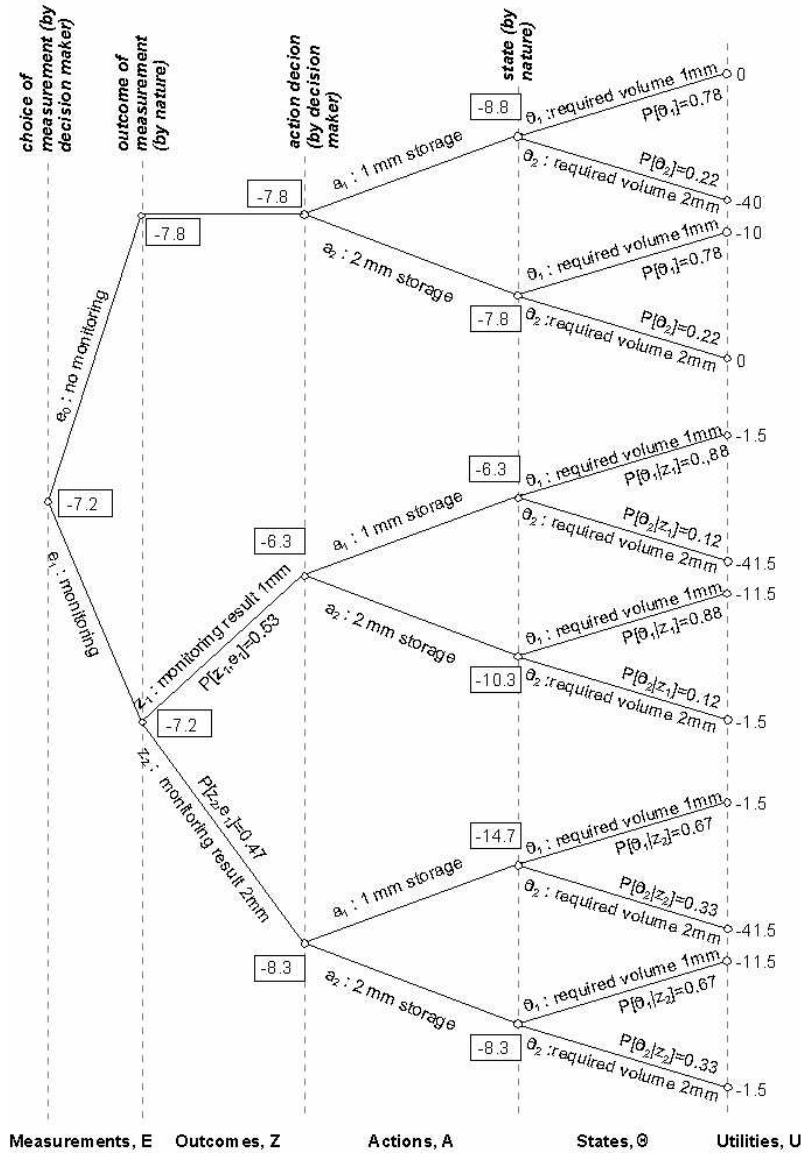


Figure 5 Example of decision model in which the choice is made between monitoring or no monitoring. In this particular case, given costs and probabilities, it appears to be worthwhile to carry out measurements prior to decisions on investments in the sewer system in order to reduce the influence of overflows on receiving waters, since $E[u(e_1)] = -7,2 > E[u(e_0)] = -7,8$.

For further analysis, the prior probabilities of the outcomes of a monitoring activity need to be known. The estimated probabilities of monitoring outcomes prior to the actual monitoring are,

$$P[z_k, e] = \sum_{i=1}^2 P[z_k | \theta_i] P[\theta_i].$$

So, the prior probabilities of the outcomes of measurement activity e_1 are computed with the probabilities in Tables 1 and 3,

$$P[z_1, e_1] = 0,53; \quad P[z_2, e_1] = 0,47.$$

Finally, the expected utilities of either ‘monitoring’ or ‘no monitoring’ are,

$$E[u(e)] = \sum_k u^*(e, z_k) P[z_k, e].$$

Since the optimal decision results in the minimal risk, the optimal choice for the decision maker is to effect the measurements e_1 prior to taking a decision on the volume of storage capacity to be built,

$$E[u(e_1)] = -7.2 > E[u(e_0)] = -7.8.$$

This example is strongly simplified. To be able to use Bayesian decision-making in practice more insight in the probabilities of states of nature θ should be obtained. Besides, the prior probabilities of measurement outcomes should be known in more detail. Therefore, in the next two paragraphs the influence of uncertainties on the probabilities of CSO volumes is studied, in particular uncertainties in sewer system data and in results of calibrated hydrodynamic models.

Uncertainties caused by variabilities in data of sewer system dimensions

Uncertainties in data of the sewer system dimensions have implications for the risks related to decisions on interventions in the sewer system. The influence of these uncertainties on the probability of CSO volumes is studied by modelling the sewer system of ‘De Hoven’ as a reservoir with an external weir and a pump. The probabilities of CSO volumes make up the states of nature θ .

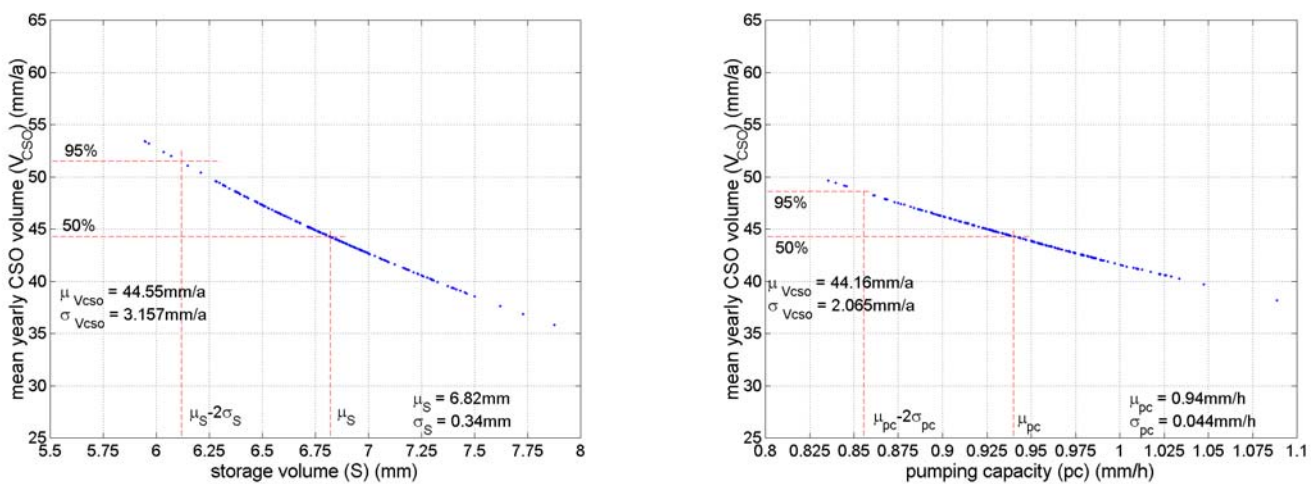


Figure 6 Influence of uncertainties in dimensions on probability distribution of CSO volume.

The influence of variabilities in four system dimensions is studied: storage volume, pumping capacity, contributing area and overflow coefficient. These sewer system dimensions are assumed to be normally distributed with known μ and σ ($\sigma=0.05*\mu$). Mean and standard deviation are based on expert judgement. Figure 6 shows that both a variability in pumping capacity and a variability in storage volume influence variabilities in yearly CSO volumes, but their influences differ. Since, a variability of 5% in pumping capacity results in a coefficient of variation of 4.7% and the same variability in storage capacity in a coefficient of variation of 7.1%. When combined with the variabilities in the other two system parameters a good estimation of the probability distribution of yearly overflow volumes prior to monitoring can be made.

Uncertainties in model parameters of a hydrodynamic model

Since 50 years of data on CSO volumes to compute the probability distribution are not available, hydrodynamic models are used to generate data on this performance parameter (Figure 3). Using a hydrodynamic model requires calibration.

The confidence intervals of the probabilities of CSO volumes are computed with Monte Carlo simulations in which the probability distribution for the residues of the calibrated hydrodynamic model of the sewer system of ‘De Hoven’ is used (Figure 7). Randomly, values of the residues are taken from the distribution after which CSO volumes are computed with a reservoir model using a long-term rainfall series (precipitation measurements of KNMI).

The results are shown in Figure 8. The figure shows a decrease in standard deviation of estimated yearly overflow volumes due to the additional information in the modelling results. The results of this analysis can be used in a decision tree to determine whether or not calibration measurements are opportune prior

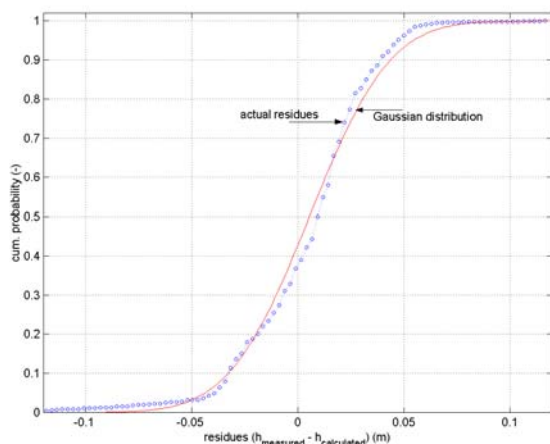


Figure 7 Probability distribution for the residues obtained after calibration of the hydrodynamic model (adapted from: Clemens, 2001). These results are used as input of the Monte Carlo simulation with the reservoir model.

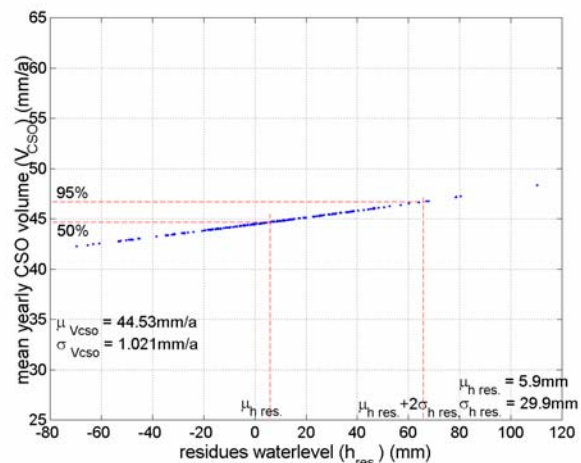


Figure 8 Influence of uncertainties in outcomes of calibrated model on probability distribution of CSO volume.

to decisions on interventions in the sewer system. The results of Monte Carlo simulations together with a loss function make decisions on monitoring activities possible. The monitoring aims at reducing the risk in decisions on CSO emission reduction.

Conclusions

The Bayesian approach offers opportunities for decision-making on CSO emission reduction. In this approach prior beliefs on the state of nature are combined with new observations of the system behaviour. The case study ‘De Hoven’ shows that uncertainties in data of the dimensions of a sewer system influence the probability distribution of CSO volumes. Uncertainties are reduced if results of a calibrated model are used. This results in significantly diminishing risks linked to decisions on investments in a sewer system. The paper describes the use of Bayes’ rule to decide on measurements prior to planning of interventions. The presented approach should be extended with continuous probability distributions instead of a set of discrete probabilities. For this purpose Monte Carlo simulations are used. Application of Bayes’ rule for decisions on monitoring prior to planning of sewer system reconstructions to reduce effects of CSOs looks promising. Further research into the subject is necessary however, to obtain methods that can be applied in practice.

Acknowledgements

The authors would like to thank HKV Consultants and the RIONED Foundation for financially supporting the research.

References

- Bayes T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, **53**, 370-418.
- Benjamin J.R. and C.A. Cornell (1979). *Probability, Statistics and Decision for Civil Engineers*. McGraw-Hill, New York.
- Chbab E.H., J.M. Van Noortwijk and M.T. Duits (2000). Bayesian frequency analysis of extreme river discharges. In: Toensmann F. and M. Koch (eds.). *River Flood Defence, Proceedings of the International Symposium on Flood Defence*. Sept. 2000, Kassel, Germany. F51-F60. Kassel, Herkules Verlag.
- Clemens F.H.L.R. and W. Von der Heide (1999). Effect of geometrical errors in hydrodynamic calculation in urban drainage. *Proceedings of the 8th ICUSD*. Aug. 1999, Sydney, Australia, 955-963.
- Clemens F.H.L.R. (2001). *Hydrodynamic Models in Urban Drainage: Application and calibration*. PhD thesis. Delft University of Technology, Delft.
- Van Noortwijk J.M., M. Kok and R.M. Cooke (1997). Optimal decisions that reduce flood damage along the river Meuse: an uncertainty analysis. In: French S. and J.Q. Smith (eds.). *The Practice of Bayesian Analysis*. p. 151-172. Arnold, London.
- Korving H. (2001). *Beslissen over metingen aan het riool*. (Decision-making on sewer system monitoring., in Dutch). Delft University of Technology, Delft.
- Pratt J.W., H. Raiffa and R. Schlaifer (1995). *Introduction to Statistical Decision Theory*. MIT Press, Cambridge.

Van Gelder P.H.A.J.M. (2000). *Statistical Methods for the Risk-Based Design of Civil Structures*. PhD thesis. Delft University of Technology, Delft.