# Decision-Making under Uncertainty: Optimal Storm Sewer Network Design Considering Flood Risk

Submitted by Si'Ao Sun to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy in Engineering
In July 2010



I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……………………………………………………………..

# Abstract

Storm sewer systems play a very important role in urban areas. The design of a storm sewer system should be based on an appropriate level of preventing flooding. This thesis focuses on issues relevant to decision-making in storm sewer network design considering flood risk.

Uncertainty analysis is often required in an integrated approach to a comprehensive assessment of flood risk. The first part of this thesis discusses the understanding and representation of uncertainty in general setting. It also develops methods for propagating uncertainty through a model under different situations when uncertainties are represented by various mathematical languages.

The decision-making process for storm sewer network design considering flood risk is explored in this thesis. The pipe sizes and slopes of the network are determined for the design. Due to the uncertain character of the flood risk, the decision made is not unique but depends on the decision maker's attitude towards risk. A flood risk based storm sewer network design method incorporating a multiple-objective optimization and a "choice" process is developed with different design criteria.

The storm sewer network design considering flood risk can also be formed as a single-objective optimization provided that the decision criterion is given a priori. A framework for this approach with a single objective optimization is developed. The GA is adapted as the optimizer. The flood risk is evaluated with different methods either under several design storms or using sampling method.

A method for generating samples represented by correlated variables is introduced. It is adapted from a literature method providing that the marginal distributions of variables as well as the correlations between them are known. The group method is developed aiming to facilitate the generation of correlated samples of large sizes. The method is successfully applied to the generation of rainfall event samples and the samples are used for storm sewer network design where the flood risk is evaluated with rainfall event samples.

# Acknowledgement

First of all, I would like to express my deep gratitude to my supervisors, Prof. Slobodan Djordjević and Prof. Soon-Thiam Khu for their exceptional supervisions on this research for the last three years. I appreciate their valuable guidance on the subject and the freedom that they gave me to develop my own interests. Prof. Djordjević offered his professional support and patience throughout the research. Prof. Khu is not only a mentor but also a good friend in private to me.

Also I would like to thank Prof. David Butler and Prof. Kapelan Zoran for engaging in some valuable discussions. I would like to thank Dr. Guangtao Fu and Dr. Albert Chen for their unconditional help whenever I went to them with a question or a problem. I would like to thank Francois Courtot, Sarah Ward and Barry Evens for their help in my English correction. I would like to thank all people from the Centre for Water Systems for their friendly support and assistance.

Then I give the acknowledgement to University of Exeter for the provision of ERS (Exeter Research Studentship), which made this research possible.

Finally I am very grateful to my parents for their selfless support and love.

# Table of Contents

# List of Publications

S. Sun, S.-T. Khu, Z. Kapelan, S.Djordjević (2010), A fast approach for multiobjective design of water distribution networks under demand uncertainty, *Journal of Hydroinformatics* (accepted)

S. Sun, S. Djordjević, S.-T. Khu (2010), Decision making in flood risk based storm sewer network design, *Water Science and Technology* (accepted)

S. Sun, S. Djordjević, S.-T. Khu (2010), A general framework for flood risk-based storm sewer network design, *Urban Water Journal* (submitted)

S. Sun, S.-T. Khu, S. Djordjević (2010), Sampling rainfall events: A novel approach to generate large correlated samples, *Hydrology Research* (submitted)

G. Fu, D. Butler, S.-T. Khu, S. Sun, (2010) Imprecise probabilistic evaluation of sewer flooding in urban drainage systems using random set theory. *Water Recourses Research* (Submitted)

S. Sun, S. Djordjević, S.-T. Khu (2009), Evaluation of flooding return period of storm sewer systems, *10$^{th}$ IWA UK Young Water Professionals conference.* London, UK.

S. Sun, S. Djordjević, S.-T. Khu (2010), Decision making in flood risk based storm sewer network design, *NOVATECH.* Lyon, France.

# List of Figures

# List of Tables

List of Abbreviations

| | |
|---|---|
| B/C | benefit/cost |
| CC | correlation coefficient |
| CDF | cumulative density function |
| DP | dynamic programming |
| DSS | decision support system |
| ENR | Engineering News-Record |
| FORM | first order reliability method |
| FSD | first degree stochastic dominance test |
| GA | genetic algorithm |
| GLUE | generalised Likelihood Uncertainty Estimation |
| GP | generalized Pareto |
| IDF | intensity-duration-frequency |
| LHS | Latin hypercube sampling |
| MCMC | Markov chain Monte Carlo |
| MCS | Monte Carlo simulation |
| MLE | maximum likelihood estimation |
| NSGA II | non dominated sorting genetic algorithm |
| NP-hard | non-deterministic polynomial-time hard |
| PDF | probability density function |
| POT | peak over threshold |
| RCC | rank correlation coefficient |
| RMSE | root mean square error |

| | |
|---|---|
| SA | simulated annealing |
| SSD | second degree stochastic dominance test |
| WDS | water distribution systems |

# List of Symbols

| Symbol | Description | SI unit |
|---|---|---|
| $A$ | contribution area of catchment | $m^2$ |
| | flow cross-sectional area | $m^2$ |
| | set | - |
| $a$ | width of a rectangular channel | m |
| $Bel$ | belief function | - |
| $C_c$ | construction cost | £ |
| $C_f$ | flood risk cost | £ |
| $C_{mh}$ | manhole cost | £ |
| $C_o$ | optimistic cost | £ |
| $C_p$ | pessimistic cost | £ |
| $C_T$ | total cost | £ |
| $C_\alpha$ | the discounted annual construction cost | £ |
| $c$ | runoff coefficient | - |
| $\boldsymbol{c}$ | correlation matrix | - |
| $c_{ij}$ | element in correlation matrix $\boldsymbol{c}$ | - |
| $corr$ | correlation coefficient | - |
| $cov$ | covariance | - |
| $E$ | expectation | - |
| $D$ | pipe diameter | m |

| | | |
|---|---|---|
| $d$ | dry duration before a rainfall event | h |
| $F$ | cumulative probability distribution | - |
| $\boldsymbol{F}$ | collection of cumulative probability distributions | - |
| $F_X$ | cumulative probability of variable $X$ | - |
| $f$ | probability density function | - |
| | general function | - |
| $H$ | flow depth | m |
| $H_\alpha$ | parameter in the Hurwicz criterion | - |
| $h$ | depth of a rectangular channel | m |
| | excavation depth | m |
| $h_B$ | the height when a levee section fails | m |
| $i$ | rainfall intensity | mm/h |
| $I$ | storage inflow | $m^3/s$ |
| $k_s$ | the equivalent sand roughness height | m |
| $L$ | pipe length | m |
| $m$ | mass function | - |
| | the order of a event in Weibull formula | - |
| $N$ | natural number | - |
| | overall number of events in Weibull formula | - |
| $n$ | Manning coefficient | - |
| $nec$ | necessity function | - |
| $O$ | storage outflow | $m^3/s$ |
| $P$ | probability function | - |

| | | |
|---|---|---|
| $p$ | probability | - |
| $P_g$ | penalty function in GA | - |
| $P_w$ | wetted perimeter | m |
| $Pl$ | plausibility | - |
| $pos$ | possibility | - |
| $Q$ | flow rate | $m^3/s$ |
| $R$ | real number | - |
| $R_h$ | hydraulic radius | m |
| $r$ | benchmark interest rate | - |
| $r_{XY}$ | correlation coefficient between variables $X$ and $Y$ | - |
| $S$ | pipe slope; Hydraulic gradient | - |
| $S_f$ | friction slope | - |
| $s_x$ | sample standard deviation | - |
| $t$ | time | s |
| $U_i$ | input or parameter with aleatory uncertainty | - |
| $V$ | mean flow velocity of pipe cross section | m/s |
| $V_E$ | excavation volume | $m^3$ |
| $V_i$ | inputs or parameter with epistemic uncertainty | - |
| $V_s$ | storage volume | $m^3$ |
| $W_i$ | inputs or parameter with aleatory and epistemic uncertainty | - |
| $w$ | wet duration of a rainfall event | h |
| $X$ | set | - |

| | | |
|---|---|---|
| | variable | - |
| $x$ | distance along a conduit | m |
| | sample | - |
| $Y$ | variable | - |
| $y$ | sample | - |
| α | parameter in α-cut of fuzzy set | - |
| | a parameter in Gumbel distribtuion | - |
| | discount factor for annual cost computation | - |
| $\beta$ | a parameter in Gumbel distribution | - |
| $\theta$ | the angle between sewer axis and a horizontal plane | ° |
| | trench side wall angle | ° |
| $\mu$ | mean value of a variable | - |
| $\sigma$ | standard deviation of a variable | - |
| $v$ | kinematic viscosity | $m^2/s$ |
| $\Phi$ | empty set | - |
| $\Omega$ | power set | - |
| $\pi(\boldsymbol{y})$ | A permutation of sample $\boldsymbol{y}$ | - |

# Chapter 1 Introduction

## 1.1 Background

Flood is one of the major natural disasters that cause significant economic losses and threatens many human lives globally every year. For example, the UK Environment Agency (2001) claimed that about 1.85 million homes, 185,000 commercial properties and approximately 5 million people are at risk from flooding in England and Wales. The potential damage of flood is especially high in urban areas due to the concentration of properties and people living in relatively small areas. Urban flooding generally results from one or more of these causes, including fluvial reason, coastal tide, high groundwater level, overland flow and hydraulic overload of artificial drainage systems (Lancaster et al. 2004).

Storm sewer systems are constructed to collect storm runoff flows from catchment surface and convey them mostly by gravity to outfalls, e.g. water courses and waste water treatment plant. Most of the older sewer systems in the UK are combined systems in which wastewater (discharged by domestic residences, commercial properties, industry, etc.) and stormwater flow together is the same pipe, while newly constructed systems since 1945 are mostly separate systems where wastewater and stormwater are kept in separate pipes (Butler and Davies 2004).

Sewer flooding, mainly caused by hydraulic overloading of storm sewer systems, is a serious issue for local government, water companies and property owners. It leads to not only economic losses to properties, but also social disruption. There is a trend of increasing number of sewer flooding events due to global climate change and urbanization (Brown and Damery 2002; Plate 2002; Ryu 2008). With global climate change, precipitations are expected to be more extreme and this may result in increase in drainage loadings. With urbanization, the increasing impermeable surfaces result in interrupted natural water courses and more water going to drainage systems. Both of these two possible changes have adverse impacts on the frequency and severity of urban flooding.

This highlights the importance of evaluating and reducing urban flooding risk. Flood risk is commonly evaluated via modelling. Modelling is a necessary tool for understanding the behaviour of the system and is a predictive tool in decision-making. An integrated modelling approach generally requires the incorporation of uncertainty analysis. It is important to understand the relative magnitude of risk and uncertainty in modelling as they are an imperative part in the understanding of the system behaviour and are essential to decision-making.

On the other hand, storm sewers are mostly laid underground, therefore they are usually very capital intensive assets both for construction and maintenance. The pursuit of cost-effective strategies for storm sewers management is always of interest. In general, the goals of low flood risk and low construction/ maintenance cost for storm sewers management contradict each other, i.e., low flood risk often means high construction/ maintenance cost and vice versa. However, the trade-off between these two objectives can still be optimized in order to attain good designs or management strategies in the sense of useing few resources to achieve good performance.

## 1.2 Motivation

Though uncertainty has been extensively discussed in the literature and uncertainty analysis is frequently applied in practice, not enough attention has been paid to the proper understanding and representation of it. The correct understanding and representation of uncertainty is the foundation on which an appropriate uncertainty analysis should be based. Furthermore, it is common that uncertainties are propagated through a model and the output of the model is of concern. The focus is always on the propagation methods instead of ensuring the coherence of the uncertainty natures between inputs and outputs. The resultant uncertainty representation could be completely different if the nature of uncertainty is wrongly treated. Hence there is a need to highlight this point.

Uncertainty and risk are related. Flood risk can never be certain as the driver of the system usually varies in time. The traditional way of designing a storm sewer system is with a deterministic approach based on a specified protection level. A more comprehensive design of a storm sewer system should take future flood risk into account. In recent decades, researchers began to suggest the explicit consideration of flood risk in the design process of flood defence systems. Literature works mostly use the expected flood risk consequence to represent the uncertain flood risk when solving problems involving flood risk. Such an approach directly treats the probabilistic flood risk as a fixed value for further analysis. However, it is necessary to show that the design process can be presented as a decision-making problem, which can be formed in different ways depending on decision makers' attitude towards risk.

The probabilistic flood risk is generally identified by model simulations with the uncertain inputs being the system driver: rainfall. The way of propagating rainfall uncertainty should depend on the available information about it. For instance, in some catchment, the intensity-depth-frequency (IDF) curves are well developed and hence should be used; while for other catchments, rainfall time series is available or the probability distributions of the variables for describing rainfall are known. In addition, the requirement for the result precision and the computation resources are different for different problems. Thus it is necessary to develop different methods for evaluating flood risk via modelling, which are dependant upon the available information as well as other matching conditions.

## 1.3 Objectives

This thesis aims to provide tools that may be used to improve decision-making in storm sewer network design subject to flood risk, based on a sound understanding of uncertainty.

The objectives of this study include:

1 To propose a general procedure for uncertainty propagation in modelling. In order to meet this objective the following issues need to be explored:
   - to explain the correct way for understanding and representing uncertainty;
   - to develop methods to propagate uncertainty.

2 To explore decision-making in storm sewer network design considering flood risk with different decision criteria. In order to meet this objective, a method for the comparison of different decision criteria for flood risk based storm sewer network design needs to be developed.

3 To develop a framework for single-objective optimization of storm sewer network design provided that the design criterion is known.

4 To develop a method for the generation of rainfall event samples represented by correlated variables, which is used for flood risk evaluation in storm sewer network design.

## 1.4 Thesis outline

This thesis is organised in 7 chapters including this introductory chapter.

Chapter 2 provides an up-to-date literature review of the issues related to this study. It consists of four subjects: uncertainty and uncertainty analysis, flood risk and related uncertainty, decision-making under uncertainty/risk and methods for storm sewer network design.

Chapter 3 is concerned with uncertainty and uncertainty analysis. Focus is placed on the understanding and representation of uncertainty and the propagation of uncertainty through modelling. The importance of proper representation of uncertainty is emphasized and the physical meanings of the representations are explained. It also presents the appropriate way to propagate uncertainty through modelling under different circumstances where uncertainty is presented by either probability or other mathematical languages. The methodology is applied to a simple case of flood discharge evaluation.

Chapter 4 defines the design of the storm sewer network considering flood risk tackled in this thesis. It can be considered as a decision-making problem that can be solved by an optimization procedure. A method incorporating a multi-objective optimization and a "choice" process is developed in order to compare different decision criteria.

Chapter 5 developed a framework for a single-objective optimization for designing a storm sewer network provided that the decision criterion is known a priori. In the framework the flood risk evaluation is located in an optimization loop. This chapter assesses the flood risk under several design storms.

Chapter 6 provides another application of the framework for the design of storm sewer network formed as a single-objective optimization. This chapter evaluates the flood risk with samples of rainfall events. A method for generating samples represented by correlated variables is introduced.

Chapter 7 summarizes the study and draws conclusions about the value of the work presented in the thesis. Possible further research directions are also given.

# Chapter 2 Literature review

This chapter presents a literature review relevant to uncertainty analysis, flood risk, decision-making and storm sewer network design. It starts with a brief description of uncertainty and uncertainty analysis in general setting; the risk evaluation and related uncertainty analysis in the area of flooding are then examined; this is followed by the review of decision-making under uncertainty/risk; lastly, the methods for storm sewer network design are investigated.

## 2.1 Uncertainty and uncertainty analysis

Uncertainty is a state of having limited knowledge where it is impossible to exactly describe existing situations or future outcomes. It is widely present in all areas in engineering and has been extensively discussed in the literature.

Though uncertainty analysis is still not a standard practice in many modelling applications and it is common to show results without uncertainty bounds to decision makers, at scientific conferences, in refereed publications or in consultancy applications, most of researchers see uncertainty analysis as an important component of good scientific practice (Pappenberger and Beven 2006). There is an increasing consensus in the incorporation of uncertainty analysis in engineering studies. For instance, Bae et al. (2004) stated that uncertainty quantification analysis has been embedded in engineering structural design instead of simply assigning safety factors over the last decade and multiple types of uncertainty in a system must be considered for a robust prediction of the system performance. Lund (2002) believed that the probabilistic analysis has largely replaced older forms of economic analysis, i.e. examining only a particular design flood when designing a flood defence infrastructure. Apel et al. (2004) asserted that flood disaster mitigation strategies should be based on a comprehensive assessment of the flood risk combined with a thorough investigation of the uncertainties associated with the risk assessment procedure.

Uncertainty can significantly affect the decision-making process and ultimately, the decisions taken. It is believed that the lack of characterization of uncertainty may yield qualitatively and quantitatively different answers from that derived from a reasoned treatment of uncertainty (Morgan and Henrion 1990). Optimal decisions can only be expected when all relevant uncertainties are taken into consideration (Aven and Pörn 1998). In this section, basic concepts relevant to uncertainty and uncertainty analysis and the ways for handing uncertainty in hydrosystems are reviewed.

### 2.1.1 The nature and source of uncertainty

There are two fundamentally different types of uncertainty characterized by the nature of uncertainty; they are aleatory uncertainty and epistemic uncertainty (Paté-Cornell 1996; Nauta 2000; Apel et al. 2004; Ross et al. 2009).

Aleatory uncertainty originates from variability in known (or observable) populations and therefore represents randomness in samples. According to Hall (2003), it can be operationally defined as a feature of the population of measurements that conforms well to a probabilistic model. Aleatory uncertainty has also been termed variability, stochastic uncertainty, objective uncertainty and Type I uncertainty.

Epistemic uncertainty results from lack of knowledge of fundamental phenomena and is related to our ability to understand, measure, and describe the system under study. It has also been called ignorance, subjective uncertainty and Type II uncertainty. It is believed that the aleatory uncertainty is a property of the system and the epistemic uncertainty is a property of the analyst (Cullen and Frey 1999).

The aleatory uncertainty cannot be reduced due to its inherent nature while the epistemic uncertainty can be further reduced, e.g. by obtaining more data or knowledge (Merz and Thieken 2005). Many researchers have recognized that aleatory and epistemic uncertainties should be treated separately (Hattis and Burmaster 1994; Hoffman and Hammonds 1994; Parry 1996; Frank 1999; Apel et al. 2004; Ross et al. 2009). Separating the two types of uncertainties may help to make more informed management decisions (Merz and Thieken 2005). Dubois (2010) believed that it is important to distinguish aleatory and epistemic uncertainties as the decision resulted from different uncertainties will be quite different: concrete action must be taken to circumvent the potential dangerous effects of inherent variability (aleatory uncertainty), whereas the best decision is probably to try and reduce the epistemic uncertainty by collecting more information due to incomplete information. The prediction of a result may depend on the way that aleatory and epistemic uncertainties are separated and a wrong assessment may occur if the distinction between them is neglected (Nauta 2000). However, explicit separation of aleatory and epistemic uncertainties may be difficult as in some cases a clear distinction is unclear because of our incomplete understanding of the system (Apel et al. 2004). In this case, Nauta (2000) pointed out, it may be better to quantify their separation improperly, than not to separate them at all.

Uncertainty arises from different sources (Apel et al. 2004). Some of the sources of uncertainty that are frequently encountered are detailed as follows:

- Inherent variation. The source of inherent variation refers to quantities that are variables inherently varying over time, space, or populations of individuals. It belongs to aleatory uncertainty which is usually characterised as random or stochastic. A typical example of this category is the amount of extreme rainfall over consecutive years.
- Statistical uncertainty (sampling uncertainty). It is common to use a data sample to infer a probability distribution of a variant. Statistical/sampling uncertainty can be introduced in this process when fitting the data sample to a specific distribution especially if the data sample is relatively small (Michele and Rosso 2001; Serinaldi 2009). Such uncertainty source arises due to the random property of the sample. Statistical/sampling uncertainty is an epistemic uncertainty source (Merz and Thieken 2005), and the level of uncertainty decreases when more data of the sample are available.
- Parameter uncertainty. When a system is simulated by a model, parameters are generally required to construct the model. Parameter uncertainty arises when the exact value of this parameter is unknown and must be estimated. It can be aleatory uncertainty or epistemic uncertainty depending on the nature of the parameter. For instance, in a rainfall-runoff model, the parameter describing the initial moisture condition of a catchment is a variable of aleatory uncertainty as it varies over time; the parameter related to the proportion of the impervious area is of epistemic uncertainty due to lack of knowledge. Additionally, if the probability distribution cannot be identified with certainty when fitting the inherent variance

of a parameter to a distribution, both aleatory and epistemic uncertainties are present.

- Model structural uncertainty. As a model is usually a simplification of the reality, it is common that errors may occur between the simulated outcome and the real data. Model structural uncertainty can be thought of as addressing the uncertainty in the appropriateness of the structure of a model (Merz and Thieken 2005). Neglecting uncertainty in model structure leads to an underestimation of the uncertainty in model predictions (Reicheert and Omlin, 1997). It generally received little attention in research compared to parameter and input data uncertainty due to the difficulty in quantifying it (Lindenschmidt et al. 2007). Model structural uncertainty usually belongs to the category of epistemic uncertainty.

### 2.1.2 Uncertainty expressed in mathematics

Mathematics provides us with the necessary language to encode uncertainty when dealing with uncertainty. Traditionally, probability theory has been widely used. It is extensively employed to describe different sources of uncertainty. For example, it is used to represent inherent variations of phenomena such as rainfall, runoff and flood (Muzik 2002; Koutsoyiannis 2004; Bocchiola and Rosso 2009), model parameters (Lei and Schilling 1994; Korving et al. 2003), sampling uncertainty (Serinaldi 2009) and model structural uncertainty (Freni et al. 2009). Though both aleatory and epistemic uncertainties are frequently described by probability distributions, they are interpreted with different physical meanings. The probability distribution for aleatory uncertainty represents the relative frequency of values, whereas when used in the context of epistemic uncertainty, it represents the degree of belief or the knowledge of a value (Voortman 2003). These two different meanings are developed from a frequency view and Bayesian's view, respectively.

Probability theory is based on the additivity axiom, which implies that the relevant evidence is a complete and consistent description of a problem. However, under circumstances involving sparse data, incomplete information or possibly inconsistent knowledge, the additivity axiom is difficult to justify (Hall et al. 1998). Compared with probability, other mathematical languages for uncertainty representation such as fuzzy sets, possibility, random sets and probability box do not suggest an objective bounding analysis where only statistically founded probability distributions are taken into account. These methods have the potential to describe incomplete objective information and model subjective judgment. Hence both probability and other uncertainty representations are useful and should be articulated with one another (Dubois 2010).

Fuzzy sets, which are firstly introduced by Zadeh (1965), are sets whose elements have degree of membership. The membership functions are often used to model the extension of some natural language describing or predicating. It permits gradual assessment of the membership of elements in a set. Fuzzy sets generally quantify epistemic uncertainty.

Possibility theory is introduced as an extension of fuzzy set theory and fuzzy logic (Zadeh 1978). Possibility usually has a shape like a fuzzy membership. Hence a kind of possibility distribution can be built from a membership function of fuzzy sets. The possibility distribution generalizes the probability by weakening the additivity axiom. It is a concise encoding of a special probability family (Dubois 2006). Probability theory uses a single number, the probability, to describe how likely an event is to occur, whereas possibility uses two concepts, the possibility and necessity of the event to measure the upper and lower likelihood of an event.

Random set theory describes set-valued stochastic process (Dubois and Prade 1991). Random sets are less precise than random variables. The focal elements (the set-value), upon which random sets are based, are a source of imprecision in the uncertain process. Random set theory is believed to be equivalent to Dempster-Shafer theory (Dempster 1967; Shafer 1976) describing evidence. A Dempster-Shafer structure on the real line is similar to a discrete probability distribution except that the locations at which the probability mass resides are sets of real values, rather than precise points. The imprecision that the random sets or the Dempster-Shafer structure presents is bounded by a lower end and an upper end, i.e., the belief function and the plausibility function.

Probability box represents a range of probability distributions (Williamson and Downs 1990). A probability box is an imprecise probability that is bounded by a lower and an upper bound. It can also be viewed as a continuous form of random sets.

The merits for different mathematical languages to represent uncertainty are recognized. For example, of all the methods for handling uncertainty, probability has by far the longest tradition and is the best understood (Hall 2003). Moreover, probability is appropriate for the quantification of all forms of uncertainty in principle (O'Hagan and Oakley 2004). On the other hand, it has been found by the scientific and engineering community that there are limitations in using only one framework (probability theory) to quantify the uncertainty in a system because of the impreciseness of data or knowledge (Bae et al. 2004). Hall (2003) believed that much of reasoning is possibilistic rather than strictly probabilistic when probing the nature of human reasoning. For instance, people reason about whether a given scenario could happen, without necessarily endeavoring to attach a probability to the likelihood of it happening. In addition, the incomplete or even contradict information can be better captured by possibility or Dempster-Shafer structure rather than probability. The imprecise probabilities are viewed as various forms of relaxation of the formal theory of probability, which intend to relieve some of the difficulties facing of probability (Wu et al. 1990). Hence the mathematical language chosen for representing uncertainty depends on the specific situation. A number of comparative discussions of different approaches for the choice of languages to represent uncertainty are in the literature (Wu et al. 1990; Dubois and Prade 1993; Helton et al. 2004).

### 2.1.3 Handling uncertainty in hydrosystems

Uncertainty is widely recognized and discussed in the area of water engineering. Various methods for uncertainty treatment or analysis have been introduced and are applied in this domain. This section reviews issues relevant to handling uncertainty in hydrosystems. Rainfall, which is the main driver of hydrosystems, is usually considered as a variable with inherent uncertainty. The annual extreme values or the peak over threshold (POT) values are generally of interest. The conventional way of assessing the inherent uncertainty in rainfall series is to fit historic data to a certain probability distribution (Grum and Aalderink 1999; Fu et al. 2010). The probability distribution fitting method can also be used in the evaluation of runoff or flood if a data record of certain length is available (Haktanir and Horlacher 1993). When identifying a probability distribution, the parameters of the distribution can be estimated by a number of methods such as weighted moments method, expected moment method (Cohn et al. 1997), maximum likelihood method (Stedinger and Cohn 1986) or Bayesian Markov Chain Monte Carlo (MCMC) method (Reis Jr. and Stedinger 2005). The statistical/sampling uncertainty can be introduced in this probability distribution fitting process due to limited number of data sample (Bao et al. 1987; Serinaldi 2009). The model structure uncertainty may arise when assuming the specific probability distribution that the data are fitted to.

Modelling has become a preferred approach to study hydrosystem, and it is recognized that uncertainty in inputs or parameters needs to propagate through the model. Monte Carlo simulation (MCS) is the most popular method in practice due to its simple principle and ease application (Grum and Aalderink 1999; Kuczera 1999; Hofer et al. 2002; Rahman et al. 2002; Apel et al. 2004; Kwon et al. 2007). Analytic method, which derives the probability distribution of model outputs using analytic technique, is also utilised to evaluate uncertainties though a model (Kurothe et al. 1997; Goel et al. 2000). However, the application of this method is limited as most uncertainty propagation problems have no analytic solution or the analytic solution is very difficult to be elicited. The first-order reliability method (FORM), which is widely used within the area of structural engineering, is also applied to the area of water engineering when a failure probability of a system is concerned (Thorndahl and Willems 2008).

Models are imperfect because the physical phenomena are not exactly known and some variables of less importance are omitted in modelling for efficiency reasons. Hence errors may occur between the model output and the true result when employing a model to simulate the reality. As a result epistemic uncertainty may arise from the simplified model structure or the not precisely known model parameters. A conventional way to quantify uncertainty under this circumstance is to incorporate it in the parameter calibration process. If the parameters are assessed by maximum likelihood estimation (MLE), the uncertainty can be estimated through a quadratic approximation of the likelihood function (Reis Jr. and Stedinger 2005). The Generalised Likelihood Uncertainty Estimation (GLUE) is a very simple method for model uncertainty qualification (Beven and Binley 1992; Beven and Freer 2001; Beven 2006). It is based on an equifinality hypothesis allowing that different sets of parameters may be equally likely as simulators of a system. GLUE has a wide application in predictions of floods and hydraulic transport (Aronica et al. 1998; Lamb et al. 1998). The Bayesian method, combining the information in the available data and prior knowledge about the model structure or parameters, can also provide a posterior probability distribution of this model structure uncertainty (Korving et al. 2003). With the Bayesian framework, one does not have to use any approximate asymptotic assumption about the uncertainty and it can easily update newly gained information (Reis Jr. and Stedinger 2005).

The aleatory and epistemic uncertainties have different physical meanings. In line with the distinction between aleatory and epistemic uncertainties, some researchers believed that the analysis should be hierarchically structured when both aleatory and epistemic uncertainties are present in modelling. A good review of the two types of uncertainty and the mathematical treatment is given in the special issue of the Journal of Reliability Engineering and Systems Safety (Helton and Burmaster 1996).

In water engineering, Grum and Aalderink (1999) argued that epistemic uncertainty and inherent variation should be treated separately and they applied a two-layer MCS to the return period analysis of combined sewer overflow effects. Apel et al. (2004) developed a hierarchical Monte Carlo framework to evaluate the flood risk of a river: in the first level of the analysis the MCS represents the variability of the system, i.e. the aleatory uncertainty, whereas the second level analysis represents the epistemic uncertainty associated with the results of the first level. Merz and Thieken (2005) stated that the separation of aleatory and epistemic uncertainties gives a more differentiated picture of the complete uncertainty and presents the probability bounds within which the true but unknown probability distribution in flood frequency analysis lies.

The representation of uncertainty should be consistent with the available information. When more than one mathematical language is used to describe uncertainty in modelling, the question arises as how to combine different modes of representations and propagate them through a model. A straightforward way is to transform one representation of uncertainty to another, for example, probability distributions can be transformed to the form of random sets (Tonon 2004). Ross et al. (2009) identified random set-based results for the estimation of groundwater flow and transport simulation where both probability and random sets are employed for describing uncertainty. Fu et al. (2010) used random sets as a media tool to evaluate sewer flooding in urban drainage: the randomness from rainfall data is represented by probability box and the imprecision from model parameters is represented by fuzzy numbers; in the end the resultant flood evaluation is given in the form of a lower and an upper bound of cumulative probabilities, which are believed to encapsulate the unknown true probability distribution. Another way to handle different mathematical uncertainty representations is to deal with them in different dimensions. Guyonnet et al. (2003) proposed a hybrid approach for addressing a risk assessment for human exposure to pollutants with both probability and fuzzy sets being present. The probability and fuzzy sets are distinguished and propagated in two dimensions. These approaches allow a consistent representation of uncertainty with the information available.

## 2.2 Flood risk and related uncertainty

Flooding brings serious consequences to properties, environment and public safety. It can be caused by various reasons, including river overflow, drainage surcharge, coastal tide, high ground water level, overland flow, etc. The occurrence of flooding can hardly be predicted with certainty due to the stochastic nature of precipitation which drives the flood systems. Hence the term “flood risk” is generally used, with risk implying the uncertain property of the flood.

The term “risk” has been in circulation for long and is used in a number of different contexts. Many researchers believed that flood risk is determined by two or three components. The two components include the chance (or probability) of an flood event occurring and the impact (or consequence) associated with that event (Sayers et al. 2002), while the three factors consist of hazard which is the probability of occurrence of a potentially damaging phenomenon, exposure which stands for the property or life exposed to the potentially damaging phenomenon) and vulnerability which is the degree of loss of the property or life (WMO and GWP 2008). The three factors description of flood risk can be easily simplified to the two factors expression by letting the hazard equal the chance and the product of exposure and vulnerability equal the impact. As a result, flood risk is frequently described by a probability distribution, namely, the probabilistic flood consequence.

Yet, in some cases risk only concerns the expected consequence. Thus a narrower definition of flood risk is the product of the consequences of flood events and the associated probabilities. Helm (1996) warned that a simple product of probability and consequence will never be sufficient to fully describe risk, but he also claimed that it provides an adequate basis for comparison and assist in decision-making.

Risk analysis has always formed a central part of the science of hydrology and hydraulics (FRMRC 2006). A comprehensive assessment of flood risk is the premise for good flood disaster mitigation strategies and flood defense designs. The risk-based design approach which generally balances benefits and costs of strategies or designs in an explicit manner is a more complete approach in comparison to the deterministic approach by specifying an exceedance probability of the flood (Vrijling 2001; Apel et al. 2004).

Flood risk is a term subject to uncertainty itself (usually belongs to aleatory uncertainty) and the evaluation of flood risk requires uncertainty analysis. In addition, the probabilistic flood risk cannot be identified with absolute certainty as uncertainty arises from different sources in the procedure of flood risk evaluation. The uncertainty may be caused by the simplification of the model, the choice of the parameters, the assumptions made in the evaluation or the lack of data. A comprehensive flood risk analysis should take into account all relevant flooding scenarios, their associated probabilities and possible damages as well as a thorough investigation of the uncertainty associated with the risk analysis. Therefore, a flood risk analysis finally yields a risk curve, i.e. the full distribution function of the flood damages in the area under consideration, ideally accompanied by uncertainty bounds (Apel et al. 2004). For flood defence designs, all uncertainties involved in the design process should be systematically and explicitly considered, and the reliability of the performance of designed systems needs to be assessed (Tung and Yen 2005). It is fundamentally irresponsible and unethical for designers not to interpret and incorporate uncertainties into design (Smith 1994).

To many, uncertainty analysis is an additional complication that can only be incorporated at the cost of additional expense, understanding and training. However, to ignore uncertainty in any form of flood risk prediction carries an associated risk for the analyst of being wrong, and does not allow the decision maker to take account of different risks of potential outcomes (FRMRC 2006).

Risk and uncertainty analysis is currently recognized as an important area of research. Its significance in decision-making has been acknowledged. The explicit incorporation of flood risk has begun to find its way in practice. For instance, Lund (2002) developed integrated floodplain management plans using optimization based on flood risk analysis. Apel et al. (2004) investigated the complete flood disaster chain with a hierarchically structured Monte Carlo framework and at the end it provided a probabilistic economic damage caused by flood associated with uncertainty bounds. Korving et al. (2003) and Korving et al. (2009) made their decision on sewer rehabilitation based on risk assessment with uncertainty consideration. Kellagher and Sayers (2009) proposed a new procedure for assessing and managing sewerage systems focusing on consequences (probability and hazard impact) rather than achieving a specific level of performance.

When assessing flood risk, Balmforth et al. (2006) suggested three key components: input, process and output (see Fig 2.1). The exceedance flow, depth, velocity, volume and duration can be used as inputs for determining flood probability. If a sufficiently long record of historic flood stages data is available, flood probability can be determined by a statistical distribution of the flood stages. Otherwise, modelling in term of hydraulic simulation is performed to generate flood stages. One input or several inputs can be fed into the consequence part including both tangible (i.e. damage to property or the health and safety of the public) and intangible factors (i.e. environmental, social-economic impacts and loss of facility). Combining the probability and consequence with a certain flood stage, the output, a probabilistic curve of flood consequences is determined. This process is widely recognized and applied in flood risk evaluation (Apel et al. 2004; Morita 2008; Ryu 2008).

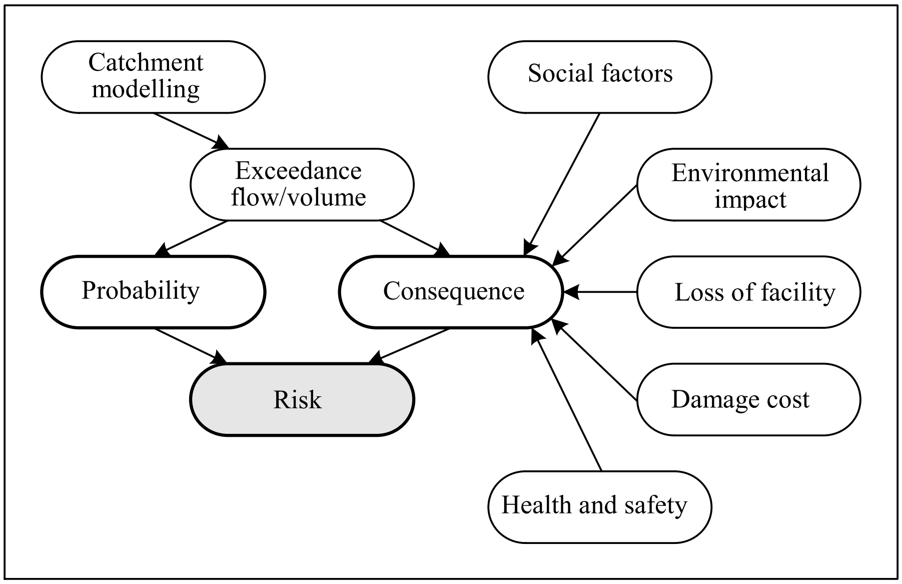



Fig 2.1 Exceedance flood risk assessment (Balmforth et al. 2006)

Referring to flood consequence, the tangible damage can be further divided into direct losses and indirect losses. The largest part of the literature on flood damages concerns direct tangible damages as it is relatively easier to be assessed compared to indirect and intangible damages. Merz et al. (2004) analyzed a data set of approximately 4000 damage records and obtained the direct monetary damage to inundated buildings and related uncertainties. Morita and Fukuda (2002) computed the damage amounts of the public and private properties as the losses from floods. Ryu and Butler (2008) expressed flood damage as percentages of flooded property values. Furthermore, researchers also worked on the quantification of indirect and intangible damages. For instance, Olsen et al. (1998) included indirect losses due to flood in their analysis for estimating the economic effects of flooding; Lekuthai and Vongvisessomjai (2001) made an attempt to assess the intangible damage in monetary terms. Jonkman et al. (2003) discussed risk measures for loss of life and calculate the flood risks for an area in the Netherlands.

## 2.3 Decision-making under risk/uncertainty

### 2.3.1 The concept

Decision-making is a mental/cognitive process resulting in an outcome leading to the selection among several alternatives. It is an important part of all science-based profession where knowledge is applied to make informed decisions. Every decision-making process produces a final choice, based on the values and preferences of the decision maker.

The typical approach of a decision-making process consists of four phases (Simon 1977; Abebe and Price 2005), namely, “intelligence” which refers to the gathering of data and information and the formation of a problem, “design” that contains multiple solution options to resolve the problem, “choice” from the alternatives that seems to best resolve the problem and “review” of whether the selected option is appropriate. Some researchers combined the last two phases as one stage so that the decision-making follows three main stages including intelligence, design and choice (Guariso et al. 1996; Quintero et al. 2005; Ryu 2008).

Very few decisions are made with absolute certainty because complete knowledge about all the alternatives is seldom possible. Moreover, it is common that the evaluations of alternatives contain inherent uncertainty. In this sense, decision-making is sometimes also defined as a process of sufficiently reducing uncertainty and doubt about alternatives to allow a reasonable choice to be made from them (Harris 2009).

Decision-making under risk is a frequently encountered in the area of civil engineering. For example, structural design usually relies on reliability analysis. It treats both the load and the resistance of a structure as stochastic variables and computes the failure probability during the design, construction and life span of the structure (Tsompanakis et al. 2002). The mitigation strategies for seismic damages on structures such as bridges require to evaluate and compare the cost effectiveness based on risk (Padgett et al. 2010). When designing a flood defence system, risk-based design methods is suggested as the appropriate level of protection is ideally obtained by balancing the cost of the protection against the risk reduction of the protection area (Voortman 2003). The effective decision-making in sewer rehabilitation should take risk and uncertainty into account (Korving et al. 2009). The water distribution systems (WDS) should be designed based on the evaluation of total WDS risk which is defined as the probability of all nodes satisfying minimum pressure head in the network (Kapelan et al. 2006).

### 2.3.2 Methodologies

In general, risk can only be reduced /minimized rather than be eliminated completely. For instance, the absolute flooding prevention is impossible for a flood defence system because of the stochastic nature of extreme precipitation (Verworn 2002).

A traditional way to deal with decision-making under risk/uncertainty is to predefine a protection/design level. In structural design, the reliability or safety factors are generally given in standards. In flood defence design, a system is usually designed for a certain flood return period.

Another simple but popular strategy for risk based decision-making is to use "expectation". With this strategy, the expected value of each possible alternative or option is estimated. The decision is then made based on these expectations. The decision maker using expectation to represent the probabilistic value adopts a risk neural attitude. Individuals may have different attitudes towards risk. Risk-averse and risk-seeking are the other two attitudes. For most public investment decision makers, the fear of possible failure of intended project performance caused by uncertainty frequently translates into a more cautious and conservative attitude. The usage of safety factor in many engineering designs is a perfect example of this behavior (Tung et al. 1993).

Other non-mainstream methods are also applied in the decision-making in engineering projects under risk. Economic analysis making using of benefit and cost has long been applied to assess different designs. Probabilistic Benefit/Cost (B/C) analysis has been computed (Park 1984). It can be employed to assess the probability of an investment being feasible when the target value of the selected performance measure is set (Tung et al. 1993). The net benefit (benefit-cost) is also hired as an important indicator for performance evaluation. Al-Futaisi et al. (1999) applied four decision criteria, namely, maximization of expected net benefits, maximization of expected net benefits with variance penalty for uncertainty, maximization of net benefits with probabilistic constraint on net benefit and maximization of net benefits with probabilistic constraints on marginal return, to explore the performance of designs identified by different decision models with economic risk. In addition, Tung et al. (1993) used the concept of stochastic dominance, which was first proposed in economics, to evaluate the economic merit of water resources projects under uncertainty, taking into account both the full range of probabilistic economic performance and the decision-maker's risk attitude.

It is worth mentioning that the advancement of computer tremendously enhanced traditional decision-making (Guo 2007). With the enormously increasing computation power, it is possible to examine more alternatives, execute better system performance and use different computation intensive techniques to improve decision-making. The decision support system (DSS), defined as interactive computer-based aids designed to assist people in solving complex tasks aiming to support and improve the decision process (Hackathorn and Kenn 1981) is a typical example. However, it should also be noted that computers typically assist decision makers in the decision-making process by improving efficiency or effectiveness of the process rather than replacing people's judgment (Keen and Morton 1978).

## 2.4 Storm sewer network design

### 2.4.1 Development of storm sewer network design

The development of urban drainage is closely associated with the development of human civilization. Guo (2007) divided periods of the development of the drainage design into five stages according to the techniques applied to it:

i. This stage is the period before the mid nineteenth century when sewer design was merely an "art" mainly based upon experience and intuition rather than a 'science' based on knowledge of the laws of physics.
ii. This is the period from the middle to the late of the nineteenth century. In this period, the designs appeared to be more "theory" or "science" oriented with increasing knowledge of the flow behaviour in sewers. However, in this stage the primary focus is on the hydraulics of pipe flows, without careful consideration of the impacts of processes above ground.
iii. This stage is the first half of the $20^{th}$ century characterized with the recognition of the importance of hydrologic processes. In the meantime, several hydraulic equations, including the renowned Saint Venant equations were established.

iv. This stage begins from the mid 1960s when the advancement of computer began to greatly enhance modelling and optimization in this area. A number of mathematical models, employing unsteady flow routing, were established to simulate the hydrological and hydraulic processes of urban drainage. The concept of optimal sewer design, involving economical factors in addition to engineering consideration, was proposed. As a result comprehensive cost-effective designs incorporating simulation models and optimization technologies became computationally tractable and flourished.

v. In this stage, from the last decade of 20$^{th}$ century or so, the scope of sewer system design has been greatly expanded to involve a wider spectrum, e.g. environment, ecology, climate change, control and maintenance aspects. More comprehensive simulation models, such as SWMM (Rossman 2008) and InfoWorks (Wallingford 2001), became common tools for sewer systems management. Furthermore, innovative optimization techniques have emerged and been broadly applied in this stage.

Though the storm sewer design has progressed a lot along the five stages, it is still an area of very active research nowadays and has great potential to be improved towards more effective and comprehensive designs, for example, the majority of current practice still use a deterministic approach without considering risk; the real-world application suffers from an intensive computation cost and the global optimization is not adopted. It appears that the research in this area is going to be a long term exercise (Guo 2007).

### 2.4.2 Precipitation, hydrological and hydraulic process in storm sewer network design

Precipitation drives storm sewer flooding, thus is an important input in storm sewer network design. Hydrological and hydraulic processes are the two sub-processes of urban drainage. Hydrological process is concerned with the water movement above ground, e.g. rainfall-runoff process while hydraulic process mainly deals with pipe flows. A typical storm sewer network design mainly considers three components: precipitation, hydrological process and hydraulic process in pipes. Models are generally employed to assist storm sewer network design.

### 2.4.2.1 Precipitation for storm sewer network design

As a storm sewer network is designed to collect and convey stormwater for the purpose of preventing flooding, precipitation of the protecting area is an essential factor for the storm sewer network design.

The observation of precipitation is the origin of all our knowledge about rainfall. Generally historical records of precipitation allow derivation of relationship between rainfall event properties (mainly intensity, duration and frequency).

Precipitation is mostly considered in flood defence design by “design events”. Storm sewer networks are conventionally designed under design storms. A design event does not necessarily mimic nature, but is merely a convenient way for designing safe and economical infrastructure (Chadwick and Morfett, 1993). A convenient form of rainfall information is the IDF relationship. A typical set of IDF curves is given in Fig 2.2 where it can be seen that for an event with a particular return period, rainfall intensity and duration are inversely related. With the IDF, if a storm sewer network is required to be designed under certain return period rainfall, design storms of that return period with different durations should all be tested for the design.

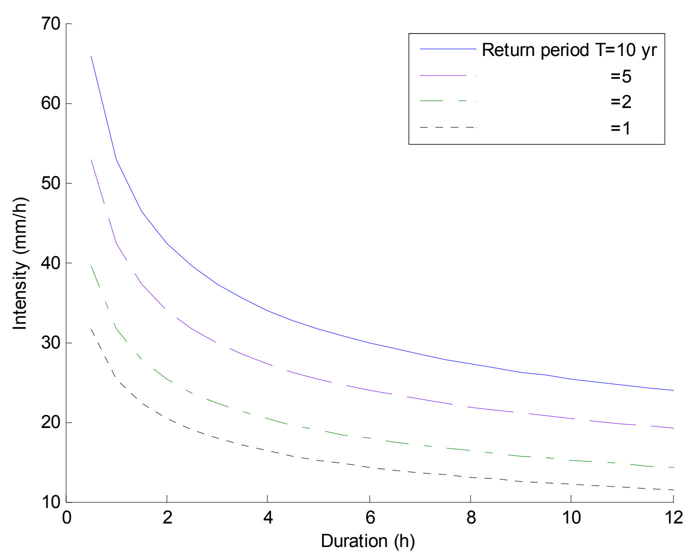



Fig 2.2 Typical intensity-duration-frequency curves

The choice of design storm return period determines the degree of protection provided by a system. In practice the design storm return period is adopted simply on the basis of judgement and precedent. Standard practice in the UK is to use storm return period of 1 year or 2 years for most schemes with 5 years being adopted where property in vulnerable areas would be subject to significant flood damage; higher periods up to 25 years may be adopted for city centre sewers (Butler and Davies 2004).

When designing a storm sewer network (pipe sizes and slopes), the conventional method includes two main basic steps after design storms have been identified:

i. Determine the peak flow rate in storm sewer pipes according to hydrological modelling (see Section 2.4.2.2); and
ii. Determine pipe sizes and slopes satisfying the requirement of conveying peak flow rate without causing any overflow or surcharge according to pipe hydraulic modelling (see Section 2.4.2.3).

Gupta et al. (1983), Yen et al. (1984), Elimam et al. (1989), Charalambous and Elimam (1990), Mays (2001), Liang et al. (2004), etc. gave details of the traditional design method.

#### 2.4.2.2 Hydrological process

A hydrological process generates surface runoff that goes into drainage systems. Various models are proposed to simulate the rainfall-runoff process. The models can be classified into empirical models and physical reality based models (also called process based methods).

The analysis and design of urban drainage systems was traditionally, and still often is, executed using the Rational Method (Lyngfelt 1991). The Rational Equation is the simplest method to determine peak discharge of rainfall runoff for designing drainage. It is an empirical method commonly used for sizing storm sewer systems. The rational equation is expressed as:

$$Q_{\mathrm{p}} = ciA \tag{2.1}$$

where $Q_{\mathrm{p}}$ is peak runoff flow rate, $c$ is runoff coefficient, $i$ is rainfall intensity, $A$ is the contribution area.

The rational equation is a prevailing method for storm sewer network design due to its simplicity. However, its application is limited as it is accurate only for small catchments (Butler and Davies 2004) and it does not consider the spatial and temporal variation of rainfall intensity.

A physical reality based hydrological process transforms a rainfall hyetograph to a surface runoff hydrograph. It mainly involves two principal parts: losses due to interception, depression storage, infiltration and evaporation and transformation from the effective rainfall to an overland flow hydrograph by surface routine (Butler and Davies 2004). The losses can be represented in different ways. The interception and evaporation are often neglected in urban drainage modelling; the depression storage is generally expressed as an equivalent rainfall depth depending on the surface ground type and slope; the infiltration in urban drainage models can be estimated with Horton's equation. More commonly, a runoff coefficient is specified for various surface types and the effective rainfall is produced by deducting the loss from the total rainfall (Butler and Davies 2004). When generating the runoff hydrograph from effective rainfall, conceptual models such as unit hydrographs and reservoir models can be utilized. With unit hydrograph method, the unit hydrograph represents the outflow hydrograph resulting from a unit depth of effective rainfall. The runoff hydrograph generated from storms of different intensities can be identified by linear addition of results from unit rainfall. In a reservoir model, the catchment is treated as a reservoir or several clustered reservoirs. A reservoir model is based on the premise of water balance:

$$\frac{dV_s}{dt} = I - O \tag{2.2}$$

where $V_s$ is storage volume, $t$ is time, $I$ is the storage inflow, $O$ is storage outflow. Additionally, a more physically based model to generate overland runoff is to solve an appropriate approximation of mass and momentum conservation equations such as the kinematic equation.

### 2.4.2.3 Hydraulic modelling

The hydraulic performances within sewer pipes are highly nonlinear and dynamic. The steady flow in a pipe can be described by Manning's equation under free surface condition:

$$V = \frac{1}{n} R_{\mathrm{h}}{}^{2/3} S^{1/2} \tag{2.3}$$

or Colebrook White equation for full pipe flow:

$$V = -2\sqrt{2gSD}\,\log_{10}\left(\frac{k_s}{3.7D} + \frac{2.51v}{D\sqrt{2gSD}}\right) \tag{2.4}$$

where $V$ is the mean flow velocity of pipe cross section; $R_{\mathrm{h}}$ is the hydraulic radius; $S$ is pipe slope (Hydraulic gradient); $n$ is the Manning roughness coefficient; $D$ is pipe diameter; $k_s$ is the equivalent sand roughness height; $v$ is kinematic viscosity.

The pipe diameters can be determined from Eq(2.3) or Eq(2.4) if the design flow of pipes, the roughness coefficient and the slope of the pipes are known. This is the traditional way to design pipe sizes and is still in use now. However, the pipes are designed individually in this method. Hence this method ignores the mutual interactions between pipes and the storage capacity of manholes. In this sense, a simulation tool that considers the system as a whole and permits instantaneous evaluation of the situations throughout the system is more comprehensive.

A simulation model for pipe flows is usually built based on the conservation of mass and momentum. The model using Saint Venant equations is the most accepted approach to describe the unsteady pipe flows. They are represented by continuity equation:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0 \tag{2.5}$$

and momentum equation:

$$\frac{1}{gA}\frac{\partial Q}{\partial t} + \frac{1}{gA}\frac{\partial}{\partial x}\left(\frac{Q^2}{A}\right) + \cos\theta\frac{\partial h}{\partial x} - \left(S - S_f\right) = 0 \tag{2.6}$$

where $t$ is time; $x$ is distance along the conduit; $Q$ is flow rate; $A$ is flow cross-sectional area; $H$ is flow depth; $\theta$ is the angle between sewer axis and a horizontal plane; $S$ and $S_f$ are the conduit slope and friction slope, respectively.

Originally, the Saint Venant equations represent flow behavior conditioned on the free surface assumption. They are retained effective for pressurized flow by introducing the Preissmann slot into pipe soffit. The Preissman slot is a conceptual vertical and suitably narrow slot providing a conceptual free surface condition for the pressurized flow. The Eq(2.5) and Eq(2.6) are still effective in surcharge conditions with the slot (Butler and Davies 2004).

In the momentum conservation Eq(2.6), the first term is the local acceleration term; the second term is the convective acceleration term; the third term is the pressure term; the fourth term is the gravity effect and the last term represents the resistance due to fiction. The dynamic routing solves the full Saint-Venant equations with all terms. When some effects do not influence the behavior evidently, the term representing them can be ignored for simplicity reasons. For example, when Eq(2.6) only considers the last two terms, it becomes Eq(2.7) where the relationship between the depth and flow is unique:

$$S - S_f = 0 \tag{2.7}$$

Eq(2.5) and Eq(2.7) forms kinematic wave model.

In general, it is difficult or impossible to solve these equations analytically. Thus numerical schemes including finite difference method, finite volume method and finite element method are developed to solve them. The finite difference method is the most popular numerical method in mainstream sewer system models such as SWMM and Infoworks.

### 2.4.3 Optimization methods for storm sewer network design

As the storm sewer network is generally capital intensive, it is always of great interest for designers to pursue a cost effective design. Furthermore, if other factors such as the flood risk, environmental issues, developing sustainability, etc. are of concern in design, it is also important to optimally dispose all the resources and attain an optimal design that well balances different objectives.

Various early optimization techniques have been applied to sewer network design. Elimam et al. (1989) applied linear programming to minimize the cost of the system whist ensuring no surcharge or flood with a fixed design discharge in each pipe. Joneja et al. (1978) employed non-linear programming to solve the optimization problem. Mays and Wenzel (1976), Yen et al. (1984) and Kulkarni and Khanna (1985) used dynamic programming (DP) in the design of sewer networks. These methods, optimized the cost of the system in a certain extent, however, are still far away from the global systematical optimization due to the individual consideration for each pipe in the system.

Heuristic methods, generally driven by artificial intelligence, have hugely facilitated engineering optimization problems. Being flexible for both continuous and discrete problems and without any assumption about the landscape of optimization objectives, the heuristic method has performed well on sewer network optimization. Among these techniques, the genetic algorithm (GA) (Holland 1975) has been the most popular algorithm applied in this area with great success (Walters and Lohbeck 1993; Parker et al. 2000; Liang et al. 2004; Afshar et al. 2005; Afshar 2006). Other optimization techniques, such as the ant colony optimization method (Afshar 2010), the particle swarm optimization method (Izquierdo et al. 2008) and the cellular automata (Guo et al. 2007) have also been applied to this task successfully.

## 2.5 Summary

This chapter reviews subjects relevant to decision making of optimal storm sewer network design considering flood risk, which provides a technical background for further discussion and methodology development in this thesis.

The problem of optimal storm sewer network design consists of many components. Though researches are carried out in this area intensively, yet there are gaps towards complete understanding and ideal solution of the problem. For instance, uncertainty is widely divided into aleatory and epistemic uncertainties, but the separate treatment of them in practice does not bring enough attention and an effective methodology to realise this task is lacked; The storm sewer network design is conventionally considered with predefined acceptable flood levels but is not comprehensively discussed as a decision making problem under the stochastic characteristic of flood risk; Optimization is applied to storm sewer network design but usually without flood risk consideration. Hence it is necessary to explore these aspects in order to improve the decision making in storm sewer network design.

# Chapter 3 Uncertainty understanding, representation and propagation: handling uncertainty in flood evaluation

Uncertainty is generally present in engineering and has been widely discussed in the literature. In scientific practice, it is often necessary to include uncertainty analysis in an integrated approach. However, in many cases, not enough attention is paid on the appropriate understanding and representation of uncertainty. It is very usual to propagate uncertainty through a model but the way of doing so often lacks profound thinking. This chapter mainly focuses on the understanding and representation of uncertainty and on the propagation of uncertainty in modelling. It starts with a review of the rules and characteristics of mathematical languages for describing uncertainty. The mathematical language serves as a basic tool in uncertainty analysis. Then, the understanding and representation of uncertainty are illustrated. In the third section the propagation of uncertainty through modelling is discussed. The inputs may be represented by different languages according to available information. The propagation methods are given under different situations when uncertainties are represented either by probability or by imprecise probability. It is highlighted that the coherence of the uncertainty nature should be ensured during the propagation. A general procedure for uncertainty propagation is proposed in the fourth section. The principle and methodology in this chapter is generally applicable in any area involving uncertainty. A simple flood evaluation case application is given in the last section.

## 3.1 Mathematical languages for uncertainty

A formal language is generally required in order to encode uncertainty. The frequently used mathematical languages for uncertainty representation including probability, fuzzy set, possibility and random set have been reviewed in Chapter 2. In this section the technical background for the mathematical languages is recalled. Focus is on the axioms on which the theories are based and on the understanding of the languages with simple examples if necessary. It sets the basis for further discussion about uncertainty representation and propagation in the subsequent sections.

### 3.1.1 Probability

Probability is the traditional way to represent uncertainty and it is best understood. O'Hagan and Oakley (2004) showed their firm opinion that the uniquely suitable construct to describe uncertainty is probability, though this is not agreed by some other researchers (Hall 2003).

Probability is a function:

$$P : A \to [0,1] \tag{3.1}$$

where $A$ is the universal set containing all relevant events. For any set $X \subset A$, it satisfies:

Axiom 1: $P(X) \geq 0$

Axiom 2: $P(A) = 1$

Axiom 3: For any sequence of disjoint sets $X_i \in X : P(\cup X_i) = \sum_i P(X_i), i \in N$. This is the additivity axiom of probability.

In probability theory, the probability density function (PDF) $P(x)$ (see Eq(3.1)) or the cumulative density function (CDF)

$$F_X(x) = P(X \leq x) \tag{3.2}$$

are usually employed to represent a random variable.

The probability theory is a classical method to quantify uncertainty. There are basically two interpretations of probability according to the nature of uncertainty. The first is the frequentistic interpretation, which defines probability as the relative frequency. Aleatory uncertainty is interpreted within this scope. The second is from the Bayesian viewpoint, which defines probability as a degree of belief without the definition of a strict relative frequency concept. Epistemic uncertainty described by probability is under this category.

### 3.1.2 Fuzzy sets

Fuzzy sets are sets whose elements have a degree of membership. It permits a gradual assessment of the membership of the elements in a set. The fuzzy set theory has applications in a wide range of domains in which information is incomplete or imprecise. A fuzzy set is a pair ($A$, $m$) where $A$ is a set and

$$m : A \rightarrow [0,1] \tag{3.3}$$

For each $x \in A$, $m(x)$ is the grade of the membership of $x$ in ($A$, $m$). $x$ is not included in the fuzzy set ($A$, $m$) if $m(x)$=0, while $x$ is fully included if $m(x)$=1.

Fig 3.1 gives a typical fuzzy expression of an unknown parameter, where the true value of the parameter is an unknown fixed value.

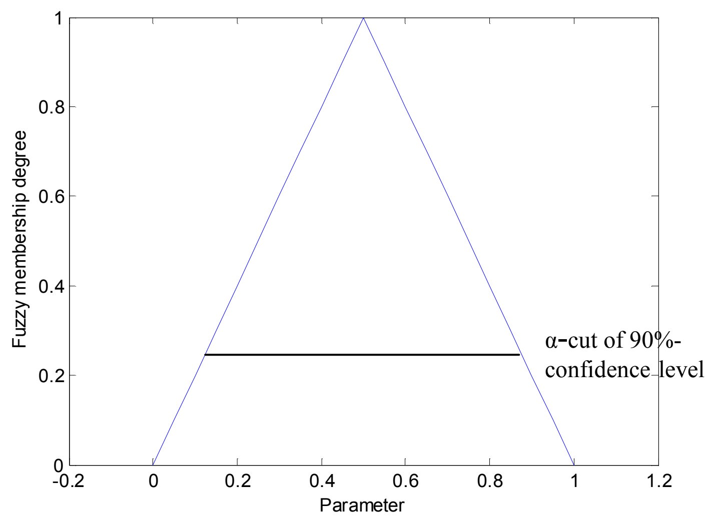



Fig 3.1 Fuzzy set expression of an unknown parameter

The fuzzy set theory is not a theory of uncertainty challenging probability theory, but a complementary choice for it. It is an abstract entity defined by a membership function, which is a pure mathematical concept without operational meaning (Dubois 2010). Fuzzy set is generally used to represent epistemic uncertainty as it represents the membership of elements usually in one's belief.

### 3.1.3 Possibility

Possibility is introduced as an extension of fuzzy sets and fuzzy logic (Zadeh 1978). The degree of the membership in the fuzzy set theory can also be viewed as possibility. However, possibility expresses more than fuzzy set does. A distribution of possibility $pos(x)$ is a function from the universal set $A$ to [0,1]

$$pos: A \rightarrow [0,1] \tag{3.4}$$

It has the rules that:

Axiom 1: $\text{pos}(\Phi) = 0$

Axiom 2: $\text{pos}(A) = 1$

Axiom 3: $\text{pos}(U \cup V) = \max(\text{pos}(U), \text{pos}(V))$ for any disjoint subsets $U$, $V \in$ A.

Possibility distribution can usually have the shape of a fuzzy membership function as Fig 3.2 (a) shows. The information is described by means of several intervals with various levels of confidence (an interval can be obtained by finding the set of values with their possibility being greater than a specified possibility). Such a description is more satisfactory than using a single interval (Dubois 2006).

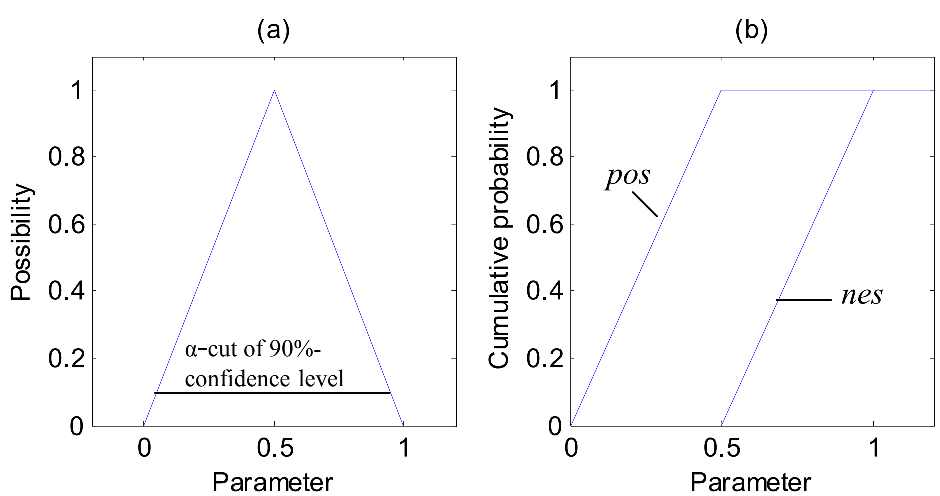



Fig 3.2 A possibility distribution and its possibility and necessity

A possibility distribution can also be seen as a concise encoding of a special probability family (Dubois 2010). Possibilities are numbers that generally stand for upper probability bounds. The necessity measure defined as:

$$nes(U) = 1 - pos(\overline{U}) \tag{3.5}$$

stands for the lower probability bound, where $\overline{U}$ is the complement of $U$. Thus possibility can be transformed to a form with imprecise probability enclosed by two bounds as shown in Fig 3.2 (b). The imprecise probability can be seen as a two-dimension uncertainty representation. When a probability distribution (representing the first dimension of uncertainty) cannot be elicited with certainty, an interval (representing the second dimension of uncertainty) is used to represent each percentile of the probability distribution.

Possibility can support various interpretations. It can be understood either as an objective notion referring to inherent properties of a system or as an epistemic notion referring to the state of one's knowledge (Hacking 1975; Dubois 2006). If a parameter that the possibility describes contains aleatory uncertainty, Fig 3.2 (b) represents the imprecise probability where the probability of the inherently varied parameter lies. The imprecision is caused by the incomplete knowledge. If a parameter is a fixed value not having any aleatory uncertainty, Fig 3.2 (b) represents an imprecise probability on epistemic uncertainty. The belief on the intervals containing the parameter is with various confidence levels.

### 3.1.4 Random sets and Dempster-Shafer theory

The theory of random set (Dubois and Prade 1991) is a theory of set-valued stochastic process, which is equivalent to the Dempster-Shafer theory (Dempster 1967; Shafer 1976) describing evidence. The elements of a random set are intervals or sets rather than precise point values.

Let $X$ be a universal non-empty set containing all the possible values of a variables $x$, and $\Omega(X)$ be the power set of $X$, i.e. the set of all the subsets of $X$. A random set is defined as a pair ($\Omega$, $m$):

$$m : \Omega \rightarrow [0,1] \tag{3.6}$$

having the rules that $m(\Phi) = 0$ and $\sum_{A \in \Omega} m(A) = 1$, where each set $A$ contains some possible values of the variables $x \in X$ and the value $m(A)$ expresses the probability that $x \in A$ but does not belong to any other subsets of $\Omega$. The element ($A$, $m$) on $X$ assigns a probability to all the subsets of $\Omega$, while classical probability theory only considers the singleton subsets of $X$. Thus random sets theory can be viewed as a generalization of probability theory that allows the consideration of imprecision in the set definition of an event. The related imprecision of this probability can be bounded at the lower end by the belief function:

$$Bel(E) = \sum_{A \subseteq E} m(A) \tag{3.7}$$

and at the upper end by the plausibility function($Pl$):

$$Pl(E) = \sum_{A \cap E \neq \Phi} m(A) = 1 - Bel(\overline{E}) \tag{3.8}$$

The belief measures the minimum amount of evidence that fully supports $x \in E$ . Similarly the plausibility measures the maximum amount of evidence that could be linked with the event $E$. The belief and plausibility functions are equal to the necessity and possibility functions in the possibility theory, respectively. The uncertainty representation by a random set has the form of imprecise probability. Similar to the presentation by possibility, it can describe parameters with or without aleatory uncertainty.

### 3.1.5 Probability box

Williamson and Downs (1990) introduced probability box to represent a range of probability distributions. Assuming $X$ is a variable, let $\underline{F}_X$ and $\overline{F}_X$ be nondecreasing functions from the real line $R$ into [0,1], the interval $[\underline{F}_X, \overline{F}_X]$ is a probability box or p-box. $\{F_X : \underline{F}_X < F_X < \overline{F}_X\}$ is a class of CDFs which are bounded by the CDF $\underline{F}_X$ and $\overline{F}_X$. $\underline{F}_X$ and $\overline{F}_X$ are also called coherent lower and upper probabilities in Hall (2006).

A probability box describes an imprecise probability bounded by $\underline{F}_X$ and $\overline{F}_X$. Inversely, a probability box can be induced from the probability family $\boldsymbol{F}$ by:

$$\underline{F}_X(x) = \inf_{F \in F} F(x); \overline{F}_X(x) = \sup_{F \in F} F(x) \quad \forall x \in R \tag{3.9}$$

Clearly $\boldsymbol{F}$ is a subset of $\{F_X : \underline{F}_X < F_X < \overline{F}_X\}$ .

With the form of being bounded by upper and lower probability distributions, probability box and random sets, Dempster-Shafer structures are essentially equivalent (Ferson et al. 2003). It is easy to convert possibility and random sets to the form of probability box.

Probability box is also a two-dimension representation of uncertainty. A typical probability box is shown in Fig 3.3.

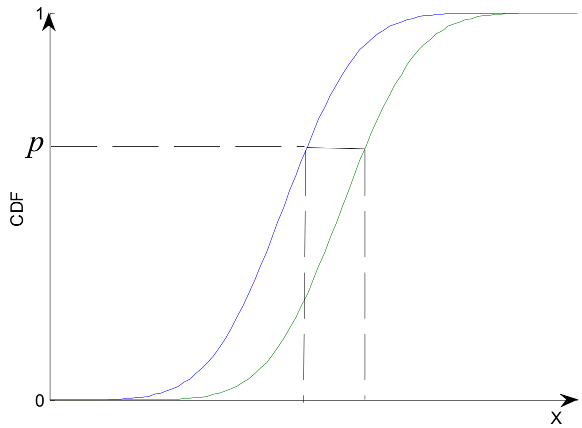



Fig 3.3 Sampling from a probability box

## 3.2 Uncertainty understanding and representation

### 3.2.1 Understanding uncertainty: interpreting aleatory and epistemic uncertainties

As stated in the literature review, aleatory and epistemic uncertainties are fundamentally different. However, not enough attention has been paid on the interpretation of the distinction between the two types of uncertainty. The mathematical tools are often used to represent uncertainty without a careful consideration of the physical meaning of the distributions (Nauta 2000).

The proper understanding of the physical meaning of uncertainty is one of the most fundamental problems when dealing with uncertainty. Consider the following two cases (uncertainty is represented by probability):

- The flood (volume or height) of a catchment over time applies to a probability distribution $F_1$;
- The flood of a catchment of 10-year return period applies to a probability distribution $F_2$ according to one's knowledge.

Although each case employs a probability distribution to describe uncertainty, the understandings are completely different. In the former case, the flood is a variable with inherent uncertainty over time. The distribution $F_1$ represents aleatory uncertainty, which is objective and will not change with one's subjective awareness. In the latter case, the 10- year flood is a fixed value, but one can only "guess" the range of the fixed value with a probability distribution $F_2$ due to one's incomplete knowledge. This epistemic uncertainty can be reduced when more information is available.

If epistemic uncertainty is also present in the first case, it makes the third case:

- The flood of a catchment over time applies to a probability distribution $F_1$. However, one cannot elicit $F_1$ with certainty due to the lack of knowledge. The uncertainty of this knowledge is described by a probability box $\boldsymbol{F}$. It is believed that the real but unknown $F_1$ lies in the range bounded by the probability box.

With this approach, two different uncertainties are described in two dimensions. In the above description, as the probability distribution $F_1$ is believed to lie in the probability box $\boldsymbol{F}$, $F_1$ represents aleatory uncertainty in the inner dimension and the range of $\boldsymbol{F}$ represents epistemic uncertainty in the outer dimension. It is natural to separate the two different types of uncertainty as this separation reasonably reveals one's belief on the true but unknown probability distribution.

### 3.2.2 Representing uncertainty: one- or two-dimension?

The use of a two-dimension uncertainty representation in the above section naturally brings forward the question: is it necessary to describe the information with "the probability of a probability"? For example, if someone tells that "the chance of the probability being 50% to have a rainy day tomorrow is high", does it make any sense? In this prediction, "the probability being 50%" is already an unsure description, is there any need to add another dimension of uncertainty "the chance being high" to this uncertain expression? Most of the time, it seems meaningless. The more sensible way in this case should be: the chance of tomorrow being rainy is high or the probability of tomorrow being rainy is 50%. However, phenomena are sometimes described with "the probability of a probability". For instance, the discharge of a runoff of T-year return period is usually given with a confidence level of 95%. It is a very basic but easily ignored issue where "the probability of a probability" (a two-dimension representation for uncertainty) is necessary and where the expression should be avoided for the reason of explicit meaning. This section is dedicated to shed some light on this issue.

It is essential to define rigorously the end point or target of the assessment, which is either a fixed value or an inherent varied value (Hoffman and Hammonds 1994). This step indicates whether aleatory uncertainty is present. If the end point of an assessment is a fixed value, it implies that the aleatory uncertainty is absent. Otherwise, aleatory uncertainty is under consideration in the problem. Under each condition, the problem can be further divided into subcategories according to the presence of epistemic uncertainty. They are discussed respectively in the following:

- When both aleatory and epistemic uncertainties are absent. The result is a fixed value.
- When aleatory uncertainty is absent and epistemic uncertainty is present. The result is expected to be a fixed value, but it will be expressed with uncertainty due to the lack of knowledge. The uncertainty can be described in the form of probability, fuzzy set, possibility or random sets.
- When aleatory uncertainty is present and epistemic uncertainty is absent. The result is an inherent varied value and is generally described by a probability distribution.
- When both aleatory and epistemic uncertainties are present. The result is expected to be an inherent variable that can be described by a probability distribution. However, due to the lack of knowledge, it is generally represented in the form of an imprecise probability. In this case a two-dimension uncertainty representation is commonly utilised.

It is clear from the above analysis that the expression of "the probability of probability" is generally required when both aleatory and epistemic uncertainties are present. This approach separates the two fundamentally different uncertainties in two dimensions: the inner dimension describes the aleatory uncertainty indicating the inherent variation, while the outer dimension represents the epistemic uncertainty revealing the imperfect knowledge. Some studies have stated the importance of separating these two uncertainties (Hoffman and Hammonds 1994; Hofer 1996; Hora 1996; Nauta 2000; Merz and Thieken 2005).

Merz and Thieken (2005) believed that the superposition of aleatory and epistemic uncertainties may lead to erroneous inferences. They gave a simple example to demonstrate this point. A population of river levee sections is located in the study area. Furthermore, a levee section is assumed to fail if the river water level exceeds a breaching height $h_B$. $h_B$ is considered as a random variable, which varies from section to section. On the other hand, for a given levee section there is epistemic uncertainty because it may not be possible to determine $h_B$ exactly. The variability of $h_B$ in the levee section population is described by the PDF $f$ ($h_B$). But the PDF is not known with certainty because of the presence of epistemic uncertainty. Fig 3.4 depicts this situation where the PDF itself illustrates the aleatory uncertainty while percentile PDF illustrates the epistemic uncertainty. Fig 3.4 A shows a situation where aleatory uncertainty dominates over epistemic uncertainty. $h_B$, the height when the levee sections fail, is relatively certain but $h_B$ varies due to the large variation of the levee population. In case B a relatively homogeneous population is supposed, but there is only little knowledge about the breaching process, where, epistemic uncertainty dominates over natural uncertainty. Combining both types of uncertainty yields a hybrid distribution. Case A and Case B give the same PDF.

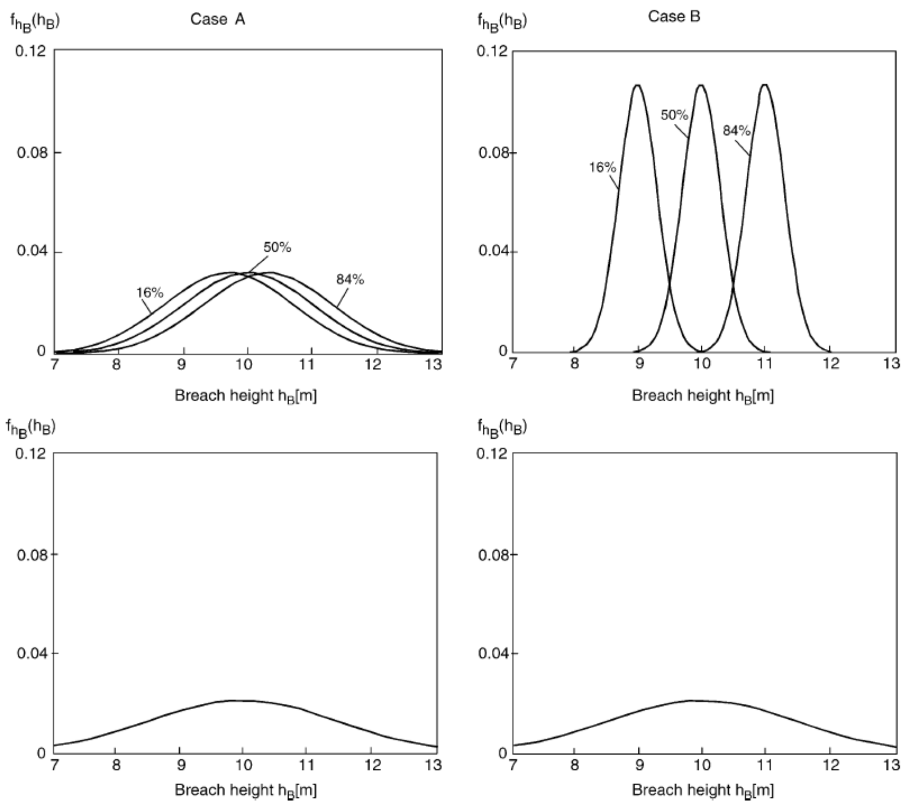


Fig 3.4 Combining aleatory and epistemic uncertainty (Merz and Thieken 2005)

This example shows that the separation of the two types of uncertainty may indicate more informed knowledge. However, it does not mean that the hybrid distribution is completely meaningless or useless. It has the physical meaning being the expected probability distribution averaged over the epistemic uncertainty. When uncertainties are propagated through a model in order to identify an uncertain variable of concern, one-dimension uncertainty analysis is generally less computationally consuming than two-dimension analysis. Hence the hybrid distribution from one-dimension analysis is sometimes valuable when the result is not required with high precision or the computation resource is limited for two-dimension analysis.

It is straightforward that the uncertainty representation should be one dimensional if there is only one type of uncertainty (being either aleatory or epistemic). For instance, in the previous case when talking about the weather of next day, if the aleatory uncertainty can be ignored and only epistemic uncertainty is present, the probability representation of the weather should be one dimensional. However, epistemic uncertainty is sometimes described by imprecise probability such as possibility or random sets when experts or analysts assign intervals to values with different confidence levels or when opinions of many experts are integrated. As mentioned in preceding sections that the imprecise probability can be viewed as a two-dimension uncertainty representation, this is an exception that one type of uncertainty is represented by two dimensions. The analyst should have a clear understanding about what the one dimension or two dimensions of uncertainty stand for, as it is essential to dispose uncertainty in the right dimension according to its nature when propagating uncertainty through a model. This point will be illustrated later in Section 3.3 of this chapter.

In conclusion, the choice of a one- or two- dimension uncertainty representation depends on the specific situation:

- If both aleatory and epistemic uncertainties are present, it is generally necessary to describe the situation with a two-dimension representation; however, a one-dimension uncertainty representation may be adopted instead due to the requirement of a coarse result or limited computational resources.
- Otherwise, one-dimension uncertainty representation is generally recommended. However, an imprecise probability may be utilised to describe epistemic uncertainty alone.

## 3.3 Uncertainty propagation

It is very common that a variable is studied through modelling when direct data are lacking. For instance, the record of real flood series is usually not long enough for a good flood risk evaluation; hence a model is usually built in order to obtain more data from model simulations. This section mainly discusses uncertainty propagation through a model.

A model can be written in its general form as:

$$Y=f\,(U_1, U_2,\dots, V_1, V_2,\dots, W_1, W_2,\dots) \tag{3.10}$$

where $Y$ is the model output, being the variable of concern, $U_i$ represents input or parameter with aleatory uncertainty, $V_i$ represents input or parameter with epistemic uncertainty, $W_i$ represents input or parameter with both aleatory and epistemic uncertainties.

When propagating uncertainty, the separation of aleatory and epistemic uncertainties should be made with care. Though it is argued that to separate aleatory and epistemic uncertainties is not easy, this is how the information should be provided. The meaning of the result could be very different if the nature of the uncertainty is wrongly treated. It may lead to wrong decision-making if the result is mistakenly understood. The uncertainty propagation should follow a basic principle of ensuring the coherence in the nature of uncertainties. More specifically, the aleatory/epistemic nature of all uncertainty sources in model inputs leads to the same nature of uncertainty in the outputs.

When there is only one type of uncertainty present, the situation is easy to handle. For example, for a model containing only aleatory uncertainty:

$$Y=f\,(U_1,\ U_2,\ldots) \tag{3.11}$$

the output $Y$ is a variable of inherent uncertainty. Similarly, if a model propagates only epistemic uncertainty:

$$Y=f\,(V_1,\ V_2,\ldots) \tag{3.12}$$

the output $Y$ is a variable of epistemic uncertainty describing the analyser's belief in what the real unknown fixed value $y$ can possibly be. If epistemic uncertainty of parameters or inputs is expressed by imprecise probability, the output expression has also a two-dimension representation.

If both types of uncertainty are present as Eq(3.10) shows, the output consists of both uncertainties according to the coherence principle. In the rest of this section, the situation with both aleatory and epistemic uncertainties present is discussed. Section 3.3.1 discusses the case in which all uncertainties are described by probability; Section 3.3.2 mainly discusses the case in which uncertainties are represented by different mathematical languages including probability, fuzzy sets, possibility, random sets and probability box. The sampling based methods are employed to propagate uncertainty for their simple principle and ease of use.

### 3.3.1 Probabilistic evaluation

#### 3.3.1.1 Separating aleatory and epistemic uncertainties (two-dimension uncertainty propagation)

When aleatory uncertainty is present, the required end point is a variable that can usually be represented by a probability distribution. If epistemic uncertainty is also present, the true end point probability distribution is impossible to know due to the lack of knowledge. A reasonable way to represent this situation is to use numerous alternative representations for the true but unknown distribution. Hence a two dimensional uncertainty representation of the resultant end point is generally required: the inner dimension handles the aleatory uncertainty denoting the inherent variation, while the outer dimension handles the epistemic uncertainty denoting the possible distributions. In the propagation process, uncertainties of different natures should be propagated in its right dimension, i.e., the uncertainty of inputs $U_1$, $U_2$,…is propagated in the inner dimension, the uncertainty of inputs $V_1$, $V_2$,…is propagated through the outer dimension, and the epistemic and aleatory uncertainties of inputs $W_1$, $W_2$,…are propagated through the inner and outer dimensions, respectively.

More specifically, if a sampling based technique (for example, MCS) is employed to propagate uncertainty, the MCS is performed in two dimensions (see Fig 3.5). In the outer dimension, parameters of epistemic uncertainty are sampled; in the inner dimension, with a set of specified parameters of epistemic uncertainty, MCS is executed by sampling parameters of aleatory uncertainty. Thus the outcome of each inner MCS is a probability distribution under a specified set of parameters of epistemic uncertainty. Sets of parameters of epistemic uncertainty are sampled in the outer dimension and each replicate in the outer loop of the simulation executes an entire MCS in the inner dimension. Consequently the outcome of the outer MCS is a group of probability distributions. These alternative distributions permit subject confidence statements about the true but unknown assessment of the end point.

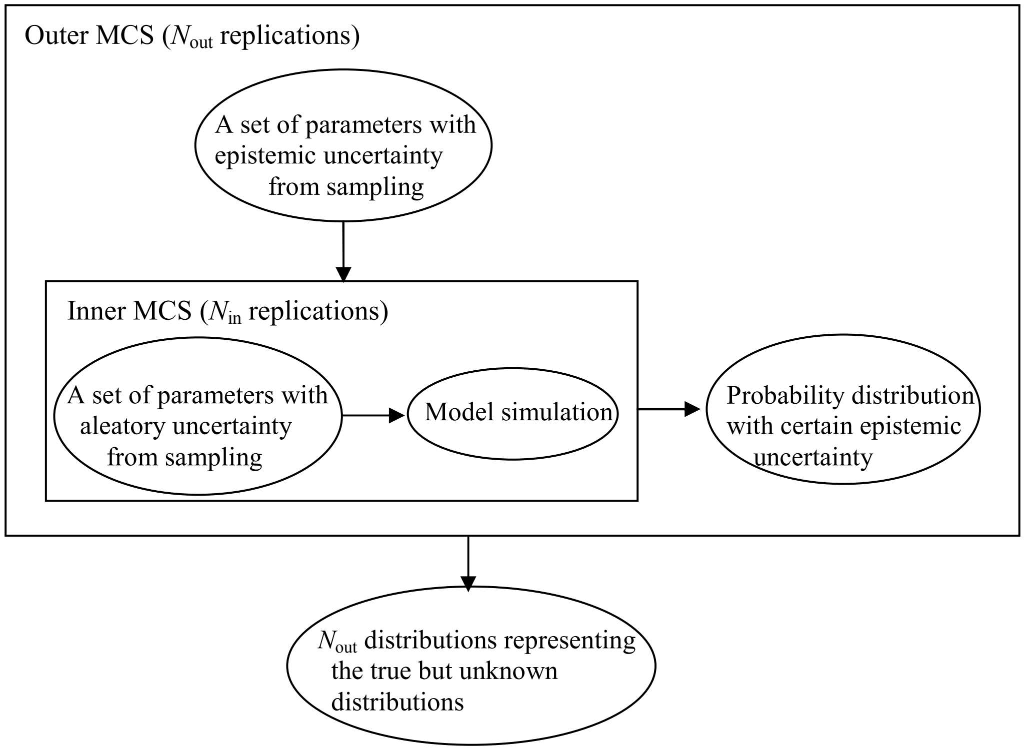



Fig 3.5 Two-dimension uncertainty propagation through a model

#### 3.3.1.2 Pooling aleatory and epistemic uncertainties (one-dimension uncertainty propagation)

Some researchers believed that the pooling of aleatory uncertainty and epistemic uncertainty leads to meaningless result with little engineering value (Grum and Aalderink 1999). However, it does have its physical meaning when combining uncertainties of different natures in one dimension. In addition, the separation of aleatory and epistemic uncertainties in uncertainty propagation model relying on a two-dimension analysis is usually very computationally demanding. Hence it is worth exploring the one-dimension analysis pooling the two types of uncertainty. If both uncertainties are propagated simultaneously, one resultant probability distribution can be identified in the end. This distribution physically represents the expected probability distribution of the inherent-varied output averaged over the epistemic uncertainty:

$$E(Y/V_1, V_2, \ldots V(W_1), V(W_2)) \tag{3.13}$$

where $V(W_1)$ is the epistemic uncertainty part of the input $W_1$. With this approach, the alternative probability distributions of the true but unknown distribution are averaged according to analyst's belief. This approach is of importance as the expectation is usually very helpful for decision-making under uncertainty.

However, this approach should be distinguished from the approach of directly using expected values of parameters of epistemic uncertainty because:

$$E(Y/V_1, V_2, \ldots) \neq Y/(\mathrm{E}(V_1), \mathrm{E}(V_2), \ldots) \tag{3.14}$$

The right side of Eq(3.14) is the expected distribution over the epistemic uncertainty, while the left side is a distribution obtained by assuming the unknown fixed parameters being their mean value. Their difference will be observed in the case study of this chapter. Though the both sides of Eq(3.14) are obtained from a one-dimension uncertainty propagation, the left side is useful in assisting decision-making if the expected distribution over epistemic uncertainty is wanted, but the right side does not have much physical meaning unless the output is not sensitive to the parameters of epistemic uncertainty.

The results from the two sides of Eq(3.14) could be very different from each other when the function $Y$ of epistemic uncertainty is highly nonlinear. A simple example is given here: let $Y= V^4+V^3$, $V$ apply to a uniform distribution in [1, 3]. It is not difficult to obtain the analytic results: $\mathrm{E}(Y)$=34.2, $Y(\mathrm{E}(V))$=24.

In Fig 3.4, it can be observed that the probability distributions averaged over epistemic uncertainty are identical though the dominances of the uncertainty in the two cases are different. It reveals that a part of information is lost when aleatory and epistemic uncertainties are not separated.

### 3.3.2 Imprecise probabilistic evaluation

It is not necessary that different uncertainties are all described by probability. For example, probability theory is not necessarily appropriate especially when expert opinion is regarded as a suitable source of information (Ross et al. 2009); and it is believed that the linguistic expert knowledge is better captured by fuzzy set. In this section, uncertainty propagation through modelling with uncertainty represented by different mathematical languages is discussed. The probability box, which represents an imprecise probability holding the unknown true probability, is utilised as a basic tool in the propagation process. The imprecision of probabilities in the term of intervals in probability box represents epistemic uncertainty. Other languages are converted to probability box with the coherence of the nature of uncertainty. Uncertainties are propagated through a model with MCS method. The resultant evaluation is represented by a probability box denoting the inherent varied output with imprecise description.

Efforts have been made to integrate uncertainties represented by different languages in modelling (Guyonnet et al. 2003, Diego 2006; Fu et al. 2010). However, not enough attention has been paid on the distinction between aleatory and epistemic uncertainties in the propagation process. The propagation of aleatory and epistemic uncertainties in wrong dimensions may lead to completely different results.

#### 3.3.2.1 Probability box for propagating uncertainty

The probability box can be viewed as a set-valued probabilistic evaluation, whose observations are intervals rather than precise point values as for probability.

The inverse functions of the lower and upper bounds $\underline{F}_X$ and $\overline{F}_X$ are introduced:

$$\underline{F}_X^{-1}(p) = \{x \mid \underline{F}_X(x) = p\} \quad \forall p \in [0,1] \tag{3.15}$$

$$\overline{F}_X^{-1}(p) = \{x \mid \overline{F}_X(x) = p\} \quad \forall p \in [0,1] \tag{3.16}$$

From Eq(3.15) and Eq(3.16), for any cumulative probability $p \in [0,1]$, there is an interval $[\underline{F}_X^{-1}, \overline{F}_X^{-1}]$ having a one to one relationship with *p*, as Fig 3.3 shows. This relationship is important for sampling from a probability box. Instead of a real number variable in the general probability distribution, the probability box is a set-valued variable. It indicates a two-dimension uncertainty representation: the variation of any probability distribution embraced in the probability box describes the aleatory uncertainty of the output whereas the set instead of a precise point value for each percentile indicates the epistemic uncertainty.

It is known from previous sections that an imprecise probability can sometimes represent a two-dimension epistemic uncertainty. In these cases, the imprecise probability should be converted to an interval with certain confidence level instead of a probability box in order to ensure the coherence of the nature of uncertainties during the propagation (see section 3.3.2.2). Such a conversion discards one dimension of the two dimensions of the epistemic uncertainty representation. Otherwise, a three-dimension uncertainty in coherence with the aleatory uncertainty and the two-dimension epistemic uncertainty will be present in output. A three-dimension uncertainty representation is generally difficult to understand and the interpretation of it does not accord with one's thinking customs.

#### 3.3.2.2 Conversion to probability box

1) Converting from probability

The conversion from probability to probability box should be different depending on the nature of the uncertainty that the probability represents.

- When the probability represents aleatory uncertainty. The probability distribution can be regarded as a special case of probability box, where the bounds $\underline{F}_X$ and $\overline{F}_X$ of the probability box overlap each other. Consequently the class of probability distributions $P(X)$ in the probability box converges to one distribution in probability and the interval corresponding to any cumulative probability $p$ the in probability box converges to a singleton $F_X^{-1}(p)$ in probability.
- When the probability represents epistemic uncertainty. As the interval at each percentile in probability box represents epistemic uncertainty, the probability should be converted to an interval with a confidence level in coherence of its nature. For example, if the epistemic uncertainty of an input can be described by a probability distribution in Fig 3.6 (a) and a confidence level of 90% is wanted, the interval of [–1.64, 1.64] from the percentiles of 5%-95% is identified to express the epistemic uncertainty. The interval with a form of a probability box is shown in Fig 3.6 (b). When sampling from this converted probability box, the interval of [–1.64, 1.64] is always obtained for any $p$. It is very different from sampling from a probability box representing aleatory uncertainty where a singleton $F_X^{-1}(p)$ is identified corresponding to a certain $p$ (see previous paragraph).

(a) (b)

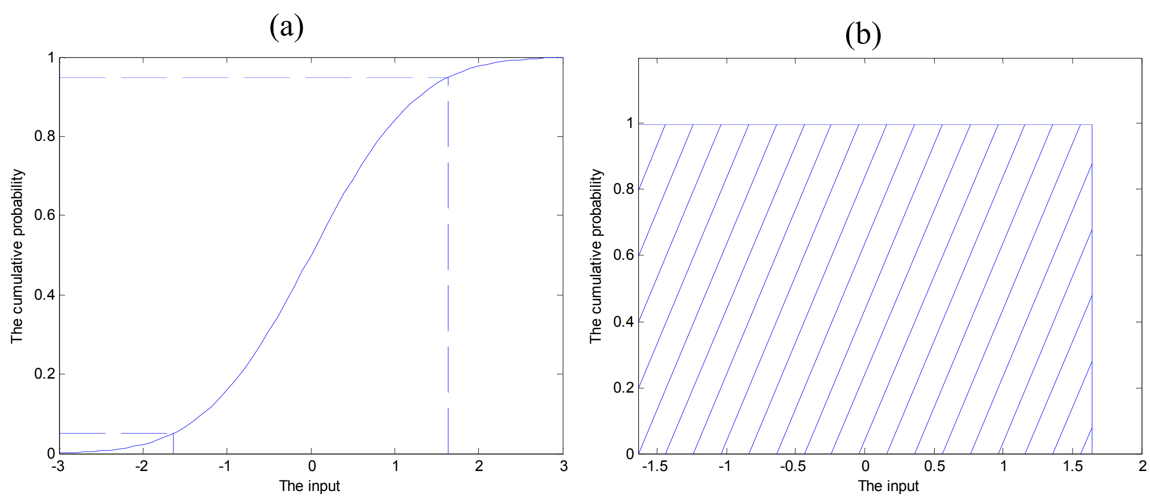



Fig 3.6 Converting probability representing aleatory uncertainty to probability box

2) Converting from fuzzy set

Fuzzy set generally describes epistemic uncertainty. An interval representing the epistemic uncertainty can be obtained with an $\alpha$-cut approach. The $\alpha$-cut of a fuzzy number is defined as the set containing all the values with membership degree no less than $\alpha \in [0,1]$. The confidence level of the true value being located in an $\alpha$-cut of the fuzzy number is 1–α. This definition is reasonable as the confidence increases when more values of lower membership are included in the $\alpha$-cut with the decrease of α. Tonon and Bernardini (1998) employed this concept to represent the probability of the true value of a parameter lying in the determined interval. The interval indicating epistemic uncertainty in Fig 3.1 with the α-cut of a 90% confidence level is represented in the form of a probability box in Fig 3.7.

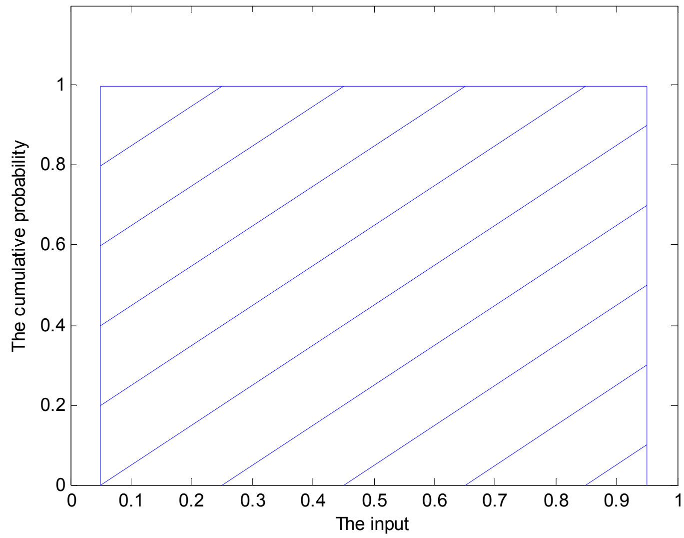



Fig 3.7 Converting fuzzy sets in Fig 3.1 to probability box with 90% confidence level

3) Converting from possibility or random set

As stated in preceding sections, possibility/random sets can support both an objective notion referring to inherent properties (aleatory uncertainty is involved) and an epistemic notion referring to the state of one's knowledge (aleatory uncertainty is not involved). The conversions from them to probability box are different depending on the type of notion that they support.

- If possibility/random sets support an objective notion. The possibility can be converted to a probability box bounded by a lower and an upper CDF as Fig 3.2 (b) shows. The random sets can be easily converted to a probability box bounded by a lower and an upper CDF determined by Eq(3.7) and Eq(3.8).
- If possibility/random sets support a subjective notion, where aleatory uncertainty is not involved. An interval should be obtained in the end to represent the epistemic uncertainty. For possibility, the $\alpha$-cut of the uncertain number is defined in the same way as defined in fuzzy number. It is the set containing all the values with possibility no less than $\alpha \in [0,1]$. The confidence level of the true value being located in an $\alpha$-cut is 1-$\alpha$. For random sets, the plausibility function, which is an equivalent to possibility, is determined by Eq(3.8). An interval is then determined with $\alpha$-cut approach similar to the conversion from possibility.

### 3.3.2.3 Propagating uncertainty using MCS

After all uncertainties are converted to an appropriate form of probability box, the MCS is employed to propagate uncertainties with the form of probability box through a model. For each uncertain input, *n* random numbers are generated from a uniform distribution on [0, 1] (*n* is the number of samples in MCS) and they are mapped to the intervals $[\underline{F}_X^{-1}, \overline{F}_X^{-1}]$ using Eq (3.15) and Eq (3.16). For parameters only representing epistemic uncertainty, it is not necessary to execute the steps of generating random numbers and mapping them to intervals as they are always represented by the same interval with any *p* as shown in Fig 3.6 (b) and Fig 3.7. The model output is evaluated for *n* times and the result of each evaluation is an interval due to imperfect knowledge of the inputs represented by intervals. Generally the maximum and minimum value of each interval of the output should be identified by an optimization method. However, in some special cases when there is a monotonic relationship between the inputs and the output, the range of the interval of the output can be simply obtained by evaluating the model with the endpoints of the inputs. After the *n* intervals of the output are obtained, the intervals are statistically analyzed in order to derive the probability box of the output. The intervals of the output can be viewed as random sets with the mass function *m* in Eq(3.6) for each interval being 1/*n*. In the end the upper and lower bounds of the probability box of the output can be identified using Eq(3.7) and Eq(3.8).

The resultant probability box is usually interpreted as the probability distribution of the output with a confidence level. It is worth mentioning that a probability box is a continuous form of random sets. Thus the idea of using the probability box to propagate uncertainties is equivalent to the proposal of using infinite random set for calculating bounds of probability proposed by Alvarez (2006).

## 3.4 A general process for uncertainty propagation through modelling

A general procedure for uncertainty propagation through modelling is presented in this section. It mainly includes the following stages:

i. Model construction. In this stage, the problem is formulated where the model is built and the output of interest is identified.

ii. Uncertainty sources investigation. This stage examines the inputs of the model and identifies the possible uncertainty sources. Uncertainty arises from different sources. The resultant uncertainty description of a required end point (model output) should be an integration of different sources of uncertainty.

iii. Uncertainty nature investigation. Following the last stage, each uncertainty source is investigated for its nature, being aleatory, epistemic or combined.

iv. Uncertainty representation. In this stage, uncertainty of inputs is represented in an appropriate way, with a mathematical language described in Section 3.1 considering its nature and the available information.

v. Uncertainty propagation. According to the representation format of uncertainty, uncertainties are propagated through a model with an appropriate method described in Section 3.3. Attention should be paid on ensuring the coherence of the nature of uncertainties when propagating them in different dimensions.

vi. Result interpretation. The resultant evaluation of the model output should be interpreted appropriately. A good result interpretation is compulsory and helpful in decision-making.

## 3.5 Applications

In this section, a simple flood estimation model is employed to illustrate and demonstrate the methods for uncertainty propagation through a model with both the probability approach and the imprecise probability approach.

The flood estimation model is a load-resistance system. The load part of the system is the runoff generated from a catchment determined by the Rational Method Eq (2.1). The resistance of the system is a drainage system to carry the rainfall water. Its capacity is considered as the discharge rate of a main channel of the system represented by Manning's equation as Eq(2.3) shows. The hydraulic radius $R_h$ in the equation is the ratio of the channel's cross-sectional area of the flow to its wetted perimeter:

$$R_{\mathrm{h}} = A / P_{\mathrm{w}} \tag{3.17}$$

where $A$ is the cross sectional area and $P_w$ is the wetted perimeter. The cross sectional area is assumed to be rectangular. Substituting Eq(3.17) into Eq(2.3), the capacity of the channel represented by a volumetric flow rate is:

$$Q_{\mathrm{V}} = \frac{1}{n}(ah)^{\frac{5}{3}} S^{\frac{1}{2}} / (a+2h)^{\frac{2}{3}} \tag{3.18}$$

where $Q_V$ is the volumetric flow rate, $a$ and $h$ are the width and the depth of the channel, respectively.

The overflow volume of the system is considered as the flood volume. Thus the overflow rate of the system is with concern:

$$f = Q_{\mathrm{p}} - Q_{\mathrm{V}} = \mathrm{c}iA - \frac{1}{n}(ah)^{\frac{5}{3}} S^{\frac{1}{2}} / (a+2h)^{\frac{2}{3}} \tag{3.19}$$

$f$ is the flood discharge rate which is a function determined by inputs and parameters. In this case, only parameter/input uncertainty is considered. Other uncertainty sources such as model uncertainty and statistical uncertainty are ignored.

Parameters without uncertainty consideration are given: the drainage area $A$ is $8\times10^4$ m$^2$; the bottom slope of the channel $S$ is 0.05; the cross sectional area of the channel is 1m×0.5m. The mathematical description of parameters with uncertainty depends on the nature of parameters and the available information. The choice of the method for the propagation of uncertainties through a model with a probabilistic approach or an imprecise probabilistic approach corresponds to the adoption of mathematical languages for the parameter descriptions. Both methods are applied in this section.

### 3.5.1 Probabilistic evaluation

Under this category, all uncertainties are described by probability. Both aleatory and epistemic uncertainties are present. The annual extreme rainfall intensity (mm/h) is represented by a Gumbel distribution:

$$P(X \le x) = \exp\left(-\exp\left(\frac{\alpha - x}{\beta}\right)\right) \tag{3.20}$$

where α and $\beta$ are the parameters to determine the Gumbel distribution. The exact value of α and $\beta$ cannot be identified with certainty due to incomplete information. The epistemic uncertainty of $\alpha$ and $\beta$ are both described by normal distributions, with the means being $\mu_\alpha = 6.5$ , $\mu_\beta = 2.4$ and the standard deviations being $\sigma_\alpha = 0.1\mu_\alpha$ , $\sigma_\beta = 0.1\mu_\beta$ . The value of the runoff coefficient relies on the catchment characteristics. It is assumed that the properties of the catchment are reasonably stable so that the aleatory uncertainty of the runoff coefficient can be ignored. However, the epistemic uncertainty is present in $c$ and it is believed to be uniformly distributed in the interval of [0.4, 0.6]. The Manning coefficient $n$ of the channel also present epistemic uncertainty and is described by a uniform probability distribution in [0.012, 0.014].

#### 3.5.1.1 Two-dimension uncertainty propagation

The aleatory and epistemic uncertainties are distinguished in two dimensions in the uncertainty propagation process. The epistemic uncertainty is disposed in the outer dimension while the aleatory uncertainty is disposed in the inner dimension. The Latin Hypercube sampling (LHS) method is employed for the reason of computational efficiency. The number of samples in inner dimension uncertainty analysis for the inherent variation of flood discharge rate is $10^4$. The sampling number is identified through limited sensitivity analysis: the probability distributions of the inherently varied flood discharge rate tend to be stable when the size of samples approaches $10^4$. The number of samples in the outer dimension analysis for epistemic uncertainty is adopted as a relatively small number $10^3$, as the analysis is performed for a confidence level, which in general does not require high precision. The overall number for model evaluations is the product of the numbers of the evaluations in both dimensions, which is $10^7$ in this case.

The probability distribution of the flood discharge rate is evaluated by the well-known Weibull formula:

$$F(y_m) = \frac{m}{N+1} \tag{3.21}$$

where $y_m$ is the $m$th order statistical value of all the events, $F(y_m)$ is the cumulative distribution function, and $N$ is the overall number of events.

The two-dimension uncertainty analysis result is illustrated in Fig 3.8. It is a combination of $10^3$ probability distributions revealing the possible flood discharge distributions. The result of a 5%-95% confidence level within the epistemic uncertainty compared with the result of the whole range is represented in Fig 3.9. The negative flood discharge rate represents the surplus capacity of the channel. The flood discharge rate of the system with different return periods can be read from the result. For example, the 20-year flood discharge rate is between 23 $m^3/h$ and 264 $m^3/h$ with a 90% confidence. The interval revealing epistemic uncertainty can only be induced with the increase knowledge about the system.

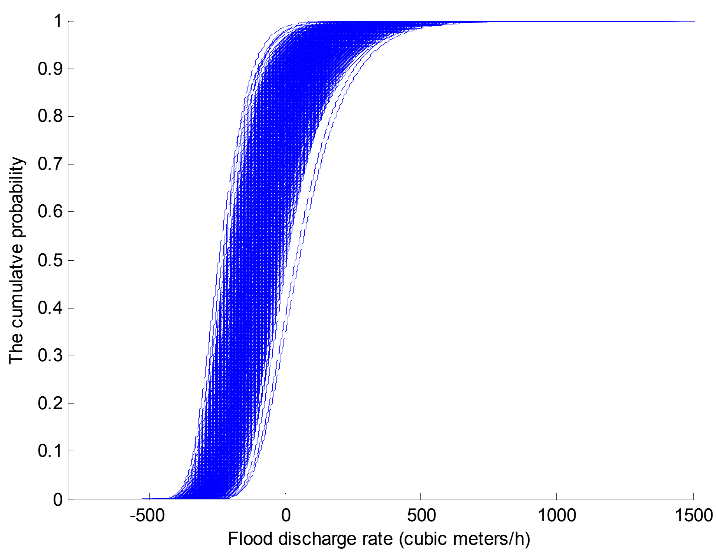



Fig 3.8 Two-dimension probabilistic uncertainty propagation for flood discharge

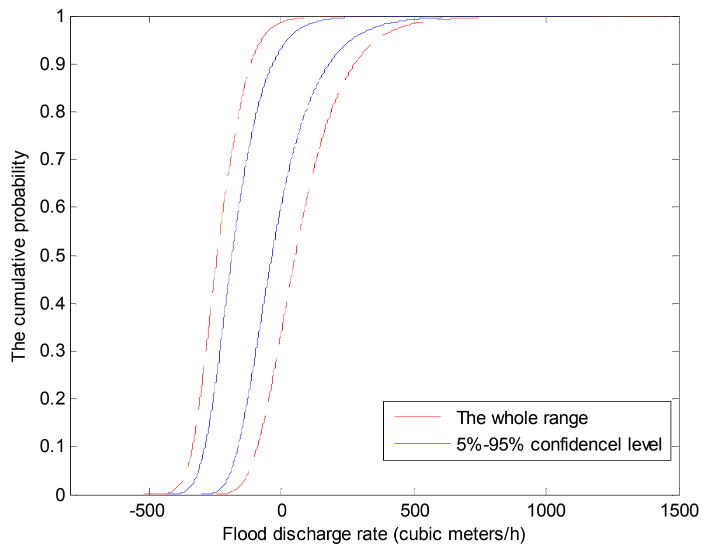



Fig 3.9 5%-95% confidence level by two-dimension probabilistic uncertainty propagation

#### 3.5.1.2 One-dimension uncertainty propagation

If the computational resource is limited or if the precision requirement of the evaluation is not high, one-dimension analysis pooling both aleatory and epistemic uncertainties may be applied instead of a two-dimension uncertainty analysis. With one-dimension propagation, both uncertainties are fed into the model and are propagated simultaneously. In the end one probability distribution is obtained representing the expected flood discharge distribution over epistemic uncertainty.

LHS with $5\times10^5$ samples, identified by limited sensitivity analysis, is performed and the resultant probability distribution as well as the 5%-95% confidence level from the previous two-dimension analysis is presented in Fig 3.10. The expected distribution is bounded by the 5%-95% confidence interval. The probability distribution ignoring epistemic uncertainty, where the parameters with epistemic uncertainty are considered as fixed values (the mean values are used as the fixed values), is also shown in the same figure for comparison. In this case these two probability distributions do not significantly deviate from each other. The probability distribution ignoring epistemic uncertainty generally underestimates flood discharges in case of extreme events and overestimates flood discharges in case of small events.

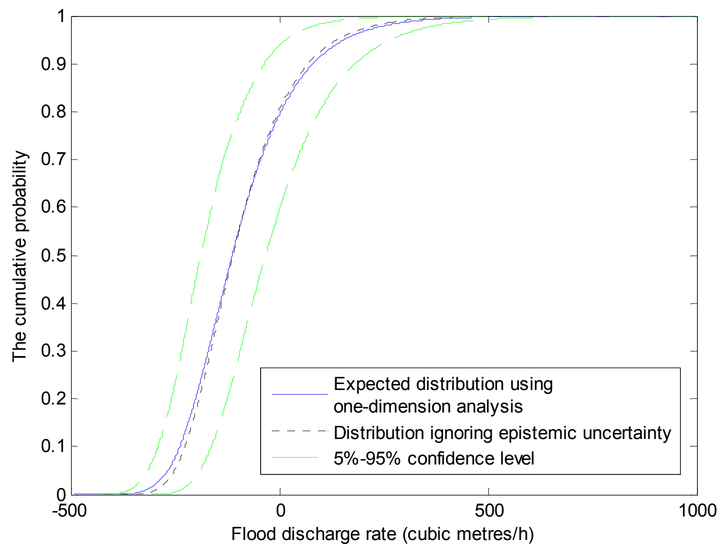



Fig 3.10 One-dimension uncertainty propagation for both aleatory and epistemic uncertainties

### 3.5.2 Imprecise probabilistic evaluation

This section studies the same model for flood estimation as in the above section, provided that some uncertainties are represented by different mathematical languages. The annual extreme rainfall intensity is described by probability box with Gumbel distribution Eq(3.20). The parameter $\alpha$ lies in the interval [6.0, 7.0] and $\beta$=2.4. The probability box of the extreme rainfall intensity is given in Fig 3.11. The epistemic uncertainty characterizing the Manning coefficient is described by fuzzy set in Fig 3.12. The $\alpha$-cut of a 90% confidence level is accounted for epistemic uncertainty analysis. The runoff coefficient is given with epistemic uncertainty. There are ten independent and equally reliable information sources, each of which believes that the true runoff coefficient lies in an interval, listed in Table 3.1. There are different ways to combine evidence from different information sources (refer to Sentz (2002) for a detailed review). This work views the independent sources as random sets with equal mass for each interval being 0.1. The plausibility function of this parameter is derived according to Eq(3.8) and is illustrated in Fig 3.13 as well as its $\alpha$-cut of 90% confidence, which is used as the interval for the epistemic uncertainty analysis. The probability box for the runoff coefficient with its belief and plausibility identified from Eq(3.7) and Eq(3.8) are also given in Fig 3.14.

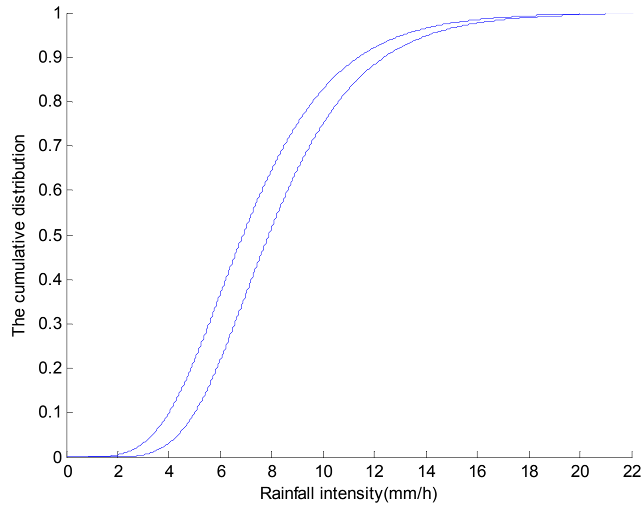



Fig 3.11 Imprecise probabilistic description of rainfall intensity

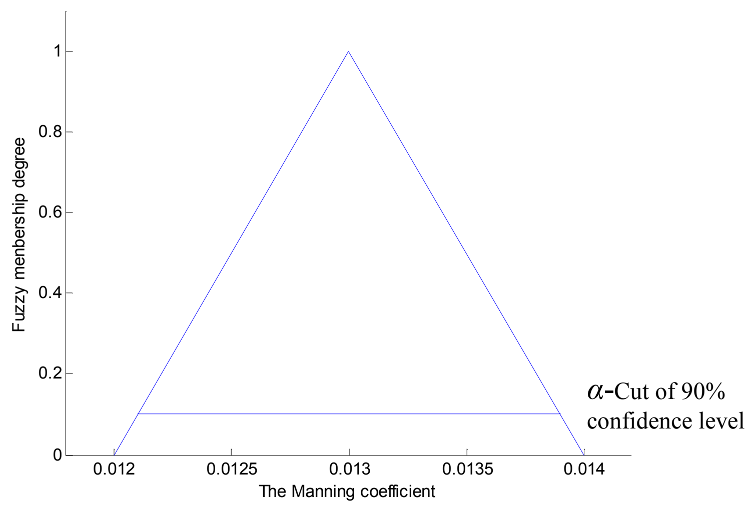



Fig 3.12 Fuzzy set expression of Manning coefficient and its $\alpha$-cut of 90% confidence level

Table 3.1 Intervals of the runoff coefficient from different evidence sources

| No | Interval of the runoff coefficient |
|---|---|
| 1 | [0.50, 0.55] |
| 2 | [0.45, 0.50] |
| 3 | [0.40, 0.60] |
| 4 | [0.45, 0.60] |
| 5 | [0.55, 0.60] |
| 6 | [0.55, 0.65] |
| 7 | [0.40, 0.45] |
| 8 | [0.50, 0.60] |
| 9 | [0.45, 0.60] |
| 10 | [0.50, 0.65] |

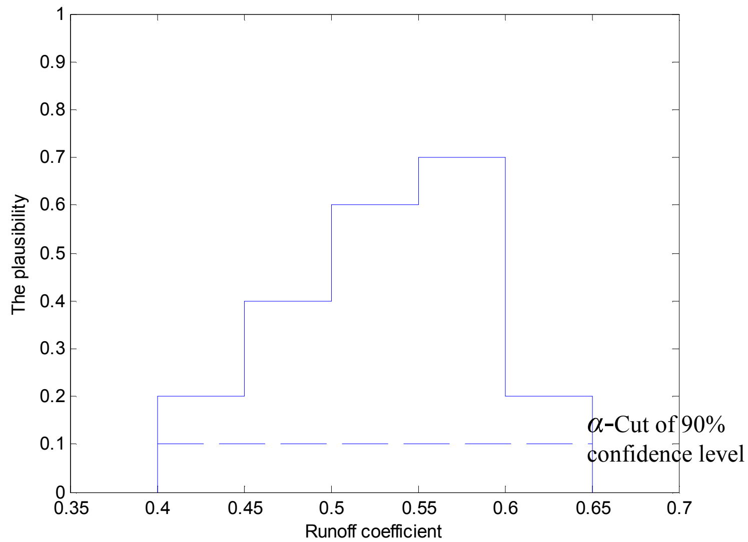



Fig 3.13 Plausibility of runoff coefficient and its $\alpha$-cut of 90% confidence level

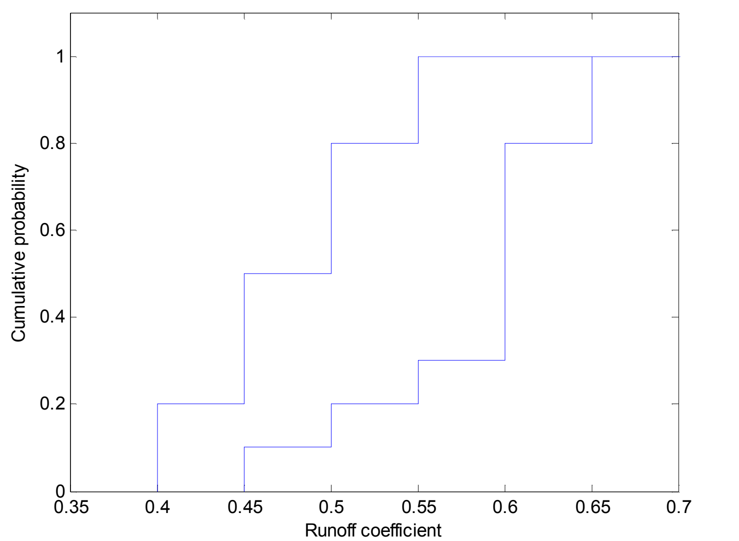



Fig 3.14 Belief and plausibility of runoff coefficient

LHS with $5\times10^5$ samples are generated. Each sample consists of intervals of the 3 uncertain parameters. The maximum and minimum values of the flood discharge rate with intervals of each sample are estimated. The extreme values are identified by evaluating the model at the endpoints of the parameter intervals as there is a monotonic relationship between the parameters and the model output. The maximum and minimum obtained from a sample forms an interval, which can be viewed as a random set of the unknown output variable with mass being $1/n$ ($n=5\times10^5$). The random sets are statistically analysed using Eq(3.7) and Eq(3.8). In the end the upper and lower bound of the probability box of the model output can be obtained in Fig 3.15. The range of the probability box describes the analyst's belief that the real probability distribution of flood discharge rate lies in the box with a 90% confidence level. If a percentile of the flood volume rate is of interest, the evaluation of it can be read from the figure, for example, the flood of a 20-year return period is evaluated as 0-354 $m^3/h$ with a 90% confidence level. Different confidence levels of epistemic uncertainty are computed and showed in the same figure. As can be observed, the interval expands when a higher level of confidence is required. It agrees with the common sense that the chance of a greater range to hold the true probability distribution is larger than that of a smaller range.

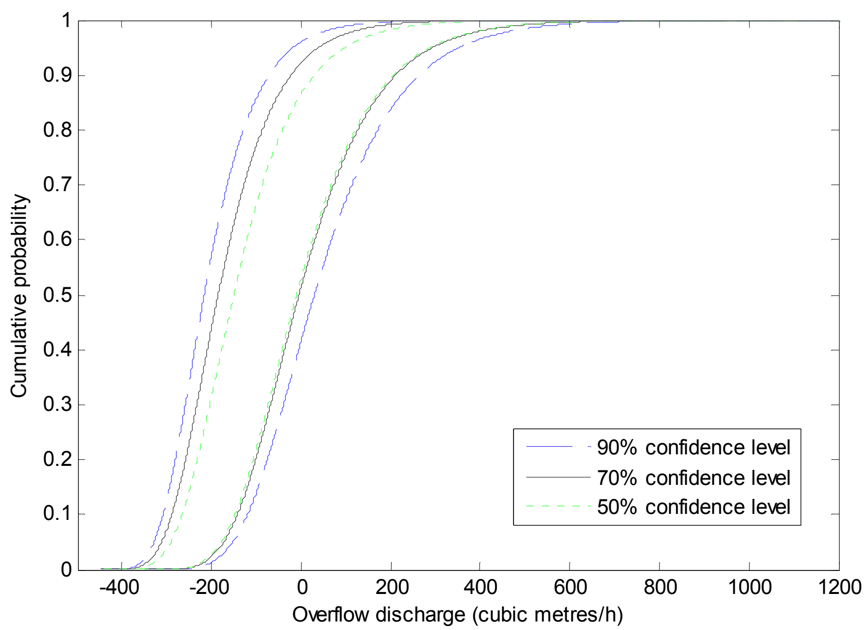



Fig 3.15 Probability boxes of the imprecise probabilistic evaluation with different confidence levels

In order to demonstrate the importance of the separation of aleatory and epistemic uncertainties in the propagation process, the epistemic uncertainty in the runoff coefficient and in the Manning coefficient is also propagated in the wrong dimension mixed with the aleatory uncertainty. In this case, instead of using a fixed interval to represent epistemic uncertainty, samples of intervals are generated for uncertain parameters represented by fuzzy sets and random sets. Intervals are sampled with different α for the fuzzy set represented parameter; and for the random sets represented parameter, intervals are sampled from its probability box form with different cumulative probability $p$ as Fig 3.14 shows. The result is shown in Fig 3.16 as well as the 90% and 50% confidence levels of the result from previous method for comparison. As this approach mixes the aleatory and epistemic uncertainty in the inner dimension analysis while the outer dimension still represents epistemic uncertainty, the result does not have any explicit physical meaning.

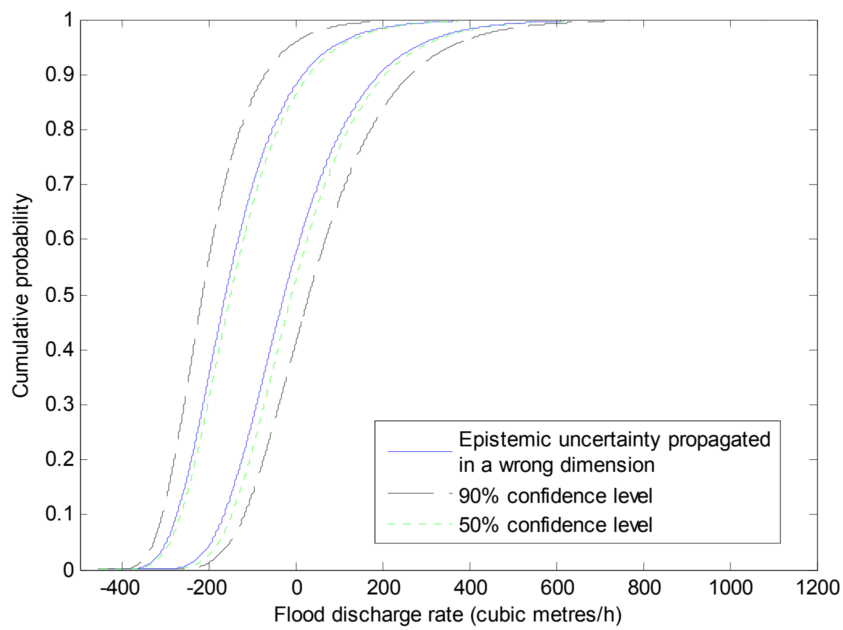



Fig 3.16 Probability box with epistemic uncertainty propagating in a wrong dimension

## 3.6 Conclusions

Uncertainty is generally present in almost all the areas in scientific practice. It is widely recognized that uncertainty analysis is an indispensable component in an integrated approach. In this chapter, focus is placed on the understanding and representation of uncertainty and uncertainty propagation through modelling.

It is important to distinguish aleatory and epistemic uncertainties when understanding uncertainty. Proper understanding is presented when different mathematical languages are employed to represent uncertainty. The appropriate representation of uncertainty is essential as it is also critical to the understanding and interpretation of uncertainty. The question about “whether a two-dimension or a one-dimension uncertainty representation is required” is illustrated in this chapter. It depends on the types of uncertainty involved in the analysis. One-dimension uncertainty is generally used when only aleatory or epistemic uncertainty is present. A two-dimension uncertainty expression distinguishes the two different types of uncertainty and has explicit meanings when both aleatory and epistemic uncertainties are present. The inner dimension represents aleatory uncertainty whereas the outer dimension describes epistemic uncertainty. The combination of the two different types of uncertainty represented by probability has the physical meaning being the expected probability distribution over the epistemic uncertainty.

The uncertainty propagation through a model is discussed in detail under different conditions when probability or other mathematical languages are employed for uncertainty representation. It is emphasized that the coherence of uncertainty nature should be ensured in the propagation process. Otherwise the resultant uncertainty representation in the model output can be significantly different and may lead to erroneous decision-making.

A general process for uncertainty propagation through modelling is proposed in this chapter. A simple case for the evaluation of the flood discharge of a catchment is studied. Uncertainties are propagated using different mathematical languages to demonstrate the idea and the proposed methodologies.

# Chapter 4 Decision-making in storm sewer network design considering flood risk

This chapter defines the problem of the storm sewer network design discussed in this thesis. It is regarded as a decision-making problem that can be solved by an optimization procedure. The flood risk, represented by a probability distribution for its stochastic character, is considered in the design. Alternatives with probabilistic evaluations are compared in order to make a decision leading to a resultant design. Several frequently used decision criteria are introduced to assist decision-making and they are applied to the storm sewer network design considering flood risk.

## 4.1 General issues

### 4.1.1 The decision-making process

As stated in literature review, decision-making mainly consists of three stages, i.e. intelligence, design and choice (Guariso et al. 1996; Quintero et al. 2005; Ryu 2008). The phase of intelligence can be interpreted as the formalization of the problem. In this stage, the environment for the decision is searched and equations are defined to represent the original situation. It allows a decision maker to find data and information available for the problem together with the models that define the situation. The phase of design is to construct alternatives to the problem. In this stage, models are developed to analyse the various alternatives. In the phase of choice, the generated alternatives are analysed and evaluated by a specified rule, and the results of alternatives are compared. In the end a particular choice is made in this stage. This may be the most difficult step as different aspects should be considered. Specified criteria will be selected for evaluating alternatives. In this thesis the focus is placed on the stage of "design" where models are developed to evaluate alternatives and on the stage of "choice" to select a proper design.

Decision-making depends on the availability of alternatives as well as the criteria applied to select a suitable solution. It ultimately culminates in the selection of a suitable design. Solutions are evaluated in relation to objectives and the decision maker is able to review and select a solution. In due course, it should be able to address all possible consequence of alternatives in order to inform and support a robust decision. As decision-making processes are generally complex, there is no one procedure applicable to all problems and situations.

### 4.1.2 Conventional methods for economical storm sewer network design

Storm sewer networks play a very important role in water management in urban areas. Without efficient drainage, storm water may cause frequent urban flooding which threatens properties, environment and public safety. Therefore, it is essential to guarantee a reliable serviceability of the storm sewer systems. On the other hand, being mostly laid underground, sewer networks are generally very capital intensive. Consequently, when a storm sewer system is designed, an appropriate design level for the protection of flood is very important as an under-designed level may bring unwelcome and unintended flood while an over-designed level can result in a waste of public funds. This encourages a pursuit of cost-effective strategies.

The design of storm sewer networks can be seen as a decision-making process which can always be formulated as an optimization problem. However, the specified criterion in the "choice" stage in the decision-making varies depending on the requirement of the design and the available computational resources. The storm sewer networks are first designed aiming to minimize the construction costs whilst ensuring no surcharge or flood with a fixed design discharge in each pipe (Mays and Wenzel 1976; Yen et al. 1984; Kulkarni and Khanna 1985; Elimam et al. 1989; Liang et al. 2004). This approach is easy to be executed but it suffers from the shortcoming that the performance of each pipe in the network is considered individually. The ignorance of systemic capacity of the storm sewer network may result in inappropriate evaluation of the designs. In this sense, the method making use of a design storm provides a better solution (Guo et al. 2007). With this method, a good system performance (e.g. no surface flood occurrence) is required under a design storm of a specified return period. The storm sewer network is designed aiming to minimize the construction cost whilst ensuring the network can convey the design storm without causing any flood. However, with this approach, it is difficult to guarantee that the specified return period of the design storm is predefined appropriately. In addition, though the concept of a design storm is used, our greatest interest is really in the return period of flooding (Butler and Davies 2004). A storm of T-year return period leading to a flood of T-year is based on the assumption that the rainfall is the only uncertainty source of a system. If some other inherent uncertainty sources such as the initial conditions of the catchment, the runoff coefficient, etc. influence the resultant flooding situation, the concept of a specified flood of certain return period (which is usually a function of storms and other influencing factors) should be considered instead of a design storm. In such cases, a storm sewer network is designed to minimize the construction cost whist ensuring that flood occurs under a specified return period.

### 4.1.3 Problem formulation

The design of storm sewer network has been greatly studied when increasing computer power sheds light on water engineering research. The storm sewer network design includes many aspects: designing the network layout, sizing the pipes, designing the slopes, locating water storage tanks and so on. In this thesis, the storm sewer network design is formulated as a pipe sizing and slope designing problem with a fixed network layout.

The alternatives for network design, formulated in the "design" process in decision-making, depend on the decision variables which are the pipe diameters (usually selected from a variety of discrete pipe sizes) and the pipe slopes (usually from a range of continuous values). An optimization is utilised in order to pursue a cost-effective design. The objective of the optimization formulated in the "choice" process depends on the decision criterion and will be discussed in next section.

There are several usual constraints that a network design should satisfy. The size of a downstream pipe is required to be not less than that of its upstream pipes:

$$D_i \le D_{i\mathrm{down}} \quad \forall i \tag{4.1}$$

where $D_i$ is the diameter of pipe $i$ and $D_{i\mathrm{down}}$ represents the diameter of the downstream pipe of pipe $i$. The excavation depth should not exceed some value due to the limits of the excavation equipment and technology:

$$h_i \le h_{\max} \quad \forall i \tag{4.2}$$

where $h_i$ is the excavation depth at manhole $i$, $h_{\max}$ is the allowed maximum excavation depth. Additionally, the height of the surface cover over pipes is required to be not less than certain value in order to make sure that the pipes are below the frost depth and they stay away from excessive live loads on the surface:

$$h_i^{'} \geq h_{\min}^{'} \quad \forall i \tag{4.3}$$

where $h_i^{'}$ is the surface cover height over the pipe near the manhole *i*, $h_{\min}^{'}$ is the minimum surface cover height allowed. The excavation depth is computed from the pipe cover height at the same place by adding the size of the pipe on it (pipe thickness is neglected here for simplicity):

$$h_i = h_i^{'} + D_j \tag{4.4}$$

Manhole depths are sometimes considered as the decision variables instead of slopes. It will not affect the optimal design, as the vertical network layout is fully determined by either one of them.

As mentioned in the previous chapter, it is known that uncertainty is widely present in nature. Grounded in the area of storm sewer network design, a decision-making involving two dimensions of uncertainty would be particularly complicated and the optimization process with a two-dimension uncertainty analysis for each alternative will be very computation-demanding if both aleatory and epistemic uncertainties are taken into account. Therefore, this study only considers aleatory uncertainty in the optimization assuming that our knowledge about the system is good. With regard to the aleatory uncertainties that lead to flood risk, storm is considered as the main uncertainty source, i.e. the stochastic flood consequence is mainly caused by the variation of rainfall; other factors are considered without uncertainty.

## 4.2 Taking flood risk into account

### 4.2.1 Construction cost and flood risk

The two important issues in designing a storm sewer network are the construction cost and the flood risk that may occur in the future. In this work, the construction cost is considered as a fixed value that can be expressed by an algebraic formula for a given storm sewer network. Flood risk is an uncertain value having a probability distribution due to the stochastic nature of the drivers of the system.

If these two values are added on an annual basis, the sum represents the possible total cost on the urban flooding area. Fig 4.1 gives a typical probability distribution of the sum: the shaded part represents the construction cost which is invariable, while the upper white part is the flood damage/consequence cost that varies for different years corresponding to the extreme rainfall of a specific year.

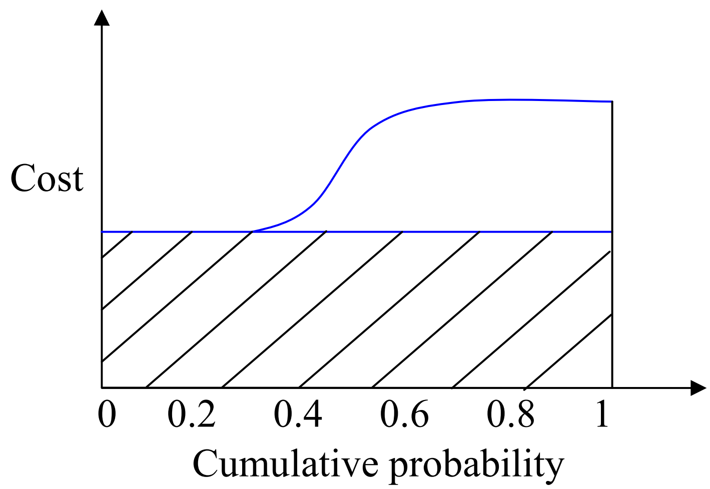



Fig 4.1 Probabilistic total cost of a storm sewer network

As a result, for each alternative storm sewer network in the design, there is a probabilistic curve representing its possible total annual cost. Hence the ranking of the candidate networks requires the comparison between probabilistic values, which, however, is not as straightforward as the comparison between fixed values.

The risk-based approach is criticised for that it necessitates the quantification of the consequences caused by flood in monetary terms, however, it is one of the essential parts of information which is necessary for a rational decision-making (Jonkman et al. 2004). Despite the limitation, the risk-based method can still provide significant rational information to the decision makers.

### 4.2.2 Probabilistic flood risk evaluation

As stated in the literature review, when there is not sufficiently long enough data of flood stages, probabilistic flood risk can be obtained through modelling in terms of hydrologic and hydraulic processes. Rainfall and other parameters are fed into the model and flood stage indicators such as exceedance flow, depth, volume and duration are identified for determining flood probability. After this, flood stages are mapped to flood damages with a flood stage-damage function. As stated in literature review, sewer flooding can be caused by different reasons. This study focuses on flood caused by hydraulic overloading in storm sewer systems.

The form of rainfall in the modelling can be continuous real or simulated series, samples of rainfall from probabilistic distributions, or design storms. The continuous rainfall series provides the most realistic information about rainfall. With it, the initial conditions of a rainfall event, automatically determined in the simulation process, do not need to be considered as varied inputs such as when using rainfall event samples. However, the continuous rainfall series usually contains a lot of dry weather periods or periods with low rainfall intensities, which have no contribution to flooding. Hence using continuous rainfall series for model simulation is usually computationally demanding. When sampling method is utilised to simulate rainfall events, the probability distribution of extreme events is firstly identified from annual maximum events or POT by statistical analysis. The samples are then generated from the derived probability distribution. After the flood stages are obtained by model simulations either using continues rainfall series or samples of rainfall events, they are mapped to flood consequences and are statistically analysed. In this chapter, the probability distribution of flood risk for a network is computed under design storms (the sampling method for the rainfall simulation will be described in Chapter 6). In this approach, the return periods of design storms are directly assigned to that of flood stages under the iso-frequency hypothesis, implying that a design storm of T-year leads to the formation of a T-year flood. The flood damage of T-year is identified from T-year flood stage-damage mapping. The probabilistic flood damage is approximated by consequences from several design events of certainty return periods placed on the probabilistic curve. Compared to methods with rainfall series or samples of rainfall events, this approach usually requires much fewer simulations of storm sewer network performance. The process is presented in Fig 4.2:

- Fig 4.2 (a) shows the relationship between return period of rainfall events and their cumulative probabilities;
- Under each design storm, the flood depth is obtained using storm sewer network performance simulation model (Fig 4.2 (b));
- The curve of Fig 4.2 (c) gives the mapping relationship between flood depths and flood consequences;

- Integrating all information from (a), (b) and (c), the flood risk curve, giving the flood damage versus cumulative probability, is presented in Fig 4.2 (d).

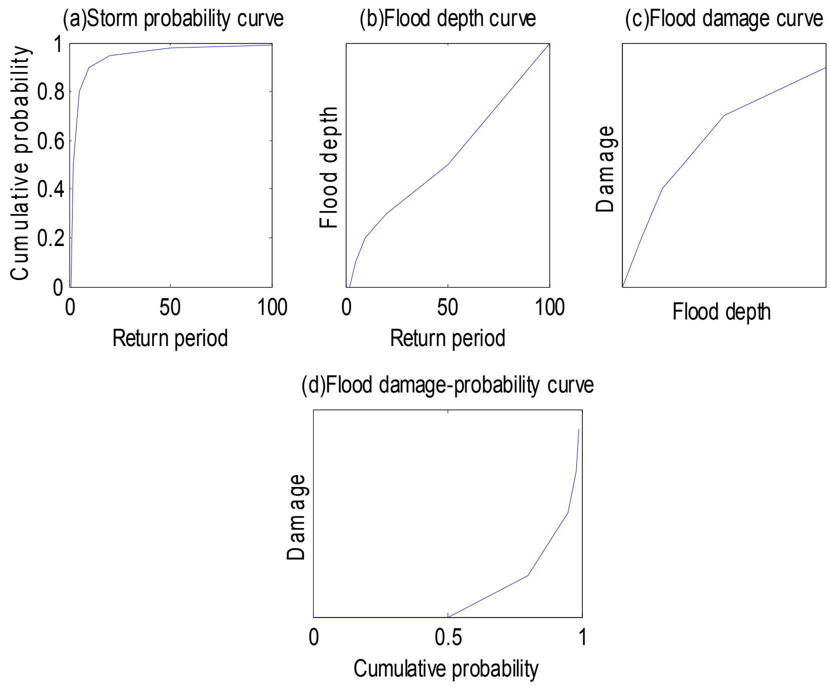



Fig 4.2 The procedure to define the flood risk curve

There are some points worth noting in the above process: 1) in this work, flood depth is chosen as the flood stage. Alternative options can be flood volume, flood water velocity, flood duration or even a combination of two or more parameters (as shown in Fig 2.1). In a specific study, the choice of the parameters representing the flood stage relies on how the flood consequence relates to those parameters. Although many studies have been focused on the damages of flooding (Smith 1994; Oliveri and Santoro 2000; Balmforth et al. 2006), no general agreement is available on the evaluation of the consequence. 2) There are many existing models for the simulation of the performance of storm sewer networks, such as SWMM, SIPSON (Djordjević et al. 2005), Infoworks and so on. Flood can be simulated by a simple atop volume, one-dimensional flow or two-dimensional surface flow. The choice of the model for simulation is related to the project objective, modelling sophistication, required accuracy, available computation recourses and so forth. 3) The flood consequence evaluation requires taking into account all the relevant effects caused by the flood, including tangible and intangible damages if possible. In this work, only property damages are considered in order to simplify the problem but it will not lose the generality of the methodology.

Design storm hydrograph can either be given directly or be created from IDF curves using the alternating block method (Chow et al. 1988). The alternating block method is briefly described (see Fig 4.3 from Morita (2008)): the design storm of T-year return period is wanted. Rain intensities *a* for 10 min, *b* for 20 min, and *c* for 30 min and so on are read from the IDF curve representing the design storm (Fig 4.3 (a)) . The hyetograph of that design storm is then arranged in a way shown in Fig 4.3 (b).

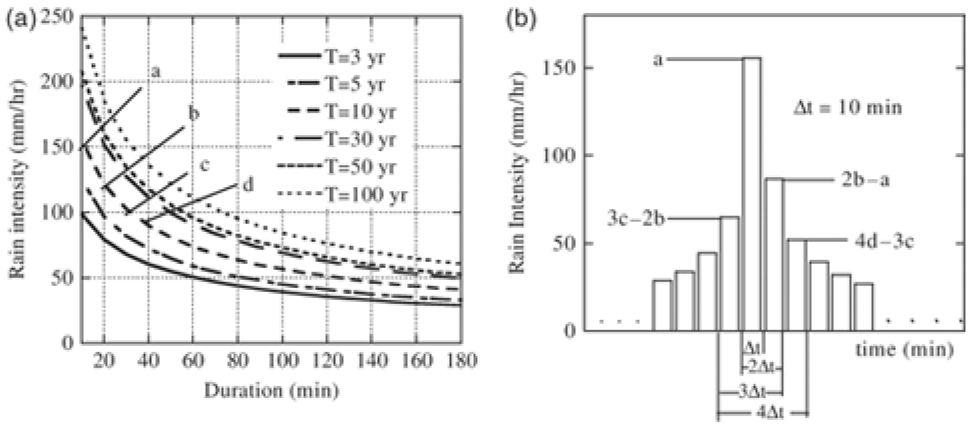



Fig 4.3 IDF relationship for hyetograph of certain return period and corresponding hyetograph by the alternating block method (Morita 2008)

### 4.2.3 Decision-making criteria

In the storm sewer network design, each alternative design can be evaluated with a construction cost and a probabilistic flood risk representing its possible flood consequence. The comparison between designs with probabilistic values is not straightforward. A design criterion, determining decision makers' preference between any two designs, is required to form the optimization objective in the decision-making process. The conventional way to handle this problem is to predefine an acceptable level of risk usually in the form of a required flood return period. Hence the probabilistic risk is turned to a deterministic constraint in the optimization. However, it is more and more recognised that this conventional approach may not be appropriate as the acceptable level of risk is predefined according to experience or precedent. Decision-making under uncertainty is widely discussed in water management (Tung et al. 1993; Korving et al. 2003). Yet no agreement on it has been achieved. In this section, some frequently used criteria for decision-making when risk is involved are discussed.

#### 4.2.3.1 The design storm based method

The conventional way to design a storm sewer network needs to predefine a level of service, usually in the form of securing no flood occurring with a certain return period. When rainfall is the only uncertainty source considered in the system causing the flood risk (i.e., other uncertainty sources can be ignored), the sewer network can be designed to transport a design storm of the specified return period without causing any flood. Using this concept, the design storm based method attempts to minimize the construction cost whist assuring no flood occurs under the specified design storm. This method turns the flood risk consideration into a constraint in the optimization.

#### 4.2.3.2 Criterion based on expected/mean flood risk

When designing storm sewer networks, an adequate design level should be based on a good balance between the construction cost and the probabilistic risk that may occur. The expected/mean value is widely used in decision-making in flood risk-based water recourses management (Goodman 1984; Korving et al. 2003; Morita 2008; Ryu and Butler 2008; Dawson et al. 2008; Korving et al. 2009). It simply and reasonably uses expectation/mean to represent the probabilistic risk. The optimization objective of the problem is to minimize the expected total cost spent on the system:

$$\min \; C_{\mathrm{T}} = \int_0^1 (C_{\mathrm{f}}(x) + C_{\mathrm{c}})dx \tag{4.5}$$

where $C_{\mathrm{T}}$ is the total cost, $C_{\mathrm{f}}(x)$ is the flood consequence which changes with the cumulative probability $x$, and $C_{\mathrm{c}}$ is the construction cost.

Using the expectation to represent the probabilistic risk is logical especially when evaluating a long-term phenomenon, as the average/overall effect of the system over time can be well characterized by the expectation.

#### 4.2.3.3 The Hurwicz criterion

The Hurwicz criterion can represent a range of risk attitudes from the most optimistic to the most pessimistic (Taha 2007). The Criterion weighs the lower and upper bounds of all alternatives by the respective weights $H_\alpha$ and (1-$H_\alpha$) where $0 \le H_\alpha \le 1$. The optimization objective is formed based on the total probabilistic cost showed in Fig 4.1:

$$\min \left(H_\alpha C_\mathrm{o} + (1 - H_\alpha)C_\mathrm{p}\right) \tag{4.6}$$

where $C_\mathrm{o}$ and $C_\mathrm{p}$ are the optimistic and pessimistic total cost. In this chapter, the probabilistic total costs at the percentile of 5% and 95% are respectively used as the optimistic and pessimistic values.

The value of $H_\alpha$ reflects decision-makers' attitude towards optimism or pessimism. A preference rating of 0 indicates a complete pessimism, while a rating of 1 specifies a complete optimism. This method can flexibly include decision-makers' risk-seeking or risk-averse altitude in the formulation the optimization objective. However, it suffers from the drawback that the parameter $H_\alpha$ is subjectively determined.

#### 4.2.3.4 Criterion based on stochastic dominance

The concept of stochastic dominance was initialised in the area of finance. It is able to deal with comparisons between probabilistic values. Tung et al. (1993) firstly applied it to the evaluation of water resources projects. The first-degree stochastic dominance test (FSD) and the second-degree stochastic dominance test (SSD) are generally performed. For the application of the tests on the decision-making in storm sewer network design, the total probabilistic costs shown in Fig 4.1 representing the possible annual total costs of alternatives are compared. The FSD checks if the value of the CDF of one candidate network is monotonically superior or equal to that of another. If it is, the former network is preferred. If the FSD test is indecisive, the SSD test, which is based on an attitude of risk-averse, can follow. Project *a* dominates projects *b* if for all the level of total cost $C_\mathrm{T}$, there is:

$$\delta F_{b-a}^{(2)}(C_\mathrm{T}) = \int_{C_\mathrm{T}}^{+\infty} \left[F_b(x) - F_a(x)\right] dx \geq 0 \tag{4.7}$$

where $F_{b-a}^{(2)}$ represents the second-degree difference of cumulative probability between *a* and *b*, $F(x)$ is the cumulative probability at *x*. Fig 4.4 shows that curve *a* dominates curve *b* according to FSD and SSD.

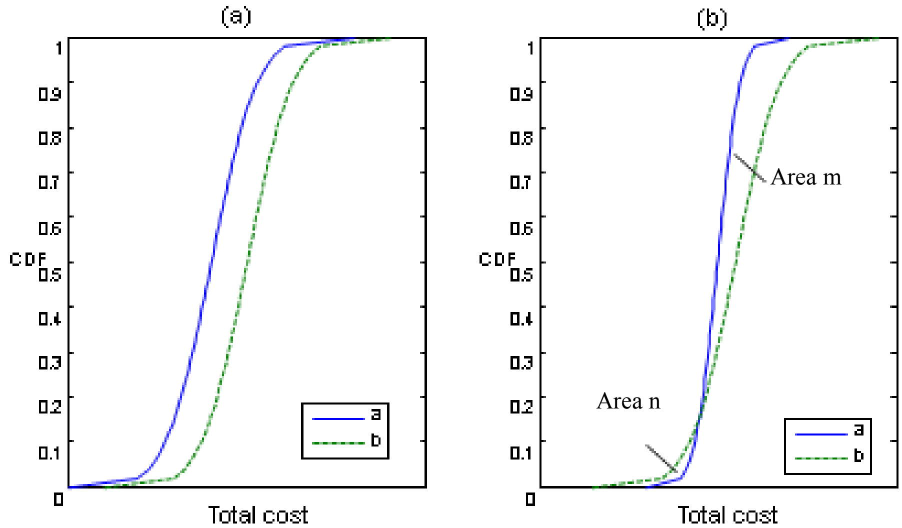



Fig 4.4 The stochastic dominance criterion

However, even with SSD, it sometimes fails to give a better alternative over another when comparing two designs. For example, in Fig 4.4 (b), if the intersect area *n* is larger than the area *m*, no favoured choice can be selected between *a* and *b*.

## 4.3 Multiple-objective optimization

### 4.3.1 Multiple-objective optimization formulation

When a decision criterion is determined for the storm sewer network design, the decision-making problem can be formulated as a single-objective optimization. The decision variables are the sizes and the slopes of the pipes. The optimization objective is specified according to the decision criterion. The single-objective optimization will be solved later in Chapter 5 and 6.

This chapter focus on the probabilistic values comparison in decision-making of storm sewer network design. Hence a multiple-objective optimization algorithm is introduced for the purpose of better comparison among the different design criteria. The minimization of the construction cost and the minimization of the expected flood risk are considered as the two objectives. The probabilistic flood risk is represented by the expected flood risk in the multiple-objective optimization assuming that the favourability of a candidate network in relation to flood risk can be fully represented by the expected flood risk. This is based on the assumption that the capacity of storm sewer networks has monotonic relationships, i.e.

if one network provides better protection under a certain design storm than another network, it always has less flood than the other one does with any other storms. Fig 4.5 presents two probabilistic flood risk under this assumption: the curves should not intersect. This assumption is reasonable as the network represented by the lower curve has better capacity in the sense of transporting water and this better capacity can be observed under all the storms.

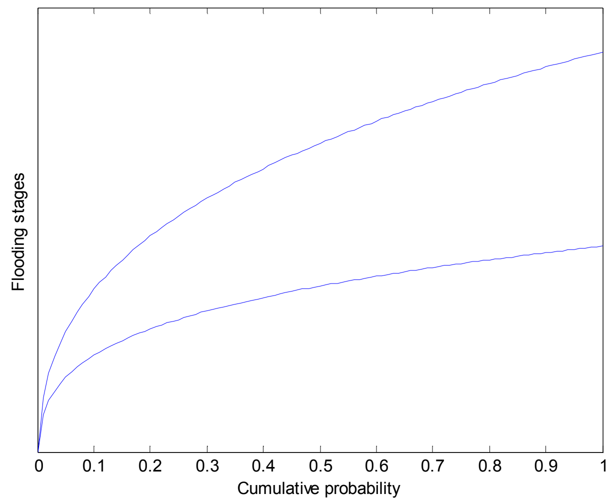



Fig 4.5 Two probabilistic flood risk curves under monotonic relationship assumption

In the end of the optimization procedure, a Pareto front of designs revealing the trade-off between the construction cost and the flood risk of alternative networks are presented. A best design is then selected from the designs in the Pareto front according to a design criterion.

The reason for performing a multiple-objective optimization instead of single-objective optimizations in this chapter is given: the optimization only needs to run once for the multiple-objective optimization algorithm and can be applied to all the decision criteria later on while the optimization algorithm needs to be performed for each decision criterion if the single-objective optimization is employed. Moreover, an approximate optimization method (for example, GA) is utilised as it is impossible to apply the enumeration method to solve this optimization problem. The stochastic optimization method may result in errors for different runs of the procedure. Hence the comparison of different design criteria using different runs of single-objective optimizations may not be reliable.

### 4.3.2 NSGA Ⅱ

For complicated optimization problems that are unsolvable with analytic methods, the true solution can only be guaranteed with full enumeration of all possible solutions. However, for continuous problems, the enumeration approach is inapplicable; and for discrete problems, it is computationally unaffordable even for a small-size problem with enumeration. The storm sewer network design is a complicated optimization problem that involves both continuous and discrete decision variables. Therefore, an efficient optimization technique is required to solve the problem.

NSGA Ⅱ (non-dominated sorting genetic algorithm) is a well-known multi-objective optimization algorithm (Deb et al. 2002). It is developed on the basis of GA (Holland 1975), which is originally designed for a single-objective optimization. The main steps in original GA usually involve:

- encode the decision variable (traditionally using binary strings but other encodings are also possible);
- generate the initial population where each individual, called chromosome, is represented by a set of parameter values that completely describe a solution and evaluate the fitness value of each chromosome with respect to the objective function measuring how good a chromosome is;
- select individuals based on fitness values;
- exert operation algorithms on the selected individuals to generate next generation of chromosomes. It usually includes crossover and mutation;
- repeat the steps of evaluating fitness of individuals and generating next generation until a convergence criterion is attained.

As NSGA Ⅱ involves comparisons of more than one objective, some steps of NSGA Ⅱ are different from those of GA. When evaluating the chromosomes in relation to optimization objectives, chromosomes are ranked based on dominance using the non-dominated sorting method. A chromosome is not dominated by another if the other chromosome is not superior to it with all the optimization objectives. The chromosomes with the same rank according to non-dominated sorting method are further evaluated with diversity estimation using the crowding distance method. Parent chromosomes are selected from the current generation based on ranks and diversity estimations. Child chromosomes are generated by operation algorithms from parent chromosomes. The algorithm combines the current and child chromosomes. The next generation are then produced based on the estimation of all the chromosomes. NSGAⅡis able to find solutions with good spread and convergence in most problems and is well recognised for its computational capacity. It has been widely applied in various disciplines such as in reservoir system optimization (Reddy and Kumar 2007), water distribution system designs (Kapelan et al. 2005), groundwater monitoring design (Reed and Minsker 2004) and so on in water engineering. In this work NSGAⅡis used as the multi-objective optimization algorithm.

## 4.4 Applications

### 4.4.1 Case studies

In this chapter, the methodology is applied to two storm sewer network designs: a synthetic network and a real network of Miljakovac, in Belgrade, Serbia. The two cases were studied by Guo et al. (2007) to determine the pipe sizes in the network under a specific design storm.

#### 4.4.1.1 The synthetic network

The synthetic network has a simple layout (see Fig 4.6) and simplified system features. It consists of 29 circular pipes, 29 standard manholes, and 1 outfall with a free outflow boundary condition. All pipes have the same Manning roughness coefficient of 0.013 and different lengths of 100, 200 or 300 meters. Pipe diameters can be chosen from 0.15m to 1.20m with 0.075 or 0.15 increments. Slopes $S_i$ should be values in the continuous interval [0.0015, 0.05]. The surface of the area is assumed to be horizontal. A subcatchment with area of $5\times10^3$ m$^2$ contributes to each manhole. Each manhole is connected to a street with the size being 100m$\times$5m. The surface flood is simply simulated by considering the streets as ponds on top of manholes. Flood depth on each street is computed as the maximum water depth from atop area.

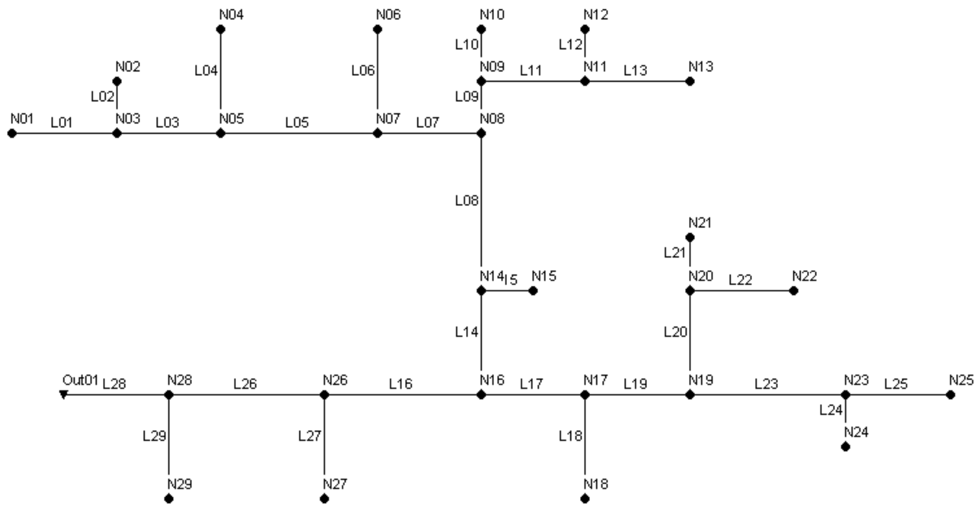



Fig 4.6 The configuration of the synthetic storm sewer network

The surface cover over pipes should be not less than 0.5m and the maximum excavation depth is 10m. The construction cost of the storm sewer system mainly consists of: (1) pipe cost; (2) earthwork; (3) manhole construction fees.

Table 4.1 Unit costs of storm sewer pipes for the synthetic network

| Pipe diameter m | unit cost £/m |
|---|---|
| 0.150 | 33.54 |
| 0.225 | 44.27 |
| 0.330 | 54.14 |
| 0.375 | 69.83 |
| 0.450 | 92.20 |
| 0.600 | 139.54 |
| 0.750 | 192.79 |
| 0.900 | 252.29 |
| 1.050 | 311.25 |
| 1.200 | 442.93 |

Table 4.1 gives the unit cost of pipes of different sizes. The unit cost of earthwork is 180 £/m$^3$, including excavation and backfill. The cross section shape of the trench is a trapezium: the width of trench bottom is $b$=$D_i$+0.5; the trench depth changes along a pipe; let $h'$ be the trench depth of a pipe at the upstream end: $h'$=$D_i$+$h_c$, where $h_c$ is the pipe cover height; the angle of the trench side wall is $\theta$=45°. The excavation volume along a pipe is integrated along a pipe length. After simplifying the integration, it has the form:

$$V_E = \frac{1}{3}\tan\theta S^2 L^3 + \frac{1}{2}(bS + 2h'\tan\theta S)L^2 + (bh' + h'^2\tan\theta)L \tag{4.8}$$

The manhole construction cost follows a function of its depth:

$$C_{mh}(h) = 292.80h + 123.21 \tag{4.9}$$

It is assumed that the storm sewer network is designed for 70 years usage. The construction price is discounted to annual cost by formula:

$$C_a = \frac{1-\alpha}{1-\alpha^n} C \tag{4.10}$$

where $C_a$ is the discounted annual construction cost, $C$ is the total construction cost, and $\alpha$ is the discount factor which can be calculated from:

$$\alpha = \frac{1}{1+r} \tag{4.11}$$

in which the benchmark interest rate $r$ is 5%.

The IDF curves of 1, 2, 5, 10, 20, 50, 100, and 200 years return period for this area are given as Fig 4.7 shows.

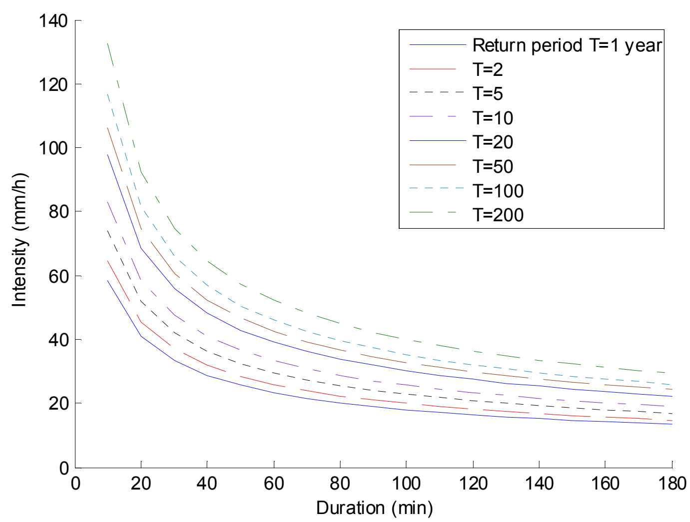



Fig 4.7 The IDF of catchment for synthetic storm sewer network design

There are 5 houses on each side of the 100 meters long street. The value of each house is £5 $\times 10^5$. All houses in the study area are assumed to be 1 or 2 stories. The damage curve is presented in Table 4.2 giving the relationship between flood depths and damages in the form of percentages of the property value (Oliveri and Santoro 2000).

Table 4.2 Percentages of total value of the damaged property (Oliveri and Santoro 2000)

| Property type | Flood depth (m) | | | | |
|---|---|---|---|---|---|
| | 0.25 | 0.50 | 0.75 | 1.00 | 1.25 |
| 1 or 2 stories, no basement (%) | 4.8 | 7.8 | 12.5 | 15.6 | 17.8 |
| 4 storeys, no basement (%) | 5.3 | 7.5 | 8.8 | 9.0 | 9.7 |

#### 4.4.1.2 The Miljakovac Network

The urban catchment Miljakovac is situated in Belgrade, Serbia. Its area covers $2.55\times10^5$ $m^2$. The storm sewer network consists of 112 circular pipes, 112 standard manholes, and 1 outfall. The layout of the network is shown in Fig 4.8.

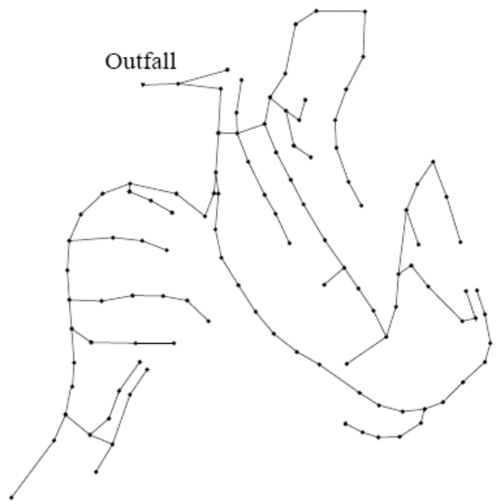



Fig 4.8 The configuration of the Miljakovac network

The construction cost function is built according to Heaney et al. (2002), which mainly includes pipeline installation expenses, trench excavation costs, bedding costs and manhole costs. The pipeline installation expenses, the bedding cost for different pipe sizes, the manhole cost function, as well as the unit excavation cost are given by Heaney et al. (2002) at the price level of January 1999. Costs are updated to 2009 values according to the US Engineering News-Record (ENR) construction cost index. The units are converted into metric units and the US dollars are exchanged to UK pounds at an exchange rate of 0.559 pounds per dollar. The lookup tables for pipeline installation expenses and for bedding costs are shown in Table 4.3.

The soil type of the area is clay and the trench excavation volume is evaluated according to Eq(4.8), but with the pipe slope S substituted by the relative slope $S_r=S-S_g$, where $S_r$ is the relative slope and $S_g$ is the ground surface slope. Let the side slope of the trench $\tan\theta=1$(Heaney et al. 2002). The unit excavation cost, including backfill and blasting, is 199.97£/$m^3$. The cost of a manhole is related to the depth of the manhole with the function:

$$C_{mh} = 1165h^{0.9317} \tag{4.12}$$

Table 4.3 Lookup table for pipeline and bedding cost for the Miljakovac network (updated from Heaney et al. 2002)

| Diameter (m) | Pipeline cost (£/m) | Trench width(m) | Bedding cost (£/m) |
|---|---|---|---|
| 0.203 | 24.63 | 0.610 | 11.42 |
| 0.254 | 30.92 | 0.610 | 12.50 |
| 0.305 | 37.73 | 0.610 | 13.76 |
| 0.381 | 48.21 | 0.914 | 18.50 |
| 0.457 | 54.76 | 0.914 | 21.33 |
| 0.610 | 78.87 | 1.219 | 30.42 |
| 0.762 | 97.48 | 1.219 | 35.37 |
| 0.914 | 143.59 | 1.829 | 52.85 |
| 1.219 | 213.82 | 2.134 | 73.81 |
| 1.524 | 309.72 | 2.438 | 98.00 |
| 1.829 | 470.35 | 3.048 | 135.63 |

On each street there are 8 houses, each of which is worth £$5\times10^5$. The pipe slopes are allowed to be in the interval of [0.003, 0.25], considering the ground slope in some area of the catchment is very steep. The surface cover over pipes should not be less than 0.5 meters and the excavation depth should not be more than 12 meters.

It is assumed that the storm sewer network is designed for 70 years use. The discounted formula and parameters are set the same as in the synthetic network.

The IDF curves have the form of $i = a/\ (duration + b)^c$, where $a$, $b$ and $c$ are parameters. The curves of different return period are shown in Fig 4.9.

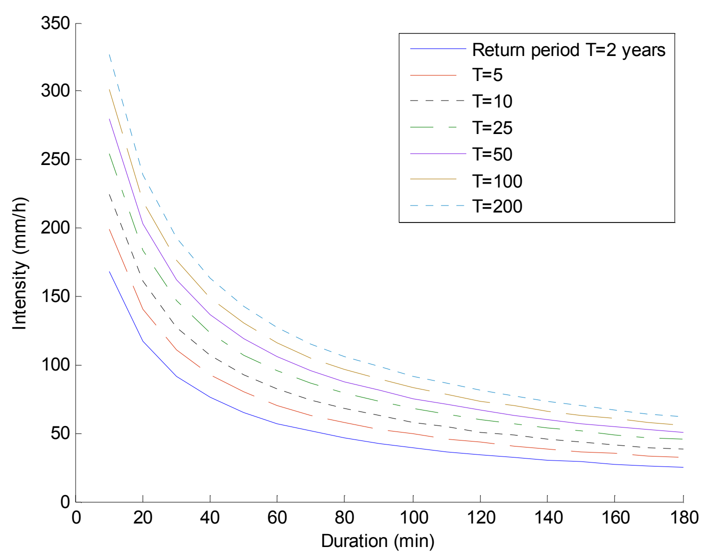



Fig 4.9 The IDF curve of catchment for the Miljakovac network

The damage curve describing the relationship between flood depths and flood damages is given in Table 4.2 (Oliveri and Santoro 2000).

## 4.4.2 Results and discussions

### 4.4.2.1 The synthetic network

It is an optimization problem with constraints that the designs should satisfy the maximum excavation, minimum pipe cover and pipe size requirements. As this chapter mainly focuses on the decision-making with various design criteria rather than on the optimization process, the way of handling constraints in GA is detailed in the next chapter. The main schemes and parameters set in NSGAII are listed in Table 4.4. The software SWMM is employed as the storm sewer hydraulic performance simulator. Surface flood is simply simulated by assuming a tank on top of each manhole.

Table 4.4 Main schemes and parameters used in NSGAII

| Population | Generations | Selection | Genetic operator | Crossover rate | Mutation rate |
|---|---|---|---|---|---|
| 200 | 1000 | Tournament selection (Parent chromosome=100; Tournament number=2) | Simulated binary crossover & Polynomial mutation | 0.9 | 0.1 |

The Pareto front of the two-objective optimization is illustrated in Fig 4.10. The flood risk is represented by the expected flood consequence, which is a characteristic value representing the probabilistic flood risk. From Fig 4.10, the construction cost starts from a threshold value. This is consistent with the existing minimum spending on the infrastructure due to the constraints of the problem. At the beginning of the trade-off between the construction cost and the expected flood risk, the increase of the construction cost is very efficient at reducing the expected flood risk cost. However, when reaching a certain level, where the capacity of the storm sewer network is adequate, further growth of the construction cost does not lead to significant reduction of the flood risk.

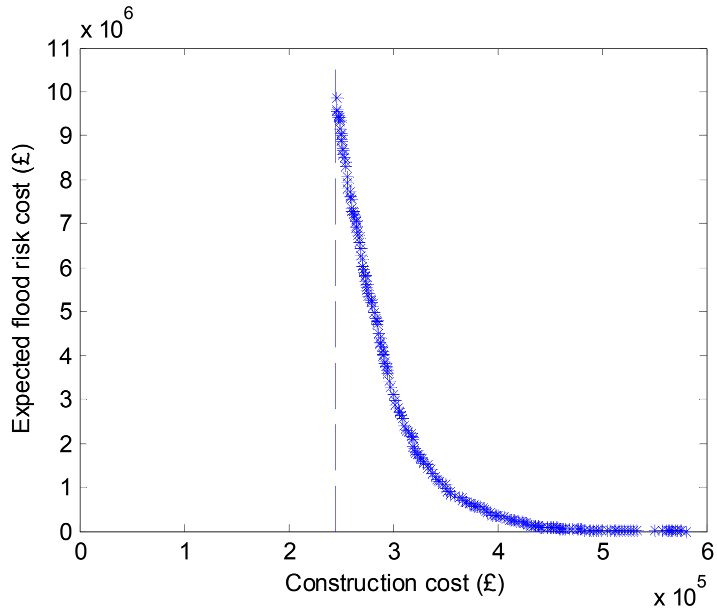


Fig 4.10 The Pareto front of multiple-objective optimization for the synthetic network

The multi-objective optimization provides a set of good designs by trading between construction cost and flood risk. Further decision-making method is required to select one best alternative for the design.

Different decision criteria listed in previous section are applied to assist the selection of a best design among the designs of the optimal Pareto front. For the design storm based method, a return period of 10 years is assumed: the construction cost is minimised with the condition that no flood occurs under the 10-year design storm. For the Hurwicz criterion, the parameter $H_\alpha$ needs to be specified. Let $H_\alpha$=0.5. The main characters of the best design obtained according to different design criteria are summarized in Table 4.5. The designs of their pipe sizes and slopes are listed in Table 4.6. The probabilistic total cost of each design, which gives the full information of a design, is represented in Fig 4.11 (the tails of the curves are magnified in the middle of the figure).

Table 4.5 Designs according to different criteria for the synthetic network

| Decision criterion | Chosen design | | |
|---|---|---|---|
| | Construction cost (£) | Expected flood risk (£) | Return period |
| Design storm based method | 513,562 | 12,302 | 10 |
| Expected cost | 483,487 | 31,191 | 2 |
| The Hurwicz criterion | 519,057 | 7,362 | 20 |
| Stochastic dominance | No best design can be given | | |

Table 4.6 Optimal designs for the synthetic network with different criteria

(a) Using design storm based method

| Pipe | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Diameter/m | 0.6 | 0.75 | 0.75 | 0.9 | 0.9 | 0.6 | 1.05 | 1.05 | 0.75 | 0.75 |
| Slope /% | 0.15 | 0.16 | 0.19 | 0.19 | 0.15 | 0.5 | 0.15 | 0.16 | 0.77 | 0.22 |
| Pipe | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Diameter/m | 0.75 | 0.75 | 0.6 | 1.05 | 0.375 | 1.2 | 0.9 | 0.375 | 0.9 | 0.75 |
| Slope /% | 0.17 | 0.15 | 0.15 | 0.15 | 1.53 | 0.16 | 0.41 | 0.71 | 0.15 | 0.17 |
| Pipe | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | |
| Diameter/m | 0.45 | 0.45 | 0.9 | 0.45 | 0.45 | 1.2 | 0.45 | 1.2 | 0.3 | |
| Slope /% | 0.76 | 0.4 | 0.18 | 0.36 | 0.3 | 0.15 | 1.12 | 0.15 | 1.47 | |

(b) Using expected value based method

| Pipe | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Diameter/m | 0.45 | 0.6 | 0.75 | 0.6 | 0.75 | 0.6 | 0.9 | 1.05 | 0.75 | 0.6 |
| Slope /% | 0.15 | 0.22 | 0.15 | 0.16 | 0.16 | 0.35 | 0.16 | 0.16 | 0.46 | 0.52 |
| Pipe | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Diameter/m | 0.75 | 0.75 | 0.75 | 1.2 | 0.375 | 1.2 | 0.9 | 0.6 | 0.75 | 0.75 |
| Slope /% | 0.15 | 0.17 | 0.15 | 0.15 | 1.8 | 0.15 | 0.41 | 0.63 | 0.22 | 0.21 |
| Pipe | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | |
| Diameter/m | 0.45 | 0.6 | 0.75 | 0.45 | 0.6 | 1.2 | 0.375 | 1.2 | 0.375 | |
| Slope /% | 0.39 | 0.15 | 0.15 | 0.55 | 0.15 | 0.15 | 1.27 | 0.2 | 1.37 | |

(c) Using the Hurwicz criterion

| Pipe | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Diameter/m | 0.6 | 0.6 | 0.9 | 0.6 | 0.9 | 0.6 | 1.05 | 1.05 | 0.75 | 0.6 |
| Slope /% | 0.16 | 0.24 | 0.15 | 0.15 | 0.16 | 0.37 | 0.16 | 0.17 | 0.46 | 0.78 |
| Pipe | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Diameter/m | 0.75 | 0.75 | 0.75 | 1.05 | 0.45 | 1.2 | 0.9 | 0.45 | 0.9 | 0.75 |
| Slope /% | 0.15 | 0.15 | 0.15 | 0.15 | 1.76 | 0.15 | 0.36 | 0.58 | 0.22 | 0.32 |
| Pipe | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | |
| Diameter/m | 0.45 | 0.6 | 0.9 | 0.45 | 0.6 | 1.2 | 0.45 | 1.2 | 0.375 | |
| Slope /% | 0.24 | 0.15 | 0.16 | 0.62 | 0.19 | 0.15 | 1.28 | 0.2 | 1.42 | |

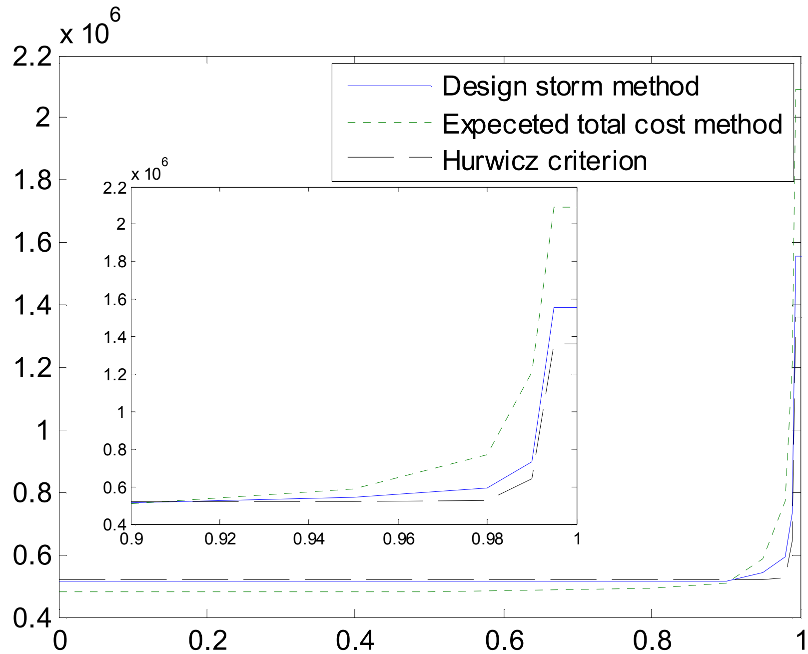



Fig 4.11 The probabilistic total cost for the synthetic network designs according to different criteria

In this case, the Hurwicz criterion (when parameter $H_\alpha$ =0.5) gives the most conservative design. The expected cost based criterion offers the cheapest design but with highest flood risk. The traditional design storm method (if the return period is set to be 10 years) provides a design between the former designs in terms of risk. These conclusions may not hold when these criteria are applied to other cases, especially when some parameters such as the return period of design storm and $H_\alpha$ can be adjusted based on decision makers' attitude towards risk. The stochastic dominance based criterion fails to give a best design, i.e. there is no design that dominates all others. This is because this criterion is actually very strict in the sense that a dominance test is conducted over the whole range of the possible values. In order to give an appropriate design, further decision-making criterion is required.

The preference of designs of the Pareto front according to the two flood risk based criteria, i.e. the expected flood risk based criterion and the Hurwicz criterion are represented in Fig 4.12. As the uncertain flood risk is represented by different characteristic values, the preferences of designs according to the two criteria are different.

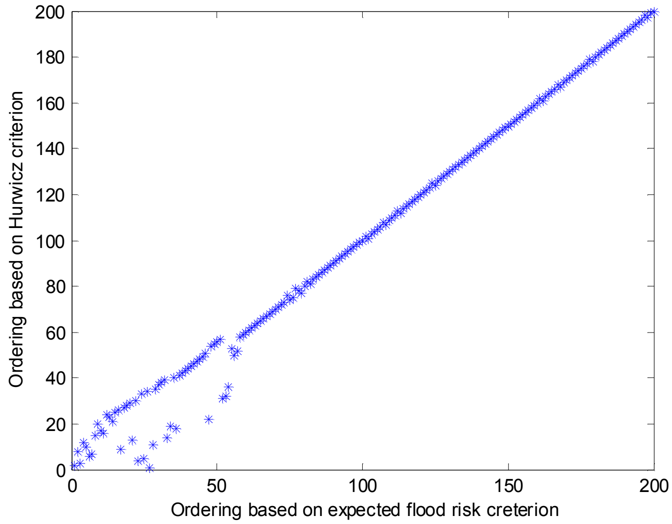



Fig 4.12 Comparison of decision-making for the synthetic network according to the expected flood risk based criterion and the Hurwicz criterion

The sensitivity of the parameter $H_\alpha$ in the Hurwicz criterion is studied. Different values are assigned to $H_\alpha$. The main characteristics of the best design obtained according to the Hurwicz criterion with different $H_\alpha$ are listed in Table 4.7. The decision-maker is more inclined to risk-seeking when $H_\alpha$ increases. As a result the network with larger $H_\alpha$ is designed with lower construction and higher risk. In this case, when $H_\alpha \leqslant 0.5$, the design is not sensitive to the variation of $H_\alpha$. It is worth mentioning that the attitude towards risk can also be adjusted through the definition of percentiles of the probabilistic value assigned to the most optimistic and pessimistic values.

Table 4.7 Sensitivity analysis of $H_\alpha$ in the Hurwicz criterion for the synthetic network design

| $H_\alpha$ | Chosen design | | |
|---|---|---|---|
| | Construction cost (£) | Expected flood risk (£) | Return period |
| 0 | 519,057 | 7,362 | 20 |
| 0.25 | 519,057 | 7,362 | 20 |
| 0.5 | 519,057 | 7,362 | 20 |
| 0.75 | 483,487 | 31,191 | 5 |
| 1 | 431,835 | 135,512 | <1 |

### 4.4.2.2 The Miljakovac network

The main schemes and parameters in NSGAII are set the same as in the synthetic network design listed in Table 4.4.

The Pareto front of the two-objective optimization is drawn in Fig 4.13. The same conclusion can be obtained as concluded in the synthetic network: The construction cost starts from a threshold value. The increase of the construction cost is very efficient at reducing the expected flood risk cost when the construction cost is low. After the capacity of the storm sewer network reaching certain level, further growth of the construction cost does not significantly reduce the flood risk cost.

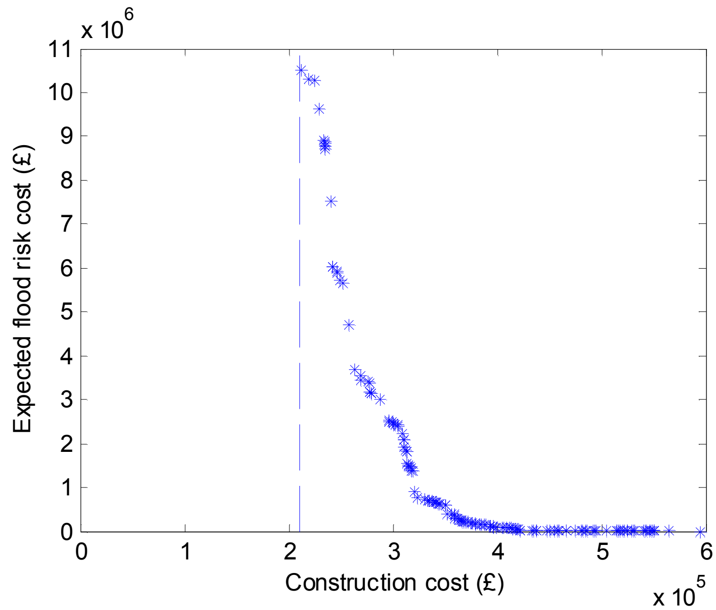



Fig 4.13 Pareto front of multiple optimization of the Miljakovac network design

The best designs from the optimal Pareto front are determined according to different decision criteria listed in previous section. The parameters in the design criteria are set the same as in the synthetic network: for the design storm based method, a return period of 10 years is assumed; the construction cost is minimised with the condition that no flood occurs under the design storm; for the Hurwicz criterion, let $H_\alpha$ =0.5. The main characteristics of the best designs obtained from different design criteria are listed in Table 4.8. The pipe sizes and slopes of the designs are not presented for legibility reasons. The probabilistic total cost of each design from different decision criteria is represented in Fig 4.14.

Table 4.8 Designs according to different criteria for the Miljakovac network

| Decision criterion | Chosen design | | |
|---|---|---|---|
| | Construction cost (£) | Expected flood risk (£) | Return period |
| Design storm based method | 480,171 | 8,837 | 10 |
| Expected cost | 422,228 | 20,169 | 5 |
| The Hurwicz criterion | 422,228 | 20,169 | 5 |
| Stochastic dominance | No best design can be given | | |

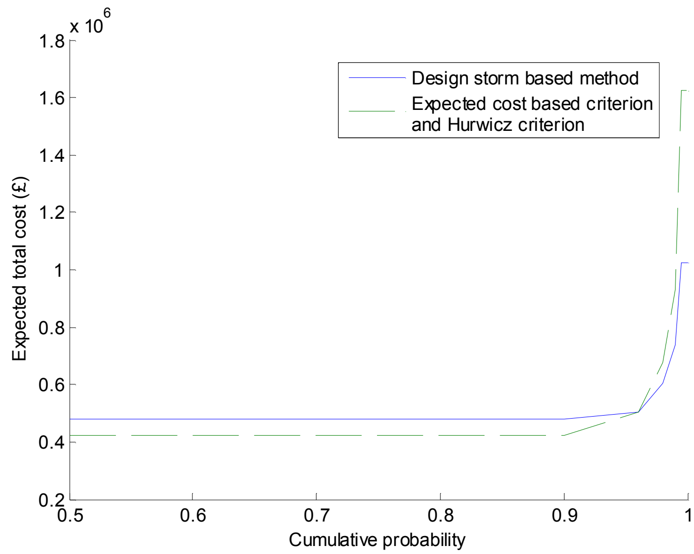



Fig 4.14 The probabilistic total cost curve for designs according to different criteria

In this case, the stochastic dominance criterion also fails to identify a best design from the Pareto front. This, again, shows the strictness of this criterion. The expected total cost and the Hurwicz criterion give the same resultant design, while the design storm of 10-year return period criterion gives a more conservative design.

The preference of designs of the Pareto front according to the expected flood risk based criterion and the Hurwicz criterion are presented in Fig 4.15.

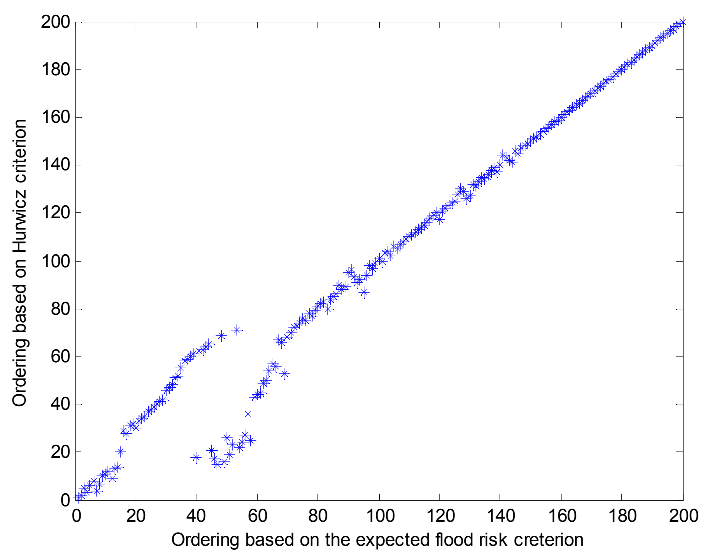



Fig 4.15 Comparison of the Miljakovac network designs according to the expected flood risk based criterion and the Hurwicz criterion

The sensitivity of the parameter $H_\alpha$ in the Hurwicz criterion is also studied in this case. The main characteristics of the best designs obtained by the Hurwicz criterion with different $H_\alpha$

are listed in Table 4.9. The trend of the increasing construction cost and decreasing flood risk when $H_\alpha$ increases can be observed. In this case, the design is not sensitive when $0.25 \leqslant H_\alpha \leqslant 0.75$.

Table 4.9 Sensitivity analysis of $H_\alpha$ in the Hurwicz criterion for the Miljakovac network design

| $H_\alpha$ | Chosen design | | |
|---|---|---|---|
| | Construction cost (£) | Expected flood risk (£) | Return period |
| 0 | 450,417 | 12,837 | 5 |
| 0.25 | 422,228 | 20,169 | 5 |
| 0.5 | 422,228 | 20,169 | 5 |
| 0.75 | 422,228 | 20,169 | 5 |
| 1 | 352,135 | 395,142 | <2 |

## 4.5 Conclusions

The storm sewer network design can be seen as a decision-making process, which can always be formulated as an optimization approach. The two important issues when designing a storm sewer network are the construction cost and the probabilistic flood risk that may occur. This chapter considers storm sewer network design, focusing on decision-making under risk consideration with different design criteria.

A process for the evaluation of probabilistic flood risk under several design storms is presented. The probabilistic flood damage is approximated based on the assumption that a T-year design storm results in a T-year flood. Some details in the procedure such as the choice of the system performance simulation model, the parameters denoting flood stages, the optimizer, etc. can be changed with an integrated consideration of different project objective, accuracy requirement and computational cost.

There is a trade-off relationship between construction cost and flood risk, i.e., usually a higher investment in the system results in a lower flood risk. Due to the uncertain property of the flood risk which is commonly described by a probability distribution, decision-making about the comparison between alternative designs is not straightforward. There is no unified way for decision-making under risk consideration. Several frequently used design criteria are introduced and discussed in this chapter. The traditional approach for storm sewer network design with a predefined design storm turns the probabilistic flood risk to a constraint in the optimization. The expected value based method uses a single statistic, with the physical meaning being expected/mean value to describe the possible flood consequence. This is appropriate especially when it applies to a long-term evaluation. The Hurwicz criterion makes use of two statistic values and weights them. The weights can reflect the attitude of decision makers towards risk. The stochastic dominance method considers flood consequence over the whole range for the FSD test and uses the SSD test with a risk-aversion attitude for further consideration if no conclusion can be drawn in the first stage. It is a strict criterion as a result it sometimes fails to offer a decision. There is no agreement about which criterion is better. The available methods for decision-making considering risk are not limited to the methods discussed here. The chosen criterion for use depends on the major concerns of a specific problem.

This chapter designs a storm sewer network with a multiple-objective optimization followed by a “choice” process making use of a decision criterion. The optimization identifies a Pareto front of optimal designs in terms of low construction cost and low flood risk. The “choice” stage then selects a best design from the Pareto front. The multiple-objective optimization is applied for the purpose of better comparison between different decision-making criteria. The designs selected according to different decision-making criteria are different as a result of different aspects being regarded by different criteria.

It is worth mentioning that once a decision criterion is determined for decision-making a priori, the storm sewer network design can be solved by a single-objective optimization, whereby the optimization objective is formed based on the determined decision criterion. The single-objective optimization will be formed and discussed in the following two chapters.

# Chapter 5 Optimal storm sewer network design: flood risk evaluated with several design storms

This chapter investigates the storm sewer network design considering flood risk with a single-objective optimization. Following the previous chapter, a decision criterion is required a priori for the formulation of the problem. The expected/mean flood risk based criterion is employed as the decision criterion to form the optimization objective considering that the storm sewer network is generally designed for a long time period usage and the expected value based criterion is particularly suitable for evaluating a long term effect. However, the choice of the criterion will not affect the generality of the methodology. A framework for the optimization of flood risk based storm sewer network design is presented. This chapter also estimates flood risk of a storm sewer network under limited number of design storms as performed in the previous chapter. GA is introduced as an optimizer for the single-objective optimization and its adapted form for this specific problem is detailed. The methodology is applied to two case studies to demonstrate its effectiveness.

## 5.1 Introduction

The traditional way for flood management is to perform an economic analysis by examining a particular design flood. The concept of risk-based approach has been around for decades. It is widely recognised that the risk-based method is sounder than the old forms of the traditional approach (Lund 2002; Voortman 2003; Morita 2008; Korving et al. 2009).

Flood risk is generally a combination of the probability of a flood occurring and the impact associated with the flood. As known in last chapter, the uncertain property of the flood risk makes the decision-making not unique. A decision criterion is required for the formulation of the problem. In practice, the expectation of the probabilistic flood risk is very often used to represent the uncertain flood risk. Hence the terminology "flood risk" also refers to the expected flood consequence in a narrower sense (Freeze et al. 1991; Korving et al. 2003; Morita 2008; Ryu and Butler 2008; Korving et al. 2009). In this chapter, the expected/mean flood risk based criterion is employed to form a single-objective optimization for the storm sewer network design.

## 5.2 The framework for single-objective optimization of storm sewer network design

### 5.2.1 Problem formulation

The storm sewer network design is a pipe sizing and slope designing problem with a fixed network layout, the same as described in the last chapter. The decision variables should satisfy the constraints represented in Eq(4.1)–Eq(4.3). The objective of the optimization of storm sewer network design considering flood risk is to balance the construction cost against the flood risk reduction. In this chapter the uncertain flood risk is represented by its expectation. Thus the single objective is formulated to minimise the sum of the construction cost and the expected flood risk as shown in Eq(4.5), which is a function of the decision variables:

$$\min f(D_i, S_i) = C_{\mathrm{c}}(D_i, S_i) + C_{\mathrm{f}}(D_i, S_i), \qquad i = 1, 2, \cdots n \tag{5.1}$$

where $f(D_i, S_i)$ is the objective function to be minimized with the decision variables being pipe diameters $D_i$ chosen from a discrete set of available values and slopes $S_i$ that lies in a continuous interval [$S_{\min}$, $S_{\max}$]; $C_{\mathrm{c}}(D_i, S_i)$ is the construction cost of the network determined by decision variables; $C_{\mathrm{f}}(D_i, S_i)$ is the expected flood risk cost which is also a function of the network design; $n$ is the number of pipes in the network.

### 5.2.2 The Framework

An outlined framework incorporating the major components for the optimal storm sewer network design considering flood risk is presented in Fig 5.1. The framework is mainly an optimization loop, where a single objective optimizer evaluates and improves trail solutions with respect to the optimization objective. For each evaluation for a trial solution executed in the optimizer, the construction cost and the flood risk of the network are both assessed. Usually the construction cost can be expressed by an algebraic formula, while the flood risk evaluation requires calling for computationally demanding system performance simulations. As the flood risk evaluation is located within the optimization loop in the framework, assuming that $n_o$ trail solutions are needed for the optimization process and $n_r$ storm sewer system performance simulations for the flood risk evaluation, the required overall number of simulations of system performance for this problem is $n_o \times n_r$. The computation is very intensive if both of the optimization process and the flood risk evaluation need a large number of evaluations. In this chapter, the probabilistic flood risk is approximated by the simulation of storm sewer system performance under several design storms (see Fig 4.2). The use of design storms significantly reduces the number of hydraulic performance simulations in comparison with using a sampling based method or the method involving simulations of a long time series. However, this method is computationally efficient at a price of a low precision of flood risk evaluation as it approximates the probabilistic flood risk with several points placed on the risk curve.

It is worth mentioning that in some cases, sampling methods may be more appropriate for the flood risk evaluation, for example when the available information about rainfall is given by probability distributions rather than design storms or when the precision of flood risk evaluation is required high. Moreover, the sampling method can flexibly include other uncertainty sources in the risk evaluation process while it is difficult for design storm based method. Regarding its main drawback being computationally expensive, we can always expect that the computational capacity will increase as time goes by. The flood risk evaluated using sampling method will be introduced in the next chapter.

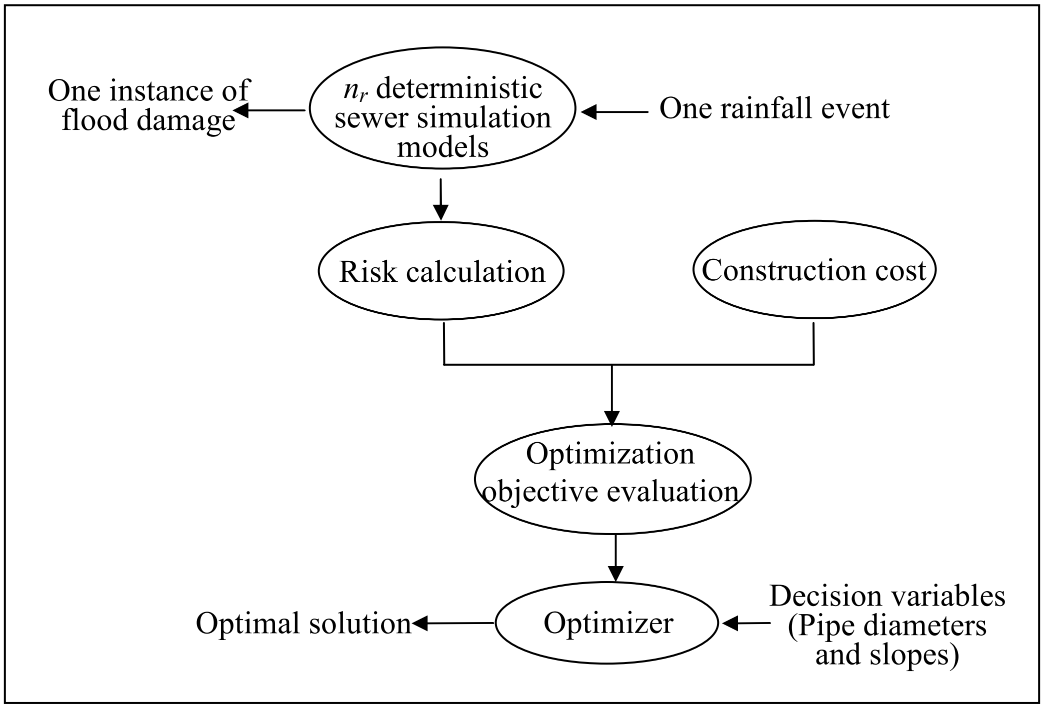



Fig 5.1 The framework for flood risk based storm sewer network design

## 5.3 Optimization by adapted GA

The GA, initially inspired by evolutionary biology, is one of the most popular heuristic optimization methods widely applied in many areas. GA is categorized as a global searching optimization method and is applicable for both continuous and discrete problems. It is employed in this chapter as the optimization algorithm for the single objective design of storm sewer network design. The brief description of original GA has been given in Chapter 4.

In the optimal storm sewer network design problem defined above, the design variables include pipe diameters and pipe slopes. Accordingly a chromosome in GA consists of two parts: the first part represents pipe diameters and the second part represents pipe slopes. The chromosomes are encoded with decimal strings in this work.

Generally GA is designed for non-constrained optimizations. If a constrained optimization is considered, a penalty function approach is usually hired to penalize the fitness of solutions when a constraint is violated. With the penalized cost, candidate solutions that violate constraints are judged inferior. As there are three constraints

Eq(4.1), Eq(4.2) and Eq(4.3) considered in our storm sewer network design problem, it is even difficult to attain a solution that satisfies all constraints when incorporating the three constraints into fitness evaluation with penalty functions simultaneously. Therefore in this work the constraints Eq(4.2) and Eq(4.3) are dealt with penalty, whereas the constraint Eq(4.1) is considered implicitly in the optimization process, i.e. the chromosomes are designed and operated in a way that implicitly satisfies it. As the pipe slopes are not involved in the constraint Eq(4.1), the implicit consideration of this constraint only involves the part of chromosomes representing pipe diameters. The initialization and the operation (including crossover and mutation) parts of GA for pipe diameters are adapted in the following:

- Initialization: When initialising the population, instead of generating all pipe sizes randomly, a pipe is randomly selected to start the size assignment process. A random valid size value is given to the starting pipe; its upstream and downstream pipes are then identified. With uneven probabilities, the upstream and downstream pipes are assigned valid random values satisfying that the upstream pipes should not be larger than the starting pipe and downstream pipes should not be smaller than it. Probability *p* is introduced as a parameter to guarantee that the initialization is well distributed in the whole possible space. For instance, let $p$=0.9, assuming there are 10 available pipe sizes, the first pipe is randomly selected and assigned size 4, for its upstream pipe, the probability of its size being 4 is 0.9, being size 3 is $(1-0.9)\times 0.9$, being size 2 is $(1-0.9)\times(1-0.9)\times 0.9$, and being size 1 is $(1-0.9)\times(1-0.9)\times 0.1$. The probability *p* is hired to prevent pipe sizes adopting extreme values early in the process. The pipe size assignment of upstream or downstream pipes of known pipes is repeated until it reaches the end of the network.
- Crossover: A crossover on two individuals is performed as follows: a pipe is randomly selected and the corresponding pipe sizes of the two individuals are compared. The upstream of the selected pipe in the new generated individual is given the corresponding sizes of the individual presenting the smaller-size in the previous comparison, while the downstream of the selected pipe is given the corresponding sizes of the other individual. In this way, the new generated individual still satisfies the constraint in Eq(4.1).

- Mutation: The mutation process on an individual is as follows: pipes are selected randomly in the network for mutation with a given probability. The range of allowable values for a selected pipe is identified referring to the constraint in Eq(4.1). The pipe is then randomly assigned a size from the valid values.

When evaluating the fitness values of individual chromosomes, the fitness value of each individual is combined of two parts: the first part is given by Eq(5.1) and the second part corresponds to the penalty on the violation of the constraints in Eq(4.2) and Eq(4.3). Considering that the objective evaluation Eq(5.1) based on several simulations of storm sewer system performance is much more computationally demanding than the evaluation of the constraints in Eq(4.2) and Eq(4.3) with algebraic expressions, the following process is proposed in order to reduce computations. In the optimization process, if a trial solution badly violates one of the constraints, its fitness is directly assigned a large penalty value, without evaluating its objective function Eq(5.1). The procedure of the adapted GA for the application of the optimal storm sewer network design considering flood risk is presented in Fig 5.2.

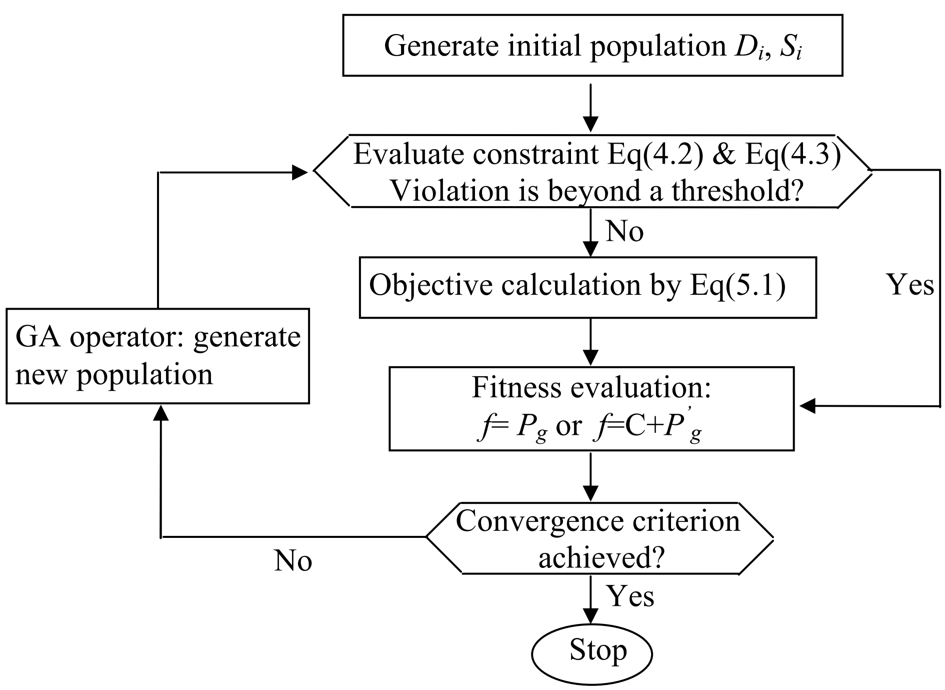



Fig 5.2 The procedure of adapted GA for optimal storm sewer network design

From Fig 5.2, if a chromosome representing a candidate solution violates the constraints in Eq(4.2) and Eq(4.3) more than a threshold value defined beforehand, a relatively large penalty is given to the fitness without evaluating its construction cost and flood risk:

$$f = P_g \tag{5.2}$$

where $f$ is the fitness value and $P_g$ is a relatively large penalty for violating a constraint Eq(4.2) or Eq(4.3), which is set to be 2-3 order higher than a roughly evaluated construction cost.

Only the "good enough" candidates with respect to the constraints of Eq(4.2) and Eq(4.3) can go to the next stage of objective function evaluation by adding the penalty of violation to the objective Eq(5.1):

$$f = C + P_g' \tag{5.3}$$

where $P_g'$ is the penalty for violating constraints in Eq(4.2) and Eq(4.3), which is proportional to the violation.

## 5.4 Applications

The methodology of the single-objective optimization for the storm sewer network design making use of several design storms is applied to the synthetic network and the Miljakovac network design described in the previous chapter.

### 5.4.1 The synthetic network

The parameters and schemes set in GA are given in Table 5.1. The parameters are identified by limited sensitivity analysis studies. The elite scheme is used in GA in order to guarantee that good individuals will not be lost during the involution of the optimization.

Table 5.1 Main schemes and parameters set in GA for the synthetic network design

| Population size | Selection method | Crossover rate | Mutation rate | Elite size | Total generation |
|---|---|---|---|---|---|
| 200 | Roulette | 0.8 | 0.05 | 2 | 1000 |

As GA is a stochastic optimization algorithm, the program was run 10 times and the best result is presented. The objective values of all generations are plotted in a logarithm scale, as shown in Fig 5.3(a). Fig 5.3 (b) is the objective values from the 50th generation with a linear scale. The objective value decreases as the search goes on. At the beginning the objective value is much higher than the objective values of subsequent generations. This is because all chromosomes violate the constraints in Eq(4.2) and Eq(4.3) badly for the first several generations, and are directly assigned a high penalty value.

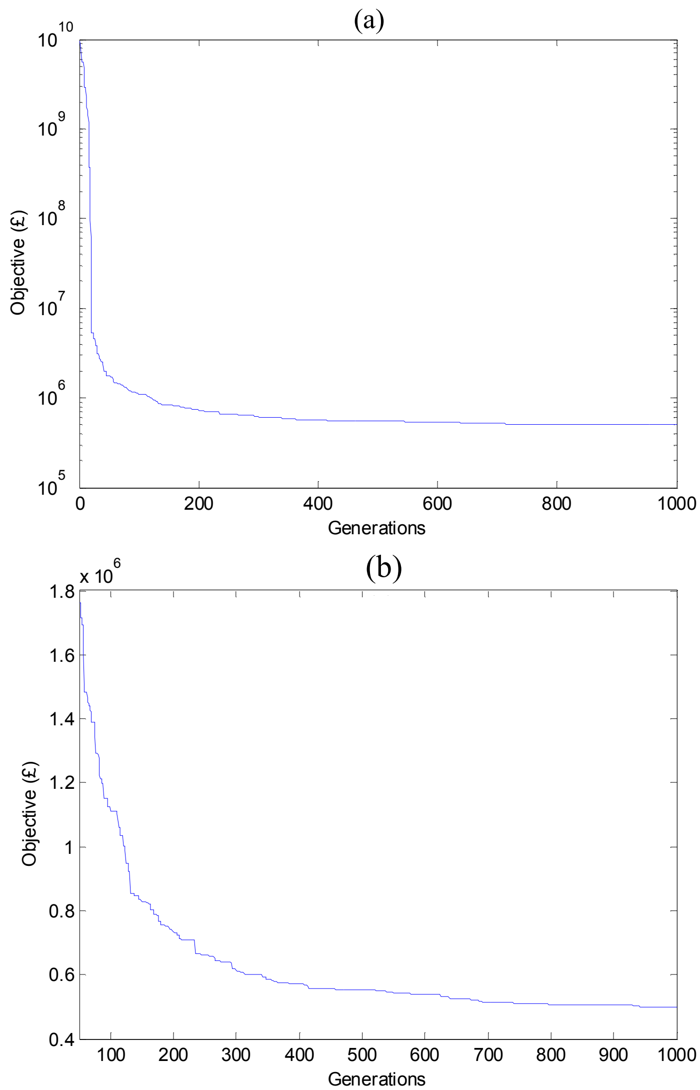



Fig 5.3 The optimization process for the synthetic network design with flood risk evaluated under several design storms

The objective value of the optimal design is £499,108. The obtained optimal design is listed in Table 5.2.

Table 5.2 Optimal design from single objective optimization for the synthetic network with flood risk evaluated under several design storms

| Pipe | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Diameter/m | 0.6 | 0.75 | 0.9 | 0.45 | 0.9 | 0.45 | 1.05 | 1.05 | 0.9 | 0.45 |
| Slope /% | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.58 | 0.15 | 0.15 | 0.76 | 0.17 |
| Pipe | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Diameter/m | 0.75 | 0.6 | 0.6 | 1.05 | 0.6 | 1.05 | 0.9 | 0.6 | 0.9 | 0.75 |
| Slope /% | 0.15 | 0.26 | 0.15 | 0.15 | 0.834 | 0.15 | 0.53 | 0.27 | 0.15 | 0.15 |
| Pipe | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | |
| Diameter/m | 0.75 | 0.75 | 0.75 | 0.45 | 0.6 | 1.05 | 0.375 | 1.05 | 0.375 | |
| Slope /% | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 1.27 | 0.36 | 1.49 | |

In order to identify the flood return period of the designed storm sewer system, the system performances of the design network are simulated under design storms. Any surface water is considered as flood occurring. In this case, no flood appears under the design storm of 5 years return period and some flood occurs with the design storm of 10 years return period under the designed storm sewer system. Hence the flood return period of this designed system is between 5 to 10 years. It is believed that the average annual cost on the system can be expected to reach the minimum when the storm sewer network is designed to resist a design storm of this specific return period. In comparison with the traditional storm sewer network design method where the return period of a design storm is required to be given beforehand according to experience or experts' opinion, this approach considers flood risk explicitly and is based on a good balance between the construction cost and the flood risk. The traditional design method may cause inappropriate capital spending because it subjectively specifies a protection level of the system.

Fig 5.4 shows the probabilistic flood damage curve for the obtained optimal design. The curve is approximated by linear interpolation of the flood damage under different design storms. The curve is extended horizontally when the flood return period exceeds 200 years (top right corner at the diagram in Fig 5.4). Indications are given here regarding the range of the design storms chosen for the risk-based optimal storm sewer network design. Concerning the lower bound of the return period of the design storms, the design storm of a return period smaller than the expected return period of the network can be excluded, e.g. if the designer believes that the network is designed to resist a 10-year storm at least, the design storms utilised in this methodology can start from return period of 10 years.

Concerning the upper bound, the suitable upper limit of the return period being used is difficult to identify beforehand. After the selection of the network design it is necessary to complete the probabilistic flood damage curve past the selected return period. If the additional curve under the area is significant comparable to the overall flood risk cost, design storms of higher return period should also be included. However, for simplicity reasons and limited available information about the design storms this part is not executed in this thesis.

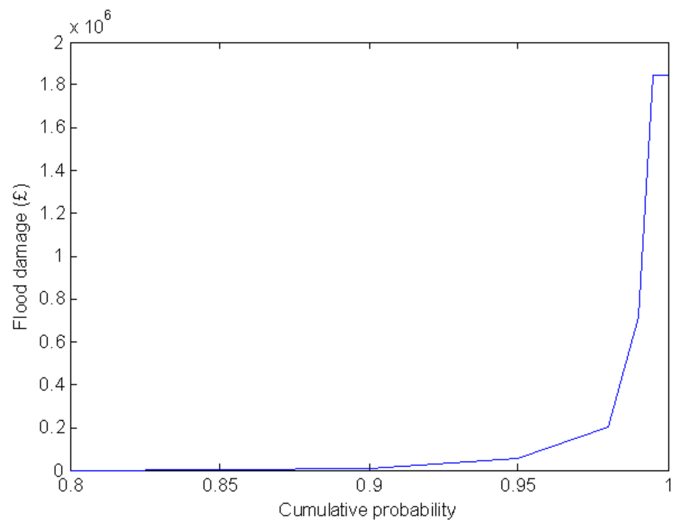



Fig 5.4 The probabilistic flood damage for the synthetic network design from single objective optimization with flood risk evaluated under several design storms

The construction cost and flood risk cost for the obtained design network are respectively £472,791 and £26,317 for an annual year. The designs obtained by the single-objective approach and by the multiple-objective approach with the same design criterion are listed in Table 5.3 for comparison. Though the decision criteria used are the same, the two designs from different approaches are not identical as GA and NSGA II are stochastic optimization algorithms. Furthermore, the different ways of formulating the problem in single- and multiple-objective optimization change the landscape of the objective function thus it may have an influence on the optimization result. In this case, the design obtained by single-objective approach is better than that by multiple-objective approach in the sense of minimizing the expected total cost. The difference between the total annual costs obtained from the two approaches is around 3%.

Table 5.3 Comparison between the synthetic network designs obtained by single-objective and multiple-objective approaches

| | Design by single objective | Design by multiple objective |
|---|---|---|
| Construction cost (£) | 472,791 | 483,487 |
| Expected flood risk (£) | 26,317 | 31,191 |
| Total cost (£) | 499,108 | 514,678 |
| Return period (year) | 5-10 | 2-5 |

The number of the flood risk evaluations for each generation in the single-objective optimization is shown in Fig 5.5. No simulation is called for the first 20 generations, which shows that all the initial random candidate solutions violate constraints in Eq(4.2) and Eq(4.3) badly. This result also appears in Fig 5.3 that shows large fitness values for the first twenty generations, corresponding to the constraint violation penalties. As the reproduction goes on, more and more candidate solutions are acceptable with respect to the constraints in Eq(4.2) and Eq(4.3) and the objective function is evaluated. After 40 generations, the number of flood risk evaluations for each generation varies from 72 to 132. The overall number of the flooding risk evaluations for the whole searching process is 99,715. The computation effort is less than a half compared with $2\times10^5$ evaluations being required if no threshold is set to evaluate the flood risk.

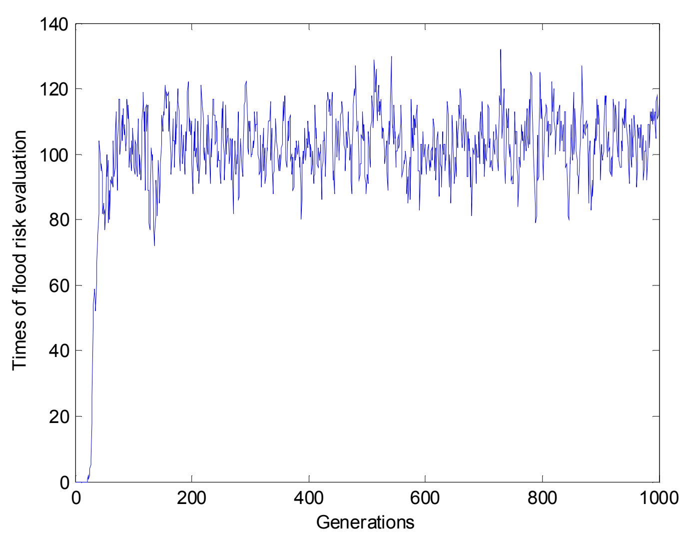



Fig 5.5 Number of objective evaluations for the synthetic network design from single objective optimization with flood risk evaluated under several design storms

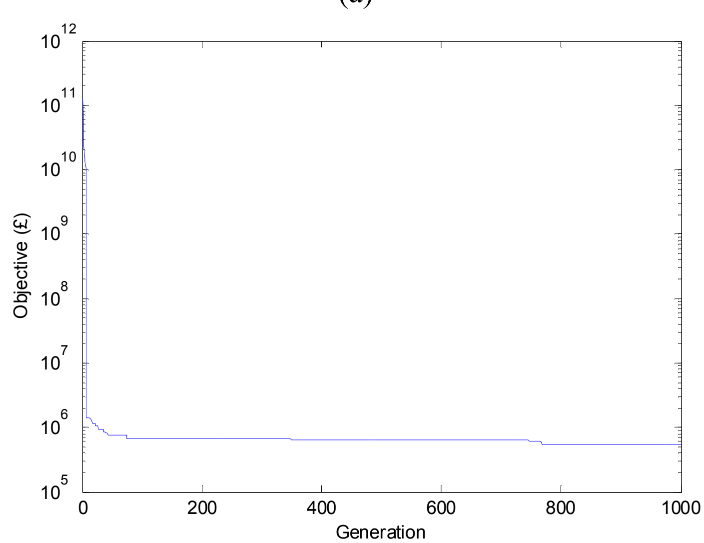
### 5.4.2 The Miljakovac network

The parameters and schemes used in GA in this case are set as the same in the synthetic network as shown in Table 5.1.

The best result of 10 runs is shown in Fig 5.6. The objective value of the optimal design improves as the search goes on. The optimal overall cost on the storm sewer system including the construction cost and the flood risk cost for an annual year is £541,423. Fig 5.7 shows the probabilistic flood damage curve for the obtained optimal design. The network is designed to resist a design storm of return period between 5 years to 10 years. The construction cost and the flood risk cost for the obtained optimal network design are £478,563 and £62,860, respectively. The designs obtained by the single-objective and multiple-objective optimization are compared in Table 5.4. In this case, the design obtained by the multiple-objective approach is better than that by the single-objective approach in the sense of minimizing the total cost. The difference between the total costs obtained from the two approaches is around 5%.

Fig 5.8 presents the number of flood risk evaluations executed for each generation in the optimization process. The overall number of flood risk evaluations is 93,663. Around half of the computational effort is saved compared with an algorithm with no threshold conditioning on the flood risk evaluation.

(a)



(b)

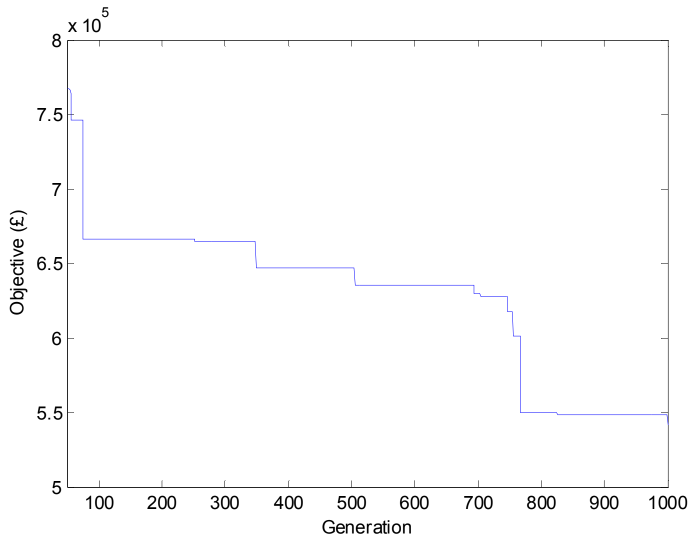



Fig 5.6 Optimization process for the Miljakovac network design with flood risk evaluated under several design storms

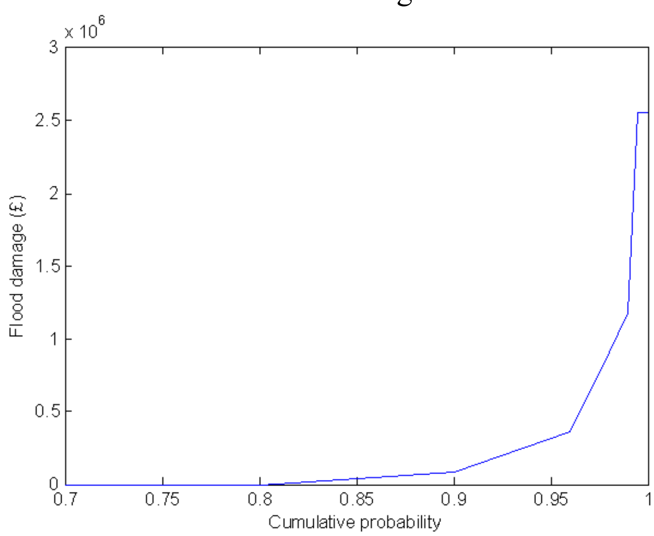



Fig 5.7 The probabilistic flood damage for the Miljakovac network design with flood risk evaluated under several design storms

Table 5.4 Comparison between the Miljakovac network designs from single-objective and multi-objective optimization approaches

| | Design by single objective | Design by multiple objective |
|---|---:|---:|
| Construction cost (£) | 478,563 | 422,228 |
| Expected flood risk (£) | 62,860 | 20,169 |
| Total cost (£) | 541,423 | 514,678 |
| Return period (year) | 5-10 | 2-5 |

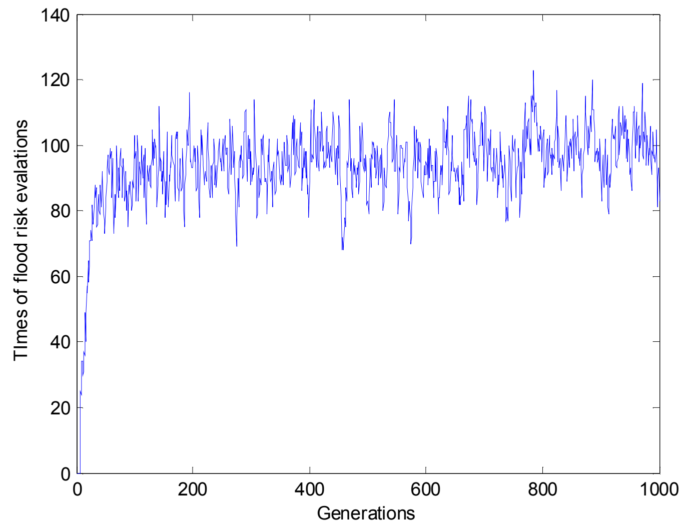



Fig 5.8 Number of objective evaluations for the Miljakovac network design

## 5.5 Conclusions

This chapter presents a framework for the design of a storm sewer network considering flood risk under a single-objective optimization. The expected value based criterion is employed as the decision criterion to formulate the optimization objective. However, the choice of the decision criterion will not affect the generality of framework. This approach allows the identification of the appropriate design level by balancing the construction cost against the flood risk reduction.

In the framework, the flood risk is computed within the optimization loop. Without affecting the general framework, the probabilistic flood risk can either be evaluated under design storms or using sampling based method or rainfall time series, depending on the available information about the main driver of the system (rainfall), the computation resources and the requirement of the problem. When evaluating the probabilistic flood risk, this chapter directly assigns the return periods of design storms to the return periods of floods. The probabilistic flood risk can be obtained through a limited number of storm sewer system performance simulations under the design storms. This approach is therefore significantly more computationally efficient than using sampling based methods or methods involving continuous time series for the flood risk evaluation of a storm sewer system.

The storm sewer network design is a constrained optimization problem incorporating both continuous and discrete decision variables. The original GA is adapted to solve this particular problem. The trial solutions for the flood risk assessment are selected or discarded based on constraint evaluations. Such threshold selection is proved to significantly improve the computational efficiency.

The framework for a risk-based optimal storm sewer network design is successfully applied to the design of two different networks: a synthetic network and the Miljakovac network. The applications demonstrate the effectiveness of the methodology.

The proposed framework can easily be extended to the design of storm sewer systems with storage elements, pumping stations, etc. by incorporating these elements into the decision variables.

# Chapter 6 Optimal storm sewer network design: flood risk evaluated with sampling method

This chapter solves the flood risk based storm sewer network design with a single-objective optimization under the framework proposed in Chapter 5. The decision criterion determining the single optimization objective also employs the expected value based criterion, where the probabilistic flood risk is represented by its expectation. Different from chapter 5 evaluating flood risk under several design storms, this chapter assesses the flood risk with sampling method. Sampling-based method is widely applied in many areas due to its simple principle and easy application, though it is usually computationally costly. A novel method for generating correlated samples is introduced. It is adapted from a literature method providing that the marginal distributions of variables as well as the correlations between them are known. The group method is proposed in order to facilitate the generation of correlated samples of large sizes. The method is generally applicable to any area where correlated samples are required. In this chapter, it is applied to the generation of rainfall event samples; the samples are then used in the evaluation of flood risk for storm sewer network design.

## 6.1 Generating correlated samples

### 6.1.1 Related issues

#### 6.1.1.1 Background

When modelling hydrosystems, it is common that the variables or inputs are not independent of each other. For example, the intensity and the duration of a rainfall event are usually observed to be negatively correlated; the peak, the volume and the duration of runoff are probably dependent from each other; a regional cross correlation among the precipitation of different regions is possibly presented when regional effects are considered, etc.

If these variables are generated via sampling, it is important to ensure that the dependency of the samples is either incorporated in the sampling process, or maintained in the resultant samples through adjustments. Any dependencies among the variables must be considered when solving such problems as substantial biases can result if correlations are neglected (Smith et al. 1992). Kapelan et al. (2005) asserted that neglecting demand correlation in water distribution system design under uncertainty may lead to underdesign of systems. Douglas et al. (2000) believed that a dramatically different interpretation would have been achieved if regional cross-correlation had been ignored when analyzing the trends in flood and low flows in the US. Grimaldi and Serinaldi (2006) stated that for a complete analysis of the three main characteristics of a flood event, i.e. peak, volume and duration, full understanding of these variables and relationships is necessary. Kanso et al. (2006) found out a clear correlation between the parameters of maximum mass and the erosion parameter in the urban runoff quality modelling. Yue (2000) pointed out that the severity of the damage caused by a storm is in fact a function of the correlated storm peak and total amount.

Sampling based technique is one of the most frequently used approaches to generate required variables for model simulations. If the variables are correlated, a traditional way of considering the dependence between variables is to use classical families of multivariate probability distributions such as the normal, log-normal, and exponential distributions. However, such an approach suffers from the limitation that the behaviours of the multiple variables must be characterized by the same parametric family of the univariate distributions. Another possibility to consider dependence between variables is via copula, which is a joint distribution function that can capture relationships between variables. Copula has been utilised to construct the correlation structure among rainfall variables (Salvadori and Michele 2006; Salvadori and Michele 2007; Zhang et al. 2007). When working with copulas, the choice of a good fitting dependence structure is important. However, the experience of choosing a suitable copula to describe certain data or phenomena is limited, though this is still an area of active research.

Pearson correlation coefficient (CC) and the Spearman rank correlation coefficient (RCC) are another two possibilities widely used to measure the dependence between variables (Helton and Davis 2003). The CC $r_{XY}$ between two random variables $X$ and $Y$ with expected values $\mu_X$ and $\mu_Y$ and standard deviations $\sigma_X$ and $\sigma_Y$ is computed as:

$$r_{XY} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y} \tag{6.1}$$

where cov is the covariance and $E$ is the expected value operator. The CC $r_{xy}$ of samples $x_i$ and $y_i$, $i$=1,2,…$n$, which are series of measurements or samples from variables $X$ and $Y$, can be calculated by :

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}} \tag{6.2}$$

where $\bar{x}$ and $\bar{y}$ are the sample means of $X$ and $Y$, $s_x$ and $s_y$ are the sample standard deviations. CC is a value between -1 and 1, providing a measure of the strength of the linear relationship between two variables, with CC=1 denoting the case of an increasing linear relationship, and CC=−1 being the case of a decreasing linear relationship. The values in between in all other cases indicate the degree of linear dependence between variables. The closer the coefficient is to either −1 or 1, the stronger the linear correlation between the variables. The RCC is defined similarly to the CC as Eq(6.1) and Eq(6.2) but with rank-transformed data. Rank-transforming is a step to convert the original values of the samples to their orders in samples. More specifically, the smallest value of samples of a variable is given a rank of 1; the next smallest value is given a rank of 2; and so on up to the largest value which is given a rank equal to the sample size $n$. The numerical values of RCC are also between -1 and 1 but it is a measurement of the strength of the monotonic relationship between two variables.

Henceforth, CC is a measure of linearity of the relationship between variables; while RCC is a measure of the monotonicity in the relationship between variables (Conover and Iman 1981). They are both useful in describing the dependency of variables. However, in sampling-based simulations, RCC is predominantly utilised due to the general difficulty in maintaining a specified CC value when the required variables to be generated are not normally distributed (Morgan and Henrion 1990).

### 6.1.1.2 Existing methods for generating correlated samples

Methods for generating correlated samples with some specific marginal distributions such as normal distributions (Cheng 1985) and Pearson family distributions (Parrish 1990) are available. Iman and Conover (1982) proposed a distribution-free and simple widely used method for generating correlated samples. However, their method can only generate samples with given RCC, but not CC. Moreover, it is difficult to guarantee the accuracy of the resultant RCC of generated samples. Li and Hammond (1975) and Lurie and Goldberg (1998) presented some 2-step methods for generating correlated samples by (1) generating an intermediate normal sample of multivariables; and (2) transforming underlying correlated normal sample into the target non normal sample. The intermediate normal sample should have appropriate correlations which are determined by an inversion of a double integral. This is a computationally intensive procedure and a feasible solution may not be available. The correlated samples can also be generated through copulas. As stated by Genest and Rivest (1993), a natural way of specifying the distribution function is to examine the copula and marginal distributions separately. Schweizer and Wolff (1981) established that the copula accounts for all the dependence between two random variables $X$ and $Y$: if $g_1$ and $g_2$ are strictly increasing functions over the range of $X$ and $Y$, the transformed variables $g_1(X)$ and $g_2(Y)$ have the same copula as $X$ and $Y$. Regardless of the scale in which each variable is measured, the copula is able to capture the synchronized fluctuations between $X$ and $Y$. Therefore it is possible to express RCC solely in terms of the copula function. However, as CC is affected by changes of (nonlinear) scale, specifying the copula alone is not sufficient and it requires the marginal distributions to be known (Frees and Valdez 1997). Hence, the method using copula to generate samples is generally constrained to those required correlation given by RCC instead of CC.

Charmpis and Panteli (2004) and Vořechovský and Novák (2009) proposed a heuristic approach for generating correlated samples. They considered the case of sampling from a multivariate distribution with correlated marginals where the specified marginal probability distributions as well as their correlation coefficients are known. The approach includes two distinct steps: the first step is to generate univariate random samples independently from their own specified marginal probability distributions; in the second step the generated univariate samples are rearranged in a way that the values of the numbers generated in the first step do not change but the positions of these numbers change, thus the desired correlations between them can be obtained. The simulated annealing (SA) algorithm is employed to rearrange the positions of the univariate samples in order to find a suitable arrangement. Chakraborty (2006) provided some theoretical results about "how close to the target correlation" by rearranging positions of univariate samples and proposed a deterministic initialization algorithm called PERMCORR based on the theoretical results. This deterministic algorithm is claimed to speed up convergence of stochastic optimization algorithms such as SA.

### 6.1.2 Methodology

#### 6.1.2.1 Basic principle

This chapter focuses on generating correlated samples with known marginal distributions and dependences given by CC or RCC. The basic idea of the methodology is to approximate the sample correlations (either CC or RCC) to the target correlations by rearranging the positions of the samples after each marginal sample has been independently generated according to its own marginal distribution (Charmpis and Panteli 2004; Chakraborty 2006; Vořechovský and Novák 2009). For ease of referencing, if an $m$-variable sample $\boldsymbol{x}$ of size $n$ is required, the elements of $\boldsymbol{x}$ can be denoted by $x_{ij}$:

$$x = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{6.3}$$

A column of the matrix $\boldsymbol{x}_i$ is generated from the known marginal distribution and represents a marginal sample. The correlations between these $m$ marginal samples that form a matrix $\boldsymbol{c}$ is required to approximate the target matrix $\boldsymbol{c}^*$: $\mathbf{c} \cong \mathbf{c}^*$. The elements of the two correlation matrices $\boldsymbol{c}$ and $\boldsymbol{c}^*$ are respectively denoted as $c_{ij}$ and $c_{ij}^*$ ($i, j$=1,2,…$m$ ). As $c$ and $c^*$ are symmetric matrices and their diagonal elements are equal to unity, only the elements below the leading diagonal are involved in the rearrangement procedure. The objective of the procedure is to minimize the difference between $c_{ij}$ and $c^*_{ij}$ ($i>j$). The positions of $\boldsymbol{x}_1$ can be kept unchanged because of the symmetry of the problem. The permutations of $\boldsymbol{x}_2, \boldsymbol{x}_3,\ldots, \boldsymbol{x}_m$ are then determined one at a time, e.g.:

- $\boldsymbol{x}_2$ is permuted letting $c_{12}= c_{12}^*$;
- $\boldsymbol{x}_3$ is permuted letting $c_{13}= c_{13}^*$, $c_{23}= c_{23}^*$;
- …
- $\boldsymbol{x}_{\mathrm{m}}$ is permuted letting $c_{1m}= c_{1m}^*$, $c_{2m}= c_{2m}^*$, $c_{m\text{-}1m}= c_{m\text{-}1m}^*$;

The root mean square error (RMSE) is used to estimate the distance between the resultant correlation vector and the required correlation indices. The size of the solution space for this problem (the number of the possible permutations of $\boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$) is $(n!)^{m-1}$. Charmpis and Panteli (2004) and Vořechovský and Novák (2009) used a stochastic optimization algorithm known as the Simulated Annealing method (SA) to search a good permutation for each marginal sample that makes $\mathbf{c}_i \cong \mathbf{c}_i^*$. Random pairwise positions in a marginal sample were interchanged according to the SA scheme. Since the size of the solution space of the problem increases dramatically as the size of samples increases, Vořechovský and Novák (2009) emphasized the method is designed for small sample generations. Chakraborty (2006) believed that stochastic algorithms such as SA are not efficient near local optima and proposed a deterministic algorithm. The difference between the correlations of two neighbor permutations of two samples $\boldsymbol{x}$ and $\boldsymbol{y}$ ($\boldsymbol{x}$ and $\boldsymbol{y}$ are both a marginal sample) by swapping the positions of $i$ and $j$ in sample set $\boldsymbol{y}$ is

$$corr(x,y) - corr(x,\pi(y)) = \frac{(x_i - x_j)}{\sigma_x}\frac{(y_i - y_j)}{\sigma_y}/n \tag{6.4}$$

where $\pi(y)$ is a permutation of sample $\boldsymbol{y}$. In the deterministic algorithm, if a random pairwise transposition enables the derived correlation $\boldsymbol{c}$ to move towards the target correlation $\boldsymbol{c}^*$, the transposition is accepted. Such deterministic transpositions are continually made until $\boldsymbol{c}$ cannot be improved any more or until it achieves the required precision. This deterministic algorithm is used as the basis of the proposed methodology.

### 6.1.2.2 The group method and its theoretical basis

In the application of sampling-based method, the required size of the generated samples depends on the extreme value of certain quartile of interest and the degree of the result accuracy required. For instance, Rahman et al. (2002) generated 10,000 samples of rainfall events in order to produce relatively stable estimates of the derived flood frequency curve in the average recurrence interval range from 1 to 100 years. They stated that the required number of the generated sample events would increase by orders of magnitude if the purpose of the MCS was to estimate flood events in the extreme range or if the random variables are more independent. Hence there is a potential need for a method for generating correlated samples of large sizes.

The proposed group method is to facilitate an efficient generation of correlated samples of large sizes. It can be outlined as follows: (1) generate samples of variables from their marginal distributions; (2) rearrange the samples in increasing (or decreasing) order; (3) bunch every $k$ samples to form groups from the beginning of each univariable sample (the reason for doing this is given afterwards), where $k$ is the number of samples in a group and is divisible by $n$. Thus $(n/k)$ groups are formed; (4) adjust the positions of the groups while keeping the interior positions of the samples within a group.

With this approach the dimension of the solution space of the problem is reduced. When $k$ equals 1, the algorithm is equivalent to the one described by Chakraborty (2006) as the positions of all the elements need to be determined.

If a bivariable problem (two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of length $n$ are known) is considered, as Chakraborty (2006) has demonstrated, $corr(x,\pi(y))$ is maximized as $c_{\max}$ when $x$ and $\pi(y)$ are concordant, while it is minimized to be $c_{\min}$ when they are discordant. When the group method is introduced, the resultant correlation range is unable to cover the whole range of $[c_{\min}, c_{\max}]$. The achievable correlation range depends on how the samples are grouped and they fall into two categories: (1) both $\boldsymbol{x}$ and $\boldsymbol{y}$ are arranged in an increasing order, and groups are then formed; (2) $\boldsymbol{x}$ is increasingly arranged while $\boldsymbol{y}$ is decreasingly arranged and they are then divided into groups. Let $\boldsymbol{x}'$ and $\boldsymbol{y}'$ present the formed groups, $[c'_{\min}, c'_{\max}]$ denote the range of correlations that can be achieved with the group method. In both cases the correlation is maximized when $\boldsymbol{x}'$ and $\pi(y')$ are concordant, while it is minimized when they are discordant. Hence, $c'_{\min} > c_{\min}$ in case (1) because the elements inside a group of $\boldsymbol{x}'$ and $\pi(y')$ keeps the monotonic increasing relationship though groups are discordant. For the same reason, $c'_{\max} < c_{\max}$ in case (2). Therefore when forming the groups, attention should be paid to make the target correlation $c^*$ in the achievable interval $[c'_{\min}, c'_{\max}]$.

If the samples are grouped with random orders, it is difficult to identify the concordant or discordant order of the groups, thus difficult to identify $[c'_{\min}, c'_{\max}]$. Moreover, as will be shown later, it is also difficult to obtain the achievable precision of the resultant correlation if the samples are reordered randomly when forming the groups.

After forming the groups, the group positions are rearranged using the deterministic algorithm. Pairwise groups are randomly chosen and their positions are exchanged only if the derived correlation is improved towards the target correlation. The upper bound of the precision that can be achieved by the group method is determined by the distance of any permutation of groups to its nearest neighbor (as proposed by Chakraborty (2006)). Let $\tau$ be the permutation obtained from $\pi$ by swapping the positions of the groups $i$ and $j$ in sample set $\boldsymbol{y}$, the difference between the two neighbor permutations is

$$corr(x,\pi(y)) - corr(x,\tau(y)) = \sum_{l=1}^{k} \frac{(x_i^{(l)} - x_j^{(l)})}{\sigma_x} \frac{(y_i^{(l)} - y_j^{(l)})}{\sigma_y} / n \tag{6.5}$$

where $l$ denotes the interior positions of samples in a group. If a permutation $\varepsilon(\boldsymbol{x})$ of $x$ is consecutively ordered: $x_{(1)} \le \ldots \le x_{(n)}$, let $\delta_x$ denote the largest difference between $x_{(i)}$ and $x_{(i+k)}$ scaled by $\sigma_x$, define $\delta_x = \frac{1}{\sigma_x} \max_{\mathrm{k}+1 \le i < n} (x_{(i)} - x_{(i-\mathrm{k})})$, and the $i$th value belongs to the $t$th group. Similarly define $\delta_y = \frac{1}{\sigma_y} \max_{k+1 \le i < n} (y_{(i)} - y_{(i-\mathrm{k})})$. Thus an upper bound of the achievable precision by the group method is:

$$
\begin{aligned}
\delta &= \max_{\pi} \min_{i,j} \left| \sum_{l=1}^{k} \frac{(x_i^{(1)} - x_j^{(1)})}{\sigma_x} \frac{(y_{\pi(i)}^{(1)} - y_{\pi(j)}^{(1)})}{\sigma_y} \right| / n \\
&\le \max_{\pi} \min_{t} \left| \sum_{l=1}^{k} \frac{\max_{k+1\le m\le n} (x_{(m)} - x_{(m-k)})}{\sigma_x} \frac{(y_{\pi(t)}^{(1)} - y_{\pi(t\text{-}1)}^{(1)})}{\sigma_y} \right| / n \\
&= \delta_x \max_{\pi} \min_{t} \left| \sum_{l=1}^{k} \frac{(y_{\pi(t)}^{(1)} - y_{\pi(\text{t-1})}^{(1)})}{\sigma_y} \right| / n \\
&\le \delta_x \left| \sum_{l=1}^{k} \frac{\max_{k+1\le i\le n} (y_{(i)} - y_{(i-k)})}{\sigma_y} \right| / n \\
&\le \delta_x \delta_y /(n/k)
\end{aligned}
\tag{6.6}
$$

The following rule shows how close $c$ can get to $c^*$ for a bivariable problem with the group method.

i. If $c^* \ge c_{\max}$ then $c = c_{\max}$.

ii. If $c^* \le c_{\min}$ then $c = c_{\min}$.

iii. If $c^* \in [c_{\min}, c_{\max}]$, then $\left|c - c^*\right| < \frac{1}{2}\delta$.

While $c^* \in [c_{\min}, c_{\max}]$, the precision of the achievable sample correlations is associated with the number of the groups ($n/k$) as shown in Eq(6.6). The precision decreases as the number of the samples in a group increases. However, the dimension of the solution space decreases from $n!$ to $(n/k)!$ when the group method is introduced. Thus there is a trade off between the achievable precision and the computational efficiency when using the group method. In practical engineering use, the precision of the obtained samples correlation coefficients is usually not required to be as precise as possible. For instance, one usually provides limited accuracy about the target correlation coefficient (say two- or three- decimal accuracy) with good confidence. In addition, it can be seen later in the application that the required precision is easy to be achieved even with a small value of the number of groups such as 50.

Fig 6.1 shows the procedure of the group method for generating correlated bivariable samples. It is not necessary to compute each correlation after a pairwise exchange of group positions using the correlation definition in Eq(6.2). The new correlation can be derived by adding the difference in Eq(6.5) to the current correlation. The algorithm for more than two variables is very similar, only the comparison is executed between more correlation coefficients. As mentioned before, the acceptance of the transpositions can be determined by RMSE.

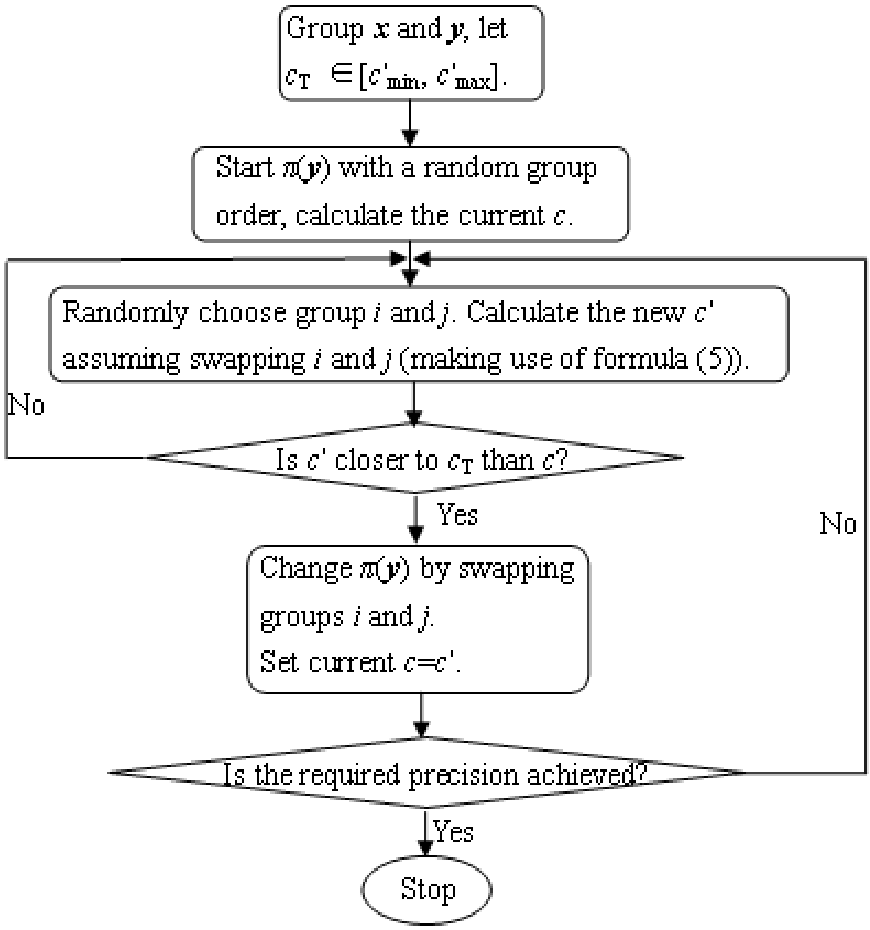



Fig 6.1 The procedure of the group method for generating bivariate samples

## 6.2 Applications of generating correlated samples

In this section, the proposed method is applied to two applications of sampling rainfall events of large sizes. Rainfall events are represented by two correlated variables, i.e., the rainfall depth and the duration or three correlated variables, i.e., the rainfall dry period, the wet period and the average intensity. This section also generates the correlated rainfall event samples for the synthetic storm sewer network design considering flood risk.

### 6.2.1 Rainfall events with correlated rainfall depth and duration

As the focus here is on the generation of the correlated samples, the marginal probability distributions of the variables representing rainfall as well as their correlation coefficients are assumed to be known a priori. Such distributions and the correlation coefficients between the variables can be obtained from statistical analysis of the real rainfall events in an actual design. The most commonly used distributions for describing rainfall variables are exponential distributions (Bacchi et al. 1994; Kurothe et al. 1997; Goel et al. 2000), Gumbel distributions (Koutsoyiannis and Baloursos 2000; Coles et al. 2003) and general Pareto distributions (Salvadori and Michele 2006). For simplicity, the distribution of rainfall depth and duration are both characterized by the Gumbel distribution Eq(3.20) in this case but such simplification has no impact on the results of the analysis.

In this example, the parameters are set as follows: $\alpha$=28mm and $\beta$=8 for the rainfall depth variable and $\alpha$=60min and $\beta$=12 for the rainfall duration variable. The CC between the rainfall depth and the duration is 0.78 (the CC is obtained by Thorndahl and Willems (2008) from the analysis of 18 years rainfall data).

A total number of 100,000 samples are drawn for the rainfall depth and the duration from their marginal distributions using general Monte Carlo sampling method. The required precision on the CC is 0.001, i.e. the algorithm stops when the difference between the obtained correlation and the target correlation is less than 0.001. Groups of 50, 100, 500, 1000, 5000, 10000, 50000 and 100000 are applied. Due to the stochastic nature contained in part of the algorithm, the proposed method was run 10 times for each number of groups and the average steps of the 10 runs are listed in Table 6.1. It is observed that the number of steps towards the resultant samples increases as the number of the groups increases. This agrees with the previous analysis on computational efficiency with the group method.

Table 6.1 Average computational steps for sampling rainfall events represented by two variables with group method

| Number of groups ($n/k$) | Average steps |
|---|---|
| 50 | 326 |
| 100 | 490 |
| 500 | 2326 |
| 1000 | 4,700 |
| 5,000 | 23,529 |
| 10,000 | 47,991 |
| 50,000 | 239,249 |
| 100,000 | 477,650 |

The typical scatter plots of the samples obtained by different number of groups are presented in Fig 6.2. All figures show an obvious linear correlated relationship between the two variables. The samples obtained from small group number tend to cluster. This is due to the fact that the samples in a group always keep their relative close positions in the searching algorithm when samples are grouped. This phenomenon also demonstrates that the marginal distributions and correlation coefficients only determine a rough trend of the samples, i.e. there is still some freedom to adjust how samples distribute under this trend. As additional subjective constraints (the way of forming groups) are introduced to the grouping procedure, samples from larger group number tend to distribute more dispersed than those from smaller group number. Therefore, the group number being large is recommended in practical use provided the computational efficiency is not an issue.

Fig 6.3 considers the case where the rainfall depth and duration are sampled from the same marginal distributions with a desired CC being 0.12. It is expected that the resultant samples in Fig 6.3 (with lower CC values) are distributed more uniformly or scattered than samples in Fig 6.2.

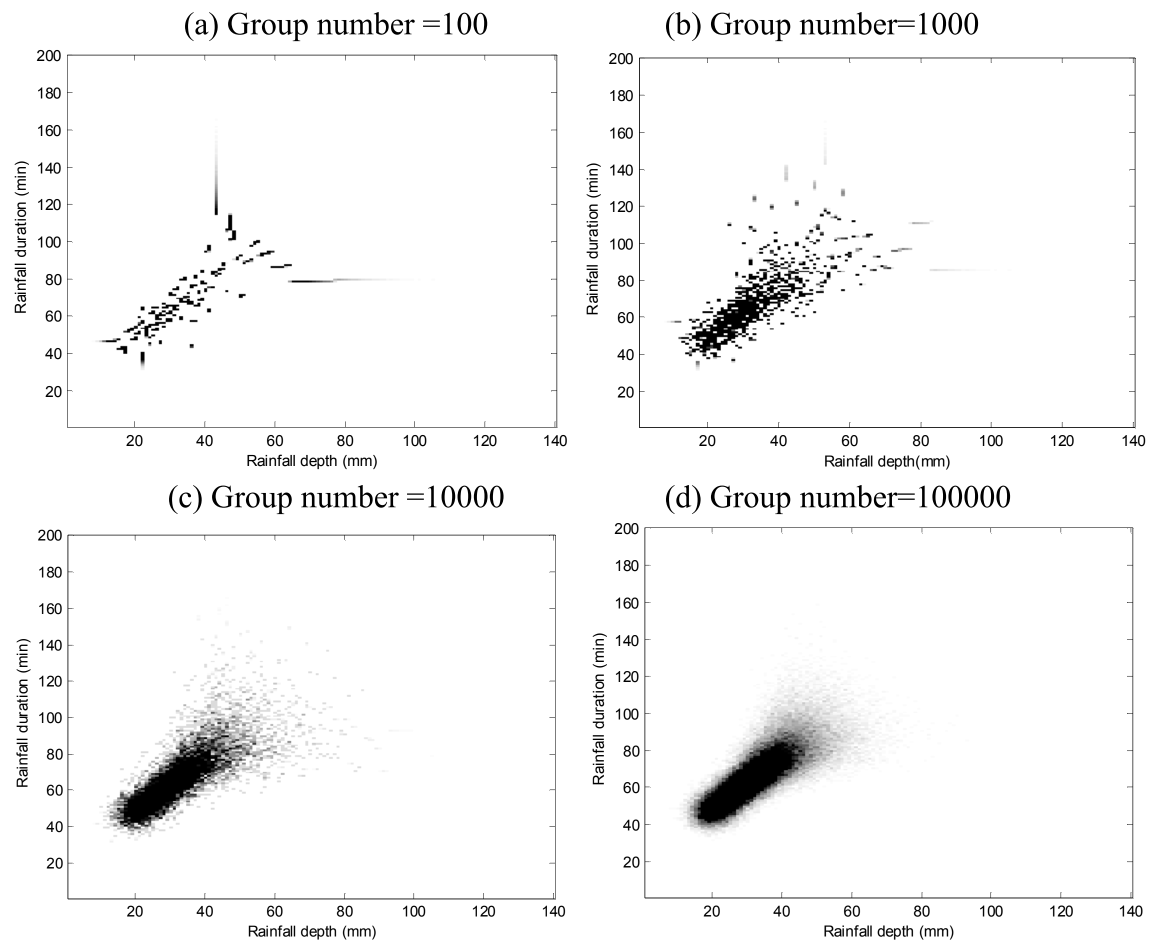


Fig 6.2 100,000 rainfall samples generated from different number of groups (CC=0.78)

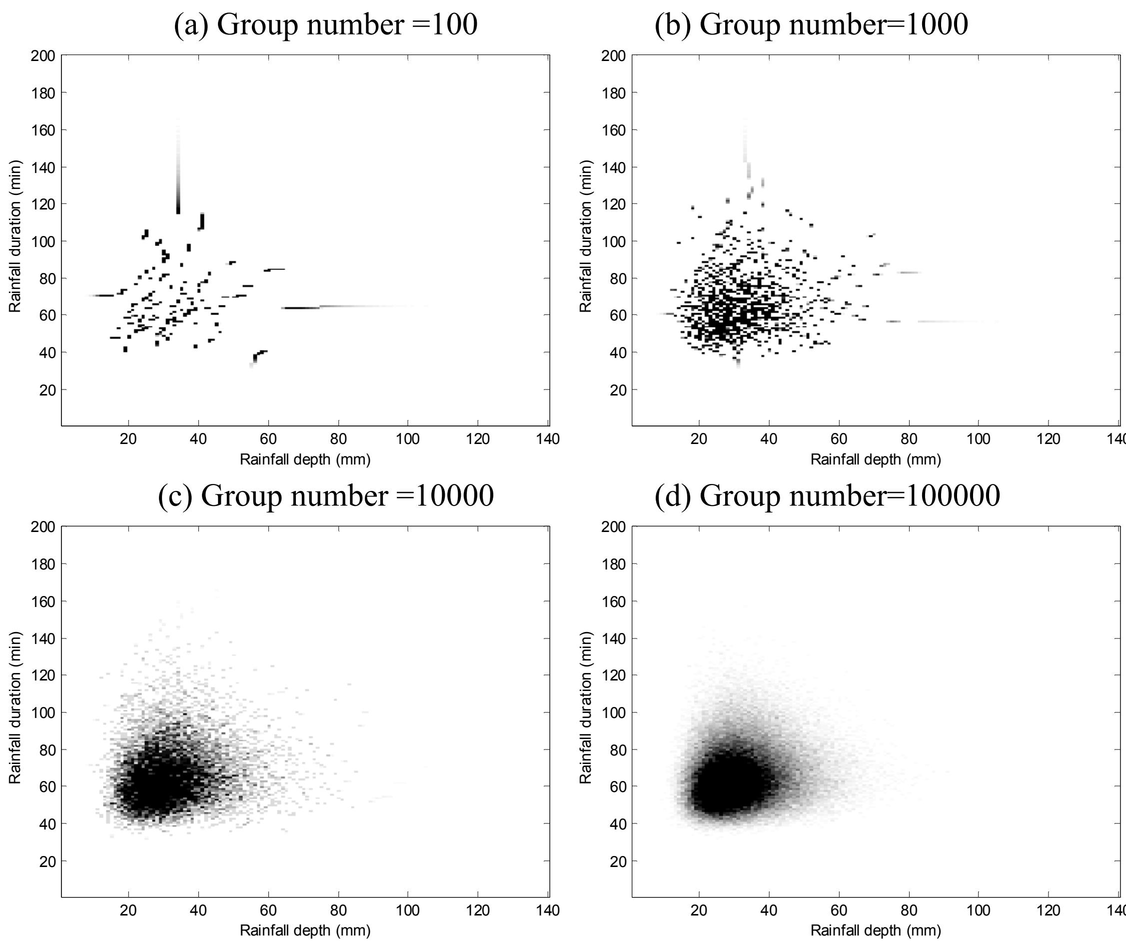



Fig 6.3 100,000 rainfall samples generated from different number of groups (CC=0.12)

### 6.2.2 Rainfall events with correlated dry duration, wet duration and intensity

In some cases, the dry period before the rainfall is also important in addition to the information about the rainfall itself. For instance, a flood may happen under a relatively small rainfall event shortly after another rainfall event but may not occur under a large rainfall event with a long time dry period earlier. Furthermore, the dry period is a crucial factor to determine the contaminated conditions of water after rains. Thus a rainfall characterizing a dry period (modeled as a duration *d* reporting no rainfall) followed by a wet period (modeled as a rectangular pulse having average intensity *i* and duration *w*) is frequently used as a coarse representation of rainfall. In this section, samples of rainfall events characterized by these three variables are generated.

The distributions of $d$, $i$ and $w$ are assumed to apply to the generalized Pareto (GP) distribution:

$$F(x) = \begin{cases} 1-\left(1-\dfrac{k}{c}(x-x^*)\right)^{1/k}, & x \ge x^* \\ 0, & x < x^* \end{cases} \tag{6.7}$$

The parameters for each variable of the CDF are listed in Table 6.2 (Salvadori and Michele 2006) from summer season rainfalls of Scoffera station, Italy). The CCs between the variables are assumed as Table 6.3 shows.

Table 6.2 The marginal GP parameters for rainfall variables

| Parameters | Dry duration (h) | Intensity (mm/h) | Wet duration (h) |
|---|---|---|---|
| $K$ | -0.46 | -0.49 | -0.05 |
| $C$ | 55.88 | 0.92 | 6.46 |
| $x^*$ | 7 | 0 | 0 |

Table 6.3 Correlated coefficients among rainfall variables

| Parameters | $d$ & $i$ | $i$ & $w$ | $d$ & $w$ |
|---|---|---|---|
| Correlation coefficient | 0.3 | -0.15 | -0.2 |

10,000 and 100,000 rainfall events are separately generated. Different group numbers are studied. The program is run 10 times for each number of groups. The required precision is set to be 0.001. In this case, the positions of samples from $i$ are first rearranged while positions of samples from $d$ are kept. When the correlation between $d$ and $i$ is achieved, the positions of samples from $w$ are then rearranged in order to achieve the required correlations between $w$ and $i$ and between $w$ and $d$. The average steps taken (including steps of arranging both marginal samples) to achieve the required precision for different groups are listed in Table 6.4. The average steps for the generation of 10,000 samples is generally less than the generation of 100,000 samples when they use the same number of groups. This is because a random step of a sample of smaller size generally walks more towards the required correlations than that of a sample of a larger size does. For generating samples of the same size, the average steps increase as the number of the groups increase.

Table 6.4 Average computational steps for sampling rainfall events represented by three variables with group method

| Number of Groups ($n/k$) | Average steps | |
|---|---|---|
| | 10,000 samples | 100,000 samples |
| 50 | 461 | 694 |
| 100 | 543 | 846 |
| 500 | 2,040 | 3,099 |
| 1000 | 3,549 | 5,998 |
| 5000 | 17,667 | 28,140 |
| 10,000 | 35,957 | 53,329 |
| 50,000 | - | 260,572 |
| 100,000 | - | 478,777 |

### 6.2.3 Samples for storm sewer network design

Owing to the intensive computation required for a large number of simulations of storm sewer system performance, the sampling method used for flood risk evaluation is only applied to the synthetic network design with a small number of rainfall event samples. As our focus is to demonstrate the methodology, the small number of samples will not affect the generality of the method. However, the number of the samples should be increased if a high precision result is required. The sampling based method for flood risk evaluation in storm sewer network design is introduced because it has many advantages over the method making use of design storms such as its simple principle and its flexibility for incorporating other uncertainty resources. In addition, though it is computation intensive, we can always expect that the computational capacity will increase potentially as time goes by.

The rainfall events are simply modelled with a rectangular shape and are fully characterized with rainfall depth and duration. The two variables are both described with the Gumbel distribution and the parameters are set the same as in section 6.2.1. The correlation coefficient is 0.78. The LHS is applied with a sample size of 200. The group number is 200. The obtained correlated samples are shown in Fig 6.4.

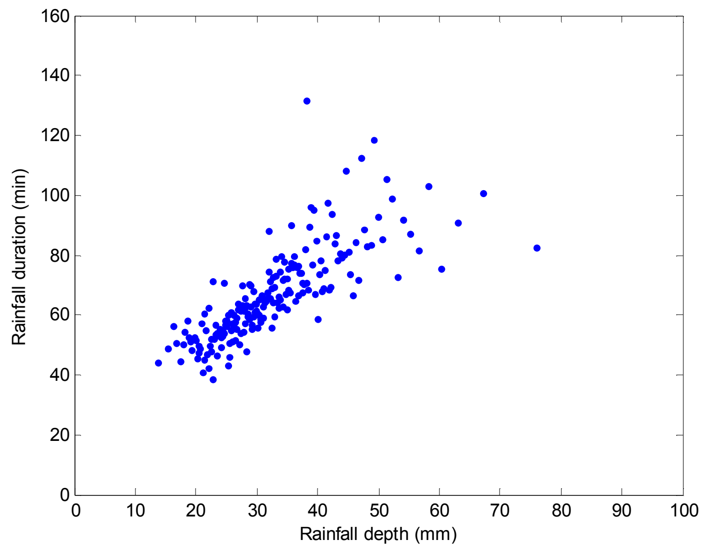



Fig 6.4 200 rainfall samples generated for synthetic network design

Fig 6.5 (a) presents rainfall event samples in the form of average intensity and duration. The rainfall events are sorted to form different “layers” as shown in Fig 6.5 (b) according to the non-dominated sorting technique, which is also used in NSGAⅡ. This approach facilitates the purpose for saving computation in flood risk evaluation. When evaluating the flood risk of one candidate storm sewer network, the rainfall events in the layer with the highest average intensities and longest durations are executed firstly and those in the next layer with relative lower intensities and shorter durations are then executed and so forth. If no flood occurs for a rainfall event in one layer, the storm sewer network simulation can stop at this layer by assuming that no flood will occur under rainfall events in the left layers. This assumption is reasonable as the consequence caused by a rainfall with higher intensity and longer duration should be worse than that caused by a relatively smaller rainfall. In this case, the 200 rainfall event samples are divided into 32 layers.

(a)

(b)

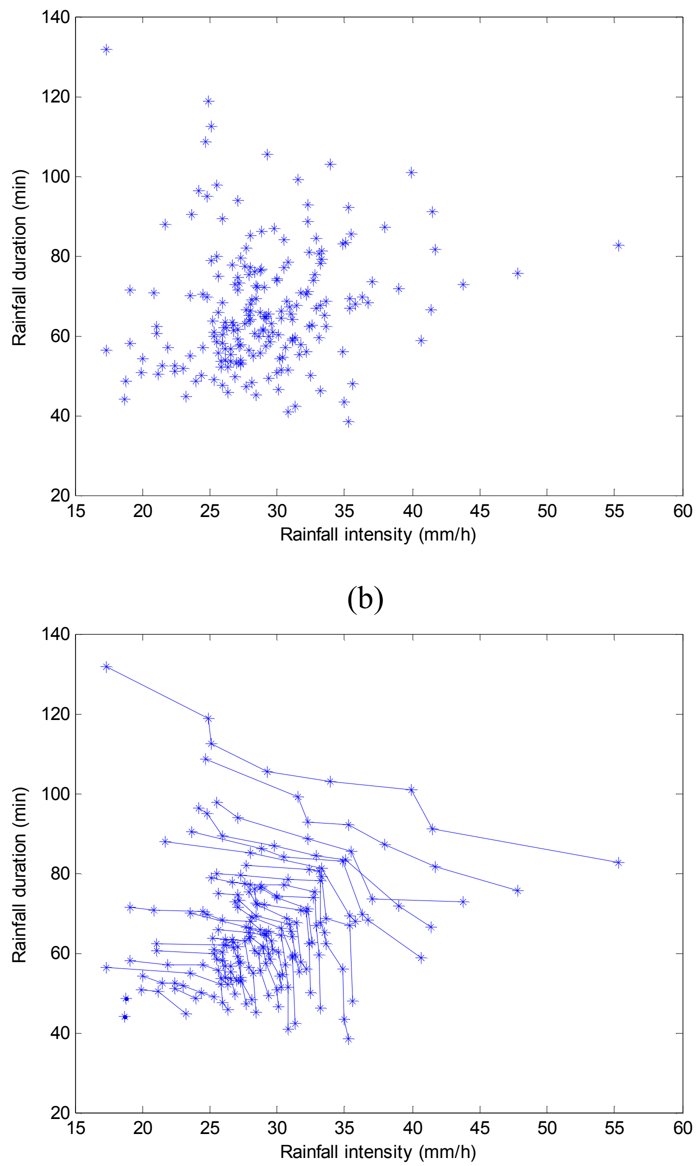



Fig 6.5 200 rainfall samples sorted by the non-dominated sorting technique

## 6.3 Optimal flood risk based storm sewer network design using sampling method: the synthetic network

The framework of the single-objective optimization for the storm sewer network design described in Section 5.2 is also applied in this chapter. The methodology is applied to the synthetic storm sewer network in this chapter. The network layout, design constraints, construction cost and damage curve are all set to be the same as in Chapter 3 and 4.

Generally the probabilistic flood risk can be identified with Weibull formula Eq(3.21). In this case the probabilistic flood risk is represented by the expected flood risk where the expected value based design criterion is adopted. The expectation of flood risk can be approximated as the mean consequence of all the flood events:

$$\overline{C}_{\mathrm{f}} = \sum_{i} C_{\mathrm{f}i}/n \tag{6.8}$$

where $\overline{C}_{\mathrm{f}}$ is the expected flood risk, $n$ is the number of the simulated flood events.

The optimization process is performed with adapted GA as described in Chapter 5. The parameters set in GA are given in Table 6.5. The optimization stops at the generation of 500.

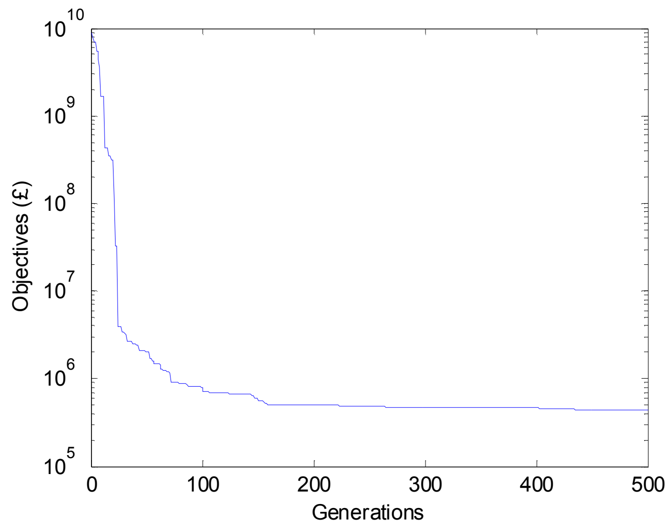
Table 6.5 Schemes and parameters set in GA for the synthetic network design with flood risk evaluated using sampling method

| Population size | Selection method | Crossover rate | Mutation rate | Elite size | Total generation |
|---|---|---|---|---|---|
| 200 | Roulette | 0.8 | 0.05 | 2 | 500 |

The GA program was run 10 times. The average computational time for each run of the program is around 359 hours on a PC with an Intel 2.4 GHz processor, 2 GB RAM and the MS Windows XP operating system. It can be seen that the computation of storm sewer network optimization involving sampling method is very intensive.

The best result of the ten running is presented. The objective values of all generations are plotted in a logarithm scale, as shown in Fig 6.6(a). Fig 6.6(b) is the objective values from the 50$^{th}$ generation with a linear scale. The objective value decreases as the search goes on. At the beginning the objective value is much higher than the subsequent values due to all chromosomes violating the constraints in Eq(4.2) and Eq(4.3) badly and being directly given a high penalty.

(a)




(b)

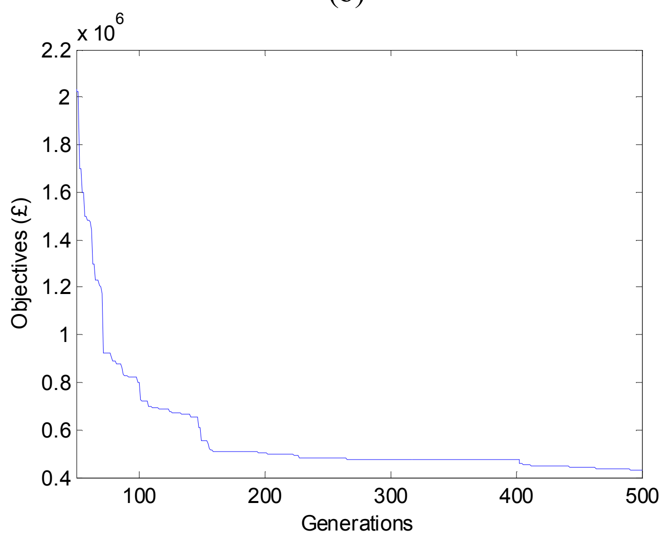



Fig 6.6 The optimization process for the synthetic network design with flood risk evaluated using sampling method

The total cost of the design network is £433,129. The construction cost and the expected flood risk are £430,191 and £2,938 respectively for an annual year. The probabilistic flood damage is show in Fig 6.7 according to Eq(3.21). There is no flood under the 188$^{th}$ event and some flood occurs for the 189$^{th}$ event in the 200 sample events. Hence the flood return period of the designed storm sewer network can be evaluated as around 1/(1-188/201)= 15 years. The obtained optimal design is listed in Table 6.6.

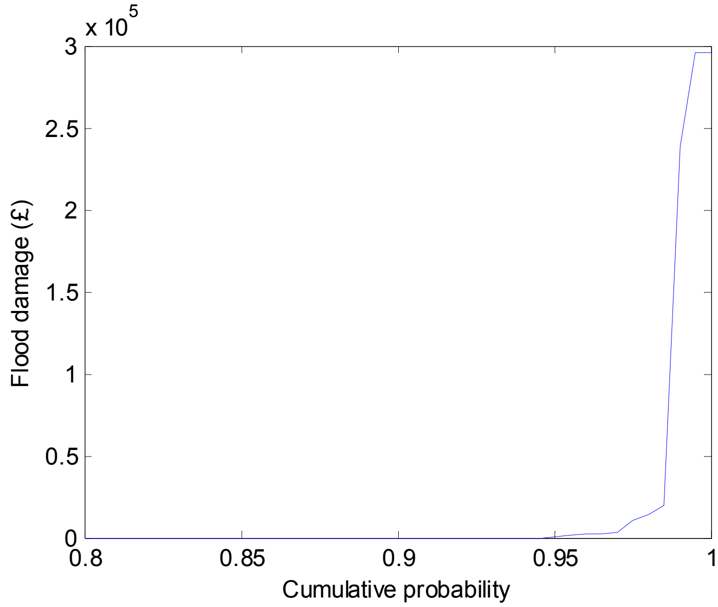



Fig 6.7 The probabilistic flood damage of the designed synthetic network

Table 6.6 Optimal design from single-objective optimization for the synthetic network with flood risk evaluated using sampling method

| Pipe | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Diameter/m | 0.375 | 0.375 | 0.6 | 0.375 | 0.75 | 0.45 | 0.75 | 1.05 | 0.6 | 0.45 |
| Slope /% | 0.16 | 0.18 | 0.15 | 0.15 | 0.15 | 0.23 | 0.15 | 0.15 | 0.67 | 0.23 |
| Pipe | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Diameter/m | 0.6 | 0.3 | 0.375 | 1.05 | 0.3 | 1.05 | 0.9 | 0.75 | 0.75 | 0.6 |
| Slope /% | 0.15 | 0.22 | 0.15 | 0.15 | 1.89 | 0.15 | 0.16 | 0.15 | 0.34 | 0.15 |
| Pipe | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | |
| Diameter/m | 0.45 | 0.375 | 0.75 | 0.375 | 0.3 | 1.05 | 0.45 | 1.05 | 0.225 | |
| Slope /% | 0.21 | 0.34 | 0.15 | 0.15 | 0.23 | 0.15 | 1.09 | 0.25 | 1.13 | |

The number of performed flood risk evaluations for each generation is shown in Fig 6.8. The overall number is 47,251. The computation is about half of the evaluations required if no threshold is set for the flood risk evaluation, which is $10^5$ (see Chapter 5). The overall number of simulations for the storm sewer performance under one rainfall event is shown in Fig 6.9 (the solid line). The performance under rainfall events of relatively low intensity and short duration are not simulated with the technique dividing samples into different layers. The computation is further deducted by this approach. The reference number of the simulations for storm sewer performance obtained by multiplying the number of flood risk evaluations with 200 (the sample size) is also showed for comparison in the same figure. The overall number of actual simulations is 3,987,723.

Compared to 9,450,200 simulations required without the layer dividing technique, around 58% of the computational time is saved.

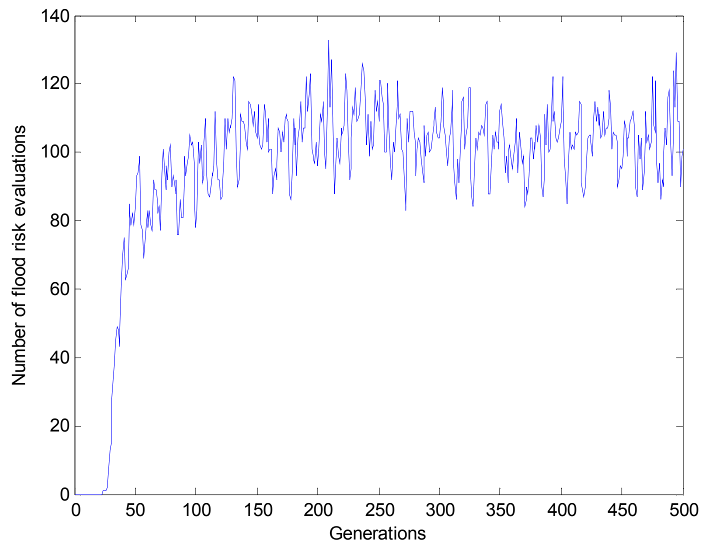



Fig 6.8 Number of objective evaluations for the synthetic network design with flood risk evaluated using sampling method

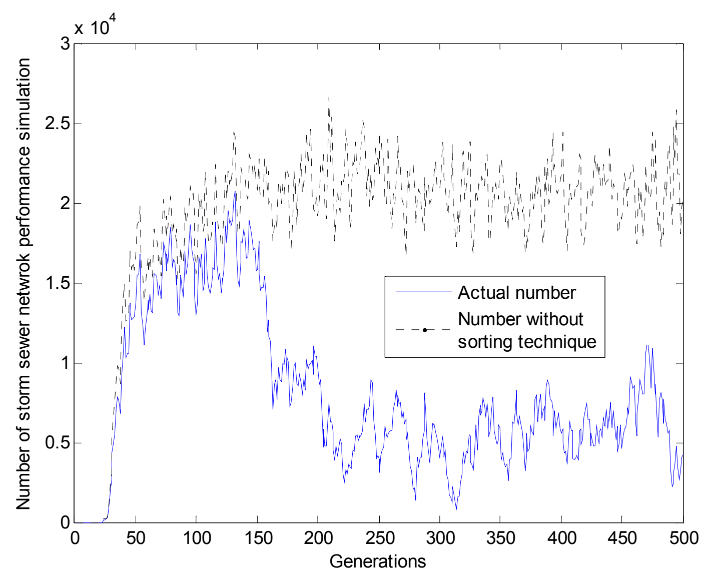



Fig 6.9 Number of simulations of storm sewer network performance for the synthetic network design with the sampling method

## 6.4 Conclusions

When modelling hydrosystems, it is common that the model is driven by several variables which may be correlated with each other. If these variables are generated using sampling method, it is important to ensure the dependency of the samples being incorporated in the resultant samples. A novel method of generating correlated samples with known marginal distributions and desired correlations (either CC or RCC) is introduced in this chapter. The group method is developed to facilitate the efficient generation of correlated samples of large sizes.

The basic idea of generating correlated samples is based on “adjusting the correlations by rearranging the relative positions of samples”. For the group method, the theoretic achievable precision is studied. In engineering practice, the precision requirement is usually not very high (say 2 or 3 decimal digit). A small number of groups such as 50 can generally satisfy the required precision. Due to the fact that the solution space can be dramatically decreased when the group method is introduced, the group method speeds up the searching process. The group method works more efficiently when the number of groups decreases.

The correlated samples are more likely to cluster when the number of groups is small, though it still reveals the correlated relationship between samples according to the target CC. This phenomenon also shows that the profile of how samples distributing is not solely determined by marginal distributions and correlation coefficients. Provided that the computational expense is not an imposing constraint, the large group number is recommended in practical use as the samples distribute more dispersed. The method successfully generates samples of rainfall events presented by variables with knowing marginal distributions and correlated coefficient.

The storm sewer network design is studied with flood risk estimated by sampling based method. The synthetic network design is applied. The optimal design is a balance between the construction cost and the flood risk. The probabilistic flood risk of a candidate network is evaluated using samples of rainfall events described by correlated variables. The threshold set for the flood risk evaluation (as described in chapter 5) and the layer dividing technique are applied in the optimization of the storm sewer network design. Both approaches effectively reduce the computational cost.

Lastly, the flood risk can be evaluated either under design storms or with sampling method. The choice of the method should depend on the available information, problem requirement and computational sources. The sampling method is extensively used in different areas due to its simple principle and easy application. However, its application in the storm sewer network design is generally computationally intensive.

# Chapter 7 Conclusions

## 7.1 Summary

This thesis primarily discusses decision-making in optimal storm sewer network design considering flood risk. Around this main topic, related issues are studied:

1. Uncertainty and uncertainty analysis are considered as an indispensable part in an integrated approach in scientific practice. This thesis discusses how to handle uncertainty in a general setting. The focus is placed on the understanding and representation of uncertainty and on the propagation of uncertainty through modelling. The methods are applied to a simple flood risk evaluation example.
2. The decision-making in the storm sewer network design considering flood risk is studied. The decision-making is not exclusive but depends on the decision-makers' attitude towards risk. Several commonly used design criteria under uncertainty are explored for the storm sewer network design. A methodology incorporating a multiple-objective optimization and a choice process is developed for the optimal storm sewer network design considering flood risk. The methodology is applied to two network designs.
3. A single-objective optimization for the storm sewer network design considering flood risk is explored provided the design criterion is known a priori. A framework for this task is developed. In this framework, the flood risk evaluation is located in the optimization loop. The flood risk is assessed under two different situations: with design storms and with samples of rainfall events.
4. A method for generating samples represented by correlated variables is developed. It is applied to the generation of rainfall events samples, which are used for flood risk evaluation in storm sewer network design.

## 7.2 Conclusions

The main conclusions regarding the discussion on uncertainty understanding and representation and uncertainty propagation are as follows:

1 The appropriate understanding of uncertainty is essential as it is critical to decision-making. It may stand for completely different information if uncertainty is wrongly understood or interpreted. The question about "whether a two-dimension or a one-dimension uncertainty representation is required" is illustrated.

2 The physical meanings of different uncertainty representations are given. With a two dimension uncertainty description, the inner dimension usually represents aleatory uncertainty while the outer dimension represents epistemic uncertainty. If both aleatory and epistemic uncertainties propagate simultaneously through a model within one dimension, the result is the expected distribution of the inherently varied model output over epistemic uncertainty.

3 The method for uncertainty propagation through a model is developed based on a sampling technique. The propagation processes are discussed under two situations: uncertainty is described by probability and by other mathematical languages. It is emphasized that the coherence of uncertainty nature should be ensured in the propagation.

4 A general process for uncertainty analysis through modelling is proposed.

5 The proposed methods with uncertainty represented by both probability and other mathematical languages for uncertainty propagation are applied to a simple flood evaluation case. It demonstrates that the methods represent and propagate uncertainties with explicit physical meanings.

Regarding the decision-making in the storm sewer network design considering flood risk, the conclusions are as follows:

1 The storm sewer network design can be seen as a decision-making process, which can be solved with an optimization approach. The uncertain property of the flood risk makes the decision-making not unique.

2 A methodology is developed for decision-making in storm sewer network design considering flood risk under different decision criteria. The methodology mainly includes a multi-objective optimization and a “choice” process. The NSGA Ⅱ is employed as the optimizer.
3 The methodology is applied to two storm sewer network designs. Different decision criteria provide different designs as different attitudes towards risk are adopted with different decision criteria.
4 The storm sewer network design can also be formed as a single objective optimization if the decision criterion is provided a priori.

Regarding the single objective optimization for storm sewer network design considering flood risk, the conclusions are as follows:
1 A framework for storm sewer network design using single objective optimization is presented. The frequently used decision criterion based on the expected value is adopted for flood risk evaluation. However, the choice of the criterion will not affect the generality of the methodology.
2 GA is employed as an optimizer for the single-objective optimization and its adapted form for the storm sewer network design makes the computation efficient.
3 The flood risk is evaluated via modelling. The probabilistic flood risk can be assessed either under design storms or with sampling methods. The choice of the evaluation method depends on the available information about the rainfall, the requirement of the result precision and the computation sources.
4 The use of design storms significantly reduces the required number of simulations of the storm sewer system performance in comparison with the sampling method. However, this computational efficiency is at a price of being an approximate evaluation. The sampling method is more flexible, for example, it is also applicable to cases when other uncertainty sources in addition to rainfall need to be considered. Moreover, we can always expect that the computation capacity increases potentially as time goes by.

The conclusions regarding the introduced method for generating correlated samples can be drawn as follows:

1 A novel method for generating correlated samples with known marginal distributions and desired correlations (either CC or RCC) is introduced in this chapter. The basic idea of generating correlated samples is about "adjusting the correlations by rearranging the relative positions of marginal samples".
2 The group method is developed to facilitate the efficient generation of correlated samples of large sizes. The correlated samples are more likely to cluster when the number of groups is small. This phenomenon also demonstrates that the profile of the distribution of samples is not solely determined by marginal distributions and correlation coefficients. Therefore, provided the computational efficiency is not an imposing constraint, a large group number is recommended in practical use because the samples are then distributed in a more dispersed manner (i.e. with less clustering).
3 The method for the generation of correlated samples successfully generated rainfall events samples and the samples are applied to the flood risk evaluation in the synthetic storm sewer network design.

## 7.3 Future research directions

Recommendations for future research are suggested in respect of two aspects: uncertainty representation and propagation and optimal storm sewer network design.

### 7.3.1 Uncertainty representation and uncertainty propagation

This thesis discusses the understanding and representation of uncertainty. It is strongly suggested to treat aleatory and epistemic uncertainties differently, not accounting for some real difficulties when facing real problems, such as the difficulty of distinguishing aleatory and epistemic uncertainties in some cases and the difficulty in understanding of the distinction of them for some decision makers. The solutions for solving these difficulties can be explored in the future work.

Uncertainty propagation in this thesis uses sampling based method. Other method such as reliability analysis method or analytic method can be studied in the future.

In addition, the sensitivity analysis for different sources of uncertainty of different nature on the result can be explored. It can provide decision makers with information about the importance of different sources. Moreover, it can possibly reduce some uncertainty analysis by ignoring uncertainties that do not significantly affect the output variation.

In the simple case application for uncertainty propagation in Chapter 3, the sources of uncertainty all come from parameter uncertainty. A method for propagating different sources of uncertainty should be developed in order to allow a systematic uncertainty analysis.

### 7.3.2 Optimal storm sewer network design

The storm sewer network design in this thesis is mainly concerned with the construction cost and the flood risk due to property damage. A more integrated consideration of the design can be incorporated in the future work, for instance, by including environment, ecology, energy, sustainability and operation.

When formulating the storm sewer network design in this work, the constraints only consider the pipe diameter progression, the minimum pipe cover and the maximum excavation depth. In practice more constraints such as the minimum and maximum flow velocity need to be taken into account in the design. A way to effectively incorporate more constraints in the optimization can be developed in the future work.

The applications in the thesis are executed assuming that the unit material cost for construction, property cost and the damage curve are given a priori. The sensitivity of these factors on the design can be examined in order to know how these factors can affect the resultant design.

The study used GA as the optimizer aiming to find the optimal design. However, GA is a stochastic optimization algorithm and there is no guarantee that the obtained design is the true optimal global optimum. The objective of the study does not focus on the optimization algorithm itself, but on the integrated framework. The optimization quality can be enhanced with more sensitivity analysis for GA parameters, more population or more generations, even with other heuristic optimization methodologies.

The storm sewer network design in this thesis takes flood risk (aleatory uncertainty) into account. It is suggested in this work not to include epistemic uncertainty in design as the problem becomes excessively complicated if two dimensions of uncertainties are considered in a design. However, in this way the problem formulation is not comprehensive as some information is missing. Ways of incorporating the other dimension of uncertainty can be explored in the future work. Otherwise, a better illustrated reasoning should be given for ignoring epistemic uncertainty.

# Reference

Abebe, A.J., and Price, R.K. (2005). "Decision support system for urban flood management." *Journal of Hydroinformatics*, 7, 3-15.

Afshar, M.H. (2006). "Application of a Genetic Algorithm to storm sewer network optimisation." *Scientia Iranica*, 13(3), 234-244.

Afshar, M.H., Afshar, A., Mariño, M.A., and Darbandi, A.A.S. (2005). "Hydrograph-based storm sewer design optimisation by genetic algorithm." *Canadian Journal of Civil Engineering Optimisation*, 33(3), 319-325.

Afshar, M.H. (2010). "A parameter free continuous ant colony optimization algorithm for the optimal design of storm sewer networks: constrained and unconstrained approach." *Advances in Engineering Software*, 41(2), 188-195.

Al-Futaisi, A., Stedinger, J.R., and ASCE, M. (1999). "Hydrologic and economic uncertainties and flood-risk project design." *Journal of Water Resources Planning and Management*, 125(6), 314-324.

Alvarez, D.A. (2006). "A Monte Carlo-based method for the estimation of lower and upper probabilities of events using infinite random sets of indexable type." *Fuzzy Sets and Systems*, 160(3), 384-401.

Apel, H., Thieken, A.H., Merz, B., and Bloschl, G. (2004). "Flood risk assessment and associated uncertainty." *Natural Hazards and Earth System Sciences*, 4(2), 295-308.

Aronica, G., Hankin, B., and Beven, K. (1998). "Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data." *Advances in Water Resources*, 22(4), 349-365.

Aven, T., and Pörn, K. (1998). "Expressing and interpreting the results of quantitative risk analyses: review and discussion." *Reliability Engineering & System Safety*, 61(1-2), 3-10.

Bacchi, B., Becciu, G., and Kottegoda, N.T. (1994). "Bivariate exponential model applied to intensities and durations of extreme rainfall." *Journal of Hydrology*, 155(1-2), 225–236.

Bae, H.-R., Grandhi, R.V., and Canfield., R.A. (2004). "Epistemic uncertainty quantification techniques including evidence theory for large-scale structures." *Computers and Structures*, 82(13-14), 1101-1112.

Balmforth, D., Digman, C., Kellagher, R., and Butler, D. (2006). *Designing for Exceedance in Urban Drainage-- Good Practice*, London: CIRIA.

Bao, Y., Tung, Y., and Hasfurther, V.R. (1987). "Evaluation of uncertainty in flood magnitude estimator on annual expected damage costs of hydraulic structures." *Water Resources Research*, 23(11), 2023-2029.

Beven, K. J. (2006). "A manifesto for the equifinality thesis " *Journal of Hydrology* 320(1-2), 18-36.

Beven, K.J., and Binley, A.M. (1992). "The future of distributed models: model calibration and uncertainty prediction." *Hydrological Processes*, 6(3), 279-298.

Beven, K.J., and Freer, J. (2001). "Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology." *Journal of Hydrology*, 249(1-4), 11-29.

Bocchiola, D., and Rosso, R. (2009). "Use of a derived distribution approach for flood prediction in poorly gauged basins: a case study in Italy." *Advances in Water Resources*, 32(8), 1284-1296.

Brown, J.D., and Damery, S.L. (2002). "Managing flood risk in the UK: towards an integration of social and technical perspectives." *Transactions of the Institute of British Geographers*, 27(4), 412-426.

Butler, D., and Davies, J.W. (2004). *Urban Drainage*, E&FN Spon, London.

Chadwick, A., and Morfett, J. (1993). *Hydraulics in Civil and Environmental Engineering*, 2nd ed., E&FN Spon, London.

Chakraborty, A. (2006). "Generating multivariate correlated samples." *Computational Statistics*, 21(1), 103-119.

Charalambous, C., and Elimam, A.A. (1990). "Heuristic design of sewer networks." *Journal of Environmental Engineering*, 116(6), 1181-1199.

Charmpis, D.C., and Panteli, P.L. (2004). "A heuristic approach for the generation of multivariate random samples with specified marginal distributions and correlation matrix." *Computational Statistics*, 19(2), 283-300.

Cheng, R.C.H. (1985). "Generation of multivariate normal samples with given sample mean and covariance matrix." *Journal of Statistical Computation and Simulation*, 21(1), 39-49.

Chow, V.T., Madment, D.R., and Mays, L.W. (1988). *Applied Hydrology. International edition. New York: MacGraw-Hill.*

Cohn, T.A., Lane, W.L., and Stedinger, W.G. (1997). "An algorithm for computing moments-based flood quantile estimates when historical flood information is available." *Water Resources Research*, 33(9), 2089-2096.

Coles, S., Pericchi, L.R., and Sisson, S. (2003). "A fully probabilistic approach to extreme rainfall modelling." *Journal of Hydrology*, 273(1-4), 35-50.

Conover, W.J., and Iman, R.L. (1981). "Rank transformations as a bridge between parametric and nonparametric statistics." *The American Statistician*, 35(3), 124-129.

Cullen, A.C., and Frey, H.C. (1999). *Probabilistic techniques in exposure assessment, A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*, Plenum Press, New York, London.

Deb, K., IEEE, A. m., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). "A fast and elitist multiobjective genetic algorithm: NSGA-II." *Transactions on Evolutionary Computation*, 6(2), 182-197.

Dawson, R. J., Speight, L., Hall, J.W., Djordjevic, S., Savic, D., and Leandro J., (2008). "Attribution of flood risk in urban areas." Journal of Hydroinformatics. 10(4), 275–288.

Dempster, A.P. (1967). "Upper and lower probabilities induced by a multivalued mapping." *Annals of Mathematical Statistics*, 38(2), 325-339.

Djordjević, S., Prodanovic, D., Maksimovic, C., Ivetic, M., and Savic, D. (2005). "SIPSON  Simulation of interaction between pipe flow and surface overland flow in networks." *Water Science and Technology*, 52(5), 275-283.

Douglas, E.M., Vogel, R.M., and Kroll, C.N. (2000). "Trends in floods and low flows in the United States: impact of spatial correlation." *Journal of Hydrology*, 240(1-2), 90-105.

Dubois, D. (2006). "Possibility theory and statistical reasoning." *Computational Statistics & Data Analysis*, 51(1), 47-69.

Dubois, D. (2010). "Commentary: representation, propagation, and decision issues in risk analysis under incomplete probabilistic information." *Risk Analysis*, 30(3), 361-368.

Dubois, D., and Prade, H. (1991). "Random sets and fuzzy interval analysis." *Fuzzy Sets and Systems*, 42(1), 87-101.

Dubois, D., and Prade, H. (1993) "Fuzzy sets and probability: misunderstandings, bridges and gaps." *Proceeding of the Second IEEE Conference on Fuzzy Systems*, 1059-1068.

Elimam, A.A., Charalambous, C., and Ghobrial, F.H. (1989). "Optimum design of large sewer networks." *Journal of Environmental Engineering*, 115(6), 1171-1190.

Environment Agency (2001). "Lessons learned: the autumn 2000 floods." Environment Agency, Bristol.

Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D.S., and Sentz, K. (2003). "Constructing Probability Boxes and Dempster-Shafer Structures." Sandia National Laboratories, California

Frank, M.V. (1999). "Treatment of uncertainties in space nuclear risk assessment with examples from cassini mission applications." *Reliability Engineering & System Safety*, 66(3), 203-201.

Frees, W.E., and Valdez, A.E. "Understanding relationships using copulas." *32nd Actuarial Research Conference*, University of Calgary, Canada.

Freeze, R.A., Massmann, J., Smith, L., Sperling, T., and James, B. (1991). "Hydrogeological decision analysis: 1 A framework." *Ground Water*, 28(5), 738-766.

Freni, G., Mannina, G., and Viviani, G. (2009). "Urban runoff modelling uncertainty: Comparison among Bayesian and pseudo-Bayesian methods." *Environmental Modelling & Software*, 24(9), 1100-1111.

FRMRC. (2006). "Implementation plan for library of tools for uncertainty evaluation." FRMRC Research Report UR2.

Fu, G., Butler, D., Khu, S.-T., and Sun, S. (2010). "Imprecise probabilistic evaluation of sewer flooding in urban drainage systems using random set theory (submitted)." *Water Resources Research*.

Genest, C., and Rivest, L.P. (1993). "Statistical inference procedures for bivariate archimedean Copulas." *Journal of American Statistical Association*, 88(423), 1034-1043.

Goel, N.K., Kurothe, R.S., Mathur, B.S., and Vogel, R.M. (2000). "A derived flood frequency distribution for correlated rainfall intensity and duration." *Journal of Hydrology*, 228(1-2), 56-67.

Goodman, A.S. (1984). *Principles of Water Resources Planning* Prentice-Hall, Englewood Cliffs, NJ.

Grimaldi, S., and Serinaldi, F. (2006). "Asymmetric copula in multivariate flood frequency analysis." *Advances in Water Resources*, 29(8), 155–1167.

Grum, M., and Aalderink, R.H. (1999). "Uncertainty in return period analysis of combined sewer overflow effects using embedded Monte Carlo simulations." *Water Science and Technology* 39(4), 233-240.

Guariso, G., Hitz, M., and Werthner, H. (1996). "An integrated simulation and optimization modelling environment for decision support." *Decision Support Systems*, 16(2), 103-117.

Guo, Y. (2007). "Efficient Optimal Design of Storm Sewer Networks Based on Cellular Automata and Genetic Algorithms," University of Exeter, Exeter.

Guo, Y., Walters, G.A., Khu, S.-T., and Keedwell, E.C. (2007). "A novel cellular automata based approach to storm sewer design." *Engineering Optimization*, 39(3), 345-364.

Gupta, A., Mehndiratta, S.L., and Khanna, P. (1983). "Gravity wastewater collection systems optimization." *Journal of Environmental Engineering*, 109(5), 1195-1209.

Guyonnet, D., Bourgine, B., Dubois, D., Fargier, H., Come, B., and Chiles, J.-P. (2003). "A hybrid approach for addressing uncertainty in risk assessment." *Journal of Environmental Engineering*, 129(1), 68-78.

Hackathorn, R.D., and Kenn, P.G.W. (1981). " Organizational Strategies for Personal Computing in Decision Support Systems." *MIS Quarterly*, 5, 21-27.

Hacking, I. (1975). "All kinds of possibility." *Philosophical Review*, 84, 321-347.

Haktanir, T., and Horlacher, H.B. (1993). "Evaluation of various distributions for flood frequency analysis." *Hydrological Sciences Journal*, 38(1), 15-32.

Hall, J.W. (2003). "Handling uncertainty in the hydroinformatic process." *Journal of Hydroinformatics*, 5, 215-232.

Hall, J.W. (2006). "Uncertainty-based sensitivity indices for imprecise probability distributions." *Reliability Engineering & System Safety*, 91(10-11), 1443-1451.

Hall, J.W., Blockley, D.I., and Davis, J.P. "Non-additive probabilities for representing uncertain knowledge." *Hydroinformatics' 98*, 1101-1108.

Harris, R. (2009). "Introduction to Decision Making." http://www.virtualsalt.com/crebook5.htm

Hattis, D., and Burmaster, D.E. (1994). "Assessment of variability and uncertainty distributions for practical risk analyses." *Risk Analysis*, 14(5), 713-730.

Heaney, J. P., Sample, D., and Wright, L. (2002). "Costs of Urban Stormwater Control." Water Supply and Water Resources Division National Risk Management Research Laboratory Edison, NJ 08837.

Helm, P. (1996). "Integrated Risk Management for Natural and Technological Disasters." *Tephra*, 15(1), 4-13.

Helton, J.C., and Burmaster, D.E. (1996). "Treatment of aleatory and epistemic uncertainty in performance assessments for complex systems." *Reliability Engineering & System Safety*, 54(2-3), 91-258.

Helton, J.C., and Davis, F.J. (2003). "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems." *Reliability Engineering & System Safety*, 81(1), 23–69.

Helton, J.C., Johnson, J.D., and Oberkampf., W.L. (2004). "An exploration of alternative approaches to the representation of uncertainty in model predictions." *Reliability Engineering & System Safety,* 85(1-3), 39-71.

Hofer, E. (1996). "When to separate uncertainties and when not to separate." *Reliability Engineering & System Safety*, 54(2-3), 113-118.

Hofer, E., Kloos, M., Krzykacz-Hausmann, B., Peschke, J., and Woltereck, M. (2002). "An approximate epistemic uncertainty analysis approach in the presence of epistemic and aleatory uncertainties." *Reliability Engineering & System Safety,* 77(3), 229-238.

Hoffman, F.O., and Hammonds, J.S. (1994). "Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability." *Risk Analysis*, 14(5), 707-712.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*, USA: The University of Michigan Press.

Hora, S.C. (1996). "Aleatory and epistemic uncertainty in probability elicitation with an example from hazardous waste management" *Reliability Engineering & System Safety*, 54(2-3), 217-223.

Iman, R., and Conover, W.J. (1982). "A distribution-free approach to inducing rank correlation among input variables." *Communications in Statistics - Simulation and Computation*, 11(3), 311-334.

Izquierdo, H., Montalvo, I., Perez, R., and Fuertes, V.S. (2008). "Design optimization of wastewater collection networks by PSO." *Computer and Mathematics with Applications*, 56(3), 777-784.

Joneja, G.S., Agarwal, S.K., and Khanna, P. (1978). "Optimisation methods provide money –saving design data." *Water and Sewage Works*, 125(12), 56-58.

Jonkman, S.N., Brinkhuis-Jak, M., and Mok, M. (2004). "Cost benefit analysis and flood damage mitigation in the Netherlands." *Heron*, 49(1), 95-111.

Jonkman, S.N., Gelder, P.H.A.J.M.v., and Vrijling, J.K. (2003). "An overview of quantitative risk measures for loss of life and economic damage" *Journal of Hazardous Materials*, 99(1), 1-30.

Reis Jr., D.S., and Stedinger, J.R. (2005). "Bayesian MCMC flood frequency analysis with historical information." *Journal of Hydrology*, 313, 97-116.

Kellagher, R., and Sayers, P. (2009). "Sam-system-based analysis and management of urban flood risks: a new procedure for performance assessment of sewerage systems." HR Wallingford report SR700.

Kanso, A., Chebbo, G., and Tassin, B. (2006). "Application of MCMC–GSA model calibration method to urban runoff quality modelling." *Reliability Engineering & System Safety* 91(10-11), 1398–1405.

Kapelan, Z., Savic, D. A., and Walters, G.A. (2005). "Multiobjective design of water distribution systems under uncertainty." *Water Resources Research*, 41(11), W11407.

Kapelan, Z., Savic, D.A., Walters, G.A., and Babayan, A.V. (2006). "Risk- and robustness-based solutions to a multi-objective water distribution system rehabilitation problem under uncertainty." *Water Science and Technology*, 53(1), 61-75.

Keen, P.G.W., and Morton, M.S.S. (1978). *Decision Support Systems: An Organizational Perspective* Addison-Wesley Publishing, Menlo Park.

Korving, H., Noortwijk, J.M., Gelder, P.H.A.J.M., and Parkhi, R.S. (2003). "Coping with uncertainty in sewer system rehabilitation." *Safety and Reliability*, 959-967.

Korving, H., Noortwijk, J.M.v., Gelder, P.H.A.J.M., and Clemens, F.H.L.R. (2009). "Risk-based design of sewer system rehabilitation." *Structure and Infrastructure Engineering*, 5(3), 215-227.

Koutsoyiannis, D. (2004). "Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation." *Hydrological Sciences*, 49(9), 575-590.

Koutsoyiannis, D., and Baloursos, G. (2000). "Analysis of a long record of annual maximum rainfall in Athens, Greece, and design rainfall inferences." *Natural Hazards* 22(1), 31-51.

Kuczera, G. (1999). "Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference." *Water Resources Research*, 35(5), 1551-1557.

Kulkarni, V.S., and Khanna, P. (1985). "Pumped wastewater collection systems optimization." *Journal of Environmental Engineering*, 111(5), 589-601.

Kurothe, R.S., Goel, N.K., and Mathur, B.S. (1997). "Derived flood frequency distribution of negatively correlated rainfall intensity and duration." *Water Resources Research*, 33(9), 2103–2107.

Kwon, H.-H., Moon, Y.-I., and Khalil, A.F. (2007). "Nonparametric Monte Carlo simulation for flood frequency curve derivation: an application to a Korean watershed." *Journal of the American Water Resources Association*, 43(5), 1316-1328.

Lamb, R., Beven, K., and Myrabo, S. (1998). "Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model." *Advances in Water Resources*, 22(4), 305-317.

Lancaster, J.W., Preene, M., and Marshall, C.T. (2004). "Development and Flood Risk-guidance for the construction industry." CIRIA, London.

Lei, J.H., and Schilling, W. (1994). "Parameter uncertainty propagation analysis for urban rainfall runoff modelling." *Water Science and Technology*, 29(1-2), 145-154.

Lekuthai, A., and Vongvisessomjai, S. (2001). "Intangible flood damage quantification." *Water Resources Management*, 15(5), 343-362.

Li, S.T., and Hammond, J.L. (1975). "Generation of pseudorandom numbers with specified univariate distribution and correlation coefficients." *IEEE transactions on systems, man and Cybernetics*, 5, 557-561.

Liang, L.Y., Thompson, R.G., and Young, D.M. (2004). "Optimising the design of sewer networks using genetic algorithms and tabu search." *Engineering. Construction and Architectural Management*, 11(2), 101-112.

Lindenschmidt, K.-E., Fleischbein, K., and Baborowski, M. (2007). "Structural uncertainty in a river water quality modelling system." *Ecological Modelling*. 204, 289-300.

Lund, J.R. (2002). "Floodplain planning with risk-based optimization." *Journal of Water Resources Planning and Management*, 128(3), 202-207.

Lurie, P.M., and Goldberg, M.S. (1998). "An approximate method for sampling correlated random variables form partially-specified distributions." *Management Science*, 44(2), 203-218.

Lyngfelt, S. (1991). "An Improved Rational Method for Urban Runoff Application." *Nordic Hydrology*, 22, 149-160.

Mays, L.W., and Wenzel, H.G. (1976). "Optimal design of multi-level branching sewer systems." *Water Resources Research*, 12(5), 913-917.

Mays, L.W. (2001). *Storm Collection Systems Design Handbook*. McGraw-Hill.

Merz, B., Kreibich, H., Thieken, A., and Schmidtke, R. (2004). "Estimation uncertainty of direct monetary flood damage to buildings." *Natural Hazards and Earth System Sciences,* 4, 153-163.

Merz, B., and Thieken, A.H. (2005). "Separating natural and epistemic uncertainty in flood frequency analysis." *Journal of Hydrology*, 309(1-4), 114-132.

Michele, D., and Rosso, R. (2001). "Uncertainty assessment of regionalized flood frequency estimates." *Journal of Hydrologic Engineering*, 6(6), 453-459.

Morgan, M.G., and Henrion, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, New York.

Morita, M. (2008). "Flood risk analysis for determining optimal flood protection levels in urban river management.". *Journal of Flood Risk Management*, 1(3), 142-149.

Morita, M., and Fukuda, T. (2002) "Decision support system for flood control facility planning based on inundation simulation and flood damage estimation using GIS." *Proceedings of 9th International Conference on Urban Drainage*, Portland.

Muzik, I. (2002). "A first-order analysis of the climate change effect on flood frequencies in a subalpine watershed by means of a hydrological rainfall-runoff model." *Journal of Hydrology*, 267, 65-73.

Nauta, M.J. (2000). "Separation of uncertainty and variability in quantitative microbial risk assessment models." *International Journal of Food Microbiology*, 57(1-2), 9-18.

O'Hagan, A., and Oakley, J.E. (2004). "Probability is perfect, but we can't elicit it perfectly " *Reliability Engineering & System Safety*, 85(1-3), 239-248.

Oliveri, E., and Santoro, M. (2000). "Estimation of urban structural flood damages: the case study of Palermo " *Urban Water Journal*, 2(3), 223-234.

Olsen, J.R., Beling, P.A., Lambert, J.H., and Haimes, Y.V. (1998). "Input-output economic evaluation of system of levees." *Journal of Water Resources Planning and Management*, 124(5), 237–245.

Padgett, J.E., Dennemann, K., and Ghosh, J. (2010). "Risk-based seismic life-cycle cost–benefit (LCC-B) analysis for bridge retrofit assessment " *Structural Safety*, 32(3), 165-173.

Pappenberger, F., and Beven, K. J. (2006). "Ignorance is bliss: or seven reasons not to use uncertainty analysis." *Water Resources Research*, 42, W05302.

Park, C.S. (1984). "Probabilistic benefit-cost analysis" *The Engineering Economist*, 29(2), 83-100.

Parker, M.A., Savic, D.A., Walters, G.A., and Kapelan, Z. "SewerNet: A Genetic Algorithm application for optimising urban drainage systems." *International Internet Conference on Urban Drainage*.

Parrish, R.S. (1990). "Generating random deviates from multivariate Pearson distributions." *Computational Statistics & Data Analysis*, 9(3), 283-295.

Parry, G.W. (1996). "The characterization of uncertainty in probabilistic risk assessment of complex systems." *Reliability Engineering & System Safety*, 54(2-3), 119-126.

Paté-Cornell, M.E. (1996). "Uncertainties in risk analysis: Six levels of treatment." *Reliability Engineering & System Safety*, 54(2-3), 95-111.

Plate, E.J. (2002). "Flood risk and flood management." *Journal of Hydrology*, 267(1-2), 2-11.

Quintero, A., Konaré, D., and Pierre, S. (2005). "Prototyping an intelligent decision support system for improving urban infrastructures management." *European Journal of Operational Research*, 162(3), 654-672.

Rahman, A., Weinmann, P.E., Hoang, T.M.T., and Laurenson, E.M. (2002). "Monte Carlo simulation of flood frequency curves from rainfall." *Journal of Hydrology*, 256(3-4), 196-210.

Reddy, M.J., and Kumar, D.N. (2007). "Multi-objective differential evolution with application to reservoir system optimization." *Journal of Computing in Civil Engineering*, 21(2), 136-146.

Reed, P.M., and Minsker, B.S. (2004). "Striking the balance: long-term groundwater monitoring design for conflicting objectives." *Journal of Water Resources Planning and Management*, 130(2), 140-149.

Reichert. P., and Omlin, M. (1997). "On the usefulness of overparameterized ecological models." *Ecological Modelling*, 95, 289-299.

Ross, J.L., Ozbek, M. M., and Pinder, G.F. (2009). "Aleatory and epistemic uncertainty in groundwater flow and transport simulation." *Water Resources Research*, 45, W00B15.

Rossman, L.A. (2008). *Storm Water Management Model User's Manual(version 5.0)* U.S. Environment Protection Agency, Cincinnati, USA.

Ryu, J. (2008). "Decision Support for Sewer Flood Risk Management," Imperial College London, London.

Ryu, J., and Butler, D. (2008). "Managing Sewer Flood Risk" 11th International Conference on Urban Drainage, Edinburgh, UK.

Salvadori, G., and Michele, C.D. (2006). "Statistical characterization of temporal structure of storms." *Advances in Water Resources*, 29(6), 827-842.

Salvadori, G., and Michele, C.D. (2007). "On the use of copulas in hydrology: theory and practice." *Journal of Hydrologic Engineering*, 12(4), 369-380.

Sayers, P., Gouldby, B., Simm, J., Meadowcroft, I., and Hall, J. (2002). "Risk, performance and uncertainty in flood and coastal defence – a review." *R&D Technical Report FD2302/TR1*, HR Wallingford Ltd.

Schweizer, B., and Wolff, E.F. (1981). "On nonparametric measures of dependence for random variables." *Annals of Statistics*, 9(4), 870-885.

Sentz, K. (2002). "Combination of evidence in Dempster-Shafer theory." Binghamton.

Serinaldi, F. (2009). "Assessing the applicability of fractional order statistics for computing confidence intervals for extreme quantiles." *Journal of Hydrology*, 376 (3-4), 528-541.

Shafer, G. (1976). *A mathematical theory of evidence*, Pinceton University Press, Princeton.

Simon, H.A. (1977). *The New Science of Management Decision*, Prentice-Hall., New York.

Smith, D.I. (1994). "Flood damage estimation- a review of urban stage-damage curves and loss functions." *Water SA*, 20(3), 231-238.

Smith, E.A., Ryan, P.B., and Evans, S.J. (1992). "The effect of neglecting correlations when propagating uncertainty and estimating the population distribution of risk." *Risk Anal*, 12(4), 467-474.

Stedinger, J.R., and Cohn, T.A. (1986). "Flood frequency analysis with historical and paleoflood information." *Water Resources Research*, 22(5), 785-793.

Taha, A.H. (2007). *Operations research : an introduction. 8th edn*, Upper Saddle River, N.J. , Pearson Education.

Thorndahl, S., and Willems, P. (2008). "Probabilistic modelling of overflow, surcharge and flooding in urban drainage using the first-order reliability method and parameterization of local rain series." *Water Research*, 42(1-2), 455-466.

Tonon, F. (2004). "Using random set theory to propagate epistemic uncertainty through a mechanical system " *Reliability Engineering & System Safety*, 85(1-3), 169-181.

Tonon, F., and Bernardini, A. (1998). "A random set approach to the optimization of uncertain structures." *Computers and Structures*, 68(6), 583-600.

Tsompanakis, Y., Papadopoulos, V., Lagaros, N.D., and Papadrakakis, M. "Reliability analysis of structures under seismic loading." *Fifth World Congress on Computational Mechanics*, Vienna, Austria.

Tung, Y.K., Wang, P.Y., and Yang, J.C. (1993). "Water resource projects evaluation and ranking under economic uncertainty." *Water Resources Management*, 7(4), 311-333.

Tung, Y.K., and Yen, B.C. (2005). *Hydrosystems Engineering Uncertainty Analysis*, ASCE Press and McGraw-Hill New York, USA.

Verworn, H.R. (2002). " Advances in urban-drainage management and flood protection." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 360(1796), 1451-1460.

Voortman, H.G. (2003). "Risk-Based Design of Large-Scale Flood Defence Systems." PrintPartners Ipskamp BV, the Netherlands.

Vořechovský, M., and Novák, D. (2009). "Correlation control in small-sample Monte Carlo type simulations I: A simulated annealing approach." *Probabilistic Engineering Mechanics*, 24 (3), 452-462.

Vrijling, J.K. (2001). "Probabilistic design of water defense systems in the Netherlands." *Reliability Engineering & System Safety*, 74(3), 337–344.

Wallingford Software. (2001). "InfoWorks CS Technical Review."

Walters, G.A., and Lohbeck, T. (1993). "Optimal layout of tree networks using genetic algorithms " *Engineering Optimisation*, 22(1), 27-48.

Williamson, R.C., and Downs, T. (1990). "Probabilistic arithmetic. I. numerical methods for calculating convolutions and dependency bounds." *International Journal of Approximate Reasoning* 4(2), 89-158.

WMO, and GWP. (2008). "Urban Flood Risk Management: A Tool for Integrated Flood Management." APFM technical document No11, Flood Management Tools Series.

Wu, J.S., Apostolaskis, G.E., and Okrent, D. (1990). "Uncertainties in system analysis: probabilistic versus nonprobabilistic theories." *Reliability Engineering & System Safety*, 30(1-3), 163-181.

Yen, B.C., Cheng, S.-T., Jun, B.-H., and L.Werzel, M. (1984). "Illinois Least Cost Sewer System Design Model. User's Guide." Austin, Tex.

Yue, S. (2000). "Joint probability distribution of annual maximum storm peaks and amounts as represented by daily rainfalls" *Hydrological Sciences Journal*, 45(2), 315–326.

Zadeh, L.A. (1965). *Fuzzy set*, Inf. Control 8.

Zadeh, L.A. (1978). "Fuzzy sets as a basis for a theory of possibility." *Fuzzy Sets and Systems*, 100(s1), 3-28.

Zhang, L., Singh, P.V., and ASCE, F. (2007). "Gumbel-hougaard copula for trivariate rainfall frequency analysis." *Journal of Hydrologic Engineering*, 12(4), 409-419.