

# Flood analysis of urban drainage systems: Probabilistic dependence structure of rainfall characteristics and fuzzy model parameters

Guangtao Fu and Zoran Kapelan

## ABSTRACT

Flood analysis of urban drainage systems plays a crucial role for flood risk management in urban areas. Rainfall characteristics, including the dependence between rainfall variables, have a significant influence on flood frequency. This paper considers the use of copulas to represent the probabilistic dependence structure between rainfall depth and duration in the synthetic rainfall generation process, and the Gumbel copula is fitted for the rainfall data in a case study of sewer networks. The probabilistic representation of rainfall uncertainty is combined with fuzzy representation of model parameters in a unified framework based on Dempster–Shafer theory of evidence. The Monte Carlo simulation method is used for uncertainty propagation to calculate the exceedance probabilities of flood quantities (depth and volume) of the case study sewer network. This study demonstrates the suitability of the Gumbel copula in simulating the dependence of rainfall depth and duration, and also shows that the unified framework can effectively integrate the copula-based probabilistic representation of random variables and fuzzy representation of model parameters for flood analysis.

**Key words** | copula, dependence structure, evidence theory, flood analysis, fuzzy set, urban drainage system

**Guangtao Fu** (corresponding author)  
**Zoran Kapelan**  
Centre for Water Systems,  
College of Engineering,  
Mathematics and Physical Sciences,  
University of Exeter,  
North Park Road,  
Harrison Building,  
Exeter,  
UK  
E-mail: [g.fu@exeter.ac.uk](mailto:g.fu@exeter.ac.uk)

## INTRODUCTION

Urban flooding is a major social and environmental issue in urban areas and can result in significant damage to properties and infrastructure, environmental pollution and traffic interruption. According to OFWAT (2002), sewer flooding, so-called localised flooding from overloaded/blocked sewer systems, is the second-most serious issue facing UK water companies after drinking water quality. It has an estimated cost of £270 million a year in England and Wales alone (POST 2007). Most sewer systems in the UK are combined sewer systems (CSS), and were designed on the basis of simple deterministic methods such as the rational method or the time-area method (Butler & Davies 2011). Using these methods, the system capacity is designed to convey dry weather flows and surface runoff from the rainfall with a specific return period. However, these methods can only ensure that the surcharge occurrence is less than that of

the rainfall, and flooding is not included in the calculation (Thorndahl & Willems 2008). There is therefore a real need to assess hydraulic performance of the sewer system by analysing sewer flood or combined sewer overflow discharge frequency statistics based on hydrodynamic simulation (Schmitt *et al.* 2004; Fu *et al.* 2011; Sun *et al.* 2012; Fu & Butler *in press*).

Understanding rainfall characteristics, in particular the dependence between rainfall variables, is very important in hydrological modelling, particularly for flood risk assessment. For example, prior research has demonstrated the importance of the dependence relationship between rainfall event duration and average intensity in deriving flood-frequency distributions (Goel *et al.* 2000) and in reproducing observed intensity–frequency–duration extreme rainfall statistics for risk-based design (Heneker *et al.* 2001).

Fontanazza *et al.* (2011) investigated the impact of the dependence between rainfall depth, duration and maximum intensity on synthetic rainfall generation and sewer flood estimation. Many bivariate or multivariate distributions have been applied to describe rainfall characteristics, for example, the Gumbel bivariate exponential distribution (e.g. Bacchi *et al.* 1994; Yue 2000; Kao & Govindaraju 2007b). Pearson's correlation coefficient  $\rho$  cannot measure the dependence between other types of the marginals appropriately, except in the case of Gaussian distributions (Nelsen 2006). Further, the marginal distributions for different variables are generally assumed to be of the same type or are transformed into the same type. This severely limits the flexibility to represent rainfall characteristics and their relationships (De Michele & Salvadori 2003; Zhang & Singh 2007; Vandenberghe *et al.* 2010).

In addition to stochastic rainfall uncertainty, there are epistemic uncertainty sources that result from incomplete knowledge of fundamental phenomena. For example, different Bayesian inference methods have been developed for the treatment of rainfall and model structural uncertainties building on the prior knowledge of uncertainties (e.g. Renard *et al.* 2010). Fuzzy sets can also be used to capture expert knowledge and imprecise data (Zadeh 1965). It is not uncommon that uncertainties of different natures have to be considered simultaneously in the uncertainty analysis process (Hall *et al.* 2007; Ross *et al.* 2009; Fu & Kapelan, 2011; Sun *et al.* 2012). For example, probability distributions are used to represent rainfall uncertainty when sufficient data are available, while intervals or fuzzy sets are used to describe the uncertainty in model parameters from different expert opinions or measurement errors. Merz & Thielen (2005) argued that the stochastic and epistemic uncertainties should be separated from each other to better understand their contributions to the total uncertainty in model output. The Dempster–Shafer theory of evidence (Dempster 1967; Shafer 1976) provides a unified framework in which uncertainties are represented separately using different characterisation methods and are propagated through a combined random set for uncertainty analysis (Hall 2003; Tonon 2004; Hall *et al.* 2007; Fu *et al.* 2011).

This paper investigates the dependence structure between rainfall depth and duration in a unified uncertainty analysis framework for sewer flood analysis that

accommodates both probabilistic and fuzzy uncertainties. The unified framework is built on a previously developed methodology on the basis of the Dempster–Shafer theory (Fu *et al.* 2011), in which probability distributions of rainfall variables and fuzzy sets of model parameters are handled simultaneously. This new methodology provides a single mathematical framework to handle uncertainties of different natures, allowing uncertainty to be modelled in the most appropriate way (Hall 2003).

This paper further develops the unified framework by using copulas to handle the dependence structure between correlated rainfall quantities. Copulas are a powerful tool to represent joint multivariate distributions (Sklar 1959; Nelsen 2006). In recent years, the use of copulas in hydrology has gained increasing attention with an attempt to describe the probabilistic structures of random variables, for example rainfall duration, intensity and depth (e.g. De Michele & Salvadori 2003; Kao & Govindaraju 2007a; Zhang & Singh 2007; Fontanazza *et al.* 2011; Fu & Butler *in press*) and flood duration, peak flow and flood volume (e.g. Zhang & Singh 2006; Chowdhary *et al.* 2011). These prior studies have demonstrated the advantage of copulas in flexibly representing the dependence structure among random variables and using different types of marginal distributions for multivariate distributions in hydrology (Kao & Govindaraju 2007b; Vandenberghe *et al.* 2010).

The copula-based uncertainty analysis framework is demonstrated using a case study of urban drainage systems in calculating the cumulative distribution functions (CDFs) or exceedance probabilities of flood quantities. In this paper, the dependence between rainfall depth and duration is represented using the Gumbel family of Archimedean copulas. The model parameters are represented using fuzzy sets that can more precisely reflect incomplete knowledge and data about the parameter values. The Monte Carlo simulation method is then used to generate synthetic rainfall events on the basis of the simulated Gumbel copula, with the assumption of homogeneity in climate and catchment characteristics. The two types of uncertainty – stochastic and fuzzy – are propagated through the urban drainage model using the copula-based uncertainty analysis framework. As a result, the model outputs (flood depth and volume) are generated in the form of lower and upper CDFs (or exceedance probabilities), which provide best

estimates given various stochastic and epistemic uncertainties considered.

The results obtained in this study show the suitability and flexibility of the Gumbel family in simulating the dependence of rainfall depth and duration and the advantage of the copula method in sewer flood analysis. The results also show that the unified framework can effectively integrate the copula tool to simulate the dependence structure between random rainfall variables.

## METHODOLOGY

This copula-based framework is built on the Dempster-Shafer theory of evidence (Dempster 1967; Shafer 1976) and the copula theory (Sklar 1959). The former provides a framework that allows for probabilistic and fuzzy representations of uncertainty, and the latter provides a flexible way to simulate the correlation between two random variables. In the framework, a Monte Carlo simulation method is developed to propagate the uncertainties through a model. The Latin Hypercube Sampling (LHS) method (McKay *et al.* 1979) is used to improve computational efficiency (Fu *et al.* 2011). A brief introduction is given below to the above theories and methods, and the reader is referred to the provided references for more details.

### Copulas

Copulas are joint probability functions of random variables that are expressed in terms of marginal distribution functions and an associated dependence structure. For two random variables  $X$  and  $Y$ , their marginal CDFs are represented by

$$u = F_X(x) \quad \text{and} \quad v = F_Y(y) \quad (1)$$

where  $u$  and  $v$  are uniformly distributed random variables and  $u, v \in [0, 1]$ . The joint CDF  $H_{XY}(x, y) = P(X \leq x, Y \leq y)$  describes the probability of two events:  $X \leq x$  and  $Y \leq y$ . According to Sklar's theorem (Sklar 1959), the CDF  $H_{XY}(x, y)$  can be represented as

$$H_{XY}(x, y) = C(u, v) \quad (2)$$

where  $C(u, v)$  is called a copula and can be uniquely determined when  $u$  and  $v$  are continuous. Through Equation (2), it is easy to see that the copula is actually a multivariate distribution function with uniform marginals (Nelsen 2006).

The definition in Equation (2) provides two main advantages in deriving the joint probability function  $H_{XY}(x, y)$ : (1) the marginals can be determined using different distributions, and (2) the dependence structure can be described separately from the marginals, which allows for complex multivariate distributions to be built to model stochastic phenomena such as rainfall without the knowledge of marginal distributions.

There are many families of copulas that represent different dependence structures. Archimedean copulas are of special interest for hydrologic analyses, and include the popular Gumbel, Frank and Cook-Johnson families. The Archimedean copulas can be expressed as

$$C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)) \quad (3)$$

where  $\varphi$  is a generator that is a convex decreasing function satisfying  $\varphi(1) = 0$  and  $\lim_{t \rightarrow 0} \varphi(t) = \infty$ . Assuming the generator  $\varphi(t) = (-\ln t)^\theta$ , Equation (3) can be rewritten as

$$C(u, v) = \exp \left\{ - \left[ (-\ln u)^\theta + (-\ln v)^\theta \right]^{1/\theta} \right\} \quad (4)$$

where  $\theta$  is a dependence parameter and  $\theta \in (1, \infty)$ . Equation (4) is referred to as the Gumbel family of copulas. This family of copulas has been used to represent the statistical dependence of rainfall and flood variables (e.g. Zhang & Singh 2006 & 2007; Balistocchi & Bacchi 2011; Chowdhary *et al.* 2011), and was chosen to describe the relationship between the rainfall depth and duration for the case study catchment considered in this paper.

The parameter  $\theta$  can be estimated using several different methods, for example the maximum likelihood method and the Inference Functions for Margins (IFM) method (Chowdhary *et al.* 2011). The parameter can also be derived from the Kendall  $\tau$  coefficient. In statistics, Kendall's  $\tau$  is used to determine the ordering of two measured quantities and can be estimated from observations. There is a close

form between  $\tau$  and  $\theta$  for Gumbel copula:

$$\tau = 1 - \theta^{-1} \quad (5)$$

To verify the appropriateness of the chosen copula, a non-parametric (empirical) approach can be used to estimate the empirical copula for comparison (Kao & Govindaraju 2007a; Zhang & Singh 2007): (1) assume an intermediate random variable  $z$  whose samples can be transformed from the observations  $z_i = \{\text{number of } (x_j, y_j) \text{ such that } x_j < x_i \text{ and } y_j < y_i\} / (N-1)$  for  $i = 1, 2, \dots, N$ ; (2) estimate the probability for  $z$  as the proportion of  $z_i < z$ . The distribution plots or Q-Q plots can be plotted for examination, which provides a good indication if the theoretical copula fits the data well.

### Dempster–Shafer theory of evidence

The Dempster–Shafer theory of evidence was developed by Dempster (1967) and Shafer (1976), and is regarded as a generalisation of probability theory. It is equivalent to the random set theory (Kendall 1974). It allocates probability masses to subsets rather than singletons of a given universe, which allows it to consider imprecision in the sense of epistemic uncertainty. A brief introduction of the evidence theory is provided below; for more information see Dubois & Prade (1991), Hall (2003), Alvarez (2009), Fu *et al.* (2011) and Sun *et al.* (2012).

Let  $X$  be a universal non-empty set containing all the possible values of a variable  $x$ , and  $P(X)$  the power set of  $X$ , i.e. the set of all the subsets of  $X$ . Distinct from probability theory, evidence theory defines a mapping from the family of non-empty elements  $\mathfrak{S}$  of  $P(X)$

$$m: \mathfrak{S} \rightarrow [0, 1] \quad (6)$$

such that  $m(\emptyset) = 0$  and

$$\sum_{A \in \mathfrak{S}} m(A) = 1 \quad (7)$$

where  $A \in P(X)$  for which  $m(A) > 0$  is called focal element, and  $m$  is called the basic probability assignment. Each set  $A$

contains some possible values of the variable  $x \in X$  and the value  $m(A)$  expresses the probability that  $x \in A$  but does not belong to any subsets of  $A$ . This does not exclude that some elements of  $A$  contribute to the probability of subset  $B \in P(X)$  so that  $A \cap B \neq \emptyset$ .

A random set  $(\mathfrak{S}, m)$  on  $X$  assigns a probability to all the subsets of  $X$ , while classical probability theory only considers the singleton subsets of  $X$ . This is designed to deal with the uncertainty where the information is not sufficient to permit the probability assignment to single events. Due to the imprecise nature of this formulation, it is impossible to calculate the precise probability of a subset  $E \subset X$ , i.e.  $P(E)$ . Instead, the related imprecision of this probability can be bounded at the lower end by the belief function *Bel* (Dempster 1967; Shafer 1976):

$$\text{Bel}(E) = \sum_{A \subseteq E} m(A) \quad (8)$$

and at the upper end by the plausibility function *Pl*:

$$\text{Pl}(E) = \sum_{A \cap E \neq \emptyset} m(A) = 1 - \text{Bel}(\bar{E}) \quad (9)$$

where  $\bar{E}$  is the complement of  $E$ . The belief  $\text{Bel}(E)$  measures the minimum amount of evidence that fully supports  $x \in E$ , i.e. which cannot be removed from  $E$  because the summation in Equation (8) only involves  $A$  such that  $A \subseteq E$ . Similarly, the plausibility  $\text{Pl}(E)$  measures the maximum amount of evidence that could be linked with the event  $E$ , i.e. which could be counted into  $E$  because the summation in Equation (9) involves all  $A$  such that  $A \cap E \neq \emptyset$ .

The evidence theory provides a valuable theoretical framework for modelling and decision making in a situation where the information available does not entail a purely probabilistic treatment of uncertainty and has an advantage in simultaneously handling a range of different types of uncertainty information, for example intervals, probability distributions and fuzzy sets (Hall 2003). It has been applied to a number of diverse areas such as global climate change prediction (Hall *et al.* 2007), urban drainage modelling (Fu *et al.* 2011; Sun *et al.* 2012), detection of pipe bursts in water distribution systems (Bicik *et al.* 2011), groundwater

transport simulation (Ross *et al.* 2009) and mechanical and structural system design (Tonon 2004).

### Uncertainty analysis framework

In this paper, a Monte Carlo simulation method is applied first to sample the fuzzy sets of uncertain model parameters and the joint probability distribution of rainfall variables. The samples generated this way are then propagated through an urban drainage model under the framework of evidence theory (Alvarez 2009). The detailed steps of the framework are as follows.

1. Assume that  $n$  sample points  $\mathbf{u}_i = (u_i^1, \dots, u_i^j, \dots, u_i^d) \in (0, 1]^d$  ( $i = 1, \dots, n; j = 1, \dots, d$ ) are generated from the independent uniform distributions on  $(0, 1]$  using the LHS technique (McKay *et al.* 1979), where  $d$  represents the number of uncertainty variables considered.
2. Generate samples for correlated rainfall variables  $(X, Y)$  (Kao & Govindaraju 2007b). Taking two independent random pairs, for instance the first two samples  $(u_i^1, u_i^2)$ , the samples for  $X$  can be derived by solving the inverse function of the marginal CDF  $F_X(x)$ , i.e. by assuming  $F_X(x) = u_i^1$ . For variable  $Y$ , the samples can be generated by assuming  $P[Y \leq y|X = x] = u_i^2$  because  $P[Y \leq y|X = x]$  is independent of  $P[X \leq x]$ . This requires that  $v_i$  is first solved through a simplified expression with copulas  $\partial C(u, v)/\partial u = u_i^2$  and the samples for  $Y$  are calculated from its corresponding inverse distribution by  $F_Y(y) = v_i$ . In this way, the correlated random samples  $(x_i, y_i)$  can be generated from their margins  $(u_i^1, v_i)$ . The sample  $x_i$  or  $y_i$  can be regarded as a set  $A_i^1 = \{x_i\}$  or  $A_i^2 = \{y_i\}$  with a single element, respectively.
3. Generate samples from fuzzy sets. The  $\alpha$ -cut of a fuzzy set is used to generate samples. The  $\alpha$ -cut of a fuzzy set is defined as the set containing all the values  $x$  with membership degree no less than  $\alpha \in [0, 1]$ , i.e.  $S = \{x|\mu_S(x) \geq \alpha\}$ , where  $\mu_S(x)$  is the fuzzy membership function. As a result, a series of sets  $S_1, S_2, \dots, S_n$  are generated for a fuzzy set. The samples can be obtained by deriving the interval at  $u_i^j$ -level cut of the fuzzy number, i.e.  $A_i^j = \{x_j|\mu_S(x_j) \geq u_i^j\}$  ( $i = 1, \dots, n; j = 1, \dots, d$ ).

4. Through steps (2) and (3), one set is generated for each sample  $u_i^j$ . The joint focal element  $A_i$  can therefore be constructed as  $A_i = A_i^1 \times \dots \times A_i^d$  in the space  $X$  ( $i = 1, \dots, n$ ). As a result, a random set  $(F_n, m)$  is generated using the LHS sampling where  $F_n = \{A_1, A_2, \dots, A_n\}$  and  $m(A_i) = 1/n$ , implying that an equal weight is assigned to the sampled elements as they are generated randomly.
5. The joint random set  $(F_n, m)$  contains all the information from all uncertainty sources of different types. The image of all focal elements in the joint random set can be calculated according to the random set extension principle by Dubois & Prade (1999). The reader is referred to the work by Fu *et al.* (2011) for more details about random set propagation.

The lower and upper cumulative probabilities (belief and plausibility measures) can be constructed on the basis of the propagated random set of the model output  $H$  – flood quantities (flood depth or volume) in the case study. Assume firstly that the flood depth/volume axis is partitioned into adjoining intervals  $[h_1, h_2], [h_2, h_3], \dots, [h_s, h_{s+1}]$  denoted  $R_1, R_2, \dots, R_s$ , respectively. According to Equations (8) and (9), the lower and upper CDFs  $\underline{F}(h)$  and  $\bar{F}(h)$  at some point  $h$  defined on the domain  $[h_1, h_{s+1}]$  can be obtained as follows (Hall *et al.* 2007):

$$\underline{F}(h) = \text{Bel}(\{H \leq h\}) = \sum_{h \geq \sup(R_j)} \rho(R_j) \quad (10)$$

$$\bar{F}(h) = \text{Pl}(\{H \leq h\}) = \sum_{h \geq \inf(R_j)} \rho(R_j) \quad (11)$$

These derived bounds represent the best possible knowledge of flood quantities, given all kinds of uncertainties in variables, and should embrace the unknown true CDF of flood quantities. The spread of the bounds represents the extent of imprecision and incompleteness in uncertainty representations and can only be reduced when more knowledge is available. The CDFs can be converted into exceedance probabilities of flood quantities.

## CASE STUDY

### Urban drainage system

A real-life urban drainage system in the UK, as shown in Figure 1, is used to demonstrate the uncertainty analysis framework developed in this paper with particular focus on the copula part of the methodology. The total catchment area is about 200 hectares, serving a population of 4,000. The sewer system model consists of 265 nodes, 265 pipes, 2 outfalls and 1 weir, and has a total conduit length of 22,482 m. The pipe gradients vary from 0.0001 to 0.0439. Flows are diverted downstream via two outfalls; one is connected to a wastewater treatment works (Outfall 1) and the other to a combined sewer overflow (Outfall 2), and both flows are eventually discharged into a river.

The storm water management model (SWMM), developed by the US Environmental Protection Agency, was used for hydrologic simulation of rainfall-runoff in the urban catchment and for hydrodynamic simulation of in-sewer transport through the urban drainage system. The sewer system model has been calibrated for flood evaluation in prior studies; for more details the reader is referred to the work of Fullerton (2004).

### Rainfall data

Rainfall depth and duration were used to analyse statistical characteristics of the actual rainfall events in the case study of urban catchment, and a series of 10 years' rainfall data were used in this study. To analyse the rainfall data, a

2-hour time interval between two consecutive events was used to separate individual rainfall events. It should be noted that a threshold of 2 mm is applied to rainfall depth to remove small rainfall events (simulation results show that no flooding occurs in the sewer system under these events). A total of 650 events were used for simulation, as shown in Figure 2(a). There is a high frequency of low rainfall depth although the upper range is close to 50 mm. Similarly, most events have a short duration but some have a duration of up to 1,800 min.

The Generalised Extreme Value Distribution (GEV) has been used to describe extreme random variables and is used in this study to fit rainfall depth and duration data. The marginal distribution function is expressed as

$$F(x) = \exp\left(-[1 + \xi(x - \mu)/\sigma]^{-1/\xi}\right) \quad (12)$$

where  $\xi \neq 0$ ,  $\sigma > 0$  and  $\mu$  are shape, scale and location parameters, respectively.

The fitted distributions for rainfall depth and duration are given in Table 1. Three goodness-of-fit tests, i.e. Kolmogorov Smirnov (K-S), Anderson Darling (A-D) and Chi-square ( $\chi^2$ ) tests, show that the fitted distributions cannot be rejected at the significance level of 5%. Table 1 shows the test statistics for the two fitted distributions. Further, Figures 2(b) and 2(c) show the histograms of the events and fitted marginal probability distributions. A comparison with the empirical distributions shows a good agreement. The correlation relationship between the two rainfall variables is investigated using copulas and the results are analysed in the results section.

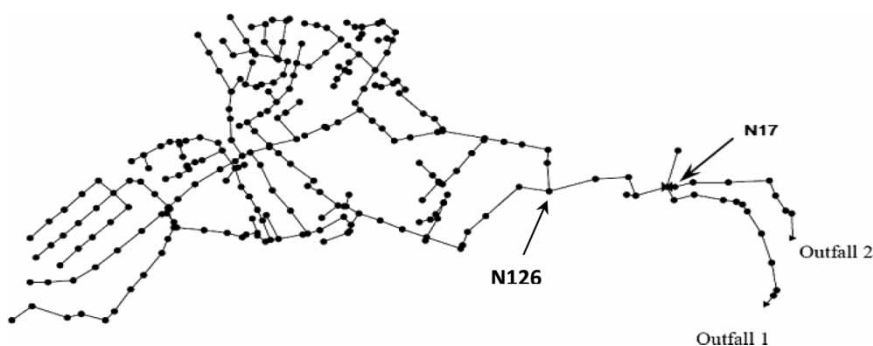
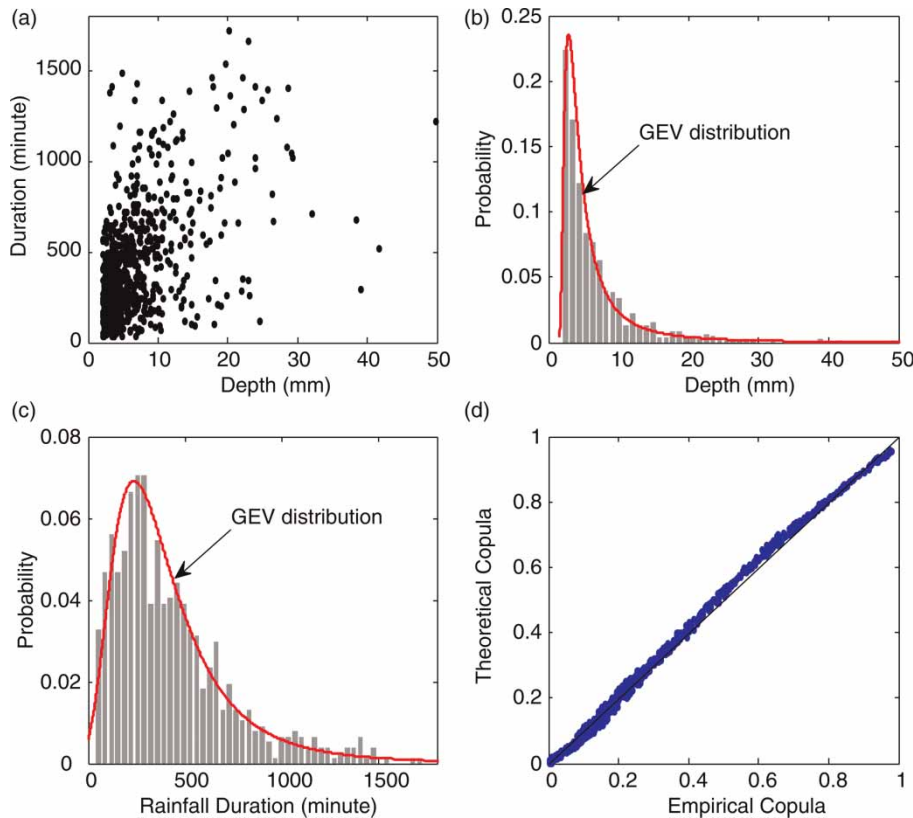


Figure 1 | Layout of the case study network.



**Figure 2** | Rainfall depth and duration marginal distributions and their copulas.

**Table 1** | Fitted marginal distributions for rainfall variables and goodness-of-fit statistics

Variables	GEV distribution parameter	K-S		$\chi^2$		A-D $A_n^2$
		$K_n$	p-value	$\chi_n^2$	p-value	
Rainfall duration	$\xi = 0.105, \sigma = 189.12, \mu = 267.9$	0.026	0.832	3.126	0.959	0.433
Rainfall depth	$\xi = 0.680, \sigma = 1.746, \mu = 4.594$	0.052	0.091	16.226	0.062	2.369

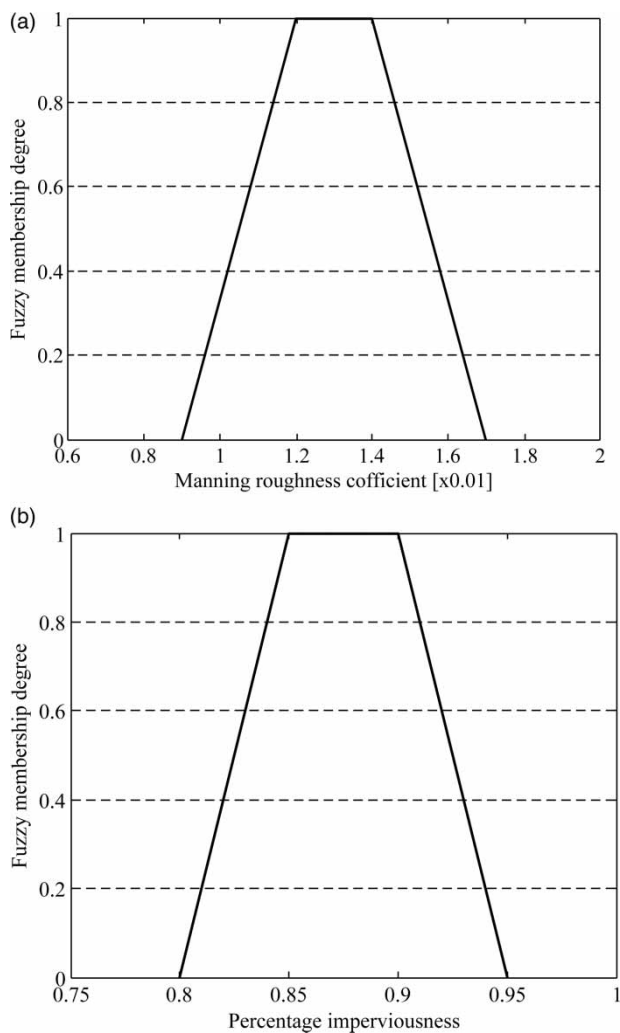
### Uncertain model parameters

Urban drainage models have a wide range of parameters. Choi & Ball (2002) categorised these parameters into two groups: measured parameters (i.e. those that can be physically measured such as catchment areas and pipe diameters) and inferred parameters (those that cannot be measured but can be inferred from the application of a model such as catchment imperviousness and Manning roughness). The measured parameters are normally assumed to be fixed, and the inferred parameters need to

be determined in terms of goodness-of-fit measures on the basis of available data during a model calibration process. According to the work of Fullerton (2004), the Manning roughness of the pipes and percentage imperviousness of the subcatchments are the most sensitive model parameters in the case study catchment; they are therefore considered as uncertain parameters to demonstrate the proposed methodology. Similarly other model parameters can be easily incorporated in the methodology if necessary.

The SWMM model parameters, Manning roughness and percentage imperviousness are assumed to be of the

epistemic type of uncertainty. After a certain period of service of a sewer system, the roughness coefficients of the pipes are difficult to estimate given complex pipe aging processes. The imprecise information can be better described using a fuzzy set instead of a probability distribution due to lack of data (Revelli & Ridolfi 2002; Hosseini & Ghasemi 2012). Similarly, the catchment percentage imperviousness, which has a direct influence on runoff calculation, has substantial uncertainty and is also represented by fuzzy sets in this study. A trapezoidal shape for the two model parameters is assumed, as shown in Figure 3.



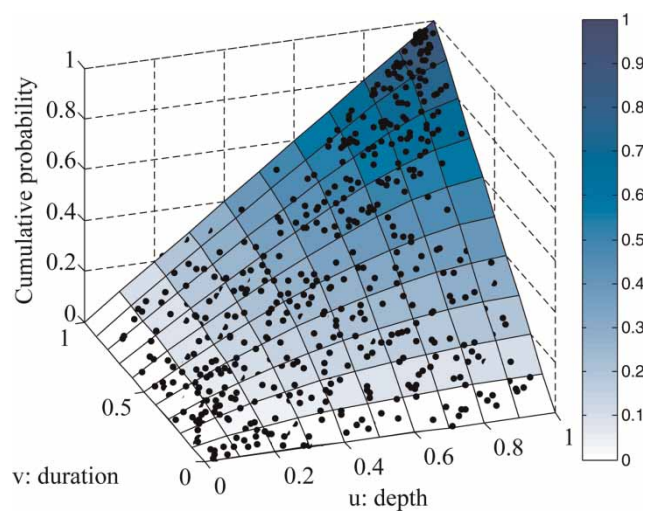
**Figure 3** | Fuzzy representation of model parameters: (a) Manning roughness coefficient and (b) percentage imperviousness.

## RESULTS AND DISCUSSION

### Dependence structure of rainfall depth and duration

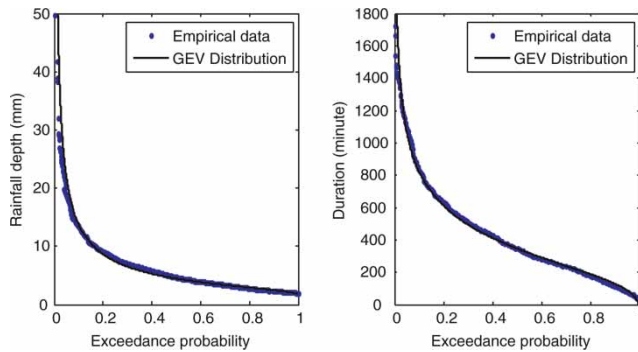
The Gumbel family of copulas are used to represent the joint distribution of rainfall depth and duration on the basis of their marginal distributions. The parameter  $\theta$  of Gumbel copula is estimated using the Maximum Likelihood method, and has a value of 1.492. Recall that parameter  $\theta$  can also be derived according to the relationship between  $\theta$  and  $\tau$  as given in Equation (5). In this case, the empirical value is calculated as 1.505. It can be seen that these two values have very good agreement. The Q-Q plot shown in Figure 2(d) provides a comparison between the parametrically estimated copula and empirical copula. The diagonal straight line represents a perfect match between the copulas. It can be seen that the quantiles of the two copulas are in good agreement.

Figure 4 visualises the fitted copula (shaded surface) together with the empirical copula (points). It should be noted that the variables  $u$  and  $v$  represent the transformed random variables  $X$  and  $Y$  (rainfall depth and duration) in the unit hypercube, and have the same ranks as  $X$  and  $Y$ . Figure 5 show the Q-Q plots for the two rainfall variables. The GEV distributions are plotted using a set of 10,000 samples generated using the copula method. The empirical distributions are based on the identified rainfall events,



**Figure 4** | Three-dimensional plots for theoretical and empirical copulas of rainfall depth and duration.





**Figure 5** | Theoretically fitted probability distributions for rainfall variables based on the Gumbel copula. The GEV distributions are plotted using a set of 10,000 samples generated using the copula method.

i.e. 650 events. These Q–Q plots show the fitted GEV distributions representing the historical data well, although the probability errors at the low tail exceedance increase as seen in the magnified areas.

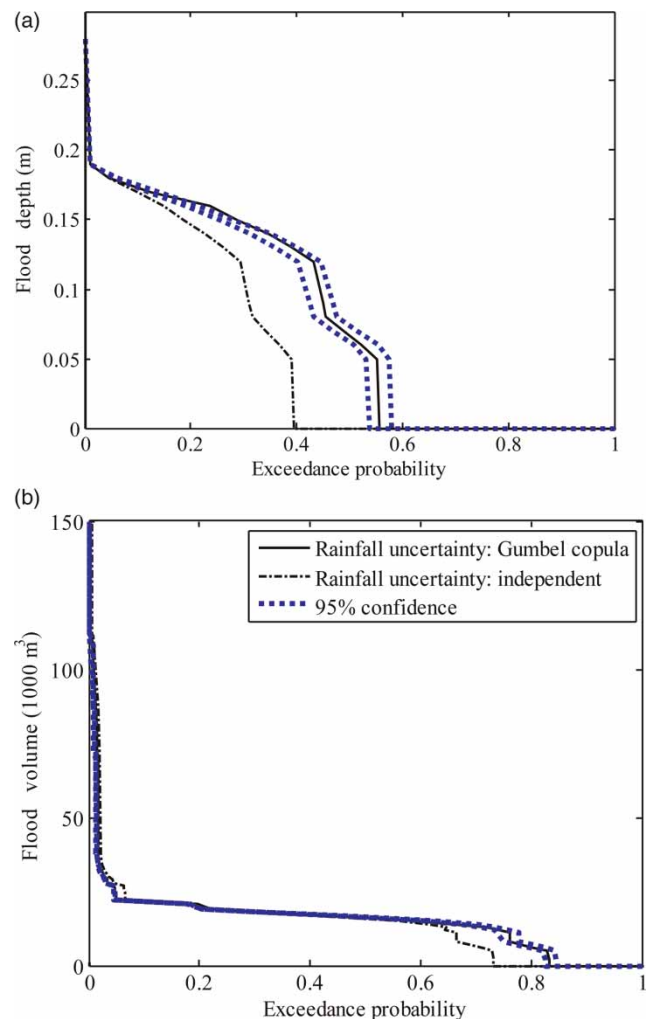
### Impact of rainfall depth–duration dependence

The theoretically fitted Gumbel copula was used to generate a large set of 10,000 samples for rainfall depth and duration. The number of samples used here has proven to be sufficient because the CDFs derived barely vary with the increase in the number of samples. Synthetic rainfall events were produced by applying a rectangular pulse with duration as the width and average rainfall intensity as the height, and they were then used as inputs to the sewer system model to calculate flood volume at the different nodes. Uniform rainfall intensity is commonly assumed in some simple rainfall–runoff methods, such as the Rational method (Butler & Davies 2011). It was also assumed here because the focus of this study is to demonstrate the copula method. However, different rainfall profiles can easily be incorporated in the stochastic-epistemic framework as an additional stochastic uncertainty source (Sun *et al.* 2012). The uncertainties in the SWMM model parameters are not considered in this case.

Manning roughness and percentage imperviousness are set to the default values 0.013 and 0.875, respectively, which were adapted from the calibrated model by Fullerton (2004).

Flood depth at node N126 (Figure 1) is selected to show the cumulative probabilities obtained from the copula method as it is one of the critical nodes that have a high

occurrence of flooding (Fullerton 2004). In addition, the total flood volume is calculated as the total amount of water flowing out of all network manholes during a rainfall event and is used to represent the overall system performance in terms of flooding. Figures 6(a) and 6(b) show the exceedance probabilities (represented by the solid lines) of flood depth at node N126 and total flood volume respectively. It can be seen from Figure 6(a) that the probability of flooding occurring (i.e. flood depth > 0) for any rainfall event is 0.56 at node N126. In other words, the probability of no flooding occurring (i.e. flood depth  $\leq 0$ ) has a value of 0.44. According to Figure 6(b), the probability of flooding occurring is 0.83 for the sewer system, i.e. flooding in at least



**Figure 6** | Exceedance probabilities of flood quantities considering stochastic rainfall uncertainty only: (a) flood depth at node N126 and (b) total flood volume in the sewer system.

one node. Clearly the probability of flooding in the entire sewer system is higher than at a single node such as N126 because flooding might occur at other nodes even when node N126 is not flooded during an event. These high probabilities of 'failure' at node N126 or system level are caused by the expansion of the network to the (left) upstream due to urban development (Fullerton 2004).

For comparison, Figures 6(a) and 6(b) also show the exceedance probabilities (the dash-dot line) when rainfall depth and duration are assumed independent. In this case, the probability of flooding at node N126 is 0.4, which is lower than in the case of correlation. This implies that the flood depth is underestimated without considering the correlation between rainfall depth and duration. Similar results are found for total flood volume as shown in Figure 6(b), i.e. the total flood volume is underestimated when the correlation between rainfall depth and duration is ignored, particularly at the low levels of flood volume. This can be explained by the fact that, in the case of dependence, higher rainfall depths are more likely related to low durations, generating more extreme rainfall events.

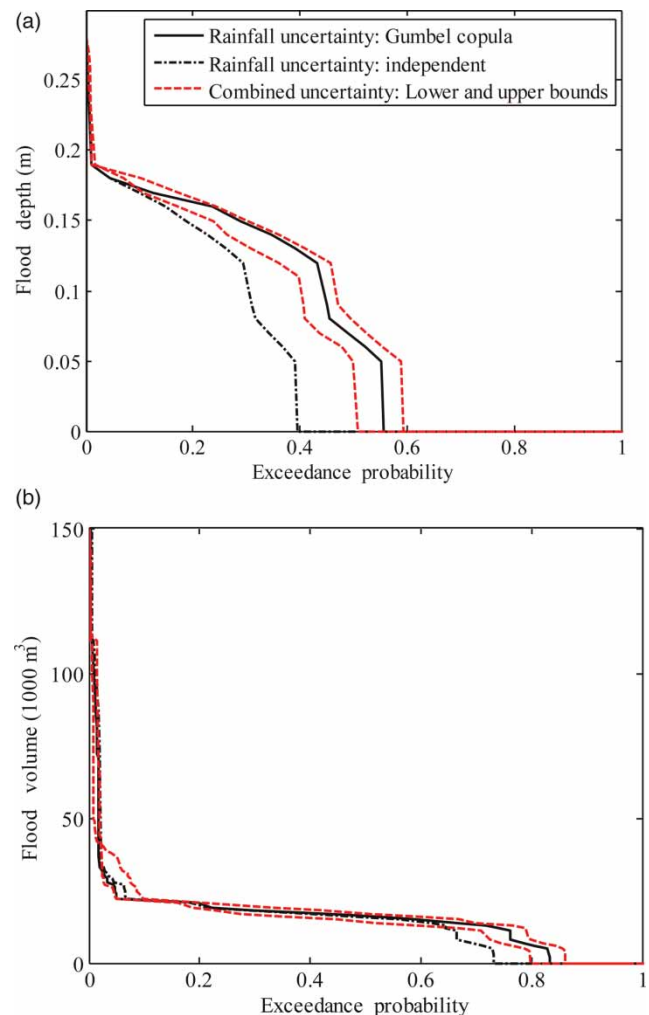
There is uncertainty in the estimation of the parameter  $\theta$  of Gumbel copula. The Maximum Likelihood method can provide an estimate of the related uncertainty. The 95% confidence interval of parameter  $\theta$  is used here to analyse the uncertainty of copula fitting on the model outputs – flood depth and volume. Figures 6(a) and 6(b) show the confidence bounds of exceedance probabilities, represented by the dotted lines. For both cases of flood depth and volume, the impact of copula-fitting uncertainty is insignificant when compared with the underestimation from the assumption of independence. This implies that it is more important to consider the dependence structure between rainfall depth and duration.

### Combined impact of rainfall and model parameter uncertainties

The uncertainty from the model parameters is considered as epistemic, and is combined with the stochastic uncertainty from rainfall to calculate the exceedance probabilities of flood depth in this case. The fuzzy sets for Manning roughness and percentage imperviousness shown in Figure 3 are sampled and combined with the samples from the Gumbel

copula using the Monte Carlo method described in the 'Methodology' section.

The constructed lower and upper exceedance probabilities of flood quantities are represented by dashed lines in Figure 7. According to the lower and upper exceedance probabilities of flood depth at node N126 (see Figure 7(a)), the probability of no flooding occurring (i.e. flood depth  $\leq 0$ ) at this node lies in the range [0.51 0.59]. In other words, when a rainfall event occurs, the probability of flooding at this node is from 0.41 to 0.49. Similarly, the probability bounds for any specific flood depth can also be derived. For example, the likelihood of flood depth greater than 0.15 m is confined to the range [0.24 0.31]. Similarly,



**Figure 7** | Lower and upper exceedance probabilities of flood quantities considering the combined uncertainty of rainfall variables and model parameters: (a) flood depth at node N126 and (b) total flood volume in the sewer system.

Figure 7(b) shows the lower and upper exceedance probabilities of total flood volume. The probability of no flooding occurring in the entire system lies in the range [0.15 0.20]. These probability gaps represent the belief interval given the substantial uncertainties in model parameters, and can be reduced only when more data or knowledge are available. For example, the lower and upper probability distributions are reduced to the single distribution represented by the solid line when the model parameters are known precisely.

Figure 7 shows the comparison of the impacts from model parameter uncertainty with those from the dependence of rainfall variables. It can be seen from Figure 7(a) that the CDF of flood depth derived from the assumption of independence is outside the probability bounds derived from the combined uncertainty of rainfall variables and model parameters. Similar results are also observed for total flood volume in Figure 7(b). This illustrates that the dependence structure between rainfall variables is more significant than the uncertainty of model parameters for flood frequency analysis in this case study.

The impact of model parameter uncertainty on the derived CDF of flood depth can also be compared with the impact from the uncertainty of copula fitting by comparing the lower and upper CDFs in Figure 7 with the 95% CDF bounds in Figure 6. In both cases of flood depth and volume, the bounds in Figure 6 are bracketed by the bounds in Figure 7. The comparison shows that the model parameter uncertainty is slightly more important than the uncertainty in parameter  $\theta$  of Gumbel copula, although both of them are significantly less important than the dependence structure of rainfall variables.

## Flood risk

Flood risk estimation is one of the important steps for flood management. Using the probabilities derived for flood quantities, flood risk can be estimated by combining the relevant costs related to the probabilities. The costs or consequences of flooding vary a great deal depending on the extent of flooding and value of the local environments and communities that are affected. With the aim of demonstrating the benefits of the copula-based framework, the flood costs in this study are presumably represented (i.e. surrogated) by flood depth or flood volume for the entire system: a unit

Table 2 | Comparison of flood risks

Cases	Node N126	Sewer network
Independent rainfall depth and duration	0.60	0.70
Correlated rainfall depth and duration	0.87	0.85
Combined uncertainty: lower bound	0.82	0.78
Combined uncertainty: upper bound	0.95	1.06

cost of 1.0 is assumed for a flood depth of 0.5 m at node N126 and a unit cost of 1.0 is assumed for a system level flood volume of 150,000 m<sup>3</sup>, and a linear relationship is assumed for other depth/volume levels assuming zero unit costs for zero flood depth/volume. The risk of flooding at a network node is then estimated as the product of probability of flooding at that node and the consequence expressed as aforementioned normalised flood depth. The risk of flooding at the system level is estimated as the product of probability of non-zero flood volume and the consequence expressed as aforementioned normalised flood volume, both evaluated at the system level.

Table 2 shows the calculated risks for three cases discussed above, i.e. independent rainfall depth and duration, correlated rainfall depth and duration and combined uncertainty of rainfall and model parameters. At both node N126 and system level, the independent case has a low risk compared with the correlated case. This confirms the underestimation of flood risk with the independence assumption in this case study. The lower and upper bounds of risks provide an estimation of the impact of model parameter uncertainties. It can be seen that the impact is less significant when compared with the independence assumption. It should be noted that the flood risks at node and system levels are not comparable due to the assumption of different unit costs. This assumption allows a good comparison between the three cases for node N126 whose risks are rather small should the same unit cost be assumed for flood volume and depth.

## CONCLUSIONS

This paper investigates the impact of the dependence structure between rainfall depth and duration on estimation of

flood quantities in a sewer network using copulas, and demonstrates how the copula method can be integrated in a unified framework to consider stochastic and epistemic uncertainties simultaneously. The method is tested on a case study of urban drainage systems for estimation of the exceedance probabilities of flood depth and volume.

The dependence structure between rainfall depth and duration for the rainfall events in the case study catchment can be represented well using the Gumbel family of Archimedean copulas with their marginal distributions described by the Generalised Extreme Value distributions. In turn, this provides a simple way to construct the joint probability distribution of rainfall depth and duration which is often ignored in practical and other applications. This is important because, as shown by the case study results obtained, the dependence structure of rainfall depth and duration (represented by Gumbel copulas) can have substantially more important impact on the estimation of flood depth/volume than the uncertainty in the SWMM model parameters.

Through integration of the copula-based probabilistic representation of random rainfall variables and fuzzy representation of model parameters, the uncertainty in flood quantities is represented in the form of lower and upper cumulative probabilities or exceedance probabilities. The gaps between the lower and upper probabilities arise from the epistemic uncertainty sources while the uncertainty represented by the distributions arises from the stochastic uncertainty sources. These cumulative probabilities can be used for flood risk estimation as demonstrated in the case study. The results obtained show that this form provides an effective way to investigate separate and combined impacts of uncertainties of different natures on the model outputs.

Although the copula-based uncertainty analysis framework is demonstrated using a case study of a sewer network, it can be easily applied to other uncertainty analysis applications where correlated random variables and fuzzy variables co-exist. Having said this, further tests on more complex problems are required to investigate the computational efficiency of uncertainty propagation and the implications of the uncertainty analysis framework on decision making.

## REFERENCES

- Alvarez, D. A. 2009 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**, 384–401.
- Bacchi, B., Becciu, G. & Kottegoda, N. T. 1994 Bivariate exponential model applied to intensities and durations of extreme rainfall. *Journal of Hydrology* **155** (1–2), 225–236.
- Balistracchi, M. & Bacchi, B. 2011 Modelling the statistical dependence of rainfall event variables through copula functions. *Hydrology and Earth System Sciences* **15**, 1959–1977.
- Bicik, J., Kapelan, Z., Makropoulos, C. & Savic, D. A. 2011 Pipe burst diagnostics using evidence theory. *Journal of Hydroinformatics* **13** (4), 596–608.
- Butler, D. & Davies, J. W. 2011 *Urban Drainage*. Spon Press, London, 3rd edn.
- Choi, K. S. & Ball, J. E. 2002 Parameter estimation for urban runoff modeling. *Urban Water* **4** (1), 31–41.
- Chowdhary, H., Escobar, L. & Singh, V. P. 2011 Identification of suitable copulas for bivariate frequency analysis of flood peaks and flood volumes. *Hydrology Research* **42** (2–3), 193–216.
- De Michele, C. & Salvadori, G. 2003 A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *Journal of Geophysical Research* **108** (D2), 4067–4077.
- Dempster, A. P. 1967 Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics* **38**, 325–339.
- Dubois, D. & Prade, H. 1991 Random sets and fuzzy interval analysis. *Fuzzy Sets Systems* **42**, 87–101.
- Fontanazza, C. M., Freni, G., La Loggia, G. & Notaro, V. 2011 Definition of synthetic rainfall events for urban flooding estimation: the integration of multivariate statistics and cluster analysis. *12th International Conference on Urban Drainage*, Porto Alegre, Brazil.
- Fu, G. & Butler, D. Copula-based frequency analysis of overflow and flooding in urban drainage systems. *Journal of Hydrology* in press.
- Fu, G. & Kapelan, Z. 2011 Fuzzy probabilistic design of water distribution networks. *Water Resour. Res.* **47**, W05538.
- Fu, G., Butler, D., Khu, S. T. & Sun, S. 2011 Imprecise probabilistic evaluation of sewer flooding in urban drainage systems using random set theory. *Water Resources Research* **47**, W02534.
- Fullerton, J. N. 2004 A simulated modelling approach for storm water flow optimisation. PhD thesis, University of Exeter, UK.
- Goel, N. K., Kurothe, R. S., Mathur, B. S. & Vogel, R. M. 2000 A derived flood frequency distribution for correlated rainfall intensity and duration. *Journal of Hydrology* **228** (1–2), 56–67.
- Hall, J. 2003 Handling uncertainty in the hydroinformatic process. *Journal of Hydroinformatics* **5**, 215–232.

- Hall, J., Fu, G. & Lawry, J. 2007 Imprecise probabilities of climate change: aggregation of fuzzy scenarios and model uncertainties. *Climatic Change* **81** (3–4), 265–281.
- Heneker, T. M., Lambert, M. F. & Kuczera, G. 2001 A point rainfall model for risk-based design. *Journal of Hydrology* **247** (1–2), 54–71.
- Hosseini, S. M. & Ghasemi, A. 2012 Hydraulic performance analysis of sewer systems with uncertain parameters. *Journal of Hydroinformatics* **14** (3), 682–696.
- Kao, S. C. & Govindaraju, R. S. 2007a A bivariate frequency analysis of extreme rainfall with implications for design. *Journal of Geophysical Research* **112**, D13119.
- Kao, S. C. & Govindaraju, R. S. 2007b Probabilistic structure of storm surface runoff considering the dependence between average intensity and storm duration of rainfall events. *Water Resources Research* **43**, W06410.
- Kendall, D. G. 1974 Foundations of a theory of random sets. In: *Stochastic Geometry* (E. F. Harding & D. G. Kendall, eds). Wiley, New York, pp. 322–376.
- McKay, M. D., Conover, W. J. & Beckman, R. J. 1979 A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239–245.
- Merz, B. & Thielen, A. H. 2005 Separating natural and epistemic uncertainty in flood frequency analysis. *Journal of Hydrology* **309** (1–4), 114–132.
- Nelsen, R. B. 2006 *An Introduction to Copulas*. Springer, New York.
- OFWAT 2002 Flooding from sewers, a way forward: consultation. Report 18/20, Office of Science and Technology, London, UK.
- POST 2007 POST note 289 – Urban flooding. Report of the Parliamentary Office of Science and Technology. Parliament UK, London.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M. & Franks, S. W. 2010 Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research* **46**, W05521.
- Revelli, R. & Ridolfi, L. 2002 Fuzzy approach for analysis of pipe networks. *Journal of Hydraulic Engineering* **128** (1), 93–101.
- Ross, J. L., Ozbek, M. M. & Pinder, G. F. 2009 Aleatoric and epistemic uncertainty in groundwater flow and transport simulation. *Water Resources Research* **45**, W00B15.
- Schmitt, T. G., Thomas, M. & Ettrich, N. 2004 Analysis and modeling of flooding in urban drainage systems. *Journal of Hydrology* **299** (3–4), 300–311.
- Shafer, G. 1976 *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ.
- Sklar, A. 1959 *Fonctions de repartition à n dimensions et leurs marges*. Publications de l'Institut de Statistique de l'Université de Paris 8, pp. 229–231.
- Sun, S., Fu, G., Djordjevic, S. & Khu, S. T. 2012 Separating aleatory and epistemic uncertainties: probabilistic sewer flooding evaluation using probability box. *Journal of Hydrology* **420–421**, 360–372.
- Thorndahl, S. & 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 and System Safety* **85**, 169–181.
- Vandenberghe, S., Verhoest, N. E. C. & De Baets, B. 2010 Fitting bivariate copulas to the dependence structure between storm characteristics: A detailed analysis based on 105 year 10 min rainfall. *Water Resources Research* **46**, W01512.
- Yue, S. 2000 The Gumbel mixed model applied to storm frequency analysis. *Water Resources Management* **14** (5), 377–389.
- Zadeh, L. 1965 Fuzzy sets. *Information and Control* **8**, 338–353.
- Zhang, L. & Singh, V. P. 2006 Bivariate flood frequency analysis using the copula method. *Journal of Hydrologic Engineering* **11** (2), 150–164.
- Zhang, L. & Singh, V. P. 2007 Bivariate rainfall frequency distributions using Archimedean copulas. *Journal of Hydrology* **332**, 93–109.

First received 22 March 2012; accepted in revised form 15 August 2012. Available online 16 November 2012