Copula-based frequency analysis of overflow and flooding in urban drainage systems

Guangtao Fu*, David Butler

Centre for Water Systems, College of Engineering, Mathematics and Physical Sciences,
University of Exeter, North Park Road, Harrison Building, Exeter EX4 4QF, UK
*Corresponding author. Email: g.fu@exeter.ac.uk; Phone: +44 (0)1392 723692; Fax: +44 (0)1392 217965

Abstract:

The performance evaluation of urban drainage systems is essentially based on accurate characterisation of rainfall events, where a particular challenge is development of the joint distributions of dependent rainfall variables such as duration and depth. In this study, the copula method is used to separate the dependence structure of rainfall variables from their marginal distributions. Three one-parameter Archimedean copulas, including Clayton, Gumbel, and Frank families, are fitted and compared for different combinations of marginal distributions that cannot be rejected by statistical tests. The fitted copulas are used, through the Monte Carlo simulation method, to generate synthetic rainfall events for system performance analysis in terms of sewer flooding and Combined Sewer Overflow (CSO) discharges. The copula method is demonstrated using an urban drainage system in the UK, and the cumulative probability distributions of maximum flood depth at critical nodes and CSO discharge volume are calculated. The results obtained in this study highlight the importance of taking into account the dependence structure of rainfall variables in the context of urban drainage system evaluation and also reveal the different

impacts of different dependence structures on the probabilities of sewer flooding and CSO volume.

Keywords: Bivariate distribution, combined sewer overflow, copula, frequency analysis, sewer flooding, urban drainage system

1. Introduction

Urban drainage systems are used in urban areas for flood and pollution control through collection and conveyance of stormwater and dry weather flow (DWF) to receiving waters and wastewater treatment plants. Most systems in the UK and many other countries are combined sewer systems, in which both DWF and stormwater flow in a single pipe network. Such combined sewer systems have two common issues: sewer flooding and combined sewer overflow (CSO) discharges when the flow exceeds the available system capacity (Butler and Davies, 2011). Their economic, social and environmental impacts have been discussed in detail in the literature (e.g., Schmitt et al., 2004; Fu et al., 2009; Andres-Domenech et al., 2010). Most sewer systems in the developed countries were constructed many decades ago and designed using simple deterministic methods on the basis of design rainfall events, which are usually related to a specified return period and generated from intensity-duration-frequency curves (Hvitved-Jacobsen and Yousef, 1988; Butler and Davies, 2011). System performance is affected by many factors that may have changed over time such as system characteristics, land use and climate change. Thus, there is a need to assess the performance of sewer systems regarding sewer flooding and CSO discharging in a changed situation (Korving et al., 2002; Schmitt et al., 2004; Thorndahl and Willems, 2008). This is also driven by strict regulations such as the Water Framework Directive in the member states of EU to improve receiving water quality through better utilization of the sewer system capacity.

Many different approaches have been developed for frequency analysis of sewer flooding or CSO discharges in an urban drainage system, for example, analytical probability methods (Benoist and Lijklema, 1989; Adams and Papa, 2000), Bayesian methods (Korving et al., 2002), first-order reliability methods (Thorndahl and Willems, 2008), and imprecise probability methods (Fu et al., 2011). In these methods, the historical rainfall series available are separated into rainfall events, and probability distributions of some rainfall variables are then used to characterise the stochastic nature of rainfall. For example, rainfall depth and duration are often used in the literature (e.g., Vandenberghe et al., 2010; Zegpi and Fernández, 2010; Fu et al., 2011).

In many cases, rainfall variables are related, however, due to the difficulty and complexity in generating the joint probability distributions of rainfall variables, the dependence structure between rainfall variables is not explicitly considered in many studies (e.g., Adams and Papa, 2000; Thorndahl and Willems, 2008; Andres-Domenech et al., 2010). Research has shown that the assumption of independence can have a significant effect on the frequency distributions of flood or CSO discharges and may lead to erroneous results (Benoist and Lijklema, 1989). Thus many efforts have been made to consider the correlation relationships between rainfall variables (Córdova and Rodríguez-Iturbe, 1985; Yue, 2000) and to analyse the implications for hydrologic design (Kao and Govindaraju, 2007b).

Most recently, there is increasing attention on the use of copulas as a flexible tool to quantify the dependence structure between correlated variables in the fields of hydrology and water engineering (e.g., De Michele and Salvadori, 2003; Kao and Govindaraju, 2007a; Zhang and Singh, 2007; Zegpi and Fernández, 2010; Vandenberghe et al., 2010 and 2011). The use of copulas enables to model the probabilistic dependence structure, independently of marginal distributions, and thus allows for multivariate random events to be described using different types of marginal distributions. This represents a significant advantage compared to conventional multivariate analysis as many variables from hydrological phenomena cannot be described using the same type of probability distributions. An important application of copulas is modelling the stochastic nature of rainfall and flood using historical data (Favre et al., 2004; Vandenberghe et al., 2010). Copulas also provide a convenient way to generate samples of correlated rainfall variables, thus they can be used for flood frequency analysis in conjunction with the Monte Carlo simulation method (e.g., Kao and Govindaraju, 2007a; Fontanazza et al., 2011).

The primary aim of this paper is to investigate the use of copulas for assessing the hydraulic performance of a combined sewer system in an urban catchment, which explicitly capture the dependence structure between rainfall depth and duration. The hydraulic performance of the sewer system is represented by the maximum water level over the ground surface (flood depth) at critical manholes and the volume of CSO discharges during a rainfall event. The latter can be used as a performance indicator for receiving water quality as long as its limitations are understood (Lau et al., 2002), although recent research suggests the performance of a sewer system can be better considered in the context of integrated urban wastewater systems (Rauch et al., 2002; Fu et al., 2008; Fu et al., 2009). In this study, the dependence between rainfall depth and duration is represented using the Archimedean copulas, and the Monte Carlo simulation method is then used to generate synthetic rainfall events for system performance analysis. The copula method is demonstrated using an urban drainage system in the UK, and the cumulative probability functions (CDF) of flood depth at one critical node and CSO overflow volume are calculated. The results show the suitability and flexibility of the Archimedean copulas in simulating the dependence of rainfall depth and duration, and the impacts of dependence structure on the performance of urban drainage systems.

2. Methodology

2.1. Concept of copulas

Copulas can be described as multivariate CDFs with standard uniform marginals and represents the dependence structure of random variables. For two random variables *X* and *Y*, their marginal cumulative distribution functions are represented by

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

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 $Y \leq y$. The bivariate CDF $H_{XY}(x, y)$ can be represented as

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

where $C(u, v)$ is called a copula and can be uniquely determined when u and v are continuous. Through Eq. (2), it is easy to see that the copula is actually a multivariate distribution function with uniform marginals (Nelsen, 2006). This provides two main advantages in determining $H_{XY}(x, y)$: (1) the marginals can be determined using different distributions, and (2) the dependence structure can be described separately from the marginals, which allows for building complex multivariate distributions to model stochastic phenomena such as rainfall without the knowledge of marginal distributions.

There are many families of copulas that represent different dependence structures. The one-parameter Archimedean copulas are of special interest for hydrologic analyses, and the general expression of Archimedean copulas can be written as

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

where φ, called a generator, is a convex decreasing function defined in [0, 1], satisfying $\varphi(1) = 0$ and $\lim_{t \to 0} \varphi(t) = \infty$. Using different forms of the function φ, different families of Archimedean copulas can be generated, for example, the Gumbel, Frank and Clayton families. These copulas can describe a wide range of dependence level, from negative to positive, and have been used to describe the rainfall characteristics in previous studies (e.g., Kao and Govindaraju, 2007b; Zhang and Singh, 2007) and thus are selected to describe the relationship between rainfall depth and duration for the case study catchment in this study.

Only recently, use of copulas in hydrology has gained substantial attention with an intent of describing the probabilistic structures of random variables such as rainfall and flood. The work by De Michele and Salvadori (2003) was perhaps the first application in hydrology, and they used the Frank family of Archimedean copulas to describe the dependence between rainfall intensity and duration. Favre et al. (2004) used the Archimedean copulas to analyse the joint distribution of flood peak flow and volume in two Canadian river catchments. Zhang and Singh (2007) compared several different Archimedean copulas including Gumbel and Frank families to simulate the joint distributions between rainfall intensity, depth and duration. Vandenberghe et al. (2010) applied a number of copulas to investigate the dependence structure of storm variables on the basis of a 105-year rainfall series. There are few applications to urban drainage systems with one exception of the work by Fontanazza et al. (2011), which applied the copula approach to generate synthetic rainfall events for urban flood estimation but focused on analysing the impacts of hyetographs. The work described in this paper will look at the impacts of different copulas on the frequency of sewer flooding and CSO overflow in the urban drainage system.

2.2. Copula fitting

For Archimedean copulas, the simplest method to estimate the parameter θ is through a concordance measurement - Kendall's τ - which is a rank correlation coefficient, defined to measure the orderings of two measured quantities. Kendall's τ is defined in the interval [-1, 1], where 1 represents total concordance, -1 represents total discordance, and 0 represents zero concordance. According to the work by Nelsen (2006), the relationship between parameter θ and Kendall's τ can be determined for the three Archimedean families.

Particularly, a closed-form expression can be derived for Clayton and Gumbel families.

In addition to the non-parametric method describe above, there are some parametric methods available for parameter estimation, such as the conventional Maximum Likelihood (ML) method, Inference Function for Margins (IFM) method (Joe, 1997) and Canonical Maximum Likelihood (CML) method (Genest et al., 1995), and Minimum Distance Methods. For more information, the reader is referred to the following studies (e.g., Genest and Favre, 2007; Chowdhary et al., 2011; Nazemi and Elshorbagy, 2012). The IFM method was used in this study as it has a better performance compared with others according to our preliminary tests. More importantly, it allows to explore the impacts of the choice of parametrically estimated marginal CDFs on copula fitting as prior research has shown that a number of marginal CDFs may not be rejected for rainfall variables under several statistical tests (Fu et al., 2005). The root mean square error is a good indicator of goodness of fit, and can be calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\{C(u_i, v_i) - C_n(u_i, v_i)\right\}^2} \tag{4}$$

where C_n is the empirical copula. For this measure, the smaller the values, the better the copula fits the data. Formal hypothetical tests are increasingly used to evaluate the goodness-of-fit of different copulas (e.g., Berg, 2009; Genest et al., 2009; Nazemi and Elshorbagy, 2012). The

Cramér-von Mises statistic is chosen to compare an estimated copula C with the empirical copula C_n:

$$T_n = n \int_{[0,1]^2} \{C(u,v) - C_n(u,v)\}^2 \, dC(u,v) = \sum_{i=1}^{n} \{C(u_i, v_i) - C_n(u_i, v_i)\}^2 \qquad (5)$$

The p-values for T_n are approximated using the bootstrapping method described by Berg (2009) and Genest et al. (2009). 10,000 random samples are used in this study according to Genest and Favre (2007). Higher p-values are desired as they represent higher suitability of the chosen copulas.

In addition to the statistics, graphical methods can be used to verify the appropriateness of a fitted copula by comparison to the empirical distribution. The empirical distribution can be estimated using a nonparametric (empirical) approach (e.g., Genest and Rivest, 1993; Kao and Govindaraju, 2007a; Zhang and Singh, 2007): (1) assume an intermediate random variable z whose samples can be transformed from the n observations

$$z_i = \frac{1}{n} \{\text{number of } (x_j, y_j) \text{ such that } x_j < x_i \text{ and } y_j < y_i\},\ i = 1, 2, \ldots, n \qquad (6)$$

(2) estimate the empirical distribution K_n using

$$K_n(t) = \frac{1}{n} \{\text{number of } z_i \text{ such that } z_i < t\},\ i = 1, 2, \ldots, n \qquad (7)$$

The theoretical distribution of Archimedean copulas, i.e., $K(t) = P[C(u,v) \le t]$, can be derived using the generating function

$$K(t) = t - \frac{\varphi(t)}{\varphi'(t)}, \quad 0 < t \le 1 \qquad (8)$$

The distribution functions $K(t)$ and $K_n(t)$ transform the two dimension into one dimension, allowing for visual comparison of the empirical and theoretical copulas. The distribution plots or

Q-Q plots can be plotted for examination. This can provide a good indication if the theoretical copula fits the data well.

Tail dependence analysis is critical to investigate the magnitude of dependence in the upper and lower tails of a bivariate distribution (e.g., Poulin et al., 2007). It also helps identify the most suitable copula by emphasising on the joint occurrence of extreme values (Nazemi and Elshorbagy, 2012). The tail dependence can be represented by a coefficient. For the Gumbel copula, the upper tail dependence coefficient is

$$\lambda_U = 2 - 2^{1/\theta} \tag{9}$$

Amongst several estimators of the coefficient, the following estimator was proposed by Frahm et al. (2005):

$$\hat{\lambda}_U = 2 - 2\exp\left[\frac{1}{n}\sum_{i=1}^{n}\log\left(\sqrt{\log\frac{1}{U_i}\log\frac{1}{V_i}} \Big/ \log\frac{1}{\max(U_i, V_i)^2}\right)\right] \tag{10}$$

where (U_i, V_i) (i=1, 2, …, n) are random samples generated from a copula. This estimator has an advantage that no parameters (threshold values) are required for calculation and thus are used in this study.

2.3. Monte Carlo sampling

On the basis of the copula method, Monte Carlo simulations can be used to generate samples for correlated random variables. A general procedure is to generate correlated pairs (u, v) using a copula and then transform them to real values based on the inverse marginal CDFs. This is normally conducted using the following steps (Kao and Govindaraju, 2007a): (1) generate the independently uniformly distributed random pairs (u, t); (2) solve v through a simplified expression with copulas $P[V \le v | U = u] = \frac{\partial C(u, v)}{\partial u} = t$, and thus obtain correlated pairs (u, v); (3)

assume $F_X(x)=u$ and solve x using the inverse function of *X*, and similarly assume $F_Y(y)=v$ and solve y. In this way, the correlated random samples (x,y) can be generated from their margins (u,v).

3. Case study

3.1. The catchment

An urban drainage system in a Scotland town is used to demonstrate the copula methodology developed in this paper. The total catchment area is about 200 hectares, serving a population of 4,000. A combined sewer system provides the infrastructure for draining the catchment runoff and routing the wastewater from the town to the treatment facilities. The sewer system consists of 265 nodes, 265 pipes, 2 outfalls and 1 weir, and has a total conduit length of 22,482 metres. 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 and the other to a combined sewer overflow, and both flows are eventually discharged into a river. Fig. 1 shows a satellite image of the study area and provides the layout of the sewer system described.

The storm water management model (SWMM), developed by the U.S. 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 system model has been well calibrated for the purpose of flood evaluation in the work by Fullerton (2004), and has been used for system design and uncertainty analysis (Fu et al., 2011; Sun et al., in revision).

3.2. Rainfall data

A 10-year time series of 5-minute rainfall from one rain gauge station is used in this study. To analyse statistical characteristics of the actual rainfall events in the case study of urban catchment, independent rainfall events are separated from the time series using the concept of inter-event time (IETD) definition, i.e., the time interval between two consecutive events should be no less than a pre-determined IETD. According to the maximum concentration time of the catchment, IETD was set to 20 minutes such that the runoff response from an individual event is not affected by any other. A total of 3405 events were identified from the rainfall series. As the events with a total amount of 3mm cannot generate sewer flooding and CSO discharges, so they are not considered for analysis in this study, and the number of events is reduced to 570, with an average of 57 rainfall events per year.

The general characteristics of rainfall events are described by rainfall depth and duration. Fig. 2 shows the scatter plots of the 570 events with marginal histograms of rainfall depth and duration. There is a high frequency of low rainfall depth, about 50% of the rainfall events have a very small depth less than 6 mm. The average rainfall depth is 7.4 mm, but the maximum is up to 42 mm. Similarly, most events last a short period, although about 10 % have a duration over 800 minutes. The average rainfall duration over the 570 events is 399 minutes with a maximum of 1725 minutes. The two variables are related to some extent, with a Kendall's tau = 0.27. As can be seen from the marginal histograms, the two variables follow a rather different marginal distribution. This clearly demonstrates the need to separate the marginal distributions and dependence structure in the joint distribution of the two rainfall variables so that the marginal distributions can be simulated by different types of distribution.

4. Results and discussion

4.1. Marginal distributions

The following commonly used distributions are used to fit the rainfall depth and duration data: Generalized Pareto (GP), Generalized Extreme Value (GEV), Log-Logistic (Log-log) and Gamma, according to previous studies (e.g., Kao and Govindaraju, 2007b; Vandenberghe et al., 2011). The above functions are fitted using the maximum likelihood estimation method that maximises the log-likelihood function. In the calculation, the maximum number of iterations is specified as 100 and the accuracy of the estimation is set to 1.0×10^{5}.

Three goodness-of-fit tests, i.e., Kolmogorov Smirnov (K-S), Anderson Darling (A-D) and Chi-Square (χ^2) tests, are used to determine if the data follow one of the specified distributions (the null hypothesis H_0). The hypothesis is evaluated at the 5% significance level. The critical values at this level are 0.056 for K-S, 16.919 for χ^2, and 2.502 for A-D, respectively. Smaller statistic values indicate better fit to the data. The hypothesis regarding a specific distribution is rejected at the significance level if the test statistic (K_n, A_n^2, or χ_n^2) is greater than the relevant critical value given above. Appendix A provides the CDFs and goodness-of-fit tests used.

The statistics of K-S (K_n), A-D (A_n^2) and Chi-Square (χ_n^2) for rainfall duration and depth are provided in Tables 1 and 2, respectively. The *p*-values measure the amount of information that is against the null hypothesis H_0, and are also provided here. The smaller the p-values, the more evidence we have against H_0. For rainfall duration, the three distributions, i.e., GEV, Log-log and Gamma, cannot be rejected with all of the three tests and have a decreasing ranking according to the statistics values. Similarly, for rainfall depth, the distribution GP best fits to the data, followed by Log-log and GEV. All of the three distributions cannot be rejected with all of the three tests. According to the A-D test, Gamma and GP are rejected for rainfall duration and depth, respectively. The A-D test is stricter than the K-S test possibly because it gives more weight to the distribution tails. The results confirm that in many cases it is not possible to determine one single best distribution particularly when a relatively short series of data is available (Korving et al., 2002; Kao and Govindaraju, 2007b; Fu et al., 2011). This study considers all the distributions that cannot be rejected to investigate the bivariate distribution using copulas.

4.2. Dependence structure

The selected marginal distributions for rainfall duration and depth are used to fit the Archimedean copulas using the CML method. The parametrically estimated values of parameter θ are provided in Table 3, along with their 95% confidence intervals. Recall that parameter θ can also be derived according to the relationships between θ and τ , and the values for Gumbel, Frank and Clayton copulas are 1.371, 2.589 and 0.741, respectively. It can be seen that the parametric estimates for Gumbel and Frank copulas are in a good agreement with those from the non-parametric method, and are bracketed in the relevant 95% confidence intervals. For the Clayton copula, however, the non-parametric estimate is significantly bigger than the parametric estimates, and is out of the 95% confidence intervals. This is possibly because the rainfall data as shown in Fig. 2 illustrate greater dependence in the upper tail than in the lower tail. On the contrary, Clayton copula exhibits greater dependence in the lower tail than in the upper tail. The Gumbel copula is an asymmetric Archimedean copula with greater dependence in the upper tail than in the lower. The Frank copula is a symmetric Archimedean copula. Thus these two copulas are more appropriate for describing the dependence structure between rainfall duration and depth.

Table 4 shows the resulting RMSE and Cramér-von Mises statistic values. The statistic confirms the inappropriateness of the Clayton copula as it has a lower *p*-value for most the marginal combinations. This statistic shows a more significant difference compared with RMSE. Amongst the three marginal distributions of rainfall depth, GP has the worst performance in terms of copula fitting although this distribution is the best in the marginal distribution fitting according to the statistics. This implies that it is important to consider the goodness-of-fit of both marginal distributions and copulas in order to achieve the best overall performance in constructing a joint distribution of multi-variables. For the three distributions of rainfall duration, there is no significant difference in copula fitting. It can be seen that the Gumbel and Frank copulas are in good agreement.

Table 5 shows the upper tail dependence coefficients for Gumbel copulas. It can be seen that the estimated coefficient values are very close to the theoretical ones. The estimator proposed by Frahm et al. (2005) has a high accuracy. More importantly, this implies that the choice of marginal distributions has no impacts on tail dependence, which is mainly controlled by copula selection as expected.

Fig. 3 shows the Q-Q plots of Gumbel and empirical copulas for different marginal distribution combinations. The *x*-axis represents the cumulative probability of empirical copula and *y*-axis represents that of Gumbel copula. The diagonal straight line represents a perfect match between the parametrically estimated copula and empirical copula. Generally the Q-Q plots confirm the results revealed from the statistic values in Table 4. That is, the GP vs. GEV and GP vs. Gamma pairs provide the worst copula fitting results, while all the other distribution combinations provide a rather good fitting. The pair Log-log vs. GEV is chosen as the base case to analyse the frequency of flooding and CSO discharges and compare the impacts of marginal distributions and copulas.

To understand the structure of dependence, Fig. 4 visualizes the CDF and probability distribution function (PDF) of the Gumbel copula on the basis of the Log-log vs. GEV combination. The variables *u* and *v* represent the transformed random variables *X* and *Y* (rainfall depth and duration) in the unit hypercube, respectively, and have the same ranks as *X* and *Y*. Fig. 4a shows the fitted copula (shaded surface) together with the empirical copula (points). The strong dependence in the upper tail is clearly illustrated in Fig. 4b.

Selection of the most suitable copula is a complex process and need to consider several different measures including statistics, graphical approaches, and comparison to empirical copulas and data regarding dependence types. A single measure may fail to identify the inappropriate copulas, leading to an overestimate or underestimate of the probability of sewer flood and CSO discharges as demonstrated below.

4.3. Flood and overflow frequency

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 is very conservative compared to the previous study by Fu et al. (2011) and can provide very stable simulation results. The 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 the flood depth at different nodes. We recognise the impact of different rainfall profiles on the frequency of flood and overflow (Fontanazza et al., 2011; Sun et al., 2012), but this is out of scope of this study.

Fig. 5 shows the cumulative probabilities of flood depth at one critical node N126 and overflow volume at the CSO. For the copula results, the dependence structure is represented by the Gumbel copula and marginal distributions for rainfall depth and duration are represented by Log-Log and GEV, respectively. In Fig. 5a, the probability of no flooding occurring (flood depth=0) has a value of 0.43. In other words, the probability of flooding at this node is 0.57 and this is equivalent to the probability of system ‘failure’ in terms of sewer flooding. Note that this probability represents the probability for each rainfall event because the way of rainfall events is simulated in this study. This high number of ‘failures’ at this critical node is caused by the expansion of the network to the (left) upstream due to urban development (Fullerton, 2004). Similarly in Fig. 5b, the probability of no CSO discharges (CSO volume=0) is estimated at 0.92.

For comparison, Fig. 5 also shows the results when rainfall depth and duration are assumed as independent. In this case, the probability of sewer flooding, having a value of 0.47, is lower than in the case of correlation. This implies that the flood depth is under-estimated without considering the correlation between rainfall depth and duration. This under-estimation has a more significant impact on flood depth when compared with the uncertainties in the fitted copula parameter, as demonstrated with the 95% confidence intervals as shown in Fig. 5. For CSO volume, the differences between correlation and independence are also significant, but mainly lie in the regions of high cumulative probabilities. This is because the probability of CSO discharge is much lower than sewer flooding, that is, CSO discharges can only results from more ‘extreme’ rainfall events.

4.4. Impacts of copulas and marginal distributions

Recall that different marginal distributions for rainfall variables cannot be rejected with the statistical tests in this case study. To investigate the impacts of different marginal distributions,

Fig. 6 shows the CDFs of flood depth and CSO volume from three marginal distributions of rainfall duration, i.e., GEV, Log-log and Gamma, combined with the Log-log distribution of rainfall depth. For flood depth, the three CDFs are roughly the same, which implies that the impacts of the different marginals are negligible when compared with other uncertainties such as the copula parameter estimation, as shown in Fig. 5. For CSO volume, similarly the impacts of different marginals are small, although there are some differences in the upper tails, reflecting the importance of distribution tails in rainfall event simulation.

Different copulas were compared for the CDFs of flood depth and CSO volume and results are shown in Fig. 7. Note that according to the copula fitting results the Clayton copula is not appropriate to describe the dependence structure of rainfall depth and duration, but it is used here for the purpose of demonstration of its potential impacts. The Clayton copula overestimates the probability of sewer flooding and CSO discharge, i.e., the performance of the sewer system. This can be explained by the dependence structure of the Clayton copula: greater dependence in the lower tail than in the upper tail, which results in more small synthetic rainfall events. Conversely, the cumulative probabilities of sewer flooding and CSO discharges estimated by the Gumbel copula are smaller than those from Clayton and Frank copulas, because more extreme events are generated as a result of the greater dependence in the upper tails.

Clearly the copulas have more significant impacts on the CDFs of sewer flooding and CSO volume than the marginal distributions, when comparing the results in Fig. 5 and Fig. 6. This implies the importance of considering the dependence structure of rainfall variables when evaluating the system performance of urban drainage systems through synthetic events based methods.

The impact of dependence structure of rainfall variables on system performance can be illustrated by calculating the return periods of the sewer system. Fig. 8 shows the return periods of CSO volumes from the three copulas. The return periods of CSO volumes are derived using the cumulative probabilities in Fig. 7b, considering the average 57 rainfall events per year.

5. Conclusions

This paper highlights the importance of considering the dependence structure of rainfall variables in the context of system performance of urban drainage systems using copulas. The copula method is demonstrated using an urban drainage system in the UK to calculate the cumulative probability distributions of flood depth and CSO volume. The rainfall characteristics in the case study are represented by two variables: rainfall depth and duration. The marginal distributions of these variables are simulated using GP, GEV, Log-log and Gamma, and the dependence structure is represented by the following three one-parameter Archimedean copula families: Gumbel, Frank, and Clayton. On the basis of the copula approach, the Monte Carlo simulation is used to generate synthetic rainfall events to evaluate the probabilistic system performance, represented by the CDFs of flood depth and CSO volume. This new methodology is promising in that it provides a simpler way to construct the joint distribution for rainfall variables by separating the dependence from their marginal distributions, and thus provides a basis for performance evaluation of urban drainage systems. The following conclusions are presented on the basis of this study:

1. It is necessary to consider all the marginal distributions that cannot be rejected by statistical tests for copula fitting using the IFM method, rather than choose the best ranked distributions. As revealed by the case of bivariate copulas, the pair of the best

fitted marginal distributions of rainfall depth and duration cannot produce the best overall performance in constructing the joint distribution of rainfall depth and duration.

2. Copula identification should be based on several different measures including statistics, graphical approaches, and comparison to empirical copulas and data regarding dependence types. A single measure may fail to identify the inappropriate copulas, leading to an overestimate or underestimate of the probability of sewer flood and CSO discharges.
3. The results obtained show the copulas, i.e., the dependence structures, have more significant impacts on the CDFs of sewer flooding and CSO volume than the marginal distributions. Different copulas affect different parts of the CDFs of sewer flooding and CSO volume, i.e., those with higher or lower return periods.

The copula method has the flexibility and advantage in building complex, bivariate probability distributions of rainfall depth and duration for system performance analysis. The results provide a more accurate probabilistic evaluation of sewer flooding and CSO discharges based on the characterization of the dependence structure of rainfall depth and duration. This provides crucial information for more accurate estimation of design storms and the associated risks.

Appendix A: Cumulative distribution functions and test statistics

The marginal distribution functions used in this paper are given in equations (A.1)-(A.4).

Generalized Extreme Value Distribution (GEV)

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

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

Generalized Pareto Distribution (GP)

$$F(x) = 1 - \left[1 + \xi(x - \mu)/\sigma\right]^{-1/\xi} \tag{A.2}$$

where ξ≠0, σ>0 and μ are shape, scale and location parameters, respectively.

Log-Logistic Distribution (Log-log)

$$F(x) = \left[1 + (\beta / x)^{\alpha}\right]^{-1} \tag{A.3}$$

where α>0 and β>0 are parameters.

Gamma Distribution

$$F(x) = \Gamma_{(x-\gamma)/\beta}(\alpha) / \Gamma(\alpha) \tag{A.4}$$

where α>0, β>0 and γ are shape, scale and location parameters, respectively. Γ is the Gamma function

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt \qquad \text{for } (\alpha>0) \tag{A.5}$$

and Γ_z is the incomplete Gamma function

$$\Gamma_z(\alpha) = \int_0^{z} t^{\alpha-1} e^{-t} dt \qquad \text{for } (\alpha>0) \tag{A.6}$$

The goodness of fit for the above distributions is considered using three different test statistics: Kolmogorov-Smirnov K_n, Anderson-Darling A_n^2 and Chi-Square χ_n^2:

$$K_n = \max_{1 \le i \le n}\left(F(x_i) - \frac{i-1}{n}, \frac{i}{n} - F(x_i) \right) \tag{A.7}$$

$$A_n^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln F(x_i) + \ln(1 - F(x_{n-i+1}))\right] \tag{A.8}$$

$$\chi_n^2 = \sum_{i=1}^{k}\frac{(O_i - E_i)^2}{E_i} \tag{A.9}$$

where n is the number of data, x_i is the ith sample (i=1, 2, ... n), and F is the cumulative distribution being texted. As the Chi-Square test is based on binned data, the total number of bins k is determined using the following empirical equation:

$$k = 1 + \log_2 n \tag{A.10}$$

O_i is the observed frequency for bin i and E_i is the expected frequency for bin i calculated by

$$E_i = F(x_2) - F(x_1) \tag{A.11}$$

where x_1 and x_2 are the lower and upper limits for bin i.

References

Adams, B.J., Papa, F., 2000. *Urban Stormwater Management Planning With Analytical Probabilistic Models*. 358 pp., John Wiley & Sons, West Sussex, U.K.

Andres-Domenech, I., Munera, J.C., Frances, F., Marco, J.B., 2010. Coupling urban event-based and catchment continuous modelling for combined sewer overflow river impact assessment. Hydrol. Earth Syst. Sci. 14 (10), 2057-2072.

Benoist, A.P. and Lijklema, L., 1989. A methodology for the assessment of frequency distributions of combined sewer overflow volumes. Water Res. 23(4), 487-493.

Berg, D. 2009. Copula goodness-of-fit testing: an overview and power comparison. Eur J Finance 15(7 & 8): 675-701.

Butler, D., Davies, J.W., 2011. *Urban Drainage*, 3rd ed. Spon Press, London.

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.

Córdova, J. R., Rodríguez-Iturbe, I., 1985. On the probabilistic structure of storm surface runoff. Water Resour. Res. 21 (5), 755– 763.

De Michele, C., and Salvadori, G., 2003. A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. J. Geophys. Res. 108(D2), doi:10.1029/2002JD002534.

Favre, A., El Adlouni, S., Perreault, L., Thiémonge, N., Bobée, B., 2004. Multivariate hydrological frequency analysis using copulas. Water Resour. Res. 40, W01101, doi:10.1029/2003WR002456.

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.

Frahm, G., Junker, M., and Schmidt, R., 2005. Estimating the tail dependence coefficient: Properties and pitfalls. Insur. Math. Econ., 37(1), 80-100.

Fu, G., Butler, D., Khu, S.T., 2008. Multiple objective optimal control of integrated urban wastewater systems. Environ. Model. Softw. 23, 225-234.

Fu, G., Butler, D., Khu, S.T., 2009. The impact of new developments on river water quality from an integrated system modelling perspective. Sci. Total Environ. 407(4), 1257-1267.

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 Resour. Res. W02534, 47, doi:10.1029/2009WR008944.

Fullerton, J.N., 2004. *A simulated modelling approach for storm water flow optimisation*, PhD thesis, Centre for Water Systems, University of Exeter, UK.

Genest, C., Favre A.C., 2007. Everything you always wanted to know about copula modelling but were afraid to ask. J Hydrol Eng 12(4), 347-368.

Genest, C., K. Ghoudi, and L. P. Rivest, 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika 82(3), 543- 552.

Genest, C., Rémillard, B., and Beaudoin, D., 2009. Goodness-of-fit tests for copulas: a review and a power study. Insurance: Mathematics and Economics 44, 199-213.

Genest, C., Rivest, L., 1993. Statistical inference procedures for bivariate Archimedean copulas. Journal of the American Statistical Association, 88 (424), 1034-1043.

Hvitved-Jacobsen, T., Yousef, Y.A., 1988. Analysis of rainfall series in the design of urban drainage control systems. Water Res. 22 (4), 491-496.

Joe, H., 1997. *Multivariate models and dependence concepts*. Chapman and Hall, London.

Kao, S.C., Govindaraju, R.S., 2007a. Probabilistic structure of storm surface runoff considering the dependence between average intensity and storm duration of rainfall events. Water Resour. Res. 43, doi:10.1029/2006WR005564.

Kao, S.C., Govindaraju, R.S., 2007b. A bivariate frequency analysis of extreme rainfall with implications for design. J. Geophys. Res., 112, D13119, doi:10.1029/2007JD008522.

Korving, H., Clemens, F., van Noortwijk, J., van Gelder, P., 2002. Bayesian estimation of return periods of CSO volumes for decision-making in sewer system management. In: Proc. of 9th International Conference on Urban Drainage, Portland, Oregon, USA.

Lau, J., Butler, D., Schütze, M., 2002. Is combined sewer overflow spill frequency/volume a good indicator of receiving water quality impact? Urban Water 4 (2), 181-189.

Nelsen, R.B., 2006. *An Introduction to Copulas*, Springer, New York.

Nazemi, A., Elshorbagy, A., 2012. Application of copula modelling to the performance assessment of reconstructed watersheds. Stochastic Environmental Research & Risk Assessment 26(2), 189-205.

Poulin, A., Huard, D., Favre, A.C., Pugin, S., 2007. Importance of tail dependence in bivariate frequency analysis. J. Hydrol. Eng. 12(4), 394-403.

Rauch, W., Bertrand-Krajewski, J.L., Krebs, P., Mark, O., Schilling, W., Schütze, M., Vanrolleghem, P.A., 2002. Mathematical modelling of integrated urban drainage systems. Water Sci. Technol. 45(3), 81-94.

Schmitt, T.G., Thomas, M., Ettrich, N., 2004. Analysis and modeling of flooding in urban drainage systems. J. Hydrol. 299 (3-4), 300-311.

Sun, S., Fu, G., Djordjevic, S., Khu, S.T., 2012. Separating aleatory and epistemic uncertainties: probabilistic sewer flooding evaluation using probability box. J. Hydrol. 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 Res. 42(1-2), 455-466.

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 Resour. Res. 46, W01512, doi:10.1029/2009WR007857.

Vandenberghe, S., Verhoest, N.E.C., Onof, C., De Baets, B., 2011. A comparative copula-based bivariate frequency analysis of observed and simulated storm events: A case study on Bartlett-Lewis modeled rainfall. Water Resour. Res., 47, doi:10.1029/2009WR008388.

Yue, S., 2000. The Gumbel mixed model applied to storm frequency analysis. Water Resources Management 14(5), 377-389.

Zegpi, M., Fernández, B., 2010. Hydrological model for urban catchments – analytical development using copulas and numerical solution. Hydrol. Sci. J. 55(7), 1123-1136.

Zhang, L., Singh, V.P., 2007. Bivariate rainfall frequency distributions using Archimedean copulas. J. Hydrol. 332, 93-109.

Figure captions

Fig. 1 - Layout of the case study network.

Fig. 2 - Scatter plot of rainfall depth and duration with marginal histograms.

Fig. 3 - Q-Q plots of Gumbel and empirical copulas for different marginal distribution combinations.

Fig. 4 - Three dimensional plots for the theoretically fitted Gumbel copula.

Fig. 5 - Cumulative probabilities of flood depth and CSO volume from Gumbel copula.

Fig. 6 – Impacts of different marginal distributions on the cumulative probabilities of flood depth and CSO volume.

Fig. 7 – Comparison of Gumbel and Frank copulas.

Fig. 8 – Return periods of CSO discharge volumes.

Table 1 - Test Statistics of the fitted CDFs for rainfall duration*.

Distribution	Distribution Parameter	K-S		χ^2		A-D
		K_n	p-value	χ_n^2	p-value	A_n^2
GEV	ξ=0.105, σ=189.12, μ=267.9	0.026	0.832	3.126	0.959	0.433
Gamma	α=2.0653, β=193.05, γ=0	0.034	0.509	8.421	0.492	0.423
Log-Log	α=2.872, β=386.08; γ=-54.80	0.036	0.427	14.177	0.116	0.975
GP**	ξ=-0.229, σ=399.22, μ=73.82	0.055	0.063	-	-	107.39

*The CDFs are provided in Appendix A.

**This distribution is rejected at the significant level of 5% with the A-D test.

Table 2 - Test statistics of the fitted CDFs for rainfall depth[*].

Distribution	Distribution Parameter	K-S		χ^2		A-D
		K_n	p-value	χ_n^2	p-value	A_n^2
GP	ξ=0.190, σ=3.823, μ=1.941	0.020	0.910	5.393	0.80	0.496
Log-Log	α=1.375, β=2.413, γ=2.976	0.046	0.176	14.894	0.094	1.804
GEV	ξ=0.680, σ=1.746, μ=4.594	0.052	0.091	16.226	0.062	2.369
Gamma[**]	α=5.813, β=5.263, γ=3.0	0.052	0.091	-	-	5.813

[*]The CDFs are provided in Appendix A.

[**]This distribution is rejected at the significant level of 5% with the A-D test.

Table 3 - Estimated parameters of copulas and their 95% confidence intervals.

Depth	Duration	Gumbel		Frank		Clayton	
GEV	GEV	1.375	[1.285 1.466]	2.472	[1.971 2.974]	0.443	[0.317 0.570]
	Log-log	1.399	[1.309 1.489]	2.412	[1.916 2.909]	0.393	[0.273 0.513]
	Gamma	1.372	[1.280 1.465]	2.546	[2.032 3.059]	0.413	[0.288 0.539]
Log-log	GEV	1.377	[1.280 1.465]	2.509	[1.997 3.021]	0.375	[0.257 0.494]
	Log-log	1.403	[1.284 1.470]	2.447	[1.940 2.953]	0.340	[0.227 0.453]
	Gamma	1.375	[1.311 1.496]	2.591	[2.067 3.114]	0.354	[0.237 0.471]
GP	GEV	1.382	[1.281 1.470]	2.845	[2.290 3.400]	0.507	[0.344 0.671]
	Log-log	1.403	[1.288 1.476]	2.447	[1.940 2.953]	0.340	[0.227 0.453]
	Gamma	1.385	[1.311 1.496]	3.063	[2.502 3.623]	0.473	[0.323 0.623]

Table 4 - Goodness-of-fit of copulas for different combinations of marginal distributions.

Depth	Duration	Gumbel			Frank			Clayton		
		RMSE	T_n	p-value	RMSE	T_n	p -value	RMSE	T_n	p -value
GEV	GEV	0.016	0.139	0.372	0.014	0.110	0.534	0.016	0.146	0.388
	Log-log	0.018	0.184	0.239	0.015	0.130	0.431	0.016	0.142	0.394
	Gamma	0.017	0.173	0.268	0.019	0.196	0.229	0.020	0.237	0.162
Log-log	GEV	0.012	0.078	0.726	0.013	0.095	0.622	0.020	0.240	0.152
	Log-log	0.014	0.114	0.491	0.014	0.118	0.488	0.021	0.241	0.150
	Gamma	0.015	0.128	0.427	0.018	0.193	0.234	0.025	0.345	0.070
GP	GEV	0.065	2.397	0	0.070	2.816	0	0.060	2.030	0
	Log-log	0.014	0.114	0.488	0.014	0.118	0.488	0.021	0.241	0.156
	Gamma	0.066	2.521	0	0.075	3.223	0	0.060	2.083	0

Table 5 – The upper tail dependence coefficients for Gumbel copulas.

Depth	GEV			Log-log			GP		
Duration	GEV	Log-log	Gamma	GEV	Log-log	Gamma	GEV	Log-log	Gamma
λ_U	0.344	0.359	0.343	0.346	0.361	0.344	0.349	0.361	0.351
$\hat{\lambda}_U$	0.344	0.363	0.345	0.349	0.361	0.346	0.347	0.360	0.350

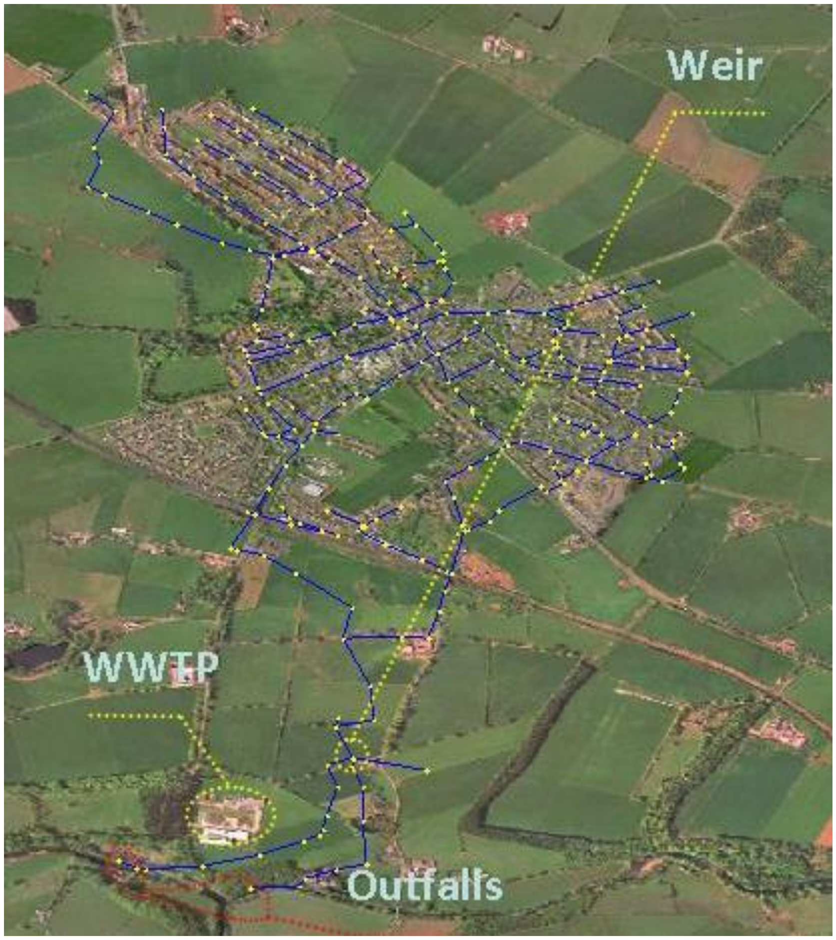
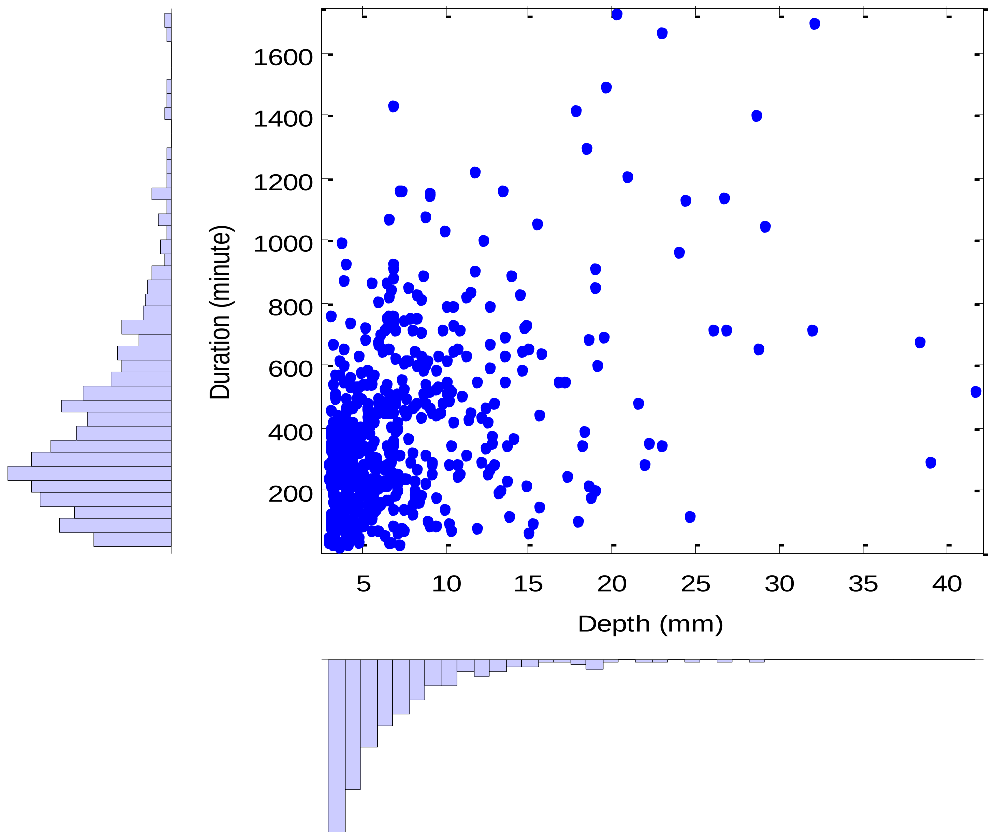


Fig. 1 - Layout of the case study network.



588

589 **Fig. 2 - Scatter plot of rainfall depth and duration with marginal histograms.**

590

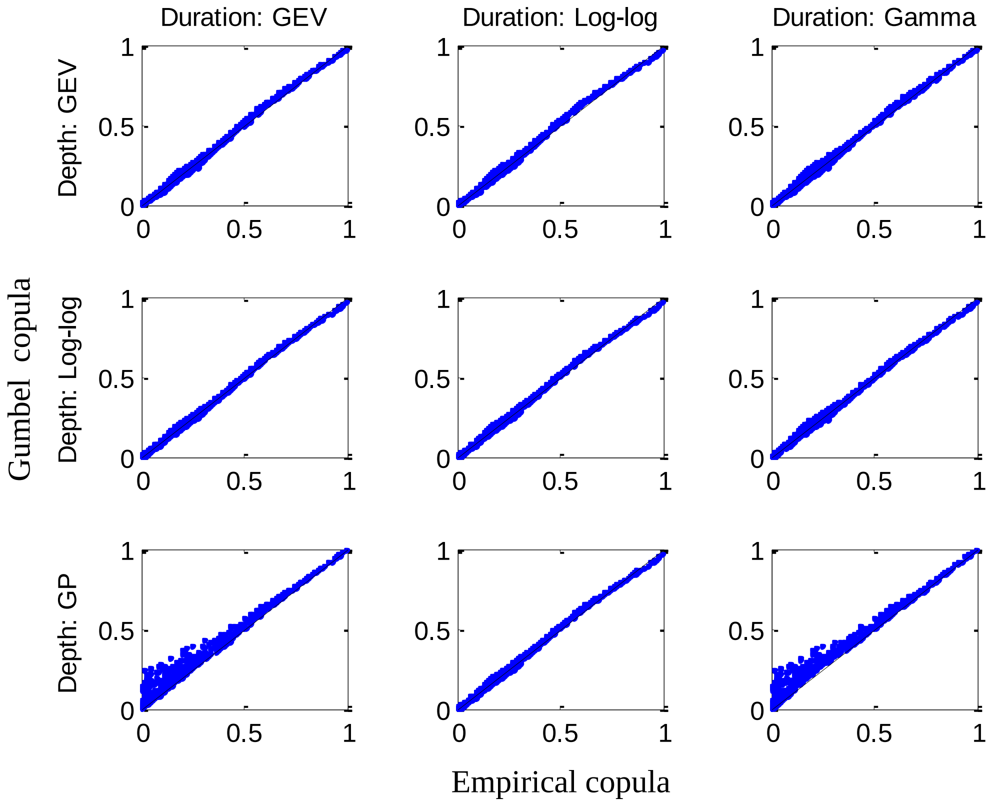


Fig. 3 - Q-Q plots of Gumbel and empirical copulas for different marginal distribution combinations.

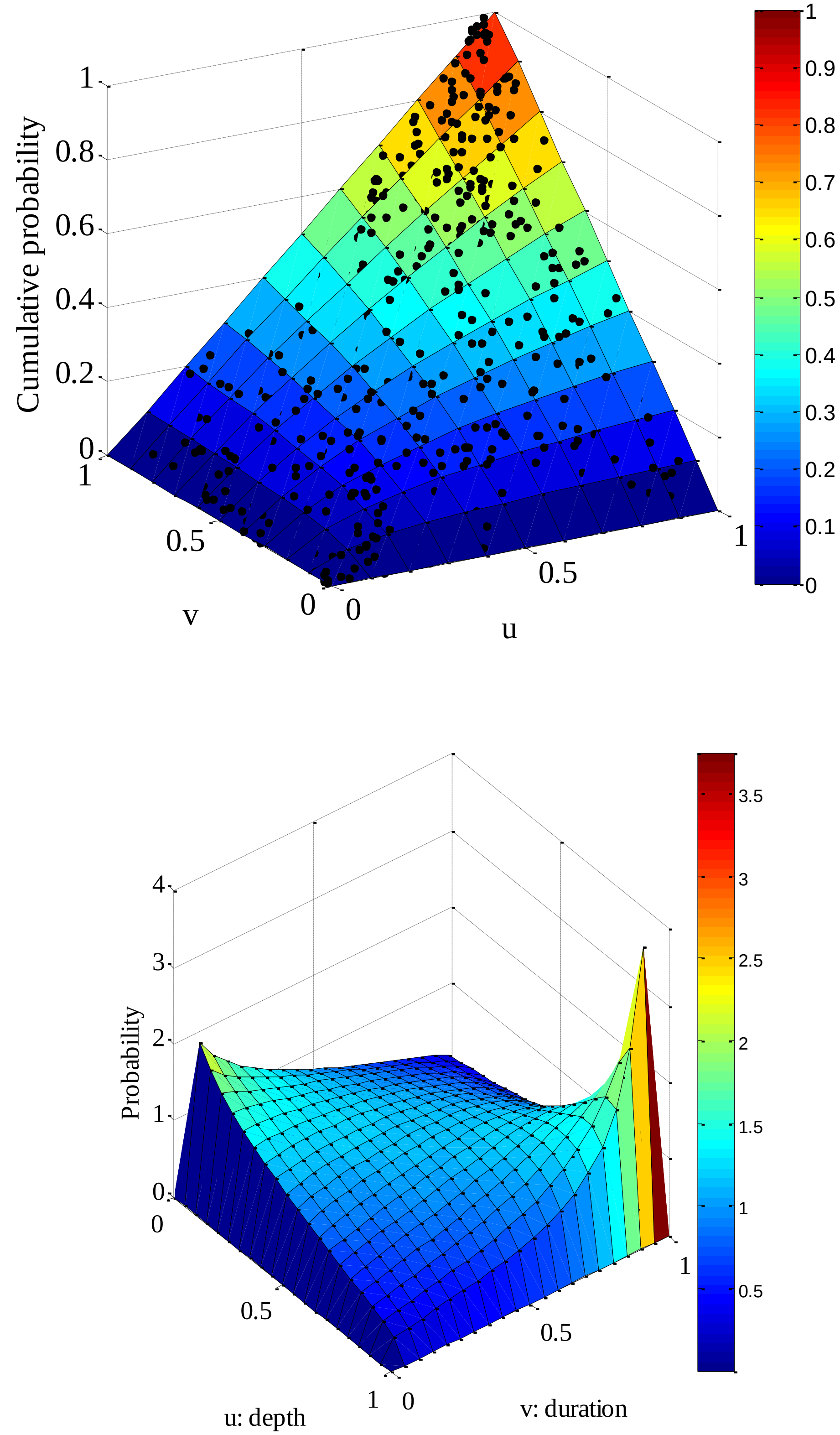


Fig. 4 - Three dimensional plots for the theoretically fitted Gumbel copula.

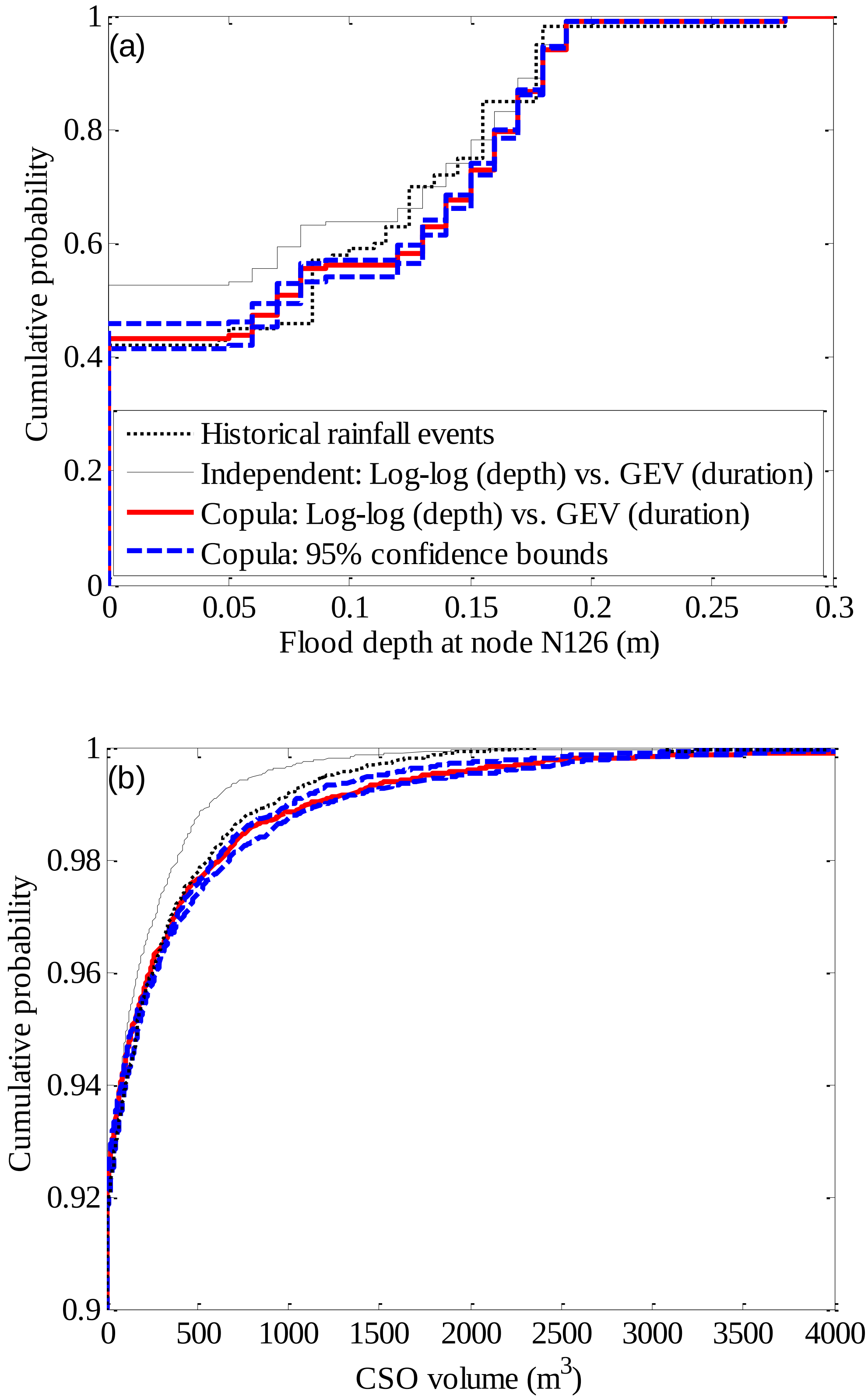


Fig. 5 - Cumulative probabilities of flood depth and CSO volume from Gumbel copula.

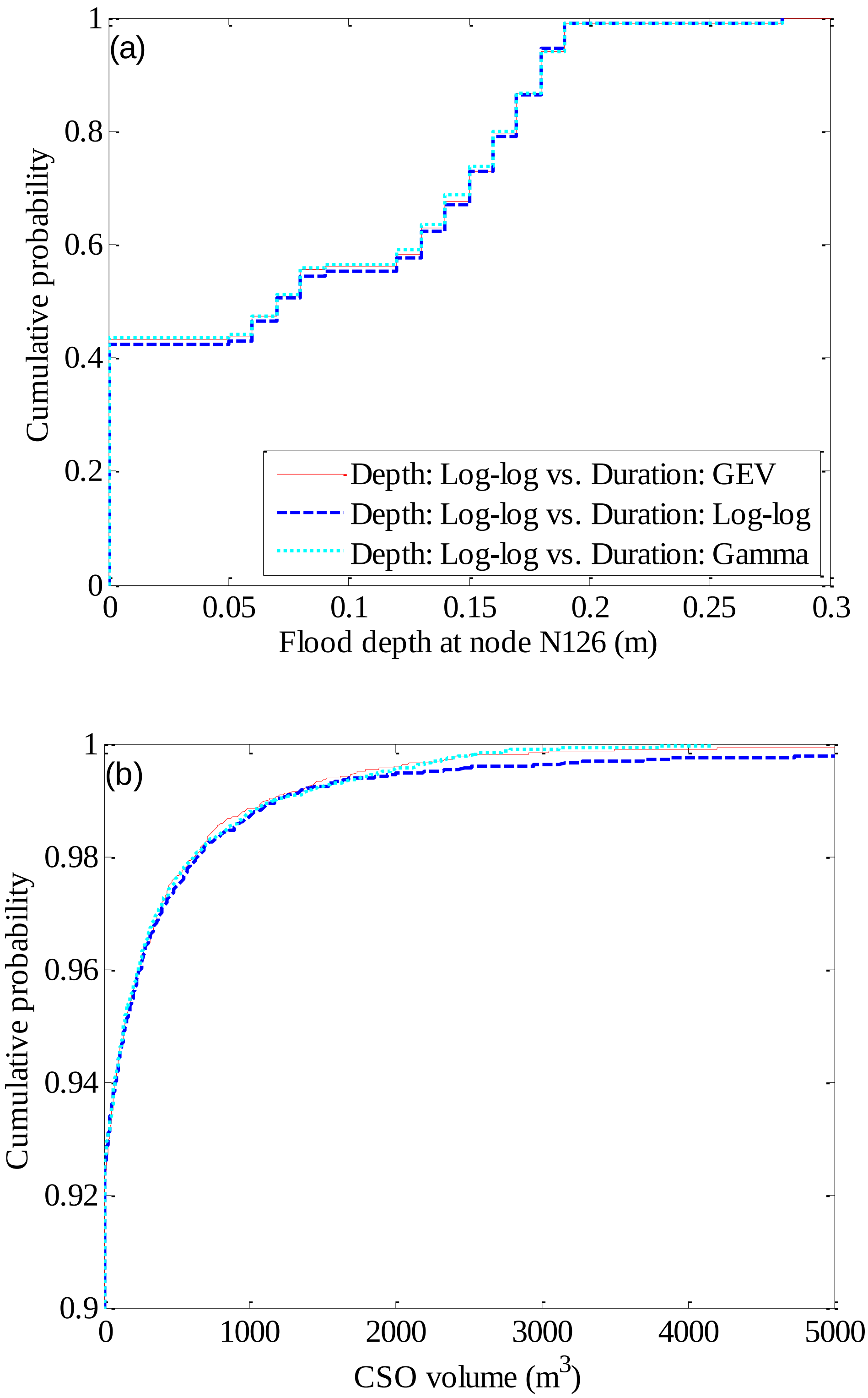


Fig. 6 – Impacts of different marginal distributions on the cumulative probabilities of flood depth and CSO volume.

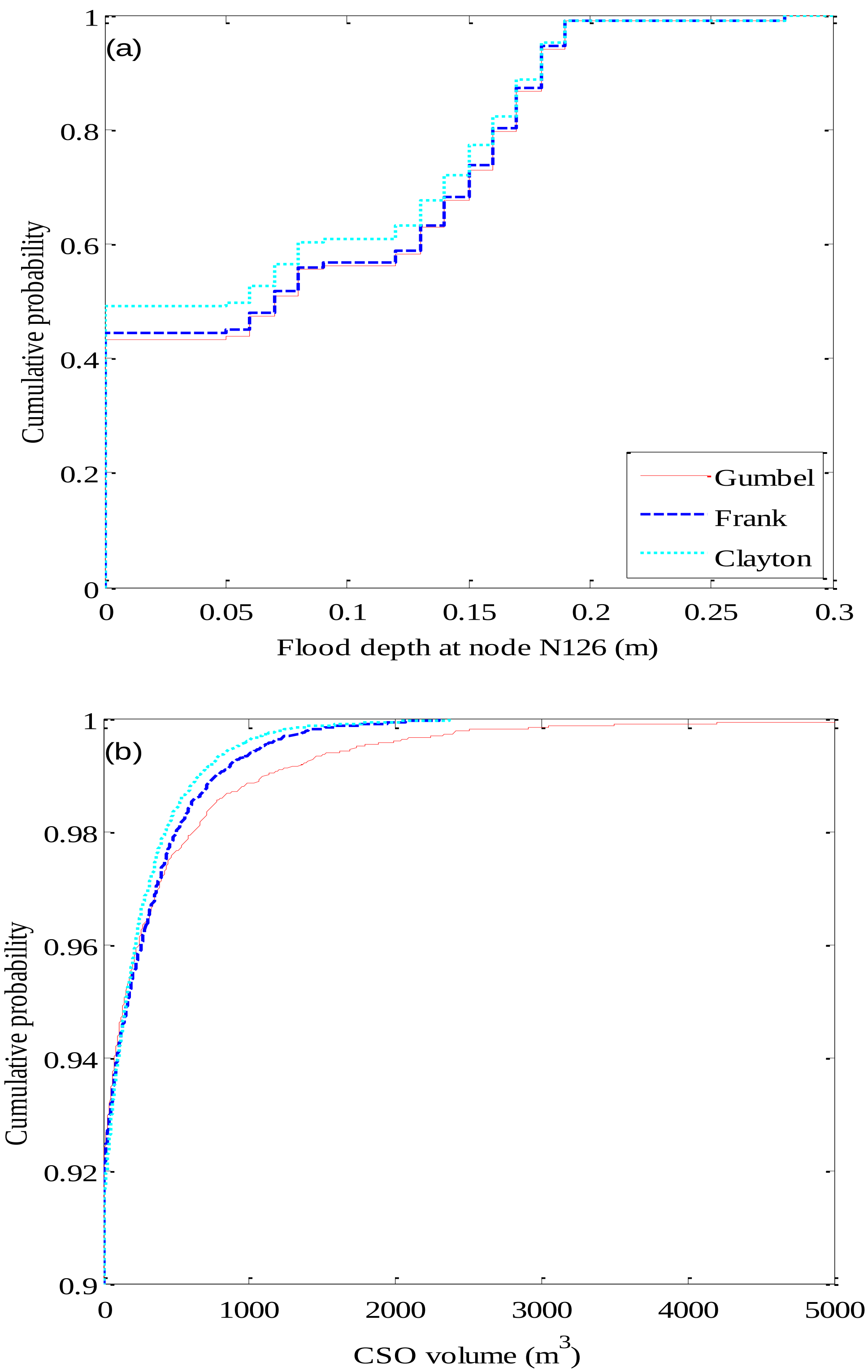


Fig. 7 – Comparison of Gumbel and Frank copulas.

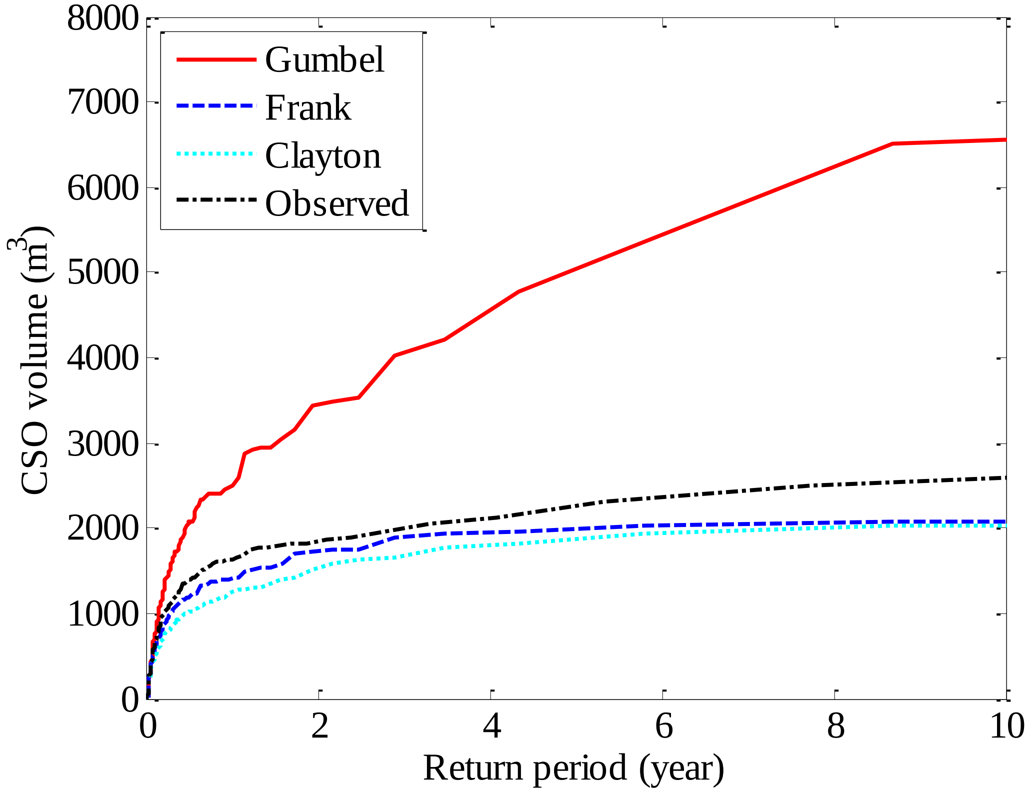


Fig. 8 – Return periods of CSO discharge volumes.