Journal of Hydrology 510 (2014) 49–58

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

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

ARTICLE INFO

Article history:
Received 16 February 2012
Received in revised form 17 June 2013
Accepted 6 December 2013
Available online 15 December 2013
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Harald Kunstmann, Associate Editor

Keywords:
Bivariate distribution
Combined sewer overflow
Copula
Frequency analysis
Sewer flooding
Urban drainage system

SUMMARY

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 and the different impacts of dependence structure and marginal distributions on system performance are analysed. 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 dependence structure and marginal distributions on the probabilities of sewer flooding and CSO volume.

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). Sewer 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 re-assess the performance of sewer systems regarding sewer flooding and CSO discharging rather than rely on the design capacity in a changed situation (Korving et al., 2002; Schmitt et al., 2004; Thorndahl and Willems, 2008).

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

* Corresponding author. Tel.: +44 (0)1392 723692; fax: +44 (0)1392 217965.
E-mail address: g.fu@exeter.ac.uk (G. Fu).

0022-1694/$ - see front matter
http://dx.doi.org/10.1016/j.jhydrol.2013.12.006

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, 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 separate impacts of the rainfall depth and duration marginal distributions and their dependence structures on the hydraulic performance of a combined sewer system using copulas. 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 Distribution Functions (CDF) of flood depth at one critical node and CSO overflow volume are calculated. The results obtained from the urban drainage case study show the suitability and flexibility of the Archimedean copulas in simulating the dependence of rainfall depth and duration, and the significant 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

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

where u and v are uniformly distributed random variables and $u, v \in [0, 1]$. The joint CDF $H_{XY}(x, y) = P(X \leqslant x, Y \leqslant y)$ describes the probability of two events: $X \leqslant x$ and $Y \leqslant y$. 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

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

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; Vandenberghe et al., 2011) 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 (non-parametric) 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. According to the work by Nelsen (2006), the relationship between parameter θ and Kendall's τ exists for the three Archimedean families. Particularly, a closed-form expression can be derived for Clayton and Gumbel families (see Kao and Govindaraju, 2007b).

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