Parameter optimization and uncertainty assessment for rainfall frequency modeling using an adaptive Metropolis-Hastings algorithm
Xingpo Liu, Chengfei Xia, Yifan Tang, Jiayang Tu and Huimin Wang

ABSTRACT
A new parameter optimization and uncertainty assessment procedure using the Bayesian inference with an adaptive Metropolis-Hastings (AM-H) algorithm is presented for extreme rainfall frequency modeling. an efficient MCMC sampler is adopted to explore the posterior distribution of parameters and calculate their uncertainty intervals associated with the magnitude of estimated rainfall depth quantiles. Also, the efficiency of AM-H and conventional maximum likelihood estimation (MLE) in parameter estimation and uncertainty quantification are compared. And the procedure was implemented and discussed for the case of Chaohu city, China. Results of our work reveal that: (i) The adaptive Bayesian method, especially for return level associated to large return period, shows better estimated effect when compared with MLE. It should be noted that the implement of MLE often produces overly-optimistic results in the case of Chaohu city. (ii) AM-H algorithm is more reliable than MLE in terms of uncertainty quantification, which yields relative narrow credible intervals for the quantile estimates to be instrumental in risk assessment of urban storm drainage planning.

Key words | adaptive Metropolis-Hastings (AM-H) algorithm, frequency distribution, Markov Chain Monte Carlo (MCMC), parameter optimization, parameter uncertainty analysis, rainfall frequency modeling

HIGHLIGHTS
• Bayesian method with adaptive Metropolis-Hastings algorithm (AM-H) shows better results compared with maximum likelihood estimation (MLE).
• Bayesian MCMC approach offers an attractive framework to quantify parameter posterior distribution.
• AM-H is more reliable than MLE on uncertainty quantification, which yields relative narrow credible intervals.
INTRODUCTION

Effective urban hydrologic engineering planning, design and operation are very much dependent on reliable rainfall frequency modeling, such as intensity/depth-duration-frequency (IDF/DDF) curves or rainfall intensity formula, which can summarize the return levels (i.e., intensities and depths) of extreme rainfall for the continuum of durations (e.g., 5-min to 24-h) and specified return periods (e.g., Chow et al. 1988; Cheng & AghaKouchak 2014; Sarhadi & Soulis 2017; Parvez & Inayathulla 2019). In practice, when the recorded rainfall series are long enough, IDF/DDF curves or rainfall intensity formula can generally be determined via frequency analysis of extreme rainfall (Porras & Porras 2001). While it is recognized that the true frequency distribution of extreme rainfall at any region is unknown, a reasonably screened two- or three-parameter theoretical distribution model is typically selected to quantifying distributional behavior of extreme events. In Ghana, for instance, Endreny & Imbeah (2009) employed two independent sets of historical rainfall data using two approaches of frequency distribution analysis to estimate parameters of regional IDF. Further, some analysts have established various formulae for design storm through the IDF/DDF curves. For example, Chen (1985) derived a generalized IDF formula for any location in the United States in the light of three basic rainfall depths, that is, R110 (1 h, 10-year rainfall depth), R2410 (24 h, 10-year rainfall depth), and R1100 (1 h, 100-year rainfall depth).

However, the raises the following question: how to produce accurate and reliable prediction of frequency and magnitude of extremes rainfalls with a given duration from rainfall IDF/DDF curves? has been an urgent and broad research subject in the past few decades (e.g., Willems 2000; Langousis & Veneziano 2007; Tung & Wong 2013; Langousis et al. 2016).

The critical element in the construction of IDF/DDF curves is the estimation of the distribution parameters. Classical used statistical methods include the maximum

NOMENCLATURE

AM-H Adaptive Metropolis-Hastings
MCMC Markov Chain Monte Carlo Method
IDF/DDF Intensity/Depth-Duration-Frequency
K-S Kolmogorov-Smirnov
NSE Nash-Sutcliffe Efficiency Coefficient
RMSE Root Mean Square Error
MAE Maximum Absolute Error
R Coefficient of Correlation
GEV Generalized Extreme Value distribution
P-III Pearson Type III distribution
LP-III Log-Pearson Type III distribution
Logn(3p) Logarithmic normal distribution (3-Parameter)
EVI Gumbel distribution
Exp Exponential distribution
MLE Maximum Likelihood Estimation
PDF Probability Density Function
CDF Cumulative Density Function

GRAPHICAL ABSTRACT
likelihood estimation method (MLE), the method of moments (Stedinger et al. 1993) and the method of the L-moments (Huard et al. 2010). Among which the MLE is the most common approach for estimating model parameters because of its relatively high efficiency and accuracy, especially for plausible modeling during extreme value frequency analysis (e.g., Cannon 2009; Villarini et al. 2010), meanwhile, maximum likelihood estimates are asymptotically unbiased (or consistent). However, imperfections of MLE may cause poor estimates when the selected sample size is short or when outliers exist (Madsen et al. 1997).

Bayesian analysis as an alternative estimation method compared with previous classical methods, provides a flexible and coherent framework for extreme value modeling which proves superiority in (a) Coupling known data with additional types of information based on functions of prior parameter density, and (b) quantifying uncertainties of estimates to make prediction (Ouarda & El-Adlouni 2011).

Nevertheless, the implement of Bayesian approach in extreme value analysis seems to be difficult due to imperfection in computing capacity and algorithms in the early 1990s. In the past two decades, Bayesian analysis has been widely applied, with the introduction of the theory of Markov Chain Monte Carlo (MCMC) addressing the limitations arising from computational aspects (Gelfand & Smith 1990). Also, some researchers have made dramatic contributions to promoting MCMC techniques in the field of extreme value modeling. For example, Coles & Powell (1996) reviewed the developments of Bayesian approaches in extreme value modeling, and also discussed the modeling ideas using MCMC techniques as well as explored the feasibility of performing Bayesian analysis in the case of MLE fails.

After the estimation of the parameters, a difficulty in producing accurate predictive estimation of extreme rainfall is that the propagation of parameter uncertainty. In general, the parameter uncertainty is due to return periods of interest used in practical engineering usually to hundreds of years, whereas records at disposal are most of the time insufficient or missing, leading to selected probability distribution can not fit well to the observed data. (e.g., Yen & Tang 1976; Tung 1987; Yu & Cheng 1998). It should be emphasized that whether the unacceptable risk of flooding caused by the underestimation of extreme values, or the construction of costly infrastructures caused by the overestimation will be the price to pay for overlooking uncertainty in estimation of the extremes of rainfall (Huard et al. 2010).

Over the past few decades, some works have explicitly described uncertainties in IDF/DDF curves. They broadly rely on two distinct theoretical hypothesises. The first one is a classical frequency statistics in which the parameters are treated as unknown scalars. Quantification of uncertainty is mainly based on the asymptotic approximation which characterized the uncertainty by calculating confidence intervals at certain level of significance (e.g., Coles 2001; Hailegeorgis et al. 2013). However, it performed poorly in quantile estimates due to be not the true posterior distributions (Reis & Stedinger 2005). The second one is a Bayesian method, the main difference from the first method in that the parameters are treated as random variables. If parameters of posterior distribution are investigated using MCMC sampler, its estimation allows by nature uncertainty quantification through providing the most likely distribution for the parameters based on the data. For examples, Coles et al. (2005) reiterated that conventional technologies (e.g., MLE) do not fully considered the uncertainties of models and predictions, and are confined to produce overly optimistic appraisals of future extreme events. And they also pointed out that Bayesian methods have theoretical and technical advantages in the evaluation of uncertainties. Huard et al. (2010) generated IDF relationships based on Bayesian analysis demonstrating the extent of uncertainties of IDF curves. Also, Mélese et al. (2018) proposed a regional study of uncertainties in IDF relationships derived from point-rainfall maxima base on two generalized extreme value models. Their work shows that the Bayesian framework is more robust than the frequentist one to the starting point of the estimation procedure.

Just like other optimization algorithms, some adjustments could be made to improve performance of MCMC sampler and accelerate convergence to the stationary target status. Specific operations include (a) Picking an initial starting point, which is located in the high-density region of the posterior distribution, in order to limit the length of the burn-in period (Renard et al. 2012). (b) Considering the difficulty of calculating the covariance matrix, AM-H is utilized to modify the jump size and direction by changing the characteristics of the jump distribution during each iteration (Renard et al. 2006). (c) Every $k$-value is retained to reduce the autocorrelation in the well-converged Markov chains (Artur et al. 2017).

Till date, Bayesian analysis still has great application potential in estimation of IDF/DDF relationships’ parameters and addressing uncertainties in parameters that are, crucial in design of urban flood prevention facilities. This study presents two innovative aspects: (i) The application of AM-H algorithm is relatively novel in extreme rainfall frequency modeling; (ii) The uncertainties of probability distribution parameters were often not taken into...
consideration for the conventional rainfall frequency modeling algorithms including MLE. The developing adapted Bayesian inference offers an attractive framework to optimize these parameters and quantify their uncertainties in rainfall frequency analysis.

The rest of this paper is as follows: Section 2 gives a detailed introduction to the rainfall frequency analysis methods and theoretical background for Bayesian MCMC approach inference with AM-H. The study area and rainfall data used are briefly introduced in Section 3 and then Section 4 reveals characteristics of the distribution of the maximum daily rainfall by comparing the fitting results of six candidate distribution models in detail. This is followed by Section 5, in which the performance of the AM-H is compared to the MLE with a quadratic approximation in the two aspects including parameters estimation and uncertainty quantification. Finally, the main findings on advantages of the AM-H and limitations of the MLE are summarized in Section 6.

METHODOLOGY

Overview of the workflow

Establishment of the formula of daily rainfall depth for Chaohu city based on extreme rainfall (i.e., annual maximum) frequency analysis involve a sequence of processes. A schematic representation of the procedure is illustrated in Figure 1.
Rainfall frequency analysis method

Annual maximum method

Sampling is the process of reasonably extracting a series of values from existing records to form a sample that can be used as the basis for frequency analysis.

The most widely used annual maximum method as the sampling method for this study (e.g., Vidal 2013). The so-called annual maximum method refers to in the \( n \)-years observed series, only the largest value will be collected in each year, regardless of whether the second event in a year exceeding annual maxima of other years.

Empirical frequency curve

The empirical frequency methods are the commonly employed in hydrologic/ hydraulic engineering design and planning. Specifically for this method used in this paper are described below, firstly, Let us assume that the total size of the sample is \( n \) and that the values are arranged in descending order, that is, \( X = \{x_{1n}, x_{2n}, \ldots, x_{n|n} | x_{1n} \geq x_{2n} \geq \ldots \geq x_{n|n} \} \), then values of the empirical frequency with specified rainfall depth is computed using the following form:

\[
P_e = \frac{m}{n+1} \times 100\% \quad m = 1, 2, \ldots, n
\]

where \( P_e \) are values of the empirical frequency, that is, the frequency of obtaining rainfall depth larger than a specified depth for a given duration, \( m \) denotes any serial number of rainfall depth in the range of 1 to \( n \), \( n \) is the size of data.

Probability distribution models

Probability distribution models (or theoretical distribution curves) are usually combined with the empirical frequency curves to meet the requirement for high precision design in hydrosystem infrastructures in the frequency analysis.

Six alternative distribution functions that are widely performed in frequency analysis of extreme events, namely: Generalized Extreme Value, Three-Parameter Lognormal, Pearson Type III, Log-Pearson Type III, Two-Parameter Exponential, as well as Gumbel, to model the annual maximum rainfall series.

Pearson type III. Let \( x \) be a random variable that only takes continuous real values. and the Pearson type III probability density function (PDF), can be expressed as (Vogel & McMartin 1991):

\[
f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left( x - \gamma \right)^{\alpha-1} \exp(-\beta(x - \gamma)) \quad (\gamma < x < \infty, \ \alpha > 0, \ \beta > 0)
\]

(2)

\[
\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x) dx
\]

(3)

In addition, the parameters \( \alpha, \beta \) and \( \gamma \) are related to the first three months of the variable \( x \) as follows:

\[
\mu = \gamma + \frac{\alpha}{\beta}
\]

(4)

\[
\sigma^2 = \frac{\alpha}{\beta^2}
\]

(5)

\[
\phi = \frac{2\beta}{|\beta|^{1/2}}
\]

(6)

Then, its cumulative distribution function (CDF), is given by:

\[
F(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left( \frac{x - \gamma}{\beta} \right)^{\alpha-1} \exp(-\beta(x - \gamma)) dx
\]

(7)

where \( \alpha, \beta \) and \( \gamma \) are, respectively, shape parameter, scale parameter and location parameter.

Logpearson type III. As early as 1967, the U.S. Water Resources Council recommended the LP-III distribution to be applied to flood frequency studies. Subsequently, Bobee (1975) considered that the LP3 distribution is more flexible than the P. III.

The essence of LP-III distribution is to assume that the logarithmic transformation of the random variable obeys P-III distribution, and the PDF and CDF of a LP-III distribution are defined respectively by:

\[
f(x) = \frac{1}{\beta x \Gamma(\alpha)} \left[ \ln (x) - \gamma \right]^{\alpha-1} \exp \left( \frac{-\ln(x) - \gamma}{\beta} \right)
\]

(8)

\[
(\alpha > 0, \ \beta \neq 0)
\]

\[
F(x) = \frac{\Gamma(\ln(x) - \gamma) / \beta^{\gamma(\alpha)}}{\Gamma(\alpha)}
\]

(9)
where $\Gamma(\cdot)$ is defined as the gamma function:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x) \, dx \quad (x > 0, \ \alpha > 0, \ \beta > 0) \quad (10)$$

where $\alpha, \beta, \gamma$ are, respectively, shape parameter, scale parameter and location parameter.

**Lognormal.** If the random variable $y = \ln(x)$ is normally distributed, then $x$ is called lognormal distributed. The PDF and CDF for random variable $x$ are introduced as follows:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\ln(x) - \mu}{\gamma}\right)^2\right) \quad (11)$$

$$F(x) = \frac{1}{2} \left[ 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln(x) - \mu} \exp\left(-\frac{1}{2} \left(\frac{y}{\gamma}\right)^2\right) \, dy \right] \quad (12)$$

where $\sigma, \mu, \gamma$ are, respectively, scale parameter, location parameter and shape parameter.

**EVI.** Gumbel distribution function is derived based on the extreme value theorem, also termed as the Extreme Value Type I distributions (EVI), and Gumbel (1958) is widely used distribution in hydrology owing to its suitability for maxima modeling. The PDF and CDF of a Gumbel distribution are shown respectively as follows:

$$f(x) = \frac{1}{\sigma} \exp\left(-\frac{x - \mu}{\sigma}\right) \exp\left(-\exp\left(-\frac{x - \mu}{\sigma}\right)\right) \quad (13)$$

where

$$z = \frac{x - \mu}{\sigma} \quad (14)$$

$$F(x) = P(X \leq x) = \exp(-\exp(-z)) \quad (15)$$

where $P(X \leq x)$ is the probability of an event not exceeding $x$, $x > 0, -\infty < x < \infty$, $x$ denotes observed rainfall. $\mu, \sigma$ are, respectively, location parameter, and scale parameter.

**Exponential.** In rainfall frequency analysis, exponential distribution (Exp) is also a commonly used type of probability distribution. The PDF and CDF are defined respectively by:

$$f(x) = \lambda \exp(-\lambda(x - \gamma)) \quad (\gamma \leq x < +\infty) \quad (16)$$

$$F(x) = 1 - \exp(-\lambda(x - \gamma)) \quad (17)$$

where $x$ represents the observed rainfall value, $\lambda(\lambda > 0), \gamma(\gamma \geq 0)$ are, respectively, scale parameter, and location parameter.

**Generalized extreme value.** According to the extreme value theory, the GEV distribution is a family of continuous probability distributions. The GEV distribution is often applied to model the smallest or largest value among a large set of independent, identically distributed random values representing measurements or observations. The PDF and CDF for random variable $x$ are introduced as follows:

$$f(x) = \frac{1}{\sigma} \left[ 1 - \left( \frac{x - \mu}{\sigma} \right)^{\frac{k}{\lambda}} \right]^{-1} \exp\left(-\left(1 - \left(\frac{x - \mu}{\sigma}\right)^{\frac{k}{\lambda}}\right)^{\frac{1}{k}}\right) \quad (18)$$

$$F(x) = \exp\left(-\left(1 - \left(\frac{x - \mu}{\sigma}\right)^{\frac{k}{\lambda}}\right)^{\frac{1}{k}}\right) \quad (19)$$

where $x$ denote the random variable representing the observed rainfall value, $\sigma, k, \mu$ are the scale, the shape and the location parameter, respectively.

**Theoretical background of the optimization algorithms**

**Bayesian MCMC framework**

In the Bayesian framework, the current status of knowledge is assessed through integrating existing the data of interest, collecting new information to resolve issues, and updating to account for both new and old data. Further, The Bayesian analysis is based on identifying a probability distribution model for the observed data $(x)$ and a vector $(\theta)$ of unknown random parameter. The well known Bayes’ theorem states that:

$$p(\theta|x) = \frac{f(x|\theta)p(\theta)}{\int f(x|\theta)p(\theta)d\theta} \propto f(x|\theta)p(\theta) \quad (25)$$

where $p(\theta|x)$ is the PDF of the posterior distribution for parameter $(\theta)$, and $p(\theta)$ denotes the PDF of the prior distribution for $\theta$, and $f(x|\theta)$ is the conditional density of the distribution.
for observed series \((x_1, x_2, \ldots, x_n)\), which is denoted as follows:

\[
f(x|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \tag{26}
\]

Also, \(\pi(\theta|x)\) can summarize the uncertainty associated to the parameter \(\theta\) based on observed data and the prior information. In Bayesian analysis, parameters of posterior distribution are explored using MCMC sampler, it could be obtained a large number of samples for which could be treated as the realization of posterior distribution. It includes the marginal parameter distributions, whose sample properties (e.g. sample variance) can be used to quantify estimation uncertainties. Notably, before the posterior distribution computed, normal distribution is always adopted as prior distribution for its characteristic of large variance if parameters to be estimated are non-informative. In addition, it can achieve Bayesian credible intervals, which is similar to the confidence interval in conventional framework (e.g., MLE), to estimate the extreme rainfall quantiles. Further, to put Bayes’ theorem into practice, we need to consider two aspects: (i) likelihood function; (ii) the calculation of posterior distribution (Equation (25)).

However, it is difficult to compute analytical form for the posterior distribution (Equation (25)) due to most of models which are contributed complex high-dimensional joint density of parameters. This issue can be addressed by using the method of Markov Chain Monte Carlo (MCMC), which provides an effective way of simulating from complex models.

Adaptive Metropolis-Hastings algorithm

MCMC simulation is based on the fact that the height of the posterior density can be estimated from the simulated a specific set of parameter values at convergence of the Markov chains (Gelman et al. 2004). Further, the algorithms rely on Bayesian MCMC method are fundamentally Monte Carlo integration by simulating from Markov Chains.

Among the various techniques based on MCMC, the Metropolis-Hastings (M-H) sampling developed by Metropolis et al. (1953) is an attractive choice for generating observations from the posterior density, and has been proved successful in a large number of practices (e.g., Marshall et al. 2004; Kavetski et al. 2006). Generally, M-H with symmetric jumping distributions can be illustrated to simulate as follows:

(i) Start with an initial value \(\theta^{(0)}\), and strictly greater than zero;
(ii) At each iteration-step \(i\),

1. Specify a candidate \(\theta^*\) from the jumping distribution \(J_i(\theta^*|\theta^{(i-1)})\);
2. Derive a probability \(\alpha_i\) to accept the candidate \(\theta^*\);

\[
\alpha_i = \min \left[1, \frac{\pi(\theta^*|x)}{\pi(\theta^{(i-1)}|x)} \right]
\]

\[
= \min \left[1, \frac{f(x|\theta^*)\pi(\theta^*)}{f(x|\theta^{(i-1)})\pi(\theta^{(i-1)})} \right] \tag{27}
\]

3. Accept or reject the proposal \(\theta^*\), i.e., set

\[
\theta^{(i)} = \begin{cases} 
\theta^*, & \text{with probability } \alpha_i, \\
\theta^{(i-1)}, & \text{otherwise.} 
\end{cases} \tag{28}
\]

To improve the performance of the M-H, appropriate optimization is needed for the MCMC sampler using integrating adaptive algorithm. So, an adaptive Metropolis-Hastings algorithm (AM-H) is developed. Firstly, the starting point is initialized by MLE. If the initial value is not properly selected, which may cause the chains fail to reach convergence after hundreds of iterations. Secondly, applying the AM-H, we obtain a Markov chain of length 2,000 with start value to limit the initial burn-in period, and then the remaining chain of 10,000 iterations converges to the posterior distribution. Thirdly, it is noted that the resulting chains are deemed to be identically distributed but not independent, since there exists significant autocorrelation in the chains. Therefore, an appropriate lag \(k\)-value (set lag \(k = 5\) in this study, i.e., every 5th iteration of the remaining 10,000 iterations is selected for the estimation of the posterior density) to reduce dependence of a random walk from the MCMC sampler make sure the final chains generated using AM-H are considered independent for posterior distribution.

Maximum likelihood estimation

MLE is commonly adopted to estimate the parameters in the selected distribution. When \(\theta\) denotes the unknown parameters, further, \(L(\theta)\) (that is the likelihood function of \(\theta\)) can be defined as:

\[
L(\theta) = \prod_{i=1}^{n} f_x(f_i|\theta) \tag{29}
\]
where $n$ is the size of sets and $x_1, x_2, \ldots, x_n$ are the observed data. The purpose is to maximize the likelihood function for the given series (Lee et al. 2010).

After determining the best estimates, the uncertainty of the parameters need also to be evaluated. According to the assumption of linearity and normality, quantization of uncertainty is performed based on the quadratic approximation by the likelihood function (Bickel & Doksum 1977). The normal approximation confidence interval (two-sided) with significance $\alpha$ for $\theta$ are expressed as:

\[
\hat{\theta} \sim N(\theta, I^{-1})
\]

\[
\Pr \left[ \hat{\theta} - \frac{Z(1 - \frac{\alpha}{2})}{\sqrt{I(\hat{\theta})}} \leq \theta \leq \hat{\theta} + \frac{Z(1 - \frac{\alpha}{2})}{\sqrt{I(\hat{\theta})}} \right] \approx 1 - \alpha
\]

where $\Pr[\text{\textperiodcentered}]$ stands for the probability of the rainfall event, $I(\hat{\theta})$ is a positive constant marked as the curvature at $\hat{\theta}$, which is the estimate of $\theta$, and $\sqrt{I(\hat{\theta})}$ is an estimate of the standard error of $\hat{\theta}$ (Meeker & Escobar 1998). What is more, $\alpha$ denotes a significant level.

**STUDY AREA AND DATA USED**

Chaohu city is located at Anhui province, China, covering the coordinates of 31.16°N to 32°N, 117.25°E to 117.58°E and the study area has an area of 2,046 km² (Figure 2). And this region with abundant sunshine, subtropical monsoon climate. The annual average sunshine duration are about 2,106 hours, and there is a frost-free period of about 250 days. Average annual rainfall is between 1,000.3 mm and 1,157.1 mm. Furthermore, due to the influence of the monsoon climate, fifty percent of the annual mean rainfall are mainly concentrated in the summer months (June to September).

In the light of the explicit requirement in the Code for Design of Outdoor Wastewater Engineering (GB50014-2006, edition. 2016) established by Minstry of Housing and Urban-Rural Development of China (MOHURD), the annual maximum method as a sampling method is feasible to be applied to the designs of flood control facilities. Also, in storm design analysis, most of hydrologists consider to choose the daily (24-h) events as the logical basis for stormwater structures design due to the accurate rainfall data can be available for daily scale (e.g., Kunkel et al. 1999; AlHassoun 2011). Thus, the annual maximum daily rainfall series (1957–2015) was extracted from the archives of a meteorological station in Chaohu city, as the source of the current rainfall frequency analysis.

**FREQUENCY ANALYSIS OF THE ANNUAL MAXIMUM DAILY RAINFALL**

In this section, the suitability of different probability distributions for the frequency analysis need to be evaluated. Then the form of distribution will be selected that curve not only provides the best fit to the plotted empirical points, but will also yield accurate quantile estimates.
Analysis of empirical frequency curve

The annual maximum of daily rainfall for the period from 1957 to 2015 in millimeters is shown in Figure 3(a). In Figure 3(a) we are mindful of annual rainfall pattern are highly variable, with the highest value of 251.8 mm being that of year 2008. Meanwhile, the result of the trend detection analysis based on ensemble empirical mode decomposition (EEMD) method (More details of the properties of EEMD can be found in Wu & Huang (2009)), indicates that annual maximum daily rainfall amounts in Chaohu city has a trend of increasing year by year in extreme rainfall during the last 59 years.

Before determining the best-fit distribution model to fit the annual maximum daily rainfall data, the plotting positions of the empirical frequency curve of the sample need to be calculated using Equation (1) first. It can be seen intuitively from Figure 3(b) that the maximum daily rainfall depths are found to decrease with increasing exceedance probability.

Selection of the probability distribution

Once we established the empirical frequency curve, six candidate distributions considered for quantile estimation, including: GEV, Logn(3p), P-III, LP-III, Exp, as well as EVI were all statistically analyzed for fitting the given sample. Subsequently the best-fit distribution was selected based on determined evaluation criterion, including Kolmogorov–Smirnov (K-S) goodness of fit test combined with RMSE, NSE, MAE as well as R.

The fitting results of the above distributions to empirical frequency points are intuitively illustrated in Figure 4. The horizontal axis represents the maximum daily rainfall observations whereas the perpendicular axis denotes the exceedance probability. In terms of visual verification, almost all the distributions exhibit the higher accuracy of estimation with exception of Exp (solid line), which consistently gave lower or higher estimates of the quantiles in different phases. This is especially obvious when there are lower values than the observed ones.

Using simple Q-Q diagnostic plots to evaluate the goodness of fit of the selected distributions (Buuren & Fredriks 2001). As shown in Figure 5, more than 95% variation of selected sample can be described using the fitted distributions except for Exp (some scattered points in the end of line do not fit well along the dash line). Additionally, it should be emphasized that these plots allow the visual estimation of the goodness of fit between empirical points and each distribution.

For the purpose of selecting the optimal distribution, the visual fitting comparison, although necessary, is highly subjective. In order to overcome this subjectivity, several
commonly statistical methods are available for further evaluating the fitted performance of each distribution.

During this period, the performances of our candidate distributions were first compared based on a goodness of fit test using the K-S test (at 5% level of significance). The results obtained by K-S test are described in Table 1. Calculating the $D$-value for all distributions, it can be found that the top 5 candidates (P-III, LogN (5p), GEV, EVI, LP-III), which successfully pass the K-S test ($D < D_\alpha = 0.1737$), also the GEV is the one having the smallest D-value. It is worth noting that Exp, which do not pass the K-S test ($D = 0.1739 > D_\alpha = 0.1737$), is exclude. This result directly proves our conjecture that the maximum daily rainfall data for Chaohu city do not fit Exp. In addition, Table 1 also lists the estimated values of parameters using the method of MLE.

Due to the results of K-S test can only exclude the inappropriate distribution (Exp). Therefore, it is necessary to adopt the quantitative statistical criteria (i.e., RMSE, NSE, MAE, R) to further investigate fitting performance of top 5 distributions, and the result are presented in Table 2. It can be observed that there is no much difference in these distributions, however, except for the indicator ‘MAE’, all the indicators of GEV are better performance than that of other distributions.
Hence, the results described above suggested that GEV could be employed as the most appropriate distribution function to describe annual maximum daily rainfall series for Chaohu city.

Parameter optimization of the best-fit distribution

After conducting distribution selection, considering the fact that the parameters calculated by MLE may have poor performance, the corresponding GEV cannot express the numerical characteristics of rainfall sample well. Therefore, Bayesian MCMC with the AM-H was adopted as a recommendation method to optimize the parameters in the GEV.

The following points require consideration before proceeding with the work of Bayesian MCMC. In order to converge the stationary chain and reduce dependence of the burn-in period (i.e., a process of removal of the unstable initial data) within the sample, the AM-H was coded by MATLAB platform to obtain Markov chains and executed to generate the 12,000 estimates of the parameters in the GEV. Also, we supposed the simulated values follow the posterior distribution, and the mean value of the posterior density (i.e., posterior mean) is defined as the final estimate of each parameter. Further, the ‘Geweke’ value, an indicator which reflects the convergence of the Markov chains (Geweke 1991), was introduced, the closer the Geweke value to one, the better the convergence.

In the Bayesian framework, the posterior distribution of the GEV can be described by Equation (25).

$$\pi(\sigma, k, \mu | x) = \frac{f(x|\sigma, k, \mu)\pi(\sigma, k, \mu)}{\int_0^\infty \int_0^\infty \int_0^\infty f(x|\sigma, k, \mu)\pi(\sigma, k, \mu)d\sigma dk d\mu}$$

(32)

where $f(x|\sigma, k, \mu)$ is the likelihood function of the GEV, and $\pi(\sigma, k, \mu)$ denotes the prior distribution.

The iterative process of each individual Markov chain corresponding to parameter is shown in Figure 6 after running the algorithm. In the figure, the estimates of parameters from 0 to 2,000 iterations are eliminated due to these values are unstable (burn-in = 2,000). Only the 10,000 estimates (from 2,001 to 12,000 iterations) were performed to calculate the mean values. Also, the ‘Geweke’ value of the first two chain all reach above 0.985, and the ‘Geweke’ value of the last chain is 0.862. this result suggests that convergence had been achieved for each chain.

Further, it is straightforward to see that the marginal histogram (the first row of Figure 7) composed of posterior densities of each parameter is approximately normal distribution. This phenomenon demonstrates again that the Markov chains of estimated components of the parameter can converge to the target state well. In addition, from the generated chains we calculated the relative distances of the chain points from the origin and counted the points that are inside given probability limits. Then the scatter matrix plot is generated and displayed shown in the bottom two rows of Figure 7, this scatterplot is presented traces of MCMC sampling, nearly all scatters which are within the 95% credible interval.

### Table 1 | Summary statistics regarding K-S test and parameter estimation for all tested distribution functions fitted to the rainfall sample (at 5% level of significance)

| Distribution | $D$-value | Order | Critical value ($D^*_{\alpha/n}$) | Check | Location | Scale | Shape |
|--------------|-----------|-------|-----------------------------------|-------|----------|-------|-------|
| P-III        | 0.0722    | 4     | 0.1737                            | Yes   | 27.839   | 31.466| 2.133 |
| Logn (3P)    | 0.0711    | 3     | 0.1737                            | Yes   | 4.321    | 9.143 | 0.515 |
| GEV          | 0.0668    | 1     | 0.1737                            | Yes   | 72.953   | 31.64 | –0.108 |
| EVI          | 0.0781    | 5     | 0.1737                            | Yes   | 74.091   | 36.159| –     |
| LP-III       | 0.0677    | 2     | 0.1737                            | Yes   | –3.259   | 0.027 | 283.33 |
| Exp          | 0.1739    | –     | 0.1737                            | No    | 31.7     | 0.0158| –     |

### Table 2 | Statistical comparison between distributions fitting to the empirical frequency curve

| Statistics | P-III | Logn (3P) | GEV | EVI | LP-III |
|------------|-------|-----------|-----|-----|--------|
| RMSE       | 0.0261| 0.0239    | 0.0236| 0.0266| 0.0237 |
| NSE        | 0.9916| 0.9929    | 0.9931| 0.9912| 0.9930 |
| MAE        | 0.0581| 0.0623    | 0.0646| 0.064 | 0.0650 |
| $R^2$      | 0.9958| 0.9965    | 0.9969| 0.9962| 0.9968 |

Hence, the results described above suggested that GEV could be employed as the most appropriate distribution function to describe annual maximum daily rainfall series for Chaohu city.
Immediately after, the mean values of posterior density after iterations as the final estimated results of the three parameters are summarized in Table 3 with the values obtained using MLE in comparison. In the light of the results of evaluation indicators, AM-H and MLE method have the approximate capacity of point estimation. But from Figure 8,
we can observe that the position of the theoretical curve obtained by the AM-H estimates with some reasonable adjustments, compared with that of MLE.

Summarizing the above results, parameters estimated by AM-H were used for the GEV to yield a more confident relationship between rainfall depth and theoretical frequency of annual maximum daily rainfall series in Chaohu city, and the detailed summary of annual maximum daily rainfall depth against the exceedance probability as tabulated in Table 4.

**DAILY RAINFALL DEPTH FORMULA FOR CHAOHU CITY**

Code for Design of Outdoor Wastewater Engineering (GB50014-2006, edition. 2016) indicates that the adopted return period varies between 3 and 5 years for design of storm drainage networks in Chaohu city, and 20–30 years for waterlogging prevention design. However, greater return periods have to be chosen when considering to lower the risks of potential damage (provide a sufficient protection level) or when more stringent design criteria are required for certain functional operations. In doing that, our follow-up work is to derive the daily rainfall depth formula for Chaohu city, applicable for high (large) return periods or lower exceedance probabilities.

**Determination of the formula**

The daily rainfall depth formula as a derived equation in power-law form describing the relationship between daily rainfall and theoretical frequency (return period). Herein, we give a detailed procedure of the derivation of the formula.

First, in the light of the results of each performance evaluation indicator in Section 4.2, the best model form (GEV) is chosen from the CDF of the top five probability distributions (better fitting models) as a reference for deriving daily rainfall formula.
The second step is to develop the daily rainfall formula using CDF of the GEV, the process works as follows.

As introduced in Section 2.2.3, the CDF can be illustrated as follows:

\[ F(x) = P(x \leq x_p) = \exp \left[ -\left(1 - k \frac{x - \mu}{\sigma} \right)^{\frac{1}{k}} \right] \]  

where \( x \) denotes observed rainfall, \( x_p \) denotes benchmark values.

In order to obtain the probability of a rainfall event exceeding \( x_p \), it follows from Equation (34) that.

\[ P(x > x_p) = 1 - F(x) = 1 - \exp \left[ -\left(1 - k \frac{x - \mu}{\sigma} \right)^{\frac{1}{k}} \right] \]  

Then the equation of return period \( T \) (i.e., the reciprocal of the probability of exceedence) is obtained from.

\[ T = \frac{1}{P(x > x_p)} = \frac{1}{1 - \exp \left[ -\left(1 - k \frac{x - \mu}{\sigma} \right)^{\frac{1}{k}} \right]} \]  

Finally, the quantile rainfall formula will be in the following form.

\[ D(24, T) \equiv x = \sigma \left\{ -\ln \left(1 - \frac{1}{T} \right)^{\frac{1}{k}} - 1 \right\} + \mu \]  

where, \( D(24, T) \) (in mm) describes the required annual maximum daily rainfall depth for return period \( T \) (in years), when parameters of the formula are presented. location (\( \mu \)), scale (\( \sigma \)) and shape (\( k \)) are parameters to be estimated, which constitute the parameter vector \( \theta = (\mu, \sigma, k) \).

**Parameter estimation and uncertainty analysis**

In this section, Bayesian method is adopted to estimate the parameters and quantify the uncertainty in the daily rainfall depth formula, when compared with MLE.

Firstly, in the case of meeting the settings mentioned in Section 4.3, Bayesian MCMC work was carried out again, that is, the posterior distributions of parameters were acquired using MCMC sampling with AM for estimates. Also, the posterior distribution of the formula can be described by Equation (37).

\[ \pi(\sigma, \mu, k|x) = \frac{f(x|\sigma, \mu, k)\pi(\sigma, \mu, k)}{\int_0^\infty \int_0^\infty \int_0^\infty f(x|\sigma, \mu, k)\pi(\sigma, \mu, k)d\sigma d\mu dk} \]  

where \( f(x|\sigma, \mu, k) \) is the likelihood function of the formula, and \( \pi(\sigma, \mu, k) \) denotes the prior distribution.

Figure 9 shows the iterative process of each Markov chain corresponding to the parameters. Also, in the first row of Figure 10, we can observe that the frequency histograms approximately match the normal distribution. Meanwhile, the trace of MCMC sampling is presented in the scatter plot (the bottom two rows of Figure 10).

In terms of the point estimates of the parameters, the posterior mean value of the AM-H and the mode value of the MLE were listed in Table 5. It can be seen from the fitting results (various indicators) that more accurate performance is yielded by the AM-H proposed to estimate...
the parameters in rainfall depth formula as compared with estimates for the MLE. Also, as can be clearly seen from Figure 11(a), the fitted curve of the AM-H is more consistent with the observed rainfall data compared with that of the MLE. Another important finding is that the AM-H acquires high accuracy when estimating the quantile of rainfall depth with large return periods. In the case of the 200-year return period, from the result of the quantile extrapolation, estimation by the AM-H is 346.447 mm and that of the MLE is 332.985 mm. It is indicated that the MLE underestimates the rainfall quantile certainly to some extent when the maximum rainfall with large return periods. What is more, this underestimation will become more serious as the return period become larger.

Quantification of the uncertainty interval in parameters except for estimating the optimal values is of great significance. The results of the quantile estimation including uncertainty at return periods of 10-, 30-, 50-, 100-, 200-, 500-year are summarized in Table 6. Note that the uncertainty caused by the AM-H or the MLE is defined as the difference between values of the upper (97.5%) and lower (2.5%) bound, respectively, we can find that the range of the uncertainty of the AM-H is remarkably reduced over all the return periods compared with that of the MLE.
For instance, in the period of 500-year, the credible interval by AM-H is [387.211, 461.985], but by MLE the interval is [302.758, 515.723].

On the other hand, the curves that relate the return periods and their corresponding return levels (rainfall depth) are plotted (Figure 11(b)). The light-grey bar which represents the estimates for the 2.5% centile curve, and 97.5% centile curve, indicates the 95% credibility interval for the rainfall quantiles obtained by the AM-H. Meanwhile, the dotted lines which represent the quantile estimates of the 2.5% and 97.5% centile curve, shows the 95% confidence level for the rainfall quantiles obtained using the MLE with a quadratic approximation. It proves that the Bayesian MCMC analysis with the AM-H is more reliable than MLE in the aspect of quantifying the uncertainty, which provides a 95% credible interval that is as narrow as possible while still being correct estimates (i.e., on average 95% of the observed data falling within the filled region).

**CONCLUSIONS**

Evaluation of the depth (intensity) and the frequency of occurrence of the extreme rainfall events are crucial for water resources planning and management as well as the reliability-based hydraulic design of urban infrastructures, such as determination of the required capacity of detention basins. At the same time, quantification of uncertainty is also a key component of evidence-based decision making in water resources engineering in order to avoid overly optimistic results.

This study employed a Bayesian MCMC method to investigate annual maximum daily rainfall series in Chaohu city from 1957 to 2015. To this end, the adaptive Metropolis-Hastings (AM-H) algorithm was apply to address the extreme rainfall frequency analysis and derive the formula for daily rainfall depth. Furthermore, efficiency of AM-H algorithm was compared to MLE with a quadratic approximation.
approximation in the two aspects including parameters estimation and uncertainty quantification. The major conclusions drawn from this study are as follows:

1. From the results of the quantitative goodness of fit tests, the Generalized Extreme Value distribution (GEV) is concluded as the best-fit distribution to fit the annual maximum daily rainfall events of Chaohu city.

2. It can be seen from the results of the parameter estimation and quantile estimation in rainfall frequency analysis, the Bayesian MCMC inference with AM-H is shown to be superior to MLE when point estimates are required with reasonable accuracy. Also, in the aspect of the uncertainty analysis, because of the Bayesian approach is based on the posterior distribution of observed data, AM-H can guide the curve fitting of the frequency curve by introducing a relatively credible interval. Furthermore, it is found that severe underestimation results of the extreme rainfall quantile caused by MLE can occur when the extrapolation is required at larger return period for Chaohu city case. Additionally, it should be emphasized that Bayesian approach is versatile and flexible in that researchers can easily be tailored for any chosen rainfall frequency analysis process.

3. Formula for daily rainfall depth of Chaohu city is given in the present study, which is of the form

\[ D(24, T) = 31.743 \left\{ -\ln \left( 1 - \frac{1}{T} \right)^{-0.1718} - \frac{1}{0.1718 + 2.2769} \right\} \]

And it will help to estimate the required daily rainfall depth for high return periods (e.g., 100- to 500-year) with more reliability.

ACKNOWLEDGEMENTS

This study has been sponsored by the National Natural Science Foundation of China [Grant No. 51008191]. In addition, the authors sincerely thank the editors and anonymous reviewers for offering constructive criticisms to improve this manuscript.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories.

REFERENCES

AliHassoun, S. A. 2011 Developing an empirical formula to estimate rainfall intensity in Riyadh region. *Journal of King Saud University - Engineering Sciences* **23**, 81–88.

Artur, T. S., Maria, M. P., Mauro, N. & Wilson, F. 2017 A Bayesian peaks-over-threshold analysis of floods in the Itajaí-açu River under stationarity and nonstationarity. *Stochastic Environmental Research and Risk Assessment* **31**, 185–204.

Bickel, P. J. & Doksum, K. A. 1977 *Mathematical Statistics- Basic Ideas and Selected Topics*. Holden-Day, San Francisco, USA.

Bobee, B. B. & Robitaille, R. 1977 The use of the Pearson type 3 and log Pearson type 3 distributions revisited. *Water Resources Research* **13**, 427–443.

Buuren, S. V. & Fredriks, M. 2001 Worm plot: a simple diagnostic device for modelling growth reference curves. *Statistics in Medicine* **20**, 1259–1277.

Cannon, A. J. 2009 A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimatology. *Hydrological Processes* **24**, 673–685.

Chen, C. L. 1985 Rainfall intensity–duration–frequency formulas. *Journal of Hydraulic Engineering-ASCE* **109**, 1603–1621.

Cheng, L. & AghaKouchak, A. 2014 Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. *Scientific Reports*. doi.org/10.1038/srep07093.

Chow, V. T., Maidment, D. R. & Mays, L. W. 1988 *Applied Hydrology*. McGraw-Hill, New York.

Coles, S. 2001 *An Introduction to Statistical Modeling of Extreme Values*. Springer, London. doi.org/10.1007/978-1-4471-3675-0

Coles, S. G. & Powell, E. A. 1996 Bayesian methods in extreme value modelling: a review and new developments. *International Statistical Review/Revue Internationale de Statistique* **64**, 119.

Coles, S., Percicchi, L. R. & Sisson, S. 2003 A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology* **273**, 35–50.

Endreny, T. A. & Imbeah, N. 2009 Generating robust rainfall intensity–duration–frequency estimates with short-record satellite data. *Journal of Hydrology* **371**, 182–191.

Gelfand, A. & Smith, A. 1990 Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. 2004 *Bayesian Data Analysis*, 2nd edn. Chapman & Hall/CRC, New York.

Geweke, J. F. 1991 Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Staff Report* **4**, 169–193.

Gumbel, E. J. 1958 *Statistics of Extremes*. Columbia University Press, New York.

Hailegeorgis, T. T., Thorolfsson, S. T. & Alfredsen, K. 2013 Regional frequency analysis of extreme precipitation with consideration of uncertainties to update (IDF) curves for the city of trondheim. *Journal of Hydrology* **498**, 305–318.
Huard, D., Mailhot, A. & Duchesne, S. 2010 Bayesian estimation of intensity–duration–frequency curves and of the return period associated to a given rainfall event. *Stochastic Environmental Research and Risk Assessment* **24**, 337–347.

Kavetski, D., Kuczera, G. & Fanks, S. W. 2006 Bayesian analysis of input uncertainty in hydrological modeling: 1. theory. *Water Resources Research* **42**. doi.org/10.1029/2005wr004568.

Kunkel K, E., Andsager, K. & Easterling, D. R. 1999 Long-term trends in extreme precipitation events over the contiguous United States and Canada. *Journal of Climate* **12**, 2515–2527.

Langousis, A. & Veneziano, D. 2007 Intensity-duration-frequency curves from scaling representations of rainfall. *Water Resources Research* **43**. doi.org/10.1029/2006wr005245.

Langousis, A., Mamalakis, A., Puliga, M. & Deidda, R. 2016 Threshold detection for the generalized Pareto distribution: review of representative methods and application to the NOAA NCDC daily rainfall database. *Water Resources Research* **52** (4), 2659–2681.

Lee, W. S., Lee, K. S., Kim, S. U. & Chung, E. S. 2010 The development of rating curve considering variance function using pseudo-likelihood estimation method. *Water Resources Management* **24**, 321–348.

Liu, X., Xia, C., Chen, Z., Chai, Y. & Jia, R. 2020 A new framework for rainfall downscaling based on EEMD and an improved fractal interpolation algorithm. *Stochastic Environmental Research and Risk Assessment* **34**, 1147–1173.

Madsen, H., Rasmussen, P. & Rosbjerg, D. 1997 Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 1: at-site modeling. *Water Resources Research* **33**, 747–757.

Marshall, L., Nott, D. & Sharma, A. 2004 A comparative study of markov chain monte carlo methods for conceptual rainfall-runoff modeling. *Water Resources Research* **40**. doi.org/10.1029/2003wr002378.

Meeker, W. Q. & Escobar, L. A. 1998 *Statistical Methods for Reliability Data*. John Wiley&Sons, New York, USA.

Mélie, V., Blanchet, J. & Molinier, G. 2008 Uncertainty estimation of intensity-duration-frequency-relationships: a regional analysis. *Journal of Hydrology* **558**, 579–591.

Metropolis, N., Rosenbluth, A. W., Teller, A. H. & Teller, E. 1953 Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1092.

Ouarda, T. & El-Adlouni, S. 2011 Bayesian nonstationary frequency analysis of hydrological variables1. *Journal of the American Water Resources Association* **47**, 496–505.

Parvez, M. & Inayathulla Dr., M. 2019 Statisical analysis of rainfall for development of intensity-duration- frequency curves for Upper Cauvery Karnataka by log-normal distribution. doi.org/10.26438/ijsrms/v6i5.1233.

Porras Sr., P. J. & Porras, P. J. 2001 New perspective on rainfall frequency curves. *Journal of Hydrologic Engineering* **6**, 82–85.

Reis Jr., D. S. & Stedinger, J. R. 2005 Bayesian MCMC flood frequency analysis with historical information. *Journal of Hydrology* **315**, 97–116.

Renard, B., Garreta, V. & Lang, M. 2006 An application of Bayesian analysis and MCMC methods to the estimation of a regional trend in annual maxima. *Water Resources Research* **42**, W12422.

Renard, B., Sun, X. & Lang, M. 2012 Bayesian methods for non-stationary extreme value analysis. *Water Science and Technology Library* **39–95**. doi.org/10.1007/978-94-007-4479-0_3.

Sarhadi, A. & Soulis, E. D. 2017 Time-varying extreme intensity-duration-frequency curves in a changing climate. *Geophysical Research Letters* **44** (3), 2454–2463.

Stedinger, J., Vogel, R. & Georgiou, E. 1993 *Handbook of Hydrology: Frequency Analysis of Extreme Events. Chap Frequency Analysis of Extreme Events*. McGraw-Hill, pp. 18.1–18.66.

Tung, Y. 1987 Uncertainty analysis of national weather service rainfall frequency atlas. *Journal of Hydraulic Engineering* **113**, 179–189.

Tung, Y. & Wong, C. 2013 Assessment of design rainfall uncertainty for hydrologic engineering applications in Hong Kong. *Stochastic Environmental Research and Risk Assessment* **28**, 583–592.

Vidal, I. 2013 A Bayesian analysis of the Gumbel distribution: an application to extreme rainfall data. *Stochastic Environmental Research and Risk Assessment* **28**, 571–582.

Villarini, G., Smith, J. A. & Napolitano, F. 2010 Non-stationary modeling of a long record of rainfall and temperature over Rome. *Advances in Water Resources* **33**, 1256–1267.

Vogel, R. W. & McMartin, D. E. 1991 Probability plot goodness-of-fit and skewness estimation procedures for the Pearson type 3 distribution. *Water Resources Research* **27**, 3149–3158.

Willems, P. 2000 Compound intensity/duration/frequency-relationships of extreme precipitation for two seasons and two storm types. *Journal of Hydrology* **233**, 189–205.

Wu, Z. & Huang, N. E. 2009 Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis* **1**, 1–41.

Yen, B. C. & Tang, W. H. 1976 Risk–safety factor relation for storm sewer design. *Journal of Environment Engineering* **102**, 509–516.

Yu, P. S. & Cheng, C. J. 1998 Incorporating uncertainty analysis into a regional IDF formula. *Hydrological Processes* **12**, 713–726.