A Factorial Bayesian Copula Framework for Partitioning Uncertainties in Multivariate Risk Inference

Y. Fan[1*], K. Huang[2], G.H. Huang[3,4*], Y.P., Li[4]

[1] Department of Civil and Environmental Engineering, Brunel University, London, Uxbridge, Middlesex, UB8 3PH, United Kingdom

[2] Faculty of Engineering and Applied Sciences, University of Regina, Regina, SK, Canada

[3] Institute for Energy, Environment and Sustainable Communities, University of Regina, Regina, Saskatchewan, Canada S4S 0A2

[4] School of Environment, Beijing Normal University, Beijing, China, 100875

*Correspondence: Dr. Y. Fan
Department of Civil and Environmental Engineering,
Brunel University London,
Uxbridge, Middlesex, UB8 3PH, United Kingdom
Tel: +44 1895265717
Email: yurui.fan@brunel.ac.uk

Dr. G. H. Huang
Institute for Energy, Environment and Sustainable Communities,
University of Regina, Regina, Saskatchewan, Canada S4S 0A2,
Tel: +1 306 585-4095;
Fax: +1 306 585-4855;
E-mail: huang@iseis.org

Abstract:

In this study, a factorial Bayesian copula (FBC) method is proposed to quantify parameter uncertainty in copula-based models and then reveal their contributions to the multivariate hydrologic risk inference. In detail, Bayesian inference and factorial analysis are integrated into copula-based multivariate risk models to (1) quantify parameter uncertainties, (ii) reveal their individual and interactive effects, and (iii) identify their detailed contributions on uncertain risk inference. Streamflow observations at Xiangxi and Wei River basins of China are used to illustrate the applicability of FBC. The results indicate that imprecise parameters in marginal distributions and the dependence structure would lead to extensive uncertainties in predictive joint return periods and failure probabilities. Also, individual and interactive effects of parameters are well revealed through multilevel factorial analysis, and the detailed contributions of one parameter to different failure probabilities under different service time scenarios are identified.

Keywords: Flood risk; Copula; Markov chain Monte Carlo; Factorial Analysis; Uncertainty

1. Introduction

Flooding, as one of the most frequently occurred natural hazards, has taken a devastating societal and economic toll over the world, leading to a large number of fatalities and property losses. (Kidson and Richards, 2005; Karmakar and Simonovic, 2009; Fan et al., 2015a, b; Huang et al., 2019; Lindenschmidt and Rokaya, 2019; Wu et al., 2019). Assessment and management of flood risks concerns many government agencies, academic institutions as well as individual stakeholders. However, hydrometeorological processes are recognized as multivariate phenomena characterized by dependent multi-attribute properties (Sarhadi et al., 2016). Accordingly, a typical flood event generally presents multiple features such as peak discharge, hydrograph volume and duration. Univariate risk analyses, mainly focusing on flood peaks, cannot procure a full description of the probability of occurrence of the hydrological event (Chebana and Ouarda, 2011; Requena et al., 2013). Consequently, multivariate approaches are recommended by many studies since they can involve a number of non-independent variables for the characterization of a flood (The European Parliament and The Council, 2007; Li et al., 2015; Fan et al., 2016a, b; Salvadori et al., 2016).

The applications of multivariate hydrologic risk analysis are growing dramatically since the introduction of copulas in hydrology and geosciences (Serinaldi, 2013). De Michele and Salvador (2003) initially introduced the concept of copulas into hydrological simulation to describe the dependence between storm duration and average rainfall intensity. After that, a great number of research works have been proposed for multivariate hydrologic simulation through copula functions, such as multivariate flood frequency analysis (Zhang and Singh 2006; Sraj et al., 2014); drought assessments (Song and Singh 2010; Kao and Govindaraju 2010); storm or rainfall dependence analysis (Zhang and Singh 2007; Vandenberghe et al. 2010);

streamflow simulation (Lee and Salas 2011; Kong et al., 2015; Fan et al., 2017). Copulas can model nonlinear dependence between two or more depended variables with different marginal distributions and relax assumptions of same family of distributions and linear relationship in traditional multivariate techniques (Zhang and Singh, 2006; Genest and Favre, 2007; Karmakar and Simonovic, 2009; Sraj et al., 2014; Huang et al., 2017).

One major issue in hydrologic risk analysis is the presence of uncertainties, resulting from model selection and parameter estimation. There are two primary sources of uncertainty: (1) natural uncertainty stemming from variability of the underlying stochastic process, and (2) epistemic uncertainty coming from incomplete knowledge about the system under study (Merz and Thieken, 2005). In particular, the limited sample size of hydrological data implies large uncertainty on the extreme quantiles (Serinaldi, 2013). Uncertainty assessment is a prominent aspect in univariate frequency analysis and it is also quite crucial in multivariate framework. Several research works have been proposed for evaluating uncertainties in copula-based multivariate risk framework (Serinaldi, 2013; Dung et al. 2015; Zhang et al., 2015). For instance, Serinaldi (2013) proposed a Monte Carlo-based approach to generate confidence intervals around p-level curves to account for uncertainties in joint quantile. These research works demonstrate that significant uncertainties exist in copula-based hydrologic risk assessment. However, one major issue to be addressed is to characterize the major sources that lead to uncertainties in multivariate risk predictions. In a multivariate framework, uncertainties in both marginal distributions and dependence structures may lead to varied risk predictions. However, few research studies have been reported to answer which source contributes most to the uncertainty in the resulting risk predictions.

Consequently, this study aims to propose a factorial Bayesian copula (FBC) framework to quantify uncertainties in multivariate risk analysis and further partition sources in the uncertain risk inference. This approach integrates copula model, Bayesian parameter estimation and multilevel factorial analysis into a framework. In detail, the multivariate risk inference models in terms of joint return periods and failure probabilities (FPs) are established based on copulas. Parameter uncertainties in marginal distributions and dependence structure are quantified by a Bayesian based Markov Chain Monte Carlo (MCMC) algorithm. Finally, individual and interactive effects of parameter uncertainties are revealed by multilevel factorial analysis. Flood data at three hydrological gauge stations in China are used to illustrate the applicability of the proposed method.

2. Methodology

The proposed FBC approach mainly consists of three components, including: (i) establishment of copula-based multivariate risk assessment model, (ii) quantification of parameter uncertainties and (iii) characterization of contributions of parameters to uncertainties in risk inference. Figure 1 illustrates the detailed procedures for FBC.

2.1. Copula-based Multivariate Risk Assessment

In the hydrology context, many hazard events may present with multivariate characteristics. For instance, floods are generally characterized by its peak and volume values, while droughts have multi-attribute of severity and duration. Also, these multi-attributes in one hydrological hazard are usually correlated. To reveal the dependence among the multiple features in one hazard and characterize the associated risk in a multivariate framework, the copulas, initially introduced into hydrology by Favre et al. (2004), have been widely used. The applicability of copulas is mainly attributed to their flexibility in modelling dependence among correlated variables with different marginal distributions. Also, a variety of dependence structures, such as asymmetry, nonlinear and tail dependence, are able to be captured by copulas (Sarhadi et al., 2016).

Consider one hydrological hazard has d correlated attributes (e.g. peak, volume and duration for a flood) with each one denoted by a random variable X_i ($i = 1, 2, \ldots, d$). If the corresponding probability distributions are denoted as $F_1(x_1|\gamma_1)$, $F_2(x_2|\gamma_2)$, …, $F_d(x_d|\gamma_d)$, where $\gamma_1, \gamma_2, \ldots, \gamma_d$ are parameters in probability distributions, the joint probability distribution of $X_1, X_2, \ldots, X_d$ can be expressed as (Nelsen, 2006):

$$F(x_1, ..., x_d \mid \gamma_1, ..., \gamma_d, \theta) = C(F_1(x_1 \mid \gamma_1), ..., F_d(x_d \mid \gamma_d) \mid \theta) \tag{1}$$

where $C(.)$ is a copula function; θ is the parameter in the copula function describing dependence among those correlated variables. More details on theoretical background and properties of various copulas can be found in Nelsen (2006).

Through the probability distribution in Equation (1), some features of the hazardous event can be revealed. Firstly, the concept of return period (RP) is of great importance in water resources and civil engineering for (i) designing and sizing hydraulic structures, (ii) identifying dangerous events, (iii) making rational making, and (iv) assessing related risk (Salvadori et al., 2013). In a multivariate context, the RP of one specific hazardous event should consider the interaction among different attributes in the hazardous scenarios, leading to multivariate RP. A number of literatures have been proposed to characterize multivariate RP in hydrologic issues (Salvadori et al., 2007; 2011; 2013, 2016; Fan et al., 2016a, b). In general, consider one kind of hydrological extreme (denoted as $\boldsymbol{X}$) with d attributes (i.e. $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$), three categories of multivariate RP are widely used for hydrologic risk assessment.

(i) "OR" case: $T^{OR} = \{(x_1, x_2, ..., x_d) \in R^d : x_1 > x_1^* \vee x_2 > x_2^* \vee ... \vee x_d > x_d^*\}$, which indicates at least one element surpass the predefined threshold. Based on the copula function, the multivariate RP in "OR" case can be expressed as:

$$T^{OR} = \frac{\mu}{1 - C(F_1(x_1 \mid \gamma_1), ..., F_d(x_d \mid \gamma_d) \mid \theta)} \tag{2}$$

where μ denotes the average time between two adjacent events under consideration.

(ii) “AND” case: $T^{AND} = \{(x_1, x_2, ..., x_d) \in R^d : x_1 > x_1^* \wedge x_2 > x_2^* \wedge ... \wedge x_d > x_d^*\}$ which indicates at all elements in the extreme events should exceed the corresponding thresholds. Based on the copula function, the multivariate RP in “AND” case can be expressed as:

$$T^{AND} = \frac{\mu}{\hat{C}(\bar{F}_1(x_1 \mid \gamma_1), \bar{F}_2(x_2 \mid \gamma_2), ..., \bar{F}_1(x_d \mid \gamma_d) \mid \theta)} \tag{3}$$

where $\hat{C}$ is multivariate survival function of the X_i’s proposed by Salvadori et al. (2013; 2016), and $\bar{F}_i(x_i \mid \gamma_i) = P(X > x_i) = 1 - F_i(x_i \mid \gamma_i)$. Following Salvadori et al. (2013; 2016), and the Inclusion-Exclusion principle proposed by Joe (2014), the multivariate survival function $\hat{C}$ can be obtained by:

$$\hat{C}(\mathbf{u}) = \bar{C}(1 - \mathbf{u}) \tag{4}$$

and

$$\bar{C}(\mathbf{u}) = 1 - \sum_{i=1}^{d} u_i + \sum_{S \in \mathrm{P}} (-1)^{\#(S)} C_S(u_i : i \in S) \tag{5}$$

(iii) “Kendall” case: The Kendall RP is to characterize the hydrologic disasters exceeding a critical layer defined by (Salvadori et al., 2011): $L_t^F = \{\boldsymbol{x} \in R^d : F(\boldsymbol{x}) = t\}$. The Kendall RP can be expressed as (Salvadori et al., 2011):

$$T^{Kendall} = \frac{\mu}{1 - K_C(t)} \tag{6}$$

where K_C is the Kendall’s distribution function associated with C, which can be expressed as:

$$K_C(t) = P(C(F_1(x_1 \mid \gamma_1), ..., F_d(x_d \mid \gamma_d) \mid \theta) \leq t) \tag{7}$$

In addition to the multivariate RP, Serinaldi (2015) recently proposed the notion of failure probability (FP) to provide more coherent, general and well devised tools for risk assessment and communication. In general, the failure probability p_M to indicate the occurrence of a critical event for at least one time in M years of design life can be defined as (Serinaldi, 2015):

$$p_M = 1 - \prod_{j=1}^{M}(1 - p_j) = 1 - (F(x_d))^M \tag{8}$$

Similar to the multivariate RP concept, the failure probability (FP) in a multivariate context can also be characterized in "OR", "AND", and "Kendall" scenarios expressed by the following equations. For a given critical threshold $\boldsymbol{x}^* = \{x_1^*, x_2^*, ..., x_d^*\}$, the failure probabilities violating this critical value can be expressed as (Salvadori et al., 2016):

$$p_T^{OR} = 1 - (C(F_1(x_1^* \mid \gamma_1), F_1(x_2^* \mid \gamma_2), ..., F_d(x_d^* \mid \gamma_d) \mid \theta))^T \tag{9}$$

$$p_T^{AND} = 1 - (1 - \hat{C}(\overline{F}_1(x_1^* \mid \gamma_1), \overline{F}_2(x_2^* \mid \gamma_2), ..., \overline{F}_1(x_d^* \mid \gamma_d) \mid \theta))^T \tag{10}$$

$$p_T^{Kendall} = 1 - (P(C(F_1(x_1^* \mid \gamma_1), F_1(x_2^* \mid \gamma_2), ..., F_d(x_d^* \mid \gamma_d) \mid \theta) \le t))^T \tag{11}$$

where p_T^{OR}, p_T^{AND}, and $p_T^{Kendall}$ respectively denote the failure probability in "AND", "OR" and "Kendall" cases. T indicate the service time of the facilities under consideration.

Focusing on a bivariate case, the joint RP and the associate failure probability in "OR", "AND", and "Kendall" scenarios can be formulated as (Salvadori et al., 2007, 2011; Graler et al., 2013; Sraj et al., 2014; Serinaldi, 2015):

$$T_{u_1,u_2}^{OR} = \frac{\mu}{1 - C_{U_1U_2}(u_1, u_2 \mid \theta)} \tag{12}$$

$$T_{u_1,u_2}^{AND} = \frac{\mu}{1 - u_1 - u_2 + C_{U_1U_2}(u_1, u_2 \mid \theta)} \tag{13}$$

$$T_{u_1,u_2}^{Kendall} = \frac{\mu}{1 - P(C_{U_1U_2}(u_1^*, u_2^*) \le t)} \tag{14}$$

$$p_T^{OR} = 1 - (C_{U_1U_2}(u_1^*, u_2^* \mid \theta))^T \tag{15}$$

$$p_T^{AND} = 1 - (u_1^* + u_2^* - \hat{C}_{U_1U_2}(u_1^*, u_2^* \mid \theta))^T \tag{16}$$

$$p_T^{Kendall} = 1 - (P(C_{U_1U_2}(u_1^*, u_2^* \mid \theta) \le t))^T \quad (17)$$

where $u_1 = F_1(x_1 \mid \gamma_1), u_2 = F_2(x_2 \mid \gamma_2)$, $u_1^* = F_1(x_1^* \mid \gamma_1), u_2^* = F_2(x_2^* \mid \gamma_2)$, (x_1^*, x_2^*) means the bivariate threshold.

2.2. Uncertainty Quantification of Parameters by Bayesian Inference

From Section 2.1, it is noticed that in the multivariate risk analysis framework through copulas, parameters in both marginal distributions and copulas may produce significant impacts on the resulting multivariate RPs and failure probabilities. In particular, extensive uncertainties may be involved in the copula-based multivariate risk assessment framework due to: (i) the inherent uncertainty in the flooding process, (ii) uncertainty in the selection of appropriate marginal functions and copulas and (iii) statistical uncertainty or parameter uncertainty within the parameter estimation process (e.g. the availability of samples) (Serinaldi, 2013; Zhang et al., 2015). In this study, the inherent uncertainty in the copula-based multivariate model will be quantified by Bayesian analysis. The Bayesian approach has been widely applied for uncertainty quantification since it can incorporate various sources of information into a singly analysis through Bayes' theorem. Given the prior probability density and observations, the posterior distribution can be derived through Bayes' theorem, which is expressed as:

$$\pi(\theta \mid X) = \frac{L(\theta \mid X)\pi_0(\theta)}{\int L(\theta \mid X)\pi_0(\theta) d\theta} \quad (18)$$

where $\pi_0(\theta)$ signifies the prior parameter distribution, and $L(\theta \mid X)$ denotes the likelihood function. $\int L(\theta \mid X)\pi_0(\theta) d\theta$ is the normalization constant. $\pi(\theta \mid X)$ is the posterior probability density function. $X = (x_1, x_2, \ldots, x_d)$ is the observation vector.

Consider the multivariate distribution expressed by Equation (1), it is noticed that the posterior distribution for the parameters in marginal distributions (i.e. γ_i, $i = 1, 2, \ldots, d$)

and copula (i.e. θ) can be derived as follows:

$$\pi(\gamma_1,\gamma_2,...,\theta\,|\,\mathbf{x}) \propto L(\gamma_1,\gamma_2,...,\theta\,|\,\mathbf{x})\pi_0(\gamma_1,\gamma_2,...,\theta)$$
$$= L(\gamma_1,\gamma_2,...,\theta\,|\,\mathbf{x})\pi_0^1(\gamma_1)\pi_0^2(\gamma_2)...\pi_0^d(\gamma_d)\pi_0^c(\theta) \qquad (19)$$

where $L(\gamma_1,\gamma_2,...,\theta\,|\,\boldsymbol{x})$ is the likelihood function of the observation $\boldsymbol{x}$ ($\boldsymbol{x}$ = x_1, x_2, …, x_d), $\pi_0^i(\gamma_i)$ are the prior distributions for parameters in marginal distribution, and $\pi_0^c(\theta)$ indicates the prior for parameters in the copula.

For the probability distribution expressed by Equation (1), the corresponding probability density function (PDF) can be derived as (Aas et al., 2009):

$$f(x_1,...,x_d\,|\,\gamma_1,...,\gamma_d,\theta) = c(u_1,u_2,...,u_d\,|\,\theta)\,f_1(x_1\,|\,\gamma_1)\,...\,f_d(x_d\,|\,\gamma_d) \qquad (20)$$

where $c(.)$ indicate the copula density, $f_i(.)$ means the PDF of marginal distribution, and $u_i = f_i(x_i\,|\,\gamma_i)$. Consequently, the likelihood function $L(\gamma_1,\gamma_2,...,\theta\,|\,\boldsymbol{x})$ can be formulated as:

$$L(\gamma_1,\gamma_2,...,\theta\,|\,\mathbf{x}) = \prod_{j=1}^{n} c(u_1^j,u_2^j,...,u_d^j\,|\,\theta)\,f_1(x_1^j\,|\,\gamma_1),...,\ f_d(x_d^j\,|\,\gamma_d) \qquad (21)$$

where n is the total number of observations.

In spite of the likelihood function, the determination of prior distributions is an essential step in any Bayesian analysis, as shown in Equations (18) and (19). The key issues in setting up a priori distribution include (i) the information going into the prior distribution; (ii) the properties of the resulting posterior distribution (Gelman, 2002). In general, uninformative and informative priors are the two types of widely used prior distributions, in which uninformative priors are adopted for the situation of no information available for the prior and informative priors can provide some specific information about the variable. However, with well-identified parameters and large sample sizes, reasonable choices of prior distributions will have minor effects on posterior inferences (Gelman, 2002). In this study, the normal informative priors are assumed in terms of the parameters in marginal distribution and dependence structure.

In many situations, analytic solutions for Equation (18) are not possible. Thus, the Markov chain Monte Carlo (MCMC) techniques are used to approximate sampling realizations of the posterior distribution. A number of MCMC algorithms have been proposed in both statistical and hydrological literatures. In this study, the Metropolis-Hastings (MH) algorithm will be adopted for quantifying the posterior distributions of the parameters in the copula-based multivariate hydrologic risk framework. The MH algorithm, initially proposed by Metropolis et al. (1953) and then extended by Hastings (1970), is widely used in hydrologic context (e.g. Viglione et al., 2013; Zhang et al., 2016). In MH algorithm, a proposal density $q(.)$ is used to generate a new sample $\boldsymbol{\Theta}$', given the current state $\boldsymbol{\Theta}$. Such a new sample is either accepted or rejected through the Metropolis acceptance probability:

$$\alpha(\boldsymbol{\Theta},\boldsymbol{\Theta}') = \min[\frac{q(\boldsymbol{\Theta}\,|\,\boldsymbol{\Theta}')\pi(\boldsymbol{\Theta}'\,|\,\boldsymbol{x})}{q(\boldsymbol{\Theta}'\,|\,\boldsymbol{\Theta})\pi(\boldsymbol{\Theta}\,|\,\boldsymbol{x})},1] \tag{22}$$

where is the posterior distribution and $q(.)$ is the proposal density function. $\boldsymbol{\Theta}$ is the parameters to be quantified. In this study, the parameters in both marginal distributions and copulas are estimated simultaneously, and thus $\boldsymbol{\Theta} = (\gamma_1, \gamma_2, \ldots, \gamma_d, \theta)$. Based on the sampling realizations from MCMC, the uncertainty and credibility intervals can be analyzed. Also, the uncertainties in the resulting multivariate RP and failure probabilities will be characterized.

In the copula-based multivariate risk assessment framework, there are several options for both the marginal distributions (e.g. Lognormal, Pearson Type III, and generalized extreme value distributions) and copula model (i.e. Gaussian, Archimedean copulas). An essential step before inferencing hydrologic risk is to choose the most appropriate model (i.e. marginal distribution and copula) to match the observed flood data. In this study, the Deviance Information Criterion (DIC) will be employed to help choose the most appropriate model. The index of DIC is developed by Spiegelhalter et al. (2002), which is a specific measure designed for model selection under Bayesian inference and can be thought of as a Bayesian alternative to the standard Akaike information criterion (AIC) (Sarhadi et al., 2016). In the Bayesian inference through MCMC, the DIC value can be formulated as (Spiegelhalter et al., 2002):

$$DIC = D(\overline{\Theta}) + 2 p_D \tag{23}$$

where

$$D(\Theta) = -2 \log L(data \mid \Theta) \tag{24}$$

$$p_D = \overline{D} - D(\overline{\Theta}) \tag{25}$$

2.3. Uncertainty Partition through Multilevel Factorial Analysis

Due to the uncertainties existing in the parameters in the copula-based multivariate risk assessment model, the predictive risk (e.g. multivariate RP or failure probabilities) values for a give hazardous event also present uncertain features. However, in the copula-based multivariate risk assessment model, there are two kinds of parameters, including the parameters in marginal distributions describing randomness of attributes in the hazardous event and the parameters in copulas describing dependence structures among attributes. Moreover, uncertainties in these parameters interact among each other, leading to intensified uncertainty in predictive risk. Therefore, multilevel factorial analysis will be used to characterize the contributions of parameters in marginal distribution, copula and their interactions to the uncertainty in the resulting risk inference values.

In factorial analysis, an experimental design is employed to account for all combinations of the levels of factors to help visualize the single effects of factors with discrete values (or levels) and their interactive effects on a response variable (Wang et al., 2015). For instance, consider a copula-based bivariate risk assessment model which has two marginal distributions (*A* and *B*) and one copula (*C*). The parameters in the two marginal distributions are assumed to be respectively denoted as γ^A with *a* levels and γ^B with b levels and the parameter in the copula is denoted with θ^C with *c* levels. The predictive risk (denoted as *R*) of the copula model can be fitted in response to the parameters γ_A, γ_B, θ_C and replicates n, which can be expressed as:

$$R_{ijkl} = \mu + \theta_i^C + \gamma_j^A + \gamma_k^B + (\theta^C\gamma^A)_{ij} + (\theta^C\gamma^B)_{ik} + (\gamma^A\gamma^B)_{jk} + (\theta^C\gamma^A\gamma^B)_{ijk} + \varepsilon_{ijkl} \begin{cases} i = 1,2,...,c \\ j = 1,2,...,a \\ k = 1,2,...b \\ l = 1,2,...,n \end{cases} \tag{26}$$

where μ denotes the overall mean effect; $\theta_i^C, \gamma_j^A, \gamma_k^B$ respectively indicate the effect for parameter θ^C in the copula at the ith level, parameter γ^A in the first marginal distribution at the jth level, and parameter γ^B in the first marginal distribution at the kth level; $(\theta^C\gamma^A)_{ij}, (\theta^C\gamma^B)_{ik}$ and $(\gamma^A\gamma^B)_{jk}$ indicate interactions between factors θ^C and γ^A, θ^C and γ^B, as well as γ^A and γ^B, respectively; $(\theta^C\gamma^A\gamma^B)_{ijk}$ denotes the interaction of factors θ^C, γ^A and γ^B; ε_{ijkl} means the random error component.

Based on Equation (26), the total variability of the predictive risk can be decomposed into its components parts as follows:

$$SS_T = SS_{\theta^C} + SS_{\gamma^A} + SS_{\gamma^B} + SS_{\theta^C\gamma^A} + SS_{\theta^C\gamma^B} + SS_{\gamma^A\gamma^B} + SS_{\theta^C\gamma^A\gamma^B} + SS_e \tag{27}$$

and

$$SS_T = \sum_{i=1}^{c}\sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{l=1}^{n} R_{ijkl}^2 - \frac{R_{....}^2}{abcn} \tag{28}$$

$$SS_{\theta^C} = \frac{1}{abn}\sum_{i=1}^{c} R_{i...}^2 - \frac{R_{....}^2}{abcn} \tag{29}$$

$$SS_{\gamma^A} = \frac{1}{bcn}\sum_{j=1}^{a} R_{.j..}^2 - \frac{R_{....}^2}{abcn} \tag{30}$$

$$SS_{\gamma^B} = \frac{1}{acn}\sum_{k=1}^{b} R_{..k.}^2 - \frac{R_{....}^2}{abcn} \tag{31}$$

$$SS_{\theta^C\gamma^A} = \frac{1}{bn}\sum_{i=1}^{c}\sum_{j=1}^{a} R_{ij..}^2 - \frac{R_{....}^2}{abcn} - SS_{\theta^C} - SS_{\gamma^A} \tag{32}$$

$$SS_{\theta^C\gamma^B} = \frac{1}{an}\sum_{i=1}^{c}\sum_{k=1}^{b} R_{i.k.}^2 - \frac{R_{....}^2}{abcn} - SS_{\theta^C} - SS_{\gamma^B} \tag{33}$$

$$SS_{\gamma^A\gamma^B} = \frac{1}{cn}\sum_{j=1}^{a}\sum_{k=1}^{b} R_{.jk.}^2 - \frac{R_{....}^2}{abcn} - SS_{\gamma^A} - SS_{\gamma^B} \tag{34}$$

$$SS_{\theta^C\gamma^A\gamma^B} = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{a}\sum_{k=1}^{b} R_{ijk.}^2 - \frac{R_{....}^2}{abcn} - SS_{\theta^C} - SS_{\gamma^A} - SS_{\gamma^B} - SS_{\theta^C\gamma^A} - SS_{\theta^C\gamma^B} - SS_{\gamma^A\gamma^B} \tag{35}$$

$$SS_e = \sum_{i=1}^{c}\sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{l=1}^{n} R_{ijkl}^2 - \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{a}\sum_{k=1}^{b} R_{ijk.}^2 \tag{36}$$

where $R_{ijk.} = \sum_{l=1}^{n} R_{ijkl}$, $R_{ij..} = \sum_{k=1}^{b}\sum_{l=1}^{n} R_{ijkl}$, $R_{.jk.} = \sum_{i=1}^{c}\sum_{l=1}^{n} R_{ijkl}$, $R_{i.k.} = \sum_{j=1}^{a}\sum_{l=1}^{n} R_{ijkl}$, $R_{i...} = \sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{l=1}^{n} R_{ijkl}$, $R_{.j..} = \sum_{i=1}^{c}\sum_{k=1}^{b}\sum_{l=1}^{n} R_{ijkl}$, $R_{..k.} = \sum_{i=1}^{c}\sum_{j=1}^{a}\sum_{l=1}^{n} R_{ijkl}$ $R_{....} = \sum_{i=1}^{c}\sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{l=1}^{n} R_{ijkl}$. Then the contributions of parameter uncertainties in marginal distributions and dependence structures can be calculated as:

(1) Contributions of parameters in marginal distributions A and B

$$\eta_A = SS_{\gamma^A} / SS_T \tag{37}$$

$$\eta_B = SS_{\gamma^B} / SS_T \tag{38}$$

(2) Contribution of parameter in the dependence structure

$$\eta_C = SS_{\theta^C} / SS_T \tag{39}$$

(3) Contributions of parameter interactions

$$\eta_{AC} = SS_{\theta^C\gamma^A} / SS_T \tag{40}$$

$$\eta_{BC} = SS_{\theta^C\gamma^B} / SS_T \tag{41}$$

$$\eta_{AB} = SS_{\gamma^A\gamma^B} / SS_T \tag{42}$$

$$\eta_{ABC} = SS_{\theta^C\gamma^A\gamma^B} / SS_T \tag{43}$$

(4) Contribution of internal variability

$$\eta_e = SS_e / SS_T \tag{44}$$

3. Applications

The proposed FBC approach is applied for hydrologic risk analysis at three locations in China, one at the Xingxi River and two at the Wei River. The detailed descriptions for these two catchments are provided in the Supplementary Materials. Observed daily streamflow data Xingshan station (at Xiangxi River) and Xianyang and Zhangjiashan gauging stations (at Wei River) are applied for hydrologic risk analysis. Figure 2 show the locations of these three gauging stations. Although a flood is generally characterized by its peak, volume and duration, the multivariate flood risk in terms of peak and volume will be applied to demonstrate the applicability for the proposed FBC method. Other multivariate risk indices for peak-duration, volume-duration and peak-volume-duration can similarly be characterized by the developed FBC method. Based on the daily stream flow data, the flood peak applied is defined as the maximum daily flow over a period and the associated flood volume is considered as the cumulative flow during the flood period. In current study, the flood characteristics are obtained based on the annual scale. This means that one flood event is identified in one year. The detailed method to identify the flood peak and the associated flood volume can be found in studies by Yue (2000, 2001). Table 1 shows some descriptive statistics values of the considered variables (peak discharge, Q; hydrograph volume, V). In detail, the daily discharge data from 1961 to 2010 are used to analyze the potential flooding risk in Xiangxi River, which means 50 flood peak and volume values (i.e. one flood peak and volume in each year) are generated. Similarly, 47 and 55 flood events are characterized at the Xianyang and Zhangjiashan station, respectively.

Place Figure 2 and Table 1 here

----

4. Results Analysis

4.1. Parameter Estimation and Model Selection

There are a number of potential models for both marginal distributions and dependent structures. In this study, three models including generalized extreme value (GEV), lognormal (LN) and Pearson Type III (PIII) distributions are selected to fit the distributional characteristics of flood peak and volume at the Xingshan, Zhangjiashan and Xianyang stations. Meanwhile, three Archimedean copulas involving Joe, Gumbel and Frank copula are applied to reflect the dependence structure between those two flood attributes (i.e. peak and volume). The parameters in both marginal distributions and copula are quantified through MH-based MCMC approach and the most appropriate model is determined by the values of DIC.

For each gauge station, the combinations of marginal distributions and copula functions will lead to a total number of 27 potential risk assessment models. For each model, the MH-based MCMC algorithm is run for 50,000 iterations in which the first 30% samples are neglected as burn-in and the rest ones are used to quantify the posterior distributions of model parameters. The Geweke's diagnostic is applied to guarantee the convergence of the Markov chain for each parameter (the absolute value of test statistic is less than 1.96) (Plummer et al., 2015). The associated DIC values are calculated based on the posterior samples in which the model with a minimum DIC value is chosen as the most appropriate one for further risk inference.

Table 2 presents the results for model selection for the three streamflow gauge stations (i.e. Xinshan, Zhangjiashan and Xianyang). The results indicate that for each flood attribute (i.e. peak and volume), the lognormal distribution will be applied to quantify its distributional characteristic. In particular, the lognormal distribution will be used for all the two flood attributes at all three gauge stations. In terms of the dependence structure between flood peak and volume, the Gumbel copula will be used for the Xiangxi River (i.e. Xingshan station), while the Frank copula is chosen for Wei River (i.e. Zhangjiashan and Xianyang stations). Figure 3 shows the posterior distributions for the parameters in the copula-based risk assessment model at the three gauge stations. The first row indicates the parameter posteriors for the lognormal distribution models for flood peak at the three stations while the second row presents the posterior distributions for the flood volume models. The last row in Figure 3 shows the parameter posteriors in the copula functions. The results in Figure 3 suggest that the uncertainty in model parameters can be well reflected by the MCMC algorithm with relative small deviations. Such results can also be demonstrated by the 95% predictive intervals (PIs) for each parameter as presented in Table 2.

Place Figure 3 and Table 2 here

4.2. Uncertainty in Risk Inferences

Parameter uncertainties in both marginal distributions and the dependence structure will also lead to uncertainty in the risk inference results. Figure 4 describes the inferred flood peak and volume generated based on the lognormal model. Since parameters in all lognormal models have some degrees of uncertainty (as shown in Figure 3 and Table 2), the inferred flood peak and volume values with specific return periods exhibit obvious uncertainties. Particularly, these uncertainties would be more extensive as the increase in the predefined return period.

Place Figure 4 here

In addition to uncertainties in predictive flood peak and volume, parameter uncertainties in both marginal distributions and the copula function also lead to imprecise values for joint risks of peak and volume. Figure 5 describes uncertainties for joint RP in OR at the three stations. It is observed that the predictive joint RP in OR has small uncertainty for a small flood event (i.e. small peak or small volume), while considerable uncertainties exist in the predictive joint RP of OR even for a flood with an actual joint RP of 20 years. For the predictive joint RP in AND, more extensive uncertainties exist than that for the joint RP in OR, as shown in Figure S1 in the Supplementary Materials. Noticeable uncertainties exist in the predictive joint RP of AND even for a minor flood event with a 5-year joint RP of AND. For some extreme flood events (e.g. with a 200-year joint RP of AND), the predictive joint RP in AND can be remarkably large. For the joint RP in Kendall, the uncertainties in the predictive values are not as remarkable as those for joint RP in AND, as presented in Figure S2 in the Supplementary Materials. However those uncertainties are still noticeable even for a moderate flood event with a 50-year joint RP in Kendall.

Place Figures 5 here

In additional to the joint RP in AND, OR and Kendall, the FP provides more consistent way for multivariate hydrological or environmental risks (Serinaldi, 2015; Salvodor et al., 2016). The formulations for calculating these FPs are expressed by Equations (9) – (11) and Equations (15) – (17). In this study, the critical thresholds for flood peak and volume are inferred based on their corresponding probability distributions with a predefined return period of 500 (i.e. p = 0.998) and the service time of one hydraulic infrastructure (e.g. a dam or river levee) ranges from 30 to 100 years. The associated FPs in OR, AND and Kendall, corresponding to the above critical peak and volume thresholds as well as service time scenarios, are shown in Figure 6. It is observed that, as a result of uncertain parameters in both marginal distributions and the dependence structure, noticeable uncertainties exist in the predictive FPs of OR, AND and Kendall. More specifically, the degree of uncertain of FPs would increase significantly if the service time of one hydraulic infrastructure increases.

Place Figure 6 here

4.3. Interactions of parameter uncertainties

Great uncertainties exist in the inferred risk values (i.e. joint RP and FPs) due to imprecise estimations for the parameters in the copula-based risk assessment model. However, one unclear issue is that how the parameter uncertainties and their interactions impact the risk inference of the copula-based risk model. Therefore, a multilevel factorial design (expressed as Equations (26) – (36)) is proposed to characterize the main and interactive effects of parameters in marginal distributions and copula function on the resulting risk inference values. In detail, a 3^5 factorial design is proposed in which the five parameters considered as the factors, in which two of them (denoted as *P_par1*, *P_par2*) are the parameters in the lognormal distribution for flood peak, two (denoted as *V_par1*, *V_pa*r2) are the parameters in the lognormal distribution for flood volume, and the last one is the parameter (i.e. *Cop_par*) in the copula function. Each factor has three levels (i.e. 0.05, 0.5, and 0.95) identifies as the 5%, 50% and 95% quantiles of its posterior samples from MCMC. Three responses are considered in the factorial design, which are corresponds to the three failure probabilities (i.e. OR, AND, and Kendall). A multi-way analysis of variance (ANOVA) is further employed to identify the statistical significance of all parameters and their interactions in the copula-based risk assessment model.

Figure 7 presents the main effect plots and full interactions plot matrices for parameters on the FP in OR at Xingshan, Xianyang and Zhangjiashan stations. It is noticed that both the main effect plots and the interaction plot matrices at the three stations have similar patterns, implying that the parameters' individual and interactive effects on the FP are nearly independent with the location of one gauge station. More specifically, the changes of the two scale parameters (i.e. *P_par2* and *V_par2*) lead to more changes in response (i.e. FP in OR) than the changes of the two location parameters (i.e. *P_par1* and *V_par2*). This indicates that the scale parameter in lognormal distribution has a more effect on the inferred FP of OR than the location parameter. Moreover, the resulting failure probability does not have a visible change as the copula parameter change from its low level to its high level, suggesting that the parameter in the copula function may has insignificant effect on the prediction of failure probability in OR. For the interactive effects among the five parameters, the full interactions plot matrices show that the interactive curves between the copula parameter (i.e. Cop_par) and the location parameter of peak (i.e. *P_par1*) are parallel at the three levels, indicating an insignificant interaction of these two parameters on the inferred failure probability of OR. Similar characteristics are also observed for the interactions between the copula parameter and the other three parameters (i.e. *P_par2*, *V_par1*, *V_par2*). Other interactive curves are observed to be intersected at the three levels, implying significant interactive effects of those parameters on the results risk inference. Table 3 provides the results of ANOVA table for the failure probability in OR. The results indicate statistical insignificance for AE, BE, CE and DE, which is consistent with the full interaction plot matrices. The individual effect of the copula parameter (i.e. *Cop_par*) is statistical insignificant at the Xianyang and Zhangjiashan stations, while such an effect is statistical significant at the Xingshan station. However, this parameter leads to less sum of squares than the other four parameters at all three cases, implying a least main effect among those five parameters, which is also observed from Figure 7.

Place Figure 7 and Table 3 here

In terms of the FP in AND, the results in Figure S3 in the Supplementary Materials indicate that both individual effects of the parameters and their interactions have similar features for all three stations, which is also observed in Figure 7. Also, for the individual effects, the two scale parameters (i.e. *P_par2* and *V_par2*) lead to steeper lines than the lines corresponding to the two location parameters (i.e. *P_par1* and *V_par1*), suggesting more significant effects of the scale parameters than those from the location parameters. It is also noticed that among the five parameters, the copula parameter (i.e. *Cop_par*) contributes least individual effects on the variation of FP in AND. Furthermore, significant interactive effects are observed among the parameters in the two marginal distributions (i.e. *P_par1*, *P_par2*, *V_par1*, and *V_par2*) which are indicated by intersecting lines in Figure S3. Some interactions also occur between the copula parameter and the parameters in marginal distributions, which are different from those for the FP in OR described in Figure 7. These interactions may probably because that, besides the joint probability values, the cumulative probabilities of flood peak and volume are also used to derive the FP in AND (i.e. Equations (10) and (16)). This leads to more chances for occurrence of visible interactions between the copula parameter and others. The above findings can be further demonstrated by the results of ANOVA for failure probability in AND. As presented in Table S1 in the Supplementary Materials, the copula parameter results into least sum of squares, suggesting least individual effects among the five parameters. Also, statistical significance for parameter interactions can be observed except to the interactions of AE and CE.

Figure S4 in the Supplementary Materials presents the main effects plot and full interactions plot matrices for parameters on the FP in Kendall at the three gauge stations, and Table S2 gives the associate ANOVA results. It can be found that the results of failure probability in Kendall have similar characteristics at the three sites, which are also found for the previous two failure probabilities (i.e. AND, OR). Also, the copula parameter results into least sloped curve and least sum of squares, implying least main effect on the inferred failure probability in Kendall. For the interactive effects among the five parameters, intersection lines are observed for all pairs of parameters, suggesting apparent interactive effects, which are also demonstrated by the ANOVE results in Table S2.

4.3 Contributions of Uncertainty Sources

In the copula-based risk assessment model, parameters in marginal distributions pose significant individual and interactive effects on the inferred failure probabilities, while the parameter in the copula function has least main effects (e.g. statistical insignificance in some cases) and the interactions for this parameter and others are sometime insignificant. As a result of parameter uncertainties, the predictive failure probabilities exhibit noticeable uncertainties as shown in Figure 6. Such uncertainties become more significant with the increase in service time. However, two more issues to be explored are that (i) how much these parameters contribute to the variation of the inferred risk values and (ii) do these contributions change significantly for the failure probabilities with different service time scenarios. Consequently, further multi-level factorial analysis is conducted in response to failure probabilities with multiple service time scenarios in order to identify (i) the contributions of parameters and their interactions on the variations or uncertainties in predictive failure probabilities, and (ii) how these contributions change with the variation in service time. In detail, to get more reliable quantification, five levels are considered for each parameter which is identified as 5%, 25%, 50%, 75% and 95% quantiles of the parameter's posterior samples. Four service time scenarios, namely 30, 50, 70, and 100 years, are under consideration to answer the change of one contribution with varied service time scenario. Moreover, only a single replicate is conducted for the factorial analysis, in which three and higher-way interactions are combined to give an estimate of internal error in the associated ANOVA. Finally, the contributions of parameters in marginal distributions and copula are characterized based on Equations (37) – (39), and the interactive effects of these parameters are combined together by summing results of Equations (40) – (44).

Figure 8 shows detailed contributions of model parameters on uncertainty in predictive failure probabilities of OR at Xingshan, Xianyang and Zhangjiashan stations. It can be observed that, even though some discrepancies exist for contributions of model parameters at different stations, the results show similar features, in which the scale parameters (i.e. *P_par2*, and *V_par2*) give more contributions (larger than 25%) in risk inference than of the location parameters (i.e. *P_par1* and *V_par1*), and the parameter in copula function pose least impact (less than 1%). However, in terms of the interaction of model parameters, parameter interactions provide more impact (more than 14%) at the Xiangshan station than that (less than 8%) at the Xianyang and Zhangjiashan stations. This may be probably because that the Xingshan station is located in the Xingxi River basin with a northern subtropics climate while the Xianyang and Zhangjiashan stations are located Wei River basin experienced a semi-arid and sub-humid continental monsoon climate. Moreover, the detailed contribution of one parameter would not change significantly for different service time scenarios. For instance, as the service time changes from 30 to 100 years, the contribution of *P_par1* at Xingshan station ranges from 7.93% to 8.42%.

For the failure probability in AND, the parameters in marginal distributions and the copula function have different effects with those parameters' impacts on the failure probability in OR. As presented in Figure S5 in the Supplementary Materials, the scale parameters (i.e. i.e. *P_par2*, and *V_par2*) in marginal distributions also pose significant impacts on uncertainties in the failure probabilities in AND. However, the detailed contributions are less than those parameters on the failure probabilities in OR shown in Figure 8. For instance, the scale parameters contribute 30.51% and 26.91% respectively to the predictive uncertainty in failure probability in AND with a service time of 30-year at Xingshan Station, while these two parameters give contributions of 39% and 28.56% respectively for the failure probability in OR. Such decreases of individual effects of scale parameters are mainly due to the remarkable increasing effects of interactions, which increase from about 15% to more than 25% at Xingshan station and from less than 10% to more than 30% at Xianyang and Zhangjishan stations. The significant increase of the interactive effect also leads to visible decrease in the contributions of location parameters (i.e. *P_par1* and *V_par1*). The contributions of the copula parameter are still neglectable even though they increase slightly from less than 0.1% for failure probabilities in OR to about 1% for the failure probabilities in AND.

For the uncertainty partition in the failure probability of Kendall, it can be noticed from Figure S6 in the Supplementary Materials that the scale parameters have most significant contributions on the failure probabilities in Kendall, followed by the interaction of model parameters, the location parameters, and the copula parameter. Moreover, such contribution partition does not change explicitly in response to the change in service time scenarios. However, compared with parameters' contributions to the failure probabilities in OR and AND, the contributions of parameters in marginal distributions (i.e. *P_par1*, *P_par2*, *V_par1*, and V*_par2*) to the failure probabilities of Kendall are generally larger than those contributions to failure probability in AND but less than the contributions to failure probability in OR. In comparison, the contribution of parameters' interaction to the failure probability in Kendall is less than the contribution to the failure probability in AND but larger than its contribution to the failure probability in OR. This is probably due to the differences in calculation of failure probabilities in OR, AND, and Kendall (i.e. Equations (9) – (11) and Equations (15) – (17)), in which parameters' interaction has more chance to pose a significant impact on the variations of the inferred risks in OR and AND than the risk in OR.

5. Conclusions

In this study, a factorial Bayesian copula (FBC) approach has been proposed to quantify parameters' uncertainties, reveal individual and interactive effects of parameters and further characterize contributions of these parameters on the inferred risk values. The developed FBC approach integrates copula-based risk assessment model, Bayesian inference and factorial analysis into a general framework. In detail, a Bayesian-based Markov chain Monte Carlo approach is employed to quantify parameter uncertainties in the copula-based risk inference model; multi-level factorial analysis is proposed to reveal individual and interactive effects of model parameters on risk inference, and the associated analysis of variance (ANOVA) is further proposed to identify the contributions of model parameters and their interaction on the predictive risk values.

To illustrate the applicability of the proposed FBC approach, flood observations at three gauge stations have been used to reveal parameters' uncertainty and their contributions to the joint risk of flood peak and volume. The joint RPs and the associated FPs in OR, AND and Kendall are considered as the risks of interest. Based on those case studies, some findings can be concluded:

1. Parameter uncertainty is one of the unavoidable factors to be well identified in multivariate risk analysis, and imprecise parameters in copula-based models can lead to great uncertainties in all inferred joint RPs and FPs in OR, AND and Kendall.

2. For different risk indices, the main and individual effects of parameters have some different features. In general, the scale parameters in the lognormal distributions of peak and volume have more individual effects than the location parameters for all three failure probabilities, followed by the location parameters and the parameter in copula function. But all interactive effects of the copula parameter and other parameters in marginal distributions are generally statistical insignificant for the FP in OR, while some of them are statistical insignificant for the FP in AND, and all of them are significant for the FP in Kendall. Moreover, such features are almost independent with the location of a gauge station.

3. For the detailed contributions, parameters in marginal distributions of peak and volume pose most contributions for the FP in OR, followed by their contributions to FPs in Kendall and AND. In contrast, the parameters' interaction has a more impact for FP in AND than its contributions to Kendall and OR. The copula parameter has

least contributions for all three FPs even through it increases from less than 0.1% to about 2% for FPs of OR to Kendall. The above contribution characterization does not change visibly for FPs with different service time scenarios.

The presence of uncertainties would pose significant impact on risk inferences within both univariate and multivariate contexts. Many studies have been reported to quantify uncertainties in hydrological risk analysis (e.g. Serinaldi, 2013; Zhang et al., 2015; Dung et al., 2015; Fan et al., 2018). However, as an extension of previous studies, the major contributions in this study is that the proposed FBC method cannot effectively quantify parameter uncertainties in the copula-based multivariate risk inference model, but also characterize the individual and interactive effects of those uncertainties on the resulting risk inferences. Also, the develop FBC approach can help track the major contributors (e.g. parameter uncertainties in marginal distributions) to the resulting uncertainties in risk inferences. Such results would be helpful to find potential pathways for uncertainty reduction in hydrological risk inferences.

The applicability of FBC has been illustrated through multivariate flood risk inference under consideration of the dependence between flood peak and volume. Nevertheless, this method can also be applied for other risk assessment issues such the compound hydroclimatic extremes (e.g. drought and heat waves (Sun et al., 2019), soil moisture and precipitation (AghaKouchak, 2015)), water quality (Shi and Xia, 2017), air pollution (Sak et al., 2017), and so on.

Acknowledgement

This research was supported by the National Key Research and Development Plan (2016YFC0502800, 2016YFA0601502), the Natural Sciences Foundation of China (51190095, 51225904), and the Natural Science and Engineering Research Council of Canada.

Appendix A. Abbreviation

AIC	Akaike information criterion
AND	All elements in the extreme events should exceed the corresponding thresholds
ANOVA	Analysis of variance
FBC	Factorial Bayesian copula
DIC	Deviance Information Criterion
FP	Failure probabilities
GEV	Generalized extreme value distribution
LN	Lognormal distribution
MCMC	Markov chain Monte Carlo
MH	Metropolis-Hastings algorithm
OR	At least one element surpass the predefined threshold
PDF	Probability density function
PI	Predictive interval
PIII	Pearson Type III distribution
RP	Return period
TGR	Three Gorges Reservoir

References

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2009), Pair-copula constructions of multiple dependence, Insurance: Mathematics and Economics, 44, 182–198.

AghaKouchak, A. (2015), A multivariate approach for persistence-based drought prediction: Application to the 2010–2011 East Africa drought, Journal of Hydrology, 526, 127–135

Chebana F., and Ouarda T.B.M., (2011). Multivariate quantiles in hydrological frequency analysis. Environmetrics, 22(1), 63-78.

De Michele C, Salvadori G (2003) A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. Journal of Geophysical Research, 108(D2), 4067, doi:10.1029/2002JD002534.

Dung N.V., Merz B., Bardossy A., Apel H., (2015). Handling uncertainty in bivariate quantile estimation – An application to flood hazard analysis in the Mekong Delta. Journal of Hydrology, 527, 704-717.

Fan Y.R., Huang W.W., Huang G.H., Huang K., Zhou X., (2015a). A PCM-based stochastic hydrologic model for uncertainty quantification in watershed systems. *Stochastic Environmental Research and Risk Assessment*, 29(3) 915-927.

Fan Y.R., Huang W.W., Li Y.P., Huang G.H., Huang K., (2015b). A coupled ensemble filtering and probabilistic collocation approach for uncertainty quantification of hydrological models. Journal of Hydrology, 530, 255-272.

Fan Y.R., G.H., Huang, K., Huang, (2018). Uncertainty Analysis for Multivariate Hydrologic Risk in the Xiangxi River in Three Gorges Reservoir Area, China, Engineering, 4(5), 617-626.

Fan Y.R., Huang W.W., Huang G.H., Huang K., Li Y.P., Kong X.M., (2016a). Bivariate hydrologic risk analysis based on a coupled entropy-copula method for the Xiangxi River in the Three Gorges Reservoir area, China, 125 (1-2), 381-397

Fan Y.R., Huang W.W., Huang G.H., Li Y.P., Huang K., (2016b). Hydrologic Risk Analysis in the Yangtze River basin through Coupling Gaussian Mixtures into Copulas. Advances in Water Resources, 88, 170-185.

Fan Y.R., Huang G.H., Baetz B.W., Li Y.P., Huang K., (2017). Development of a Copula‐based Particle Filter (CopPF) Approach for Hydrologic Data Assimilation under Consideration of Parameter Interdependence. Water Resources Research, 53(6), 4850-4875.

Favre, A.-C., S. El Adlouni, L. Perreault, N. Thiémonge, and B. Bobée (2004), Multivariate hydrological frequency analysis using copulas, Water Resources Research 40, W01101, doi:10.1029/2003WR002456

Gelman A., Prior distribution. In: Abdel H. El-Shaarawi and Walter W. Piegorsch, editors Encyclopedia of Environmetrics, John Wiley & Sons, Vol. 3,2002, p. 1634–1637.

Genest C, Favre AC. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4): 347–368.

Graler B., van den Berg M.J., Vandenberghe S., Petroselli A., Grimaldi S., De Baets B., Verhoest N.E.C., (2013). Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrology and Earth System Sciences* 17: 1281–1296.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their application. Biometrika 57, 97–109.

Huang K., Dai L.M., Yao M., Fan Y.R., Kong X.M., (2017). Modelling dependence between traffic noise and traffic flow through an entropy-copula method, Journal of Environmental Informatics, 29(2), 134 – 151, doi:10.3808/jei.201500302

Huang G.W., Liu H., Li X., and Ma M., (2019). Exploring Drivers of Nitrate Contamination of Drinking Water in an Arid Region of China. Journal of Environmental Informatics, 32(2), 105-112.

Joe, H. (2014), Dependence Modeling With Copulas, Chapman and Hall, London, U. K.

Kao S.C., Govindaraju R.S., (2010). A copula-based joint deficit index for droughts. *Journal of Hydrology*, 380, 121-134.

Karmakar S., Simonovic S.P., (2009). Bivariate flood frequency analysis. Part 2: a copula-based approach with mixed marginal distributions. *Journal of Flood Risk Management*, 2, 32-44.

Kidson R., Richards K.S., (2005). Flood frequency analysis: assumption and alternatives. *Progress in Physical Geography*, 29(3), 392-410.

Kong X.M., Huang G.H., Fan Y.R., Li Y.P., (2015). Maximum entropy-Gumbel-Hougaard copula method for simulation of monthly streamflow in Xiangxi river, China. *Stochastic Environmental Research and Risk Assessment*, 29, 833-846

Lee T, Salas JD. 2011. Copula-based stochastic simulation of hydrological data applied to Nile River flows. *Hydrology Research* 42(4): 318–330.

Li Z., Huang G.H., Fan Y.R., Xu J.L., (2015). Hydrologic Risk Analysis for Nonstationary Streamflow Records under Uncertainty. Journal of Environmental Informatics, 26(1), 41-51.

Lindenschmidt K.E., and Rokaya P., (2019). A Stochastic Hydraulic Modelling Approach to Determining the Probable Maximum Staging of Ice-Jam Floods. Journal of Environmental Informatics, 34(1) 45-54

Merz B, Thieken AH (2005) Separating natural and epistemic uncertainty in flood frequency analysis. *Journal of Hydrology* 309(1–4):114–132

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E., (1953). Equation of state by fast computing machines. Journal Chemical Physics 21, 1087–1092

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

Plummer, M., Best, N., Cowles, K., Vines, K., Sarkar, D., Bates, D., and Almond, R.: Package coda, URL http://cran.r-project.org/web/packages/coda/coda.pdf, accessed January 25, 2015, 2015.

Requena A., Mediero L., Garrote L., (2013). A bivariate return period based on copulas for hydrologic dam design: Accounting for reservoir routing in risk estimation. Hydrology and Earth System Sciences 17(8):3023-3038

Salvadori G., De Michele C., Durante F., (2011). On the return period and design in a multivariate framework. *Hydrology and Earth System Sciences* 15: 3293–3305

Salvadori, G., F. Durante, and C. De Michele (2013), Multivariate return period calculation via survival functions, Water Resources Research, 49, 2308–2311, doi:10.1002/wrcr.20204

Salvadori G., De Michele C., Kottegoda N.T., Rosso R., (2007). *Extremes in Nature: an Approach using Copula*. Springer: Dordrencht; 292

Salvadori, G., F. Durante, C. De Michele, M. Bernardi, and L. Petrella (2016), A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities, Water Resources Research, 52, 3701–3721, doi:10.1002/2015WR017225.

Sak H., Yang G., Li B., and Li W., (2017). A copula-based model for air pollution portfolio risk and its efficient simulation. Stochastic Environmental Research and Risk Assessment, 31(10), 2607-2616.

Sarhadi, A., D. H. Burn, M. C. Ausín, and M. P. Wiper (2016), Time-varying nonstationary multivariate risk analysis using a dynamic Bayesian copula, Water Resources Research, 52, 2327–2349, doi:10.1002/2015WR01852

Serinaldi, F., 2013. An uncertain journey around the tails of multivariate hydrological distributions. Water Resources Research 49, 6527–6547, doi:10.1002/wrcr.2053.

Serinaldi, F. (2015), Dismissing return periods!, Stochastic Environmental Research and Risk Assessment, 29(4), 1179–1189, doi:10.1007/s00477-014-0916-1.

Shi W., Xia J., (2017). Combined risk assessment of nonstationary monthly water quality based on Markov chain and time-varying copula. Water Science and Technology, 75(3-4), 693-704.

Song S., Singh V.P., (2010). Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data. *Stochastic Environmental Research and Risk Assessment*, 24(3), 425-444.

Spiegelhalter D., Gest N.G., Carlin B.P., Van Der Linde, A., (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B, 64(4), 583-639.

Sraj M., Bezak N., Brilly M. (2014). Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava River. *Hydrological Processes*, 29(2), 225-238, DOI: 10.1002/hyp.10145

Sun C.X., Huang G.H., Fan Y.R., Zhou X., Lu C., Wang X.Q. (2019). Drought Occurring with Hot Extremes: Changes Under Future Climate Change on Loess Plateau, China. Earth’s Future, 7, 587–604

The European Parliament and The Council (2007), Directive 2007/60/EC: On the assessment and management of flood risks, Official Journal of the European Union, 116 pp

Vandenberghe S, Verhoest NEC, De Baets B. (2010). Fitting bivariate copulas to the dependence structure between storm characteristics: a detailed analysis based on 105 year 10 min rainfall. *Water Resources Research*, 46. DOI: 10.1029/2009wr007857.

Viglione, A, R. Merz, J.L. Salinas, G. Blschl (2013). Flood frequency hydrology: 3. A Bayesian analysis, Water Resources Research, 49, 675-692.

Wang S., Huang G.H., Huang W., Fan Y.R., Li Z., (2015). A fractional factorial probabilistic collocation method for uncertainty propagation of hydrologic model parameters in a reduced dimensional space. Journal of Hydrology, 529, 1129-1146.

Wu H., Chen B., Snelgrove K., Lye L.M., (2019). Quantification of Uncertainty Propagation Effects during Statistical Downscaling of Precipitation and Temperature to Hydrological Modeling. Journal of Environmental Informatics, 34(2), 139-148.

Xu Y., Huang G.H., Fan Y.R, (2016). Multivariate flood risk analysis for Wei River. Stochastic Environmental Research and Risk Assessment, 31 (1), 225-242 doi: 10.1007/s00477-015-1196-0

Yue S., (2000). The bivariate lognormal distribution to model a multivariate flood episode. Hydrological Processes, 14(14), 2575-2588.

Yue S., (2001). A bivariate gamma distribution for use in multivariate flood frequency analysis. Hydrological Processes, 15(6), 1033-1045

Zhang J.L., Li Y.P., Huang G.H., Chen X., Bao A.M., (2016). Assessment of parameter uncertainty in hydrological model using a Markov-Chain-Monte-Carlo-based multilevel-factorial-analysis method, Journal of Hydrology, 538, 471-486.

Zhang L., Singh V.P., (2006). Bivariate flood frequency analysis using the copula method. *Journal of Hydrologic Engineering*, 11, 150-164.

Zhang L, Singh VP (2007) Bivariate rainfall frequency distributions using Archimedean copulas. *Journal of Hydrology* 332(1–2):93–109

Zhang Q., Xiao M.Z., Singh V.P., (2015). Uncertainty evaluation of copula analysis of hydrological droughts in the East River basin, China. Global and Planetary Change, 129, 1-9.

Captions of Tables

Table 1. Flood characteristics for different stations

Table 2. Model selection results by DIC and parameter estimation by MCMC

Table 3. ANOVA table for failure probability in OR

Captions of Figures

Figure 1. Framework of the proposed factorial Bayesian copula approach

Figure 2. The location of the studied watersheds. Wei River is the largest tributary of Yellow river, with a drainage area of 135,000 km2. The historical flood data from Xianyang and Zhangjiashan stations on the Wei River are analyzed through the proposed FBC approach. The Xiangxi River is located in the Three Gorges Reservoir area with a drainage area of 3,200 km2. The historical data from Xingshan station is used in this study.

Figure 3. Posterior distributions of parameters in the copula models for different gauge stations

Figure 4. Uncertainty in flood peak and volume inference at the three gauge stations: both inferred flood peak and volume contain extensive uncertainties due to randomness in model parameters; such uncertainties increase significantly the increase in predefined return period

Figure 5. Uncertainty quantification of the joint return period in “OR”: the blue dash lines indicate the predictive means and the red dash lines indicate the 5% and 95% quantiles.

Figure 6. Uncertainty quantification for the failure probabilities in OR, AND and Kendall at the Xingshan, Xianyang and Zhangjiashan stations: Considerable uncertainties are observed in all failure probabilities and such uncertainties would increase with the increase in service time

Figure 7. Main effects plot and full interactions plot matrix for parameters on the failure probability in OR at the three gauge stations

Figure 8. Contributions of model parameters on uncertainty in predictive failure probabilities of OR at the three gauge stations

Table 1. Flood characteristics for different stations

Station name	period		flood variable	
			Peak (m^3/s)	Volume (m^3/(s day))
		Minimum	91	72
Xingshan	1961-2010	Median	451.5	713.3
		Maximum	1050	2430
		Minimum	139	317
Xianyang	1960-2006	Median	1350	2491
		Maximum	12380	17802
		Minimum	217	303.7
Zhangjiashan	1958-2012	Median	775	1365.3
		Maximum	3730	7576.1

Table 2. Model selection results by DIC and parameter estimation by MCMC

Stations	Attributes	Selected Distribution		Parameters			DIC
				α	β	θ	
Xingshan	Peak	LN	Mean	6.13	0.51	-	1403.77
			95% PI	[5.99, 6.27]	[0.43, 0.60]	-	
	Volume	LN	Mean	6.67	0.62	-	
			95% PI	[6.49, 6.85]	[0.52, 0.72]	-	
	Copula	Gumbel	Mean	-	-	2.34	
			95% PI	-	-	[1.86, 2.73]	
Zhangjiashan	Peak	LN	Mean	6.63	0.65	-	1677.35
			95% PI	[6.48, 6.79]	[0.58, 0.76]	-	
	Volume	LN	Mean	7.20	0.86	-	
			95% PI	[7.00, 7.41]	[0.75, 0.98]	-	
	Copula	Frank	Mean	-	-	11.13	
			95% PI	-	-	[8.49, 13.15]	
Xianyang	Peak	LN	Mean	7.17	0.86	-	1583.36
			95% PI	[6.93, 7.40]	[0.74, 0.99]	-	
	Volume	LN	Mean	7.79	0.94		
			95% PI	[7.52, 8.06]	[0.84, 1.10]	-	
	Copula	Frank	Mean	-	-	10.67	
			95% PI	-	-	[8.08, 13.61]	

Note: 95%PI: 95% predictive interval

Table 3. ANOVA table for failure probability in OR

Parameter	Xiangshan					Xianyang					Zhangjiashan				
	SS	DF	MS	F-Value	P-value	SS	DF	MS	F-Value	P-value	SS	DF	MS	F-Value	P-value
A	9679.0	2	4839.5	927.7	< 0.0001	11330.0	2	5665.0	2108.7	< 0.0001	7916.5	2	3958.2	1579.8	< 0.0001
B	46186.7	2	23093.3	4427.0	< 0.0001	47547.6	2	23773.8	8849.4	< 0.0001	55688.1	2	27844.1	11112.7	< 0.0001
C	10355.3	2	5177.7	992.6	< 0.0001	11306.2	2	5653.1	2104.3	< 0.0001	8649.6	2	4324.8	1726.1	< 0.0001
D	35232.7	2	17616.3	3377.1	< 0.0001	49436.6	2	24718.3	9200.9	< 0.0001	64601.5	2	32300.8	12891.4	< 0.0001
E	78.6	2	39.3	7.5	0.0007	1.2	2	0.6	0.2	*0.7988*	1.2	2	0.6	0.2	*0.7862*
AB	5529.8	4	1382.5	265.0	< 0.0001	3055.0	4	763.7	284.3	< 0.0001	2381.7	4	595.4	237.6	< 0.0001
AC	278.3	4	69.6	13.3	< 0.0001	180.7	4	45.2	16.8	< 0.0001	101.4	4	25.4	10.1	< 0.0001
AD	1061.0	4	265.2	50.8	< 0.0001	799.6	4	199.9	74.4	< 0.0001	768.9	4	192.2	76.7	< 0.0001
AE	1.4	4	0.4	0.1	*0.9914*	0.1	4	0.0	0.0	*0.9999*	0.1	4	0.0	0.0	*0.9999*
BC	1558.3	4	389.6	74.7	< 0.0001	765.8	4	191.4	71.3	< 0.0001	722.5	4	180.6	72.1	< 0.0001
BD	6481.6	4	1620.4	310.6	< 0.0001	3392.9	4	848.2	315.7	< 0.0001	5489.1	4	1372.3	547.7	< 0.0001
BE	19.1	4	4.8	0.9	*0.4568*	0.5	4	0.1	0.0	*0.9964*	0.5	4	0.1	0.1	*0.9946*
CD	4967.3	4	1241.8	238.1	< 0.0001	3226.5	4	806.6	300.3	< 0.0001	2589.4	4	647.3	258.4	< 0.0001
CE	2.3	4	0.6	0.1	*0.9780*	0.1	4	0.0	0.0	*0.9999*	0.0	4	0.0	0.0	*1.0000*
DE	17.3	4	4.3	0.8	*0.5088*	0.5	4	0.1	0.0	*0.9961*	0.5	4	0.1	0.1	*0.9945*
Error	1001.6	192	5.2			515.8	192	2.7			481.1	192	2.5		
Total SS	122450.3	242				131559.0	242				149392.2	242			

Note: SS, DF, MS represent sum of squares, degrees of freedom, mean square, respectively; A and B denote the two parameters in lognormal distribution (i.e. P_par1 and P_par2) for flood peak; C and D denote the two parameters in lognormal distribution (i.e. V_par1 and V_par2) for flood volume; E indicates the parameter in the copula (i.e. Cop_par). P-values greater than 0.05 are highlighted in this table.

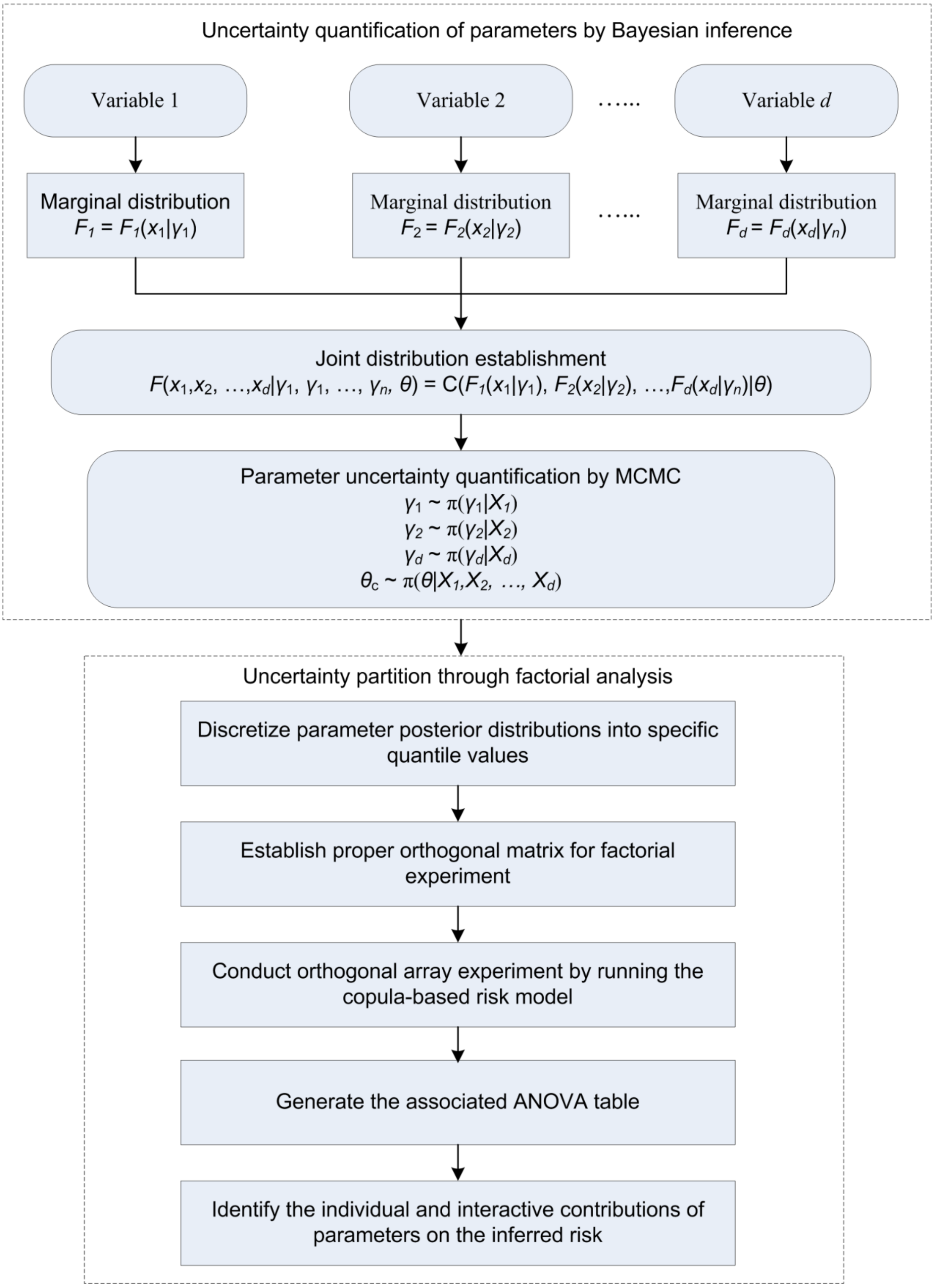

Figure 1. Framework of the proposed factorial Bayesian copula approach

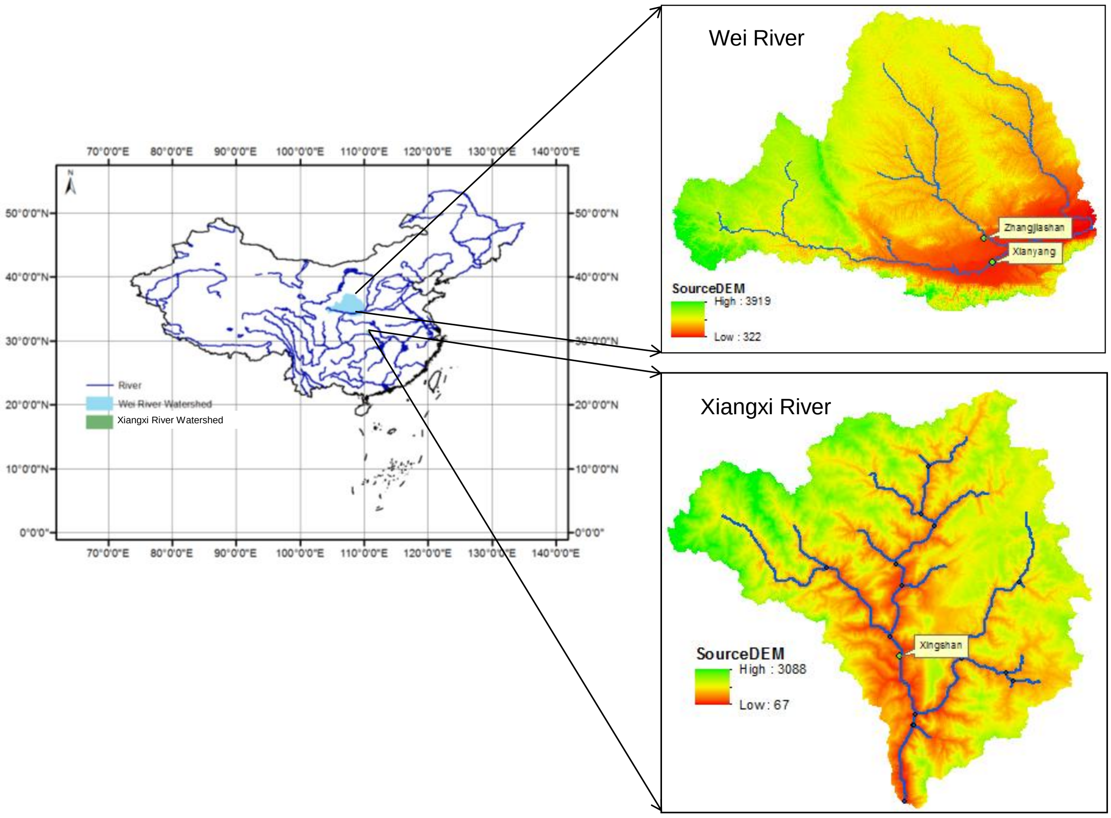

Figure 2. The location of the studied watersheds. Wei River is the largest tributary of Yellow river, with a drainage area of 135,000 km^2. The historical flood data from Xianyang and Zhangjiashan stations on the Wei River are analyzed through the proposed FBC approach. The Xiangxi River is located in the Three Gorges Reservoir area with a drainage area of 3,200 km^2. The historical data from Xingshan station is used in this study.

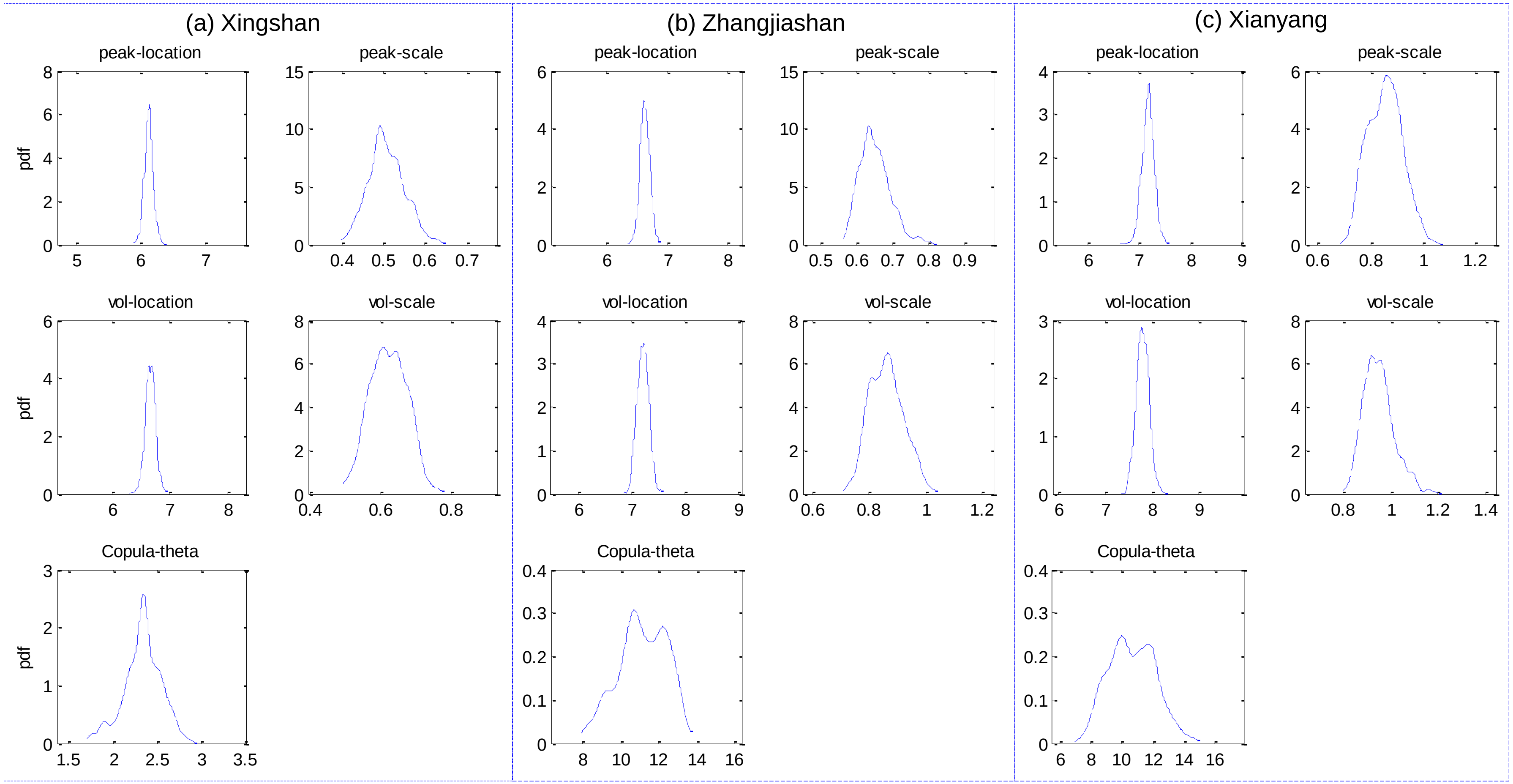

Figure 3. Posterior distributions of parameters in the copula models for different gauge statioins

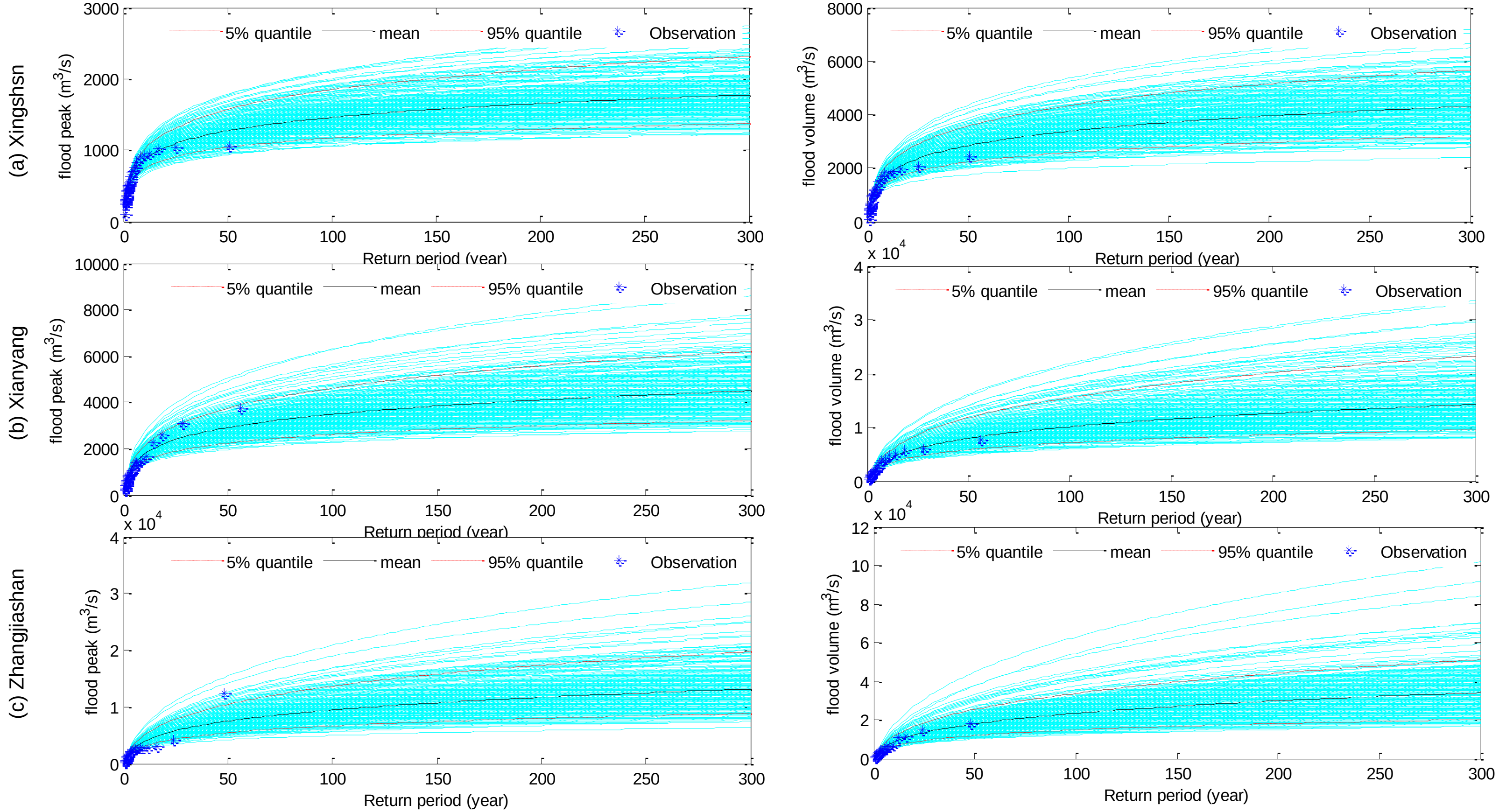

Figure 4. Uncertainty in flood peak and volume inference at the three gauge stations: both inferred flood peak and volume contain extensive uncertainties due to randomness in model parameters; such uncertainties increase significantly the increase in predefined return period

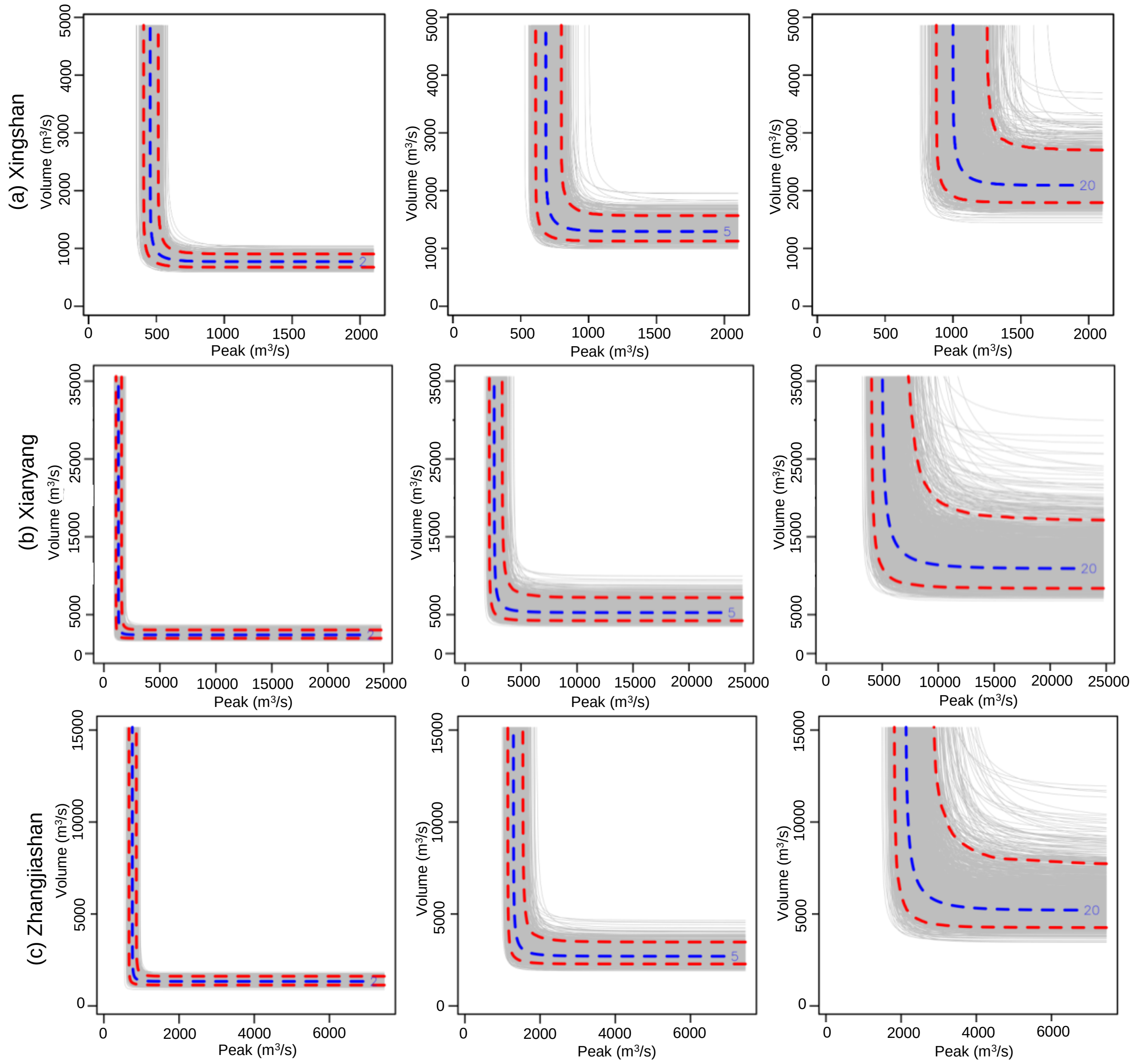

Figure 5. Uncertainty quantification of the joint return period in “OR”: the blue dash lines indicate the predictive means and the red dash lines indicate the 5% and 95% quantiles.

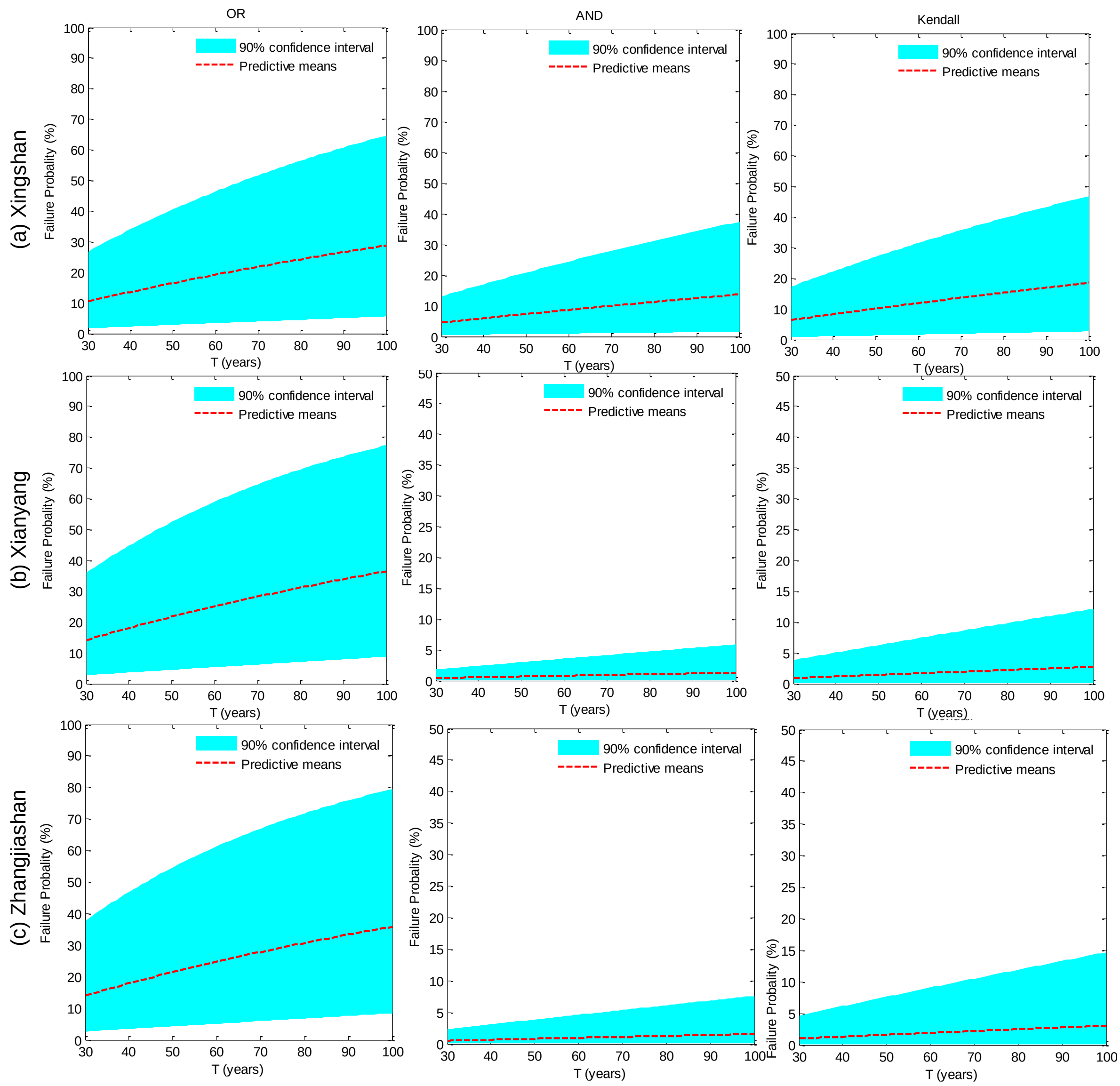

Figure 6. Uncertainty quantification for the failure probabilities in OR, AND and Kendall at the Xingshan, Xianyang and Zhangjiashan stations: Considerable uncertainties are observed in all failure probabilities and such uncertainties would increase with the increase in service time

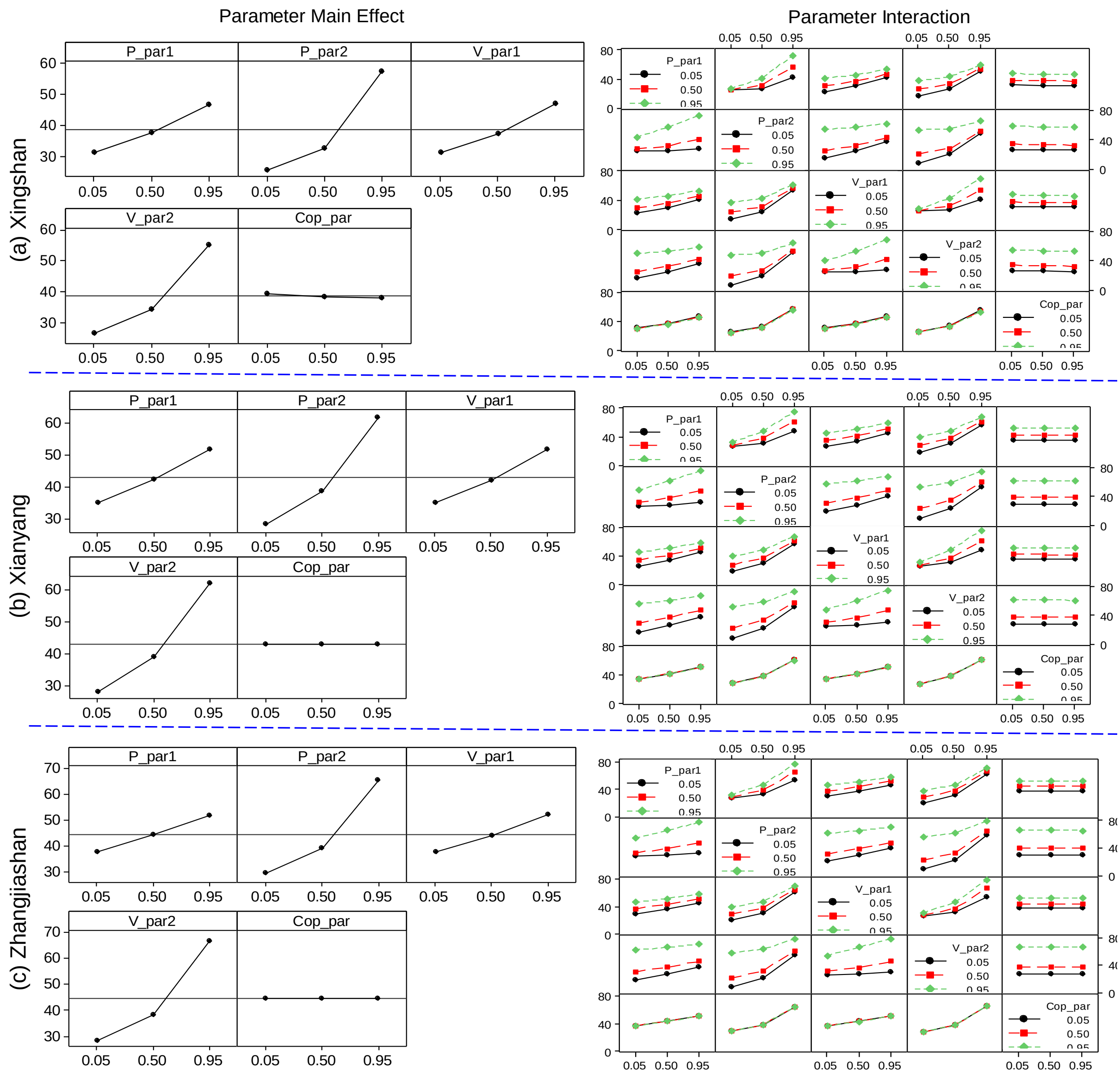

Figure 7. Main effects plot and full interactions plot matrix for parameters on the failure probability in OR at the three gauge stations

	(a) Xingshan				(b) Xianyang				(c) Zhangjiashan			
	30 year	50 year	70 year	100 year	30 year	50 year	70 year	100 year	30 year	50 year	70 year	100 year
P_par1	8.42%	8.28%	8.14%	7.93%	10.07%	10.03%	9.94%	9.78%	6.30%	6.28%	6.25%	6.18%
P_par2	39.00%	39.10%	39.13%	39.03%	35.64%	35.78%	35.73%	35.37%	37.87%	38.17%	38.25%	38.05%
V_par1	8.63%	8.64%	8.64%	8.61%	10.22%	10.09%	9.93%	9.66%	7.04%	6.92%	6.79%	6.61%
V_par2	28.56%	29.13%	29.62%	30.24%	37.99%	38.24%	38.27%	38.02%	43.67%	43.43%	42.96%	41.99%
Cop_par	0.088%	0.087%	0.086%	0.084%	0.003%	0.003%	0.002%	0.002%	0.003%	0.002%	0.002%	0.002%
Interaction	15.31%	14.76%	14.39%	14.11%	6.07%	5.86%	6.12%	7.17%	5.12%	5.20%	5.74%	7.16%

Legend (%)

1	10	20	30	40

Figure 8. Contributions of model parameters on uncertainty in predictive failure probabilities of OR at the three gauge stations