Report of International Collaborative Research (Project No.: 2021W-02)

Flood Risk Assessment with High Dimensional Vine Copulas: A Methodology Considering Spatial-Temporal Correlation of Rainfall

Xinyu Jiang[1], Hirokazu Tatano[2]

1. School of Management Wuhan University of Technology, Wuhan, China. E-mail: jxy119@whut.edu.cn

2. Disaster prevention research institute, Kyoto University, Kyoto, Japan, E-mail: tatano.hirokazu.7s@kyoto-u.ac.jp

1. Introduction

Flood risk assessment provides basic information of probability of occurrence and corresponding consequence to basin flood risk management such as dam building, city planning, flood insurance designing etc. Conventional flood risk assessment and management usually only focusing on a target area with artificially setting boundaries. Both hazard simulation (from rainfall design to inundation simulation) and exposure estimation are conducted within the target area. This treatment works well if the target area can be thought as an independent area. However, it will also lead to an underestimation of flood risk if the target area have a close relationship with other areas. On one hand, hazards which attacking the target area can be originated from multiple sources in other areas. It is the joint-behaviors of multiple flood sources rather than hazards in target area determine the risk level. On the other hand, exposures which attacked by hazards in target area can affect population and assets in other areas through economic relations. In other words, other related area may also indirectly expose to flood disaster in the target area. Therefore, spatial correlation between target area and related area should be deeply studied in flood risk assessment.

Assessing flood risk considering the relationship between target area and related areas to support integrated flood risk management is the final goal of our research. Two previous studies had been conducted. Firstly, a methodology for assessing spatial flood risk from multiple flood sources in a small river basin was proposed (Tatano, Jiang 2016; Jiang, Tatano 2019). In the proposed methodology, under the assumption of spatial homogeneity of rainfall in small river basin, the concept of critical rainfall duration, determined by the time of concentration of flooding, is used to represent the characteristics of flooding from different sources. A copula method is adopted to capture the correlation of rainfall amounts in different critical rainfall durations to reflect the correlation of potential flooding from multiple flood sources. Secondly, as for the flood risk assessment considering spatial correlation of rainfall in large river basin, taking advantage of the database for policy decision-making for future climate changes (d4PDF), a case study in Shonai river basin was conducted (Tatano, Jiang 2017). This case study directly adopted a mass of GCM simulation rainfall data which includes various spatial-temporal patterns to assess flood risk, avoided the theoretical analysis of spatial-temporal correlation of rainfall. It could be thought as an empirical study, however, not solve the spatial correlation problem for large river basin fundamentally.

In this study, we continue the research of flood risk assessment considering spatial correlation of rainfall in large river basin from a perspective of high dimensional vine copulas-based rainfall spatial-temporal correlation capturing and multivariate rainfall designing.

2. Methodology

2.1 Copula method

Copulas are functions that join, or couple, multivariate distribution functions to their one-dimensional (1D) marginal distribution functions. In the bivariate case, according to Sklar's theory (1959), the joint cumulative distribution function *H(x,y)* of any pair *(x,y)* of continuous random variables may be written in the form

$$H(x,y)=C(F(x),G(y))\ x,y\in \mathrm{R},$$

where *F(x)* and *G(y)* are continuous marginal distributions, so that C:$[0,1]^2$→[0,1] for all copulas.

A copula offers a method for measuring scale-free dependence and for constructing families of joint distribution. One of the main advantages provided by a copula is that the selection of an appropriate model for the dependence between varieties, represented by the copula, can then proceed independently from the choice of marginal distributions.

2.2 High-dimension Vine-copula

Most of copulas adopted in applications are bivariate cases. In actual cases, higher dimension copulas are often required. In our case study, multivariate copulas are required to analyze the joint distribution of flood risk sources.

Sklar's theory can easily extend to multivariate versions. Let *H* be any dimension of distribution function with margins F_1, F_2, …, F_n. Then, there exists an n-copula C:$[0,1]^n$→[0,1] such that for all x in R^n,

$$H(x_1,x_2,...,x_n)=C(F_1(x_1),F_2(x_2),...,F_n(x_n))\ x\in \mathrm{R}^n.$$

If F_1, F_2, …, F_n are all continuous, then C is unique; otherwise, C is uniquely determined for Ran*F1* × Ran*F2* × … × Ran*Fn*. Conversely, if C is an n-copula and F_1, F_2, …, F_n are distribution functions, then the function H is defined as an n-dimensional distribution function with margins *F1*, *F2*, …, *Fn*.

Vine copula is a type of multivariate copula which adopt pair copula construction (PCC) technology for multivariate copula construction. It adopts a hierarchical idea and takes advantage of density function. The modeling scheme is based on the decomposition of a multivariate density into d(d − 1)/2 bivariate copula densities, of which the first d − 1 are unconditional, and the rest are conditional. The multivariate joint density function can be written as the combination of marginal distribution and conditional distribution:

$$f(x_1,x_2...x_n)=f(x_n)\cdot f(x_{n-1}\mid x_n)\cdot f(x_{n-2}\mid x_n,x_{n-1})...f(x_1\mid x_2,...,x_n).$$

By Sklar's theorem, which shows

$$F(x_1,x_2...x_d)=\mathrm{C}(F_1(x_1),F_2(x_2)...F_n(x_n)),$$

using the chain rule, we have

$$f(x_1, x_2 \ldots x_n) = c_{1\ldots n}(F_1(x_1), F_2(x_2)\ldots F_n(x_n)) \cdot f_1(x_1)\ldots f_n(x_n),$$

where $c_{1\ldots n}$ denotes the densities of $C_{1\ldots n}$. From these formulas, a general formula that is expressed in each term can be derived:

$$f(x|\mathbf{v}) = c_{xv_j|\mathbf{v}_{-j}}(F(x|\mathbf{v}_{-j}), F(u_j|\mathbf{v}_{-j})) \cdot f(x \mid \mathbf{v}_{-j}),$$

where $\mathbf{v}_{-j}$ denotes the v-vector excluding v_j. Formula therefore can be represented in terms of bivariate copulas. Another crucial question is how to obtain the conditional distribution functions $F(x|\mathbf{v})$. Jeo shows that for every j,

$$F(x|\mathbf{v}) = \frac{\partial c_{xv_j|\mathbf{v}_{-j}}(F(x|\mathbf{v}_{-j}), F(u_j|\mathbf{v}_{-j}))}{\partial F(u_j|\mathbf{v}_{-j})}.$$

The concept of PCC is highly iterative. To understand it visually, we take 3D joint densities. For example,

$$\begin{aligned} f(x_1, x_2, x_3) &= f(x_1) \cdot f(x_2 \mid x_1) \cdot f(x_3 \mid x_1, x_2) \\ &= f(x_1) \cdot c_{1,2}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2) \cdot f(x_3 \mid x_1, x_2) \\ &= f(x_1) \cdot c_{1,2}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2) \cdot c_{2,3|1}(F(x_2 \mid x_1), F(x_3 \mid x_1)) \cdot c_{1,3}(F_1(x_1), F_3(x_3)) \cdot f_3(x_3) \end{aligned}$$

As is shown above, the 3D joint densities are represented in terms of bivariate copulas $C_{1,2}$, $C_{1,3}$, and $C_{2,3|1}$ with densities $c_{1,2}$ $c_{1,3}$ $c_{2,3|1}$. These so-called pair copulas may be chosen independently of each other. The multivariate copulas can also be expressed as

$$c(x_1, x_2, x_3) = c_{1,2}(F_1(x_1), F_2(x_2)) \cdot c_{1,3}(F_1(x_1), F_3(x_3)) \cdot c_{2,3|1}(F(x_2 \mid x_1), F(x_3 \mid x_1)),$$

where $F(x_2 \mid x_1) = \partial C(x_1, x_2) / \partial x_1$ and $F(x_3 \mid x_1) = \partial C(x_1, x_3) / \partial x_1$.

Since the decomposition is not unique, many different ways can be used for ordering variables. More detailed information about the vine copula can be found in Joe's (2011) and Schepsmeier's (2015) work.

3. Appling vine copula to rainfall design

In flood risk assessment, a key question is to define the joint behavior of different flood sources considering the complex spatial and temporal correlations. Especially, those behaviors should be identified with return period (probability of occurrence). For example, as shown in figure.1, the risk assessment area (marked by red star) will influenced by local basin rainfall (light red bar plot), the first branch basin rainfall (purple bar plot) as well as the second branch basin rainfall (deep yellow bar plot). Then, how to define the flood risk about risk assessment area through rainfall? Traditional consideration may directly use the probability analysis of local rainfall. However, it will obviously lead to an under estimation of flood risk since ignoring the influence from branch basins.

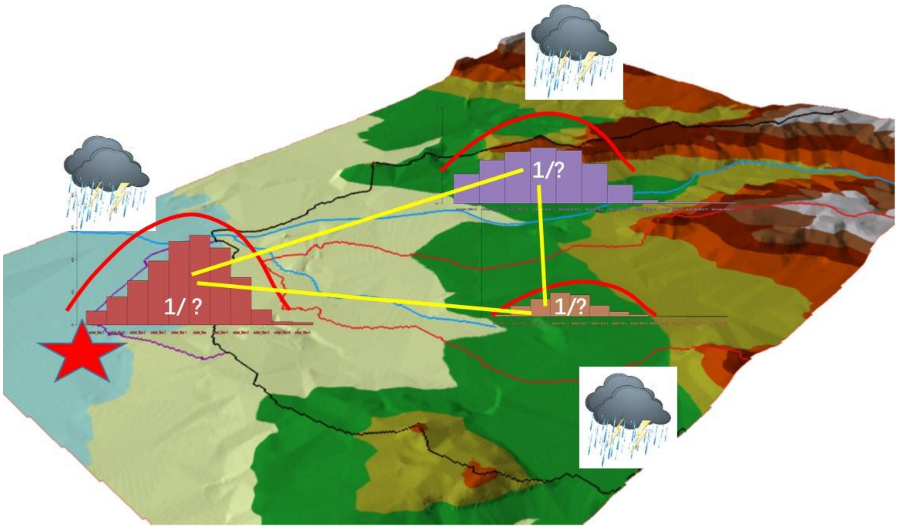

Figure 1 Spatial temporal correlation problem for flood risk assessment

It is an intuitionistic that we should build a multi-variate distribution to reflect the joint behavior of rainfalls in different space, as well as different time. As is shown in the following illustration: where F is a joint probability function, R is rainfall, A, B, C donates different basins and t1…tm donates different times of a rainfall event. There are two challenges: first, how to construct such a multi-variate distribution. Second, how to define the return period (probability of occurrence) from the multi-variate distribution.

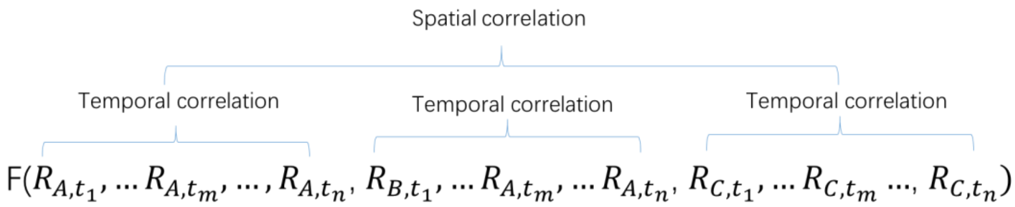

Figure 2 Joint probability of spatial and temporal correlated rainfall

For the first challenge, the high dimensional copula is used. And for the second challenge, we specifically focused on the explanation of vine copula construction technology. Vine copula is based on univariate probability and conditional probability. The univariate probability is the starting point of the construction of high dimensional probability. In our research context, the starting point should be explained as the probability of occurrence of the most significant rainfall. If thinking temporal relation is important, we should consider the time distribution of rainfall first. If thinking spatial relation is important, we should consider the spatial distribution of rainfall first. Overall, we should determine the critical space and time first, then surround the critical rainfall, considering the dependence structure of rainfall at different time and space, and therefore simulate the rainfall events.

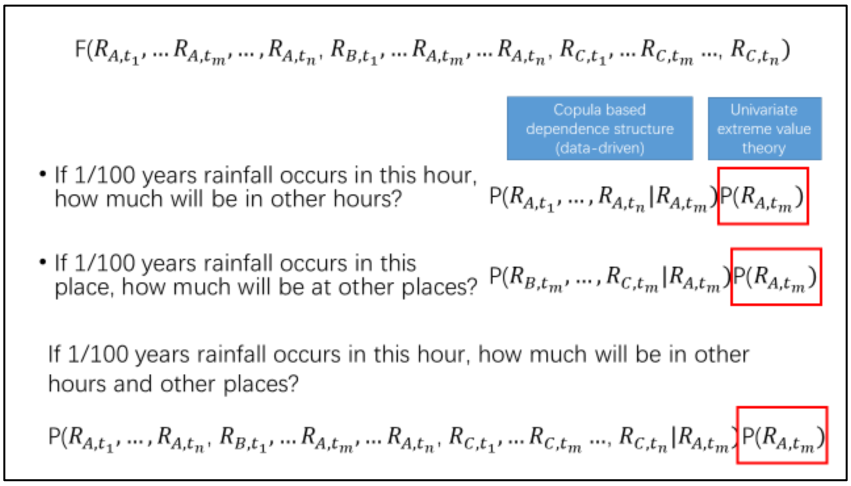

Figure 3 the relationship between vine copula, univariate extreme value distribution (critical rainfall) and dependence structure

A methodology procedure is proposed for adopting vine copula to rainfall design, as shown in figure 4. Firstly, the rainfall from multi rain gauging station are pre-processed to simultaneous rainfall event. Then, define the critical rainfall and capture spatial-temporal correlation of rainfall events using vine copulas. Thirdly, random conditional sample the correlated value from vine copula. Fourthly, fitting the marginal distributions of rainfall at each time and space. Fifthly, converting the correlated sampling value to rainfall value with the marginal distribution. Finally, finish the design of spatial temporal correlated rainfall.

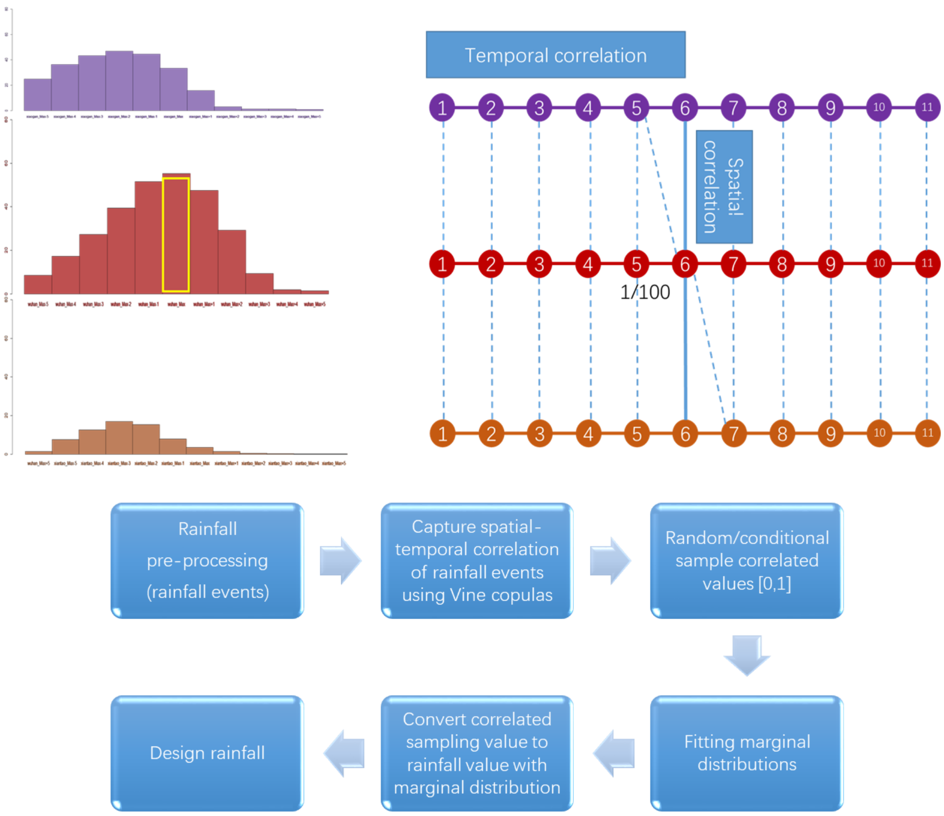

Figure 4 the procedure for adopting vine copula to rainfall design

4. Case study

A case study in the east part of Hubei province, China is conducted to demonstrate the proposed methodology. In the case study area, there are three regions. Wuhan is at the downstream where the flood risk will be originated from rainfall of Xiantao, Xiaogan as well as Wuhan itself. When considering the flood risk in Wuhan, the joint-behavior of rainfalls in Xiaogan and Xiantao also should be studied. By adopting the methodology we proposed above, spatial-temporal correlated rainfall events could be generated according to given return periods. Figure 5 shows six spatial-temporal correlated rainfall events given probability of occurrence of critical rainfall (1h maxima rainfall in Wuhan) smaller than 1/100.

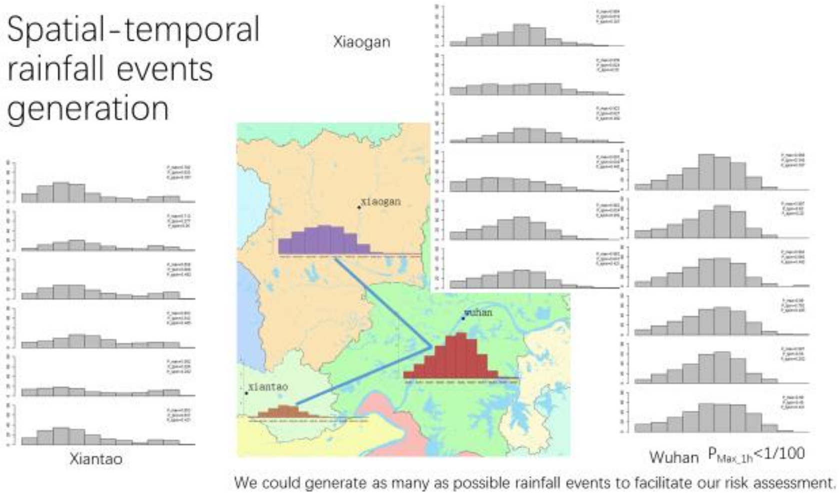

Figure 5 Vine copula based rainfall design for Wuhan and correlated areas

5. Discussion

The proposed methodology is a practical way for generating spatial-temporal correlated rainfall for integrated flood risk assessment. It is a mixture of science and art, although many quantitative analysis are adopted. For the art part (here we mean decision by people), in this methodology, set focus, decision according to correlation, select copula just by goodness of fit and pre-determine the cine structure are need. They are, in some sense, experience based. And for the science part, univariate extreme value theory, correlation theory, copula and vine copula theory are included. To improve this methodology, in the future, for practical application, validating this method through many times of the practical applications and accumulating experiences for identify key rainfall features, selection of copulas, predetermine the vine tree structures are necessary. For the theory improvement, selecting copula based on theory rather than just goodness of fit. E.g. what will be best copula for a block maximum and a non-maximum and multivariate extreme value theory will be considered (Jiang 2021, 2022a, 2022b).

We also compared the vine-copula based methodology and existing spatial-temporal statistical methods, such as deterministic prediction method, trend prediction, regression models for forecasting, generalized linear spatial-temporal regression, spatial-temporal kriging, spatial-temporal basis functions, temporal basis functions, non-Gaussian spatial-temporal gams, as well as non-Gaussian spatial-temporal models (Cressie 2015). Copula-based methods are more proper for the risk assessment research questions that emphasize on the joint probability of occurrence. However, some ideas from the classical models could be very crucial for us to improve the proposed methods, for example the consideration of trend, still as well as directions, especially when expanding the copula-based method to large number of stations. Further researches will also consider this direction.

6. Reference

Cressie N, Wikle C K. Statistics for spatio-temporal data. John Wiley & Sons, 2015.

Joe, Harry, and Dorota Kurowicka, eds. Dependence modeling: vine copula handbook. World Scientific, 2011.

Jiang, X.; Yang, L.; Tatano, H. Assessing Spatial Flood Risk from Multiple Flood Sources in a Small River Basin: A Method Based on Multivariate Design Rainfall. *Water* 2019, *11*, 1031.

Jiang, X., Design Rainfall for Flood risk assessment: A multivariate method using high dimensional vine copulas, The 11th International Conference of the International Society for the Integrated disaster risk management (IDRiM 2021), online, 2021-09-23

Jiang, X., Spatial and Temporal Correlation in Disaster Risk Assessment: Challenges between Geographic and Economic perspectives. The 12th International Conference of the International Society for the Integrated Disaster Risk Management (IDRIM 2022), online, 2022-09-22

Jiang, X., Designing spatial-temporal correlated rainfalls for integrated flood risk assessment： A vine copula-based approach, 62nd Sogo Bosai Seminar, Disaster Prevention Research Institute, Kyoto University, Japan, 2022-09-02

Tatano, H., Jiang, X., A methodology for flood risk assessment considering spatial-temporal correlation, 16th JICE research and development grant report meeting, 2016.05.25

Tatano, H., Jiang, X., A methodology for flood risk assessment considering spatial-temporal correlation, 16th JICE research and development grant report meeting, 2017.05.09

Schepsmeier U, Stoeber J, Brechmann E C, et al. Package 'vinecopula'[J]. R package version, 2015, 2(5).

Sklar A. Fonctions de repartition an dimensions etleursmarges. Publ Inst Statist Univ Paris, 1959.8:229-231.