Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia

Vol. 7, 2022, Page 52 - 55

Preliminary Study on Flood Frequency Analysis in Johor River Basin Using Vine Copula

Naqibah Aminuddin Jafry[a], Jamaludin Suhaila[b], Fadhilah Yusof[a,b], Siti Rohani Mohd Nor[c*], Nor Eliza Alias[d]
[a,b,c]Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310, Johor Bahru, Johor, MALAYSIA.
[b]UTM Centre for Industrial and Applied Mathematics (UTM-CIAM), Ibnu Sina Institute for Scientific and Industrial Research, Universiti Teknologi Malaysia, 81310, Johor Bahru, Johor, MALAYSIA.
[d]Department of Water and Environmental Engineering, School of Civil Engineering, 81310, Universiti Teknologi Malaysia, Johor Bahru, Johor, MALAYSIA.
*Corresponding author: sitirohani@utm.my

Abstract
Flood is a multi-attribute naturally occurring phenomenon that is defined by mutually associated flood parameters such as peak flow, volume, and flood duration. Numerous literatures have incorporated univariate and bivariate analysis to measure flood risk, but a more effective evaluation may be made by considering all three mutually associated flood features at the same time to avoid underestimation or overestimation of the flood events. However, trivariate flood frequency analysis based on copula technique has several drawbacks such as higher-dimensional expressions are only accessible for a minority of copula families which may not be flexible enough to depict complicated dependency relationships. This study aims to accommodate and provide solutions to the existing flood problem by investigating the occurrence of flood related to the river basins in Johor Bahru, thus incorporating the Mutual Information (MI) techniques in the vine copula selection approach. The proposed method will yield a better joint return period and provide comprehensive insights for flood risk analysis, thus, allowing practitioners to offer a better water planning and management framework.
Keywords: Vine Copula; Mutual Information; Entropy; Flood

Introduction
Start Flood is a multi-attribute naturally occurring phenomenon that is defined by mutually associated flood parameters such as peak flow, volume, and flood duration that can be extracted from the flood hydrograph [4]. Many researchers have incorporated univariate and bivariate analysis to measure flood risk, but a more effective evaluation may be made by considering all three mutually associated flood features at the same time to avoid underestimation or overestimation of the flood events. A univariate probability distribution analysis is evidently insufficient for a flood occurrence, because these random variables such as flood peak, flood volume, and flood duration are closely related to each other due to the multivariate nature of the phenomenon itself. Many studies have emphasized its lack of credibility, claiming that univariate frequency analysis approaches are incapable to adequately define inflow hydrographs or minimize flood analysis uncertainty [3]. To examine the hydrological risk of dams or the reliability of spillways, not only the flood peak but also the flood volume and duration must be considered, but that can only be handled using multivariate models that describe the flood characteristics' dependency structures [9].

Copula concept has successfully emerged in 1959 and it was first proposed by Sklar [8] to model universal dependency in multivariate data and this method can display the structure of dependence between two or more random variables. This method has widely used by researchers in many fields for example economics to describe the dependency of two variables which is stock market and hydrology to model the hydrological events such as flood and drought which are interrelated by several entities or properties that must be considered for when modelling and analyzing hydrological occurrences. Copulas are gaining popularity and have attracted much attention in the financial field because they are a flexible tool that allows the marginals to be represented using any sort of distribution without affecting the dependencies between them and the limiting assumptions of normality and linear dependency can be avoided [6]. Bárdossy and Li [1] emphasized the advantages of using copula in their study where a copula has the benefit of being invariant to strictly increasing monotonic transformations of the variables.

The biggest disadvantage of using traditional multivariate modelling is that each flood's entities must be described by the same parametric family of univariate distributions, which is obviously not the case in the real-world application since each entity has different distributions. As a result, the conventional multivariate distributions may not equip decision-makers with the most favorable outcomes and a more accurate model is needed. Although there are many copula families, higher-dimensional expressions are only accessible for a minority of them, which may not be flexible enough to depict complicated dependency relationships [9]. Vine copulas have developed as alternative families in the past few years, allowing us to bypass these limits. Tosunoglu, Gürbüz [9] compared conventional trivariate copula families (Archimedean and Elliptical) to newly suggested vine copulas in this literature and they discovered that vine copulas are the best viable models for representing the combined dependency structure of flood peak flow, volume, and duration series. Therefore, the use of vine or pair-copula constructions (or PCC) in this study will amplify the efforts of higher dimensional copulas (when there are three or more variables that need to be considered).

Materials and methods

This study aims to accommodate and provide solutions to the existing flood problem by investigating the occurrence of flood related to the river basins in Malaysia, specifically in Johor Bahru, thus adopting a vine copula model and incorporating the Mutual Information (MI) approach in vine copula section techniques to enhance the existing multivariate modelling analysis. Vine copula is an extension of conventional copula modelling approach which covers the restrictions of the traditional copula techniques. In the event where multivariate modelling for river basins can be modelled precisely and efficiently, flood control and prevention in Malaysia can be improvised thus the catastrophe can be controlled, proper planning can be executed, and the amount of loss can be reduced.

PCC is able to characterize the correlation patterns between high-dimensional response-independent variables, making them a useful and adaptable tool for analyzing the dependence structures between complicated coupled correlated variables. The major goal of utilizing the vine copula model and other multivariate copula models to describe the relationships between flood factors was to predict design events for a specific return time and explore the differences in a practical application. As a result, the recurrence period of flood in Johor River Basin can be estimated. Nevertheless, there are several drawbacks with the vine copula approach where the existing vine copula selection methods are either time-consuming or model-dependent, implying that previous procedures may have an impact on the outcome of the selection. Therefore, MI approaches which were based on copula entropy will be incorporated in constructing vine copula.

Vine copula is able to provide better outcomes in multivariate modelling analysis as it can resolve the issues and overcome the limitations of traditional copula modelling. The selection of the best copula structure under vine is crucial and time consuming since we need to choose and compute the empirical copula before selecting the vine structure. Canonical or C-vine distribution and D-vine distribution are two particular types of parametric regular vine building that are attributable to different varieties of pair-copula decomposition as can be seen in Figure 1 [2].

The D-vine structure's applicability is commonly mentioned from the existing literature owing to its higher flexibility than the C-vine structure, and it would be useful when the presence of any certain vectors that govern the number of mutual interactions within distributed observations is specified or known [5].

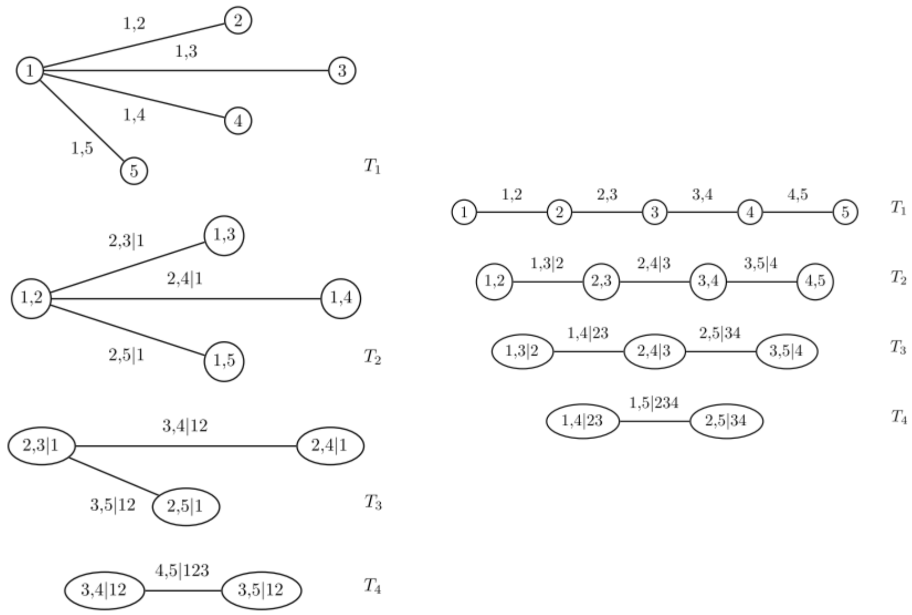

Figure 1 Examples of five-dimensional C- vine (left panel) and D-vine trees (right panel).

Thus, the incorporation of mutual information and copula entropy in vine copula construction can reduce the amount of time and the computational effort required as compared to the usual vine copula selection method. In modelling multivariate flood data with complicated patterns of dependency specifically in the tails, both MI and PCC outperform another typical multivariate copula [3]. Furthermore, the minimal information pair-copula model, especially, is more flexible and gives a better approximation of the joint probability density and related measures, making it useful for probabilistic flood hazard assessment. Additionally, Minimum Information (MI) PCC preserves the complicated multidimensional flood structure for diverse tail dependencies by accurately predicting their tail coefficient for given copulas, and it also makes it easier to simulate multivariate extremes when data length is restricted [3, 5].

This study will be conducted using daily discharge data of tributaries flowing into Johor River Basin which are Linggiu, Sayong, Penggeli, Semangar, Lebam, Telor, Panti, Temboyah, and Permadi. The datasets will be covering the period of 30 years starting from 2000 until 2020. These datasets have been collected from the Department of Irrigation and Drainage Malaysia (DID). Data analysis will be conducted using R software, and EasyFit software. The incorporation of MI in vine copula construction and traditional copula approach for flood frequency analysis were compared.

Expected results

Based on the existing literature review, the author believes that the incorporation of MI technique in the vine copula selection approach will reduce the amount of time and computational effort required to carry out this method as compared to the usual vine copula approach. Moreover, MI will assist in providing sufficient information on the uncertainty of the variables. According to a study conducted by Ni, Wang [7], as compared to the conventional tau-based technique, the MI-based approach accurately described various types of dependent structure and generated and provided more information on variables.

It was also highlighted in their study that the existing vine copula selection methods are either time-consuming or model-dependent, implying that previous procedures may have an impact on the selection outcomes. However, further work needs to be done to prove and investigate the above statement.

Conclusion

This study will broaden our understanding on multivariate flood frequency analysis using the vine copula approach, as well as the significance of modelling and estimating the flood recurrence period. For government, practitioners, and decision makers, this study will assist in mitigating and reducing the loss resulting from this natural disaster, hence providing an early warning, therefore will give ample time to the victims staying in the flood prone area and the authorities to decide the most reasonable action that should be taken. Furthermore, loss of lives caused by this disaster can also be avoided. After completing this study, the author also believes that the incorporation of MI and CE techniques in constructing vine copula will aid in obtaining the conditional probability, conditional return period and joint return periods values based on the dependency of flood properties for flood recurrence estimation in Johor River Basin, thus, proper water planning can be executed. For future work, the author can propose a way to choose the most suitable vine structure using a novel technique as the vine structure selection is still considered as a complicated task. Based on the best fitted copula distribution obtained in this study, the analysis of the vine copula for conditional probability, conditional return period, and joint return period will be calculated for the flood frequency analysis.

Acknowledgement

The authors would like to express their greatest appreciation to the Department of Irrigation and Drainage, Malaysia, for providing the daily discharge data of Johor River Basin. This work was supported by the Research University Grant with vote no: Q.J130000.2554.21H65 under Universiti Teknologi Malaysia.

References

[1] Bárdossy, A. and Li, J. 2008. Geostatistical interpolation using copulas. Water Resources Research. 44(7).

[2] Brechmann, E.C. and Schepsmeier, U. 2013. Modeling Dependence with C- and D-Vine Copulas: TheRPackageCDVine. Journal of Statistical Software. 52(3): 1-27.

[3] Daneshkhah, A., Remesan, R., Chatrabgoun, O. and Holman, I.P. 2016. Probabilistic modeling of flood characterizations with parametric and minimum information pair-copula model. Journal of Hydrology. 540: 469-487.

[4] Ganguli, P. and Reddy, M. J. 2012. Probabilistic assessment of flood risks using trivariate copulas. Theoretical and Applied Climatology. 111(1): 341-360.

[5] Latif, S. and Mustafa, F. 2020. Copula-based multivariate flood probability construction: a review. Arabian Journal of Geosciences. 13(3): 1-25.

[6] Mahfoud, M. and Michael, M. 2012. Bivariate Archimedean copulas: an application to two stock market indices. BMI Paper. 1517333.

[7] Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., Liu, J. and Xie, F. 2020. Vine copula selection using mutual information for hydrological dependence modeling. Environ Res. 186: 109604.

[8] Sklar, M. 1959. Fonctions de repartition an dimensions et leurs marges. Publ inst statist univ Paris. 8: 229-31.

[9] Tosunoglu, F., Gürbüz, F. and İspirli, M.N. 2020. Multivariate modeling of flood characteristics using Vine copulas. Environmental Earth Sciences. 79(19): 1-21.