ASM Sc. J., 11, Special Issue 3, 2018 for SANREM, 246-253

Tail Dependence between Rainfall Characteristics using Copula: The Case of Pahang state

Mohd Faris Fauzi[1], Ummul Fahri Abdul Rauf[2*] Panlop Zeephongsekul[3] and Nur Aisyah Abdul Fataf[1]

[1] Department of Mathematics, Faculty of Defence Science and Technology, National Defence University of Malaysia, 57000 Sg. Besi, Kuala Lumpur, Malaysia

[1] Department of Mathematics, Centre for Defence Foundation Studies, National Defence University of Malaysia, 57000 Sg. Besi, Kuala Lumpur, Malaysia

[3] School of Science, College Science, Engineering and Health, RMIT University, VIC 3000 Australia

Flood disaster is one of the natural disasters that frequently hits the country apart from landslides, forest fires and others. The main objective of this study is to investigate whether the rainfall variables that are commonly used in hydrological modelling, such as intensity, severity and duration at different flood prone areas, are more highly correlated at the tails. Tail dependence is one of the important approaches in extreme value analysis which describes the dependence in tail of a multivariate distribution. Traditional approach of considering the measures of dependence such as the Spearman correlation and the Pearson linear correlation are not able to accurately characterize the dependence especially on the tails of distribution. The classical approach for considering the joint distribution of rainfall characteristics using typical bivariate modeling presents some constraints that can be avoided by Copulas. Bivariate copulas are used to estimate the degree of tail dependence displayed by rainfall characteristics. Monthly rainfall data were selected from different rain-gauge stations, over a period of forty-five year from 1970 to 2014 which are located in the eastern Peninsular state of Pahang. Considering the whole tests, the Gumbel copula was selected as the best copula model that best explains the dependence structures among the rainfall variables.

Keywords: Tail dependence, rainfall characteristics, copulas, bivariate, joint distribution

I. INTRODUCTION

In modeling extremes rainfall distribution, the bivariate analysis of flood allowing to understand the extreme precipitation that could cause the occurrence of floods. Traditionally, the Spearman correlation and the Pearson linear correlation not able to accurately characterize the dependence especially on the tail of the distributions. Thus, it may not be suitable for describing extremes. In analyzing floods, many previous studies did not take the phase to analyze the tail dependency tests, even though this test was an important step in relation to extreme modeling.

Copula form is widely used to be a very useful tool in the analysis of dependency model because the joint distribution is separated from marginal distribution and interdependency between the variables. The concept of copula was introduced by Sklar (1959, p. 229). Recently the pioneering research on the used for hydrology (Zhengjun *et al.* 2017, p. 685) applications was done and copula has been widely applied in researches (Mirabbasi *et al.* 2012, p. 191 & Abdul

*Corresponding author's e-mail: ummul@upnm.edu.my

Rauf & Zeephongsekul 2014, p. 4835). The objective of dependence modeling and multivariate copulas in hydrological applications is to develop parametric families to construct for multivariate data with different dependence structures.

Tail dependence explains the amount of dependence in the tail of a bivariate distribution. Thus, tail dependence will estimate the degree of tail dependence displayed by rainfall variable in the corner of the lower-left quadrant or upper-right quadrant of a bivariate distribution. The interpretation of tail dependence for multivariate relevant to the bivariate marginal distribution functions. The tail dependence of bivariate distribution describes the limiting proportion that one margin exceeds a certain threshold given by the other margin that has already exceeded that threshold.

Extreme events require appropriate modeling and a need for good understanding of these behaviors and other copulas in the tails. The constraints of the copula approach are due to the lack of a predominant way of knowing the dependence structure of a data set matched with the appropriate copula model, the Akaike Information Criterion (AIC) was used and the goodness-of-fit approach will be analyzed to fit the data set. Heavy-tailed distributions are the basis for the diversity of events caused by precipitation and at least one of the precipitation types paired to a heavy-tailed weather record should keep with heavy-tailed (Nicholas and Alexander 2015, p. 2965). An adapted theoretical property of the tail quotient correlation coefficient, which considers the dependence between two random variables.

In the following sections, a different chapter is needed for the application of copulas to be described in detail. The next section discusses the data set used in this research, the selection of rain-gauge station and the description of the rainfall characteristics. The best marginal distributions selected from Akaike Information Criteria (AIC) in section *Marginal Distribution*. The copula selection starts with an overview of the Archimedean copula used and the goodness of fit (section *Archimedean Copula*) are addressed. Section *Tail Dependence Analysis* provides an in-depth analysis of the dependence between the different variables. Finally, conclusions are formulated in the last section.

I. CASE STUDY AREA: EASTERN OF MALAYSIA

Pahang is the third largest state in Malaysia, after Sarawak and Sabah, and the largest in Peninsular Malaysia with a total area of 36,137 km2 (13,953 sq mi). The state occupies the huge Pahang River river basin. Pahang is a state in the East Coast which has the effect of Northeast Monsoon, from November to March every year, and the level of floods that occur is dependent on the intensity and rainfall value that falls at a time during the season. In this monsoon season (2014/2015), several flood events have occurred in the state, some of which have occurred in the districts of Rompin, Pekan, Kuantan, Maran, Temerloh, Bera, Jerantut, Raub, Lipis, Cameron Highland, and Bentong.

The main source of floods identified this year was due to the high intensity of rainfall in the catchment area and upstream, which went down completely and continuously a period of 9 to 11 days in the state of Pahang. The flood phenomenon has also caused some areas, particularly near the coast, to experience dramatic increase in water which is the source of flooding in the lowlands during the monsoon season. Placement sites in lowland areas mostly in the river basins and natural floodplains are also among the factors contributing to the increasing number of evacuees, each time the flood season arrives.

Table 1. Summary of selected station in Pahang state

Station Code	Station Number	Station Name	Latitude	Longitude
EP1	3524080	Kampung Tebing Tinggi	03°31'N	102°25'E
EP2	3431099	Kampung Serambi	03°29'N	103°08'E
EP3	3732021	Kampung Soi	03°43'N	103°18'E

The precipitation time series used in this research consists of two rain-gauge stations across the state of Pahang located in eastern, Malaysia. The time series used in this research consists of 45 years of rainfall observations from 1970 to 2014, collected from the Department of Irrigation and Drainage (DID) Ampang, Malaysia, retrieved to monthly time series data for tail dependence analysis. Table 1 presents the geographic locations of the selected rain-gauge stations.

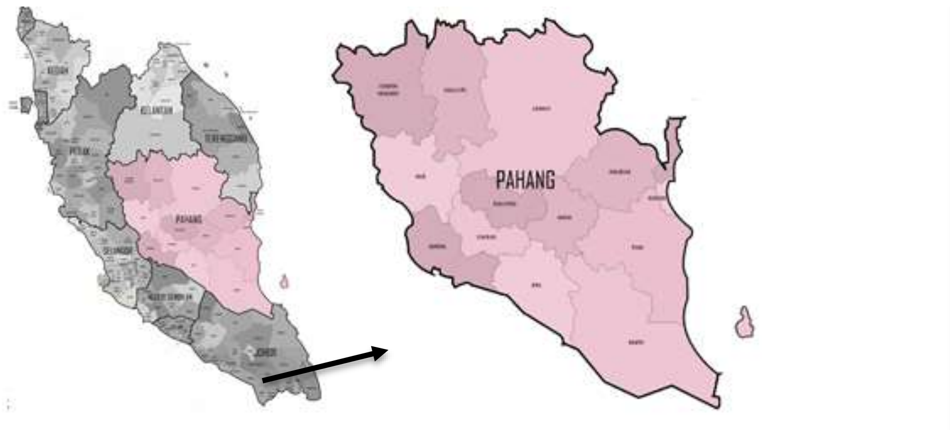


Figure 1. Maps location of Peninsular Malaysia (left) and Pahang state (right)

II. MARGINAL DISTRIBUTIONS

Marginal distribution is obtained using copula for bivariate distributions modelling. In fitting the marginal distribution of precipitation for EP1, EP2 and EP3 stations, gamma, lognormal, weibull and exponential distributions were used. For preliminary steps, the data set is fitted from distribution (Table. 2). The cumulative distribution functions (CDFs) of the distributions for both stations show that lognormal distributions fit the data better than other distributions. Thus, the best marginal distributions are selected based on Akaike

Information Criteria (AIC) that obtained using mean square error (MSE). Table 3 shows the AIC value from all distributions in this study. It shows that for EP1 and EP2 stations, lognormal distribution fits the best, hence lognormal distribution is used for modelling marginal distribution. For EP3 station show weibull distribution fits the best distribution.

Table 2. Parameter estimate for marginal distribution

Rainfall Characteristics	**Station Code**	**Marginal Distributions**						
		Gamma		Lognormal		Weibull		Exponential
		α	1/β	μ	σ	α	β	λ
Severity	EP1	2.60	0.64	1.20	0.66	**1.69**	**4.57**	0.25
	EP2	2.36	0.54	**1.25**	**0.68**	1.55	4.91	0.23
	EP3	2.02	0.46	**1.21**	**0.76**	1.50	4.87	0.23
Duration	EP1	2.89	1.03	**0.85**	**0.60**	1.70	3.17	0.36
	EP2	2.88	0.98	**0.89**	**0.61**	1.74	3.32	0.34
	EP3	2.34	0.83	**0.81**	**0.69**	1.60	3.17	0.35

Table 3. Goodness of Fit Distribution using Akaike Information Criterion (AIC)

Rainfall Characteristics	**Station Code**	**Akaike Information Criterion (AIC)**			
		Gamma	Lognormal	Weibull	Exponential
Severity	EP1	110.32	111.29	**110.05**	120.22
	EP2	151.85	**150.03**	153.83	160.53
	EP3	136.05	**135.75**	136.52	140.68
Duration	EP1	115.29	**113.19**	117.84	127.98
	EP2	121.85	**120.63**	123.88	134.96
	EP3	108.86	**107.79**	109.85	116.09

Table 3. Archimedean Copulas

Copula	$C_\theta(u,v)$	**Relationship with** τ
Clayton	$(u^{-\theta}+v^{-\theta}-1)^{-1/\theta}$	$\frac{\theta}{\theta+2}$
Frank	$-\theta^{-1}\log\left(1+\frac{(e^{\theta_u}-1)(e^{\theta_v}-1)}{e^{-\theta}-1}\right)$	$1-\frac{4}{\theta}[D_1(-\theta)-1]$
Gumbel-Hougaard	$exp\left[-(\log u)^{\theta}+(-\log v)^{1/\theta}\right]$	$1-\theta^{-1}$

III. ARCHIMEDEAN COPULA

In hydrology application and water resource management, among families of the copula, the Archimedean copula is the most popular method used in the previous study. Archimedean copulas allow modeling dependence in high dimensions with only one parameter. Archimedean copula is easily to construct and can be applied for both positive and negative correlation between the multiple variables (Zhang and Singh 2006, pp. 150-164). If X and Y are the two variables with marginal distribution functions represented as $F_X(x)$ and $F_Y(y)$, let $U = F_X(x)$ and $V = F_Y(y)$, then U and V are uniformly distributed random variables with their values u and v. Archimedean copulas can be expressed in terms of $\varphi(.)$ convex decreasing function, the Archimedean copula can be derived by:

$$C_\theta(u,v) = \varphi^{-1}(\varphi(u)+\varphi(v)) \quad (1)$$

where θ is the parameter of the copula generating function. These copulas and their generators are shown in Table 3 below

In bivariate Archimedean copula, there is a mathematical relationship between Kendall's tau coefficient, τ, and the generator functions of Archimedean copula, which is given by:

$$\tau = 1 + 4\int_0^1 \frac{\varphi_\theta(t)}{\varphi'_\theta(t)}\,dt \quad (2)$$

where t is either two variables, u and v. Graphically, the bivariate plot for three stations in this study, data is generated from the estimated Archimedean copula distribution from two rainfall characteristics of severity against the duration for both observed, shown in Figure 2.

IV. TAIL DEPENDENCE ANALYSIS

Let $X = (X_1, X_2)$ is bivariate distribution, the upper tail-dependent λ_{up} if:

$$\lambda_{up} = \lim_{v\to 1^-} P\{X_1 > F_1^{-1}(v) \mid X_2 > F_2^{-1}(v)\} > 0, \quad (3)$$

where F_1^{-1} and F_2^{-1} represent the generalized inverse distribution functions of X_1 and X_2, respectively. Therefore, $X = (X_1, X_2)$ is upper tail-independent if λ_{up} equals 0. Furthermore, we call λ_{up} the upper tail-dependence coefficient (upper TDC). Similarly, we define the lower tail-dependence coefficient, if it exists, generalization of bivariate tail dependence:

Table 4. Upper tail dependence coefficient for different copulas

Station	Copula	θ	λ_{up}	Value
EP1	Clayton	8.93	0	0
	Frank	21.69	0	0
	Gumbel	6.36	$2-2^{1/\theta}$	0.885
EP2	Clayton	5.72	0	0
	Frank	23.84	0	0
	Gumbel	5.20	$2-2^{1/\theta}$	0.857
EP3	Clayton	3.15	0	0
	Frank	15.40	0	0
	Gumbel	3.42	$2-2^{1/\theta}$	0.775

$$\lambda_{lw} = \lim_{v \to 0} P\{X_1 \le F_1^{-1}(v) \mid X_2 \le F_2^{-1}(v)\} > 0, \quad (4)$$

The upper tail dependence coefficient for parametric estimators for Archimedean copula with their value is given in Table 4. In this study, the nonparametric for λ_{up} are used in computing the upper tail dependence coefficient. They include LOG (Coles *et al.* 1999, p. 339), the secant (SEC) (Joe 1997), and the Caperaa, Fougeres, and Genest (CFG) estimators (Capéraà *et al.* 1997, p. 30). The LOG and the SEC estimators require a threshold, whereas the CFG estimator does not need a threshold and was utilized in this research, differently with the CFG estimator, the nonparametric upper tail dependence coefficient is shown by this equation:

$$\widehat{\lambda_{up}^{CFG}} = 2 - 2\exp\left\{\frac{1}{n}\sum_{i=1}^{n} log\left[\frac{\sqrt{\log\frac{1}{u_i}\log\frac{1}{v_i}}}{\log\frac{1}{\max(u_i\,,v_i)^2}}\right]\right\} \quad (5)$$

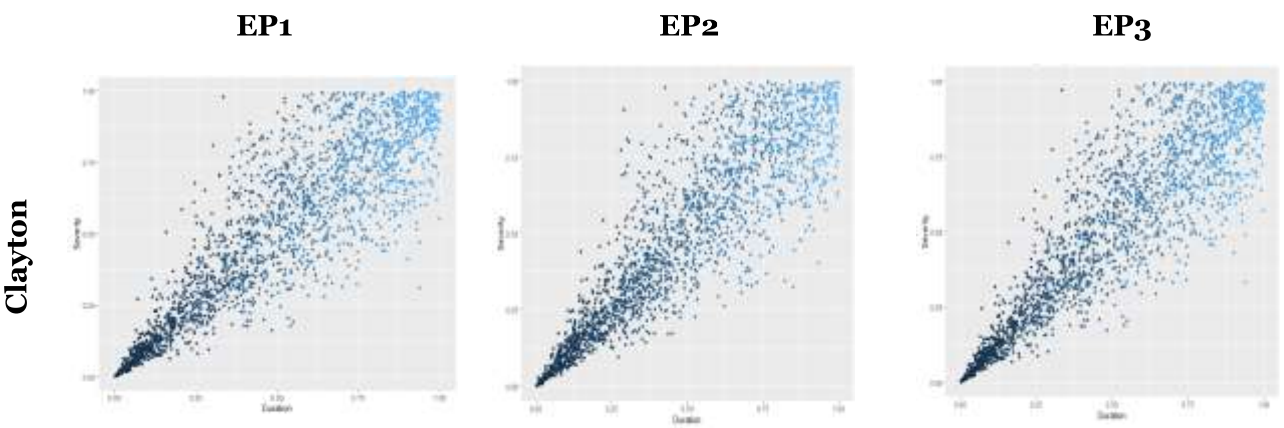
where $u_i = F_X(x_i)$ and $u_i = F_Y(y_i)$. The upper tail dependence coefficient for this study using this estimator is generated as 0.994 (EP1), 0.999 (EP2) and 0.991 (EP3), which in comparison with Table 4, indicated Gumbel–Hoggard to be the best fitted copula.

Therefore, Gumbel-Hougaard based bivariate flood distributions for the EP1 (Eq. 6), EP2 (Eq. 7) and EP3 (Eq. 8) stations can be expressed as follows:

$$exp\left[-(\log u)^{6.36} + (-\log v)^{1/6.36}\right] \quad (6)$$

$$exp\left[-(\log u)^{5.72} + (-\log v)^{1/5.72}\right] \quad (7)$$

$$exp\left[-(\log u)^{3.42} + (-\log v)^{1/3.42}\right] \quad (8)$$



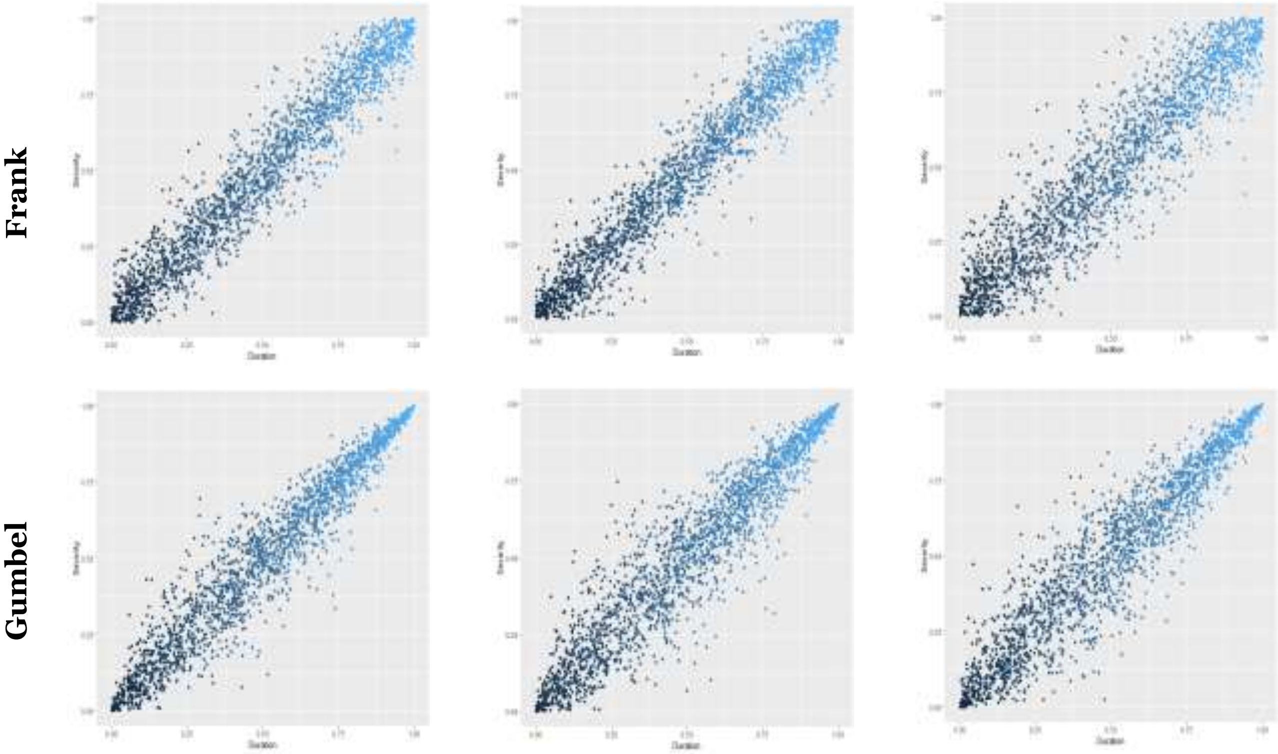


Figure 2. Copula based joint distribution for three stations, Clayton, Gumbel and Gumbel.

V. CONCLUSION

A study of rainfall variables, such as severity and duration at different flood prone areas, that are commonly used in hydrological modelling to analyze the tail dependence in extreme value analysis which describes the dependence in tail of a multivariate distribution. The copula method is used to form a derivation and conditional distribution for the selected station. The selected areas are located in EP1, EP2 and EP3 areas in Pahang. Among the three copulas of Archimedean families that are used in this study,

the best-fitted copula is Gumbel–Hoggard, chosen in terms of tail dependence and nonparametric test, CFG, which is more highly correlated at the tails on the upper part of the

VI. ACKNOWLEDGEMENTS

The funding for this research is fully supported by the Ministry of Higher Education Malaysia under the Fundamental Research Grant Schemes (FRGS /1/2015/SG04/UPNM/02/2).

[1] Sklar A. (1959). 'Fonction de re'partition a' n dimensions et leurs marges', Publications de L'Institute de Statistique, Universite' de Paris, vol. 8, pp. 229–231.

[2] Zhang Z. , Zhang C. & Cui Q. (2017). 'Random Threshold Driven Tail Dependence Measures with Application to Precipitation Data Analysis', Statistica Sinica, vol. 27, no. 2, pp. 685-709.

[3] Mirabbasi, R., Fakheri-Fard A. and Dinpashoh Y. (2012). 'Bivariate drought frequency analysis

using the copula method', Theoretical and Applied Climatology, vol. 108, no. 1-2, pp. 191–206.

[4] Abdul Rauf U. F. & Zeephongsekul P. (2014). 'Analysis of Rainfall Severity and Duration in Victoria, Australia using Non-parametric Copulas and Marginal Distributions', Water Resource Management, vol. 115, no. 13, pp. 4835–4856.

[5] Cavanaugh N. R. & Gershunov (2015). 'Probabilistic tail dependence of intense precipitation on spatiotemporal scale in observations, reanalyses, and GCMs', Climate Dynamics, vol. 45, no. 11-12, pp. 2965–2975.

[6] Zhang L. & Singh V. P. (2006). 'Bivariate flood frequency analysis using the copula method', Journal of Hydrologic Engineering, ASCE, vol. 11, no. 2, pp. 150–164.

[7] Coles S., Heffernan J. & Tawn J. (1999). 'Dependence measures for extreme value analyses', Extremes, vol. 2, no. 4, pp. 339–365.

[8] Joe H. (1997). 'Multivariate Models and Dependence Concept', 1st edn, Chapman and Hall, New York.

[9] Cap´era`a P., Foug`eres A. L. & Genest. C. (2000). 'Bivariate distributions with given extreme value attractor', Journal of Multivariate Analysis, vol. 72, no. 1, pp. 30–49.