Modelling rainfall severity and duration in north-eastern Victoria using Copulas

U.F. Abdul Rauf [a] **and P. Zeephongsekul**[a]

[a] *School of Mathematical and Geospatial Sciences,*
RMIT University, Victoria
Email: ummulfahri.abdulrauf@rmit.edu.au

Abstract: Since 2010, there have been many areas in Victoria, particularly in the north-eastern provinces, that were badly affected by heavy rain causing major flooding. Rainfall duration and severity are commonly regarded as important variables in the field of hydrology and water resource management which affect flooding intensity.

However, rainfall characteristics are often multidimensional and hence require the joint modelling of several random variables. The selected rain gauge stations are located in the flood areas that historically have large amounts of rain per year. Monthly rainfall data for an 81 year period were supplied by the Australian Bureau of Meteorology (BOM) for 2 rain gauge stations in north-eastern Victoria. Traditionally, the pair-wise dependence between variables such as intensity, duration, and severity has been described using the classical family of bivariate distributions. The main restriction of this approach is that the marginal distribution of each of the two random variables must be characterized by the same parametric family of univariate distribution. However in reality the variables do not have the same marginal distribution. In the past decade, Copula models were introduced to circumvent this limitation. Copulas have also been successfully applied to various multivariate simulation studies in civil, mechanical, offshore engineering and flood frequency analysis.

In this paper, we will first discuss the Standard Precipitation Index, (SPI) which was developed to quantify rainfall severity and duration. Secondly, estimation of univariate marginal distributions were performed and followed by the identification of the generator and parameter of appropriate copulas and, finally, the determination of the joint probability distributions are obtained. In our research, the bivariate rainfall distributions were obtained using the family of Archimedean Copulas such as Gumbel-Houggaard and Clayton family of copulas. The models are then used to describe various rainfall events in Victoria and various goodness of fit tests will be performed.

This study aims to model the joint distributions of rainfall severity and duration in north-eastern Victoria and then to apply the results in considering various aspects of recent flooding that occurred in north-eastern Victoria.

Keywords: Standard Precipitation Index, rainfall, severity, duration, Copulas

1. INTRODUCTION

The climate of Victoria is characterised by a range of different climate zones, from the warm, dry Mallee region of the northwest to the alpine snowfields in the northeast. Median annual rainfall ranges from less than 300 mm in parts of the Mallee to in excess of 2500 mm in the wettest parts of the mountainous regions (BOM, 2010).

In 2010, many areas of Victoria, particularly the north-eastern region, were severely affected by heavy rainfall causing major flooding throughout the region. Rainfall intensity, duration and its severity are classified as important variables in the field of hydrology and water resource management. However, rainfall characteristics are often multidimensional and hence require the joint modelling of several random variables. Traditionally, the pair-wise dependence between variables such as intensity, duration, and severity has been described using the classical family of bivariate distributions. The main restriction of this approach is that individual behaviour of the two variables must be characterized by the same parametric family of univariate distributions. However, in reality the variables do not have to have the same type of marginal distribution. In the past decade, Copula models, developed by A. Sklar in 1959, were introduced to circumvent this limitation and have been successfully applied in various fields such as civil, mechanical, offshore engineering in drought and flood analysis.

As the use of univariate distributions has become a standard approach in rainfall analysis, the application of multivariate distributions has not been pursued to any great extent, despite it being more relevant to these problems; see, e.g Zhang and Singh (2007), Genest and Favre (2007) and Kao (2008). Apart from the mathematical complexity incurred with working in higher dimension, one of the major reasons for the unpopularity of multivariate analysis is the limited choice of multivariate distributions (e.g. multivariate Gaussian or Gumbel's bivariate exponential distributions). These conventional multivariate distributions are associated with specific marginal (normal or exponential), and therefore not suitable for general applications (c.f. Kao (2008)). It is especially problematic for hydrologic applications, since many variables have complicated marginal distributions. With the use of copulas, these limitations can be solved. Copulas models allow multidimensional distribution to be defined based on the marginal distributions being heterogeneous and not necessarily belonging to same family of distribution. This will provide a more flexible approach to modeling multidimensional distributions. For example, when modeling the joint distribution of rainfall severity and duration, it is often necessary to model the former, but not the latter, using an extreme value distribution. Therefore it is important that a function that joins these two different marginal distributions, i.e. a copula, be available so that the appropriate bivariate distribution can be constructed from these marginal distributions. Since the choice of copulas is very extensive (refer to Nelsen, 2006), this gives rise to a large class of bivariate distributions that can be used to model these bivariate random variables.

The data is first transformed into appropriate indices by a transformation known as the Standard Precipitation Index (SPI) and these indices are subsequently used to generate the rainfall intensity and duration. The resulting severity and duration data are fitted using the family of Archimedean copulas. The final estimated models are applied to analyse various rainfall events in Victoria. This paper describes the rainfall distributions in north-eastern Victoria and, the copula was used to model the joint distributions of rainfall severity and duration in north-eastern Victoria and then to apply the results in considering various aspects of recent flooding that occurred in north-eastern Victoria.
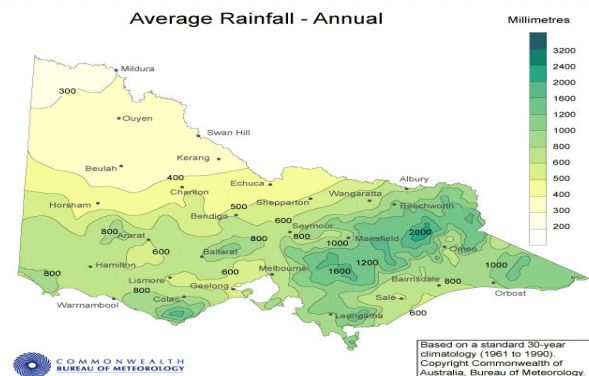

Figure 1. Annual Rainfall in Victoria

2. THE TWO STATIONS IN NORTH-EASTERN VICTORIA

Average annual rainfall (shown in Figure 1) comprises the average rainfall in Victoria per year. From the figure, it is clear that north-eastern areas are wet compared to other regions. In this study, rainfall data over 81 years were taken from two stations which are located in Harrietville and Whitlands. These two stations are considered as the wet areas in north-eastern Victoria with average of rainfall of about 1400mm per year.

The rainfall data was provided by Australian Bureau of Meteorology for a period from 1930 until 2010. Table 1 gives brief summary of these two stations.

Table 1. Summary of two selected stations in North-eastern Victoria

Station No	Station Name	Latitude	Longitude	Annual Mean (mm)
83012	Harrietville	36.89° S	147.06°E	1417.8
83032	Whitlands	36.85° S	146.32°E	1410.0

3. RAINFALL DISTRIBUTION

3.1. Overview

Rainfall intensity is one of the key issues affecting water resource management. The study of rainfall characteristics has been an area of great interest for many researchers studying rainfall patterns in Australia (Hassan and Dunn, 2011, Piantadosi *et al.* 2009, Whiting et al. 2003, Zakaria et al. 2010).

3.2. Standard Precipitation Index

The Standard precipitation Index (SPI) was first employed by McKee et al. in 1993 for defining and monitoring severe drought conditions. This index was developed by using the probability of observed precipitation as a parameter and then transformed it into an index by Dubrovsky et al. (2009). SPI is currently used to monitor drought and rainfall intensities in over 60 countries. The main advantage of SPI is that it has multiple scales that allow for temporal flexibility in the evaluation of rainfall conditions and water supply. The SPI calculation for any area or region is based on the long-term precipitation data for a desired period. This long-term record is fitted to a Gamma distribution, which is then transformed into a normal distribution so that the mean SPI for the location and desired period is zero (Edwards and McKee, 1997). The parameters of the gamma probability density function are estimated for each set of monthly rainfall data, for each time scale of interest (3-month, 6-month, 12-month, 24-month or 48-month). Referring to Thom (1966), the maximum likelihood estimator for parameters of gamma, α and β are obtained by

$$\hat{\alpha} = \frac{1}{4A}\left(1+\sqrt{1+\frac{4A}{3}}\right), \qquad \hat{\beta} = \frac{\bar{x}}{\hat{\alpha}}, \tag{1}$$

where

$$A = \ln(\bar{x}) - \frac{\sum \ln(x)}{n} \ . \qquad n = \text{number of rainfall observations} \tag{2}$$

The resulting parameters are then used to find the cumulative probability of an observed rainfall event for the given month and time scale for the selected stations. This cumulative probability is then transformed to the standard normal random variable which is the value of SPI. Detailed discussion on how to find the cumulative probability of gamma was explained by Thom (1966). Positive values of SPI indicate that the precipitations are greater than their median value and negative values indicate that they are less than the median value. As the SPI values are normalised, wet and dry events can be represented in the same way, and wet periods can also be monitored using the SPI (Dubrovsky et al, 2009). At least 30 years of continuous monthly precipitation data were needed to calculate the SPI. The data are first transformed into these indices then subsequently used to generate the rainfall severity and duration where rainfall severity is defined by

$$S = -\sum_{i=1}^{D} SPI_i \ , \tag{3}$$

and D was rainfall duration. This paper evaluates the SPI on 3- and 6-month time scales using monthly precipitation totals for two rainfall stations in north-eastern Victoria (described in Table 1) from January 1930 to December 2010 representing wet and flood areas. McKee et al. (1993, 1995) proposed a seven-category classification for the SPI: extremely wet $(S > 2.0)$, very wet (1.5 to 1.99), moderately wet (1.0 to 1.49), near normal (-0.99 to 0.99), moderately dry (-1.49 to -1.0), severely dry (-1.99 to -1.5), and extremely dry $(S < -2.0)$.

4. COPULAS

4.1. Concept of Copulas

Originally, the name of "copula" was taken to emphasize the way in which copula "couples" a joint distribution function to its univariate marginal distribution. According to Sklar' theorem (Nelsen, 2006), if $F_{X,Y}(x, y)$ be a joint distribution with margins $F_X(x)$ and $F_Y(y)$, then there exists a copula C such that

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y)) . \tag{4}$$

Since the $F_X(x)$ and $F_Y(y)$ are cumulative distribution functions, the domain of $C(F_X(x), F_Y(y))$ is the unit square. By assuming that the marginal distributions are continuous (Shiau, 2006) with the probability density functions $f_X(x)$ and $f_Y(y)$, the joint probability density functions then can be expressed as

$$f_{X,Y}(x, y) = c(F_X(x), F_Y(y)) f_X(x) f_Y(y) , \tag{5}$$

where c is density function of C, defined as

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} . \tag{6}$$

4.2. Archimedean Copulas

In this section, we discuss an important family of copulas, the family of Archimedean copulas, which will be used in this study. Details and properties of this important family of copulas can be found in the authoritative text by Nelsen (1999). Archimedean copulas originally appeared not in statistics, but rather in the study of probabilistic metric spaces, where they were used as part of the development of probabilistic version of triangle inequality. However, recently, these copulas have been applied in research in hydrology and water resource management. Zhang and Singh (2006) and Shiau (2006) used Archimedean copulas to represent the joint distribution of rainfall and drought characteristics. Archimedean copulas were very useful because they were easy to construct and possess several beneficial features (Kao, 2008). They are also versatile since different choices are available depending on the range of dependence levels that exist between the variables. According to Kao (2008) and based on the measure of dependency known as the Kendall's τ, Gumbel-Hougaard copula is suitable only $\tau > 0$, Ali-Mikhail-Haq for weaker dependence (-0.18 < τ < 0.33), while Clayton, Frank, and Genest-Ghoudi are suitable for both positive and negative dependencies.

Archimedean copulas are defined by its generator φ where $\varphi(x)$ is a continuous strictly decreasing function the maps from [0,1] to [0, ∞] such that $\varphi(1) = 0$. Then, assuming that φ^{-1} is the inverse of φ, the Archimedean copula is defined by

$$C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)) . \tag{7}$$

Various generators exist for different Archimedean copulas and these can be found in Table 4.1 of Nelsen (1999). In general, φ is dependent on a parameter θ and will therefore be denoted by φ_θ.

There is an intrinsic relationship between Kendall's tau and the generator of an Archimedean copula and it is given by

$$\tau = 1 + 4 \int_0^1 \frac{\varphi_\theta(t)}{\varphi'_\theta(t)} dt . \tag{8}$$

For example, the relationship between τ and θ for the Clayton copula is given by:

$$\theta = \frac{2\tau}{1 - \tau} . \tag{9}$$

However, the definition does not extend to a multivariate cases because there will be multiple values of tau one for each pairing. The second Archimedean copula that was used in this study is the Gumbel-Hougaard copula. The relationship between Kendall's tau, τ and the Gumbel copulas parameter θ is given by:

$$\theta = \frac{1}{1-\tau} \; . \tag{10}$$

4.3. Parameter Estimation

In this paper, the bivariate random variables used were rainfall severity and rainfall duration. The marginal distributions used were gamma and exponential for rainfall severity and rainfall duration respectively. The parameters of these distributions were estimated by the maximum likelihood method. The maximum likelihood estimations of univariate marginal are performed first, and then followed by an estimation of the copula dependence parameters using *R Software*. The package of copulas software was developed and written by Yan (2007) and this was used to obtain the copulas parameter. The package provides functions to evaluate the log-likelihood of data under the specified copula model (Yan, 2007) and to obtain maximum likelihood estimates. By assuming that the rainfall severity and rainfall duration are distributed as respective gamma and exponential distributions (Shiau, 2006), we obtained $\hat{\alpha}$ and $\hat{\beta}$ as the estimated parameters of the gamma distribution and $\hat{\lambda}$ the estimated parameter of exponential distribution. These parameters are displayed in Table 2 below.

5. JOINT RAINFALL SEVERITY AND DURATION DISTRIBUTIONS

5.1. SPI Distributions

The annual mean precipitation for Harrietville is 1417mm and Whitsland is 1410mm. The monthly SPI series is demonstrated in Figure 2 for a period from 1930 until 2010. For example, the 3 month SPI calculated for March, 1930 would have utilized the precipitation total of January, 1930 through March, 1930. For Whitsland's station, the number of extremely wet events (SPI > 2.0) occurred about 13 times within the studied period. However, for Harrietville's station, the number of extremely wet events occurred slightly less about 8 times within the period.
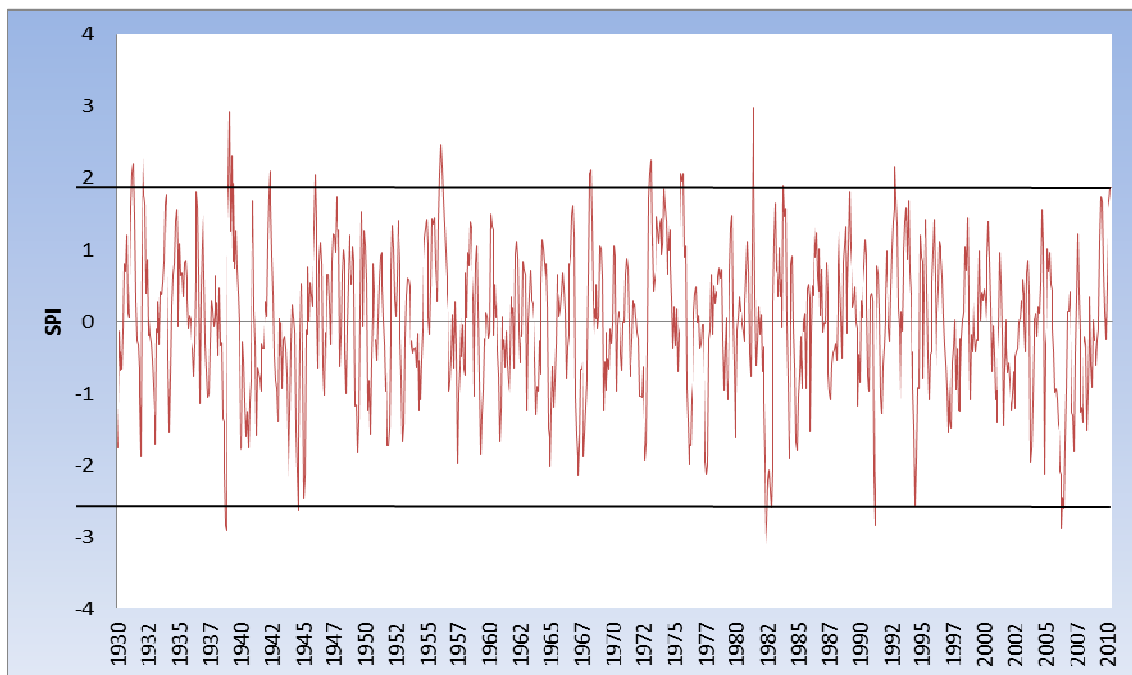

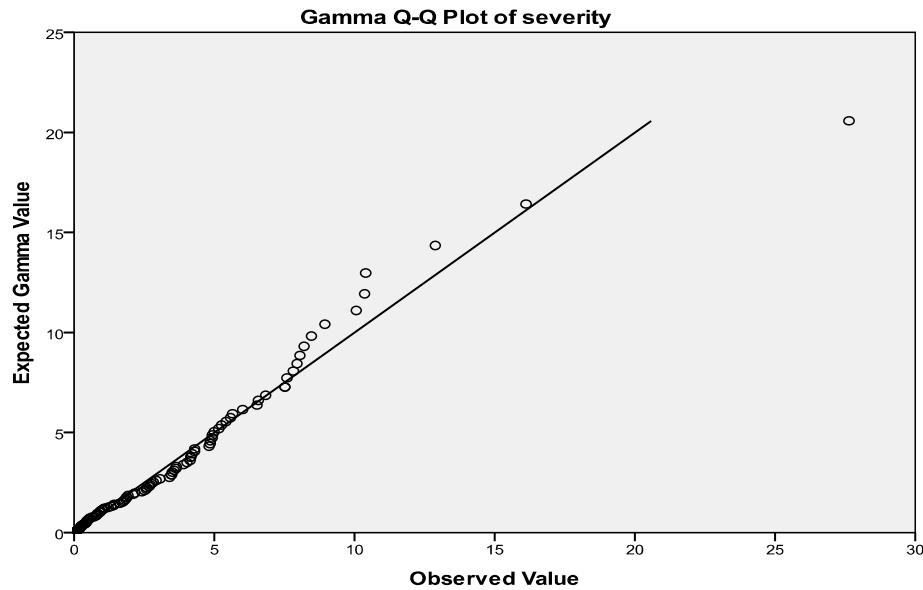**Figure 2.** Wet and Dry Events for Whitsland's Station

Figure3. Q-Q plot for 3-month SPI Readings (Harrietville's Station)

5.2. Marginal Distributions

From the Q-Q plot and cumulative gamma (exponential) distribution function for observed and fitted rainfall severity (duration), the data of rainfall severity and duration are best fitted as gamma and exponential distributions, respectively. Standard R functions were used to calculate the maximum likelihood estimate for all parameters. The $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\lambda}$ values found using the MLE for Harrietville and Whitsland are shown in Table 2. However, all set of data for rainfall severity and duration show that there are significant relationship between the two rainfall variables. Traditional modeling of bivariate data was not suitable in this case. Copula methods are used to construct the joint distribution of rainfall severity and duration.

Table 2. The values of parameter and log-likelihood function of copulas

Station	SPI Readings	Copulas	Marginal 1		Marginal 2	Parameter	Log-likelihood
			$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\theta}$	**Function**
(83012)	3 months	Clayton	1.867	0.601	0.469	8.79	-2427.38
Harrietville	6 months	Clayton	1.153	0.253	0.319	15.04	-2774.72
	3 months	Gumbel-Hougaard	1.867	0.601	0.469	4.80	-2512.77
	6 months	Gumbel-Hougaard	1.153	0.253	0.319	8.05	-2952.631
(83032)	3 months	Clayton	2.094	0.643	0.452	11.95	-2196.36
Whitsland	6 months	Clayton	1.364	0.298	0.327	15.61	-2770.18
	3 months	Gumbel-Hougaard	2.094	0.643	0.452	6.78	-2354.43
	6 months	Gumbel-Hougaard	1.364	0.298	0.327	8.51	-2781.06

5.3. Copulas for Rainfall Severity and Duration

Two Archimedean copulas are used in this paper. The choice is based on their frequent application in recent literature to describe bivariate hydrological phenomena. The selected copulas are Clayton and Gumbel-Hougaard. Since the likelihood function of Clayton copulas is greater than Gumbel-Hougaard (refer to Table 2), the former is selected as the copula to model the bivariate distribution linking rainfall severity and duration in this study. The result of the joint rainfall model using Clayton copula is:

$$F_{S,D}(d,s) = C(F_S(s), F_D(d))$$
$$= C(u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}$$

(11)

where $F_S(s)$ and $F_D(d)$ are the cumulative rainfall severity and duration. Note again that the rainfall severity and duration distribution functions the gamma and exponential respectively.

6. DISCUSSION AND CONCLUSIONS

A study of rainfall characteristics is an essential tool for water resource management. An application of the Standard Precipitation Index (SPI) is used to monitor its severity and duration.

Copula method is then applied to model the joint distribution of these bivariate variables in north-eastern Australia. The Clayton copula is the best copula to construct the dependence structure of bivariate rainfall distribution for the selected stations. The region selected for this study is located in northeastern Victoria.

In the future, studies will focus on other two-parameter copulas and also take into account other states in Australia. Further studies will also be undertaken to detect changes in other precipitation characteristics in the targeted areas.

ACKNOWLEDGMENTS

REFERENCES

BOM (Bureau of Meteorology) (2010), Average annual and monthly rainfall. Australian Government: Bureau of Meteorology *http://www.bom.gov.au/climate*

Dubrovsky, M., Svoboda, M. et al. (2009). Application of relative drought indices in assessing climate-change impacts on drought conditions in Czechia. *Theoretical and Applied Climatology,* 96(1), 155-171.

Edwards, D.C., and McKee, T.B. (1997). Characteristics of 20th Century Drought in United State at Multiple Time Scales. Climatology Report No. 97-2, Colorado State University, Fort Collins, CO, USA, pp 55.

Genest, C. and Favre, A.C. (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering,* 12(4), 347-368.

Kao, S.C. (2008). Multivariate statistical analysis of Indiana hydrologic data. Ph.D. Thesis. Purdue University, Indiana, United States.

Khan, S., Gabriel H., et al. (2008). Standard precipitation index to track drought and assess impact of rainfall on watertables in irrigation areas. *Irrigation and Drainage Systems,* 22(2), 159-177.

McKee, T.B., Doesken, N.J., and Kleist, J. (1993) The relationship of drought frequency and duration to time scales. In: Proc. 8th Conference on Applied Climatology. Anaheim, California, pp 179–184.

McKee, T.B., Doesken, N.J., and Kleist, J. (1995) Drought monitoring with multiple time scales. In: Proc. 9th Conference on Applied Climatology. Boston, Massachusetts, pp 233–236.

Nelsen, R.B. (1999). *An Introduction to Copulas*, Springer-Verlag, New York.

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

Piantadosi, J., Boland, J., et al. (2009). Generating Synthetic Rainfall on Various Timescales-Daily, Monthly and Yearly. *Environmental Modeling and Assessment,* 14(4), 431-438.

R-Software: *http://cran.r-project.org/*

Sklar, J. (1959). Fonctions de re'partition a' n dimensions et leura marges. Publ. Inst. Stat.Univ. Paris, Vol. 8, pp. 229-231.

Shiau, K. (1959). Fitting Drought Duration and Severity with Two-Dimensional Copulas. *Water Resources Management,* 20(5), 795-815.

Thom, H.C.S, (1966). Some methods of climatological analysis. W.M.O., 415, Nota tecnica 81, Geneve.

Yan, J. (2007). Enjoy the Joy of Copulas: With a Package copula. *Journal of Statistical Software,* 21(4).

Zakaria, R., Metcalfe, A.V., Howlett, P. G., Piantadosi , J. and Boland, J. (2010), Using the skew t-copula to model bivariate rainfall distribution. *ANZIAM J.,* 51, C231-C246.

Zhang, L. and Singh, V. P. (2006). Bivariate Flood Frequency Analysis Using the Copula Method. *Journal of Hydrologic Engineering,* 11(2), 150-164.

Zhang, L. and Singh, V. P. (2007). Bivariate rainfall frequency distributions using Archimedean Copulas. *Journal of Hydrology*, 332(1-2), 93-109.