

Analysis of Rainfall Severity and Duration in Victoria, Australia using Non-parametric Copulas and Marginal Distributions

Ummul Fahri Abdul Rauf · Panlop Zeephongsekul

Received: 22 April 2014 / Accepted: 18 August 2014 /
Published online: 5 September 2014
© Springer Science+Business Media Dordrecht 2014

Abstract The analysis of joint probability distributions of rainfall characteristics such as severity and duration is important in water resources management. Deriving their distributions using standard statistical techniques are often problematical due to its complexity. Standard methods usually assume that the rainfall characteristics are independent or that their marginal distributions belong to the same family of distributions. The use of copulas based methodologies can circumvent these restrictions and are therefore increasingly popular. However, the copulas and marginal distributions that are commonly used belong to specific parametric families and their adoption could lead to spurious inferences if the underlying assumptions are violated. For this reason, we recommend a nonparametric or semiparametric approach to estimate the joint distribution of rainfall characteristics. In this paper, we introduce and compare several copula-based approaches, each involving a combination of parametric or nonparametric marginal distributions conjoined by a parametric or nonparametric copula. An empirical illustration of the different approaches using rainfall data collected from six stations in the state of Victoria, Australia, demonstrated that a nonparametric approach can often give better results than a purely parametric approach.

Keywords Copulas · Parametric and nonparametric approaches · Rainfall severity and duration

U. F. Abdul Rauf · P. Zeephongsekul (✉)
School of Mathematical and Geospatial Sciences, RMIT University, Victoria, 3000, Australia
e-mail: panlop.zeephongsekul@rmit.edu.au

Present Address:

U. F. Abdul Rauf
Centre for Foundation Studies, National Defence University of Malaysia, Sungai Besi Camp, 57000
Kuala Lumpur, Malaysia
e-mail: ummul@upnm.edu.my; ummulfahri.abdulrauf@rmit.edu.au

1 Introduction

Floods occur for a number of reasons, however the primary cause behind most floods is heavy rainfall over a long period of time (McBride and Nicholls 1983; Nicholls and Wong 1990). During the past several years, the state of Victoria, Australia, has experienced severe floods, especially at higher altitudes in the western and northeastern part of the state. Heavy rainfall across north-eastern Australia contributed to this natural disaster by causing flooding in the upper reaches of many of Victoria's major rivers. These major floods caused massive destruction to existing infrastructure and resulted in hundreds of evacuations in affected areas. Regions that experienced these extreme climactic events also faced severe damages to properties and farmlands, with severe losses in incomes and consequential economic hardship. Due to the desire to understand, predict and control these recurring extreme flood events, research into areas related to occurrences of rainfall and its severity have increased significantly e.g. Pui et al. (2011), Mehrotra and Sharma (2011), Chappell et al. (2013), and Zhao et al. (2013).

The practice of using parametric distributions such as Lognormal, Weibull, Gamma, Pearson, Gumbel and other extreme value distributions to analyze rainfall and drought data has become very common among researchers in climatology and environmental studies. For example, Abdul Rauf and Zeepongsekul (2011, 2014) used parametric distributions to investigate rainfall severity and duration patterns in the state of Victoria, Australia; Shiau (2006) and Shiau and Modarres (2009) estimated the Standard Precipitation Index by fitting rainfall intensity using the Gamma distribution. As would be expected, this parametric approach does not work well for every precipitation data and appears to fit poorly near the tails of the distribution (Haghighat jou et al. 2013). To alleviate this problem, a nonparametric kernel density approach has been applied to fit precipitation data. Examples of such works are Haghighat jou et al. (2013), where nonparametric kernel density is used to estimate the annual precipitation series in Iran, and Sharma (2000), who presented a nonparametric long-term probabilistic forecast model based on estimation of the conditional probability distribution of rainfall using nonparametric kernel density estimation techniques.

Since rainfall characteristics such as intensity, severity and duration are important variables in hydrological research, deriving their joint distribution in order to study their statistical behavior is crucial. In the traditional approach, the joint distribution of these characteristics are from the same parametric family of distributions. For example, Yue (2000) model the annual maximum storm peaks and amounts by a normal distribution. In Shiau (2003), a bivariate extreme value distribution with Gumbel marginal distributions is used to model extreme flood events characterized by flood volumes and flood peaks. However, it is not realistic to postulate the same parametric marginal distribution for all characteristics since this assumption is usually not justified in practice. Sklar in 1959 introduced the concept of a copula function joining different marginal distributions into a multivariate distribution (Joe 1997; Nelsen 2006; Genest and Favre 2007). Since its introduction, copulas have been widely applied in many disciplines, especially by economists, climate scientists, hydrologists and actuarial scientists. Copulas was introduced in rainfall studies (Serinaldi et al. 2009; Kao and Govindaraju 2007, 2008, 2010) as it provides a more flexible approach that allows different types of marginal distributions to coexist and join together by a copula to form a multivariate distribution. Another advantage of using copula is the relaxation of the independence assumption which is inappropriate in modeling hydrological variables (De Michele and Salvadori 2003; Genest and Favre 2007; Zhang and Singh 2007).

The study of rainfall characteristics is one of the key research areas among researchers working in the area of water resources management. Two pertinent rainfall characteristics are the severity and duration of rainfall in a geographic region. Prior to embarking upon developing any viable water plan, the determination of the joint distribution of these two rainfall characteristics is very important and this will be an objective of this study. The first application of a nonparametric approach to copulas analysis was by Deheuvels (1979) who used empirical copula with empirical marginal distributions. Genest et al. (1995) used a semiparametric method by postulating that the copula function belongs to a parametric family whose parameter is estimated using maximum likelihood estimation. Scaillet and Fermanian (2002) applied a similar method to estimate copulas where there is temporal dependence, a situation prevalent in financial time series data. In another study, Chen and Huang (2007) proposed a bivariate kernel copula which assists in alleviating the problem of boundary bias.

The novelty of this paper is to introduce a nonparametric and a semiparametric approach to the analysis of a bivariate rainfall data model for two rainfall characteristics, namely severity and duration. This will involve using both a parametric and nonparametric copula as well as marginal distributions. To the best of our knowledge, the nonparametric and semiparametric approaches have not been exploited to any great extent and certainly not in the context of analyzing Australian rainfall data. Three approaches will be used here. The first approach combines nonparametric marginal distributions with a parametric copula. The role is reversed in the second approach where the marginal distributions are assumed parametric and the copula nonparametric. The third approach utilizes both nonparametric marginal distribution and nonparametric copula. These approaches supplement our earlier work (Abdul Rauf and Zeepongsekul 2014) where a parametric approach was used to estimate both the marginal distributions and the copula.

This paper is organized as follows. After this Introduction, Section 2 briefly describes the study area of the selected rainfall stations. Section 3 consists of the theoretical framework for the proposed models. Section 4 concludes the paper with some suggestions for future work.

2 Study Area and Data

The focus of this study is the state of Victoria, Australia. Victoria is located in the south-eastern part of Australia. Geographically, it is the smallest mainland state (refer to Fig. 1). Melbourne, the capital of Victoria, is Australia's second largest city. Melbourne has four seasons although the climate is highly variable between seasons. Summer occurs from December to February, autumn during March to May, winter from June to August and spring from September to November. Annual maximum temperatures for Melbourne occur in the summer months of January and February. During this time, the climate is hot with dry spells. Although winter has the coldest temperature October tends to be the wettest month. Rainfall varies across the state, for example in 2010, maximum rainfall (single point estimate) in Victoria ranged from 425 millimetres to 1,250 millimetres.

Given that a large amount of time series observations are required in order to construct reliable bivariate statistical models of joint distributions of rainfall duration and severity, a 61-year record of data during the years 1950–2010 is obtained from the Bureau of Meteorology (BOM) Australia, for analysis. Table 1 gives the geographical coordinates, annual mean rainfall and percentages of missing observations of the six selected rainfall stations used to collect rainfall data.

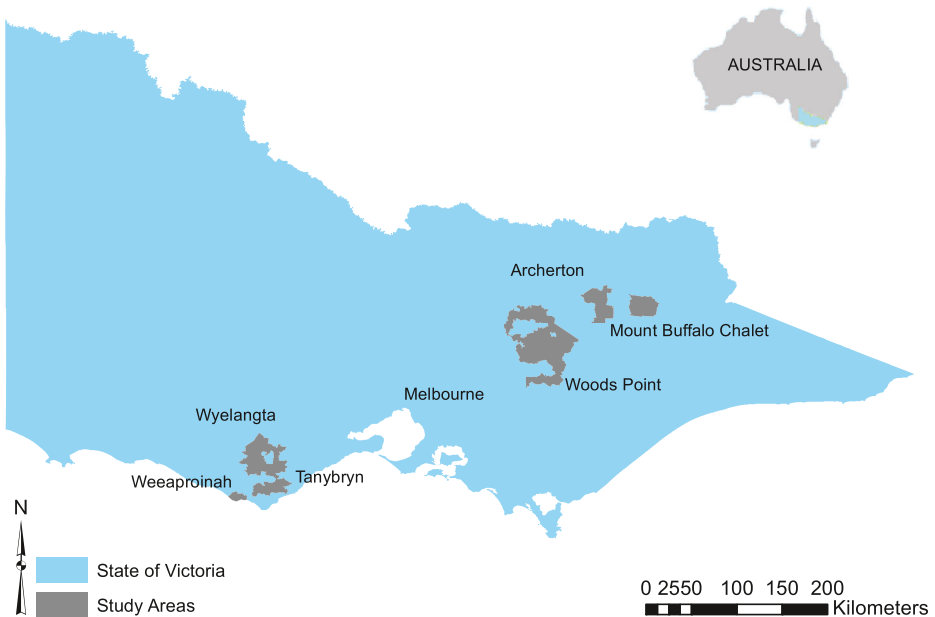


Fig. 1 Map of Victoria Australia and the coordinates of the six rainfall stations

3 Models and Data Analysis

3.1 Kernel Density Estimation

Kernel density estimation is one of the most popular non-parametric method used to estimate the probability density function (pdf) of a random variable. It is easy to apply and can often uncover structural features in the data set which a parametric approach might not reveal. Recently, nonparametric methods have been extensively applied to rainfall studies Sharma and Lall (1999), Sharma (2000), Haghightat jou et al. (2013), and Kim et al. (2006) used kernel density estimation to estimate the rainfall probability density function. In this paper, we will estimate the marginal distribution functions using both parametric and non-parametric approach. For the parametric approach, we will fit the Gamma, Weibull,

Table 1 Geographic locations of the six (6) selected stations in Victoria

Station No	Station Name	Latitude	Longitude	Annual Mean Rainfall (mm)	% Missing Observations
83000	Archerton	36.91°S	146.24°E	1368.1	0.04
83033	Woods Point	37.57°S	146.25°E	1472.5	0.10
83073	Mount Buffalo Chalet	36.72°S	146.82°E	1882.6	0.05
90076	Tanybryn	38.68°S	143.68°E	1621.9	0.19
90083	Weeaprounah	36.84°S	143.51°E	1936.1	0.00
90087	Wylangta	38.66°S	143.45°E	1949.8	0.00

Log-normal and Exponential to the data and, for the nonparametric approach, kernel density estimation is used to estimate the marginal distribution of both rainfall characteristics, i.e. its severity and its duration.

Given a random sample X_1, \dots, X_n from a population with a continuous, univariate probability density function (pdf) $f(\cdot)$ and cumulative distribution function (cdf) $F(\cdot)$, the kernel density estimator of $f(\cdot)$ is defined as

$$\hat{f}(x, h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \tag{1}$$

where $K(\cdot)$ is the kernel function and h the bandwidth. Under some mild conditions (c.f. Wand and Jones 1995), the kernel density estimator is a consistent estimator of the true pdf $f(\cdot)$. Kernel functions are symmetric unimodal functions about the origin and they satisfy the conditions $\lim_{|x| \rightarrow \infty} |x|K(x) = 0$ and $\int_{-\infty}^{\infty} x^2 K(x) dx < \infty$. The standard kernel functions are the Gaussian, Triangular, Biweight and the Epanechnikov functions c.f. (Wand and Jones 1995). The nonparametric kernel estimator of $F(x)$ is obtained by integrating (1) giving

$$\hat{F}(x, h) = \frac{1}{nh} \sum_{i=1}^n \int_{-\infty}^x K\left(\frac{u - x_i}{h}\right) du \tag{2}$$

$$= \frac{1}{n} \sum_{i=1}^n \mathbf{K}\left(\frac{x - x_i}{h}\right) \tag{3}$$

where $\mathbf{K}(x) = \int_{-\infty}^x K(u) du$.

It is well known that the choice of a kernel function does not significantly affect the quality of the approximation and in this paper we use the Gaussian kernel which is defined by

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right). \tag{4}$$

The choice of the bandwidth h , on the other hand, does affect the quality of the approximation. For the bandwidth, we adopt the value suggested by Silverman (1986) which is

$$h = 0.9An^{-1/5} \tag{5}$$

where

$$A = \frac{\min\{\text{standard deviation, interquartilerange}\}}{1.34}.$$

It was shown in Silverman (1986) that this bandwidth adapts well to the Gaussian kernel and is a robust measure of the spread of the underlying distribution of which it is estimating. Using simulated data, it was also shown that this bandwidth gives a reasonable Mean Integrated Square Error (MISE) values where

$$\text{MISE}(\hat{f}) = E\left(\int [\hat{f}(x) - f(x)]^2 dx\right) \tag{6}$$

as well as revealing the bimodality and skewness of the underlying distribution.

3.2 Standard Precipitation Index

Standard Precipitation Index (SPI) was introduced by McKee et al. (1993) for monitoring drought and was used to identify extreme drought events and evaluate their severity. SPI

measures the precipitation deviations based on the long-term precipitation data for a given period. One of the most significant aspect of this index arises from the fact that it is easy to calculate and interpretable using standard probabilistic analysis (Guttman 1998). SPI is calculated as the precipitation value y such that,

$$\Phi(SPI) = F(y) \quad (7)$$

or, equivalently

$$SPI = \Phi^{-1}(F(y)) \quad (8)$$

where $F(\cdot)$ is the cdf of the precipitation variable and $\Phi(\cdot)$ is the cumulative standard normal distribution function. The standard procedure used to compute SPI is to first fit the precipitation data to the Gamma distribution using Maximum Likelihood Estimation (MLE) method (Yusof et al. 2013), and then apply Eq. 8 to obtain the SPI values. The details SPI categories and how to monitor the rainfall characteristics (severity and duration) can be found in Yusof et al. (2013) and Abdul Rauf and Zeepongsekul (2014).

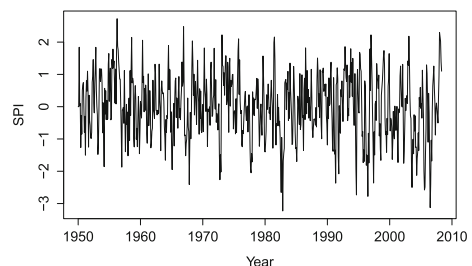
One key disadvantage in using the above approach is that for many precipitation time series data, the Gamma distribution does not provide a good fit to rainfall data. This then called upon an alternative approach and, when uncertain as to which parametric distribution to use, it is best to adopt a nonparametric approach since it imposes less restriction on the underlying distributions. A nonparametric approach has been introduced by Cancelliere et al. (2007) in drought forecasting where they proposed two methodologies for forecasting seasonal SPI. Kim et al. (2006) used a nonparametric local polynomial estimator which called upon a kernel smoother to build an explicit model for the equiprobable transformation of the cumulative distribution functions of rainfall. In this study, instead of using gamma distribution, we used kernel density estimation in fitting the precipitation data.

The input data for this study consists of monthly SPI values computed in a 3-month time scale for the period from June 1950 to June 2010. For example, the monthly SPI calculated in a 3-month time scale at the end of November deploys the precipitation total for September, October and November in that particular year. Similarly, the 3-month SPI calculated for November 1950 would have used the precipitation total of September 1950 to November 1950 in order to calculate the index. The indices used in this study were prepared by adopting the procedure employed by Kim et al. (2006). Figure 2 presents the nonparametric SPI for Acherton Station, Victoria.

3.3 Copulas and Semiparametric Models

The traditional approach in building multivariate distributions in the field of hydrology has several limitations due to the fact that hydrological variables such as rainfall severity and duration, are highly dependent and their distributions can differ significantly from each others. These restrictions can be accommodated with the introduction of copulas into the

Fig. 2 Nonparametric SPI Indices for Archerton Station, Victoria (1950-2010)



modeling process. Copulas, introduced in 1959 by A. Sklar, facilitate the modeling of the dependency between variables and allow the flexibility of choosing the marginal distributions of the individual variables. This attribute is akin to a multivariate normal distribution where the mean vector and covariance matrix jointly determines the multivariate distribution, although in this case, the marginal distributions are also normal. In the case of copulas, the marginal distributions and copula function will determine the joint distribution and does it uniquely if these marginal distributions are continuous functions. The name *copula* is derived from the Latin word *copulare* which means to *couple* or *join*. This term lends emphasis to a copula’s role in linking the univariate marginal distribution functions to form a joint distribution function. For two variables, Sklar’s theorem (Nelsen 2006) states that if $F_{X,Y}(x, y)$ is a joint distribution function of a bivariate random variables (X, Y) with marginal distributions $F_X(x)$ and $F_Y(y)$ respectively, then there exists a copula function $C(\cdot)$ such that

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y)). \tag{9}$$

If both $F_X(x)$ and $F_Y(y)$ are continuous distributions, then this copula is unique for the particular joint distribution. Differentiating (9) with respect to x and y yields the joint probability density function given by

$$f_{X,Y}(x, y) = c(F_X(x), F_Y(y)) \cdot f_X(x) \cdot f_Y(y) \tag{10}$$

where c is the copula density function defined by

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}. \tag{11}$$

Two-dimensional copulas have been applied to model hydrological and drought phenomena by a number of researchers including Shiau (2006), Zhang and Singh (2007), Serinaldi et al. (2009), and Mirabbasi et al. (2012).

This paper will focus on three specific approaches to the analysis of bivariate rainfall variables, namely severity and duration. These approaches, summarized in Table 2, are both nonparametric and semiparametric. In the table P will refer to *Parametric* and N to *Nonparametric*. The parametric distributions that will be used to fit the marginal distributions are the Gamma, Log-Normal, Weibull and Exponential distributions as these were seen in an earlier paper (Abdul Rauf and Zeepongsekul 2014) to fit the data well. In the nonparametric case, the kernel density functions discussed in Section 3.1 is used to estimate the marginal distributions. In a related paper, Reddy and Ganguli (2012) applied both parametric and nonparametric approaches for fitting marginal distributions.

3.3.1 PN Model

In this model, the marginal distributions are parametric and the copula nonparametric. The copula is based on the Beta kernel function developed by Brown and Chen (1999),

Table 2 Proposed Semiparametric Approaches

Model acronym	Marginal distributions	Copula
PN	Parametric	Nonparametric
NP	Nonparametric	Parametric
NN	Nonparametric	Nonparametric

Harrell and Davis (1982) and Chen (1999, 2000). One of the attractive features of this kernel which warranted its use here is its ability to alleviate the severe boundary bias common in many standard kernel estimators. Univariate Beta-kernel density function based on sample of uniform variables U_1, U_2, \dots, U_n with support in $[0, 1]$ is defined by

$$b(u) = \frac{1}{n} \sum_{i=1}^n K \left(U_i, \frac{u}{h} + 1, \frac{1-u}{h} + 1 \right) \tag{12}$$

where $K(\cdot, \alpha, \beta)$ denotes the Beta density function with parameters α and β given by

$$K(x, \alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} x^{\alpha-1} (1-x)^\beta \tag{13}$$

and h is the bandwidth.

For this model, we adopt the Beta kernel copula, introduced by Charpentier et al. (2006), which has the copula density function obtained from the product of Beta densities, defined by

$$\begin{aligned} \hat{c}_h(u, v) &= \frac{1}{nh^2} \sum_{i=1}^n K \left(U_i, \frac{u}{h} + 1, \frac{1-u}{h} + 1 \right) \\ &\quad \times K \left(V_i, \frac{v}{h} + 1, \frac{1-v}{h} + 1 \right). \end{aligned} \tag{14}$$

3.3.2 NP Model

In this approach, we combine nonparametric marginal distributions which utilize the kernel density function (1) with a parametric copula. The three parametric copulas belonging to the family of Archimedean copulas are given below:

1. Clayton

$$C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \quad 0 \leq \theta < \infty. \tag{15}$$

2. Frank

$$C_\theta(u, v) = -\theta^{-1} \log \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right), \quad -\infty \leq \theta < \infty. \tag{16}$$

3. Gumbel-Hougaard

$$C_\theta(u, v) = \exp \left[-((-\log u)^\theta + (-\log v)^\theta)^{1/\theta} \right], \quad 1 \leq \theta < \infty. \tag{17}$$

Note that θ , the copulas parameter, is usually estimated using Maximum Likelihood Estimation (MLE) method.

3.3.3 NN Model

In the third approach, we have a combination of both nonparametric marginal distributions and nonparametric copula. We use a nonparametric kernel density to estimate the marginal distributions of rainfall severity and rainfall duration. The Beta kernel copula introduced in Eq. 14 is employed to obtain the joint cdf of the rainfall characteristics.

3.4 Application of Models to Data

In this section, we build a comprehensive copula-based model to estimate the joint distribution of our two rainfall characteristics. To recapitulate, data from six rainfall stations that are generally considered as flood prone areas in the state of Victoria, Australia are used in this study. Archerton, Woods Point and Mount Buffalo Chalet are located in the North-eastern region of Victoria and Tanybryn, Weaaprounah and Wyelangathree are located in the South-western region.

Figure 2 shows the monthly SPI for Acherton from year 1950 to 2010. From this graph, it is apparent that Archerton regularly faces very wet event once every four years between 1950 to 1981. But, from 1981 to 1997, the region had a relatively dry period with SPI not exceeding 2.0. After this period, this area regularly faces very wet event once every 5 to 6 years with SPI exceeding 2.0 on two occasions during these years.

For the first approach (PN model), four parametric distributions, namely Gamma, Weibull, Lognormal and Exponential distributions, were fitted to rainfall severity and rainfall duration for the six stations. All parameters for the each marginal distributions are estimated from the data using the MLE method. Table 3 present the values of the estimated parameters. We note here that since a nonparametric approach has been used to compute SPI values, the results shown in the table are different from the results produced by a parametric approach in Abdul Rauf and Zeepongsekul (2014). The goodness of fit statistics used were the Schwartz Information Criterion (SIC), also known as the Bayesian Information Criterion (BIC), and Akaike Information Criterion (AIC) defined below:

$$BIC = -2 \ln(L_{max}) + k \ln(n) \tag{18}$$

$$AIC = -2 \ln(L_{max}) + \left(\frac{2nk}{n - k - 1} \right). \tag{19}$$

Here n be number of observations, k the number of parameters to be estimated and L_{max} is the maximum value of the log-likelihood function for the estimated model. For each station, the best fitted distribution for each rainfall characteristic is subsequently selected using the AIC and BIC values and these are displayed in Table 4. The parametric models

Table 3 Parameter estimate for marginal distributions

Rainfall Characteristics	Stations No.	Gamma		Lognormal		Weibull		Exponential
		$\hat{\alpha}$	$1/\hat{\beta}$	$\hat{\mu}$	$\hat{\sigma}$	\hat{a}	\hat{b}	$\hat{\lambda}$
Severity	83000	2.52	1.02	0.69	0.61	1.47	2.76	0.41
	83033	2.51	0.95	0.76	0.62	1.49	2.96	0.38
	83073	2.92	1.29	0.64	0.56	1.57	2.56	0.44
	90076	2.60	1.02	0.73	0.60	1.51	2.85	0.39
	90083	4.10	1.84	0.68	0.49	2.02	2.53	0.45
	90087	4.45	1.98	0.69	0.47	2.10	2.55	0.43
Duration	83000	3.26	1.90	0.38	0.53	1.67	1.94	0.58
	83033	3.42	1.88	0.44	0.52	1.68	2.05	0.55
	83073	3.77	2.38	0.32	0.49	1.78	1.80	0.63
	90076	3.46	2.05	0.37	0.51	1.74	1.91	0.59
	90083	5.78	3.94	0.29	0.40	2.28	1.66	0.68
	90087	6.23	4.17	0.32	0.39	2.41	1.69	0.70

Table 4 Goodness of Fit Test for Marginal Distributions- AIC and BIC Information Criterion

Stations	Rainfall Characteristics	Distributions	AIC	BIC
Archerton 83000	Severity	Gamma	244.18	248.68
		Lognormal	228.84	233.33
		Weibull	253.76	258.27
		Exponential	268.57	270.83
	Duration	Gamma	179.96	184.46
		Lognormal	165.46	169.96
		Weibull	191.82	196.31
		Exponential	217.46	219.71
Woods Point 83033	Severity	Gamma	217.94	222.13
		Lognormal	207.59	211.78
		Weibull	225.09	229.28
		Exponential	238.49	240.59
	Duration	Gamma	159.54	163.73
		Lognormal	148.36	152.55
		Weibull	170.83	175.02
		Exponential	193.64	195.74
Mount Buffalo Chalet 83073	Severity	Gamma	247.27	251.96
		Lognormal	231.71	236.40
		Weibull	259.78	264.47
		Exponential	282.50	284.85
	Duration	Gamma	176.68	181.37
		Lognormal	160.79	165.48
		Weibull	191.55	196.23
		Exponential	226.87	229.22
Tanybryn 90076	Severity	Gamma	194.43	198.45
		Lognormal	183.87	187.89
		Weibull	201.81	205.82
		Exponential	214.40	216.40
	Duration	Gamma	136.16	142.17
		Lognormal	128.03	132.05
		Weibull	147.28	151.30
		Exponential	169.78	171.79
Weeaproinah 90083	Severity	Gamma	238.31	243.12
		Lognormal	230.97	235.78
		Weibull	247.65	252.47
		Exponential	297.54	299.94
	Duration	Gamma	145.47	150.28
		Lognormal	134.99	139.81
		Weibull	161.31	166.13
		Exponential	228.45	230.85

Table 4 (continued)

Stations	Rainfall Characteristics	Distributions	AIC	BIC
Wyalangta 90087	Severity	Gamma	225.70	230.44
		Log Normal	219.81	224.55
		Weibull	235.18	239.92
		Exponential	287.99	290.36
	Duration	Gamma	138.26	143.00
		Lognormal	130.80	135.55
		Weibull	151.65	156.39
		Exponential	223.39	225.77

that best fitted the data would give the lowest values using these criteria. The table shows that the Lognormal distribution best fitted both rainfall severity and duration. Consider Wyalangta’s station as an example, we estimated its best fitted cumulative distributions for rainfall severity and rainfall duration to be

$$F_S(s) = \frac{1}{0.47\sqrt{2\pi t}} \int_0^s e^{-\frac{(\ln t - 0.69)^2}{0.44}} dt, \quad s > 0 \tag{20}$$

and

$$F_D(d) = \frac{1}{0.39\sqrt{2\pi t}} \int_0^s e^{-\frac{(\ln t - 0.32)^2}{0.30}} dt, \quad d > 0 \tag{21}$$

respectively. The Beta kernel copula density was computed using Eq. 14 with the uniform variables U_i and V_i generated using the Lognormal distributions (20) and (21) respectively.

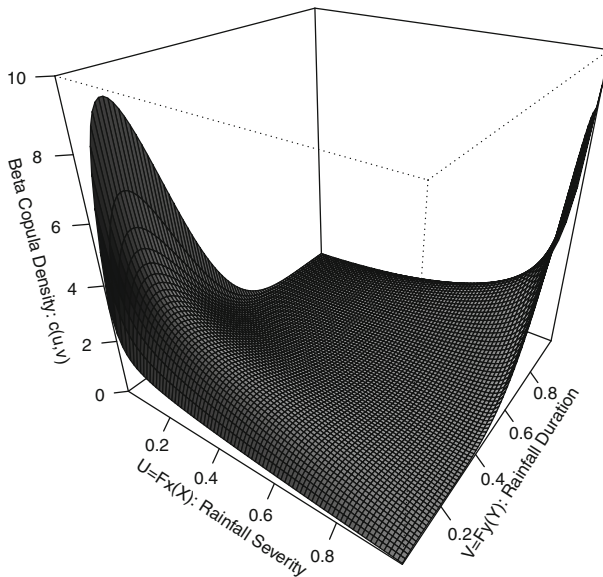


Fig. 3 Bivariate beta copula density function of rainfall severity and rainfall duration: PN Wyalangta

The resultant copula density is plotted in Fig. 3. Note that there is a bias correction at the extreme corners, i.e. with higher rainfall severity, the rainfall duration has longer duration with a higher probability and likewise at region with low rainfall intensity and duration.

We proceed with the second approach (NP model) that utilized a parametric copula model with nonparametric marginal distributions. For this purpose, the marginal distributions were estimated using kernel density functions and the three parametric Archimedean copulas given by Eqs. 15, 16 and 17 were chosen. For each of the three copulas, we estimate the parameter θ using MLE method and these are displayed in Table 5. Scatter plots of the simulated marginal distribution data were generated using the three estimated archimedean copula functions for all six stations. Due to space and page limitations, we have only displayed the plots for the first three stations (Fig. 4). The graphs show that there is a correlation between these marginal distribution data for all stations: the Frank copula demonstrates a symmetric dependence structure across all stations, while the Clayton and Gumbel-Hougaard copula were found to have a higher dependency in the tails. These dependencies were found to be higher in the left tail for the asymmetric Clayton copula, and in the right tail for the Gumbel-Hougaard copula. From the plot, we can conclude that the two rainfall characteristics show a high degree of dependence on each other and they are certainly captured by these scatterplots.

The comparison between the three copulas to determine which copula is best suited for representing the joint distributions is done with the Goodness of fit test using a variation of the Cramér-von Mises statistic, S_n introduced in Genest et al. (2009) for family of copula. S_n is here defined by:

$$S_n = \sum_{i=1}^n (C_{Ni} - C_{\theta i})^2 \tag{22}$$

Table 5 Parameter estimate for fitted Copulas

Stations	Type of Copulas	Parameter Estimates $\hat{\theta}$	Standard Error	z-value	Pr(> z)	Maximum loglikelihood
Archerton	Clayton	2.940	0.4250	6.9120	0.0000	26.48
	Frank	9.308	1.109	8.397	0.0000	32.45
	Gumbel-Hougaard	2.562	0.2633	9.729	0.0000	27.45
Woods Point	Clayton	2.550	0.4338	5.889	0.0000	17.80
	Frank	7.860	1.107	7.104	0.0000	19.96
	Gumbel-Hougaard	2.160	0.2543	8.508	0.0000	14.69
Mount Buffalo	Clayton	2.700	0.3868	6.982	0.0000	26.64
Chalet	Frank	9.160	1.053	8.7000	0.0000	35.27
	Gumbel-Hougaard	2.60	0.2556	10.19	0.0000	31.52
Tanybryn	Clayton	2.870	0.4721	6.077	0.0000	20.00
	Frank	8.960	1.222	7.33	0.0000	23.73
	Gumbel-Hougaard	2.438	0.287	8.492	0.0000	18.72
Weeaprounah	Clayton	3.011	0.4004	7.520	0.0000	32.64
	Frank	9.839	1.084	9.075	0.0000	40.30
	Gumbel-Hougaard	2.870	0.2597	10.35	0.0000	33.06
Wyelangta	Clayton	2.950	0.4136	7.129	0.0000	30.55
	Frank	10.170	1.146	8.882	0.0000	39.56
	Gumbel-Hougaard	2.620	0.263	9.985	0.0000	29.70

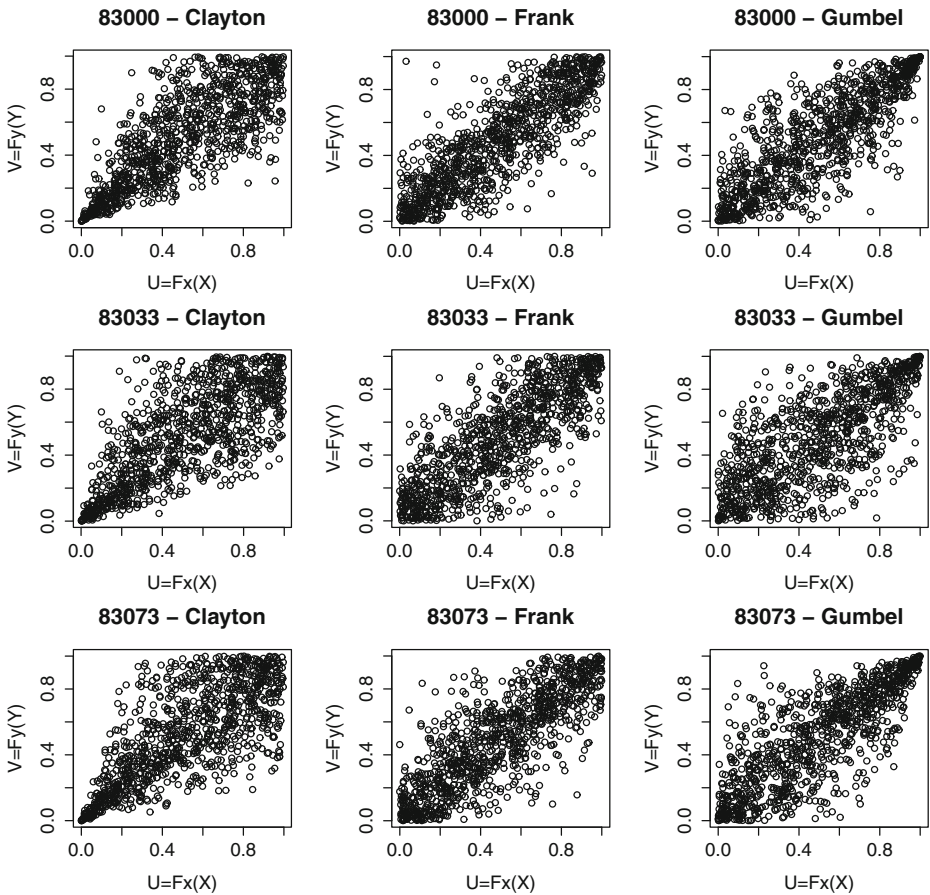


Fig. 4 Parametric copula based joint distribution for 3 stations (Archerton, Woods Point and Mount Buffalo Chalet)

where C_N is the empirical copulas (Genest et al. 2009), C_θ the estimated parametric copulas and the subscript i represents the sample number. The copula with the smallest value of S_n (hence giving the largest p-value) is chosen to represent the joint distribution of rainfall severity and duration. The values of S_n and p-values for all three copulas are presented in Table 6. For all the six stations, it is evident that the Clayton copula provides the best fit with largest p-value. The bivariate copula density function using the Clayton copula for Wyelanta station is displayed in Fig. 5, where the plot shows that Clayton copula has more probability concentrated in the left tail.

As was shown by Charpentier et al. (2006), a parametric copula would tend to underestimate the true joint distribution. In the third approach (NN), the choice of the Beta kernel is essential in order to improve the bias at the boundaries. In Fig. 6, the plot shows two peaks near the lower tails and a higher peak on the right tails of the distributions. There are some obvious disparities between this figure and Fig. 3 (PN model) and Fig. 5 (NP model) of the same station. Figure 5 shows low probability density near the tails of the surface which may not reflect the true nature of the joint distribution.

Table 6 Goodness-of-fit test for fitted Copulas using S_n

Stations	Type of Copulas	S_n	p-value
Archerton	Clayton	0.0127	0.6409
	Frank	0.3903	0.0005
	Gumbel-Hougaard	1.0152	0.0004
Woods Point	Clayton	0.0080	0.9715
	Frank	0.3338	0.0005
	Gumbel-Hougaard	0.9087	0.0005
Mount Buffalo Chalet	Clayton	0.0124	0.6369
	Frank	0.3807	0.0005
	Gumbel-Hougaard	0.9829	0.0005
Tanybryn	Clayton	0.0226	0.1304
	Frank	0.4320	0.0004
	Gumbel-Hougaard	1.127	0.0005
Weeaproinah	Clayton	0.0116	0.7038
	Frank	0.3392	0.0005
	Gumbel-Hougaard	0.9613	0.0004
Wyelangta	Clayton	0.0195	0.3162
	Frank	0.0959	0.0010
	Gumbel-Hougaard	0.1664	0.0005

To check the goodness of fit, we employ the Mean Absolute Error (MAE) between the fitted values based on each of the three approaches and the values obtained by calculating the corresponding empirical copula to the data. The MAE is defined in a similar way to S_n but using absolute instead of mean-squared deviation:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |C_{Pi} - C_{Ni}| \quad (23)$$

where C_{Pi} is the fitted copula values which were calculated based on the three approaches, and C_{Ni} are the empirical copula values (Deheuvels 1979). From Table 7, it is found that the PN and NN approaches generally give the smallest MAE values with NN dominating for most of the selected stations. This again indicates that the nonparametric approach has great merit over the parametric approach when it comes to fitting joint distribution based on large historical hydrological data. Whether it is true in general remains to be seen with new field work and additional data collection.

3.5 Return Periods

A return period is the interval of time between consecutive recurrence of an event such as severe draught, flood or extreme rainfall. Estimation of return periods of these events, characterized by various levels of severity, is an essential part of any hydrological and water planning projects. Most hydrological characteristics such as rainfall severity and its duration are highly dependent and separate analysis of each characteristic is therefore not sufficient in assessing overflow water risks or in performing flood analysis. A multivariate approach

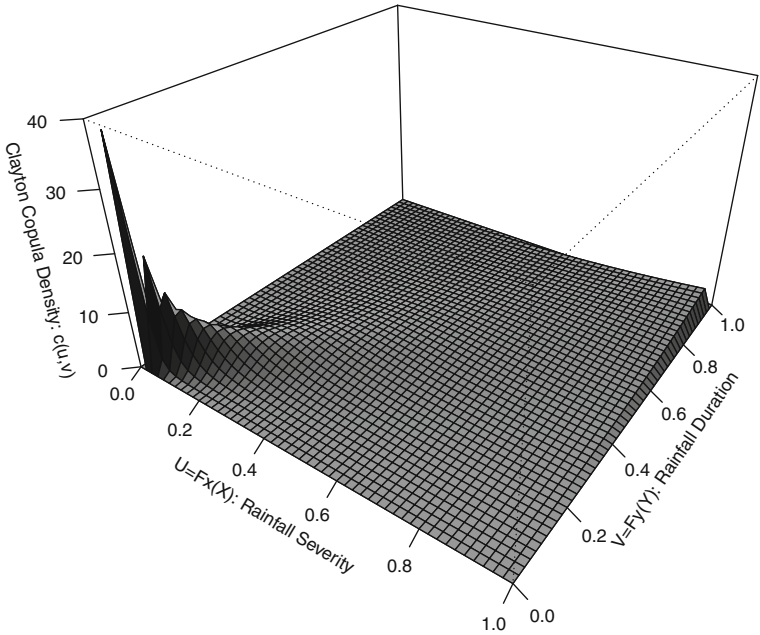


Fig. 5 Bivariate Clayton copula density function of rainfall severity and rainfall duration : NP Wyelangta

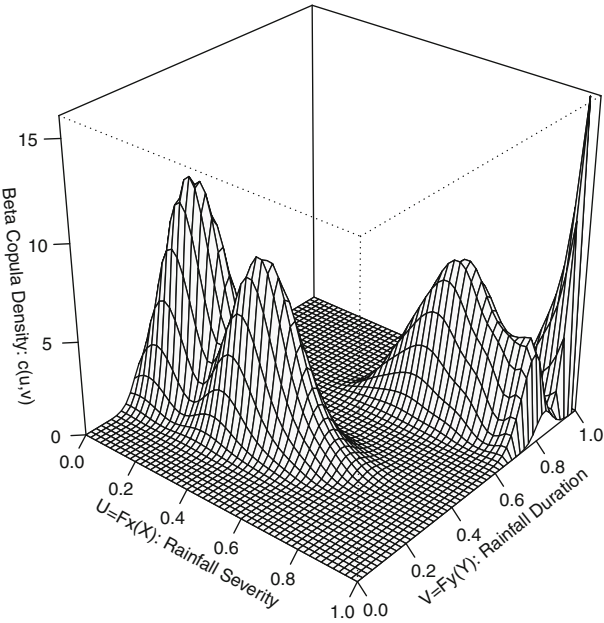


Fig. 6 Bivariate beta copula density function of rainfall severity and rainfall duration : NN Wyelangta

Table 7 Comparison of three approaches using MAE

Stations	Approach	MAE
Archerton	PN	0.0185
	NP	0.2270
	NN	0.0256
Woods Point	PN	0.2830
	NP	0.2450
	NN	0.1650
Mount Buffalo Chalet	PN	1.3560
	NP	2.0650
	NN	0.7890
Tanybryn	PN	0.3690
	NP	2.6780
	NN	1.5900
Weeaproinah	PN	2.0890
	NP	3.0250
	NN	1.2980
Wyelangta	PN	2.3870
	NP	2.9520
	NN	0.1460

is therefore an essential and a far superior method used in analyzing these events. For example, Kim (2003) obtained the joint distributions of drought duration and drought intensity using bivariate kernel estimator for estimating the joint return periods for the arid regions in Conchos River Basin, Mexico. However, the traditional approach of considering the joint distribution of rainfall characteristics using standard bivariate modeling presents some limitations as mentioned, and these can be circumvented by using Copulas. In this section, the expected return periods, using a single or joint rainfall characteristics, introduced by Shiau and Shen (2001) formulated for drought events with certain severity and duration will be applied to the rainfall data from Victoria. This will incorporate copulas and the proposed semiparametric approaches outlined in Table 2.

Let L be the extreme rainfall interarrival time, then the expected return period (Shiau and Shen 2001) for floods with severity S greater than or equal to a certain value s is given by:

$$T_s = \frac{E(L)}{1 - F_S(s)} \quad (24)$$

where $F_S(s)$ is the cumulative distribution function of S . Similarly, the expected return period for floods with rainfall duration D greater than or equal to d is given by

$$T_d = \frac{E(L)}{1 - F_D(d)} \quad (25)$$

where $F_D(d)$ is the cumulative distribution of the rainfall duration.

In this study, we used two approaches to estimate the cumulative distribution functions for rainfall severity and durations. In the first (parametric) approach, we used the Lognormal

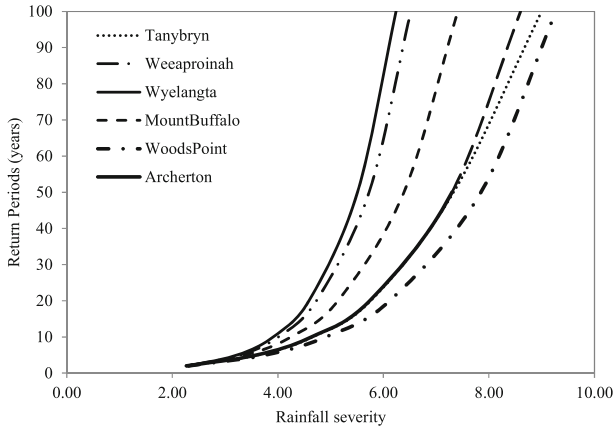


Fig. 7 Return Period for rainfall severity using Lognormal

distribution for both $F_S(s)$ and $F_D(d)$, since this distribution best fit both rainfall characteristics as can be seen from Table 4. In the second (nonparametric) approach, we used a kernel density estimator to estimate the distributions of the two rainfall characteristics.

The expected interarrival time for extreme rainfall events, $E(L)$, for the six (6) selected stations in Table 1 were estimated to be 10.0, 10.8, 9.0, 10.6, 8.9 and 9.2 months respectively. The expected return periods for severity and duration exceeding certain values were then calculated using Eqs. 24 and 25 respectively. The return periods up to 100 years against a set of abscissae of severity for both Lognormal and kernel density estimate are displayed in Figs. 7 and 8 respectively. For example, if severity exceeds 8.60 and duration exceeds 5.2, then a return period of 100 years is expected using the Lognormal distribution. Similarly, if severity exceeds 10.7 and duration exceeds 5.9, then a return period of 100 years is expected using the kernel density estimator. As expected, the graphs of the return periods

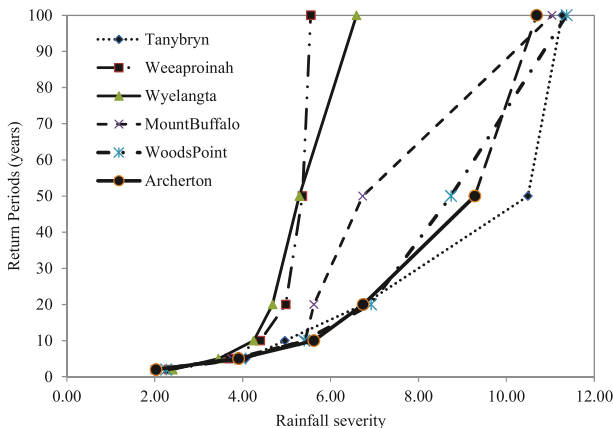


Fig. 8 Return Period for rainfall severity using kernel density estimate

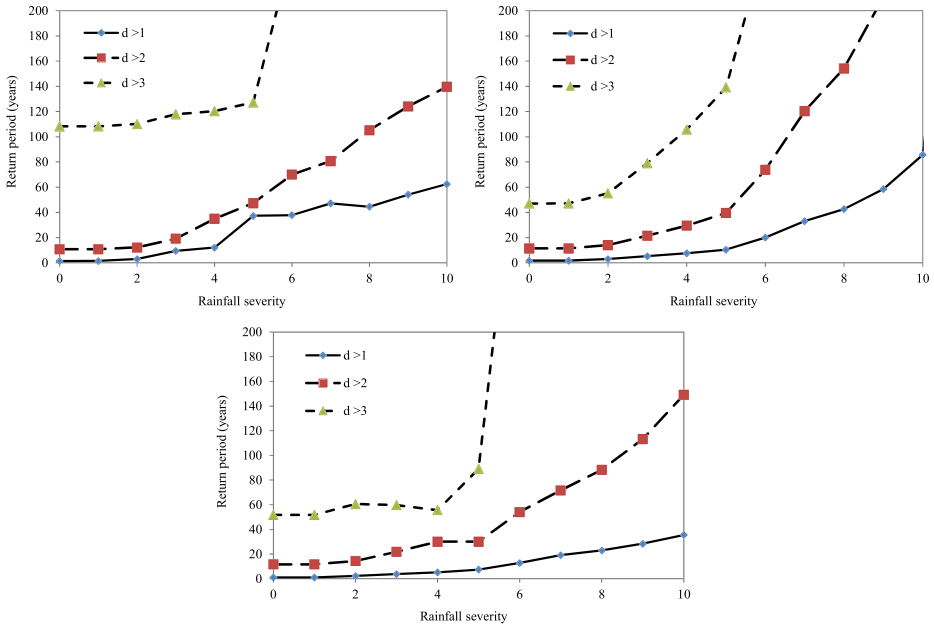


Fig. 9 Conditional return period of rainfall duration when rainfall severity exceed certain value - PN (top right), NP (top left) and NN (bottom)

rise sharply with increasing severity and duration but much more smoothly in the parametric than the nonparametric case. The results also suggest the areas with higher mean annual rainfall and lower interarrival times are more exposed to the extreme rainfall events that can cause floods.

Next, we consider the joint return and conditional return period which specifically involve the dependency between the two rainfall characteristics. The joint probability can be calculated in terms of copulas which provides the freedom for the marginal distributions to assume any appropriate parametric or nonparametric form. The following four expected return periods were introduced by Shiau (2003) and are readily interpretable using conditional probabilities:

$$\begin{aligned}
 T_{DS} &= \frac{E(L)}{P(D \geq d \text{ and } S \geq s)} \\
 &= \frac{E(L)}{1 - F_D(d) - F_S(s) + F_{DS}(d, s)} \\
 &= \frac{E(L)}{1 - F_D(d) - F_S(s) + C(F_D(d), F_S(s))}. \tag{26}
 \end{aligned}$$

$$\begin{aligned}
 T'_{DS} &= \frac{E(L)}{P(D \geq d \text{ or } S \geq s)} = \frac{E(L)}{1 - F_{DS}(d, s)} \\
 &= \frac{E(L)}{1 - C(F_D(d), F_S(s))} \tag{27}
 \end{aligned}$$

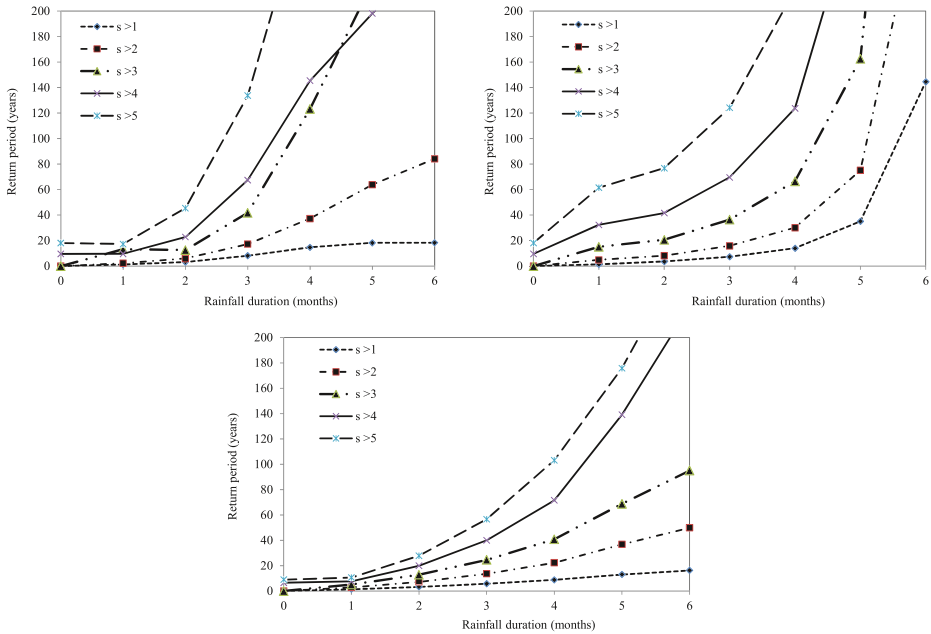


Fig. 10 Conditional return period of rainfall severity when rainfall duration exceed certain value - PN (top right), NP (top left) and NN (bottom)

T_{DS} (T'_{DS}) is the conditional return period given that the duration $D \geq d$ and (or) severity $S \geq s$. The following conditional return periods are based on further conditioning (26) on $S \geq s$ and $D \geq d$ respectively:

$$\begin{aligned}
 T_{D|S \geq s} &= \frac{T_{DS}}{1 - F_S(s)} \\
 &= \frac{T_s}{1 - F_D(d) - F_S(s) + C(F_D(d), F_S(s))}
 \end{aligned}
 \tag{28}$$

and

$$\begin{aligned}
 T_{S|D \geq d} &= \frac{T_{DS}}{1 - F_D(d)} \\
 &= \frac{T_d}{1 - F_D(d) - F_S(s) + C(F_D(d), F_S(s))}.
 \end{aligned}
 \tag{29}$$

Using all three approaches, Figs. 9 and 10 provide graphical representations of $T_{D|S \geq s}$ and $T_{S|D \geq d}$ for Archerton station given that rainfall severity and rainfall duration exceed certain values listed in the abscissae respectively. From the figures, we conclude that the results show similar trend for all cases. However, both PN and NP cases tend to show higher return periods for high severity at the same duration, while PN and NN cases tend to show a higher return periods for higher duration at the same severity.

4 Conclusions

Severe floods are worldwide phenomena which are becoming more frequent due to erratic weather conditions and climate changes. Since rainfall is the major cause of flood events, and its duration and severity influence their protraction, these two rainfall characteristics are considered important variables in hydrological studies.

In this paper, we used rainfall data from 6 selected stations from North-eastern and South-western Victoria, Australia, to analyze rainfall severity and duration prevalent in that state. A copula methodology is used to derive the joint distributions of these variables. A novelty of the paper is to employ both a nonparametric and a semiparametric approach, whereby in the first approach, we have allowed both the copula and the marginal distributions to assume nonparametric forms, while in the second approach, the copula and the marginal distributions take their turn assuming a parametric and nonparametric form respectively. This contrasts sharply with the standard approach adopted in many papers which assumes that both the marginal distributions and copulas assume parametric forms. Estimating parameters of these purely parametric models using standard techniques is often time consuming. Furthermore, legitimate concern can be raised with respect to their accuracies in case the assumptions underlying these models are violated or small data sets are involved. On the other hand, the nonparametric approach ameliorate these problems and can give better results without assuming a particular form for the marginal or copula distributions.

We began this paper by first quantifying severity through the Standard Precipitation Index (SPI). For SPI estimation, we presented an alternative approach using nonparametric kernel density by employing the Gaussian kernel to estimate the probability density function of rainfall intensity. We then apply several copulas-based approaches, each involving a combination of parametric or nonparametric marginal distributions conjoined by a parametric or nonparametric copula, to model the two rainfall characteristics. Using goodness of fit tests, we found that the Lognormal distribution provides the best fit among four parametric marginal distributions and the Clayton copula the best fit copula among the three Archimedean copulas chosen. Further, Table 7 indicates that the purely nonparametric approach (NN) generally provides a better fit to the data than the two mixed approaches. Finally, we used the three approaches to derive several return periods of severe rainfall events for the stations selected. Estimation of return period is of course crucial in water management planning and the results obtained are not unexpected. Both parametric and nonparametric approaches gave similar trend, with the parametric approach providing slightly higher return periods than the nonparametric approach.

There is much scope for applying the nonparametric approaches adopted in this paper to flood or even drought events in other regions of the world. It would also be interesting to see whether the results obtained in this paper can be replicated elsewhere. Finally, the methods can be extended without much difficulty to more than two rainfall characteristics thus making them more applicable under a wider range of flood conditions.

Acknowledgments The authors sincerely acknowledge the Bureau of Meteorology (BOM), Australia, for providing the complete monthly precipitation data that been used in this study. The work is financed by SLAB Scholarship provided by the Ministry of Higher Education of Malaysia and National Defence University of Malaysia.

References

- Abdul Rauf U, Zeepongsekul P (2011) Modelling rainfall severity and duration in north-eastern victoria using copulas. Proceedings of the 19th International Congress on Modelling and Simulation, Perth
- Abdul Rauf UF, Zeepongsekul P (2014) Copula based analysis of rainfall severity and duration: a case study. *Theor Appl Climatol* 115(1-2):153–166
- Brown BM, Chen SX (1999) Beta-bernstein smoothing for regression curves with compact support. *Scand J Stat* 26(1):47–59
- Cancelliere A, Mauro G, Bonaccorso B, Rossi G (2007) Drought forecasting using the standardized precipitation index. *Water Resour Manag* 21(5):801–819. doi:[10.1007/s11269-006-9062-y](https://doi.org/10.1007/s11269-006-9062-y)
- Chappell A, Renzullo LJ, Raupach TH, Haylock M (2013) Evaluating geostatistical methods of blending satellite and gauge data to estimate near real-time daily rainfall for australia. *J Hydrol* 493(0):105–114
- Charpentier A, Fermanian J, Scaillet O (2006) Copulas: from theory to application in finance, 1st edn, Risk Books, Torquay, UK, chap The Estimation of Copulas: Theory and Practice
- Chen SX (1999) Beta kernel estimators for density functions. *Comput Stat Data Anal* 31(2):131–145
- Chen SX (2000) Beta kernel smoothers for regression curves. *Stat Sin* 10:73–91
- Chen SX, Huang TM (2007) Nonparametric estimation of copula functions for dependence modelling. *Can J Stat* 35(2):265–282
- De Michele C, Salvadori G (2003) A generalized pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J Geophys Res Atmos* 108(D2). doi:[10.1029/2002JD002534](https://doi.org/10.1029/2002JD002534)
- Deheuvels P (1979) La fonction de dependance empirique et ses proprietes: un test non parametrique d'indpendance. *Acad Roy Bull Cl Sci* 65:274–292
- Genest C, Favre A (2007) Everything you always wanted to know about copula modeling but were afraid to ask. *J Hydrol Eng* 12(4):347–368
- Genest C, Ghoudi K, Rivest LP (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82(3):543–552
- Genest C, Rumillard B, Beaudoin D (2009) Goodness-of-fit tests for copulas: A review and a power study. *Insur Math Econ* 44(2):199–213
- Guttman NB (1998) Comparing the palmer drought index and the standardized precipitation index 1. *JAWRA J Am Water Works Assoc* 34(1):113–121
- Haghighat jou P, Akhoond-Ali A, Nazemosadat M (2013) Nonparametric kernel estimation of annual precipitation over iran. *Theor Appl Climatol* 112(1-2):193–200. doi:[10.1007/s00704-012-0727-6](https://doi.org/10.1007/s00704-012-0727-6)
- Harrell FE, Davis CE (1982) A new distribution-free quantile estimator. *Biometrika* 69(3):635–640
- Joe H (1997) *Multivariate Models and Dependence Concepts*. Chapman and Hall London
- Kao SC, Govindaraju RS (2007) A bivariate frequency analysis of extreme rainfall with implications for design. *J Geophys Res Atmos* 112(D13). doi:[10.1029/2007JD008522](https://doi.org/10.1029/2007JD008522)
- Kao SC, Govindaraju RS (2008) Trivariate statistical analysis of extreme rainfall events via the plackett family of copulas. *Water Res Res*:44
- Kao SC, Govindaraju RS (2010) A copula-based joint deficit index for droughts. *J Hydrol* 380(12):121–134
- Kim TW (2003) Nonparametric approach for estimating return periods of droughts in arid regions. *J Hydrol Eng* 8(5):237–246
- Kim TW, JB Valds, Nijssen B, Roncayolo D (2006) Quantification of linkages between large-scale climatic patterns and precipitation in the colorado river basin. *J Hydrol* 321(14):173–186
- McBride JL, Nicholls N (1983) Seasonal relationships between australian rainfall and the southern oscillation. *Mon Weather Rev* 3:1998–2004
- McKee T, Doesken N, Kleist J (1993) The relationship of drought frequency and duration to time scales. 8th Conference on Applied Climatology, Anaheim, California
- Mehrotra R, Sharma A (2011) Impact of atmospheric moisture in a rainfall downscaling framework for catchment-scale climate change impact assessment. *Int J Climatol* 31(3):431–450
- Mirabbasi R, Fakheri-Fard A, Dinpashoh Y (2012) Bivariate drought frequency analysis using the copula method. *Theor Appl Climatol* 108:191–206
- Nelsen RB (2006) *Introduction to copulas*, lecture notes statistics, vol 139, 2nd edn. Springer-Verlag
- Nicholls N, Wong KK (1990) Dependence of rainfall variability on mean rainfall, latitude, and the southern oscillation. *J Clim* 3:163170
- Pui A, Lal A, Sharma A (2011) How does the interdecadal pacific oscillation affect design floods in australia. *Water Resour Res* 47(5). doi:[10.1029/2010WR009420](https://doi.org/10.1029/2010WR009420)
- Reddy M, Ganguli P (2012) Bivariate flood frequency analysis of upper godavari river flows using archimedean copulas. *Water Resour Manag* 26(14):3995–4018. doi:[10.1007/s11269-012-0124-z](https://doi.org/10.1007/s11269-012-0124-z)

- Scaillet O, Fermanian JD (2002) Nonparametric estimation of copulas for time series. *FAME Research Paper* (57)
- Serinaldi F, Bonaccorso B, Cancelliere A, Grimaldi S (2009) Probabilistic characterization of drought properties through copulas. *Physics and Chemistry of the Earth. Parts A/B/C* 34(10-12):596–605
- Sharma A (2000) Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3 a nonparametric probabilistic forecast model. *J Hydrol* 239(14):249–258
- Sharma A, Lall U (1999) A nonparametric approach for daily rainfall simulation. *Math Comput Simul* 48(46):361–371
- Shiau J (2006) Fitting drought duration and severity with two-dimensional copulas. *Water Resour Manag* 20:795–815
- Shiau J, Shen H (2001) Recurrence analysis of hydrologic droughts of differing severity. *J Water Resour Plan Manag* 127(1):30–40
- Shiau JT (2003) Return period of bivariate distributed extreme hydrological events. *Stoch Env Res Risk A* 17:42–57
- Shiau JT, Modarres R (2009) Copula-based drought severity-duration-frequency analysis in iran. *Meteorol Appl* 16(4):481–489
- Silverman B (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall
- Wand M, Jones M (1995) *Kernel Smoothing*. Chapman and Hall
- Yue S (2000) Joint probability distribution of annual maximum storm peaks and amounts as represented by daily rainfalls. *Hydrol Sci J* 45(2):315–326
- Yusof F, Hui-Mean F, Suhaila J, Yusof Z (2013) Characterisation of drought properties with bivariate copula analysis. *Water Resour Manag* 27(12):4183–4207. doi:[10.1007/s11269-013-0402-4](https://doi.org/10.1007/s11269-013-0402-4)
- Zhang L, Singh VP (2007) Bivariate rainfall frequency distributions using archimedean copulas. *J Hydrol* 332(12):93–109
- Zhao F, Zhang L, Chiew FH, Vaze J, Cheng L (2013) The effect of spatial rainfall variability on water balance modelling for south-eastern australian catchments. *J Hydrol* 493(0):16–29. doi:[10.1016/j.jhydrol.2013.04.028](https://doi.org/10.1016/j.jhydrol.2013.04.028)