

Trivariate copulas for characterisation of droughts

G. Wong¹

M. F. Lambert²

A. V. Metcalfe³

(Received 3 August 2007; revised 4 January 2008)

Abstract

Australian agriculture is at serious risk from drought, and water resource infrastructure and management can mitigate the effects. The consequences of droughts depend on their intensity, duration and severity. These variables are correlated and the dependence structure is here described by copulas. Copulas are multivariate uniform distributions which allow for the dependence structure to be modelled independently of the marginal distributions. Trivariate Gaussian and Gumbel copulas are fitted to the data from a rainfall district in NSW. We assess the goodness of fit of the data to the different forms using several criteria. The data are best described by a Gumbel copula and three parameter Weibull marginal distributions.

Contents

1	Introduction	C307
2	Definition of drought	C308
3	Copulas	C309
4	Application	C311
4.1	Study region	C311
4.2	Fitting marginal distributions	C312
4.3	Fitting trivariate copulas	C314
4.4	Goodness-of-fit tests	C319
5	Conclusions	C320
	References	C321

1 Introduction

The Australian Bureau of Agriculture and Resource Economics estimates that the 2006 drought slashed the national winter cereal crop by 36% and cost rural Australia around \$3.5 billion [11]. Water resource infrastructure and efficient management are vital to reduce the effects of such droughts.

In hydrological studies, drought is defined as a period with substantially less rainfall than usual. The main characteristics of a drought are its duration, intensity and severity and we model the association between them with a trivariate copula. Although the concept of copulas goes back at least as far as 1959 [10], most applications in hydrology have been recent [2, 5, 12], and the only application to drought, of which we are aware, is bivariate [9]. Copulas provide a succinct quantitative characterisation of the distribution

of drought variables, that can be used for climatic and economic analyses. We define droughts in Section 2, discuss copulas in Section 3, and fit copulas to trivariate drought data in Section 4.

2 Definition of drought

The Standardized Precipitation Index (SPI) is a common indicator of drought, which has an advantage of only requiring rainfall data [7]. The calculation of the SPI from monthly rainfall totals for some chosen period (m) follows. Suppose r_t is the rainfall for month t . Calculate a backward moving average of length m :

$$y_t = \frac{r_t + r_{t-1} + \cdots + r_{t-m+1}}{m}, \quad m \leq t,$$

Then suitable probability distributions are fitted to the $\{y_t\}$ for each calendar month (j). A typical choice is the two parameter Gamma distribution with cumulative distribution function (CDF) $G_j(\cdot)$. The SPI (S_t) is defined by applying the inverse CDF of the standard Normal distribution to $G_j(y_t)$, that is, $S_t = \Phi^{-1}(G_j(y_t))$. It follows from this construction that S_t is deseasonalized and that 95% of S_t are expected to be in a range from -2 to $+2$. Droughts are usually defined as periods during which S_t is consistently below a threshold of -1 [7]. Since the S_t is defined monthly, the minimum duration of a drought is one month and its duration is the number of months until S_t increases above the threshold of -1 . The corresponding intensity is defined as the absolute value of the minimum value taken by S_t over the duration of the drought. The severity is defined as the area under the threshold over the duration of the drought.

Figure 1 shows three droughts relative to a threshold -1 of which, drought 1 has the highest intensity (I), drought 2 has the longest duration (D) and drought 3 is the most severe (severity is the area below -1).

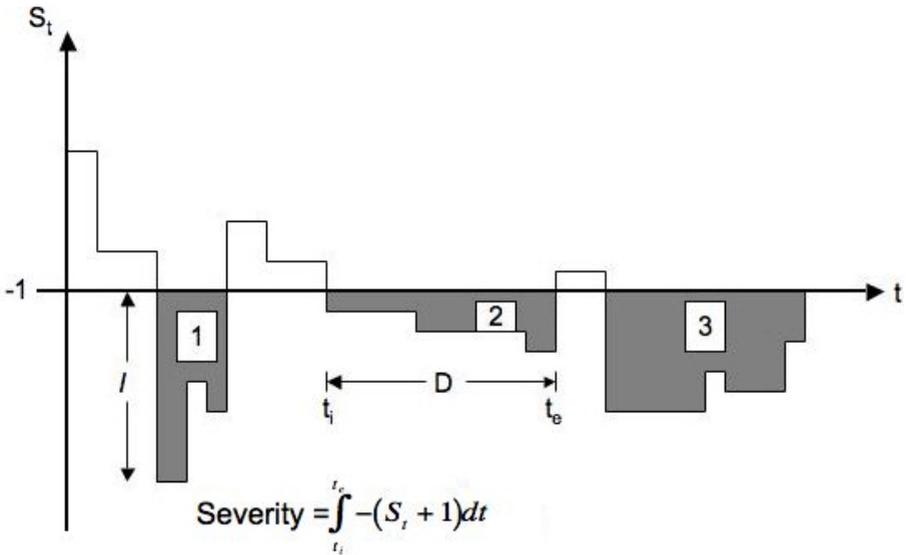


FIGURE 1: Definition of drought characteristics.

3 Copulas

A n -copula is a multivariate uniform distribution, and its cumulative distribution function C is a mapping $C : [0, 1]^n \rightarrow [0, 1]$. Sklar's Theorem [10] states that if there exists an n -dimensional CDF with univariate margins F_1, \dots, F_n , then there exists an n -copula C such that

$$H(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) = C(u_1, \dots, u_n) \quad (1)$$

where $F_k(x_k) = u_k$ for $k = 1, \dots, n$ with $U_k \sim U(0, 1)$ and F_k is continuous. Conversely, any choice of copula and F_k is a n -variable CDF. From Equation (1), we model the univariate marginal distributions F_k and the joint distribution separately.

The Archimedean family of copulas is commonly used in hydrological analyses due to its versatility. The symmetric Archimedean n -copula is of

the form

$$C(u_1, \dots, u_n) = \varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_n)), \quad (2)$$

where φ is the generator of the copula. A copula of this family is uniquely defined by its generator function. For example, the bivariate Gumbel copula $C(u, v) = \exp[-\{(-\ln u)^\theta + (-\ln v)^\theta\}^{1/\theta}]$ is defined by its generator function $\varphi(t) = (-\ln t)^\theta$. However, for $n > 2$, the symmetric copula has the consequence that correlations between any pairs of variables are identical. This is unrealistic for most applications. A form of asymmetric copula is formed by nesting symmetric copulas [1, 6] and in the three variable case

$$\begin{aligned} C(u_1, u_2, u_3) &= C_1(C_2(u_1, u_2), u_3) \\ &= \varphi_1^{-1}(\varphi_1\{\varphi_2^{-1}[\varphi_2(u_1) + \varphi_2(u_2)] + \varphi_1(u_3)\}). \end{aligned} \quad (3)$$

From Equation (3), the relationship between variables u_1 and u_2 is described by copula C_2 . The result from this relationship is then associated with u_3 through the outer copula C_1 . Note that both C_1 and C_2 are bivariate copulas with individual generators φ_1 and φ_2 respectively, but unlike u_3 , C_2 does not have a uniform distribution.

A different, widely used copula is the Gaussian copula which belongs to a family of elliptical copulas. A multivariate Gaussian copula with n variables is

$$C(u_1, \dots, u_n) = \Phi^n(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)), \quad (4)$$

where Φ^n indicates the standard multivariate normal cumulative distribution function (CDF) with correlation matrix R , and Φ^{-1} is the inverse CDF of the standard univariate normal distribution.

Tail dependence is a measure of dependence between extreme values at both upper and lower tails of a bivariate distribution and is of practical importance in hydrology. A bivariate copula C has upper tail dependence $\lambda_U \in (0, 1]$ if

$$\lim_{u \rightarrow 1} \bar{C}(u, u)/(1 - u) = \lambda_U, \quad (5)$$

where \bar{C} is the complement of the CDF C :

$$\bar{C}(u, u) = 1 - u - u + C(u, u).$$

The bivariate copula has no upper tail dependence if $\lambda_U = 0$. Similarly, C has lower tail dependence $\lambda_L \in (0, 1]$ if

$$\lim_{u \rightarrow 0} C(u, u)/u = \lambda_L, \quad (6)$$

and no lower tail dependence if $\lambda_L = 0$. For example, the parameter ρ in the bivariate Gaussian copula is the off-diagonal entry in the correlation matrix R , and the Gaussian copula has no tail dependence when $|\rho| < 1$ [1], as it will be in practical applications. In contrast, the bivariate Gumbel copula has upper tail dependence $2 - 2^{1/\theta}$ [1].

4 Application

4.1 Study region

Australia is divided into 99 relatively homogenous rainfall districts. We use monthly precipitation data from District 63 (Central Tablelands) in NSW, from 1913–2002. The monthly SPI(3) series was calculated, using the calculations in Section 2. There were 78 droughts during the 90 year period. Figure 2(a) shows the location of District 63 and Figure 2(b) shows the distribution of the time between the end of one drought and the start of another in terms of months. Correlations between the three drought characteristics and non-drought time, before or after the drought, are small, and none is statistically significant at the 5% level. Consequently, time between droughts is realistically modelled independently of the other three variables.

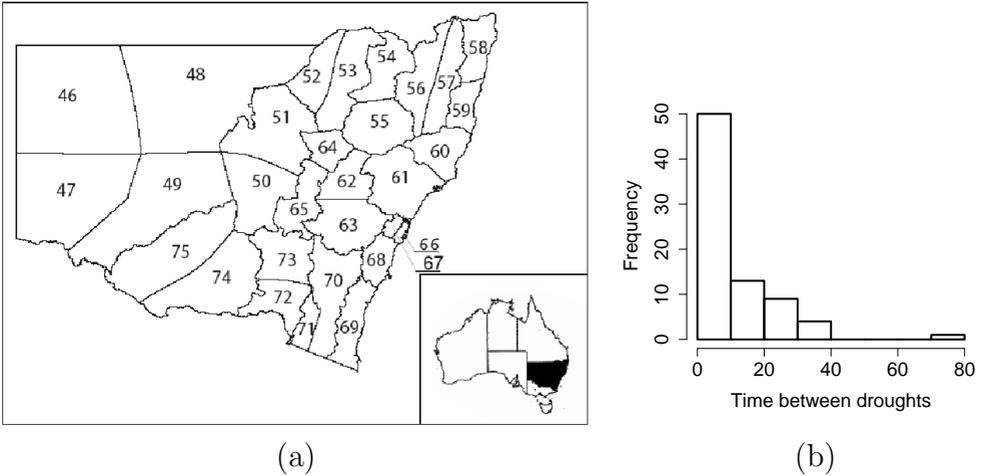


FIGURE 2: (a) New South Wales (NSW) rainfall districts; (b) Histogram of months between the end of one drought and start of the next.

4.2 Fitting marginal distributions

Drought severity, duration and intensity are fitted with the three parameter Weibull distribution with CDF

$$F(x) = 1 - \exp[-(x - L)^k / \theta], \quad L \leq x,$$

where L , θ and k are the threshold, scale and shape parameters respectively. Table 1 shows the maximum likelihood estimates of the parameters. We define u , v and w to be the respective severity, duration and intensity after transformation to uniform distributions. Table 2 displays the correlation between the drought variables. Figure 3(a)–(c) shows the probability plots of severity, intensity and duration fitted by the three parameter Weibull marginals. In Figure 3(c), points are superimposed as the data are restricted to integers. Overall, the Weibull distribution provides a reasonably good fit to the drought variables.

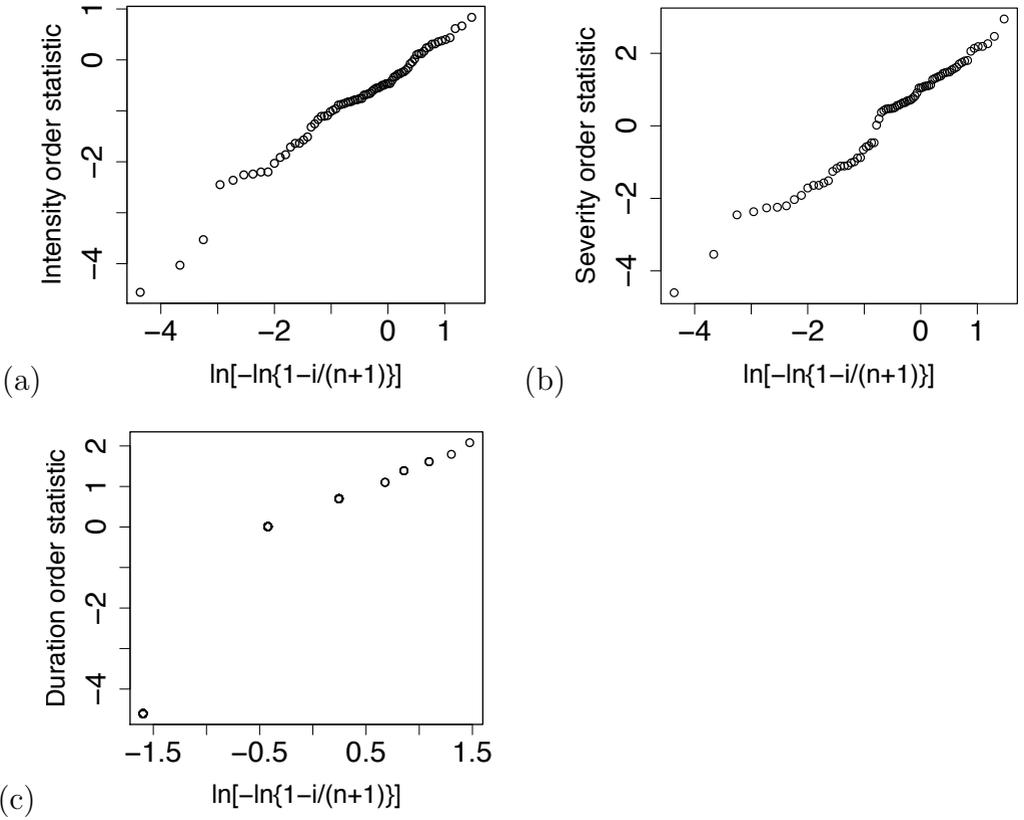


FIGURE 3: Weibull probability plots of (a) Intensity, (b) Severity, and (c) Duration.

TABLE 1: Estimated parameters of Weibull distributions.

Characteristic	Threshold	Scale	Shape
Severity	0.9904	2.463	0.8244
Duration	0.99	0.9375	0.5291
Intensity	0.99	0.688	1.311

TABLE 2: Correlations of drought variables after transformation to uniform marginals.

Correlations	Pearson	Kendall's tau
u, v	0.958	0.860
u, w	0.859	0.752
v, w	0.710	0.615

4.3 Fitting trivariate copulas

The Gaussian copula is fitted by taking the standard Gaussian inverse CDF of each element of the (u, v, w) triples and then calculating the three sample Pearson correlations: $r_{uv} = 0.934$, $r_{uw} = 0.867$ and $r_{vw} = 0.695$. Simulations of 100,000 deviates are made using the multinomial Gaussian distribution function in **R** [8]. Figure 4 is a contour plot, fitted to the simulated data, representing the bivariate copula probability density function, with the historic data shown as points. Figure 5 is a plot of the simulated data from the copula back-transformed to severity and duration. The 90% percentiles emphasize the association of the variables in the upper tails. To measure upper tail dependence, we calculate the Pearson correlation of points falling above the 90th percentiles (Table 3).

The correlations of the simulated data are close to those of the his-

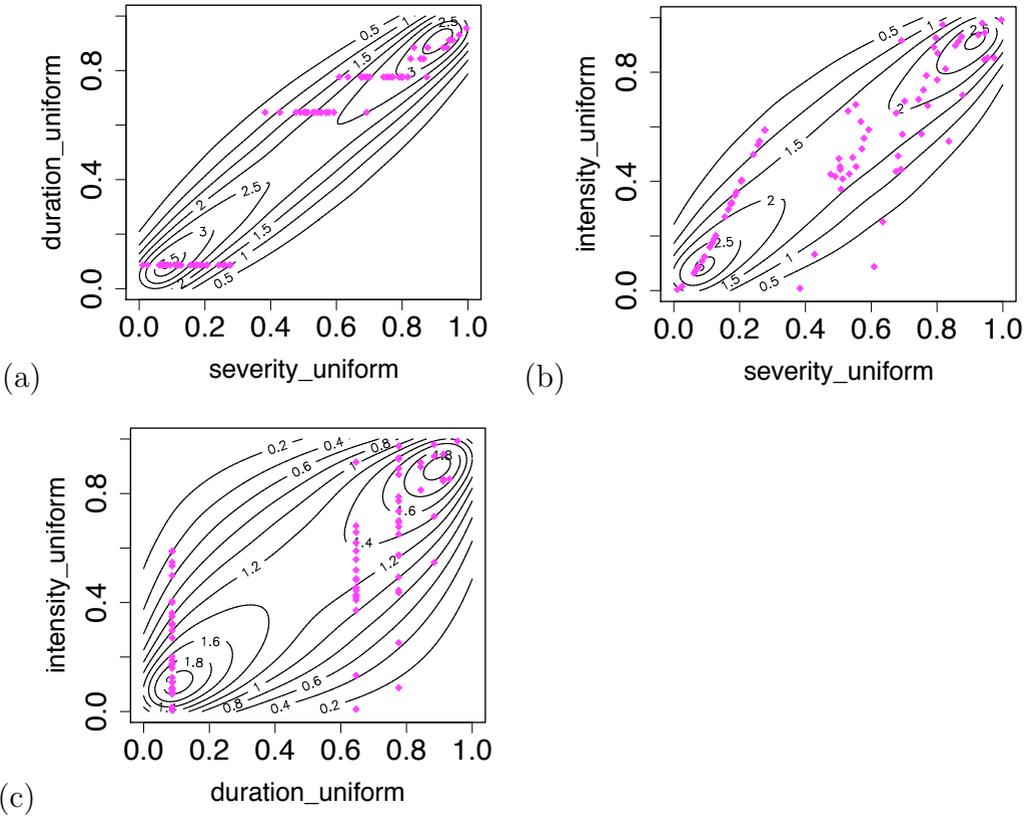


FIGURE 4: Empirical contour plot of (a) u against v , (b) u against w and (c) v against w from the Gaussian copula

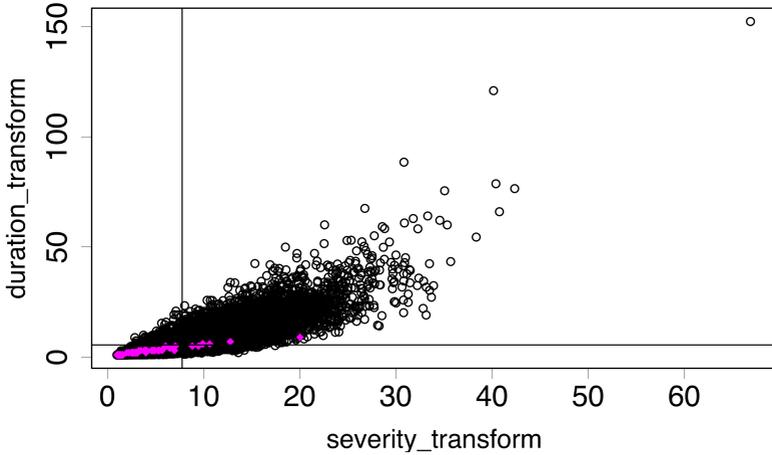


FIGURE 5: Plot of back-transformed simulations of severity against duration from Gaussian copula and 90% quantiles of margins, and historic data.

toric data. However, the upper tail correlations of the simulated data are fairly weak, especially between the pairs of severity-intensity and duration-intensity.

The trivariate Gumbel copula is

$$\begin{aligned}
 & C_1(C_2(u, v), w) \\
 &= \varphi_1^{-1}(\varphi_1\{\varphi_2^{-1}[\varphi_2(u) + \varphi_2(v)] + \varphi_1(w)\}) \\
 &= \exp[-\{[(-\ln u)^{\theta_2} + (-\ln v)^{\theta_2}]^{\theta_1/\theta_2} + (-\ln w)^{\theta_1}\}^{1/\theta_1}], \quad (7)
 \end{aligned}$$

where $\varphi_i(t) = (-\ln t)^{\theta_i}$, $i = 1, 2$, and $\theta_i \geq 1$.

θ_1 is the measure of dependence for the pairs (u, w) and (v, w) , and θ_2 represents dependence of (u, v) . For a proper three dimensional copula to exist, $\theta_1 \leq \theta_2$.

Parameters $\hat{\theta}_1$ and $\hat{\theta}_2$ were determined using Maximum Likelihood Estimates (MLE) method to give $\hat{\theta}_1 = 2.18$ and $\hat{\theta}_2 = 3.97$. We use a Gibbs sampler to simulate from this copula [1]. Let U_1, \dots, U_n have joint distri-

TABLE 3: Descriptive statistics of simulations from Gumbel and Gaussian copulas, u , v and w are severity, duration and intensity after transformation to uniform distributions.

Variable pairs	Correlation		Upper tail correl.		Proportion	
	Gumbel	Gauss	Gumbel	Gauss	Gumbel	Gauss
u, v	0.913	0.928	0.788	0.586	0.0813	0.0746
u, w	0.729	0.856	0.561	0.416	0.0660	0.0633
v, w	0.728	0.678	0.557	0.232	0.0664	0.0455

bution function C . The conditional distribution of U_k given the values of U_1, \dots, U_{k-1} is

$$\begin{aligned}
 & C_k(u_k \mid u_1, \dots, u_{k-1}) \\
 &= \Pr\{U_k \leq u_k \mid U_1 = u_1, \dots, U_{k-1} = u_{k-1}\} \\
 &= \frac{\partial^{k-1} C_k(u_1, \dots, u_k)}{\partial u_1 \cdots \partial u_{k-1}} \bigg/ \frac{\partial^{k-1} C_{k-1}(u_1, \dots, u_{k-1})}{\partial u_1 \cdots \partial u_{k-1}}. \tag{8}
 \end{aligned}$$

In our case study, u is first simulated from $U(0, 1)$ before we simulate v from $C_2(v \mid u)$ and w can be obtained from the conditional distribution of $C_3(w \mid u, v)$. Figure 6 shows the corresponding contour plots of the simulations, of length 100,000, and the historic data.

Figure 7 shows the plot of the transformed simulations of severity against duration with respect to the historic data. Table 3 gives correlations of the simulations from the trivariate Gumbel copula. From Table 3, the overall correlations are somewhat lower than for the Gaussian copula. However, the upper tail dependence is captured by the Gumbel copula. This is important in the case of extreme events where positive correlations between variables with high values have serious consequences. The proportion of points in that upper tail is also higher than for the Gaussian copula.

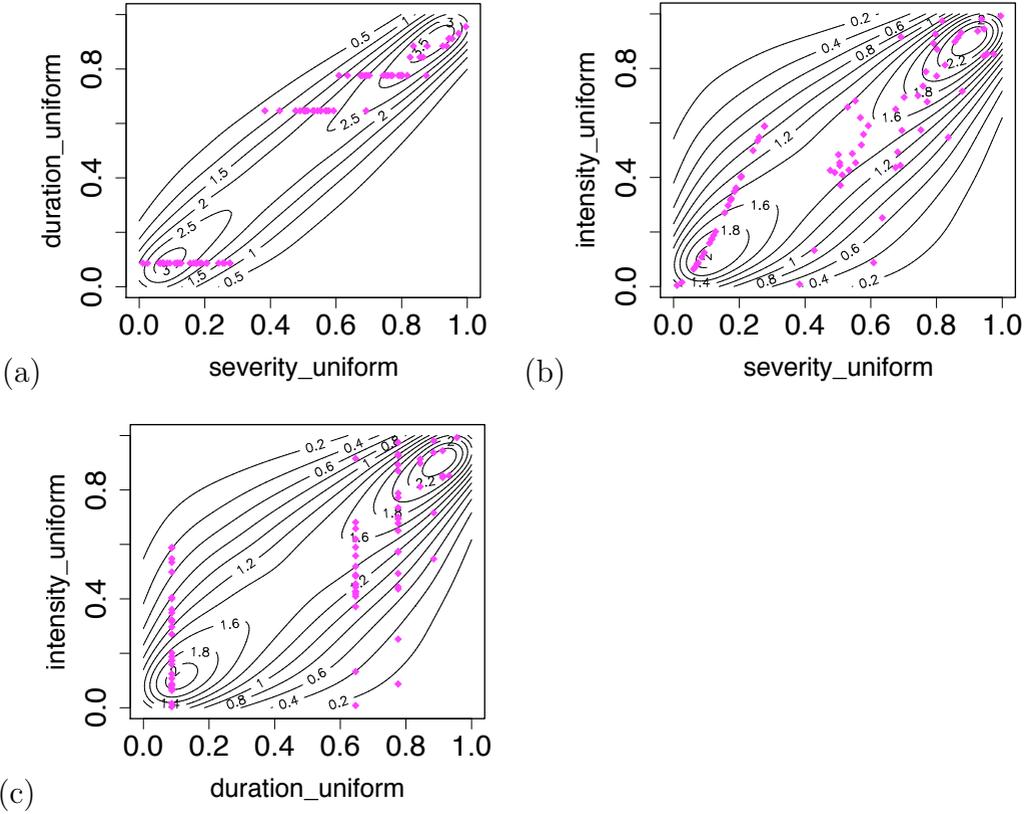


FIGURE 6: Empirical contour plot of (a) u against v , (b) u against w and (c) v against w from a Gumbel copula.

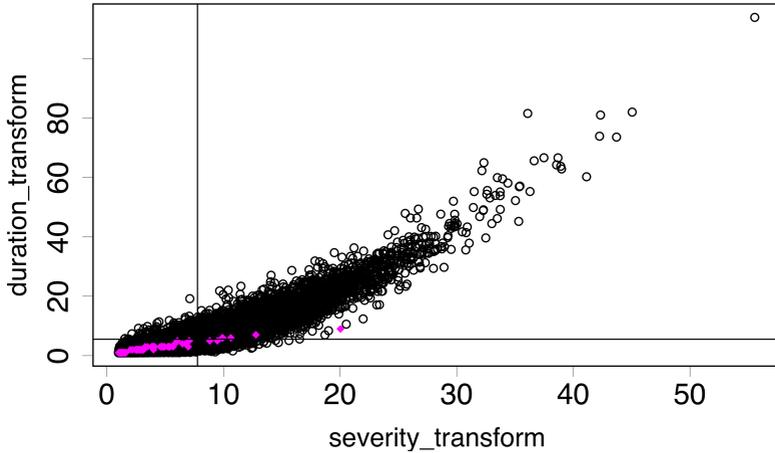


FIGURE 7: Plot of back-transformed simulations of severity against duration from Gumbel copula and 90% quantiles of margins, and historic data.

4.4 Goodness-of-fit tests

We plot probabilities calculated from the fitted copula, against comparable probabilities from the empirical copula C_n [4] in Figure 8:

$$C_n(u, v, w) = \frac{1}{n} \sum_{i=1}^n \mathbf{I} \left(\frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v, \frac{T_i}{n+1} \leq w \right) \quad (9)$$

where $\mathbf{I}(A)$ denotes the indicator function of the logical expression A . The ranks of the i th severity, duration and intensity datum is denoted by R_i , S_i and T_i respectively. The measure of fit is based on how close the points are to the diagonal line. The Gumbel copula provides a better fit than the Gaussian copula, because the points are closer to the diagonal. The lower left step linear clusters are a consequence of many droughts of one month duration.

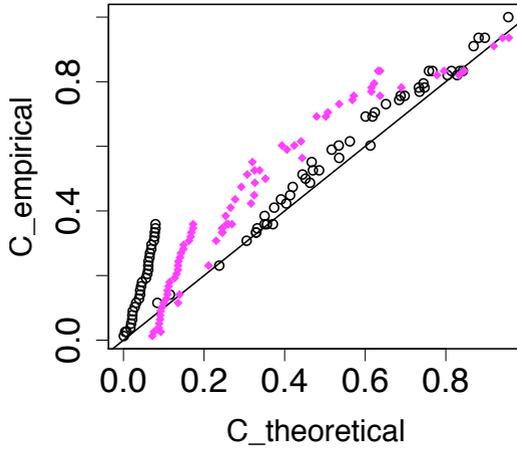


FIGURE 8: Plot of values from theoretical Gumbel copula (circles) and theoretical Gaussian copula (diamonds) against empirical copula.

5 Conclusions

The times between the end of one drought and the beginning of the next appear to be independent of the severity, duration and intensity of these droughts. Hence, drought events can be modelled with a trivariate copula and an independent distribution of times. The difference between copulas is striking in the Q-Q plot in Figure 8, but less so in the bivariate contour plots. The higher tail correlation of the Gumbel copula (Table 3) is apparent when comparing Figure 6 and Figure 4. A good fit was provided with three parameter Weibull distributions for margins and a two parameter Gumbel copula. In general, different forms of distributions can be used for each margin.

Copulas provide a description of drought characteristics in terms of a small number of parameters. Changes in these parameters over time and their relationship with climatic indices, such as the Southern Oscillation Index, can be investigated. Such analyses may provide evidence of climate change and

enable drought predictions to be made. Other advantages of copulas are that economic loss functions can be expressed in terms of the parameters and that they can be used for simulating future drought scenarios.

References

- [1] P. Embrechts, F. Lindskog and A. J. McNeil. Modelling dependence with copulas and applications to risk management. In: Rachev S. T., editor. *Handbook of heavy tailed distributions in finance*. Amsterdam, North-Holland: Elsevier, 2003. [C310](#), [C311](#), [C316](#)
- [2] A-C Favre, S. E. Adlouni, L. Perreault, N. Thiémonge and B. Bobee. Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, 2004, **40**, W01101. [doi:10.1029/2003WR002456](https://doi.org/10.1029/2003WR002456) [C307](#)
- [3] C. Genest and J. MacKay. The joy of copulas: bivariate distributions with uniform marginals. *The American Statistician*, 1986, **40**(4), 280–283.
- [4] C. Genest and A-C Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 2006, **12**(4), 347–368. [doi:10.1061/\(ASCE\)1084-0699\(2007\)12:4\(347\)](https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347)) [C319](#)
- [5] S. Grimaldi and F. Serinaldi. Asymmetric copula in multivariate flood frequency analysis. *Advances in Water Resources*, 2006, **29**, 1155–1167. [doi:10.1016/j.advwatres.2005.09.005](https://doi.org/10.1016/j.advwatres.2005.09.005) [C307](#)
- [6] H. Joe. *Multivariate models and dependence concepts*. New York: Chapman & Hall, 1997. [C310](#)

- [7] T. B. McKee, N. J. Doesken and J. Kliest. The relationship of drought frequency and duration to time scales. *In Proceeding of the 8th Conference on Applied Climatology*, 17–22 January 1993, Anaheim, CA, American Meteorological Society: Boston, MA, USA, 179–184. C308
- [8] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. 2007. <http://www.R-project.org> C314
- [9] J. T. Shiau. Fitting Drought Duration and Severity with Two-Dimensional Copulas. *Water Resources Management*, 2006, **20**, 795–815. doi:10.1007/s11269-005-9008-9 C307
- [10] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 1959, **8**, 229–231. C307, C309
- [11] The Australian. *El Nino bad news for drought-stricken east*. The Australian, 14th September. Accessed 19th September. <http://www.theaustralian.news.com.au/printpage/0,5942,2041027> C307
- [12] L. Zhang and V. P. Singh. Bivariate Flood Frequency Analysis using Copula Method. *Journal of Hydrologic Engineering*, 2006, **11**(2), 150–164. doi:10.1061/(ASCE)1084-0699(2006)11:2(150) C307

Author addresses

1. **G. Wong**, The University of Adelaide, North Terrace, South Australia 5005, AUSTRALIA.
<mailto:geraldine.wong@adelaide.edu.au>
2. **M. F. Lambert**, The University of Adelaide, North Terrace, South Australia 5005, AUSTRALIA.
3. **A. V. Metcalfe**, The University of Adelaide, North Terrace, South Australia 5005, AUSTRALIA.