Modelling and simulation of seasonal rainfall using checkerboard copulas of maximum entropy

Jonathan Borwein[a], Phil Howlett[b], Julia Piantadosi[b,*]

[a]*Centre for Computer Assisted Research Mathematics and its Applications (CARMA), University of Newcastle, Callaghan, NSW 2308, Australia.*
[b]*Scheduling and Control Group (SCG), Centre for Industrial and Applied Mathematics (CIAM), Barbara Hardy Institute (BHI), University of South Australia, Mawson Lakes, SA, 5095, Australia.*

Abstract

We propose and justify a model for seasonal rainfall at a single site using a checkerboard copula of maximum entropy for the joint probability distribution and a set of two-parameter gamma distributions for the marginal monthly distributions. The model allows correlation between individual months and thereby enables an improved model for seasonal variation. A central theme is the principle of maximum entropy which we use—subject to clearly stated assumptions—to find the most parsimonious representation for the underlying distributions required to model the relevant statistical characteristics. We used observations over 109 years from 1905 to 2013 to construct joint distributions for monthly rainfall during the wet season January-February-March-April at three locations—Cairns, Brisbane and Sydney—on the east coast of Australia. Empirical evidence from extensive simulations is used to validate each application. Simulated seasonal rainfall over 109 years exhibits a high degree of statistical variation at each site. We conclude that caution should be exercised when using observed records to study rainfall trends as records are typically too short-lived to capture long-term trends.

Keywords: Rainfall modelling and simulation, maximum entropy, checkerboard copula, gamma distribution.
2008 MSC: 15B51, 60E05, 65C05, 52A41

*Author to whom correspondence should be addressed.
Email address: julia.piantadosi@unisa.edu.au (Julia Piantadosi)

1. Introduction

We propose a model for seasonal rainfall using a checkerboard copula of maximum entropy to define a joint probability distribution for the monthly rainfall. To test our model, we used official records from the Australian Bureau of Meteorology (BoM) for the period 1905–2013 at three typical BoM stations on the east coast of Australia—031036 Cairns (Kuranda) and 040224 Brisbane (Alderly) in Queensland, and 066062 Sydney (Observatory Hill) in New South Wales. Cairns has a tropical monsoonal climate with significant wet and dry seasons. Brisbane and Sydney each have a humid sub-tropical or temperate climate with no pronounced dry season. The Köppen classification [1] for Cairns is *Am* while for both Brisbane and Sydney the classification is *Cfa*. Although the annual rainfall is relatively high and the wet season from January to April[1] is generally regarded as reliable, there is ample historical evidence of extended periods with below average rainfall.

There is consensus amongst climate scientists that summer and autumn rainfall in eastern Australia is influenced on a recurring basis by the quasi-periodic seasonal climatic events, El Niño and La Niña. During El Niño, rainfall is inhibited, and during La Niña, it is enhanced. It is therefore not especially surprising to find significant correlations in monthly rainfall at each location during the period January to April—the wettest time of the year. Our aim is to construct a parsimonious model at each location for a vector-valued random variable $\boldsymbol{X} = (X_1, X_2, X_3, X_4) \in \mathbb{R}^4$ that can be used to simulate typical monthly rainfall time series for the months of January, February, March and April. Repeated simulations with our proposed models over the same period as the observations show a high degree of variation in the key sample statistics but importantly the observed statistics lie well within the commonly accepted empirical confidence intervals established by these simulations. Our models generate stationary time series for both monthly and seasonal rainfall and the simulations suggest that even seemingly significant trends in the observed data could be due to chance alone.

To illustrate both the problem and the proposed solution we compared graphs of the observed data in the form of a time series to graphs of the corresponding simulated data generated by a joint probability distribution defined

[1]April is wetter than December at both Cairns and Sydney though not at Brisbane. On balance we chose January–April rather than December–March as a representative *wet season* for this study. Either way a similar analysis can be applied.

by the appropriate checkerboard copula of maximum entropy. The observed time series for monthly and seasonal rainfall totals over 109 years during the period January-February-March-April at BoM station 031036 Cairns (Kuranda) are shown in Figure 1. Our aim in this paper is to construct simulated time series that respect the key monthly and seasonal statistics from the observed rainfall time series. In particular the model is designed to incorporate the observed monthly correlations and thereby provide a realistic model for the seasonal variance. The simulated time series for the monthly and seasonal rainfall at BoM station 031036 Cairns (Kuranda) in a typical Trial $\#(C13, 109)$ are shown in Figure 2.

Similar time series can be displayed for the other two locations. We will show later that sample statistics at each location are highly variable in trial simulations over a period of 109 years. This certainly suggests the possibility that the observed sample statistics may not necessarily be an accurate representation of the true population statistics.

2. A brief literature review

A comprehensive review of the literature on rainfall modelling is not feasible here. Indeed we shall refer only to articles on rainfall modellng that are directly relevant to the methods used in this paper.

The principle of maximum entropy, enunciated by the physicist E. T. Jaynes [2, 3] in 1957, is fundamental to the methods used in this paper. We will apply this principle in two ways. We find a checkerboard copula of maximum entropy using the notion of discrete entropy and we justify our use of the gamma distribution by arguing that this distribution maximizes the continuous entropy when fitting a probability distribution to a set of strictly positive monthly rainfall totals. The modern notion of discrete entropy [4] was introduced by John von Neumann in his 1927 treatise on quantum mechanics in which he defined the entropy of a statistical operator $\rho = \{p_n, \psi_n\}_n$, where $p_n > 0$ and $\sum_n p_n = 1$ and where $\{\psi_n\}$ is a complete orthonormal system of basis vectors as the weighted ensemble average $S(\rho) = -k \langle \rho \log_e \rho \rangle_n = -k \,\mathrm{Tr}(\rho \log_e \rho) = -k \sum_n p_n \log_e p_n$. See ([5], pp. 348–353) for more details.

This measure was adopted in 1948 by C. E. Shannon [6] as a measure of information in the theory of communication systems. Shannon also introduced the analogous notion of continuous or differential entropy $S(f) = -\int_\Omega f(x) \log_e f(x) dx$, where $f(x) \geq 0$ and $\int_\Omega f(x) dx = 1$ for continuous

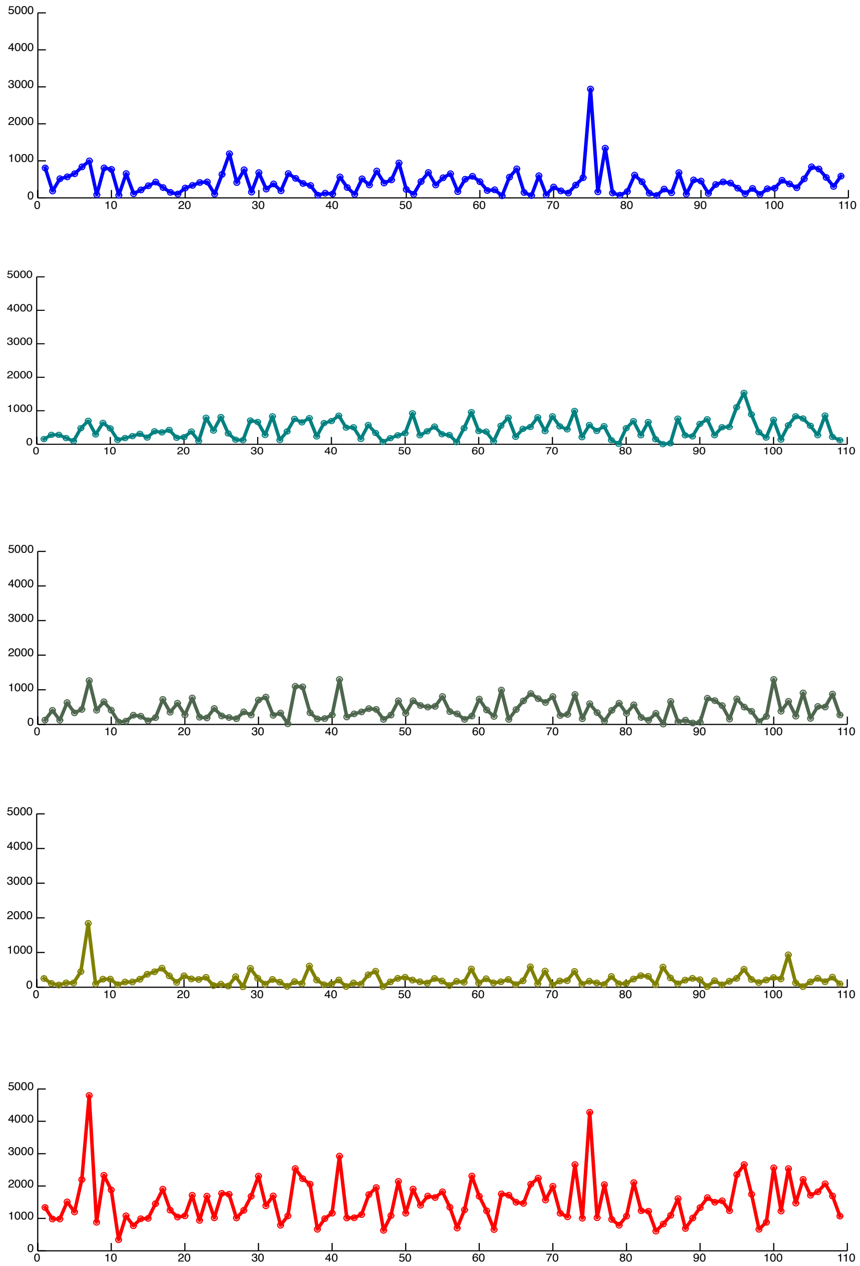


Figure 1: Time series for observed monthly and seasonal rainfall totals at BoM station 031036 Cairns (Kuranda) showing January **(top)**, February **(second top)**, March **(middle)**, April **(second bottom)** and Seasonal **(bottom)**.

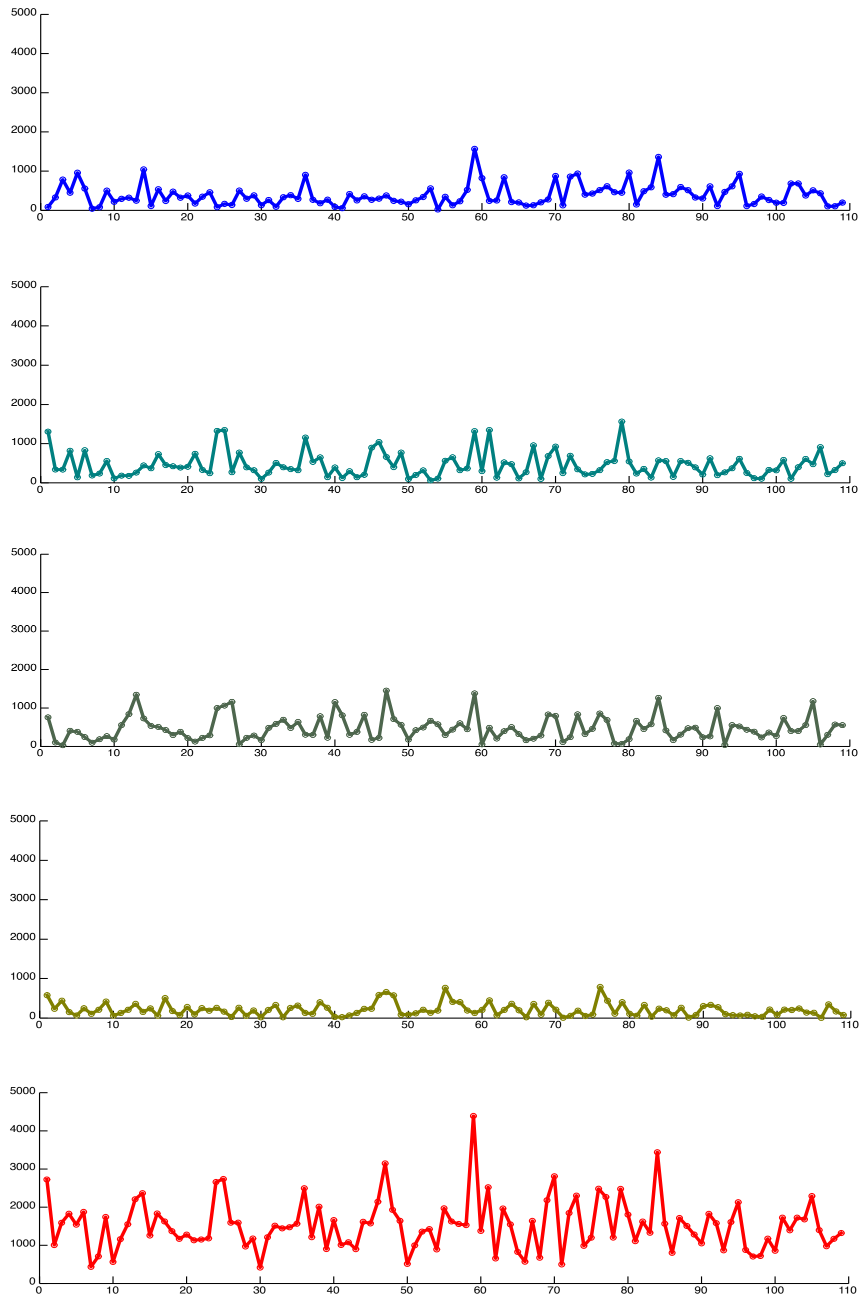


Figure 2: Time series for Trial #$(C13, 109)$ showing simulated monthly and seasonal rainfall totals at BoM station 031036 Cairns (Kuranda) showing January **(top)**, February **(second top)**, March **(middle)**, April **(second bottom)** and Seasonal **(bottom)**.

probability distributions. The entropy of a system is a measure of the inherent disorder. Entropy is maximized when the system is in the highest possible state of disorder subject to any imposed constraints. For a system with a finite number of states the entropy is maximized when all state probabilities are equal.

The early work on rainfall modelling, such as the paper by Stern and Coe [8], follows a classical style that is typical of the physical sciences. However, the focus has shifted in recent times to a pragmatic approach that is less concerned with a logical axiomatic basis and more concerned with a utilitarian outcome. In such cases there may be too much emphasis on fitting the observed data and a lack of awareness about sample variation. Such practices run the risk of *backtest overfitting*. See [7] for an extended discussion.

The most relevant recent paper to our work is a comprehensive 2005 report to the Australian Cooperative Research Centre for Catchment Hydrology by Srikanthan [9]. We will outline our concerns about the Srikanthan model but we refer readers to [10] for a more detailed discussion. Although Srikanthan describes a successful scheme for generation of daily rainfall data at multiple sites, a substantive difficulty emerges in the accumulation of simulated daily rainfall totals. This difficulty lies at the very heart of the problem we address here. Indeed Srikanthan himself makes the following critical observation.

> The generated daily rainfall amounts when aggregated into monthly and annual totals will not, in general, preserve the monthly and annual characteristics.

Consequently he implements a nested *a posteriori* correction process. The required correction tacitly acknowledges an axiomatic problem with the original model in which correlations in daily rainfall, although undoubtedly very small, are ignored. We believe it is logically inconsistent to select a gamma distribution that will generate realistic daily rainfall depths if one then intends to systematically modify the data generated by it. This inevitably means that the simulated daily rainfall depth distributions will be biased relative to the observed distributions. For additional remarks see [10]. We note also that the problem highlighted by Srikanthan can be overcome using the correlative coherence analysis proposed by Getz [11] or by using a Tweedie distribution as proposed by Hasan and Dunn [12]. However neither of these methods allows a detailed model of the individual correlations.

There is a large number of other papers that we could legitimately cite, but we mention only a few. For a more comprehensive review, we refer to Srikanthan and McMahon [13] and to an earlier review by Wilks and Wilby [14]. The over-dispersion phenomenon that bedevils the Srikanthan model [9] was studied by Katz and Parlange [15], who suggested that higher order Markov models can reduce apparent discrepancies in the number of generated wet days and the number of observed wet days. Rosenberg *et al.* [16] constructed a joint density using a Laguerre series to incorporate the correlation between successive months and hence correct the seasonal variance, but the optimal parametric structure of this model is unclear. Hasan and Dunn [12] have recently used a Tweedie distribution to model monthly rainfall. The model combines a Poisson process to generate wet and dry days and a collection of correlated gamma distributions to model daily rainfall depth. There is insufficient freedom in this model to match individual daily correlations, but it is possible to adjust the correlation parameters, so as to avoid the over-dispersion problem discussed above.

3. Stationarity for partially observed time-series

In practice, it may be possible to observe only a finite number of terms in a single realization $\{x_i\}_{i=1}^N$ of a doubly-infinite time series $\{x_i\}_{i=-\infty}^{\infty}$. In such cases Koutsoyannis [17] argues that it is difficult to tell whether an observed time series is stationary. Nevertheless, he suggests that it is useful to compare the standard deviation at scale k for the partially observed time series to the adjusted standard deviation at scale k for simulated observations of a known stationary time series over the same period. In [10] we used such a comparison to show that the seasonal rainfall for the months February-March-April at BoM station 059017 Kempsey (Wide Street) in New South Wales—between Sydney and Brisbane and also on the east coast—could not easily be distinguished from random simulations produced by a stationary distribution. We will not repeat these tests for Cairns, Brisbane and Sydney but refer to [10] for further discussion. We will nevertheless model the monthly rainfall at each of the current locations as a stationary time-series and then show retrospectively—as we did in [10] for the Kempsey study—that *apparent* trends in the observed monthly rainfall lie well within the empirical 95% confidence limits for *phantom* trends in simulated data generated by the stationary model. Thus we justify our assumption that the observed time-series can reasonably be modelled as a stationary time series.

4. Modelling monthly rainfall

The gamma distribution is often used to model rainfall accumulations. A common justification is that it is *sufficiently flexible* to model a wide range of observed data. While this may be true the argument is at best vague. The principle of maximum entropy [2, 3] provides an objective justification. This argument was presented in [10, 18] but is worth repeating here.

4.1. Maximum entropy and the gamma distribution

We assume X is a random variable and that the observed values $\{x_n\}_{n=1}^N$ are strictly positive. We seek a probability density $f : (0, \infty) \to (0, \infty)$ such that the differential entropy

$$h(f) = (-1) \int_0^\infty f(x) \log_e f(x) dx \tag{1}$$

is maximized subject to the additional constraints imposed by the observed means

$$E[X] = \overline{x} = \frac{1}{N} \sum_{n=1}^N x_n \quad \text{and} \quad E[\log_e X] = \overline{\log_e x} = \frac{1}{N} \sum_{n=1}^N \log_e x_n. \tag{2}$$

We can formulate this problem as a convex optimization with linear constraints. From the theory of Fenchel duality and the Fenchel-Young inequality [19, pp. 171-178] we have

$$\begin{aligned}
p &= \inf_{f \in L^1[(0,\infty)]} \left\{ -h(f) - 1 \mid E[1] = 1, E[X] = \overline{x}, E[\log_e X] = \overline{\log_e x} \right\} \\
&\geq \sup_{(\alpha,\beta,\kappa) \in \mathbb{R}^3} \left\{ \log_e \kappa - \overline{x}/\beta + (\alpha - 1)\overline{\log_e x} - \kappa \int_0^\infty x^{\alpha-1} e^{-x/\beta} dx \right\} \\
&= \sup_{(\alpha,\beta,\kappa) \in \mathbb{R}^3} \left\{ \log_e \kappa - \overline{x}/\beta + (\alpha - 1)\overline{\log_e x} - \kappa\, \Gamma(\alpha)\beta^\alpha \right\} \\
&= \sup_{(\alpha,\beta,\kappa) \in \mathbb{R}^3} \varphi(\alpha, \beta, \kappa) \\
&= -\log_e[\Gamma(\alpha)\beta] + (\alpha - 1)\psi(\alpha) - (\alpha + 1) = d
\end{aligned} \tag{3}$$

where the parameters α, β and κ are determined by the equations

$$\log_e \beta + \psi(\alpha) = \overline{\log_e x}, \quad \alpha\beta = \overline{x}, \quad \kappa(\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \tag{4}$$

and where $\psi(\alpha) = \Gamma'(\alpha)/\Gamma(\alpha)$ is the digamma function. The supremum and the conditions (4) are found simply by solving the equations $\partial\varphi/\partial\alpha = 0$, $\partial\varphi/\partial\beta = 0$ and $\partial\varphi/\partial\kappa = 0$. The function

$$f_{\alpha,\beta}(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}$$

which arises naturally in (3) when solving the dual optimization problem to find d is the probability density on $(0, \infty)$ for the gamma distribution with parameters α and β. If X is a random variable with this distribution we write $X \sim \Gamma(\alpha, \beta)$. In the case where α and β are determined by (4) then the additional constraints (2) are also satisfied. Since it is easy to show that $-h(f_{\alpha,\beta}) - 1 = d$ it follows that $p = d$ and that $f_{\alpha,\beta}$ is the unique solution to our original convex optimization problem. Note that the equations (4) are also the maximum likelihood equations used to estimate α and β if one has decided *a priori* to fit a gamma distribution.

Remark 4.1. One may argue with our decision to impose constraints that the mean of the random variable must equal the mean of the observed data and the mean of the logarithm of the random variable must equal the mean of the logarithm of the observed data. Indeed there are various other maximum entropy distributions that could be obtained by imposing alternative constraints. For instance a normal distribution is obtained if one insists that the mean and variance of the population should equal the mean and variance of the observed sample. A normal distribution is clearly not appropriate here because it allows negative values. We believe that the constraints used to derive the maximum entropy gamma distribution are more appropriate in this instance than any of the standard alternatives. For a more extensive discussion we refer, once again, to [10].

5. Modelling monthly rainfall using a gamma distribution

The observed monthly rainfall totals for December, January, February and March (months $i = 1, \ldots, 4$) at each location for the period 1905 to 2013 are all strictly positive. If we assume that the monthly rainfall distributions are stationary then according to the principle of maximum entropy we may reasonably propose the following null hypothesis: that the sequence of observed monthly rainfall totals at each location for month i can be modelled as the independently and successively generated outcomes of a real-valued random variable $X_i \sim \Gamma(\alpha_i, \beta_i)$ where $\alpha_i > 0$ and $\beta_i > 0$ are obtained from the observed monthly totals for month i by the method of maximum likelihood for each $i = 1, \ldots, 4$. At each location we will use the notation $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_4)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_4)$ to denote the maximum likelihood parameters. We obtained the following values.

Cairns:

$$\begin{aligned} \boldsymbol{\alpha} &= (1.8451, 1.9273, 1.9166, 1.6669), \\ \boldsymbol{\beta} &= (226.5323, 230.7580, 225.2654, 136.3662); \end{aligned}$$

Brisbane:

$$\begin{aligned} \boldsymbol{\alpha} &= (1.8683, 1.3665, 1.4058, 1.1635), \\ \boldsymbol{\beta} &= (88.5816, 121.7345, 99.1313, 77.9525); \end{aligned}$$

Sydney:

$$\begin{aligned} \boldsymbol{\alpha} &= (1.7141, 1.3498, 1.8642, 1.5012), \\ \boldsymbol{\beta} &= (62.0237, 85.8447, 70.0515, 80.4501). \end{aligned}$$

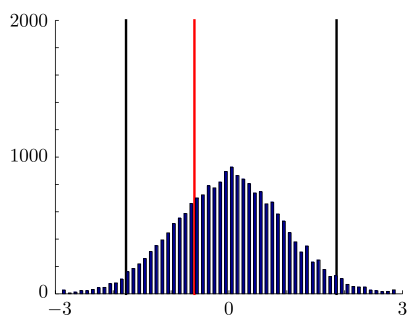
Although the independent generation of successive random numbers from a fixed gamma distribution must necessarily produce a stationary time-series it is nevertheless true that finite samples of such series may exhibit *phantom* trends. Our aim is to compare the trends (if any) in the series of observed monthly rainfalls for each month with the *phantom* trends in a large number of simulated data series of the same length generated successively and independently by the appropriate gamma distribution. We will show that the trends in observed monthly rainfall are so small that they could reasonably be regarded as *phantom* trends, due to chance alone.

We tested the null hypothesis by a linear regression on the observed time series of monthly rainfall totals and on each of 20000 simulated time series of monthly rainfall totals over the same period of $N = 109$ years with each simulated series generated by the proposed gamma distribution. Let $\{r_i(t)\}$ denote the rainfall in month i and year t. We used MATLAB to find (p_i, q_i), such that $\sum_{t=1}^{N} |r_i(t) - (p_i t + q_i)|^2$ is minimized. The slope, p_i, of this line is the trend-slope.

The trend-slopes for the observed datasets and the corresponding 95% confidence intervals for the trend-slopes of the simulated datasets at Cairns (Kuranda) were $p_1 = -0.5967 \in [-1.78, 1.86]$ in January, $p_2 = 1.8298 \in [-1.91, 1.94]$ in February, $p_3 = 0.6333 \in [-1.87, 1.85]$ in March and $p_4 = -0.4887 \in [-1.06, 1.07]$ in April. We also tested the observed seasonal rainfall against the simulated seasonal rainfall using a maximum likelihood gamma distribution $X_t \sim \Gamma(\alpha_t, \beta_t)$ with $\alpha_t = 5.7335$ and $\beta_t = 265.4147$. We found $p_t = 1.3777 \in [-3.72, 3.77]$. Similar tests were carried out at the other sites. Overall there was only one failure in a total of fifteen tests.

We conclude that there is insufficient evidence to reject the hypothesis that the observed rainfall totals are the outcomes of a stationary random variable. This means that the apparent observed trends could reasonably be regarded as due to chance alone. The results of our simulations for the trend-slopes of the monthly rainfalls at BoM station 031036 Cairns (Kuranda) are shown in Figures 3 and 4.

Although most climate scientists expect rainfall events in eastern Australia to become more extreme and although such changes could conceivably lead to more extreme monthly rainfall distributions, we believe there is currently no firm agreement about such rainfall trends.



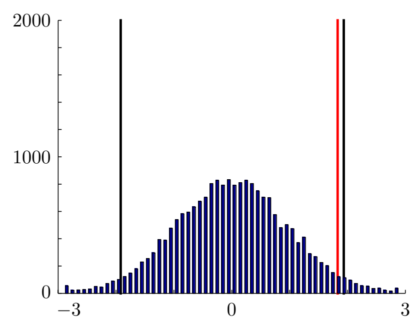


Figure 3: Trend-slope histograms for 20000 simulated rainfall datasets at BoM Station 031036 Cairns (Kuranda) generated by maximum likelihood gamma distributions for January (**left**) and February (**right**). The vertical red lines show trend-slopes for the observed datasets lying inside the empirical 95% confidence intervals for the simulated trend-slopes.

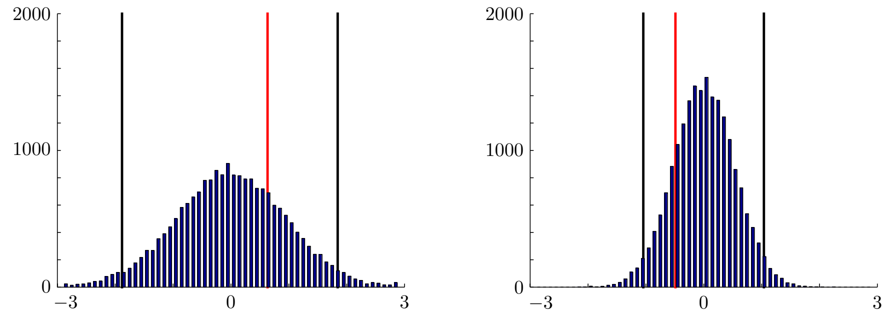


Figure 4: Trend-slope histograms for 20000 simulated rainfall datasets at BoM Station 031036 Cairns (Kuranda) generated by maximum likelihood gamma distributions for March (**left**) and April (**right**). The vertical red lines show trend-slopes for the observed datasets lying inside the empirical 95% confidence intervals for the simulated trend-slopes.

6. Q-Q Plots to test the Goodness-of-Fit for the simulated time series of monthly rainfalls

We demonstrated the goodness-of-fit for the observed monthly rainfall data to the designated gamma distributions using Q-Q plots.

Firstly, we used the designated gamma distribution to generate 1000 simulated datasets for each month at each location. Then, we plotted the simulated quantiles against the theoretical quantiles. The results for January rainfall at BoM station 031036 Cairns (Kuranda) are shown in Figure 5 on the left. These plots show the full range of variation expected for the designated gamma random variable from 1000 samples, each of size $N = 109$. By discarding the bottom 25 and top 25 values for each quantile from the simulated datasets, we found empirical 95% confidence intervals.

Secondly, we plotted the observed quantiles against the theoretical quantiles using the designated gamma distribution for each month at each location. The results for January rainfall at BoM station 031036 Cairns (Kuranda) are shown in Figure 5 on the right. We used grey bars on these plots to show the empirical 95% confidence intervals for the quantiles obtained from the simulated datasets described above.

Similar plots were produced for all months at each location. Overall the Q-Q plots showed there are no recognised statistical grounds to reject the hypothesis that the monthly rainfall totals can be modelled by the designated maximum likelihood gamma distributions.

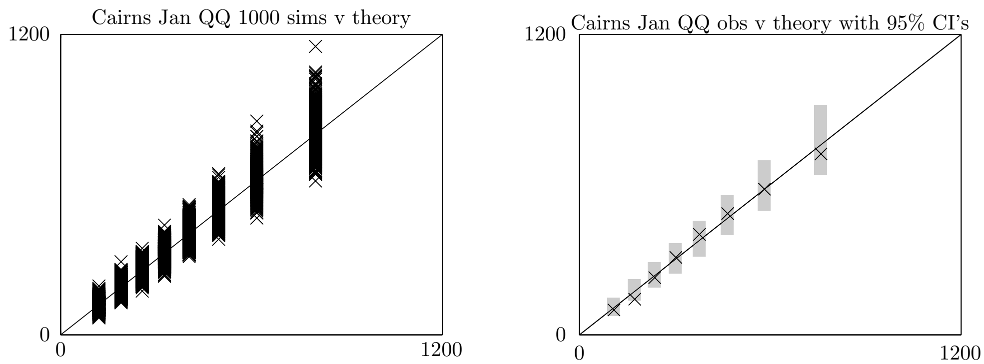


Figure 5: Q-Q plots of January quantiles at BoM station 031036 Cairns (Kuranda) for 1000 simulated datasets each covering a period of 109 years *versus* corresponding theoretical quantiles for $X_1 \sim \Gamma(\alpha_1, \beta_1)$ **(left)** and Q-Q plots for observed January quantiles at BoM station 031036 Cairns (KUranda) *versus* theoretical quantiles for $X_1 \sim \Gamma(\alpha_1, \beta_1)$ (**right**) where the vertical grey bars show empirical 95% confidence intervals for the simulated quantiles generated by $X_1 \sim \Gamma(\alpha_1, \beta_1)$.

Remark 6.1. We have been taken to task by some for not comparing our proposed model to other models currently in popular use. Nevertheless, our model is based on well-established scientific methodology. We use the principle of maximum entropy to argue that the maximum likelihood gamma distribution is the most appropriate model for rainfall accumulations in which the observed totals are strictly positive. While it is true that we make an arbitrary decision to use the sample means of the observed data and log-data as axiomatic constraints, we note that in any modelling process, some assumptions are necessary. At the very least the assumptions and the subsequent logic are clearly defined. Once the numerical parameters have been calculated we test our model by showing that the observed data lies well within the empirical 95% confidence intervals established by repeated simulations using the model. The conclusion is clear: there are no reasonable statistical grounds for rejecting the model. The argument that other models may provide a better fit to the observed data is essentially irrelevant. Indeed, this criticism embraces a fundamental misconception that an observed sample is always a true representation of the entire population. Moreover, the suggested iterative correction methods used by Srikanthan and others are subject to concerns about overfitting [7]. A legitimate criticism of our model would need to argue either that the principle of maximum entropy is inappropriate or else that we should incorporate more suitable constraints on the observed data.

6.1. The transformed data—removal of seasonal effects

If $X_i \sim \Gamma(\alpha_i, \beta_i)$ where $F_i(x) = F_{\alpha_i,\beta_i}(x)$ for month i at a given location then the transformed monthly rainfalls $U_i = F_i(X_i)$ are uniformly distributed on $(0, 1)$ and so seasonal factors are removed from the observed data. Figure 6 shows the histogram for the transformed observed January rainfall totals $u_{i,j} = F_i(x_{i,j})$ for $j = 1, \ldots, 109$ at BoM station 031036 Cairns (Kuranda). We used the binomial distribution with $N = 109$, $p = 0.1$ and $q = 0.9$ to calculate approximate 95% confidence intervals $I = (p - 1.96\sqrt{pq/N}, p + 1.96\sqrt{pq/N}) = (0.044, 0.156)$ for the heights of the bars. Similar plots were produced for all months at each location. Overall the plots showed that there were no recognised statistical grounds to reject the hypothesis that the transformed totals were uniformly distributed.

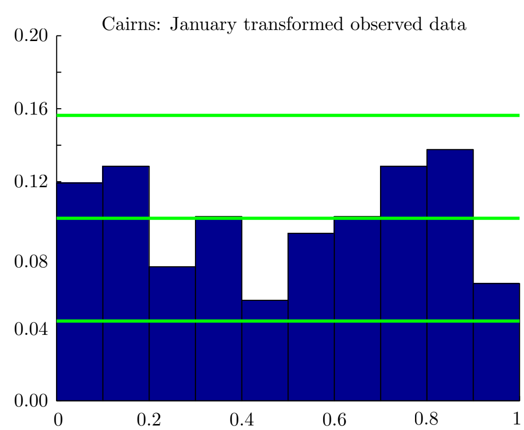


Figure 6: Histogram for the transformed observed January rainfall totals $u_{1,j} = F_1(x_{1,j})$ for $j = 1, \ldots, 109$ at BoM station 031036 Cairns (Kuranda) where $F_1 = F_{\alpha_1,\beta_1}$ is the cumulative distribution for $\Gamma(\alpha_1, \beta_1)$. The horizontal lines show the mean value and the upper and lower bounds for the 95% confidence intervals.

7. A joint probability distribution for the seasonal rainfall

The next step in the modelling process is to construct a joint probability distribution for the entire four-month time period at each location. We will do this in what we believe is the most natural way—by using the principle of maximum entropy. Past studies of rainfall accumulations over several months [15, 16] have concluded that the variance of the simulated time series of seasonal rainfall totals generated by models with independent marginal distributions is often not consistent with the observed variance. In particular if the observed monthly marginal distributions show an overall positive correlation we would expect the observed variance in seasonal rainfall to be higher than one would find with independent marginal monthly distributions. In any event we cannot expect a model with independent marginal distributions to correctly simulate the seasonal rainfall patterns generated by correlated marginal distributions.

Our aim will be to construct a joint distribution that not only preserves the desired monthly rainfall characteristics, but also replicates the observed variance in the seasonal rainfall totals. We will do this by incorporating marginal correlations using a checkerboard copula of maximum entropy.

8. Copulas with prescribed correlation

An m-dimensional *copula*, where $m \geq 2$, is a cumulative probability distribution $C(\boldsymbol{u}) \in [0, \infty)$ defined on the m-dimensional unit hypercube $\boldsymbol{u} = (u_1, u_2, \ldots, u_m) \in [0,1]^m$ for a vector-valued random variable $\boldsymbol{U} = (U_1, U_2, \ldots, U_m)$ with uniform marginal probability distributions for the real-valued random variables $U_1, U_2, \ldots, U_m$. See [20, 21]. The correlation coefficients for the joint distribution are defined by

$$\rho_{r,s} = \frac{E[(U_r - 1/2)(U_s - 1/2)]}{\sqrt{E[(U_r - 1/2)^2]\, E[(U_s - 1/2)^2]}} = 12E[U_r U_s] - 3 \qquad (5)$$

for each $1 \leq r < s \leq m$. In order to model the joint probability distribution for a vector-valued random variable $\boldsymbol{X} = (X_1, X_2, \ldots, X_m) \in (0, \infty)^m$ with known marginals $u_i = F_i(x_i)$ we simply construct uniformly distributed random variables $U_i = F_i(X_i) \in (0,1)$ for each $i = 1, 2, \ldots, m$ and use the m-dimensional copula $C(\boldsymbol{u}) = C(\boldsymbol{F}(\boldsymbol{x})) = C(F_1(x_1), F_2(x_2), \ldots, F_m(x_m))$. We say that the grade correlation coefficients for $\boldsymbol{X}$ are simply the correlation coefficients for $\boldsymbol{U} = (U_1, U_2, U_3, U_4)$ defined above. That is

$$\begin{aligned} \rho_{r,s} &= \frac{E[(F_r(X_r) - 1/2)(F_s(X_s) - 1/2)]}{\sqrt{E[(F_r(X_r) - 1/2)^2]\, E[(F_s(X_s) - 1/2)^2]}} \\ &= 12E[F_r(X_r)F_s(X_s)] - 3 \end{aligned} \tag{6}$$

for each $1 \leq r < s \leq m$. We distinguish between the Spearman rank correlation coefficients defined from the observed data $\{x_{r,j}\}_{j=1,\ldots,N}$ and $\{x_{s,j}\}_{j=1,\ldots,N}$ for $1 \leq r < s \leq m$ as the Pearson correlation coefficients of the ranks of the observed data [21] and the grade correlation coefficients $\boldsymbol{\rho} = [\rho_{r,s}]$ defined by (6). Although the two measurements are similar they are not the same. For all samples in this paper we will use the *observed* grade correlation coefficients defined by

$$\widehat{\rho}_{r,s} = \frac{\sum_{j=1}^{N}(u_{r,j} - \bar{u}_r)(u_{s,j}) - \bar{u}_s)}{\sqrt{\sum_{j=1}^{N}(u_{r,j} - \bar{u}_r)^2\, \sum_{j=1}^{N}(u_{s,j}) - \bar{u}_s)^2}} \tag{7}$$

for each $1 \leq r < s \leq m$ where $u_{i,j} = F_i(x_{i,j})$ and $\bar{u}_i = \sum_{j=1}^{N} u_{i,j}/N$ for each $i = 1, \ldots, m$. The *observed* grade correlation coefficients $\widehat{\rho}_{r,s}$ are simply the Pearson correlation coefficients for the transformed data $\{u_{r,j}\}_{j=1,\ldots,N}$ and $\{u_{s,j}\}_{j=1,\ldots,N}$ for $1 \leq r < s \leq m$. The Spearman rank correlations could be used throughout for all samples—the observed data and the simulated data—in place of the observed grade correlation coefficients. Similar results will be obtained.

9. Modelling the joint probability with a checkerboard copula

We construct a joint distribution using a checkerboard copula of maximum entropy [22, 23]. A 4-dimensional *checkerboard* copula is a probability distribution defined by subdividing the unit 4-dimensional hypercube into n^4 congruent small hypercubes with constant density on each one. If the density on $I_{ijk\ell}$ is given by $n^3 h_{ijk\ell}$ then the marginal distributions will be uniform if

$$\sum_{j,k,\ell} h_{ijk\ell} = 1\ \forall i, \quad \sum_{i,k,\ell} h_{ijk\ell} = 1\ \forall j, \quad \sum_{i,j,\ell} h_{ijk\ell} = 1\ \forall k, \quad \sum_{i,j,k} h_{ijk\ell} = 1\ \forall \ell.$$

In such cases we say that $\boldsymbol{h} = [h_{ijk\ell}]$ is quadruply-stochastic. We wish to construct a joint density in this form with the desired correlations. For sufficiently large n there are many ways that this can be done. *The principle* *of maximum entropy suggests that the best such distribution is the most disordered or least prescriptive solution—the quadruply-stochastic hyper-matrix* $\boldsymbol{h}$ *which has the most equal subdivision of probabilities but still allows the required correlations.*

Problem 9.1 (The primal problem). *Find the hyper-matrix* $\boldsymbol{h} = [h_{\boldsymbol{i}}] \in \mathbb{R}^\ell$ *where* $\boldsymbol{i} = (i_1, \ldots, i_m)$ *and* $\ell = n^m$ *to maximize the entropy*

$$J(\boldsymbol{h}) = (-1) \left[\frac{1}{n} \sum_{\boldsymbol{i} \,\in\, \{1,\ldots,n\}^m} h_{\boldsymbol{i}} \log_e h_{\boldsymbol{i}} + (m-1) \log_e n \right] \tag{8}$$

subject to the multi-stochastic constraints

$$\sum_{j \neq r,\, i_j \in \{1,\ldots,n\}} h_{\boldsymbol{i}} = 1 \tag{9}$$

for all $i_r \in \{1, \ldots, n\}$ *and each* $r = 1, \ldots, m$ *and* $h_{\boldsymbol{i}} \geq 0$ *for all* $\boldsymbol{i} \in \{1, \ldots, n\}^m$ *and the grade correlation coefficient constraints*

$$12 \left[\frac{1}{n^3} \cdot \sum_{\boldsymbol{i} \,\in\, \{1,\ldots,n\}^m} h_{\boldsymbol{i}} (i_r - 1/2)(i_s - 1/2) \right] - 3 = \widehat{\rho}_{r,s} \tag{10}$$

for $1 \leq r < s \leq m$ *where* $\widehat{\rho}_{r,s}$ *is known for all* $1 \leq r < s \leq m$.

Problem 9.1 is solved using the theory of Fenchel duality. See [22, 23] for details of the solution and [19, 24] for the underlying theory. The m-dimensional copula of maximum entropy is defined by $m(m-1)/2$ real parameters—the grade correlation coefficients—defined in equation (6).

9.1. A model for January–February–March–April rainfall at Cairns

The quadruply-stochastic hyper-matrix $\boldsymbol{h} \in \mathbb{R}^{6\times6\times6\times6}$ with $m = 4$ and $n = 6$ defining the quadrivariate checkerboard copula of maximum entropy for rainfall at BoM station 031036 Cairns (Kuranda) during the January-February-March-April season was calculated using a special MATLAB program. The copula is shown to four decimal place accuracy in Appendix A. The grade correlation coefficients were constrained by setting $\boldsymbol{\rho} = \widehat{\boldsymbol{\rho}}$ where $\widehat{\boldsymbol{\rho}}$ is the matrix of observed grade correlation coefficients. The entropy was calculated using (8) and was found to be $J(\boldsymbol{h}) \approx -0.05904$. Similar results are obtained if the copula of maximum entropy is replaced by a checkerboard normal copula although numerical calculation of the latter is considerably more difficult and the entropy is slightly less. More information about the definition and computation of these two checkerboard copulas can be found in [22, 23].

All numerical calculations were performed in MATLAB and the relevant m-files are freely available from the *CARMA* website [25] or from the corresponding author Dr. Julia Piantadosi. The MATLAB program computed the relevant hyper-matrix in 11.77 s on a MacBook Pro OS X laptop computer.

Checkerboard copulas of maximum entropy with the same dimensions were also calculated to model seasonal rainfall in January-February-March-April at BoM stations 040224 Brisbane (Alderly) and 066062 Sydney (Observatory Hill). The defining hyper-matrices can be found in the relevant MATLAB m-files on the *CARMA* website [25].

10. Summary of the key observed statistics

The key observed statistics at each site are shown in the following list.

Observations for Cairns: The respective monthly and seasonal means are given by $\bar{\boldsymbol{x}} = (418, 445, 432, 227)$ mm and $\bar{\boldsymbol{t}} = 1522$ mm. The standard deviation and variance are $s = 685$ mm and $s^2 = 469720$ mm^2. The observed grade correlation coefficients are

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & 0.0729 & -0.0191 & -0.0536 \\ & 1.0000 & 0.2645 & 0.1459 \\ & & 1.0000 & 0.1212 \\ & & & 1.0000 \end{bmatrix}.$$

Observations for Brisbane: The respective monthly and seasonal means are given by $\bar{\boldsymbol{x}} = (165, 166, 139, 91)$ mm and $\bar{\boldsymbol{t}} = 562$ mm. The standard deviation and variance are $s = 259$ mm and $s^2 = 67332$ mm^2. The observed grade correlation coefficients are

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & 0.0622 & 0.1834 & -0.0286 \\ & 1.0000 & 0.1509 & 0.1179 \\ & & 1.0000 & 0.0880 \\ & & & 1.0000 \end{bmatrix}.$$

Observations for Sydney: The respective monthly and seasonal means are given by $\bar{\boldsymbol{x}} = (106, 116, 131, 121)$ mm and $\bar{\boldsymbol{t}} = 474$ mm. The standard deviation and variance are $s = 187$ mm and $s^2 = 34959$ mm^2. The observed grade correlation coefficients are

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & 0.1197 & -0.0351 & -0.1987 \\ & 1.0000 & -0.0754 & -0.0627 \\ & & 1.0000 & -0.0269 \\ & & & 1.0000 \end{bmatrix}.$$

11. Experimental results

We tested our model using extensive trials at each location. Each trial consisted of a sequence of independently and randomly generated seasonal rainfalls using the relevant copula of maximum entropy. Thus each trial produced four contemporaneous sequences of monthly rainfall totals at each location. We conducted 20 basic trials at each site with each trial consisting of 109 successive, independently generated, random simulations of seasonal rainfall. By repeating the trials we were able to compare the observed statistics to the simulated statistics and empirically investigate the level of uncertainty in sample statistics for samples of this size. We also carried out a smaller number of extended trials over a period of 5450 years to generate larger samples with more stable sample statistics that better approximate the underlying population statistics for the model. In order to check our model for the overall seasonal rainfall at each location we also fitted a maximum likelihood gamma distribution to the observed data. Although this is a recognised method for modelling the total seasonal rainfall it cannot be used in simulation to generate seasonal rainfall totals with corresponding monthly subtotals. If one uses a maximum likelihood gamma distribution to model monthly rainfall for each month and then adds together the independently generated monthly totals to form a seasonal total then many authors [15, 16] have found that the variance of the simulated sums does not match the observed variance. The joint distribution defined by the copula of maximum entropy allows us to simulate individual monthly totals with each seasonal total. We will also show that it generally provides a much improved simulation of the seasonal variance especially in situations where there are significant correlations in monthly rainfall.

11.1. Results for BoM station 031036 Cairns

The observed seasonal totals for Cairns and the corresponding maximum likelihood gamma distribution are shown in Figure 7 on the left. To provide an initial assessment of the model using the checkerboard copula of maximum entropy we conducted 3 successive trials with each trial covering a period of 5450 years. The histogram for trial $\#(C1, 5450)$ is shown in Figure 7 on the right. The full statistics for the trial are shown below. These should be compared to the statistics for the observed rainfall shown in the previous section.

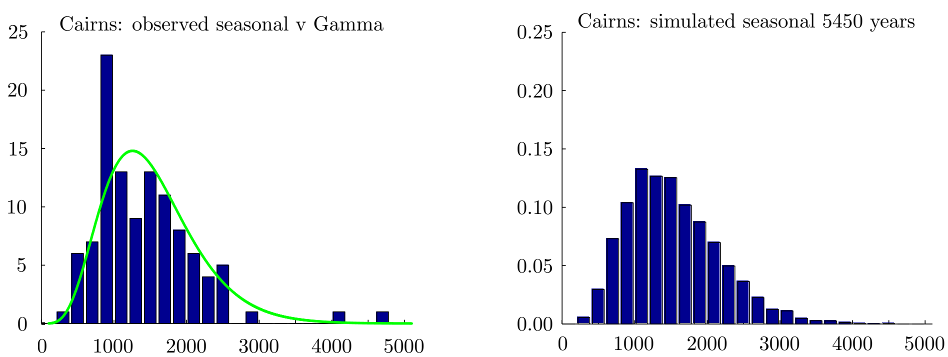


Figure 7: Histogram of observed seasonal rainfall at BoM station 031036 Cairns (Kuranda) for January-February-March-April with corresponding maximum likelihood gamma distribution (**left**) and histogram of Trial $\#(C1, 5450)$ (**right**) from 3 successive trials using the copula of maximum entropy with each trial covering a period of $N = 5450$ years showing typical sample characteristics.

Trial $\#(C1, 5450)$**:** The monthly means were $\bar{x} = (415, 450, 432, 230)$. The overall seasonal mean was 1526 with standard deviation $s = 627$ and corresponding variance $s^2 = 393470$. The observed grade correlation coefficients were

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & 0.0739 & -0.0249 & -0.0531 \\ & 1.0000 & 0.2659 & 0.1336 \\ & & 1.0000 & 0.1076 \\ & & & 1.0000 \end{bmatrix}.$$

For the seasonal rainfall at Cairns the observed mean was 1522 mm and the observed standard deviation and variance were respectively 685 mm and 469720 mm^2. If the seasonal rainfall is modelled with a maximum likelihood gamma distribution then the model mean is 1522 mm while the model standard deviation and variance are respectively 637 mm and 406070 mm^2. If the seasonal rainfall is modelled by the sum of independently distributed maximum likelihood gamma distributions then the model mean is 1522 and the model standard deviation and variance are respectively 571 mm and 325570 mm^2. For Trial $\#(C1, 5450)$ the simulated mean was 1526 and the simulated standard deviation and variance were respectively 627 mm and 393470 mm^2. It is possible to calculate a theoretical value for the variance of the copula of maximum entropy. The details can be found in [23].

In order to investigate the likely variation in sample statistics we conducted 20 successive trials with each trial covering the same period of 109 years as the observed rainfall. Histograms for Trial $\#(C9, 109)$ with the lowest mean $\bar{x} = 1386$ mm and Trial $\#(C1, 109)$ with the highest mean $\bar{x} = 1641$ mm are shown in Figure 8. We also note that Trial $\#(C15, 109)$ had the lowest standard deviation $s = 545$ mm and corresponding variance $s^2 = 296480$ mm^2 and that Trial $\#(C5, 109)$ had the highest standard deviation $s = 693$ mm and corresponding variance $s^2 = 480080$ mm^2. The time series shown earlier in Figure 2 are from Trial $\#(C13, 109)$. The full details of these trials are shown below.

Trial $\#(C1, 109)$**:** The monthly means were $\bar{x} = (433, 500, 477, 230)$ mm. The overall seasonal mean was 1641 mm with standard deviation $s = 684$ mm and corresponding variance $s^2 = 467870$ mm^2. The observed grade correlation coefficients were

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & -0.0063 & 0.0114 & -0.1573 \\ & 1.0000 & 0.2997 & 0.1069 \\ & & 1.0000 & 0.0125 \\ & & & 1.0000 \end{bmatrix}.$$

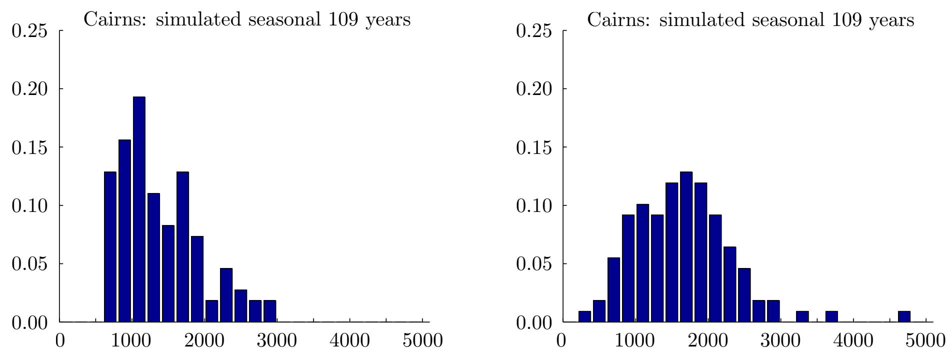


Figure 8: Selected histograms for total rainfall from 20 successive trials at BoM station 031036 Cairns (Kuranda) for January-February-March-April with each trial covering a period of $N = 109$ years using the copula of maximum entropy. Trial $\#(C9, 109)$ (**left**) shows the trial with lowest mean and Trial $\#(C1, 109)$ (**right**) shows the trial with highest mean.

Trial $\#(C5, 109)$**:** The monthly means were $\bar{x} = (413, 411, 440, 223)$ mm. The overall seasonal mean was 1487 mm with standard deviation $s =$ 693 mm and corresponding variance $s^2 = 480080$ mm^2. The observed grade correlation coefficients were

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & 0.0572 & -0.0634 & 0.0122 \\ & 1.0000 & 0.3029 & 0.2314 \\ & & 1.0000 & 0.3212 \\ & & & 1.0000 \end{bmatrix}.$$

Trial $\#(C9, 109)$ The monthly means were $\bar{x} = (387, 400, 409, 190)$ mm. The overall seasonal mean was 1386 mm with standard deviation $s =$ 561 mm and corresponding variance $s^2 = 314280$ mm^2. The observed grade correlation coefficients were

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & 0.0553 & -0.0611 & -0.0899 \\ & 1.0000 & 0.4353 & 0.0858 \\ & & 1.0000 & 0.1528 \\ & & & 1.0000 \end{bmatrix}.$$

Trial $\#(C13, 109)$ The monthly means were $\bar{x} = (389, 458, 474, 207)$ mm. The overall seasonal mean was 1528 mm with standard deviation $s =$

662 mm and corresponding variance $s^2 = 437650$ mm^2. The observed grade correlation coefficients were

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & 0.2285 & -0.0545 & -0.0167 \\ & 1.0000 & 0.0898 & 0.3402 \\ & & 1.0000 & 0.1459 \\ & & & 1.0000 \end{bmatrix}.$$

Trial $\#(C15, 109)$ The monthly means were $\bar{x} = (417, 434, 472, 212)$ mm. The overall seasonal mean was 1535 mm with standard deviation $s =$ 545 mm and corresponding variance $s^2 = 296480$ mm^2. The observed grade correlation coefficients were

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & -0.0764 & -0.0342 & -0.0453 \\ & 1.0000 & 0.1347 & 0.0713 \\ & & 1.0000 & 0.0444 \\ & & & 1.0000 \end{bmatrix}.$$

It is clear from these trials that a high degree of variation is possible in simulated seasonal rainfall for Cairns over a period of 109 years with the proposed model. We do not believe that this is an artefact of the model as the observed statistics also indicate a high standard deviation. Our results do however suggest that caution should be exercised when using the observed sample to construct a population model or to make inferences about likely trends. We also note that the variance of the proposed model is much closer to the observed variance than is the case for a model with independent marginal monthly distributions.

11.2. Results for BoM station 040224 Brisbane

The observed seasonal totals for Brisbane and the corresponding maximum likelihood gamma distribution are shown in Figure 7 on the left. To provide an initial assessment of the model using the checkerboard copula of maximum entropy we conducted 3 successive trials with each trial covering a period of 5450 years. The histogram for trial $\#(B3, 5450)$ is shown in Figure 9 on the right. The full details for the trial are shown below. These should be compared to the details for the observed totals shown earlier in the paper.

Trial $\#(B3, 5450)$**:** The monthly means were $\bar{x} = (167, 166, 140, 91)$. The overall seasonal mean was 564 with standard deviation $s = 262$ and

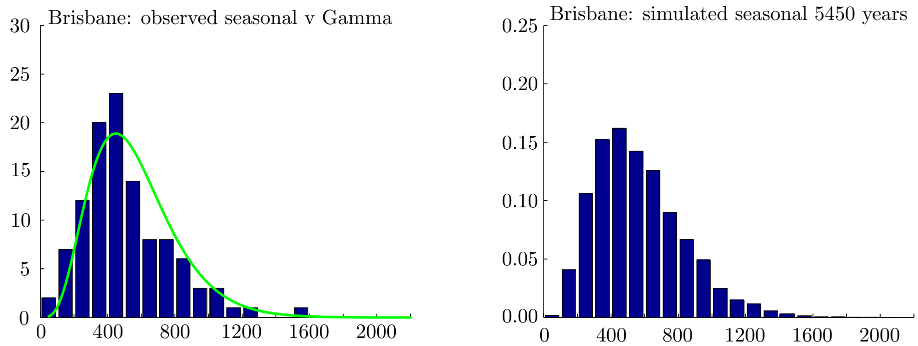


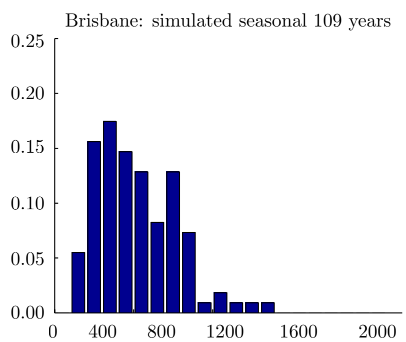
Figure 9: Histogram for observed total rainfall at BoM station 040224 Brisbane (Alderly) for January-February-March-April with corresponding maximum likelihood gamma distribution (**left**) and histogram for Trial $\#(B3, 5450)$ (**right**) from 3 successive trials using the copula of maximum entropy with each trial covering a period of $N = 5450$ years showing typical sample characteristics.

corresponding variance $s^2 = 68729$. The observed grade correlation coefficients were

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & 0.0597 & 0.1735 & -0.0334 \\ & 1.0000 & 0.1590 & 0.1158 \\ & & 1.0000 & 0.0913 \\ & & & 1.0000 \end{bmatrix}.$$

For the seasonal rainfall at Brisbane the observed mean was 562 mm and the observed standard deviation and variance were respectively 259 mm and 67332 mm^2. If the seasonal rainfall is modelled with a maximum likelihood gamma distribution the mean is 259 mm while the standard deviation and variance are respectively 252 mm and 63669 mm^2. If the seasonal rainfall is modelled by the sum of independently distributed maximum likelihood gamma distributions then the mean is 259 and the standard deviation and variance are respectively 236 mm and 55795 mm^2. For Trial $\#(B3, 5450)$ the simulated mean was 564 and the simulated standard deviation and variance were respectively 262 mm and 68729 mm^2.

In order to investigate the likely variation in sample statistics at Brisbane we conducted 20 successive trials with each trial covering the same period of 109 years as the observed rainfall. Histograms for Trial $\#(B5, 109)$ with the lowest mean $\bar{x} = 521$ mm and Trial $\#(B16, 109)$ with the highest mean $\bar{x} = 613$ mm are shown in Figure 10. We also noted that Trial $\#(B14, 109)$ had the lowest standard deviation $s = 223$ mm and corresponding variance $s^2 = 49654$ mm^2 and that Trial $\#(B3, 109)$ had the highest standard deviation $s = 302$ mm and corresponding variance $s^2 = 91168$ mm^2. The full details of these trials are shown below.



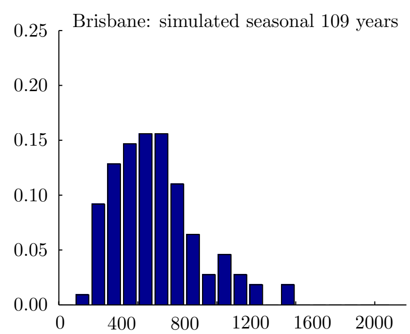


Figure 10: Selected histograms for total rainfall from 20 successive trials at BoM station 040224 Brisbane (Alderly) for January-February-March-April with each trial covering a period of $N = 109$ years using the copula of maximum entropy. Trial $\#(B5, 109)$ (**left**) shows the trial with lowest mean and Trial $\#(B16, 109)$ (**right**) shows the trial with highest mean.

Trial $\#(B3, 109)$ The monthly means were $\bar{x} = (168, 167, 148, 94)$ mm. The overall seasonal mean was 576 mm with standard deviation $s = 302$ mm and corresponding variance $s^2 = 91168$ mm^2. The observed grade correlation coefficients were

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & 0.0971 & 0.3252 & -0.0405 \\ & 1.0000 & 0.1650 & 0.2691 \\ & & 1.0000 & 0.1772 \\ & & & 1.0000 \end{bmatrix}.$$

Trial $\#(B5, 109)$ The monthly means were $\bar{x} = (150, 146, 138, 88)$ mm. The overall seasonal mean was 521 mm with standard deviation $s = 251$ mm and corresponding variance $s^2 = 62823$ mm^2. The observed grade correlation coefficients were

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & 0.1400 & 0.3551 & -0.1059 \\ & 1.0000 & 0.3009 & 0.0814 \\ & & 1.0000 & 0.1143 \\ & & & 1.0000 \end{bmatrix}.$$

Trial $\#(B14, 109)$ The monthly means were $\bar{x} = (165, 165, 162, 81)$ mm. The overall seasonal mean was 573 mm with standard deviation $s = 223$ mm and corresponding variance $s^2 = 49654$ mm^2. The observed grade correlation coefficients were

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & -0.0257 & 0.1210 & -0.0685 \\ & 1.0000 & -0.0426 & 0.0351 \\ & & 1.0000 & 0.0359 \\ & & & 1.0000 \end{bmatrix}.$$

Trial $\#(B16, 109)$ The monthly means were $\bar{x} = (192, 155, 166, 100)$ mm. The overall seasonal mean was 613 mm with standard deviation $s = 272$ mm and corresponding variance $s^2 = 74208$ mm^2. The observed grade correlation coefficients were

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & -0.1259 & 0.0845 & -0.0240 \\ & 1.0000 & 0.2538 & 0.0314 \\ & & 1.0000 & 0.1387 \\ & & & 1.0000 \end{bmatrix}.$$

11.3. Results for BoM station 066062 Sydney

The observed seasonal totals for Sydney and the corresponding maximum likelihood gamma distribution are shown in Figure 11 on the left. To provide an initial check on the model using the checkerboard copula of maximum entropy we conducted 3 successive trials with each trial covering a period of 5450 years. The histogram for trial $\#(S1, 5450)$ is shown in Figure 11 on the right. The full details for the trial are shown below. These should be compared to the details for the observed totals shown earlier in the paper.

Trial $\#(S1, 5450)$**:** The monthly means were $\bar{x} = (108, 114, 130, 123)$. The overall seasonal mean was 475 with standard deviation $s = 179$ and

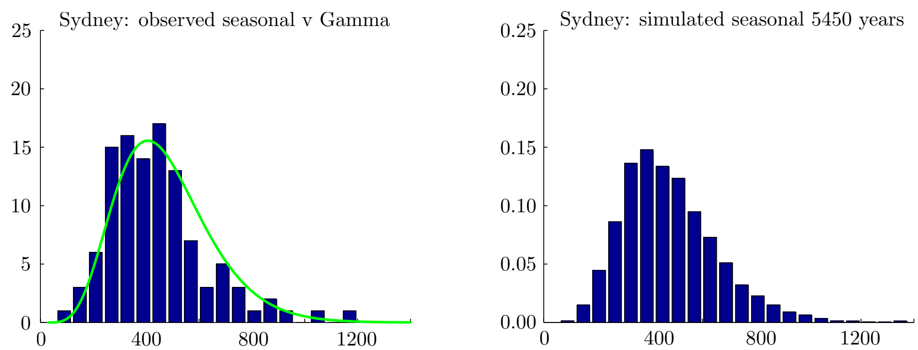


Figure 11: Histogram for observed total rainfall at BoM station 066062 Sydney (Observatory Hill) for January-February-March-April with corresponding maximum likelihood gamma distribution (**left**) and histogram for Trial $\#(S1, 5450)$ (**right**) from 3 successive trials using the copula of maximum entropy with each trial covering a period of $N = 5450$ years showing typical sample characteristics.

corresponding variance $s^2 = 32105$. The observed grade correlation coefficients were

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & 0.1148 & -0.0327 & -0.1919 \\ & 1.0000 & -0.0906 & -0.0675 \\ & & 1.0000 & -0.0133 \\ & & & 1.0000 \end{bmatrix}.$$

For the seasonal rainfall at Sydney the observed mean was 474 mm and the observed standard deviation and variance were respectively 187 mm and 34959 mm^2. If the seasonal rainfall is modelled with a maximum likelihood gamma distribution the model mean is 474 mm while the model standard deviation and variance are respectively 179 mm and 31879 mm^2. If the seasonal rainfall is modelled by the sum of independently distributed maximum likelihood gamma distributions then the mean is 474 and the standard deviation and variance are respectively 188 mm and 35405 mm^2. For Trial $\#(S1, 5450)$ the simulated mean was 475 and the simulated standard deviation and variance were respectively 179 mm and 32105 mm^2.

In order to investigate the likely variation in sample statistics at Sydney we conducted 20 successive trials with each trial covering the same period of 109 years as the observed rainfall. Histograms for Trial $\#(S9, 109)$ with the lowest mean $\bar{x} = 410$ mm and Trial $\#(S8, 109)$ with the highest mean $\bar{x} = 502$ mm are shown in Figure 12. We also noted that Trial $\#(S20, 109)$ had the lowest standard deviation $s = 148$ mm and corresponding variance $s^2 = 21994$ mm^2 and that Trial $\#(S5, 109)$ had the highest standard deviation $s = 216$ mm and corresponding variance $s^2 = 46481$ mm^2. The full details of these trials are shown below.

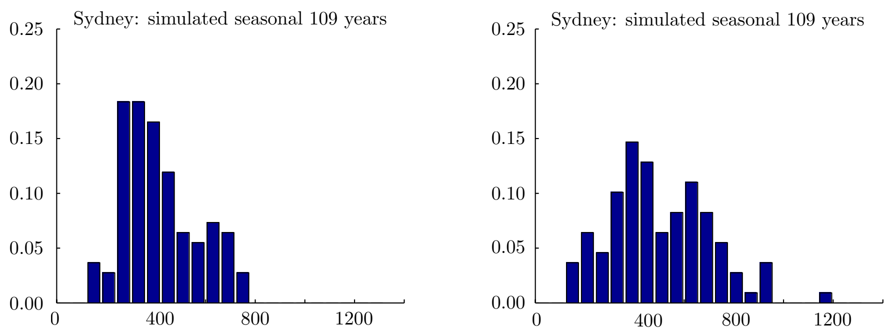


Figure 12: Selected histograms for total rainfall from 20 successive trials at BoM station 066062 Sydney (Observatory Hill) for January-February-March-April with each trial covering a period of $N = 109$ years using the copula of maximum entropy. Trial $\#(S9, 109)$ (**left**) shows the trial with lowest mean and Trial $\#(S8, 109)$ (**right**) shows the trial with highest mean.

Trial $\#(S5, 109)$ The monthly means were $\bar{x} = (109, 108, 133, 137)$ mm. The overall seasonal mean was 487 mm with standard deviation $s = 216$ mm and corresponding variance $s^2 = 46481$ mm^2. The Spearman rank correlation coefficients were

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & 0.1068 & -0.0576 & -0.1873 \\ & 1.0000 & -0.0226 & 0.0463 \\ & & 1.0000 & -0.0336 \\ & & & 1.000 \end{bmatrix}.$$

Trial $\#(S8, 109)$ The monthly means were $\bar{x} = (118, 117, 135, 131)$ mm. The overall seasonal mean was 502 mm with standard deviation $s = 202$ mm and corresponding variance $s^2 = 40603$ mm^2. The Spearman rank correlation coefficients were

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & 0.1967 & 0.0345 & -0.1905 \\ & 1.0000 - 0.0390 & -0.1193 & \\ & & 1.0000 & -0.0157 \\ & & & 1.000 \end{bmatrix}.$$

Trial $\#(S9, 109)$ The monthly means were $\bar{x} = (96, 101, 113, 100)$ mm. The overall seasonal mean was 410 mm with standard deviation $s = 150$ mm and corresponding variance $s^2 = 22394$ mm^2. The correlation coefficients were

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & 0.0667 & 0.1166 & -0.1580 \\ & 1.0000 & 0.1009 & -0.0214 \\ & & 1.0000 & -0.1633 \\ & & & 1.000 \end{bmatrix}.$$

Trial $\#(S20, 109)$ The monthly means were $\bar{x} = (115, 97, 117, 120)$ mm. The overall seasonal mean was 449 mm with standard deviation $s = 148$ mm and corresponding variance $s^2 = 21994$ mm^2. The correlation coefficients were

$$\widehat{\boldsymbol{\rho}} = \begin{bmatrix} 1.0000 & 0.0744 & -0.2276 & -0.3123 \\ & 1.0000 & -0.0898 & -0.1240 \\ & & 1.0000 & 0.1526 \\ & & & 1.000 \end{bmatrix}.$$

12. Conclusions and further work

The problem of seasonal rainfall modelling has no obvious solution because there is no standard joint distribution with marginal gamma distributions for any given set of specified grade correlation coefficients. Our model overcomes these problems. Once again we reiterate that our model is derived on a solid theoretical basis, and that the standard tests of observed data against limits set by repeated trials with the simulated data generated by the model show there is insufficient evidence to reject the model on statistical grounds. On the one hand we argue that our model is quite basic and natural in a conceptual sense—a stationary time series defined by successive independently generated rainfall totals using a joint probability distribution that replicates the observed monthly correlations. On the other hand we note that although the copula of maximum entropy may seem complicated—the defining hyper-matrix contains 1296 elements—it is nevertheless easy to compute and can be easily applied using a standard numerical calculation package such as MATLAB. There are many other models that have been proposed recently for the purpose of modelling catchment hydrology. In most cases, researchers report on the successful use of these models in the simulation of catchment rainfall. Although such models are tested extensively to ensure that the simulated data is a good approximation to the observed data, it seems there is often no clear axiomatic structural basis.

We have previously considered preliminary applications of this model to monthly and seasonal rainfall for a season with $m = 3$ months at both Kempsey [10, 18] and Sydney [22, 23] on the east coast of Australia in New South Wales. In each case we used $n = 4$ equal length subdivisions on each axis and hence were required to construct a checkerboard copula of maximum entropy defined by a 3 dimensional hyper-matrix with $4^3 = 64$ elements. In this paper we have extended the season to $m = 4$ months with $n = 6$ equal subdivisions on each axis and so the checkerboard copula is defined by a 4-dimensional hyper-matrix with $6^4 = 1296$ elements. One could probably obtain very similar results using a smaller 4-dimensional hyper-matrix with only $n = 4$ equal subdivisions on each axis giving $4^4 = 256$ elements. One reason for considering the larger number of subdivisions relates to the inequality

$$-1 + 1/n^2 \leq \rho_{r,s} \leq 1 - 1/n^2$$

for checkerboard copulas established in [22, 23]. Although these limits are not a direct problem here they could become a problem if one wished to apply this model to a problem with high pairwise correlation coefficients. We are currently looking at a model for monthly rainfall at several different locations within a small geographic region in south-eastern Australia. The grade correlation coefficients for monthly rainfall at these locations lie in the range $[0.8, 0.93]$ and so a checkerboard copula for a joint distribution in this application will certainly require $n \geq 5$. Thus the real motivation for using the larger copula was that we wanted to find out that the numerical calculations were indeed feasible.

We would like to extend these studies and the current work to model total seasonal rainfall jointly at several different locations on the east coast of Australia—perhaps at Cairns, Brisbane, Kempsey and Sydney. Although these investigations have not yet commenced we will not be surprised to find significant positive correlations in seasonal rainfall between the various locations. Thus the larger copula may become a necessity.

A final interesting and potentially more difficult question relates to a model for simultaneous generation of monthly and seasonal rainfall at several different locations in a similar geographic region—such as the east coast of Australia.. We envisage a two-tiered model with a checkerboard copula of maximum entropy at the top level to model correlations in total seasonal rainfall between different sites and a checkerboard copula of maximum entropy for each site at the bottom level—just as we have done here—to model local correlations in monthly rainfall. We would first of all use a random simulation on the checkerboard copula at the top level to generate seasonal rainfall at each site. At each site we would then use the local checkerboard copula to define a cumulative probability distribution for all possible combinations of monthly rainfall for the given seasonal total. We would then make a random selection using this distribution to find the individual monthly totals. In principle this will be no harder than the current simulation except, of course, that there will be two steps required for each outcome at each site.

13. References

[1] Peel, M. C., Finlayson, B. L., and McMahon, T. A., Updated world map of the Köppen-Geiger climate classification, *Hydrol. Earth Syst. Sci.*, **11**, 1633-1644, doi:10.5194/hess-11-1633-2007, 2007.

[2] Jaynes, E.T. Information theory and statistical mechanics, *Phys. Rev.*, **106**, 620–630, 1957a.

[3] Jaynes, E.T. Information theory and statistical mechanics ll. *Phys. Rev.* **108**, 171–190, 1957b.

[4] Von Neumann, J. *Mathematical Foundations of Quantum Mechanics*; Princeton University Press: Princeton, New Jersey, USA, 1996.

[5] Ziedler, E. *Applied Functional Analysis, Applications to Mathematical Physics*; Applied Mathematical Sciences; Springer: New York, USA, Volume 108, 1997.

[6] Shannon, C.E. A mathematical theory of communication *Bell Syst. Tech. J.*, **27**, 379–423, 623–656, 1948.

[7] David H. Bailey, Jonathan M. Borwein, Marcos López de Prado and Qiji Jim Zhu, Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance, *Notices of the AMS*, **61**, 5, 458-471, 2014.

[8] Stern, R.D.; Coe, R. A model fitting analysis of daily rainfall, *J. Roy. Stat. Soc. A*, **147**, 1–34, 1984.

[9] Srikanthan, R. *Stochastic Generation of Daily Rainfall at a Number of Sites*; Technical Report 05/7; Cooperative Research Centre for Catchment Hydrology: Melbourne, Australia, 2005.

[10] Jonathan Borwein, Phil Howlett, Julia Piantadosi, Modelling and simulation of seasonal rainfall using the principle of maximum entropy, *Entropy*, **16**, (2), 747-769, 2014.

[11] Getz, W.M. Correlative coherence analysis: Variation from intrinsic and extrinsic sources in competing populations. *Theor. Popul. Biol.*, **64**, 89–99, 2003.

[12] Hasan, M.M.; Dunn, P.K. Two Tweedie distributions that are near optimal for modelling monthly rainfall in Australia. *Int. J. Climatol.*, **31**, 1389–1397, 2011. doi:10.1002/joc.2162.

[13] Srikanthan, R.; McMahon, T.A. Stochastic generation of annual, monthly and daily climate data: A review. *Hydrol. Earth Syst. Sci.*, **5**, 633–670, 2001.

[14] Wilks, D.S.; Wilby, R.L. The weather generation game: A review of stochastic weather models. *Prog. Phys. Geogr.*, **23**, 329–357, 1999.

[15] Katz, R.W.; Parlange, M.B. Overdispersion phenomenon in stochastic modelling of precipitation. *J. Clim.*, **11**, 591–601, 1998.

[16] Rosenberg, K.; Boland, J.; Howlett, P.G. Simulation of monthly rainfall totals. *ANZIAM J.*, **46**, E85–E104, 2004.

[17] Koutsoyiannis, D. Hurst-Kolmogorov dynamics and uncertainty. *J. Am. Water Resour. Assoc.*, **47**, 481–495, 2011.

[18] Jon Borwein, Phil Howlett and Julia Piantadosi, Modelling and simulation of seasonal rainfall, In Gaertner, P., Bowden, F., Piantadosi, J. and Mobbs, K. (eds), *22nd National Conference of the Australian Society for Operations Research (ASOR 2013)*. The Australian Society for Operations Research, Adelaide, December 2013, pp. 300-306, 2013. http://www.asor.org.au/conferences/asor2013/J9/borwein.pdf

[19] Borwein, J.M.; Vanderwerff, J.D. *Convex Functions: Constructions, Characterizations and Counterexamples*; Encyclopedia of Mathematics and its Applications; Cambridge University Press: City, Country, Volume 109, 2010.

[20] Kulpa, T. On approximation of copulas. *Int. J. Math. Math. Sci.*, **22**, 259–269, 1999.

[21] Nelsen, R.B. *An Introduction to Copulas*; Lecture Notes in Statistics; Springer-Verlag: New York, NY, USA, Volume 139, 1999.

[22] Piantadosi, J.; Howlett, P.G.; Borwein, J.M. Copulas of maximum entropy. *Optim. Lett.*, **6**, 99–125, 2012. doi:10.1007/s11590-010-0254-2.

[23] Piantadosi, J.; Howlett, P.G.; Borwein, J.M.; Henstridge, J. Maximum entropy methods for generating simulated rainfall. *Numer. Algebra Control Optim.*, **2**, 233–256 , 2012. doi:10.3934/naco.2012.2.233.

[24] Borwein, J.M.; Lewis, A.S. *Convex Analysis and Nonlinear Optimization, Theory and Examples*, 2nd ed.; CMS Books in Mathematics; Springer-Verlag Inc.: New York, NY, USA, Volume 3, 2006.

[25] Centre for Computer Assisted Research Mathematics and its Applications (CARMA) website, www.carma.newcastle.edu.au/hydro.

Appendix A. The copula of maximum entropy for Cairns

The quadruply-stochastic hyper-matrix $\boldsymbol{h} \in \mathbb{R}^{6\times6\times6\times6}$ with $m = 4$ and $n = 6$ defining the quadrivariate checkerboard copula of maximum entropy for rainfall at BoM station 031036 Cairns (Kuranda) during the January-February-March-April season is given by

$$\boldsymbol{h}_{11} = [h_{11k\ell}] = \begin{bmatrix} 0.0110 & 0.0099 & 0.0087 & 0.0076 & 0.0065 & 0.0056 \\ 0.0088 & 0.0081 & 0.0074 & 0.0067 & 0.0059 & 0.0052 \\ 0.0068 & 0.0065 & 0.0061 & 0.0057 & 0.0052 & 0.0047 \\ 0.0051 & 0.0050 & 0.0049 & 0.0047 & 0.0044 & 0.0041 \\ 0.0038 & 0.0038 & 0.0038 & 0.0037 & 0.0037 & 0.0035 \\ 0.0027 & 0.0028 & 0.0029 & 0.0029 & 0.0029 & 0.0029 \end{bmatrix},$$

$$\boldsymbol{h}_{12} = [h_{12k\ell}] = \begin{bmatrix} 0.0077 & 0.0072 & 0.0066 & 0.0060 & 0.0055 & 0.0049 \\ 0.0067 & 0.0065 & 0.0062 & 0.0058 & 0.0054 & 0.0050 \\ 0.0057 & 0.0057 & 0.0056 & 0.0055 & 0.0052 & 0.0050 \\ 0.0048 & 0.0049 & 0.0050 & 0.0050 & 0.0049 & 0.0048 \\ 0.0038 & 0.0040 & 0.0042 & 0.0044 & 0.0045 & 0.0045 \\ 0.0030 & 0.0032 & 0.0035 & 0.0037 & 0.0039 & 0.0041 \end{bmatrix},$$

$$\boldsymbol{h}_{13} = [h_{13k\ell}] = \begin{bmatrix} 0.0051 & 0.0050 & 0.0049 & 0.0046 & 0.0044 & 0.0041 \\ 0.0050 & 0.0050 & 0.0050 & 0.0049 & 0.0048 & 0.0046 \\ 0.0047 & 0.0049 & 0.0050 & 0.0051 & 0.0051 & 0.0051 \\ 0.0042 & 0.0046 & 0.0049 & 0.0051 & 0.0053 & 0.0054 \\ 0.0038 & 0.0042 & 0.0046 & 0.0049 & 0.0053 & 0.0056 \\ 0.0032 & 0.0037 & 0.0042 & 0.0046 & 0.0051 & 0.0056 \end{bmatrix},$$

$$\boldsymbol{h}_{14} = [h_{14k\ell}] = \begin{bmatrix} 0.0033 & 0.0034 & 0.0034 & 0.0034 & 0.0034 & 0.0033 \\ 0.0035 & 0.0037 & 0.0039 & 0.0040 & 0.0041 & 0.0042 \\ 0.0036 & 0.0040 & 0.0043 & 0.0046 & 0.0048 & 0.0050 \\ 0.0037 & 0.0041 & 0.0046 & 0.0050 & 0.0055 & 0.0059 \\ 0.0036 & 0.0041 & 0.0047 & 0.0054 & 0.0060 & 0.0067 \\ 0.0034 & 0.0040 & 0.0048 & 0.0056 & 0.0064 & 0.0073 \end{bmatrix},$$

$$\boldsymbol{h}_{15} = [h_{15k\ell}] = \begin{bmatrix} 0.0021 & 0.0022 & 0.0024 & 0.0025 & 0.0026 & 0.0026 \\ 0.0024 & 0.0027 & 0.0029 & 0.0032 & 0.0034 & 0.0036 \\ 0.0028 & 0.0031 & 0.0036 & 0.0040 & 0.0044 & 0.0048 \\ 0.0030 & 0.0036 & 0.0042 & 0.0048 & 0.0055 & 0.0061 \\ 0.0033 & 0.0040 & 0.0048 & 0.0056 & 0.0066 & 0.0077 \\ 0.0034 & 0.0042 & 0.0053 & 0.0064 & 0.0078 & 0.0093 \end{bmatrix},$$

$$\boldsymbol{h}_{16} = [h_{16k\ell}] = \begin{bmatrix} 0.0012 & 0.0014 & 0.0016 & 0.0017 & 0.0018 & 0.0020 \\ 0.0016 & 0.0019 & 0.0021 & 0.0024 & 0.0027 & 0.0030 \\ 0.0020 & 0.0024 & 0.0028 & 0.0033 & 0.0038 & 0.0044 \\ 0.0024 & 0.0030 & 0.0037 & 0.0044 & 0.0053 & 0.0062 \\ 0.0029 & 0.0037 & 0.0046 & 0.0057 & 0.0070 & 0.0085 \\ 0.0033 & 0.0043 & 0.0056 & 0.0072 & 0.0091 & 0.0114 \end{bmatrix},$$

$$\boldsymbol{h}_{21} = [h_{21k\ell}] = \begin{bmatrix} 0.0112 & 0.0098 & 0.0084 & 0.0072 & 0.0061 & 0.0051 \\ 0.0088 & 0.0080 & 0.0071 & 0.0062 & 0.0054 & 0.0047 \\ 0.0068 & 0.0063 & 0.0058 & 0.0052 & 0.0047 & 0.0042 \\ 0.0050 & 0.0048 & 0.0046 & 0.0043 & 0.0039 & 0.0036 \\ 0.0036 & 0.0036 & 0.0035 & 0.0034 & 0.0032 & 0.0030 \\ 0.0025 & 0.0026 & 0.0026 & 0.0026 & 0.0025 & 0.0025 \end{bmatrix},$$

$$\boldsymbol{h}_{22} = [h_{22k\ell}] = \begin{bmatrix} 0.0080 & 0.0073 & 0.0066 & 0.0059 & 0.0052 & 0.0046 \\ 0.0070 & 0.0066 & 0.0061 & 0.0056 & 0.0051 & 0.0046 \\ 0.0059 & 0.0057 & 0.0055 & 0.0052 & 0.0049 & 0.0045 \\ 0.0048 & 0.0048 & 0.0048 & 0.0047 & 0.0045 & 0.0043 \\ 0.0038 & 0.0039 & 0.0040 & 0.0041 & 0.0041 & 0.0040 \\ 0.0029 & 0.0031 & 0.0033 & 0.0034 & 0.0035 & 0.0036 \end{bmatrix},$$

$$\boldsymbol{h}_{23} = [h_{23k\ell}] = \begin{bmatrix} 0.0056 & 0.0053 & 0.0050 & 0.0047 & 0.0044 & 0.0040 \\ 0.0053 & 0.0052 & 0.0051 & 0.0049 & 0.0047 & 0.0044 \\ 0.0049 & 0.0050 & 0.0051 & 0.0050 & 0.0049 & 0.0048 \\ 0.0044 & 0.0047 & 0.0048 & 0.0050 & 0.0050 & 0.0050 \\ 0.0039 & 0.0042 & 0.0045 & 0.0047 & 0.0050 & 0.0051 \\ 0.0033 & 0.0037 & 0.0040 & 0.0044 & 0.0048 & 0.0051 \end{bmatrix},$$

$$\boldsymbol{h}_{24} = [h_{24k\ell}] = \begin{bmatrix} 0.0037 & 0.0037 & 0.0037 & 0.0036 & 0.0035 & 0.0033 \\ 0.0039 & 0.0040 & 0.0041 & 0.0042 & 0.0042 & 0.0041 \\ 0.0040 & 0.0042 & 0.0045 & 0.0047 & 0.0048 & 0.0049 \\ 0.0039 & 0.0043 & 0.0047 & 0.0051 & 0.0054 & 0.0057 \\ 0.0038 & 0.0043 & 0.0048 & 0.0053 & 0.0059 & 0.0063 \\ 0.0035 & 0.0041 & 0.0048 & 0.0055 & 0.0062 & 0.0069 \end{bmatrix},$$

$$\boldsymbol{h}_{25} = [h_{25k\ell}] = \begin{bmatrix} 0.0024 & 0.0025 & 0.0026 & 0.0027 & 0.0027 & 0.0027 \\ 0.0028 & 0.0030 & 0.0032 & 0.0034 & 0.0035 & 0.0037 \\ 0.0031 & 0.0035 & 0.0038 & 0.0042 & 0.0045 & 0.0048 \\ 0.0034 & 0.0039 & 0.0044 & 0.0050 & 0.0056 & 0.0061 \\ 0.0036 & 0.0042 & 0.0050 & 0.0058 & 0.0066 & 0.0075 \\ 0.0037 & 0.0045 & 0.0054 & 0.0065 & 0.0077 & 0.0090 \end{bmatrix},$$

$$\boldsymbol{h}_{26} = [h_{26k\ell}] = \begin{bmatrix} 0.0015 & 0.0016 & 0.0018 & 0.0019 & 0.0020 & 0.0021 \\ 0.0019 & 0.0021 & 0.0024 & 0.0027 & 0.0029 & 0.0032 \\ 0.0023 & 0.0027 & 0.0032 & 0.0036 & 0.0041 & 0.0046 \\ 0.0028 & 0.0034 & 0.0040 & 0.0048 & 0.0055 & 0.0064 \\ 0.0033 & 0.0041 & 0.0050 & 0.0061 & 0.0073 & 0.0087 \\ 0.0037 & 0.0047 & 0.0060 & 0.0075 & 0.0093 & 0.0114 \end{bmatrix},$$

$$\boldsymbol{h}_{31} = [h_{31k\ell}] = \begin{bmatrix} 0.0113 & 0.0097 & 0.0082 & 0.0068 & 0.0056 & 0.0046 \\ 0.0088 & 0.0078 & 0.0068 & 0.0058 & 0.0049 & 0.0041 \\ 0.0067 & 0.0061 & 0.0054 & 0.0048 & 0.0042 & 0.0037 \\ 0.0049 & 0.0046 & 0.0042 & 0.0039 & 0.0035 & 0.0031 \\ 0.0035 & 0.0034 & 0.0032 & 0.0030 & 0.0028 & 0.0026 \\ 0.0024 & 0.0024 & 0.0023 & 0.0023 & 0.0022 & 0.0021 \end{bmatrix},$$

$$\boldsymbol{h}_{32} = [h_{32k\ell}] = \begin{bmatrix} 0.0084 & 0.0075 & 0.0066 & 0.0058 & 0.0050 & 0.0043 \\ 0.0072 & 0.0066 & 0.0060 & 0.0054 & 0.0048 & 0.0043 \\ 0.0060 & 0.0057 & 0.0053 & 0.0050 & 0.0046 & 0.0041 \\ 0.0048 & 0.0047 & 0.0046 & 0.0044 & 0.0042 & 0.0039 \\ 0.0038 & 0.0038 & 0.0038 & 0.0038 & 0.0037 & 0.0036 \\ 0.0029 & 0.0030 & 0.0031 & 0.0031 & 0.0032 & 0.0032 \end{bmatrix},$$

$$\boldsymbol{h}_{33} = [h_{33k\ell}] = \begin{bmatrix} 0.0060 & 0.0056 & 0.0052 & 0.0047 & 0.0043 & 0.0038 \\ 0.0057 & 0.0055 & 0.0052 & 0.0049 & 0.0046 & 0.0042 \\ 0.0052 & 0.0052 & 0.0051 & 0.0049 & 0.0047 & 0.0045 \\ 0.0046 & 0.0047 & 0.0048 & 0.0048 & 0.0048 & 0.0047 \\ 0.0040 & 0.0042 & 0.0044 & 0.0045 & 0.0047 & 0.0047 \\ 0.0033 & 0.0036 & 0.0039 & 0.0042 & 0.0044 & 0.0046 \end{bmatrix},$$

$$\boldsymbol{h}_{34} = [h_{34k\ell}] = \begin{bmatrix} 0.0041 & 0.0041 & 0.0039 & 0.0038 & 0.0036 & 0.0033 \\ 0.0043 & 0.0043 & 0.0043 & 0.0043 & 0.0042 & 0.0040 \\ 0.0043 & 0.0045 & 0.0047 & 0.0047 & 0.0048 & 0.0047 \\ 0.0042 & 0.0045 & 0.0048 & 0.0051 & 0.0053 & 0.0054 \\ 0.0040 & 0.0044 & 0.0049 & 0.0053 & 0.0057 & 0.0060 \\ 0.0037 & 0.0042 & 0.0048 & 0.0053 & 0.0059 & 0.0064 \end{bmatrix},$$

$$\boldsymbol{h}_{35} = [h_{35k\ell}] = \begin{bmatrix} 0.0027 & 0.0028 & 0.0029 & 0.0029 & 0.0028 & 0.0028 \\ 0.0031 & 0.0033 & 0.0035 & 0.0036 & 0.0037 & 0.0037 \\ 0.0035 & 0.0038 & 0.0041 & 0.0044 & 0.0046 & 0.0048 \\ 0.0037 & 0.0042 & 0.0047 & 0.0052 & 0.0056 & 0.0061 \\ 0.0039 & 0.0045 & 0.0052 & 0.0059 & 0.0067 & 0.0074 \\ 0.0040 & 0.0047 & 0.0056 & 0.0066 & 0.0076 & 0.0087 \end{bmatrix},$$

$$\boldsymbol{h}_{36} = [h_{36k\ell}] = \begin{bmatrix} 0.0018 & 0.0019 & 0.0020 & 0.0021 & 0.0022 & 0.0023 \\ 0.0022 & 0.0025 & 0.0027 & 0.0029 & 0.0031 & 0.0033 \\ 0.0027 & 0.0031 & 0.0035 & 0.0039 & 0.0043 & 0.0047 \\ 0.0032 & 0.0038 & 0.0044 & 0.0051 & 0.0058 & 0.0065 \\ 0.0037 & 0.0045 & 0.0054 & 0.0064 & 0.0075 & 0.0088 \\ 0.0041 & 0.0052 & 0.0064 & 0.0078 & 0.0095 & 0.0114 \end{bmatrix}.$$

The remaining elements $\boldsymbol{h}_{41}, \ldots, \boldsymbol{h}_{66}$ can be found from the formula

$$h_{ijk\ell} = h_{pqrs}$$

where $i + p = j + q = k + r = \ell + s = 7$.