

# A method for estimating the potential long-range predictability of precipitation over Western Australia

C. S. Frederiksen<sup>1</sup>S. Grainger<sup>2</sup>X. Zheng<sup>3</sup>

(Received 11 August 2008; revised 9 December 2008)

## Abstract

A methodology for estimating the potential predictability of seasonal mean climate variables where the daily data consist of dichotomous (on/off) events, such as precipitation, is described and applied to Western Australian rainfall for the period 1951–2000. The method relies on determining an estimate for the intermonth correlations for precipitation and utilizes a stochastic, two state, first order, Markov chain model fitted to the daily data.

## Contents

### 1 Introduction

C584

---

<http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/1411>  
gives this article, © Austral. Mathematical Soc. 2008. Published December 17, 2008.  
ISSN 1446-8735. (Print two pages per sheet of paper.)

<i>1</i>	<i>Introduction</i>	C584
<b>2</b>	<b>Methodology</b>	<b>C585</b>
2.1	Estimating the intraseasonal component of variability . . .	C585
2.2	Estimating intermonth correlations . . . . .	C588
2.3	Potential predictability . . . . .	C592
<b>3</b>	<b>Potential predictability of West Australian rainfall</b>	<b>C592</b>
<b>4</b>	<b>Conclusions</b>	<b>C595</b>
	<b>References</b>	<b>C596</b>

# 1 Introduction

Zheng et al. [9] proposed a methodology to estimate the potential predictability of seasonal mean climate data. This method uses monthly data within the season to estimate the amount of the total interannual variance in the seasonal mean due to chaotic processes, acting on a subseasonal time scale, and processes acting on a much longer time scale (interannual). It is therefore convenient to conceptualise the seasonal mean data as consisting of a slow component, which is potentially predictable, and a chaotic component, largely related to intraseasonal processes, which is unpredictable and referred to as the intraseasonal component [2]. The potential predictability of the climate variable is defined as the proportion of the total variance not associated with the intraseasonal component.

However, the methodology assumes stationarity in the monthly statistics based on the assumption that the underlying daily time series are approximately first order autoregressive (AR1). Such assumptions do not in general hold for climate data such as precipitation when the daily data consist of dichotomous (on/off) events. Madden et al. [6] estimated the predictability of New Zealand rainfall using a generalized chain dependent process to model the daily precipitation data. Zheng and Frederiksen [8] derived predictable

patterns of New Zealand rainfall by ranking the monthly precipitation into a finite number of categories. Here we present a new methodology that extends the methods of Madden et al. [6] and Zheng et al. [9] to allow for dichotomous events. It relies on using daily data to estimate the intermonth correlation between monthly means of the rainfall within a season. In this article we use this method to estimate the potential predictability of Western Australian seasonal rainfall.

## 2 Methodology

### 2.1 Estimating the intraseasonal component of variability

A crucial step, in the estimation of the long range potential predictability of seasonal mean climate data, is to estimate the interannual variability of the climate ‘noise’, or the intraseasonal component of interannual variability [2, 9]. For this purpose, a useful statistical model, using monthly data, is [2, 9]

$$x_{ym} = \mu_y + \epsilon_{ym}, \quad (1)$$

where  $x_{ym}$  represent monthly anomalies of a climate variable, with the annual cycle removed, in year  $y$  ( $y = 1, \dots, Y$ ) and month  $m$  ( $m = 1, 2, 3$ ) of the season;  $\mu_y$  is the seasonal population mean in year  $y$  which is associated with very slowly varying external forcing or internal dynamics and is therefore potentially predictable;  $\epsilon_{ym}$  is a monthly departure representing the intraseasonal, or essentially unpredictable, component. The vector  $\{\epsilon_{y1}, \epsilon_{y2}, \epsilon_{y3}\}$  is assumed to comprise a stationary and independent random vector with respect to year [5].

Equation (1) implies that month-to-month fluctuations, or intraseasonal variability, arise entirely from this component ( $x_{y1} - x_{y2} = \epsilon_{y1} - \epsilon_{y2}$  for

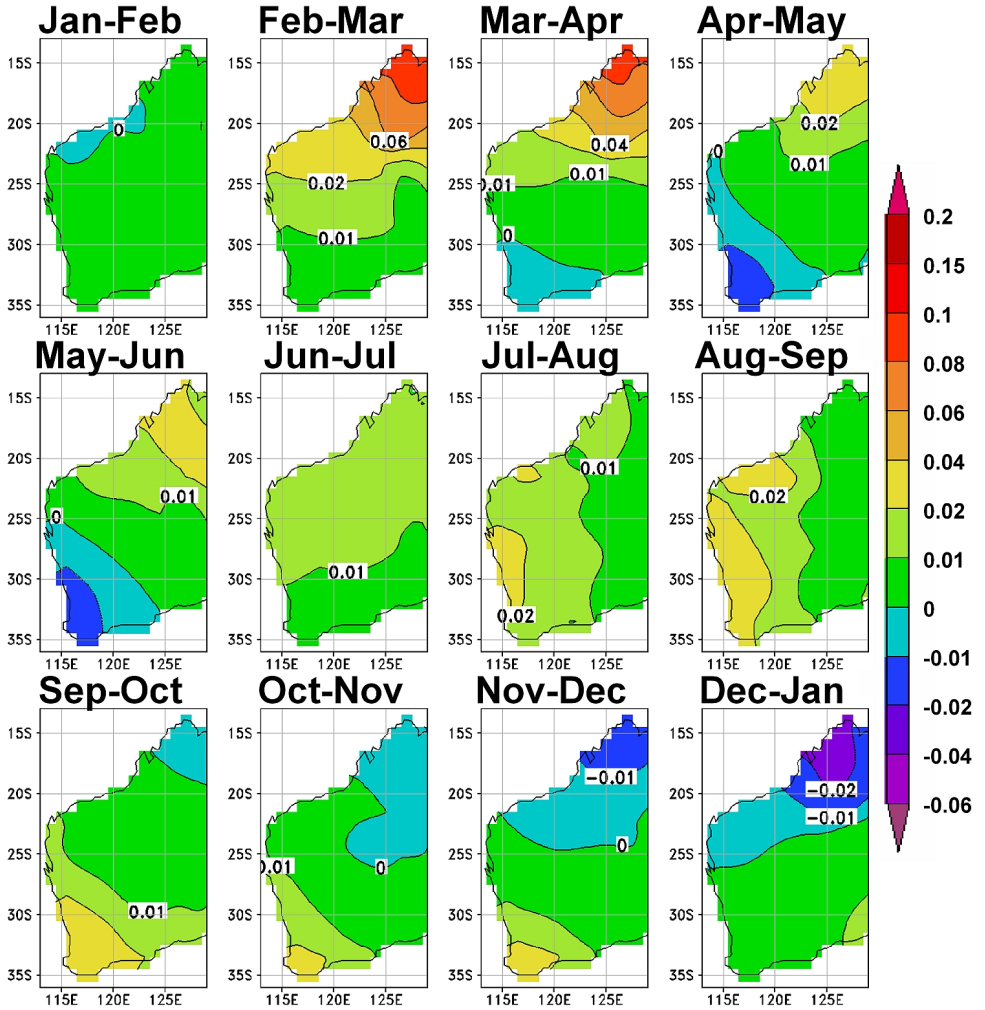


FIGURE 1: Intermont correlation  $\hat{\phi}_{12}$  for all months 1951–2000.

example). We use the convention that an average over any index is represented by a circle subscript. For example,  $x_{y\circ}$  is an average over all months in the season, and  $x_{\circ\circ}$  is an average over all months and years. The symbol  $V$  denotes the interannual variance for a single variable, for example  $V(x_{ym})$ , or the intermonth covariance for two variables, for example  $V(x_{ym}, x_{yn})$ . For the intraseasonal component,  $\sigma_m^2$  and  $\phi_{mn}$  represent the interannual variance in month  $m$  and the intermonth correlation between months  $m$  and  $n$ , respectively. Thus

$$\phi_{mn} = \frac{V(\epsilon_{ym}, \epsilon_{yn})}{\sigma_m \sigma_n}. \tag{2}$$

The variance of the seasonal mean of the intraseasonal component is then

$$\begin{aligned} V(\epsilon_{y\circ}) &= E \left[ \left( \frac{1}{3} \sum_{m=1}^3 \epsilon_{ym} \right)^2 \right] \\ &= \frac{1}{9} (\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + 2\sigma_1\sigma_2\phi_{12} + 2\sigma_2\sigma_3\phi_{23} + 2\sigma_1\sigma_3\phi_{13}), \end{aligned} \tag{3}$$

where  $E$  is the expectation taken over all years. Estimations of the second moments of the intermonth differences in climate variable  $x$  provide three equations relating the six variables in equation (3)

$$E \left[ (x_{y1} - x_{y2})^2 \right] \approx \frac{1}{Y} \sum_{y=1}^Y (x_{y1} - x_{y2})^2 \equiv a. \tag{4}$$

But, equation (1) implies that

$$E \left[ (x_{y1} - x_{y2})^2 \right] = E \left[ (\epsilon_{y1} - \epsilon_{y2})^2 \right] = \sigma_1^2 - 2\sigma_1\sigma_2\phi_{12} + \sigma_2^2 = a. \tag{5}$$

In a similar manner, it follows that

$$\sigma_1\sigma_2\phi_{12} - \sigma_1\sigma_3\phi_{13} - \sigma_2^2 + \sigma_2\sigma_3\phi_{23} = b, \tag{6}$$

$$\sigma_2^2 - 2\sigma_2\sigma_3\phi_{23} + \sigma_3^2 = c, \tag{7}$$

where

$$b \equiv \frac{1}{Y} \sum_{y=1}^Y (x_{y1} - x_{y2})(x_{y2} - x_{y3}), \quad (8)$$

$$c \equiv \frac{1}{Y} \sum_{y=1}^Y (x_{y2} - x_{y3})^2. \quad (9)$$

Earlier studies using this approach assumed that the monthly statistics of the climate variable are stationary and the underlying daily time series is an AR1 process [2, 9]. This provided an additional three equations  $\sigma_1 = \sigma_2 = \sigma_3$ ,  $\phi_{12} = \phi_{23}$  and  $\phi_{13} = 0$ , and allowed  $V(\epsilon_{y0})$  to be estimated from equation (3). However, these assumptions do not hold for precipitation which consists of dichotomous (on/off) events. In the next section we show how the three intermonth correlations  $\phi_{12}$ ,  $\phi_{23}$  and  $\phi_{13}$  are estimated to provide another three equations.

## 2.2 Estimating intermonth correlations

Here, we use a stochastic, two state, first order, Markov chain model [7] fitted to daily precipitation data to estimate the intermonth correlation between monthly means of the precipitation during a season.

Let  $\{J_{m,t} \mid t = 1, \dots, T_m\}$  denote the sequence of daily precipitation occurrences in month  $m$  with  $T_m$  days such that  $J_{m,t} = 1$  indicates a ‘wet day’, defined here as a day with at least 1 mm of rain, and  $J_{m,t} = 0$  a ‘dry day’. A first-order Markov chain model is completely characterized by the transition probability

$$P_{m,jk} = \Pr(J_{m,(t+1)} = k \mid J_{m,t} = j), \quad (10)$$

where  $j, k = 0, 1$ . The transition probabilities are estimated for each month using daily data over all years, and hence we construct a one step transition

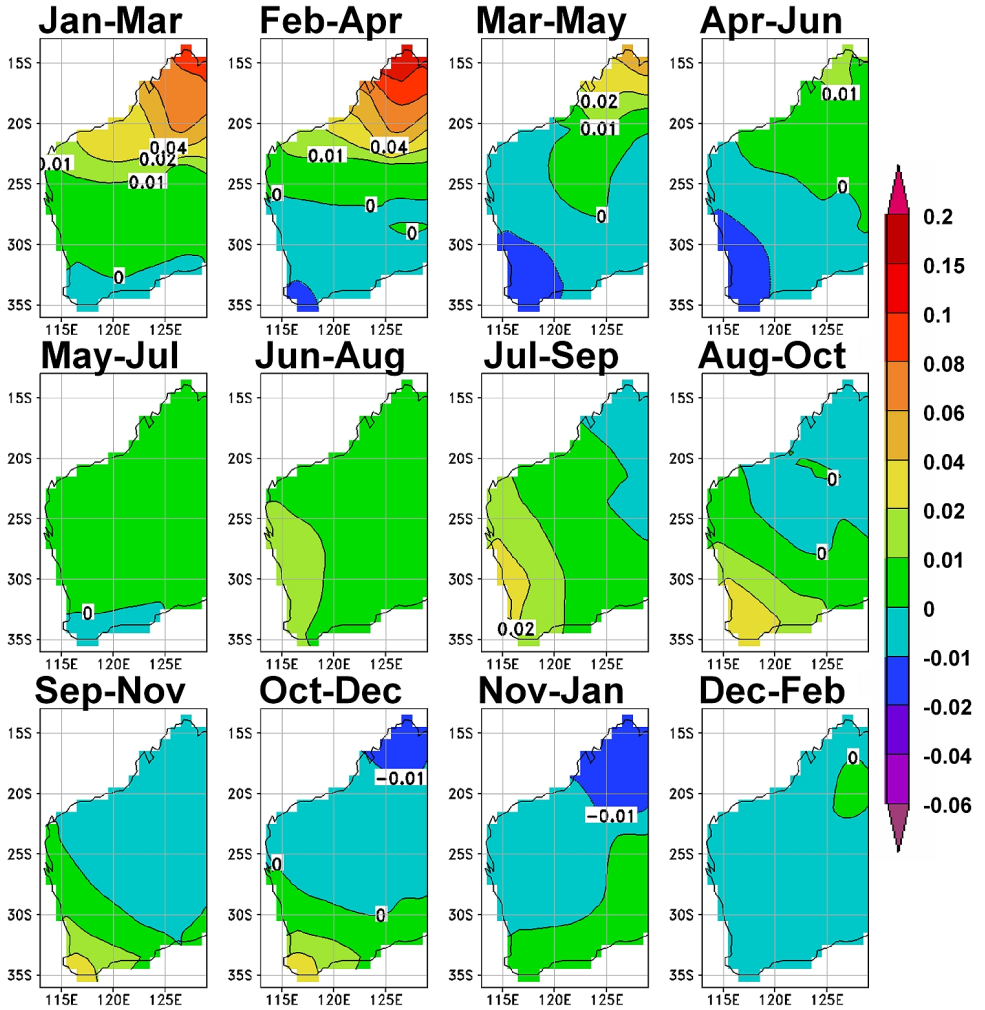


FIGURE 2: Intermont correlation  $\hat{\phi}_{13}$  for all months 1951–2000.

matrix  $\mathbf{P}_m$  for month  $m$

$$\mathbf{P}_m = \begin{pmatrix} P_{m,00} & P_{m,01} \\ P_{m,10} & P_{m,11} \end{pmatrix} = \begin{pmatrix} 1 - p_m & p_m \\ \alpha_m & 1 - \alpha_m \end{pmatrix}. \quad (11)$$

Then, the  $t$ -step transition probability is simply  $\mathbf{P}_m^{(t)} = \mathbf{P}_m^{(t-1)}\mathbf{P}_m$ , and Feller [1] showed that

$$\mathbf{P}_m^{(t)} = \frac{1}{\alpha_m + p_m} \begin{pmatrix} \alpha_m & p_m \\ \alpha_m & p_m \end{pmatrix} + \frac{(1 - \alpha_m - p_m)^t}{\alpha_m + p_m} \begin{pmatrix} p_m & -p_m \\ -\alpha_m & \alpha_m \end{pmatrix}. \quad (12)$$

Suppose now that  $\{R_{m,t} \mid t = 1, \dots, T_m\}$  is the time series of daily precipitation amounts in month  $m$ , and the ‘intensity’ of rainfall on wet days ( $R_{m,t} > 0$  or  $J_{m,t} = 1$ ) are taken to be conditionally independent and identically distributed with respect to time  $t$  [4]. Then, for any *wet* day in month  $m$ , the first and second moments of the intensity are estimated from daily data as

$$\eta_m \equiv E[R_{m,t} \mid J_{m,t} = 1], \quad (13)$$

$$\nu_m \equiv E[R_{m,t}^2 \mid J_{m,t} = 1]. \quad (14)$$

Using these estimates for  $P_{m,jk}$ ,  $\eta_m$  and  $\nu_m$ , it is possible to estimate the stationary probability of rainfall  $\pi_m$  [7], the expected rainfall amount and the expected second moment on any day in month  $m$  respectively, by

$$\pi_m \equiv \Pr(J_{m,t} = 1) = \frac{P_{m,01}}{(1 - P_{m,11} + P_{m,01})} = \frac{p_m}{\alpha_m + p_m}, \quad (15)$$

$$E[R_{m,t}] = E[R_{m,t} \mid J_{m,t} = 1] \Pr(J_{m,t} = 1) = \eta_m \pi_m, \quad (16)$$

$$E[R_{m,t}^2] = E[R_{m,t}^2 \mid J_{m,t} = 1] \Pr(J_{m,t} = 1) = \nu_m \pi_m. \quad (17)$$

By definition, the intermonth correlation between the monthly mean rain-



fall in month  $m$ , that is  $\frac{1}{T_m} \sum_{t=1}^{T_m} R_{m,t}$ , and  $n$ , that is  $\frac{1}{T_n} \sum_{\tau=1}^{T_n} R_{n,\tau}$ , is

$$\phi_{mn} \equiv \frac{E \left[ \left( \frac{1}{T_m} \sum_{t=1}^{T_m} (R_{m,t} - E[R_{m,t}]) \right) \left( \frac{1}{T_n} \sum_{\tau=1}^{T_n} (R_{n,\tau} - E[R_{n,\tau}]) \right) \right]}{\sqrt{E \left[ \left( \frac{1}{T_m} \sum_{t=1}^{T_m} R_{m,t} - E[R_{m,t}] \right)^2 \right] E \left[ \left( \frac{1}{T_n} \sum_{\tau=1}^{T_n} R_{n,\tau} - E[R_{n,\tau}] \right)^2 \right]}}. \tag{18}$$

In order to estimate equation (18), we need first to estimate  $E[R_{m,t}R_{n,\tau}]$  by

$$\begin{aligned} E[R_{m,t}R_{n,\tau}] &= E[R_{n,\tau}, J_{n,\tau} = 1 \mid J_{m,t} = 1]E[R_{m,t} \mid J_{m,t} = 1] \Pr(J_{m,t} = 1) \\ &= \eta_m \pi_m \Pr(J_{n,\tau} = 1 \mid J_{m,t} = 1) \eta_n, \end{aligned} \tag{19}$$

where, using the [2, 2] element of the  $t$ -step transition matrices in each month

$$\Pr(J_{n,\tau} = 1 \mid J_{m,t} = 1) = \begin{cases} \mathbf{P}_m^{(\tau-t)} [2, 2], & m = n = 1, 2, 3; \tau > t, \\ \mathbf{P}_m^{(T_m-t)} \mathbf{P}_n^{(\tau)} [2, 2], & n = m + 1, \\ \mathbf{P}_1^{(T_1-t)} \mathbf{P}_2^{(T_2)} \mathbf{P}_3^{(\tau)} [2, 2], & m = 1, n = 3. \end{cases} \tag{20}$$

Hence, the numerator in equation (18) is

$$\begin{aligned} &E \left[ \left( \frac{1}{T_m} \sum_{t=1}^{T_m} (R_{m,t} - E[R_{m,t}]) \right) \left( \frac{1}{T_n} \sum_{\tau=1}^{T_n} (R_{n,\tau} - E[R_{n,\tau}]) \right) \right] \\ &= \frac{1}{T_m T_n} \sum_{t=1}^{T_m} \sum_{\tau=1}^{T_n} E[R_{m,t}R_{n,\tau}] - \left( \frac{1}{T_m} \sum_{t=1}^{T_m} E[R_{m,t}] \right) \left( \frac{1}{T_n} \sum_{\tau=1}^{T_n} E[R_{n,\tau}] \right) \\ &= \eta_m \eta_n \left\{ \frac{\pi_m}{T_m T_n} \sum_{t=1}^{T_m} \sum_{\tau=1}^{T_n} \Pr(J_{n,\tau} = 1 \mid J_{m,t} = 1) - \pi_m \pi_n \right\}. \end{aligned} \tag{21}$$

Similarly, the two terms in the denominator are

$$E \left[ \left( \frac{1}{T_m} \sum_{t=1}^{T_m} (R_{m,t} - E[R_{m,t}]) \right)^2 \right]$$

$$= \frac{2\eta_m^2 \pi_m}{T_m^2} \sum_{t=1}^{T_m-1} \sum_{\tau=t+1}^{T_m} \text{Pr}(J_{m,\tau} = 1 \mid J_{m,t} = 1) + \frac{\pi_m \mathcal{V}_m}{T_m} - \pi_m^2 \eta_m^2. \quad (22)$$

Using equations (21), (22) in equation (18), we estimate  $\phi_{12}$ ,  $\phi_{23}$  and  $\phi_{13}$  and use these as our estimates of the intermonth correlations for the intraseasonal component of the precipitation. With these estimates, equations (5), (6) and (7) are solved numerically for  $\sigma_1$ ,  $\sigma_2$  and  $\sigma_3$ .

### 2.3 Potential predictability

The total interannual variability of the seasonal mean of a climate variable is estimated by the sample variance

$$\hat{V}(x_{y_o}) = \frac{1}{Y-1} \sum_{y=1}^Y [x_{y_o} - x_{o_o}]^2, \quad (23)$$

where the caret ‘ $\hat{\phantom{x}}$ ’ denotes an estimate. A commonly used definition of potential predictability [2, 9, e.g.] is the fraction of total interannual variability of the seasonal mean that is not due to the intraseasonal variance

$$\text{Potential Predictability} \equiv 1 - \frac{\hat{V}(\epsilon_{y_o})}{\hat{V}(x_{y_o})}, \quad (24)$$

which is calculated using equations (3) and (23).

## 3 Potential predictability of West Australian rainfall

Here, we illustrate the methodology by using it to estimate the potential predictability of West Australian (WA) rainfall using monthly mean and daily

data for the period November 1950 to December 2000, taken from the Australian Bureau of Meteorology National Climate Centre gridded historical dataset [3]. The data have been interpolated from the original  $0.25^\circ \times 0.25^\circ$  latitude/longitude grid to a  $1^\circ \times 1^\circ$  grid. The daily data was used to estimate  $\phi_{12}$ ,  $\phi_{23}$  and  $\phi_{13}$ .

Figures 1 and 2 show the estimated  $\hat{\phi}_{12}$  and  $\hat{\phi}_{13}$ , respectively, for each calendar month. Both figures show a very distinct annual cycle with largest correlations over the north during the summer and autumn months. During these months,  $\hat{\phi}_{12}$  and  $\hat{\phi}_{13}$  can *both* have values as large as 0.1. This contrasts with the case for a climate variable (for example, surface temperature, or the geopotential height) that is modelled as an AR1 process, for which the constraints  $0 \leq \phi_{12} \leq 0.1$  and  $\phi_{13} = 0$  apply [9]. This further illustrates why these AR1 assumptions are not applied to precipitation. For the remaining months, the magnitudes of  $\hat{\phi}_{12}$  and  $\hat{\phi}_{13}$  rarely exceed 0.02. Over southern WA, the magnitudes of both correlations rarely exceed 0.02 in all months.

Figure 3 shows the estimated potential predictability of WA precipitation for each three month season, denoted here by the first letter of each month (for example December–January–February is DJF). Again, there is a distinct annual cycle in the pattern of potential predictability. From NDJ to AMJ, the potential predictability is very low north of  $20^\circ\text{S}$ , indicating that intraseasonal processes dominate. South of  $20^\circ\text{S}$ , during these seasons, the potential predictability ranges between 20% to 70%, except for the far southwest of the state, where the potential predictability is generally much less than 20%. For the remaining seasons MJJ to OND, there is quite substantial potential predictability over the north and central regions of the state, especially near the eastern border of the state. Potential predictability over the southwest of the state, where most of the population and agriculture is located, is still very low, except during ASO and SON. This indicates that chaotic weather phenomena is mainly responsible for the seasonal rainfall in this area.

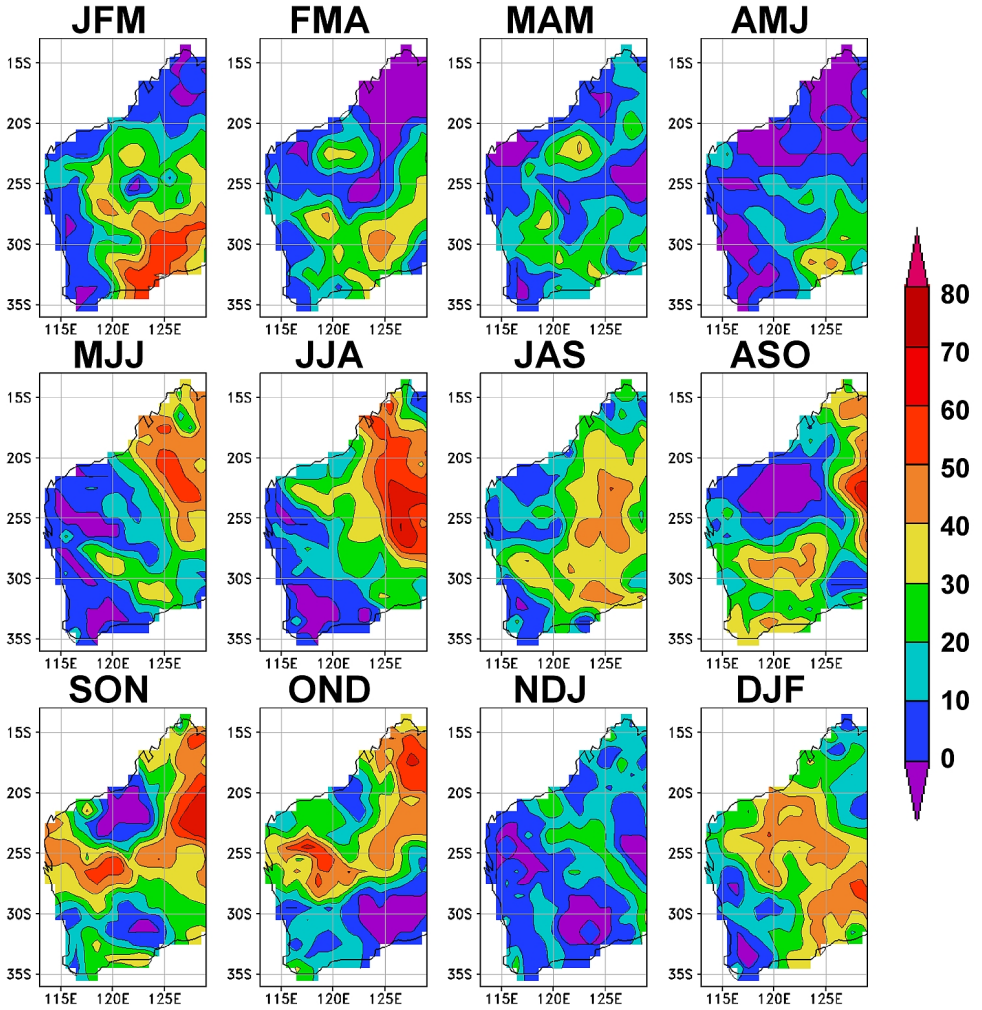


FIGURE 3: Potential predictability of Western Australian rainfall 1951–2000 in all three month seasons expressed as a percentage.

## 4 Conclusions

We have outlined a procedure for estimating the potential predictability of seasonal mean climate data where the underlying daily processes are characterized as dichotomous events, as is the case with precipitation. The proposed methodology is an extension of a recently proposed method suitable for monthly climate variables that are approximated by AR1 ‘red noise’ processes [2, 9]. The generalised formalism expressed by equations (5), (6) and (7) encompasses this earlier method.

In the more general case, estimates of the intermonth correlations allow a solution for the monthly variances and consequently the component of the total variance that is due to intraseasonal processes, and finally an estimate of the potential predictability. For rainfall events, a stochastic two state, first order, Markov chain model fitted to daily precipitation data allows these correlations to be estimated.

Using this methodology, we constructed maps of potential predictability for WA rainfall in all possible three month seasons. The potential predictability shows a distinct annual cycle with generally

1. high potential predictability over the north and centre of the state during MJJ to OND, especially on the eastern border,
2. high potential predictability in the southeast of the state during DJF to FMA, and
3. very little potential predictability in the southwest of the state except during ASO and SON when there is moderate potential predictability.

**Acknowledgements** Monthly mean and daily rainfall data for Western Australia was provided by the National Climate Centre of the Australian Bureau of Meteorology. SG is supported by the Australian Climate Change

Science Program of the Australian Department of Climate Change. xz is supported by the New Zealand Foundation for Research, Science and Technology (contract c01Xx0701). CSF acknowledges the support of the West Australian Department of Environment and Conservation under the Indian Ocean Climate Initiative Stage 3.

## References

- [1] W. Feller. *An Introduction to Probability Theory and Its Applications*. Vol. 2. John Wiley and Sons, 626pp, 1966. [C590](#)
- [2] C. S. Frederiksen and X. Zheng. Coherent Structures of Interannual Variability of the Atmospheric Circulation: The Role of Intraseasonal Variability. *Frontiers in Turbulence and Coherent Structures, World Scientific Lecture Notes in Complex Systems*, Vol. 6, Eds Jim Denier and Jorgen Frederiksen, World Scientific Publications, 87–120, 2007. [C584](#), [C585](#), [C588](#), [C592](#), [C595](#)
- [3] D. A. Jones and G. Weymouth. An Australian monthly rainfall data set. *Technical Report No. 70*, Bur. Met. Australia, 1997. [C593](#)
- [4] R. W. Katz and X. Zheng. Mixture Model For Overdispersion of Precipitation. *J. Climate*, 12:2528–2537, [doi:10.1175/1520-0442\(1999\)012;2528:MMFOOP;2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012;2528:MMFOOP;2.0.CO;2), 1999. [C590](#)
- [5] E. N. Lorenz. On the existence of extended range predictability. *J. Appl. Meteor.*, 12:543–546, [doi:10.1175/1520-0450\(1973\)012;0543:OTEOER;2.0.CO;2](https://doi.org/10.1175/1520-0450(1973)012;0543:OTEOER;2.0.CO;2), 1973. [C585](#)
- [6] R. A. Madden, D. J. Shea, R. W. Katz and J. W. Kidson. The potential long-range predictability of precipitation over New Zealand. *Int. J. Clim.*, 19:405–421, <http://ejournals.ebsco.com/direct.asp?ArticleID=UWLQYU54E2X7BYBW3DT>, 1999. [C584](#), [C585](#)

- [7] D. S. Wilks. Statistical Methods in the Atmospheric Sciences. (second edition). Academic Press, 627pp, 2006. [C588](#), [C590](#)
- [8] X. Zheng and C. S. Frederiksen. A study of predictable patterns for seasonal forecasting of New Zealand rainfall. *J. Climate*, 19:3320–3333, [doi:10.1175/JCLI3798.1](#), 2006. [C584](#)
- [9] X. Zheng, H. Nakamura and J. A. Renwick. Potential predictability of seasonal means based on monthly time series of meteorological variables. *J. Climate*, 13:2591–2604, [doi:10.1175/1520-0442\(2000\)013;2591:PPOSMB;2.0.CO;2](#), 2000. [C584](#), [C585](#), [C588](#), [C592](#), [C593](#), [C595](#)

## Author addresses

1. **C. S. Frederiksen**, Centre for Australian Weather and Atmospheric Research, Bureau of Meteorology, GPO Box 1289, Melbourne, Victoria 3001, AUSTRALIA.  
<mailto:c.frederiksen@bom.gov.au>
2. **S. Grainger**, Centre for Australian Weather and Atmospheric Research, Bureau of Meteorology, GPO Box 1289, Melbourne, Victoria 3001, AUSTRALIA.
3. **X. Zheng**, National Institute of Water and Atmospheric Research, Wellington, NEW ZEALAND