PUBLISHED VERSION

Thyer, Mark Andrew; S˘ &˙^¦æÑÕ^[¦*^ÁŒ¦¬^å

Modeling long-term persistence in hydroclimatic time series using a hidden state Markov model

http://hdl.handle.net/2440/81999

Modeling long-term persistence in hydroclimatic time series using a hidden state Markov model

Mark Thyer and George Kuczera

Department of Civil, Surveying and Environmental Engineering, University of Newcastle
Callaghan, New South Wales, Australia

Abstract. A hidden state Markov (HSM) model is developed as a new approach for generating hydroclimatic time series with long-term persistence. The two-state HSM model is motivated by the fact that the interaction of global climatic mechanisms produces alternating wet and dry regimes in Australian hydroclimatic time series. The HSM model provides an explicit mechanism to stochastically simulate these quasi-cyclic wet and dry periods. This is conceptually sounder than the current stochastic models used for hydroclimatic time series simulation. Models such as the lag-one autoregressive (AR(1)) model have no explicit mechanism for simulating the wet and dry regimes. In this study the HSM model was calibrated to four long-term Australian hydroclimatic data sets. A Markov Chain Monte Carlo method known as the Gibbs sampler was used for model calibration. The results showed that the locations significantly influenced by tropical weather systems supported the assumptions of the HSM modeling framework and indicated a strong persistence structure. In contrast, the calibration of the AR(1) model to these data sets produced no statistically significant evidence of persistence.

1. Introduction

The Australian climatic regime is influenced by numerous global climate circulations. These circulations produce the high variability and persistence that are a common feature of Australian hydroclimatic data. The challenge is to develop stochastic models that are able to reproduce this long-term persistence. The aim of this work is to present an alternative and conceptually sounder stochastic model for simulating long-term persistence in hydroclimatic time series.

There are already a number of stochastic models available for hydroclimatic time series simulation. The autoregressive (AR) and autoregressive moving average (ARMA) models are presented by *Salas* [1993] as being able to accommodate most typical cases. The lag-one autoregressive model (AR(1)) is commonly used in the Australian water supply industry to simulate annual rainfall time series [*Grayson et al.*, 1996]. These and other models such as the fractional Gaussian noise and broken line models were developed to reproduce certain statistical aspects of a time series. In contrast, this new approach considers the influence of the global climatic circulations on hydroclimatic time series.

Previous studies have identified several global climatic mechanisms that impact on the Australian climatic regime at annual and decadal timescales. The most well known is the El Niño–Southern Oscillation (ENSO), which has a quasi-cyclic occurrence [*Allan et al.*, 1996]. ENSO's relationship with Australian rainfall is significant [*Allan*, 1988; *Lough*, 1991; *Nicholls and Kariko*, 1993] but changes over time [*Kane*, 1997; *Nicholls and Kariko*, 1993]. The ENSO influence is also modulated by an interdecadal oscillation in the sea surface temperature (SST) in the central Pacific [*Power et al.*, 1999], referred to as the IPO. In addition, *Smith* [1994] showed that there are strong

correlations between Indian Ocean SST and Australian winter rainfall. In the Southern Ocean there exists the Antarctic circumpolar wave [*White and Peterson*, 1996] that describes the rotation of alternating warmer and cooler regions around the South Pole approximately every 10 years.

The interaction of these circulations is complex. As their periodicities vary, their influence could be negated or enhanced by each other at any point in time. It is believed that the cumulative effect of these circulations produces a quasi-cyclic forcing mechanism in the hydrological cycle. *Rodriguez-Iturbe et al.* [1991] modeled the land-atmosphere interaction with quasi-cyclic forcing and found that soil moisture persisted around two stable modes: wet and dry. Similarly, wet and dry cycles have been noted in many Australian rainfall time series [*Srikanthan and Stewart*, 1992]. An example is the Sydney metropolitan annual rainfall [*Linacre and Geerts*, 1997], shown in Figure 1. The rainfall time series shows that there are several periods when the annual rainfall is persistently below the long-term mean. In the cumulative residual time series a positive slope indicates an above-average (wet) rainfall period, and a negative slope indicates a below-average (dry) period. The varying occurrence and persistence of the wet and dry periods are clearly visible. They are considered to be a result of the nonlinear climate dynamics produced by the varying climatic mechanisms which influence Australian rainfall.

The hidden state Markov (HSM) model is introduced as an alternative model for generating hydroclimatic time series with long-term persistence. The model framework provides an explicit mechanism to simulate the varying wet and dry regimes. Thus the HSM model has the ability to emulate the influence that the various climatic processes have on Australian hydroclimatic data.

Previously, Markov chain models have been applied to model daily rainfall processes (*Salas* [1993] has an extensive list of references). *Zucchini and Guttorp* [1991] used a hidden state Markov model to simulate the occurrence pattern of daily rainfall at multiple sites. An alternative approach is the non-
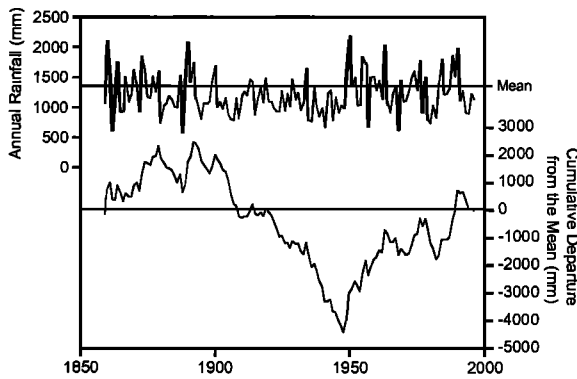
Figure 1. Sydney annual (January to December) rainfall and cumulative residual time series (1859–1996).

homogenous hidden Markov model [*Hughes and Guttorp*, 1994; *Hughes et al.*, 1999] that relates atmospheric variables to local daily precipitation processes. This framework has its limitations for long-term water resources applications. Long-term simulations of hydroclimatic inputs are needed to evaluate water supply system performance. In the model of Hughes and Guttorp and Hughes et al. this would require long-term simulations of the atmospheric variables used to classify the states. As the primary focus for the development of the HSM model is long-term water resources planning, a far simpler stochastic framework is adopted where the hidden states can be simulated without recourse to auxiliary climatic variables.

This new modeling approach for reproducing long-term persistence in hydroclimatic time series is presented as follows: Section 2 describes the structure of the HSM model, followed by an outline of the calibration procedure in section 3. The Gibbs sampler, from the family of Bayesian techniques known as Markov Chain Monte Carlo methods, is used to infer the posterior distribution of the model parameters. Four different case studies were chosen for model calibration: long-term rainfall time series from Sydney, Brisbane, and Melbourne and a paleoclimatic reconstruction of Burdekin river runoff. Each one is located in a different Australian climatic regime. Analysis of the calibration results will determine whether the HSM modeling assumptions are justified and if a strong persistence structure has been identified. The discussion includes comparison with other stochastic models and considers other issues related to the application of the HSM modeling framework.

2. Hidden State Markov Model

The HSM model framework, illustrated in Figure 2, assumes that the climate is in one of two states: wet (W) or dry (D). Each state has an independent rainfall distribution, assumed to be Gaussian in this study. The persistence in each state varies because it is governed by the state transition probabilities. This provides an explicit mechanism to replicate the variable length wet and dry cycles. If these cycles are viewed as a manifestation of nonlinear climate dynamics, then the conceptualization of HSM model may be viewed as an attempt to explain these dynamics by introducing an external variable: the climate state "wet" or "dry." This two-state mechanism concurs with that of *Rodriguez-Iturbe et al.* [1991], who noted that behavior of hydroclimatic records exhibits persistence in several distinct states with the occasional transition between the states.

The simulation of a hydroclimatic time series is a two-step

process. In the first step the climate state at year t, s_t, is simulated by a Markovian process:

$$s_t | s_{t-1} \sim \text{MARKOV}(\mathbf{P}), \tag{1}$$

where $\mathbf{P}$ is the state transition probability matrix defined by

$$\mathbf{P} = [p_{ij}] = Pr(s_t = j | s_{t-1} = i) \qquad i, j = W, D. \tag{2}$$

When the transition probabilities p_{WD} and p_{DW} are known, the remaining components of the matrix are easily calculated by $p_{DD} = (1 - p_{DW})$, $p_{WW} = (1 - p_{WD})$. Thus, hereafter, $\mathbf{P}$ shall refer to a row vector of the transition probabilities:

$$\mathbf{P} = (p_{WD}, p_{DW}). \tag{3}$$

Once the state for year t is known, the hydroclimatic variables may be simulated using

$$y_t \sim \begin{cases} N(\mu_W, \sigma_W^2) & s_t = W \\ N(\mu_D, \sigma_D^2) & s_t = D \end{cases}, \tag{4}$$

where $N(\mu, \sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2. Therefore the vector of unknown parameters for the HSM model, θ, is composed of the rainfall distribution parameters for each state, the transition probabilities, and the hidden state time series, $\mathbf{S}_N = \{s_1, s_2, \ldots, s_n\}$, where

$$\theta' = (\mu_W, \sigma_W, \mu_D, \sigma_D, \mathbf{P}, \mathbf{S}_N). \tag{5}$$

Prior to model calibration the hidden state time series is unknown (i.e., it is "hidden"). Thus it is included as a model parameter to be estimated during the calibration process.

In summary, the HSM model makes the following modeling assumptions to simulate hydroclimatic time series: (1) The distribution of hydroclimatic data is composed of two independent wet and dry state distributions. (2) Both the wet and dry state distributions are Gaussian. (3) The climate state at time t, s_t, is purely dependent on the climate state from the previous time step, s_{t-1}. (4) The probability of state transition is assumed to be stationary over time.

3. Model Calibration: Gibbs Sampler

For model calibration a Bayesian framework is used to infer the distribution of the model parameters θ for the given time series data $\mathbf{Y}_N$. This distribution is referred to as the posterior distribution of the model parameters, $p(\theta | \mathbf{Y}_N)$. For the HSM model it is not possible to derive an analytical expression for the posterior distribution. Thus Markov Chain Monte Carlo (MCMC) simulation methods are employed to draw samples from the posterior distribution. The basic idea of MCMC methods is to simulate a Markov chain iterative sequence, where at each iteration a sample of the model parameters θ is generated. Given certain conditions, the distribution of these
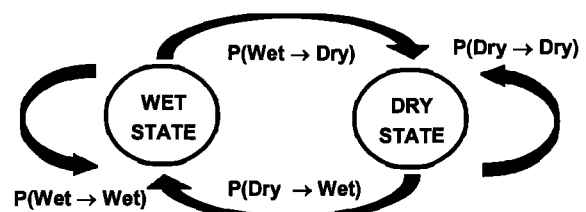

Figure 2. Model framework of the hidden state Markov (HSM) model.

samples converges to a stationary distribution which is the posterior distribution $p(\theta|\mathbf{Y}_N)$. *Tierney* [1994] provides a detailed discussion of the theoretical aspects of MCMC convergence, while a more applied explanation of MCMC methods is given by *Gelman et al.* [1995] and *Gilks et al.* [1996].

To calibrate the HSM model, the MCMC method known as the Gibbs sampler is applied. The idea in the Gibbs sampler is to simulate, in turn, the distribution of each parameter conditioned on the data and the remaining parameters (referred to as the full conditional distribution). Thus, at each iteration, each component of the parameter vector is sampled from the following conditional distribution:

$$\theta_j \leftarrow p(\theta_j|\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_d, \mathbf{Y}_N), \quad (6)$$

where d is the number of conditional distributions and θ_j can refer to either a scalar component or a subvector of the parameter vector θ. *Smith and Roberts* [1993, p. 5] note that if the parameters are highly correlated, then the convergence of the Gibbs sampler could be "painfully slow." To avoid this, highly correlated parameters are "blocked" together as a subvector of θ and sampled from a multivariate conditional distribution. For the HSM model, *Chibb* [1996] developed a method that allows the entire hidden state time series to be blocked together and sampled as a single parameter. The advantage of this is described by *Chibb* [1996, p. 81] as: "Instead of n additional blocks in the Gibbs sampler (the number required if each state is sampled from its full conditional distribution) only one additional block is required. This dramatically improves the convergence of the Markov chain induced by the Gibbs sampling algorithm." Therefore each iteration i of the Gibbs sampler proceeds as follows (note that x^i refers to the ith sample of parameter x):

$$\mathbf{S}_N^i \leftarrow p(\mathbf{S}_N|\mu^{i-1}, \sigma^{i-1}, P^{i-1}, \mathbf{Y}_N),$$

$$P^i \leftarrow p(\mathbf{P}|\mathbf{S}_N^i),$$

$$\mu_k^i \leftarrow p(\mu_k|\mathbf{S}_N^i, \sigma_k^{i-1}, \mathbf{Y}_N), \quad (7)$$

$$\sigma_k^i \leftarrow p(\sigma_k|\mathbf{S}_N^i, \mu_k^i, \mathbf{Y}_N),$$

where k refers to the hidden state, wet or dry. The hidden state time series $\mathbf{S}_N$ is sampled first because once it is known, the sampling of the remaining parameters is a relatively simple procedure. Chibb's method for sampling the entire hidden state time series as a single parameter is summarized in Appendix A for the two-state case of the HSM model. The conditional distribution of the transition probabilities $\mathbf{P}$ is purely dependent upon knowledge of $\mathbf{S}_N$, and hence the data $\mathbf{Y}_N$ are omitted. The sampling procedure for $\mathbf{P}$ is a special case of the expressions given by *Chibb* [1996] for multistate hidden Markov models. *Gelman et al.* [1995] provides the conditional distributions for the state distribution parameters μ and σ.

3.1. Choice of Priors

In this application we have little or no prior knowledge of the true parameter values. Thus diffuse or noninformative prior distributions are chosen, where a wide range of the parameter values are assigned a similar probability density. The aim of this approach is to let the data $\mathbf{Y}_N$ dominate the analysis so that inferences are unaffected by information external to the data [*Gelman et al.*, 1995].

Conjugate prior distributions were used for all the parameters. For the transition probabilities this is a beta distribution

[*Chibb*, 1996], and for the state means this is a normal distribution [*Gelman et al.*, 1995]. For both cases the parameter values chosen represented suitably diffuse proper prior distributions. For the state means the prior was bounded between 0.0 and 3000.0 as all the hydrological data used in this study were within this range. For the state standard deviations a scaled inverse χ^2 distribution [*Gelman et al.*, 1995] was used for the prior. Initially, the parameter values chosen represented an improper prior. However, when no data are sampled in a particular state, this results in the posterior becoming improper. Thus, for this case, a different set of parameter values were chosen that resulted in a relatively diffuse proper prior. Bounds of 0.0001 and 2.0 were used for the coefficient of variation (CV) as *McMahon and Mein* [1986] found that hydrological data from numerous sites around the world were within this range.

3.2. Implementation

To initialize the Gibbs sampler, arbitrary starting values must be supplied to the parameter vector. For the HSM model a heuristic method was used to provide these starting values. This method is described as follows: In the first step the hidden state time series is estimated by smoothing the data $\mathbf{Y}_N$ using a 5 year moving average filter and applying the long-term mean as a threshold to distinguish between the wet and dry states. If a smoothed value was above the threshold, it was classified as wet; if it was below, it was dry. Starting values for the transition probabilities were sampled using the method for the state transition probability matrix given by *Chibb* [1996]. The starting values for the wet and dry state mean and standard deviation were estimated from the two new data sets, Y_W and Y_D, which corresponded to the data in the wet and dry states sorted using the hidden state time series. Applying this method greatly increased the convergence rate as the Gibbs sampler was initialized at a good starting position close to the mode of the posterior distribution.

Once initialized, the Gibbs sampler is allowed to "warm up" for a given number, say, b, iterates before using the simulated output as samples from the posterior distribution $p(\theta|\mathbf{Y}_N)$. Multiple parallel paths of the Markov chain iterative sequence induced by the Gibbs sampler were used in this study. Compared to a single path, multiple paths are able to more widely explore the parameter space. For a good approximation of the posterior, 100 paths with 100 samples each were used, producing a total of 10,000 samples. For each path a warm-up of $b = 3500$ was found to be more than adequate to ensure convergence.

A number of indicators were used to monitor convergence. Time series plots of the percentiles of the sample distributions for all the parameters were inspected for signs of convergence failure. When using multiple paths, it is important to determine whether the individual paths are mixing properly in the parameter space. The R statistic, defined by *Gelman et al.* [1995], uses the within-path and between-path variation to estimate the potential scale reduction. If during an iterative sequence the individual paths are mixing slowly, then the R statistic will remain high. Thus it is monitored to see if there is any reason to believe further iterations may enhance the inference. For a review of MCMC convergence issues, refer to *Cowles and Carlin* [1996].

Figure 3. Location of hydroclimatic data sources.

4. Hydrological Data

Long-term continuous hydrological records were needed for this study. Three of the case studies were long-term rainfall records from the Sydney, Brisbane, and Melbourne metropolitan areas. The fourth case study was a paleoclimatic reconstruction of Burdekin river annual runoff. The relative location of each case study is shown in Figure 3. A summary of the details of each time series for the four case studies is given in Table 1. Each case study represents a cross section of the varying climatic regimes that exist on the eastern coast of Australia. In the north there exists a predominantly tropical climate with distinct dry and wet seasons influenced by the weather systems from the tropical Pacific. For example, 90% of the Burdekin river runoff occurs in the wet season (December to April). This influence is also evident in the Brisbane data as the majority of rainfall occurs in the summer months. Moving south, a transition occurs to a more temperate climate where the dominant weather systems originate from the Southern Ocean. Sydney is located in the middle of this transition region. This produces an even distribution of rainfall throughout the year, with the summer months influenced by the tropical systems and the Southern Ocean cold fronts dominating in winter. For the Melbourne rainfall the Southern Ocean systems are the most influential as the tropical systems rarely extend that far south.

The paleoclimatic reconstruction of Burdekin river runoff was used for the fourth case study to determine if a strong persistence structure could be identified over a significantly longer period. *Isdale et al.* [1998] used measurements of fluorescence bands in coral from the Great Barrier Reef to reconstruct the Burdekin river flow. The regression model accounted for 83% of the variability of the Burdekin river flow.

Table 1. Summary of the Details of the Time Series for the Four Case Studies

Case Study	Start Date	End Date	Length, years	Annual Statistics[a]
Sydney	Jan. 1859	April 1997	138	1221 (332)
Brisbane	Jan. 1860	June 1994	134	1146 (345)
Melbourne	Jan. 1856	May 1997	141	658 (127)
Burdekin	1644	1980	337	100 (99)

[a]For the annual statistics the empirical mean and standard deviation (in parentheses) are given.

It is important to note that the results from this reconstructed record are interpreted as indicative as there are some doubts about the regression model used to reconstruct the flow data.

5. Analysis of Calibration Results

The aim of this analysis is to evaluate whether the HSM modeling assumptions (outlined in section 2) are justified and to determine if a strong persistence structure is identified. To achieve this, the posterior distribution $p(\theta|Y_N)$ of the relevant model parameters is examined. This allows direct assessment of the parameter uncertainty, following the methodology of *Stedinger and Taylor* [1982b]. Hereafter, the term "posterior" shall refer to either the posterior distribution or the posterior density.

The first assumption is that the distribution of hydroclimatic data is a mixture of two independent wet and dry state distributions. If the difference between these two distributions is not significant, there is no justification for using a two-state model. Rudimentary statistical methods cannot be applied to test if the difference is significant because the sample data in each state change at every iteration in the Gibbs sampler. Thus the posterior difference between the means, $p(\mu_W - \mu_D|Y_N)$, the 5th percentiles, $p(5\%ile_W - 5\%ile_D|Y_N)$, and the 95th percentiles, $p(95\%ile_W - 95\%ile_D|Y_N)$ of the wet and dry state distributions is examined. If there is a low probability that there is negative difference between these wet and dry values, then there is considered to be a significant difference between the two distributions. This provides evidence to support the first assumption of a two-state framework. Figure 4 shows distribution of the posterior difference between the means and the 5th percentiles. The 95th percentiles were omitted because the results were the same for all case studies. The percentile values were calculated using the sampled mean and standard deviation and the assumption that the wet and dry state distributions were Gaussian. All the difference values have been standardized by dividing by the average of the wet and dry state means to facilitate comparison of all four case studies.

To test the second assumption that both the wet and dry state distributions are Gaussian, samples are drawn from the posterior predictive distribution and compared to the observed data. The posterior predictive distribution is defined by *Gelman et al.* [1995] as the distribution of replicated data y^{rep} simulated using the HSM model, given that the model parameters have been conditioned on the observed data. It is referred to as $p(y^{rep}|Y_N)$. In this study, Monte Carlo simulation is used to sample from this distribution. Ten thousand samples
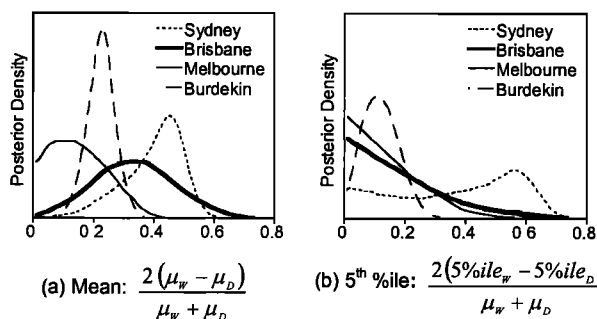

(a) Mean: $\dfrac{2(\mu_W - \mu_D)}{\mu_W + \mu_D}$ (b) 5th %ile: $\dfrac{2(5\%ile_W - 5\%ile_D)}{\mu_W + \mu_D}$

Figure 4. Posterior distribution of the standardized difference in the wet and dry state distributions for the four case studies: (a) mean and (b) 5th percentile.
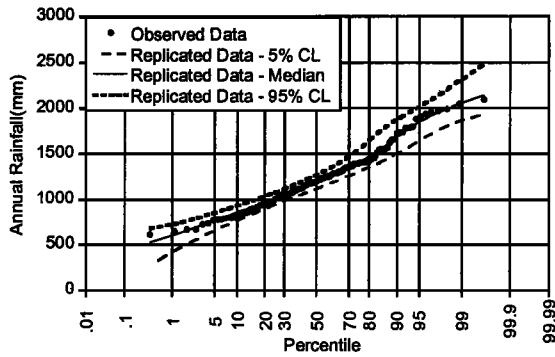
Figure 5. Normal probability plot of the observed data compared to the posterior predictive distribution of the replicated data $p(y^{rep}|\mathbf{Y}_N)$ for the Sydney annual rainfall (June–May water year). CL, confidence limit.

of the model parameters were drawn from their posteriors $p(\theta|\mathbf{Y}_N)$ and used in the HSM model to simulate 10,000 time series of length equal to the observed time series. The distribution of this replicated data was compared to the observed data. If the observed data are within the 5% and 95% confidence limits of the replicated data, this is considered a good fit to the data and hence there would be no reason to reject the assumption that the wet and dry state distributions are Gaussian. Figure 5 shows this comparison for the Sydney case study. Inspection of the posterior predictive distribution also determines whether a significant proportion of the distribution is truncated because hydrological values cannot be negative.

To test the third assumption that the current climate state s_t is dependent on the previous climate state s_{t-1}, the posteriors of the transition probabilities are examined for all four case studies (Figure 6). If the previous and current climate states are independent, there would be no persistence and the posteriors would indicate that the transition probabilities are unidentifiable. If the transition probabilities are identifiable, then this provides evidence to support the third assumption.

The number of years the climate persists in either the wet or dry state (the state residence time) indicates the strength of the persistence structure identified by the HSM model. A long state residence time indicates a strong persistence structure. The expected state residence time can be calculated from the transition probabilities (hereafter known as P_{TRANS}). As the occurrence of a state transition is a Bernoulli trial with prob-

ability P_{TRANS}, the number of years before that state transition occurs follows a geometric distribution. The expected state residence time is the mean of the geometric distribution, which is the inverse of P_{TRANS}. The axes of the plots in Figure 6 have been transformed to display the expected state residence time. As the transition probabilities have a range of 0 to 1, the expected state residence time can vary from ∞ to 1. To remove sampling noise, a matrix-smoothing algorithm was applied to these plots. An artefact of this technique is that in some cases the posteriors extend beyond the limits of ∞ to 1. This can be ignored; the overall trend of the posteriors is the important feature to note.

For each of the rainfall case studies the monthly rainfall was aggregated to produce 12 annual data series, which correspond to the 12 "water" years (January to December, February to January, etc). Each of these annual data series was calibrated to ascertain which water year exhibits the strongest wet and dry state signal. A state signal index (SSI) was developed as a measure of how well the wet and dry state "signal" is identified. The SSI is calculated as follows: In each iteration of the Gibbs sampler every year is classified as wet or dry. The posterior probability that a particular year is from the wet distribution, $p(s_t = W|\mathbf{Y}_N)$, can be calculated by counting the number of times it is classified as wet. If $p(s_t = W|\mathbf{Y}_N)$ is close to 0.5, then the year is equally likely to be in either a wet or a dry state. The SSI aggregates the $p(s_t = W|\mathbf{Y}_N)$ values for the entire time series. It is defined as

$$\text{SSI} = \frac{\sum_{t=1}^{N} |p(s_t = W|\mathbf{Y}_N) - 0.5|}{N}, \qquad (8)$$

where N is the number of years in the time series. If a high proportion of the time series have a value of $p(s_t = W|\mathbf{Y}_N)$ close to 0.5, then SSI $\to$ 0 and the wet and dry states are considered to be poorly identified. Conversely, as the number of years with a value of $p(s_t = W|\mathbf{Y}_N)$ close to 0 or 1 increases, then SSI $\to$ 0.5. Therefore the water year with highest value of SSI is presented for further analysis in this study. Figure 7 shows a time series of $p(s_t = W|\mathbf{Y}_N)$ and the corresponding SSI for each of the rainfall case studies.

For each of the rainfall case studies the autocorrelation function (up to lag 30) of the rainfall simulated from the calibrated HSM model was checked and found to be within the
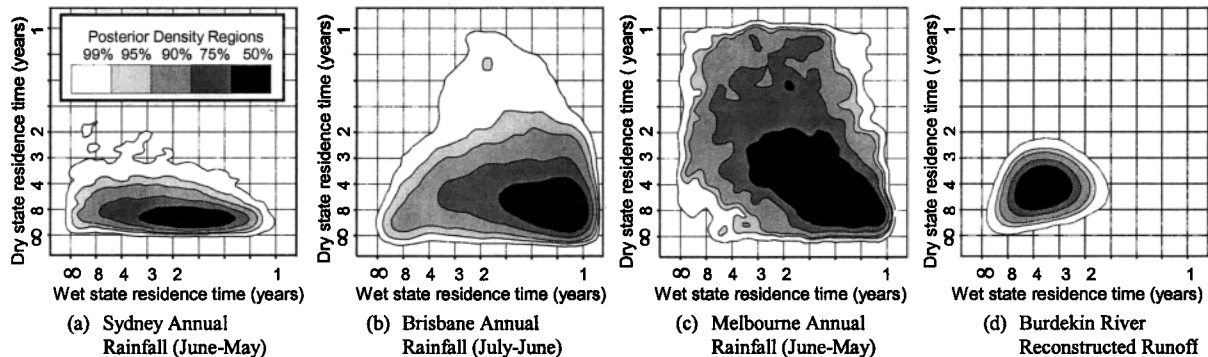

(a) Sydney Annual Rainfall (June-May) (b) Brisbane Annual Rainfall (July-June) (c) Melbourne Annual Rainfall (June-May) (d) Burdekin River Reconstructed Runoff

Figure 6. Posterior distributions of the transition probabilities for the (a) Sydney, (b) Brisbane, (c) Melbourne, and (d) Burdekin river case studies. The axes have been transformed to display the expected state residence time.
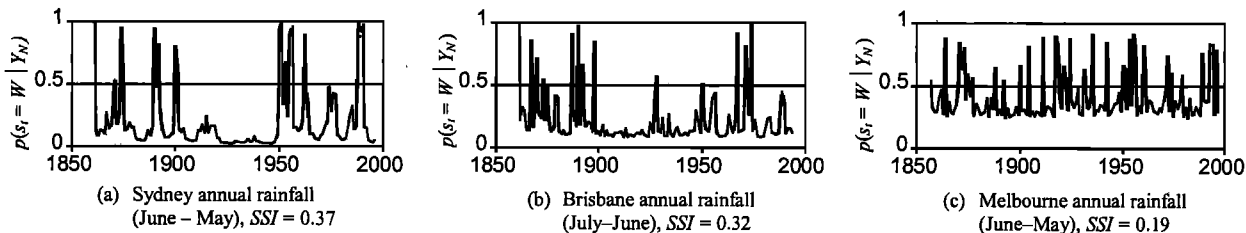
(a) Sydney annual rainfall
(June – May), SSI = 0.37

(b) Brisbane annual rainfall
(July–June), SSI = 0.32

(c) Melbourne annual rainfall
(June–May), SSI = 0.19

Figure 7. Time series plots of the posterior probability of a year being classified as wet, $p(s_t = W|Y_N)$, for the (a) Sydney, (b) Brisbane, and (c) Melbourne rainfall case studies.

95% confidence limits of the values estimated from the observed rainfall data.

5.1. Sydney Metropolitan Rainfall Data

For the Sydney rainfall case study the results support the HSM modeling assumptions and identify a strong persistence structure. Figure 7a shows that the states are well identified for the June to May water year which produced the highest SSI value of 0.374. There is a very low probability that the true value for the wet and dry means are equal (Figure 4a). The results were similar for the 95th percentiles. For the difference in the 5th percentiles (Figure 4b), although the posterior probability of the difference being negative is finite, the location of the mode indicates a significant positive difference. These results imply that there is a significant difference between the wet and dry state distributions and therefore support the first assumption. Figure 5 shows that the posterior predictive distribution is a good fit to the observed data and the proportion of the predictive distribution that is assigned negative values is negligible (<0.1%). These results support the second assumption. The posteriors of the transition probabilities, shown in Figure 6a, indicate that the parameters are identifiable. This supports the third assumption. A strong persistence structure is also identified as the mode of the posteriors corresponds to an expected dry state residence time close to 10 years.

5.2. Brisbane Metropolitan Rainfall Data

The results from Brisbane rainfall case study were similar to those from Sydney. The modeling assumptions were all justified, although the persistence structure identified was not as strong. The July to June water year had the highest SSI value of 0.323. In Figure 4b the mode is centered at zero, indicating a high probability that the wet and dry 5th percentiles are similar. However, the results for the mean (Figure 4a) and 95th percentiles still indicate that there is a significant difference between the wet and dry states, thereby supporting the first assumption. Comparison of the posterior predictive distribution to the observed data was the same as for Sydney; the results supported the second assumption. Figure 6b indicates that although the persistence is not as strong compared to Sydney, the transition probabilities are still clearly identifiable. This supports the third assumption.

5.3. Melbourne Metropolitan Rainfall Data

The results for this case study indicate that the HSM modeling assumptions cannot be justified for the Melbourne rainfall data. The June to May water year produced the highest SSI value of only 0.188. Figure 7c shows that compared to Sydney and Brisbane, the majority of years are equally likely to be from the wet or dry states. Figures 4a and 4b indicate that there is a high probability that the means and 5th percentiles of both

the wet and dry states are equal. This indicates that there is no significant wet and dry state difference, which violates the first assumption (hence the second assumption was not tested). Figure 6c indicates that the transition probabilities are poorly identified. This is significant evidence to reject the third assumption and indicates that the current climate state is probably independent of the previous climate state for the Melbourne data.

5.4. Burdekin River Reconstructed Runoff Data

For this case study the evaluation of the water year was not required as the data provided were already aggregated to the October to September water year. A log transformation of the reconstructed runoff data was necessary to ensure that the wet and dry state distributions were Gaussian. The following results are for this time series of log-transformed values, which will be referred to as the Burdekin data. Because the calibration results are similar to those for Sydney, the modeling assumptions were justified for the Burdekin data. Figures 4a and 4b show that there is a significant posterior difference between the wet and dry state distributions; thus the first assumption was not violated. The posterior predictive distribution was a good fit to the observed data, and the posteriors of the state transition probabilities are clearly identifiable (Figure 6d). It is important to note that the spread of the transition probabilities for the longer Burdekin record is greatly reduced when compared with the shorter Sydney and Brisbane rainfall time series (compare Figure 6d to Figures 6a and 6b). This indicates that the persistence structure is better identified when the length of the hydroclimatic record increases.

5.5. Summary of HSM Model Calibration Results

Table 2 shows the expected values of the posteriors for the HSM model parameters for the case studies where the assumptions of the HSM model were justified.

Table 2. Expected Values of the Posterior for the HSM Model Parameters for the Sydney, Brisbane, and Burdekin Case Studies

Case Study	Wet State		Dry State	
	Mean (s.d.), mm	Residence Time, years	Mean (s.d.), mm	Residence Time, years
Sydney	1736 (215)	1.8	1137 (238)	10.0
Brisbane	1541 (417)	1.5	1057 (261)	5.0
Burdekin[a]	161 (121)	3.4	52 (23)	4.5

[a]The Burdekin river values for the state distributions were back-transformed from the log-transformed parameter values.

6. AR(1) Model for Hydroclimatic Time Series

For the lag-one autoregressive (AR(1)) model the rainfall-generating mechanism is

$$Q_t - \mu = \phi(Q_{t-1} - \mu) + a_t, \tag{9}$$

where Q_t is the transformed annual rainfall for year t, ϕ is lag-one autoregressive coefficient, μ is mean of time series, and a_t is random shock $\sim N(0, \sigma_a^2)$. The annual rainfall is transformed using a Box-Cox transformation to ensure that the distribution of a_t is approximately Gaussian. Equation (9) shows that this year's rainfall value is a function of the previous year's rainfall and a random shock value. In contrast to the HSM model, there is no explicit mechanism for simulating the wet and dry periods.

To simulate the posteriors for the parameters of the AR(1) model, it was necessary to employ a different MCMC method called the Metropolis algorithm [Gelman et al., 1995]. The Metropolis algorithm follows the same principles as the Gibbs sampler. However, it represents a more general implementation. The major difference is that the model parameters are not sampled from their full conditional distributions. Instead, parameter samples are drawn from an arbitrary sampling distribution (usually the multivariate Gaussian), and then a posterior ratio test is used to accept or reject the samples. The Metropolis algorithm is required when it is not possible to derive the full conditional distributions required for the Gibbs sampler. This is the case for the AR(1) model given in (9). The transformation applied to the annual rainfall means that the full conditional distributions are not easily derived.

For the purposes of this comparison we are primarily interested in the posterior of the lag-one autoregressive coefficient ϕ, as it is considered a measure of the degree of the year-to-year persistence. If the ϕ value is close to zero, the simulated time series will have little or no persistence. The posterior of ϕ is shown in Figure 8 for the three rainfall case studies. For each of these time series the hypothesis that they are independent ($\phi = 0$) cannot be reasonably rejected. For Sydney, Brisbane, and Melbourne the probability that $\phi < 0$ is 9%, 40% and 18%, respectively. Thus these results provide no clear indication of a persistence structure. This contrasts with the HSM model results that indicated a strong persistence structure for the Sydney and Brisbane data.

7. Discussion

The methodology used to evaluate which water year was chosen implies that there exists a single water year which clearly had the strongest wet and dry state signal. However, this
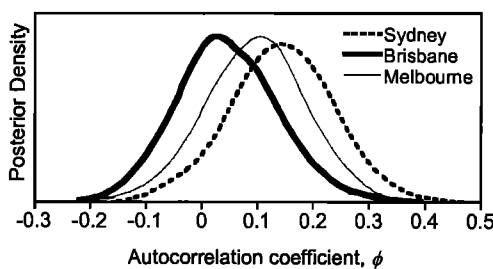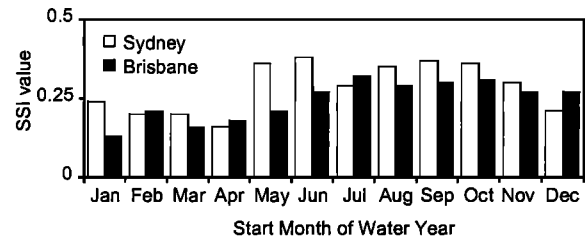
Figure 9. Values of the state signal index SSI for the 12 time series corresponding to the 12 water years for the Sydney and Brisbane rainfall case studies.

is not necessarily the case. Figure 9 shows the SSI value for each of the 12 water years. For the Sydney data there are 5 water years with similarly high SSI values. Their start months are all in the period May to October (excluding July). For the Brisbane data there are 4 water years with similarly high SSI values. Their start months are in the period from July to October. This trend indicates that the strongest wet and dry state signal exists when the water year begins in the winter period (approximately May to October). If the water year begins in the summer period, a strong wet and dry state signal is not evident. This implies that the wet and dry state persistence is primarily a summer phenomenon. As the summer rainfall is largely influenced by the tropical Pacific weather systems we could hypothesize that this is the cause of the wet and dry state persistence.

Comparing the results for the four case studies reveals that a strong persistence was identified when the hydroclimatic data were greatly influenced by the tropical Pacific weather systems (Sydney rainfall, Brisbane rainfall, and Burdekin reconstructed runoff). The Melbourne rainfall data provided no such evidence of persistence. It is known that the tropical weather systems rarely travel as far south as Melbourne. This provides more evidence to support the hypothesis that the primary cause of the wet and dry state persistence is associated with the climatic phenomena from the tropical Pacific. Further analysis with more rainfall sites would be required to fully test this hypothesis.

Of the four modeling assumptions outlined in section 2 only three were tested to verify that they were not violated. The final assumption, that the state transition probabilities are stationary over time, was not tested because a satisfactory methodology for testing this assumption has not yet been developed. It is believed that violation of this assumption would have a far lesser impact on the simulated time series when compared to the impact of using a stochastic model which does not have an explicit mechanism for replicating wet and dry periods.

In the future the HSM model could be modified so the transition probabilities or the hidden states are a function of some persistent climatic variable, such as SSTs (S. Franks, personal communication, 2000). This could potentially aid in our understanding of the climatological processes that produce the wet and dry state persistence. This modified HSM model would be similar to the class of models used by Hughes et al. [1999]. As mentioned in the introduction this type of model has its limitations for long-term water resources applications. Long-term simulations are dependent on long-term simulations of the required atmospheric variables. In contrast, the HSM model has the advantage that long-term simulations are a relatively simple procedure that do not require simulation of auxiliary climatic variables.
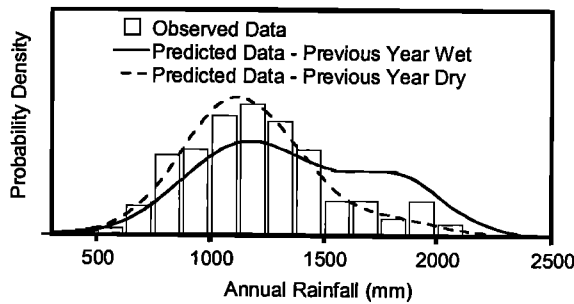
Figure 8. Posterior distribution of the autocorrelation coefficient ϕ for the AR(1) model for each of the rainfall case studies.

Figure 10. Histogram of the observed annual rainfall for Sydney (June–May water year) compared to the predicted distribution $p(\bar{y}|Y_N)$ for next year's rainfall given that the previous year's climate state was wet and dry.

To simulate future rainfall time series, the posterior predictive distribution outlined in section 5 is used. However, there is a subtle notation change. Instead of producing replicates of the observed data the distribution of the unknown future data $\bar{y}$ is predicted. Thus it is denoted as $p(\bar{y}|Y_N)$. This technique will be illustrated by simulating the predicted distribution of annual rainfall for Sydney 1 year into the future. Two different cases are used to demonstrate the influence the previous year's climate state has on the predicted rainfall distribution. The first case assumes that the previous year's climate state was wet, and the second assumes that it was dry. The two distributions are compared to a histogram of the observed data in Figure 10. For the wet state case there is a significant difference between the predicted and the observed distribution. As expected, high values for the rainfall have an increased chance of occurring. However, for the dry state case the difference is not as marked. This is because in the model calibration, approximately 85% of the Sydney observed data has a high probability of being classified as dry (refer to Figure 7a). Hence the dry state distribution closely follows the observed distribution. With a different data set the proportion of rainfall classified as dry may be less, and the predicted distribution for the dry state case would show a greater difference.

The AR(1) model results for Sydney and Brisbane were inconsistent with the HSM model. No statistically significant persistence structure was identified. The likely reason is that the AR(1) model has no explicit mechanism to simulate wet and dry periods. The AR(1) and other stochastic models (e.g., ARMA, broken line, and fractional Gaussian noise models) were previously applied because they are able to reproduce certain statistical characteristics of a time series (e.g., the Hurst phenomenon, a statistical indicator of long-term persistence; refer to Salas [1993]). In contrast, the HSM model provides a conceptually sounder model which emulates the influence of the global climatic mechanisms.

Apart from the AR(1) model, other alternatives such as a three-state HSM model or a threshold autoregressive model [Tong, 1990] could be applied. In Tong's model the hidden states could define the threshold, which is used to distinguish between two autoregressive processes. Compared to the HSM model this represents a different modeling approach for the state distributions. Both these models could potentially be viable alternatives. However, the important issue is the ability to reproduce the varying wet and dry spells, which the two-state HSM model is able to do.

In a rigorous statistical sense only part of the data set should

be used for the model calibration with the remaining portion reserved for model validation [Stedinger and Taylor, 1982a]. However, to identify and characterize long-term persistence, a long-term time series is required. Thus for the HSM model this provides a physical limitation as to how much of the data can be reserved for model validation. This is illustrated using synthetically generated data.

The aim of this analysis is to determine the approximate length of the time series required to identify a strong persistence structure. The expected values from the Sydney posteriors (refer to Table 2) were used to generate a synthetic time series because they represented a strong dry state persistence. This parameter set is denoted as set S1. Three synthetic time series of length 45, 90, and 140 years were generated (140 years is the approximate length of the original Sydney data). Figure 11 shows the posteriors of the transition probabilities when each of these synthetic series was calibrated to the HSM model. Both posteriors are close to uniform for the 45 year series. For the 90 year series the $P(\text{Wet to Dry})$ posterior is tighter, but the $P(\text{Dry to Wet})$ posterior still assigns a significant density to a wide range of values. It is only for the 140 year series that the posteriors indicate that the persistence structure is identifiable. The posterior difference between the wet and dry state distributions shows a similar trend. Only the 140 year series shows that the difference is clearly significant. These results indicate that long-term time series are required to identify a strong persistence structure.

The difference between the wet and dry state distributions is another factor that influences whether the persistence structure can be identified. If these two distributions become closer, it is harder to identify the persistence structure than if they are further apart. To illustrate this, two more parameter sets were used to generate synthetic time series. Both sets were based on set S1, except that for set S2 the wet state was increased to double the wet and dry mean difference and for set S3 the wet state mean was decreased to halve the difference. Table 3 shows the expected values and standard deviations of the posteriors for the transition probabilities for all three parameter sets. The results for the S2 set show that the persistence structure was identifiable even for a 45 year time series, while for the S3 set the synthetic parameter values were not recovered for the 140 year series. The posteriors for the remaining parameters showed that their synthetic values were recovered when the persistence structure was identified. This exercise illustrates that the difference between the wet and dry distributions influences whether the persistence structure can be identified.

To verify that the persistence identified by the HSM model could not be recovered from independent data, a synthetic
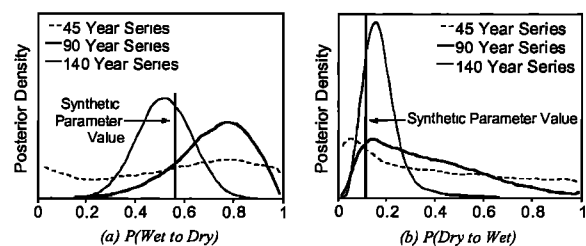

Figure 11. Posterior distributions of the transition probabilities calibrated to a synthetic time series of varying length: (a) $P(\text{Wet to Dry})$ and (b) $P(\text{Dry to Wet})$.

Table 3. Expected Values and Standard Deviations of the Posterior for the Transition Probabilities Calibrated to Synthetic Time Series of Varying Lengths

HSM Model Parameter Set[a]	Length of Synthetic Series		
	45 Years	90 Years	140 Years
P (Wet to Dry), Synthetic Parameter Value = 0.55			
S1	0.55 (0.29)	0.71 (0.15)	0.52 (0.11)
S2	0.44 (0.16)	0.52 (0.12)	0.49 (0.09)
S3	⋯	0.24 (0.23)	0.18 (0.16)
P (Dry to Wet), Synthetic Parameter Value = 0.10			
S1	0.40 (0.30)	0.36 (0.23)	0.15 (0.04)
S2	0.11 (0.06)	0.14 (0.04)	0.15 (0.04)
S3	⋯	0.41 (0.30)	0.27 (0.21)

[a]S1, Expected values of the Sydney posteriors; S2, same as S1, except wet state mean increased; S3, same as S1, except wet state mean decreased.

series was generated from a Gaussian distribution with parameters estimated from the Sydney data. When these independent data were calibrated to the HSM model, no persistence structure was identified.

In the introduction it was stated that the primary focus for the development of the HSM model is for long-term water resources planning. One specific application is for the drought risk assessment of water supply systems. Simulation of future water supply system performance requires the long-term simulation of future hydrological inputs, which the HSM model can provide. However, at this stage only single-site simulations are possible. For larger water supply systems a multisite simulation framework would be required. The development of a multisite HSM model will form part of the next stage of research. A multisite HSM model would also be able to identify regions with a common wet and dry persistence structure. The time series plots of the wetter and drier periods for Sydney and Brisbane in Figure 7 illustrate that some of the periods are common for both sets of data. Identification of these climatic regions will be a part of future research.

8. Conclusions

This paper introduces the HSM model as an alternative for modeling long-term persistence in hydroclimatic time series. Unlike the statistical models currently used (e.g., AR(1)), the conceptual basis of the HSM two-state model framework more closely simulates the influence of the global climatic mechanisms. The cumulative effect of several quasi-periodic global climate circulations (e.g., ENSO, IPO, and circumpolar vortex) produces the varying wet and dry regimes evident in Australian hydroclimatic data. The HSM model has an explicit mechanism to simulate this wet and dry persistence structure.

The Gibbs sampler, a Markov Chain Monte Carlo method, was used to calibrate the HSM model for four different case studies. For the case studies with strong influence from the tropical Pacific weather systems the results supported the assumptions of the HSM model framework and indicated the existence of a strong persistence structure. In direct contrast, for the same case studies the AR(1) calibration results indicated no clear persistence structure.

These results could potentially have significant ramifications for drought risk assessment and water resource management techniques. If the current statistical models used do not cor-

rectly identify the persistence structure in the data, this could lead to suboptimal water resource planning decisions.

Appendix A: Sampling From Conditional Distribution of Hidden State Time Series

The general method for sampling the hidden state time series for multistate Markov mixture models developed by *Chibb* [1996] is terse. This appendix presents a fuller treatment for the two-state case of the HSM model. The entire state time series is simulated using the distribution $p(S_N | Y_N, \theta)$, which is the joint posterior mass function of all the states given Y_N and θ. The derivation aims to develop a simple expression for this joint distribution. This will lead to a recursive simulation procedure where at each step, starting with the terminal state s_n, only a single state has to be drawn. For the following derivation it is convenient to adopt the notation used by *Chibb* [1996], where

$$S_N = \{s_1, \ldots, s_n\}, \qquad S_t = \{s_1, \ldots, s_t\},$$

$$S^{t+1} = \{s_{t+1}, \ldots, s_n\},$$

with a similar convention adopted for Y_N, Y_t, and Y^{t+1}.

A1. Step 1

By initially rewriting the joint distribution of the states $p(S_N | Y_N, \theta)$ and applying the conditional probability theorem repeatedly to the right-hand term a recursive expression results:

$$p(S_N | Y_N, \theta) = p(s_1, \{s_2, \ldots, s_n\} | Y_N, \theta)$$

$$= p(s_1 | \{s_2, \ldots, s_n\}, Y_N, \theta) p(\{s_2, \ldots, s_n\} | Y_N, \theta)$$

$$= p(s_1 | \{s_2, \ldots, s_n\}, Y_N, \theta) p(s_2 | \{s_3, \ldots, s_n\}, Y_N, \theta)$$

$$\cdot p(\{s_3, \ldots, s_n\} | Y_N, \theta).$$

The summary of this recursion is

$$p(S_N | Y_N, \theta)$$

$$= p(s_1 | S^2, Y_N, \theta) \cdots \cdot p(s_t | S^{t+1}, Y_N, \theta) \cdots \cdot p(s_n | Y_N, \theta).$$

$$\text{(A1)}$$

The typical term, excluding the terminal point, is therefore $p(s_t | S^{t+1}, Y_N, \theta)$.

A2. Step 2

Expand and split the Y_N term from the typical term in (A1) and apply Bayes theorem to the result. Further expand and split the S^{t+1} in the left-hand term and apply the conditional probability theorem:

$$p(s_t | Y_N, S^{t+1}, \theta) = p(s_t | Y_t, Y^{t+1}, S^{t+1}, \theta)$$

$$= p(Y^{t+1}, S^{t+1} | s_t, Y_t, \theta) p(s_t | Y_t, \theta) / p(Y^{t+1}, S^{t+1}, Y_t, \theta)$$

$$\propto p(s_{t+1}, S^{t+2}, Y^{t+1} | s_t, Y_t, \theta) p(s_t | Y_t, \theta)$$

$$\propto p(S^{t+2}, Y^{t+1} | s_{t+1}, s_t, Y_t, \theta) p(s_{t+1} | s_t, Y_t, \theta) p(s_t | Y_t, \theta).$$

The terms $p(Y^{t+1}, S^{t+1}, Y_t, \theta)$ and $p(S^{t+2}, Y^{t+1} | s_{t+1}, s_t, Y_t, \theta)$ are independent of s_t and hence become part of the normalizing constant. The Markovian property of the states means that s_{t+1} is purely dependent on knowledge of s_t; thus $p(s_{t+1} | s_t, Y_t, \theta)$ becomes $p(s_{t+1} | s_t, \theta)$. Therefore a simplified

expression for the typical term of the joint posterior density is the product of two terms:

$$p(s_t|\mathbf{Y}_N, \mathbf{S}^{t+1}, \boldsymbol{\theta}) \propto p(s_{t+1}|s_t, \boldsymbol{\theta})p(s_t|\mathbf{Y}_t, \boldsymbol{\theta}). \qquad (A2)$$

The first term is the transition probability of going from s_t to s_{t+1}, and the other term is the mass function of s_t given $\mathbf{Y}_t$. The normalizing constant of this mass function is the sum of the numbers obtained using (A2) as $s_t \in \{\text{WET, DRY}\}$.

The final stage of the calculation is to determine the mass function $p(s_t|\mathbf{Y}_t, \boldsymbol{\theta})$ given in (A2). The method developed is applied recursively for all s_t from $t = 1$ to n. Assume that the function $p(s_{t-1}|\mathbf{Y}_{t-1}, \boldsymbol{\theta})$ is available. Then repeat the following steps:

1. The prediction step involves determination of $p(s_t|\mathbf{Y}_{t-1}, \boldsymbol{\theta})$ using the total probability theorem:

$$p(s_t|\mathbf{Y}_{t-1}, \boldsymbol{\theta}) = \sum_{k \in \{\text{WET, DRY}\}} p(s_t|s_{t-1} = k, \boldsymbol{\theta})p(s_{t-1} = k|\mathbf{Y}_{t-1}, \boldsymbol{\theta}), \qquad (A3)$$

where the Markovian property of the states means that

$$p(s_t|s_{t-1}, \mathbf{Y}_{t-1}, \boldsymbol{\theta}) = p(s_t|s_{t-1}, \boldsymbol{\theta}).$$

2. The update step involves determination of $p(s_t|\mathbf{Y}_t, \boldsymbol{\theta})$ by first splitting the $\mathbf{Y}_t$ term so that it becomes $p(s_t|y_t, \mathbf{Y}_{t-1}, \boldsymbol{\theta})$ and applying Bayes theorem:

$$p(s_t|\mathbf{Y}_t, \boldsymbol{\theta}) \propto p(y_t|s_t, \mathbf{Y}_{t-1}, \boldsymbol{\theta})p(s_t|\mathbf{Y}_{t-1}, \boldsymbol{\theta}). \qquad (A4)$$

Now, the left-hand term of this result is the probability density of the rainfall at time t, y_t, given the climate state s_t. As the rainfall distribution is assumed Gaussian this is easily evaluated. The right-hand term is calculated in the prediction step. The normalizing constant for mass function given in (A4) is the sum of all the terms for $s_t \in \{\text{WET, DRY}\}$. At $t = 1$ these steps can be initialized by ignoring the prediction step and using the stationary Markovian state probabilities derived from the state transition probability matrix $\mathbf{P}$ for $p(s_1|\mathbf{Y}_0, \boldsymbol{\theta})$.

Using the expressions derived previously, the simulation of the state time series is a relatively simple procedure. First the prediction and update steps are run recursively to compute the mass functions $p(s_t|\mathbf{Y}_t, \boldsymbol{\theta})$, for all $t = 1$ to n. Sampling of the state time series starts by initially simulating s_n using $p(s_n|\mathbf{Y}_N, \boldsymbol{\theta})$. The remaining states can be simulated using the mass function $p(s_t|\mathbf{Y}_N, \mathbf{S}^{t+1}, \boldsymbol{\theta})$ calculated using the expression given in (A2).

References

Allan, R. J., El Nino-Southern Oscillation influences on the Australasian Region, *Prog. Phys. Geogr.*, 12, 4–40, 1988.

Allan, R., J. Lindsay, and D. Parker, *El Nino Southern Oscillation and Climatic Variability*, 405 pp., CSIRO Pub., Collingwood, Victoria, Australia, 1996.

Chibb, S., Calculating posterior distributions and modal estimates in Markov mixture models, *J. Econometrics*, 75, 79–97, 1996.

Cowles, M. K., and B. P. Carlin, Markov chain Monte Carlo convergence diagnostics: A comparative review, *JASA J. Am. Stat. Assoc.*, 91(434), 883–904, 1996.

Gelman, A., J. B. Carlin, H. S. Stern, and D. S. Rubin, *Bayesian Data Analysis*, 526 pp., Chapman and Hall, New York, 1995.

Gilks, W. R., S. Richardson, and D. J. Spieghalter, *Markov Chain Monte Carlo in Practice*, 486 pp., Chapman and Hall, New York, 1996.

Grayson, R. B., R. M. Argent, R. J. Nathan, T. A. McMahon, and R. G. Mein, *Hydrological Recipes: Estimation Techniques in Australian hydrology*, 125 pp., Coop. Res. Cent. for Catchment Hydrol., Melbourne, Victoria, Australia, 1996.

Hughes, J. P., and P. Guttorp, A class of stochastic models for relating synoptic scale atmospheric patterns to regional hydrologic phenomena, *Water Resour. Res.*, 30(5), 1535–1546, 1994.

Hughes, J. P., P. Guttorp, and S. P. Charles, A non-homogeneous hidden Markov model for precipitation occurrence, *Appl. Stat.*, 48(1), 15–30, 1999.

Isdale, P. J., B. J. Stewart, K. S. Tickle, and J. M. Lough, Palaeohydrological variation in a tropical river catchment: A reconstruction using fluorescent bands in corals of the Great Barrier Reef, Australia, *Holocene*, 8(1), 1–8, 1998.

Kane, R. P., On the relationship of ENSO with rainfall over different parts of Australia, *Aust. Meteorol. Mag.*, 46(1), 39–49, 1997.

Linacre, E., and B. Geerts, *Climates and Weather Explained*, 432 pp., Routledge, New York, 1997.

Lough, J. M., Rainfall variations in Queensland, Australia: 1891–1986, *Int. J. Climatol.*, 11, 745–768, 1991.

McMahon, T. A., and R. G. Mein, *River and Reservoir Yield*, 368 pp., Water Resour. Publ., Highlands Ranch, Colo., 1986.

Nicholls, N., Sea surface temperatures and Australian winter rainfall, *J. Clim.*, 2, 965–973, 1989.

Nicholls, N., and A. Kariko, East Australian rainfall events: Interannual variations, trends, and relationships with the Southern Oscillation, *J. Clim.*, 6, 1141–1152, 1993.

Power, S., T. Casey, C. Folland, A. Colman, and V. Mehta, Interdecadal modulation of the impact of ENSO in Australia, *Clim. Dyn.*, 15(5), 319–324, 1999.

Rodriguez-Iturbe, I., D. Entekhabi, and R. L. Bras, Nonlinear dynamics of soil moisture at climate scales, 1, Stochastic analysis, *Water Resour. Res.*, 27(8), 1899–1906, 1991.

Salas, J., Analysis and modeling of hydrologic time series, in *Handbook of Hydrology*, edited by D. Maidment, pp. 19.1–19.72, McGraw-Hill, New York, 1993.

Smith, A. F. M., and G. O. Roberts, Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods, *J. R. Stat. Soc., Ser. B*, 55(1), 3–23, 1993.

Smith, I., Indian Ocean sea-surface temperature patterns and Australian winter rainfall, *Int. J. Climatol.*, 14, 287–305, 1994.

Srikanthan, R., and B. J. Stewart, Analysis of Australian rainfall and rainday data with respect to climate variability and change, 254 pp., Hydrol. Branch, Bur. of Meteorol., Melbourne, Victoria, Australia, 1992.

Stedinger, J. R., and M. R. Taylor, Synthetic streamflow generation, 1, Model verification and validation, *Water Resour. Res.*, 18(4), 909–918, 1982a.

Stedinger, J. R., and M. R. Taylor, Synthetic streamflow generation, 2, Effect of parameter uncertainty, *Water Resour. Res.*, 18(4), 919–924, 1982b.

Tierney, L., Markov chains for exploring posterior distributions, *Ann. Stat.*, 22, 1701–1762, 1994.

Tong, H., *Non-linear Time Series*, 564 pp., Oxford Univ. Press, New York, 1990.

White, W. B., and R. G. Peterson, An Antarctic circumpolar wave in surface pressure wind, temperature and sea-ice extent, *Nature*, 380, 669–702, 1996.

Zucchini, W., and P. Guttorp, A hidden Markov model for space-time precipitation, *Water Resour. Res.*, 27(8), 1917–1923, 1991.

G. Kuczera and M. Thyer, Department of Civil, Surveying and Environmental Engineering, University of Newcastle, Callaghan, New South Wales 2308, Australia. (cegak@cc.newcastle.edu.au; cemat@cc.newcastle.edu.au)