Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Hydrology 275 (2003) 12–26
Journal of Hydrology
www.elsevier.com/locate/jhydrol

A hidden Markov model for modelling long-term persistence in multi-site rainfall time series 1. Model calibration using a Bayesian approach

Mark Thyer, George Kuczera*

Department of Civil, Surveying and Environmental Engineering, University of Newcastle, Callaghan, NSW 2308, Australia

Received 15 October 2001; revised 25 November 2002; accepted 6 December 2002

Abstract

A Bayesian approach for calibrating a hidden Markov model (HMM) to long-term multi-site rainfall time series is presented. Using a HMM approach for simulating long-term persistence is attractive because it has an explicit mechanism to produce long-term wet and dry periods which are a feature of many long-term hydrological time series. The ability to fully evaluate parameter uncertainty for the multi-site HMM represents an advance in the stochastic modelling of long-term persistence in multi-site hydrological time series. The challenges in applying the Bayesian Markov chain Monte Carlo (MCMC) method known as the Gibbs sampler to infer the posterior distribution of the multi-site HMM parameters are fully outlined. The specification of appropriate prior distributions was found to be crucial for the successful implementation of the Gibbs sampler. It is described how using synthetic data led to the development of an appropriate prior specification. Further synthetic data analysis showed how the benefits of space-for-time substitution for identifying the long-term persistence structure are dependent on the spatial correlation that exists in multi-site data. A methodology for handling missing data is also described. This study highlights the important role of the priors in Bayesian analysis using MCMC methods by illustrating that misleading inferences can result if the priors are inappropriately specified.

Keywords: Hidden Markov model; Multi-site framework; Gibbs sampler; Long-term persistence; Australian rainfall; Markov chain Monte Carlo

1. Introduction

Numerous stochastic models have been developed to simulate multi-site long-term hydrological data for long-term water resources planning. To ensure optimal planning decisions for drought risk assessments are made it is important that the model has a conceptual framework to reproduce the long-term persistence of the hydrological data. The most popular models are the ARMA–type processes (Grayson et al., 1996; Salas, 1993; Salas and Smith, 1981; Srikanthan and McMahon, 2000). As an alternative to the ARMA approach, Thyer and Kuczera (2000) applied a hidden Markov model (HMM) for simulating long-term persistence in single site rainfall time series. The motivation was that unlike the ARMA approach, the HMM has an explicit mechanism to produce time series with long-term wet and dry periods. These

* Corresponding author. Fax: +61-2-49216991.
E-mail addresses: markthyer@hotmail.com (M. Thyer), cegak@cc.newcastle.edu.au (G. Kuczera).

0022-1694/03/$ - see front matter © 2003 Published by Elsevier Science B.V.
doi:10.1016/S0022-1694(02)00412-2

long-term wet and dry regimes are prominent in Australian hydrological data due to the influence of quasi-cyclic global climatic mechanisms (e.g. the El Niño phenomenon). Thyer and Kuczera's (2000) results supported the notion that the HMM provides a conceptually sounder approach for simulating long-term persistence in hydrological time series than the ARMA-type processes.

This study presents a Bayesian approach for fully quantifying the parameter uncertainty of a HMM for simulating long-term rainfall time series at multiple sites. This extension to a multi-site HMM framework for modelling long-term persistence has several advantages over the single site HMM used by Thyer and Kuczera (2000). Firstly, the single site HMM is inadequate for larger water resource systems where multi-site simulations are required for drought risk assessments. Secondly, Thyer and Kuczera (2000) showed in a single-site analysis that the length of historic records limited the ability to identify long-term persistence. In a multi-site framework the long-term persistence structure is assumed to be regional. As a result, space-for-time substitution arising from the use of multi-site data may provide more information to better identify the long-term persistence structure than would a single-site analysis. Finally, by exploiting the spatial correlation that exists in multi-site data a methodology for handling missing data can be developed. This enables greater utilization of the available rainfall information for the identification of the long-term persistence structure.

Several researchers have previously applied the HMM framework for modelling hydrological time series. Jackson (1975) was one of the first to suggest a two-state HMM for modelling drought lengths in streamflow time series. However, Jackson (1975) only made limited progress with parameter estimation noting that it was a difficult problem. Thyer and Kuczera (2000) presented a full Bayesian solution to this problem for a single site HMM. In the multi-site context Zucchini and Guttorp (1991) used a HMM to describe the occurrence of wet/dry days at multiple sites. Hughes and Guttorp (1994) built on this approach to relate daily multi-site precipitation occurrence to synoptic scale atmospheric data using a nonhomogenous hidden Markov model (NHMM). This has been extended to include a model for precipitation amounts (Bellone et al., 2000; Charles et al., 1999a). The NHMM approach shows great potential for applications such as assessing climatic change because precipitation processes are related to atmospheric variables, as shown by Charles et al. (1999b). However, it is not suitable for long-term rainfall simulation as currently there is no method available for stochastically simulating long-term atmospheric data (Thyer and Kuczera, 2000). In this study the motivation for applying the HMM concept for simulating long-term persistence is for water resources applications. Therefore, a simpler approach was adopted where the hidden states could be simulated without recourse to auxiliary atmospheric variables. Furthermore, this previous research has only considered procedures which derive single value parameter estimates for the HMM. In contrast, this study will present a Bayesian procedure for fully evaluating parameter uncertainty of a multi-site HMM.

Thyer and Kuczera (2000) used a Markov chain Monte Carlo (MCMC) method known as the Gibbs sampler to evaluate the parameter uncertainty of a single site HMM. Originally it was thought that the extension of this procedure to the multi-site case would be a straightforward exercise. However, several unexpected challenges were encountered, primarily regarding the specification of appropriate priors. The purpose of this paper is to describe these problems and how they were overcome to ensure successful implementation of the Gibbs sampler. Using synthetic data it is also investigated how the expected benefit of space-for-time substitution for identifying the long-term persistence structure is influenced by the spatial correlation between sites. In the companion paper (Thyer and Kuczera, 2003) the multi-site HMM will be applied to real rainfall data. That analysis provides insight into the existence of a regional long-term persistence structure in some major water supply catchments on the east coast of Australia and also outlines some challenges, which still exist for the Bayesian calibration procedure of the multi-site HMM.

This paper is organised as follows: A description of the multi-site HMM model is given in Section 2. Section 3 outlines the calibration procedure using the Gibbs sampler approach; it includes an explanation of the difficulties faced in specifying appropriate priors and a description of the methodology used to handle missing data. In Section 4 the synthetic data analysis, which led to the development of a suitable prior specification, is presented. Included is an investigation of the effects of spatial correlation in multi-site data on the identification of long-term persistence structure. In Section 5 the discussion outlines the unresolved issues and limitations that require further investigation for the presented Bayesian calibration procedure of the multi-site HMM.

2. Multi-site hidden Markov model

The HMM framework (Fig. 1) assumes the climate is in one of two states: wet (W) or dry (D). Each state has an independent rainfall distribution. The persistence in each state is governed by the state transition probabilities. For example, if $P(\text{Dry} \rightarrow \text{Wet})$ is low there will be long-term persistence in the dry state. This provides an explicit mechanism to produce rainfall time series with long-term wet and dry periods. If these varying wet and dry periods are viewed as manifestations of nonlinear climate dynamics then the HMM conceptualization can be viewed as an attempt to simulate these dynamics by introducing an external variable—the climate state 'wet' or 'dry'. Thus, while not directly modelling the complex climatic processes the HMM can emulate their influence on long-term hydrological time series. In this multi-site HMM the climate state is assumed to be 'regional', i.e. it is common to all rainfall sites at any point in time. This notion of a regional climate state strengthens the conceptual link to the climatic mechanisms that produce the long-term wet and dry periods. These climatic mechanisms have a large-scale impact and therefore it is intuitive to think in terms of a regional climate state.

The simulation of rainfall values follows a two-step process. In the first step, the climate state at time

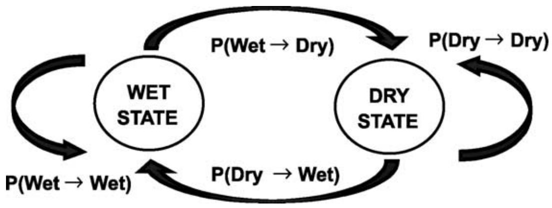


Fig. 1. Model framework of the hidden Markov model.

step 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 matrix defined by:

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

where $i, j =$ climate state (W or D). Once the climate state at time step t is known the vector of rainfall values, $\mathbf{y}_t$ at r multiple sites can be simulated using:

$$\mathbf{y}_t = \begin{pmatrix} y_t^1 \\ \vdots \\ y_t^r \end{pmatrix} \sim \begin{cases} N_r(\boldsymbol{\mu}_\mathrm{W}, \boldsymbol{\Sigma}_\mathrm{W}) & \text{if } s_t = \mathrm{W} \\ N_r(\boldsymbol{\mu}_\mathrm{D}, \boldsymbol{\Sigma}_\mathrm{D}) & \text{if } s_t = \mathrm{D} \end{cases} \tag{3}$$

where $N_r(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a multivariate Gaussian distribution in r dimensions, with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

The vector of unknown parameters for the multi-site HMM, $\boldsymbol{\theta}$, is therefore:

$$\boldsymbol{\theta}' = (\boldsymbol{\mu}_\mathrm{W}, \boldsymbol{\Sigma}_\mathrm{W}, \boldsymbol{\mu}_\mathrm{D}, \boldsymbol{\Sigma}_\mathrm{D}, \mathbf{P}, S_N) \tag{4}$$

where $S_N = \{s_1, s_2, \ldots, s_n\}$, the hidden state time series, is included with the model parameters because it is assumed unknown apriori and must be estimated during model calibration.

3. Model calibration procedure

A Bayesian framework is used to infer the posterior distribution of the model parameters, $\boldsymbol{\theta}$, for the observed time series data, $\mathbf{Y}_N^{\text{obs}}$. In this multi-site approach $\mathbf{Y}_\mathrm{N}^{\text{obs}}$ represents a matrix of the time series of N vectors of the observed rainfall at r sites. If these time series are not contiguous then all the years when the sites are missing data are denoted as $\mathbf{Y}^{\text{mis}}$ and are treated as additional unknown parameters. Hence, in the presence of missing data the posterior distribution is denoted as $\mathbf{p}(\boldsymbol{\theta}, \mathbf{Y}^{\text{mis}}|\mathbf{Y}_N^{\text{obs}})$. For the HMM it is not possible to derive an analytical expression for this posterior hence a MCMC method known as the Gibbs sampler is used to simulate values from the posterior. Further explanation of MCMC methods is provided by Chib and Greenberg (1995), Gelman et al. (1995) and Gilks et al. (1996).

3.1. The Gibbs sampler

The Gibbs sampler simulates a Bayesian posterior by drawing samples, in turn, for each component of the parameter vector from the distribution of that component conditioned on the data and the remaining parameters (referred to as the full conditional posteriors). This iterative sequence of samples is a Markov chain. Given certain conditions the distribution of these samples converges to a stationary distribution, which is the posterior. Tierney (1994) provides an extensive treatment of the theoretical aspects of MCMC convergence.

At the ith iteration of the Gibbs sampler (hereafter referred to as a Gibbs iteration), each component of the parameter vector is sampled from the following conditional posterior (note: x^i refers to the ith sample of parameter x):

$$\theta_j^i \sim p(\theta_j | \theta_{-j}^{i-1}, Y_N) \qquad j = 1, \ldots, d \tag{5}$$

where θ_{-j}^{i-1} represents the components of θ, except for θ_j, at their current values, such that $\theta_{-j}^{i-1} = (\theta_1^i, \ldots, \theta_{j-1}^i, \theta_{j+1}^{i-1}, \ldots, \theta_d^{i-1})$, where d represents the number of components of θ. The term 'component' is used because θ_j can refer to either a scalar or a subvector of θ.

The role of the priors in the Gibbs sampler is illustrated by expanding Eq. (5) using Bayes' rule:

$$p(\theta_j | \theta_{-j}^{i-1}, Y_N) = p(Y_N | \theta_j, \theta_{-j}^{i-1}) p(\theta_j | \theta_{-j}^{i-1}) \tag{6}$$

where $p(\theta_j | \theta_{-j}^{i-1})$ is the prior distribution for θ_j, which can be dependent on the other parameter values θ_{-j}^{i-1}.

Thyer and Kuczera's (2000) procedure for applying the Gibbs sampler to the single site HMM is adapted to the multi-site version by modifying the conditional posteriors to enable the sampling of the multi-site state rainfall parameters and the missing data. The missing data values are sampled from their full conditional distributions. These sampled missing data values then augment the observed data to form the complete time series, $\mathbf{Y}_N$, that would have been observed in the absence of missing data (Gelman et al., 1995), such that:

$$\mathbf{Y}_N = (\mathbf{Y}_N^{\mathrm{obs}}, \mathbf{Y}^{\mathrm{mis}}) \tag{7}$$

Thus, the full conditional posteriors for this implementation of the Gibbs sampler become:

$$\begin{aligned}
&S_N^i \sim p(S_N | \boldsymbol{\mu}^{i-1}, \boldsymbol{\Sigma}^{i-1}, \mathbf{P}^{i-1}, \mathbf{Y}_N^{i-1}) \\
&\mathbf{P}^i \sim p(\mathbf{P} | S_N^i) \\
&\mathbf{Y}^{\mathrm{mis}^i} \sim p(\mathbf{Y}^{\mathrm{mis}^i} | S_N^i, \boldsymbol{\mu}^{i-1}, \boldsymbol{\Sigma}^{i-1}, \mathbf{Y}_N^{\mathrm{obs}}) \\
&\mathbf{Y}_N^i = (\mathbf{Y}_N^{\mathrm{obs}}, \mathbf{Y}^{\mathrm{mis}^i}) \\
&\boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i \sim p(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^i | S_N^i, \mathbf{Y}_N^i)
\end{aligned} \tag{8}$$

where k refers to the hidden state, wet or dry. The sampling procedures for S_N and $\mathbf{P}$ are given in Thyer and Kuczera (2000).

3.2. Specification of priors for the state rainfall parameters

In a Bayesian framework it is generally recommended that prior distributions be chosen to represent what is known as diffuse or uninformative priors (Gelman et al., 1995) to ensure that the inferences are unaffected by information external to the data $\mathbf{Y}_N$. However, for the state rainfall parameters $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$ the additional complexity of the multivariate state distributions causes some complications for implementing the Gibbs sampler that prevent the use of uninformative priors which are independent of the data. Therefore, data-based empirical Bayes approximations as priors are developed. The reasons for being forced to adopt this less than ideal approach are outlined below.

For the state rainfall parameters it is convenient to use the parameterisation where $\boldsymbol{\mu}$ and the inverse of the covariance matrix $\boldsymbol{\Sigma}^{-1}$ are considered to be jointly unknown. In this case the corresponding joint prior density is $p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1})$. Using the relationship $p(\boldsymbol{\mu}, \boldsymbol{\Sigma} - 1) = p(\boldsymbol{\mu} | \boldsymbol{\Sigma} - 1) p(\boldsymbol{\Sigma} - 1)$ the priors for the state rainfall parameters can be specified. For $p(\boldsymbol{\mu} | \boldsymbol{\Sigma}^{-1})$, a conjugate prior density, the multivariate Gaussian density is used. Similarly for $p(\boldsymbol{\Sigma}^{-1})$ a conjugate prior, the Wishart density is used. This results in the following parameterisation for $p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1})$ (Gelman et al., 1995):

$$\boldsymbol{\Sigma}_k^{-1} \sim W_r(\nu_0, \mathbf{W}_0) \qquad \boldsymbol{\mu}_k | \boldsymbol{\Sigma}_k^{-1} \sim N_r(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}/\kappa_0) \tag{9}$$

where $W_r(\nu, \mathbf{W})$ denotes a Wishart distribution with ν degrees of freedom and scale matrix $\mathbf{W}$. As it controls the form of the precision matrix $\boldsymbol{\Sigma}^{-1}$ it is convenient to think of the prior scale matrix $\mathbf{W}_0$ as a prior precision matrix (DeGroot, 1970).

The major challenge found in implementing the Gibbs sampler was choosing suitable values for the hyperparameters: The prior degrees of freedom ν_0 and the prior precision matrix $\mathbf{W}_0$ for the Wishart distribution, the prior mean vector $\boldsymbol{\mu}_0$ and the number of prior measurements on the $\boldsymbol{\Sigma}$ scale κ_0 for the multivariate Gaussian distribution. Hyperparameter values can be chosen to produce noninformative improper priors. However, for both the Wishart and multivariate Gaussian distributions if improper priors are used the posteriors will become improper if the number of data is less than the dimension of the distribution. This can readily occur during the Gibbs iterations when the hidden state time series is sampled with the number of rainfall values in a particular state being less than the number of sites. Hence, the use of improper priors is prohibited because it can lead to improper posteriors which are not allowed in a Bayesian analysis (Diebolt and Robert, 1994).

Hyperparameter values therefore must be chosen which result in proper priors. Initially, $\boldsymbol{\mu}_0$ and $\mathbf{W}_0$ were set to arbitrary constant values. Every element of $\boldsymbol{\mu}_0$ was set to 1000.0 and the diagonals of $\mathbf{W}_0$ were set to values representative of a standard deviation of 300.0 and the off-diagonal terms were set to values that represented the spatial correlation structure based on the assumption that the correlation between sites decayed as a function of their inter-site distance. The constants ν_0 and κ_0 were set to values suitably low enough to ensure a diffuse proper prior. Using these hyperparameter values attempts to calibrate the multi-site HMM with the Gibbs sampler were unsuccessful because of the existence of computational 'trapping' states. During the Gibbs iterations when a low number of rainfall values are assigned to one state there is little or no information about the rainfall parameters for that state and they are sampled directly from their priors. If the priors are such that the sampled mean vector and covariance matrix are not reasonably close to the rainfall data there is a very low chance that any observed rainfall data will be assigned to that state in the next iteration. Thus, the process repeats itself and the chain of parameter samples becomes stuck in a so-called trapping state, and therefore the Gibbs sampler is unlikely to converge.

Several studies use the technique of reparameterisation to resolve the problem of trapping states in MCMC algorithms (Billio et al., 1999; Robert, 1996; Robert and Mengersen, 1999; Robert and Titterington, 1998). In the context of the HMM, the idea in reparameterisation is to express the rainfall parameters for one state as a perturbation of the parameters of the opposing state. This has two advantages: It allows noninformative priors to be used and more importantly it eliminates the trapping states, as all the data contribute to the sampling of all the parameters. Robert and Mengersen (1999) demonstrated this technique for a model with univariate Gaussian state distributions. Attempts were made to extend their ideas to the multivariate Gaussian state distributions of the multi-site HMM. Reparameterisation of the mean vector was theoretically possible. However, no satisfactory method for the expression of the covariance matrix for one state as a perturbation of the covariance matrix for the other state was found. Hence, for now at least, the technique of reparameterisation was not applied to the multi-site HMM.

In this analysis the approach taken was to find suitable hyperparameter values which minimize the chance of the Gibbs sampler becoming stuck in a trapping state. Robert (1996) proposed the use of empirical Bayes approximations as priors, with the following data-based hyperparameter values:

$$\boldsymbol{\mu}_0 = \bar{\mathbf{y}} \quad (10)$$

$$\mathbf{W}_0 = \frac{2\mathbf{S}^{-1}}{(\nu_0 - 3)} \quad (11)$$

where $\bar{\mathbf{y}}$ and $\mathbf{S}$ are the sample mean vector and covariance matrix of the entire data set $\mathbf{Y}_N^{\text{obs}}$. The constants ν_0 and κ_0 are set to values of 6 and 1, respectively, to reduce prior information. Eq. (11) implies a prior expectation for $\boldsymbol{\Sigma}_k$ of $\mathbf{S}/2$ as $\nu_0 \to \infty$. The priors defined by the hyperparmeter set given in Eqs. (10) and (11) will hereafter be referred to as prior P1.

It was found that the application of prior P1 provided a means for the Gibbs sampler to escape from the trapping states. The empirical Bayes approximations gave $\boldsymbol{\mu}_0$ and $\mathbf{W}_0$ a structure which resembles the actual rainfall data. This is key, because now when no rainfall data is assigned to a particular state plausible values for state rainfall parameters are still sampled from their respective priors. Thus, in the next iteration of the Gibbs sampler it is more likely that observed rainfall data will be assigned to that state and therefore the chain of parameter samples is able to escape from the trapping state.

Although these empirical Bayes approximations were able to alleviate the problem of the trapping states, in Section 4.1 it is shown that for certain synthetic data examples the value for $\mathbf{W}_0$ proposed by Robert (1996) gives results which are counterintuitive. It is not explicitly stated by Robert (1996) why the values of 2 and $(\nu_0 - 3)$ were assigned to $\mathbf{W}_0$. A more intuitive and flexible approach, which was found to be effective, was to use:

$$\mathbf{W}_0 = \frac{\mathbf{S}^{-1}}{\left(\nu_0 - \dfrac{r+1}{2}\right)} \tag{12}$$

This implies a prior expectation for $\boldsymbol{\Sigma}_k$ of $\mathbf{S}$ as $\nu_0 \to \infty$, which for this application seems more reasonable than $\mathbf{S}/2$. The $(r+1)/2$ term in the denominator was required because at low values of ν_0 the Wishart distribution is highly skewed and the expected value is significantly different from the mode of the prior density. The addition of the $(r+1)/2$ term provided the necessary adjustment to ensure that the mode is approximately equal to $\mathbf{S}$ for low values of ν_0. In addition, it was found not always suitable to set ν_0 at a constant value of 6 as proposed by Robert (1996). Alternatively, ν_0 should be adjusted depending on the number of sites to ensure an adequately diffuse proper prior—refer Section 4.1 for further explanation. The priors defined by the hyperparameter values given in Eqs. (10) and (12) will hereafter be referred to as prior P2. The reader is referred to Section 4.1 for a full explanation of why prior P2 was chosen in preference to prior P1.

3.3. Sampling distribution for the missing data values

The sampling of the missing data values from their full conditional distribution utilises the properties of multivariate Gaussian distributions. This procedure has the constraint that for every time step there must be at least one site with an observed data point. For each time step t (classified in state k) which has missing data, the vector of missing data $\mathbf{Y}_t^{\text{mis}}$ with $m < r$ values is sampled conditioned on the known values of $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ and the vector of observed values for that time step, $\mathbf{Y}_t^{\text{obs}}$, such that:

$$\mathbf{Y}_t^{\text{mis}} | \mathbf{Y}_t^{\text{obs}}, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \sim N_m(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \tag{13}$$

where the values for $\boldsymbol{\mu}_m$ and $\boldsymbol{\Sigma}_m$ are derived using $\mathbf{Y}_t^{\text{obs}}$, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$, as given by Tong (1990) and re-iterated here:

$$\begin{aligned} \boldsymbol{\mu}_m &= \boldsymbol{\mu}_k^{\text{mis}} + \boldsymbol{\Sigma}_k^{\text{mis|obs}} (\boldsymbol{\Sigma}_k^{\text{obs}})^{-1} (\mathbf{Y}_t^{\text{obs}} - \boldsymbol{\mu}_k^{\text{obs}}) \\ \boldsymbol{\Sigma}_m &= \boldsymbol{\Sigma}_k^{\text{mis}} - \boldsymbol{\Sigma}_k^{\text{mis|obs}} (\boldsymbol{\Sigma}_k^{\text{obs}})^{-1} \boldsymbol{\Sigma}_k^{\text{obs|mis}} \end{aligned} \tag{14}$$

with the partitioning of $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ to obtain the values for $\boldsymbol{\mu}_k^{\text{mis}}$, $\boldsymbol{\Sigma}_k^{\text{mis|obs}}$, etc. undertaken in the following fashion:

$$\boldsymbol{\mu}_k = \begin{bmatrix} \boldsymbol{\mu}_k^{\text{mis}} \\ \boldsymbol{\mu}_k^{\text{obs}} \end{bmatrix} \qquad \boldsymbol{\Sigma}_k = \begin{bmatrix} \boldsymbol{\Sigma}_k^{\text{mis}} & \boldsymbol{\Sigma}_k^{\text{mis|obs}} \\ \boldsymbol{\Sigma}_k^{\text{obs|mis}} & \boldsymbol{\Sigma}_k^{\text{obs}} \end{bmatrix} \tag{15}$$

3.4. Initialising the parameter vector

In applying the Gibbs sampler the parameter vector must be first initialised with arbitrary starting values. In this implementation a heuristic procedure based on the method for the single site version provided in Thyer and Kuczera (2000) with some modifications to accommodate the multi-site framework and the inclusion of missing data was used. The first step is to estimate the missing data values using the expected value based on correlations with the observed data from the multiple sites. Next, the hidden state time series is estimated by combining the rainfall information from all the multiple sites. First the rainfall values from each site were smoothed using a five-year moving average filter. These smoothed values were then standardized using their mean and standard deviation. A single time series was then created by summing the standardized smoothed values from each site at each time step. This time series is considered to represent an estimate of the relative wetness or dryness of the region as indicated by the multi-site rainfall data. Hence if a particular value of this time series was positive, the hidden state was classified as wet, if it was negative it was classified as dry. For $\mathbf{P}$ the starting values were sampled using the method given by Chib (1996). Starting values for $\boldsymbol{\mu}_{\mathrm{W}}$, $\boldsymbol{\mu}_{\mathrm{D}}$ and $\boldsymbol{\Sigma}_{\mathrm{W}}$, $\boldsymbol{\Sigma}_{\mathrm{D}}$ were calculated from the data sets, $\mathbf{Y}_{\mathrm{W}}$ and $\mathbf{Y}_{\mathrm{D}}$ derived from the complete data series, $\mathbf{Y}_N$, using the estimated hidden state time series.

3.5. Assessing convergence

Once initialised, the Gibbs sampler is allowed to continuously sample until the Markov chain induced by the sampler has converged to a stationary distribution. The most critical issue in implementing the Gibbs sampler is how to determine if convergence has been achieved. In this study multiple chains were allowed to independently explore the parameter space and a variety of diagnostic tools were monitored for signs of convergence failure. This approach, as used by Thyer and Kuczera (2000) for the single site case, was advocated by Cowles and Carlin (1996). The diagnostics included inspection of the percentiles of the sample distributions for all parameters and the R statistic (as defined by Gelman et al. (1995)), which was calculated to ensure the multiple chains were mixing properly in the parameter space.

In the multi-site case additional diagnostics were required to determine whether a particular chain was caught in a computational trapping state. Diagrams showing the Gibbs samples for each individual parameter and a diagnostic referred to as the hidden state frequency (HSF) time series were also examined. During iterations of the Gibbs sampler the HSF time series plots the relative proportion of allocation to each state for each year in the time series thus far in the iterative sequence. Essentially, the HSF time series is equivalent to the posterior of the hidden state time series but it is simply prior to the Gibbs sampler achieving convergence. By comparing the HSF time series for an individual chain to the HSF time series for all the other chains combined provided immediate and unambiguous feedback when a particular chain became stuck in a trapping state. If all the diagnostic tools indicated convergence had been achieved, the remaining simulated output was treated as samples from the posterior distribution.

3.6. Other implementation issues

In this application 10 chains were used. The reduction from the 100 chains used in single site case (Thyer and Kuczera, 2000) was found to be more computationally efficient and was not found to affect the inferences. Once convergence had been achieved, the last 1000 samples were used from each chain, producing 10,000 samples, which provides a good approximation to the posterior.

Another consideration in formulations like the two-state HMM is that the posterior distributions can have two modes which are mirror images of each other, referred to as *aliasing* (Gelman et al., 1995). During Gibbs iterations the chains can move between each mode, which can significantly decrease the rate of convergence. In the HMM this is where the wet state parameters can become the dry state parameters and vice versa. To remove the effects of aliasing, the constraint that the sampled wet state mean vector is always greater than the dry state mean vector was enforced.

4. Verification of model calibration procedure using synthetic data

In this section the Gibbs sampler will be used to calibrate the multi-site HMM using synthetic data generated by the multi-site HMM. The aim is to gain an understanding about what factors adversely affect the ability of the Gibbs sampler to identify the true parameter values, and therefore the true underlying persistence structure of the synthetic data. The main issue that will be investigated is how adding more sites to the analysis influences the ability of the Gibbs sampler to identify the true parameter values.

When an extra site is included in the analysis there are a plethora of factors that could affect whether the true synthetic parameter values can be recovered. In this study only the influence of the correlation between the sites will be examined. Multi-site synthetic rainfall data was generated using the HMM with the same hidden state time series and the same wet and dry state rainfall parameters at each site. The synthetic parameter values used for the wet and dry state mean and standard deviation for all the sites are given in Table 1. For different sets of

Table 1
State rainfall parameters used to generate multi-site synthetic data

Parameter	Synthetic value
$u_W(\sigma_W)$	1250 (300)
$u_D(\sigma_D)$	850 (250)
p_{WD}	0.15
p_{DW}	0.08

multi-site data the spatial correlation between the sites was varied, hence different wet and dry state covariance matrices were produced. Table 1 also shows the values used for the transition probabilities. These synthetic parameter values were chosen because they were typical of the posterior expected values obtained from a real data site with a well-identified two-state persistence structure (refer to the companion paper Thyer and Kuczera (2003).

In the following analysis synthetic replicates of 300 years were generated. Using the values for the transition probabilities the expected number of data in the wet $E(n_W)$ and dry $E(n_D)$ states and the expected number of transitions from either state to the other $E(\text{trans.})$ can be calculated for a time series of a given length (Thyer, 2001). A length of 300 years was chosen because for the parameter values given in Table 1 this results in $E(n_W) = 105$, $E(n_D) = 195$ and $E(\text{trans.}) = 31$, which is considered to be an adequate number of wet and dry periods to identify the synthetic parameter values. The Gibbs sampler was then applied to these synthetic replicates to see if the posteriors could identify the true parameter values.

4.1. Development of P2 prior specification

To generate the first set (S1) of multi-site synthetic data the wet and dry state correlations between every site was set to 0.95. Fig. 2 compares percentile box plots to compare the transition probability posteriors for the analyses with the number of sites varying from 1 to 5, using Robert's prior P1. The transition probability posteriors were chosen because if these posteriors indicated the transition probabilities were well identified then the remainder of parameters would also be well identified. For both p_{WD} and p_{DW} there is a clear trend of increasing posterior variance with increasing number of sites. This seems an intuitive result because when more sites are included in the analysis, more parameters are required to be identified. Every time the number of sites is increased by one to a total of r sites, the number of parameters increases by $2(r+1)$. When the number of sites increases from four to five this represents 12 additional parameters. If there is little new rainfall information from the extra site because the sites are highly correlated then it would be expected that the parameter uncertainty would increase, as is the case in Fig. 2.

One would expect that this increase in parameter uncertainty (as measured by the increased posterior variance) due to an increasing the number of sites should be able to be offset by a decrease in the prior variance of the parameters. To investigate this the prior variance on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}^{-1}$ was decreased by increasing ν_0. Given the influence of the priors on the posteriors (Eq. (6)) and the interrelationship between the posteriors of each parameter through the conditional posteriors (Eq. (8)) it would be expected that if the prior variance of the $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}^{-1}$ was decreased then the posterior variance on the p_{WD} and p_{DW} would also decrease. However, the results were not as expected. Fig. 3(a) and (b) shows that when ν_0 was increased for a four site analysis with prior P1 the posterior variance of p_{WD} and p_{DW} clearly did not decrease with a decrease in the prior variance of the $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}^{-1}$. To understand why this occurred the posteriors for σ_W and σ_D for one site is compared to the P1 prior for the two cases of $\nu_0 = 6$ and $\nu_0 = 100$ in Fig. 4(a) and (b). For the $\nu_0 = 6$ case the σ_D posterior is clearly bimodal, while for $\nu_0 = 100$ both the σ_D and σ_W posteriors are bimodal. In both Fig. 4(a) and (b) the mode of the σ_D posterior with the lower parameter value coincides with the mode of the prior P1 while in Fig. 4(b) the mode of the σ_W posterior with the higher parameter value coincides with the true synthetic parameter value. This suggests that one mode in the posteriors is induced by the prior P1 while the other is induced by the data. The question is: Why does prior P1 induce a different mode in the σ_k posteriors to the data? The explanation is that from Eq. (11) it is known that for prior P1 the expected value for $\boldsymbol{\Sigma}_k$ is $\mathbf{S}/2$ as $\nu_0 \rightarrow \infty$. Hence, the expected prior value for σ_k is the empirical estimate of the standard deviation for the entire time series s_y divided by $\sqrt{2}$. For the case of $\nu_0 = 100$ (Fig. 4(b)) the bimodal posteriors indicate that some of the Gibbs sampling chains converge to this prior mode for σ_k of

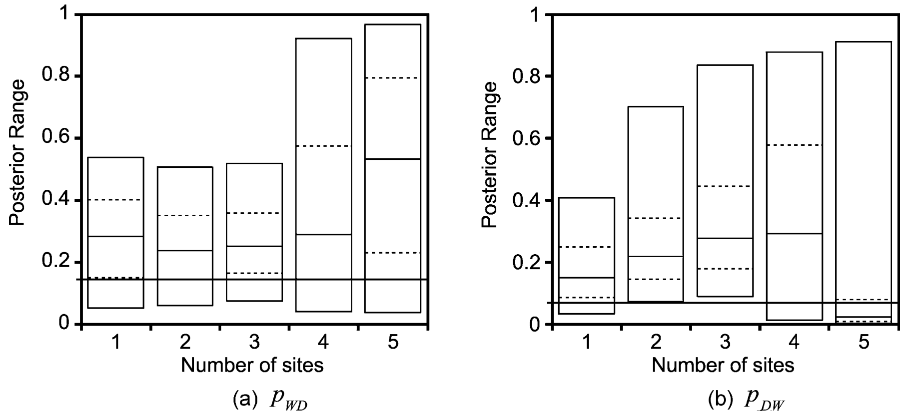


Fig. 2. Posteriors of the transition probabilities for synthetic series set S1 (0.95 correlation between sites) for a varying numbers of sites (1–5) with prior P1. Solid horizontal line indicates the true synthetic parameter value. In these percentile box plots the bottom and top of the box structure corresponds to the 5th and 95th percentiles, the middle solid line is the median (50th percentile), and the two dashed lines are the 25th and 75th percentiles of the posterior.

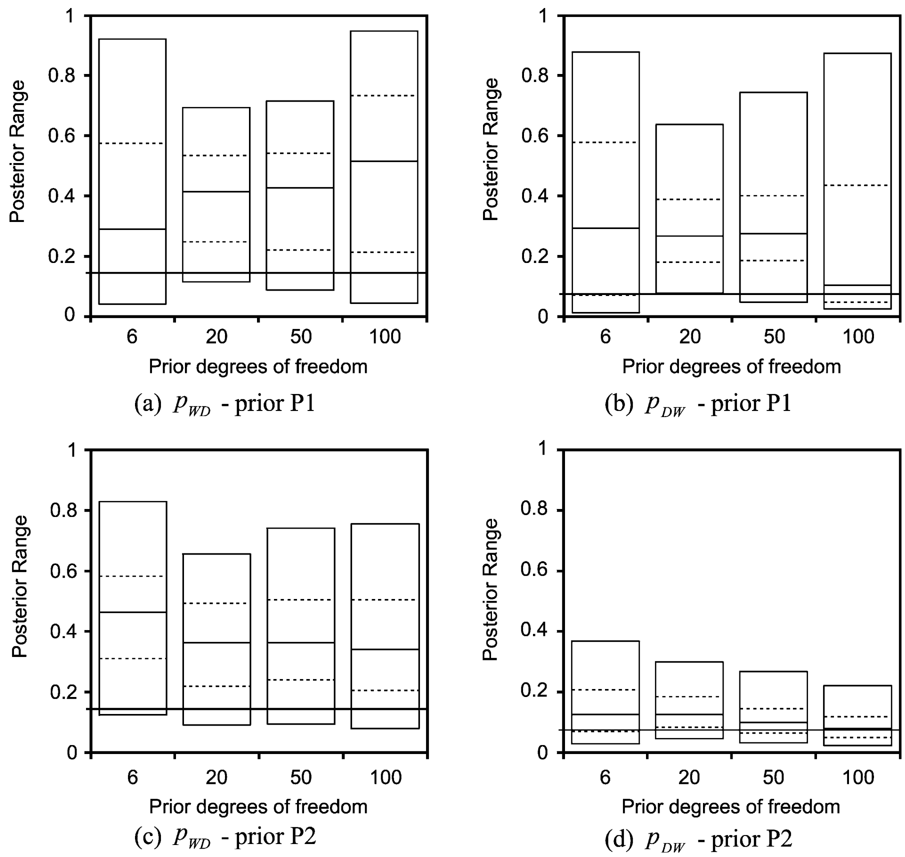


Fig. 3. Posteriors of the transition probabilities for synthetic series set S1 (0.95 correlation between sites) for the four-site analysis and varying prior degrees of freedom (a) and (b) are with prior P1, (c) and (d) are with prior P2. Solid horizontal line indicates the true synthetic parameter value. Refer to Fig. 2 for explanation of the box structure.

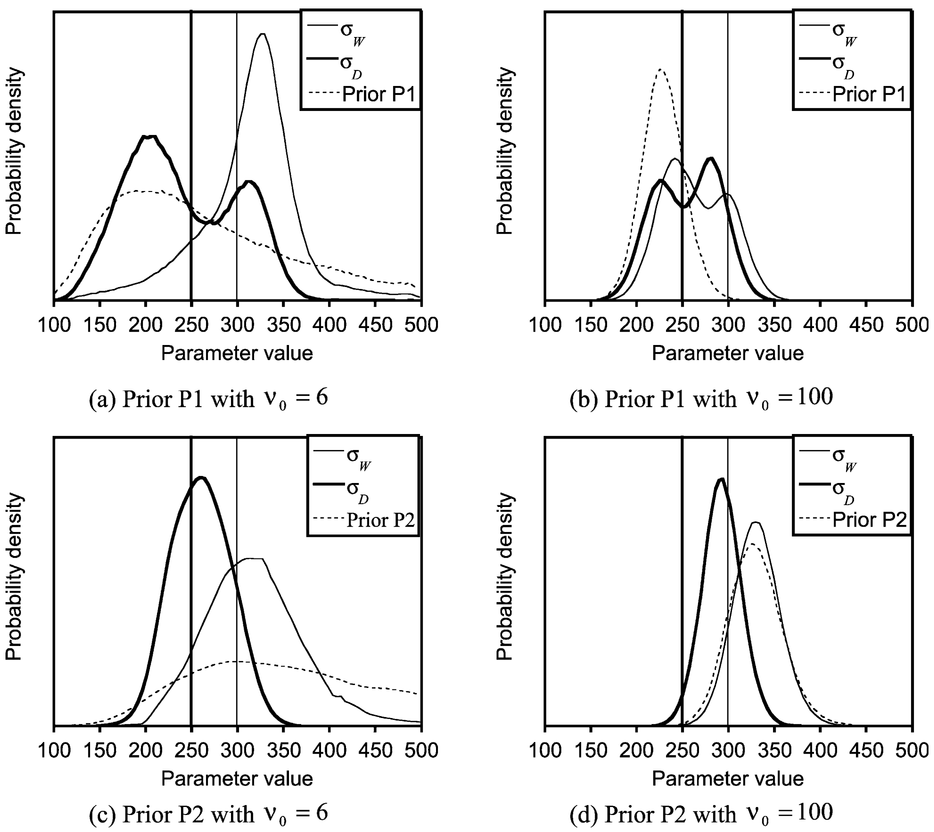


Fig. 4. Comparison of the σ_W and σ_D posteriors to the P1 and P2 priors with varying prior degrees of freedom ν_0 for one site from the four-site analysis of synthetic series set S1 (0.95 correlation between sites). Solid vertical line represents the true synthetic parameter values, $\sigma_W = 300$ and $\sigma_D = 250$.

$s_y/\sqrt{2}$ despite the fact that data may not have a true synthetic parameter value for σ_W or σ_D of $s_y/\sqrt{2}$. The high R statistic for this case also indicated the chains were not mixing properly and converging to two different modes. Even for the case of $\nu_0 = 6$ (Fig. 4(a)) the bimodal σ_D posterior indicates some of the chains are converging to the prior mode for σ_k of $s_y/\sqrt{2}$. Inspection of the $\boldsymbol{\mu}_W$ and $\boldsymbol{\mu}_D$ posteriors for this $\nu_0 = 6$ case revealed they were also bimodal. This apparent aberration where bimodal posteriors are induced by the prior suggests that the wisdom of specifying a prior expectation for $\boldsymbol{\Sigma}_k$ of $\mathbf{S}/2$ should be questioned. In the context of wet and dry state annual rainfall it is rarely the case that the expected posterior value for σ_W and σ_D was $s_y/\sqrt{2}$. Results from the single site analysis (Thyer and Kuczera, 2000) indicate σ_W was often significantly larger than s_y.

These results led to an alternative specification for the hyperparameter $\mathbf{W}_0$ known as prior P2 (refer Eq. (12)) to be trialled. In this context, the prior P2 implies a more intuitive prior expectation for $\boldsymbol{\Sigma}_k$ of $\mathbf{S}$ as $\nu_0 \to \infty$ while ensuring that for low values of ν_0 the mode of the prior density was close to $\mathbf{S}$. The results for prior P2 when the ν_0 was increased are shown in Fig. 3(c) and (d). Notice the difference from Fig. 3(a) and (b). For p_{WD} there is only a slight decrease and for p_{DW} there is clear decrease in the posterior variance as the prior variance is decreased. This is because prior P2 does not produce bimodal σ_W and σ_D posteriors, as shown in Fig. 4(c) and (d). Also, notice how for the $\nu_0 = 6$ case the modes of the σ_W and σ_D posteriors are closer to their true synthetic values with the prior P2 (Fig. 4(c)) than with prior P1 (Fig. 4(a)). Similarly Fig. 5 shows that the posterior of

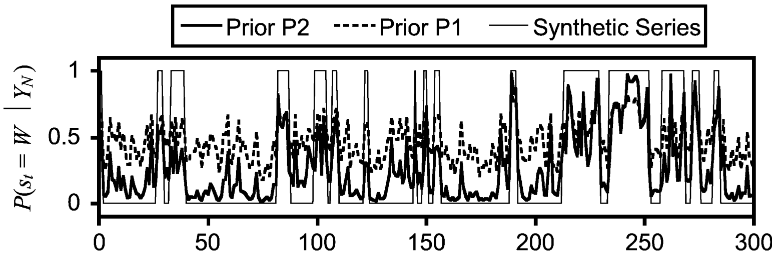


Fig. 5. Comparison of the time series plots of the posterior probability of a year being classified as wet $P(s_t = W|\mathbf{Y}_N)$ for the four-site analysis of synthetic series set S1 (0.95 correlation between sites) using prior P1 and P2 with $\nu_0 = 6$ to the true synthetic hidden state series.

the hidden state time series is closer to true synthetic time series with prior P2 than with prior P1. Based on these results it was concluded that prior P2 represents a more intuitive prior specification, which does not induce an unwanted mode in the posterior. Hence it was adopted for use in all further analysis.

For the P2 prior specification a new issue arises—it is not immediately clear what to use for ν_0. Robert (1996) gave no rationale for choosing $\nu_0 = 6$ for his prior P1. Ideally, the prior should be as diffuse as possible to minimise its influence on the posterior inferences. Given that the prior must be proper, a minimum value of $\nu_0 = r + 2$ was adopted. Using this minimum value the prior distribution was generated for a differing number of sites. It was found that as the number of sites increased the prior variance for the mean and standard deviation of each site also increased. For the prior variance to be constant as the number of sites increased it was necessary to increment ν_0 as shown in Table 2. This scheme was used in all further analysis with the synthetic series to ensure the prior had minimal influence when comparing posteriors for differing numbers of sites.

4.2. Influence of spatial correlation between sites

Fig. 6 shows the transition probability posteriors for synthetic series S1 for a varying number of sites using prior P2. Set S1 has a high correlation of 0.95. For p_{WD} there is a clear increase in posterior variance as the number of sites increases from one to four, and it then decreases for five sites. For p_{DW} the general trend is a slight decrease in the posterior variance. Fig. 7 shows the transition probability posteriors for synthetic series S2, which was generated with no correlation between the sites (independent). There is a definite decrease in the posterior variance for when the number of sites increases. For the real rainfall data used in the companion paper Thyer and Kuczera, (2003) the average spatial correlation is approximately 0.8. Fig. 8 shows the posteriors for synthetic series S3, which was generated using a correlation of 0.8 between the sites. For p_{WD} the posterior variance decreases as up to four sites are added and increases for the five-site analysis, whereas for p_{DW} the posterior variance decreases.

The general trend from Figs. 6–8 indicates that the spatial correlation between the sites has an influence on the posterior variance of the transition probabilities when more sites are included in the analysis. The higher the spatial correlation the more likely the posterior variance of the transition probabilities will increase as more sites are included in the analysis. In comparison, when there was no correlation between sites the posterior variance for both p_{WD} and p_{DW} decreased as more sites were added.

5. Discussion

In Section 1 it was stated that one of the motivations for developing the multi-site HMM was to exploit the potential benefit of space-for-time

Table 2
Prior degrees of freedom to give consistent prior variance for mean vector and covariance matrix as number of sites increases

Number of sites, r	Prior degrees of freedom, ν_0
1	$(r + 2) = 3$
2	$(r + 3) = 5$
3	$(r + 3) = 6$
4	$(r + 4) = 8$
5	$(r + 4) = 9$

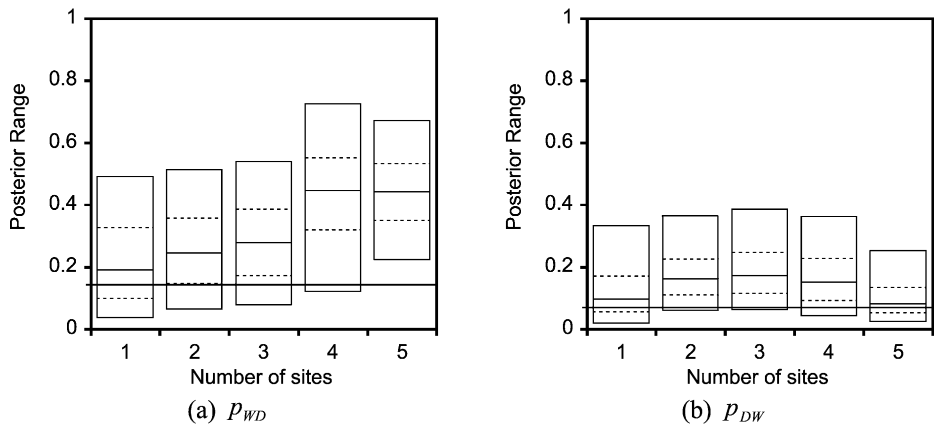


Fig. 6. Posteriors of the transition probabilities for synthetic series set S1 (0.95 correlation between sites) with varying number of sites (1–5) and prior P2 with varying prior degrees of freedom, as given in Table 2. Solid horizontal line indicates the true synthetic parameter value. Refer to Fig. 2 for explanation of the box structure.

substitution. It was believed that the extra information contained within multi-site data would make the two-state persistence structure easier to identify. From the results of the synthetic analysis presented in Figs. 6–8 it can be seen that the benefits of space-for-time substitution are dependent on the spatial correlation of multi-site data. When there was no spatial correlation the benefits of space-for-time substitution are apparent. However, when the spatial correlation is high the benefits of space-for-time substitution are significantly reduced. This is because when a single extra site is added to the analysis for a total of r sites, the number of parameters increases by $2(r+1)$. If the rainfall data from that site is highly correlated with other sites then there is little extra information provided in the multi-site rainfall data to identify these extra parameters.

Figs. 6–8 show that in some cases the true synthetic parameter value for p_{WD} is outside the 5 and 95% limits of the posterior, more so for cases with highly correlated data. There are two possible explanations for this result: Firstly, because only one replicate of 300 years was used the effects of sampling variability means that this replicate could be one of the 10% whose true synthetic parameter value is outside the 5 and 95% limits of the posterior. The other possible explanation is that the calibration procedure presented for the multi-site HMM has

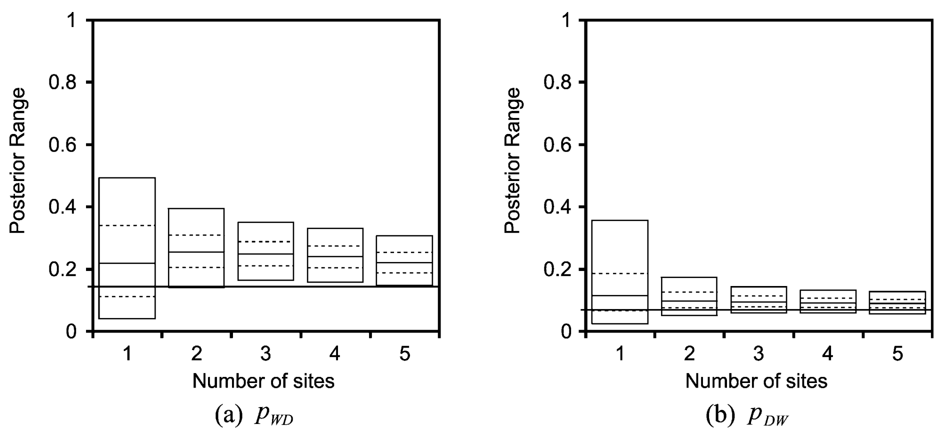
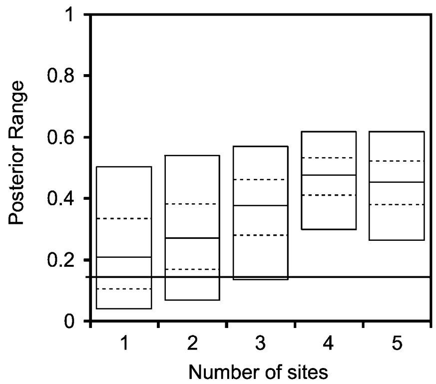
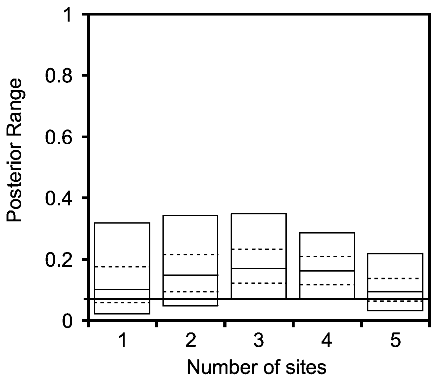


Fig. 7. Posteriors of the transition probabilities for synthetic series set S2 (independent sites—no correlation between sites) for varying number of sites with prior P2 and varying prior degrees of freedom, as given in Table 2. Solid horizontal line indicates the true synthetic parameter value. Refer to Fig. 2 for explanation of the box structure.



(a) p_{WD}



(b) p_{DW}

Fig. 8. Posteriors of the transition probabilities for synthetic series set S3 (0.8 correlation between all sites) for varying number of sites with prior P2 and varying prior degrees of freedom, as given in Table 2. Solid horizontal line indicates the true synthetic parameter value. Refer to Fig. 2 for explanation of the box structure.

difficulty recovering the true synthetic parameter value as more highly correlated sites are included in the analysis. Given these results, it is strongly recommended that prior to calibrating the multi-site HMM using the Bayesian approach outlined in this study the spatial correlation between all the sites should be determined. If the sites are highly correlated, then the procedure should be used with caution.

The limitation associated with highly correlated sites needs further work. Such work could consider alternative multi-site HMM parameterisations in which the covariance structure is constrained so as to benefit from space-for-time substitution in multi-site analysis. For example, following the approach of Hughes et al. (1999) the correlation can be expressed as a function of the distance and direction between sites.

Another motivation for the multi-site framework arose from its ability to handle missing data which enables greater utilisation of the available rainfall information. The procedure for handling missing data presented does have the following limitations. Firstly, for every time step there must be at least one site with an observed data point. Secondly, as each missing data point is modelled as an additional model parameter if one or more sites have a large number of missing data points then the additional uncertainty introduced by these extra parameters may mean the inference of the long-term persistence structure is not enhanced by including these sites. McDonald and Zucchini (1997) present an approach for computing the likelihood function of a HMM in the presence of missing data which is not subject to the above constraints. Application of this procedure using a Bayesian approach would require the use of another MCMC method, the Metropolis algorithm, instead of the Gibbs sampler and is the subject of ongoing research.

In this study synthetic data was used to investigate the effects of only one factor, the influence of the spatial correlation between sites. In a multi-site context there are several more factors that will likely affect whether the two-state persistence present in synthetic HMM data can be identified when more sites are added to the analysis. These include whether the site has the same hidden state time series and the separation of the wet and dry states. Future research should include further analysis with synthetic data to understand the influence of these factors on the identification of the two-state persistence structure.

A major problem with generalizing the results and trends from synthetic data is the issue of sampling variability. A synthetic replicate of finite length represents only one sample from an overall population. Hence, due to sampling variability two different replications of a synthetic series generated from the same process may give results with completely different trends. However, in this study the use of one replicate is justifiable because only one example is needed to demonstrate that prior P1 was found to be unsuitable for this application.

The analysis that led to development of the P2 prior specification highlights the crucial role of the priors in Bayesian MCMC methods. The existence of computational trapping states and the bimodal posteriors induced by the P1 prior (Fig. 4) are evidence of the potential dangers of using inappropriate prior specifications in the Gibbs sampling framework. Importantly, it illustrates that because of the relationship of the conditional posteriors, misspecification of the prior for only a subset of the parameters can lead to misleading inferences for the entire parameter set.

Ideally noninformative priors should be used to ensure minimal prior influence on the inferences. However, as this is not possible for the multi-site HMM it was necessary to develop a diffuse informative prior formulation (prior P2) that is not independent of the data. The use of a data-based priors is not ideal in a Bayesian context, as priors are supposed to represent information available prior to inspection of the data. The justification for the P2 prior is practically motivated. The P2 specification overcomes the problem of computational trapping states when using the Gibbs sampler and avoids the artefact of bimodal posteriors when using the P1 prior formulation suggested by Robert (1996). Without the use of diffuse data-based priors the quantification of parameter uncertainty via the Gibbs sampler and its attendant insights would not be possible.

The problem of trapping states needs to be put in perspective. In our study trapping states arose when the Gibbs sampler visited the small region of the parameter space where one-state behaviour is dominant. Without a diffuse data-based prior such as P2 the Gibbs sampler gets 'stuck' in this small region of parameter space. Nonetheless the Gibbs sampler does visit the bulk of the parameter space without the need of such assistance. The art in formulating the prior is therefore to make it as diffuse as possible while avoiding trapping states.

The important finding of this study is that the analyst needs to be aware that diffuse data-based priors may be required to help the Gibbs sampler avoid computational trapping states in some regions of parameter space. It is therefore essential that the prior and posterior distributions be compared to ensure that the prior is relatively diffuse compared to the posterior in order that the posterior be dominated by the likelihood function. The challenge remains to find a reparameterisation of the multi-site HMM that reduces the dependence of the Gibbs sampler on data-based priors.

6. Conclusions

A Bayesian approach for calibrating a multi-site HMM to long-term rainfall time series was presented. The HMM framework is attractive for modelling long-term persistence in rainfall time series because it has an explicit mechanism to emulate the influence of the dominant physical processes on long-term rainfall data. The motivation for using a multi-site approach is that large water supply catchments require multi-site simulations of hydrological inputs. In addition use of multi-site data offers the potential benefit that space-for-time substitution may lead to better identification of the long-term persistence structure; moreover it enables the development of a procedure for handling missing data.

The main focus of this paper was to describe the development of a Bayesian calibration procedure for the multi-site HMM. A MCMC method known as the Gibbs sampler was used to infer the posterior distribution of the model parameters. Several challenges had to be overcome in the implementation of this algorithm. The specification of appropriate priors for the state rainfall distributions was a crucial issue. The prior recommendations made by Robert (1996) were found to be unsuitable for this application. Using synthetic data it was shown how a suitable prior specification was developed. This paper highlights the important role of priors in Bayesian MCMC analysis and the potential pitfalls that can be encountered if the priors are not carefully chosen. A methodology for handling missing data in the multi-site framework was also presented and its limitations discussed.

Further investigation with synthetic data case studies showed that the benefit of space-for-time substitution in multi-site data is dependent on the spatial correlation of the multi-site data. The discussion outlined the unresolved issues and limitations that apply to this current implementation of the Bayesian approach for calibrating the multi-site HMM.

Acknowledgements

The authors would like to gratefully acknowledge the comments of the two anonymous reviewers. Their excellent reviews helped to sharpen these two manuscripts considerably.

References

Bellone, E., Hughes, J.P., Guttop, P., 2000. A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. Climate Research 15, 1–12.

Billio, M., Monfort, A., Robert, C.P., 1999. Bayesian estimation of switching ARMA models. Journal of Econometrics 93, 229–255.

Charles, S.P., Bates, B.B., Whetton, P.H., Hughes, J.P., 1999a. Validation of downscaling models for changed climate conditions: case study of southwestern Australia. Climate Research 12, 1–14.

Charles, S.P., Bates, B.C., Hughes, J.P., 1999b. A spatiotemporal model for downscaling precipitation occurrence and amounts. Journal of Geophysical Research 104 (D24), 31657–31669.

Chib, S., 1996. Calculating posterior distributions and modal estimates in Markov mixture models. Journal of Econometrics 75, 79–97.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis–Hastings algorithm. The American Statistician 46 (4), 327–335.

Cowles, M.K., Carlin, B.P., 1996. Markov chain Monte Carlo convergence diagnostics: a comparative review. Journal of American Statistical Association 91 (434), 883–904.

DeGroot, M.H., 1970. Optimal Statistical Decisions, McGraw-Hill, New York, p. 489.

Diebolt, J., Robert, C.P., 1994. Estimation of finite mixture distributions through Bayesian sampling. Journal of the Royal Statistical Society (Series B) 56 (2), 363–375.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.S., 1995. Bayesian Data Analysis, Chapman & Hall, London, p. 526.

Gilks, W.R., Richardson, S., Spieghalter, D.J., 1996. Markov Chain Monte Carlo in Practice, Chapman & Hall, London, p. 486.

Grayson, R.B., Argent, R.M., Nathan, R.J., McMahon, T.A., Mein, R.G., 1996. Hydrological Recipes: Estimation Techniques in Australian Hydrology, Cooperative Research Centre for Catchment Hydrology, Melbourne, Australia, p. 125.

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

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

Jackson, B., 1975. Markov mixture models for drought lengths. Water Resources Research 11 (1), 64–74.

McDonald, I.L., Zucchini, W., 1997. Hidden Markov and Other Models for Discrete Valued Time Series, Chapman & Hall, London, p. 236.

Robert, C.P., 1996. Mixtures of distributions: inference and estimation. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), Markov Chain Monte Carlo in Practice, Chapman & Hall, London, p. 441–464.

Robert, C.P., Mengersen, K.L., 1999. Reparameterisation issues in mixture modelling and their bearing on MCMC algorithms. Computational Statistics and Data Analysis 29, 325–343.

Robert, C.P., Titterington, M., 1998. Reparameterization strategies for hidden Markov models and Bayesian approaches to maximum likelihood estimation. Statistics and Computing 8 (2), 145–148.

Salas, J., 1993. Analysis and modeling of hydrologic time series. In: Maidment, D., (Ed.), Handbook of Hydrology, McGraw-Hill, New York, pp. 19.1–19.72.

Salas, J.D., Smith, R.A., 1981. Physical basis of stochastic models of annual flows. Water Resources Research 17 (2), 428–430.

Srikanthan, R., McMahon, T.A., 2000. Stochastic Generation of Climate Data: A Review, CRC for Catchment Hydrology, Monash University, Clayton, Vic, p. 42.

Thyer, M., 2001. Modelling long-term persistence in hydrological time series. PhD Thesis, Department of Civil, Surveying and Environmental Engineering, University of Newcastle, p. 360.

Thyer, M., Kuczera, G., 2000. Modelling long-term persistence in hydro-climatic time series using a hidden state Markov model. Water Resources Research 36 (11), 3301–3310.

Thyer, M., Kuczera, G., 2003. A hidden Markov model for modelling long-term persistence in multi-site rainfall time series. 2. Real data analysis. Journal of Hydrology.

Tierney, L., 1994. Markov chains for exploring posterior distributions. Annals of Statistics 22, 1701–1762.

Tong, Y.L., 1990. The Multivariate Normal Distribution, Springer, New York, p. 271.

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