Journal of Hydrology (2007) 340, 129–148

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jhydrol

# A general Bayesian framework for calibrating and evaluating stochastic models of annual multi-site hydrological data

Andrew J. Frost [a,*], Mark A. Thyer [b], R. Srikanthan [a], George Kuczera [b]

[a] *Hydrology Unit, Bureau of Meteorology, P.O. Box 1289, Melbourne 3001, Australia*
[b] *School of Engineering, University of Newcastle, University Drive, Callaghan NSW 2308, Australia*

Received 4 July 2006; received in revised form 14 February 2007; accepted 30 March 2007

**KEYWORDS**
Stochastic rainfall;
Long-term persistence;
Parameter and model uncertainty;
Hidden Markov models;
Lag-one autoregressive models;
Box–Cox transformation

**Summary** Multi-site simulation of hydrological data are required for drought risk assessment of large multi-reservoir water supply systems. In this paper, a general Bayesian framework is presented for the calibration and evaluation of multi-site hydrological data at annual timescales. Models included within this framework are the hidden Markov model (HMM) and the widely used lag-1 autoregressive (AR(1)) model. These models are extended by the inclusion of a Box–Cox transformation and a spatial correlation function in a multi-site setting. Parameter uncertainty is evaluated using Markov chain Monte Carlo techniques. Models are evaluated by their ability to reproduce a range of important extreme statistics and compared using Bayesian model selection techniques which evaluate model probabilities. The case study, using multi-site annual rainfall data situated within catchments which contribute to Sydney's main water supply, provided the following results: Firstly, in terms of model probabilities and diagnostics, the inclusion of the Box–Cox transformation was preferred. Secondly the AR(1) and HMM performed similarly, while some other proposed AR(1)/HMM models with regionally pooled parameters had greater posterior probability than these two models. The practical significance of parameter and model uncertainty was illustrated using a case study involving drought security analysis for urban water supply. It was shown that ignoring parameter uncertainty resulted in a significant overestimate of reservoir yield and an underestimation of system vulnerability to severe drought.


* Corresponding author.
*E-mail address:* a.frost@bom.gov.au (A.J. Frost).

0022-1694/$ - see front matter 
doi:10.1016/j.jhydrol.2007.03.023

## Introduction

Drought risk assessment of large multi-reservoir water supply systems relies on the ability to stochastically simulate multi-site long-term hydrological data. Identification and evaluation of an appropriate multi-site stochastic model for this task is of paramount importance for long-term water resource planning decisions. A general framework for modelling multi-site data will be developed in this paper. This framework will include several currently used multi-site models and some new approaches to compare and evaluate their performance.

Historically, numerous stochastic models have been developed to simulate multi-site long-term hydrological data. Matalas (1967) introduced a multi-site AR(1) model for modelling annual rainfall/streamflow defined by:

$$\mathbf{z}_{t+1} = \boldsymbol{\mu}_z + A(\mathbf{z}_t - \boldsymbol{\mu}_z) + \varepsilon_t \quad (1)$$

where $\mathbf{z}_t = (\mathbf{z}_{1t}, \ldots, \mathbf{z}_{dt})$ defines the vector of transformed annual rainfall/streamflow amounts over $d$ sites for a given time step $t$; $\boldsymbol{\mu}_z$ is a $d$ vector of transformed annual mean flows; $A$ is a $(d \times d)$ parameter matrix; and $\varepsilon_t$ is a $d$ vector of normal deviates randomly and independently drawn from a multivariate distribution with mean zero and covariance matrix $\Sigma$; that is,

$$\varepsilon_t \sim N(0, \Sigma) \quad (2)$$

Hydrological data often do not follow a Gaussian distribution. Matalas (1967) suggested a log transformation of streamflow data. A more general approach is to employ a Box–Cox (BC) transformation (Box and Cox, 1964). For a given time series of multi-site data, $\mathbf{Y}_T = (\mathbf{y}_1, \ldots, \mathbf{y}_T)$ where $\mathbf{y}_t = (y_{1t}, \ldots, y_{dt})$, a BC transformation is applied to the observed data value $y_{it}$ for site $i$ according to:

$$z_{it} = \begin{cases} \frac{(y_{it})^{\lambda_i} - 1}{\lambda_i} & \lambda_i \neq 0 \\ \log_e y_{it} & \lambda_i = 0 \end{cases} \quad (3)$$

where $\lambda_i$ is the BC transformation parameter associated with site $i$. This transformation is carried out independently for each site.

Matalas (1967) presented a method of moments approach to estimating the parameters of this model. Kuczera (1987) noted a practical problem with this method is that 'the estimated covariance matrix may not be positive definite, thereby preventing its decomposition, a necessary step for synthetic flow generation'. Kuczera's suggestion to overcome this problem was to use the EM algorithm which is a maximum likelihood method that ensures the covariance matrix is positive definite, even when record lengths are of unequal length.

In following years many other model structures have been applied for the synthetic generation of hydrological processes. Stationary linear autoregressive moving average (ARMA) models have found wide application (Grayson et al., 1996; Hipel and McLeod, 1994; Salas, 1993)— and are often capable of reproducing important historical statistics, including long-term related statistics such as storage and drought related statistics (Fortin et al., 2004). Even so, other models such as the fractional gaussian noise (Mandelbrot and Wallis, 1969), Broken Line (Rodriguez-Iturbe et al., 1972), Shifting Level (Klemes, 1974; Potter, 1976; Salas and Boes, 1980), and FARMA (Montanari et al., 1997) have been proposed in the hydrological literature, often with the intention of reproducing the empirically observed phenomenon of long-term persistence in hydrological series (Hurst, 1951).

The majority of this previous research does not evaluate parameter uncertainty as part of model calibration. Thyer et al. (2006) showed there is substantial parameter uncertainty due to the short data lengths of observed hydrological data for the single site models trialled. Stedinger and Taylor (1982) further developed the multi-site AR(1) model of Matalas (1967), assuming that the autoregressive Gaussian error process operates on a log transformed scale compared to the data, to incorporate approximations for parameter uncertainty. Their study showed that parameter uncertainty is more important than model uncertainty for stochastic modelling of hydrological data for the models trialled. However, their method suffers from the practical consideration reported by Kuczera (1987), that the empirical covariance may not be invertible, and also is not fully Bayesian as it does not allow for concurrent estimation of the uncertainty in the Gaussian transformation parameter (the BC transformation is used here). The change-point model was cast in a Bayesian parameter uncertainty estimation framework by Perreault et al. (2000a) and then extended to a multi-site model in Perreault et al. (2000b) for use in identification of single change-points in hydrological data series (e.g. mean/standard deviation). The shifting level model, a generalisation of the change-point model allowing multiple change-points, was further developed by Fortin et al. (2004) to use fully Bayesian parameter uncertainty estimation as opposed to method of moments (Salas and Boes, 1980). However, this has not been extended to a multi-site framework — a required input for multi-reservoir system analysis. Thus there is currently no methodology for evaluating parameter uncertainty for stochastic models of annual multi-site hydrological data. This study aims to overcome this deficiency by developing a general Bayesian parameter uncertainty estimation framework for stochastic models of annual multi-site hydrological data.

The first model structure included in this general framework is a multi-site generalization of the AR(1) model presented by Thyer et al. (2002). This is included because it is commonly used and provides a basis for evaluating alternatives. The implementation is an advance on the multi-site model presented by Kuczera (1987), Stedinger and Taylor (1982) and Matalas (1967) as full quantification of parameter uncertainty is undertaken and estimation of the parametric transformation of the data are undertaken concurrently with all other parameters. The dependence structure of the multi-site AR(1) presented here is subtly different to the previous models — it is the 'contemporaneous' AR(1) model with zero lag-1 cross-correlations (Hipel and McLeod, 1994; Salas, 1993). A major improvement over other models is the capability to generate (and calibrate to records containing) zero rainfall/streamflow in a given year — a phenomenon apparent in the more arid regions (and smaller catchments for streamflow) of Australia.

The second model structure included in this framework is a new implementation of the multi-site hidden Markov model (HMM). Thyer and Kuczera (2000) introduced a single-site two-state HMM for modelling annual hydrological time ser-

ies as an alternative to the widely used AR(1) model in an attempt to model the two-state persistence structure apparent in Australian hydrological data. Thyer and Kuczera (2003a,b) later presented a multi-site version of the HMM to enable better identification of the HMM 'hidden' state parameters. However, their implementation had several shortcomings. Firstly, the calibration procedure performed poorly if the sites were highly correlated, primarily because the number of individual inter-site correlations to be estimated grew rapidly with the number of sites. Secondly, the model was incapable of reproducing highly skewed historical data, because it did not include any transformation of data.

In this study, an alternative parameterization for the spatial correlation of the multi-site HMM is presented which considerably reduces the number of parameters and overcomes the shortcomings identified by Thyer and Kuczera (2003b). As it is included as part of a generalized modelling framework, this parameterization for the spatial correlation is also applied to the AR(1) model. Similarly, the concurrent estimation of the transformation parameter developed for the AR(1) model is also applied to the HMM. This enables the modelling of skewed data — a common feature of annual rainfall/streamflow data from Australia.

The power of multi-site analysis is its ability to utilize space-for-time substitution to overcome the limitations imposed by short data lengths of observed annual hydrological data and improve model identifiability. Casting both the HMM and AR(1) models in a generalized multi-site framework provides a far greater flexibility to evaluate alternative multi-site model parameterizations, e.g. the pooling of parameters across a region. In this study, the utility of various pooled parameter parameterisations will be evaluated.

The two models considered in this study (AR(1) and HMM) have differing assumptions regarding the dependence of one year's rainfall on those preceding it. Neither the AR(1) or HMM models satisfy the mathematical definition of persistence; that the sum of correlations over all lags is infinite (Beran, 1994, p. 6–7). The HMM was developed with the intention of capturing a different form of dependence structure to that of the AR(1) — analogous to the long-term wet and dry periods produced by phenomena such as the El-Niño Southern Oscillation. It was found in a single site study carried out by Srikanthan et al. (2001) that the HMM assumptions (e.g. two climate states) were justified for a range of sites around Australia. However, clear identification of regions where the HMM assumptions held was not possible.

Evaluation of model performance is a key feature of this generalized model framework. This comparison was undertaken using Bayesian model selection techniques and assessing the ability of the models to reproduce a range of observed statistics. To provide a practical basis for this analysis a reservoir simulation study was used to compare the models performance in terms of simulated drought risk.

The case study used to test this general framework will be rainfall data surrounding and within the Hawkesbury-Nepean catchment (the main water supply catchment for Sydney, Australia). The model performance of the multi-site HMM and AR(1), and pooled parameter variants will be compared and evaluated. This comparison will consider:

(a) if the Box–Cox transformation was required;
(b) which of the AR(1) and HMM models performed better;
(c) if pooling parameters across a region improved model identifiability; and
(d) the impact of evaluating parameter and model uncertainty on drought risks.

This paper is organized as follows: Section ''Stochastic models for annual multi-site hydrologic data'' describes the features of the stochastic models included in the generalized framework. Section ''Model calibration procedure'' describes the Bayesian calibration procedure, including the implementation of the Markov chain Monte Carlo (MCMC) methods used to evaluate parameter uncertainty. The hydrological data from the case study is described in Section ''Hydrological data''. Section ''Evaluation of model performance'' outlines the range of techniques used to evaluate the model performance. The results are presented in ''Results'' and discussed in 'Discussion'''. The discussion focuses on the implications of this paper for the future research for the modelling of multi-site stochastic hydrological data. Finally, the conclusions are presented.

## Stochastic models for annual multi-site hydrologic data

One of the goals of this paper is to develop a generalised framework for stochastic modelling of annual multi-site hydrological data, which includes both the AR(1) and HMM as special cases with different parameterisations for the temporal dependence. This section outlines the key features of the model parameterisations; further details of the individual models tested, including the pooled parameter variants will be given in Section ''Model specification''.

### Transformation of data

Both the HMM and AR(1) model have underlying Gaussian assumptions. Therefore, the BC transformation is employed here to enable modelling of non-Gaussian data. Thyer et al. (2002) recognized that when applying the BC transformation the constraint $z_{it}\lambda_i + 1 \geqslant 0$ applies during calibration and simulation. This constraint is required to ensure that real numbers $y_{it} > 0$ can be simulated for non-zero values in (1). Thyer et al. (2002) deal with this constraint by applying a truncated Gaussian distribution (according to $z_{it}\lambda_i + 1 > 0$) — with there being no probability/density associated with $z_{it}\lambda_i + 1 \leqslant 0$ — the truncated probability mass was used to normalize the distribution. The BC transformation used here differs from Thyer et al. (2002) according to:

$$y_{it} = \begin{cases} (z_{it}\lambda_i + 1)^{1/\lambda_i} & \lambda_i \neq 0, (z_{it}\lambda_i + 1)^{1/\lambda_i} > 0 \\ 0 & \lambda_i \neq 0, (z_{it}\lambda_i + 1) = 0 \\ \exp(z_{it}) & \lambda_i = 0 \end{cases} \quad (4)$$

Note that $(z_{it}\lambda_i + 1)^{1/\lambda_i} < 0$ is not accounted for here as the error process is truncated so as to ensure positive rainfall/streamflow, as detailed in the following section. This truncation method was applied to both the HMM and AR(1).

### Multi-site AR(1)

Grouping the transformed rainfall/streamflow variables $z_{it}$ as vectors $\mathbf{z}_t = (z_{1t}, \ldots, z_{dt})$, and assuming a lag-1 auto-correlated process on $\mathbf{z}_t$:

$$\mathbf{z}_t = \boldsymbol{\mu} + \boldsymbol{\varphi}(\mathbf{z}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t; \quad \boldsymbol{\varepsilon}_t \sim \mathrm{MTN}\left(0, \sum : \mathbf{z}_t\lambda + 1 \geqslant 0\right) \tag{5}$$

where $\boldsymbol{\varphi} = D(\phi_1, \ldots, \phi_d)$ is a $d \times d$ matrix with $D$ denoting a diagonal matrix containing lag-1 auto-correlation coefficients at each site; $\boldsymbol{\lambda}$ is a $d$ vector of BC transformation parameters for each site; $\boldsymbol{\mu} = [\mu_i : i = 1, \ldots, d]$ is the mean parameter vector; $\boldsymbol{\varepsilon}_t$ is an i.d. error term given by a mixed truncated Normal (Gaussian) distribution denoted MTN() with zero mean and spatial covariance matrix $\sum = \left[\rho_{ij}(\sigma_z)_i(\sigma_z)_j : i, j = 1, \ldots, d\right]$. Here $(\sigma_z)_i$ denotes the standard deviation parameter for site $i$ (in $z$-space), while the site-to-site correlation coefficient $\rho_{ij}$ will be defined subsequently. The mixed truncated normal is required given the real constraint $y \geqslant 0$, and is described in Section ''Likelihood function''.

### Multi-site HMM

The two-state HMM framework assumes the climate is in one of two states: wet (W) or dry (D). The climate state at year $t$, $r_t$, is simulated using a Markov chain process. This process is characterized by the set of one-step transition probabilities, defined as follows:

$$P = \{p_{ij}\} = \Pr(r_t = j | r_{t-1} = r),\ r, j = (\mathrm{W\ or\ D}) \tag{6}$$

Residence time in each state is governed by these state transition probabilities — the probability of changing from one state to another from one year to the next. This induces a temporal correlation similar to the ARMA(1,1) model (Akintug and Rasmussen, 2005) if there is persistence in a particular state. However, if the state does not change at all there is zero correlation. Each state $r$ = (D,W) has an independent mixed truncated Normal distribution with associated mean vector $\boldsymbol{\mu}_r = [\mu_{ir} : i = 1, \ldots, d]$, and standard deviation vector, $\boldsymbol{\sigma}_r = [\sigma_{ir} : i = 1, \ldots, d]$, for each state $r$. The number of Markovian states is chosen by the modeller — with the aforementioned studies using 2 states.

The Box—Cox transformation (4) used for the AR(1) model can also be used for the HMM — with associated state/site specific BC transformation parameter. Again, as for the AR(1), there is also a mass truncation probability associated with zero.

The multi-site HMM framework assumes that there is a single climate state across a multi-site region and the probability of transitions between states is governed by a Markovian process as provided in (6). If the transition probabilities sum to one, $p_{WD} + p_{DW} = 1$ then there is no persistence between states and the HMM degenerates to a mixture model (Lambert et al., 2003). The HMM can also produce the behaviour observed in shifting mean and change-point models (low transition probabilities can produce a single change-point within a series). Indeed, certain parameterizations of these models can be derived from the generic HMM form (Fortin et al., 2004; Sveinsson et al., 2003).

### Spatial parameterization

Previously Thyer and Kuczera (2003b) fitted individual correlation parameters $\rho_{ij}$ to model the spatial correlation. In this study, a different parameterization was used to reduce the number of model parameters $\rho_{ij}$. An exponential correlation decay function is assumed;

$$\rho_{ij} = \exp(-\|i - j\|/\omega), \quad \omega > 0 \tag{7}$$

where $\|i - j\|$ represents the absolute distance between sites $i$ and $j$, and $\omega$ is the 'correlation length' parameter. This functional correlation form has been used in rainfall modelling studies (e.g. Sanso and Guenni, 2000). This paper examines the application of this structure to rainfall data using the distance between the site locations. This spatial parameterisation was applied to both the HMM and AR(1) model.

## Model calibration procedure

For model calibration, a Bayesian framework is used to infer the posterior distribution of the model parameters, for the given time series data, $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_T)$, referred to as $p(\boldsymbol{\theta}|\mathbf{Y})$. The inference of the posterior distribution defines the Bayesian technique — for an introduction see Lee (1989) or Gelman et al. (2004). Unlike approaches using a single parameter estimate, use of the posterior results in the modeller not overstating the confidence of model predictions.

The posterior distribution is defined by Bayes equation:

$$p(\boldsymbol{\theta}|\mathbf{Y}) = \frac{f(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} f(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}} \propto f(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{8}$$

The prior $p(\boldsymbol{\theta})$ defines the distribution of parameters before calibration — and is user chosen. The likelihood function $f(\mathbf{Y}|\boldsymbol{\theta})$, the sampling distribution of $\mathbf{Y}$ given $\boldsymbol{\theta}$, defines the fit to the data for a particular parameter set (and model) and is therefore related to the model formulation. This likelihood reflects the influence of the data on parameter identification, with the posterior distribution therefore providing an update on the prior belief after taking into account the data.

### Markov chain Monte Carlo sampling: Metropolis—Hastings sampler

Sampling methods known as Markov chain Monte Carlo (MCMC) are employed to draw samples from the posterior distribution. Such sampling methods are useful when it is not possible to derive an analytical expression to sample from the posterior distribution, as is the case for the HMM and AR(1) models specified here. In cases where the likelihood can be calculated (as is the case here) an MCMC method known as the Metropolis—Hastings sampler can be applied. The algorithm used in this study is detailed within Appendix A.

Attempts have been made to improve the efficiency of the Metropolis—Hastings algorithm (e.g. Andrieu and Robert, 2001). The problem with some approaches such as the periodic updating of the covariance matrix used in a preliminary stage here (see Appendix A), and in other studies

(Haario et al., 1999; Kuczera and Parent, 1998), or with the use of the as yet unproven updating method employed by Vrugt et al. (2003), is that the chain can lose its 'detailed balance' property and proofs for convergence may not hold. The Haario et al. (2001) sampler for production of samples was chosen here as it is proven that this adaptive sampler is ergodic, that is estimates taken from it converge to the posterior.

## Gibbs sampler

It is noted that in previous application of the HMM by Thyer and Kuczera (2003a) that an alternative MCMC sampling method was employed, namely the Gibbs sampler (Geman and Geman, 1984). The advantage of the Gibbs sampler (when it is reasonably simple to compute and sample from the required conditional distributions) is that issues regarding choice of jump distribution are no longer important (Fortin et al., 2004). Rather, choice of priors which allow calculation of conditional distributions are mandatory. In some cases calculation of the full likelihood (required for the Metropolis–Hastings sampler) may not be analytically possible, which therefore would require the Gibbs sampler, or some combination of the two. The Metropolis–Hastings sampler was used in this study because it was found to overcome the problem of ''trapping states'' which hindered convergence of the Gibbs sampler as highlighted by Thyer and Kuczera (2003b).

## Transformation of mean and standard deviation parameters

Thyer et al. (2002) noted there is a practical problem when using the Metropolis–Hastings sampler with Gaussian jump distribution to estimate the posterior of the single site AR(1) model employing the BC transformation; very slow convergence occurs due to the highly correlated and irregular distributional shape of the posterior. To overcome this problem they developed parameter transformations for the mean and standard deviations based on first order approximations:

$$\mu_{ir} = \begin{cases} \frac{m_{ir}^{\lambda_i}-1}{\lambda_i} & \lambda_i \neq 0 \\ \log_e m_{ir}^{\lambda_i} & \lambda_i = 0 \end{cases} \tag{9}$$

$$\sigma_{ir} = m_{ir}^{\lambda_i - 1} s_{ir} \sqrt{1-\phi_i^2} \tag{10}$$

where $m_{ir}$ and $s_{ir}$ are the transformed mean and standard deviation parameters, respectively. This transformation was also applied throughout this study to aid convergence of the MCMC sampler. Thus the Metropolis-Hasting algorithm proceeds by sampling in the transformed parameter space of $\boldsymbol{\theta} \in (m_{ir}, s_{ir}, \ldots)$ – where $(\ldots)$ represents the other parameters which are not transformed, and then Eqs. (9) and (10) are used to transform the parameters to $(\mu_{ir}, \sigma_{ir}, \ldots)$ which are used to evaluate the likelihood function.

## Likelihood function

Use of the Metropolis–Hastings algorithm requires evaluation of the likelihood function.

### AR(1) likelihood function

Similar to single-site derivation produced in Thyer et al. (2002) Appendix A, the AR(1) likelihood is calculated according to:

$$p(\mathbf{Y}_T|\boldsymbol{\theta}) = p(\mathbf{y}_1|\boldsymbol{\theta}) \prod_{t=2}^{T} p(\mathbf{y}_t|\mathbf{Y}_{t-1}, \boldsymbol{\theta}) \tag{11}$$

where the likelihood of an observation at a single timestep, given the single timestep dependence structure and the data transformation (4), is given by:

$$\begin{aligned} p(\mathbf{y}_t|\mathbf{Y}_{t-1}, \boldsymbol{\theta}) &= p(\mathbf{y}_t|\mathbf{y}_{t-1}, \boldsymbol{\theta}) = |\partial \mathbf{z}_t / \partial \mathbf{y}_t| p(\mathbf{z}_t|\mathbf{z}_{t-1}, \boldsymbol{\theta}) \\ &= p(\mathbf{z}_t|\mathbf{z}_{t-1}, \boldsymbol{\theta}) \prod_{i=1}^{d} g(y_{it}) \end{aligned} \tag{12}$$

where:

$$g(y_{it}) = \begin{cases} y_{it}^{\lambda_i - 1} & y_{it} > 0 \\ 1 & y_{it} = 0 \end{cases}$$

Given the dependence structure (5), $p(\mathbf{z}_t|\mathbf{z}_{t-1},\boldsymbol{\theta})$ is evaluated as follows:

$$\begin{aligned} p(\mathbf{z}_t|\mathbf{z}_{t-1}, \boldsymbol{\theta}) &= p(\mathbf{z}_{t+}|\mathbf{z}_{t-1}, \boldsymbol{\theta}) P(\mathbf{z}_{t0}|\mathbf{z}_{t+}, \mathbf{z}_{t-1}, \boldsymbol{\theta}) \\ &= f_{\mathrm{MVN}}(\mathbf{z}_{t+} : \bar{\mathbf{z}}_{t+}, \Sigma_{t+}) \\ &\quad \times \int_{l_1}^{h_1} \int_{l_2}^{h_2} \cdots \int_{l_{d0}}^{h_{d0}} f_{\mathrm{MVN}}\left(\mathbf{z}_{t0} : \bar{\mathbf{z}}_{t0|+}, \Sigma_{t0|+}\right) \mathrm{d}\mathbf{z}_{t0} \end{aligned} \tag{13}$$

where $f_{\mathrm{MVN}}(\mathbf{x}:\boldsymbol{\mu}_x,\Sigma_x)$ is the multivariate Normal density function (Gelman et al., 1995, p. 474) evaluated for random vector $\mathbf{x}$, given associated mean and covariance parameters $\boldsymbol{\mu}_x$ and $\Sigma_x$. This Mixed Truncated Normal distribution (defined here) lumps the probability associated with the multivariate Normal variates satisfying $z_{it}\ \lambda_i + 1 < 0$ at $z_{it}\lambda_i + 1 = 0$.

As any subvector of a joint multivariate Normal can be written as being conditionally multivariate Normal given the remaining subvector (Gelman et al., 1995, p. 479), the vector of transformed site rainfall values $\mathbf{z}_t$ are broken into two possibly empty subsets $\mathbf{z}_t = \{\mathbf{z}_{t0}, \mathbf{z}_{t+}\}$, where $\mathbf{z}_{t0}$ and $\mathbf{z}_{t+}$ corresponding to the sites where $y_{it} = 0$ and $y_{it} > 0$, respectively. Using the relationships derived therein, the mean $\bar{\mathbf{z}}_t = \boldsymbol{\mu} + \boldsymbol{\varphi}(\mathbf{z}_{t-1} - \boldsymbol{\mu})$ vector and covariance matrix $\Sigma$ are used in calculating the unconditional parameters $\{\bar{\mathbf{z}}_{t+}, \Sigma_{t+}\}$ and conditional parameters $\{\bar{\mathbf{z}}_{t0|+}, \Sigma_{t0|+}\}$ related to the $y_{it} > 0$ and $y_{it} = 0$ values, respectively.

This integral lumps all mass relating to $y_{it} < 0$ space on $y_{it} = 0$ zero values, where the integral bounds are:

$$l_i = \begin{cases} -1/\lambda_i \\ -\infty \end{cases} \quad h_i = \begin{cases} \infty & \text{if } \lambda_i > 0 \\ -1/\lambda_i & \text{if } \lambda_i < 0 \end{cases}$$

This alteration dispenses with the need for the calculation of normalising probabilities within the likelihood for every timestep as used in the Thyer et al. (2002) formulation. The truncated mass (as opposed to density) need only be evaluated for sites/timesteps where zero rainfall/streamflow is observed.

As an analytical function for the cumulative distribution of the standard multivariate Normal distribution does not exist, the SADNRM numerical integration method introduced by Genz (1993) was used here.

**HMM likelihood function**

For the HMM evaluation of the likelihood was undertaken using the method detailed in Bengio (1999) and in this context within Frost (2004, p. 38). Due to the single timestep state dependence structure assumed with a HMM, the basis of the likelihood function is identical to that used for the AR(1) in (11). The models differ in the calculation of $p(\mathbf{y}_t|\mathbf{Y}_{t-1},\boldsymbol{\theta})$, with an iterative procedure used as follows:

$$p(r_t|\mathbf{Y}_{t-1},\boldsymbol{\theta}) = \sum_{r_{t-1}\in(W,D)} p(r_t|r_{t-1},\boldsymbol{\theta})p(r_{t-1}|\mathbf{Y}_{t-1},\boldsymbol{\theta}) \quad (14a)$$

$$p(\mathbf{y}_t|\mathbf{Y}_{t-1},\boldsymbol{\theta}) = \sum_{r_t\in(W,D)} p(\mathbf{y}_t|r_t,\boldsymbol{\theta})p(r_t|\mathbf{Y}_{t-1},\boldsymbol{\theta}) \quad (14b)$$

$$p(r_t|\mathbf{Y}_t,\boldsymbol{\theta}) \propto p(\mathbf{y}_t|r_t,\boldsymbol{\theta})p(r_t|\mathbf{Y}_{t-1},\boldsymbol{\theta}) \quad (14c)$$

In (14a) the term $p(r_t|r_{t-1},\boldsymbol{\theta})$ is simply one of the transition probabilities in (6). In (14b) the term $p(\mathbf{y}_t|r_t,\boldsymbol{\theta})$, similar to (12) is given by:

$$p(\mathbf{y}_t|r_t,\boldsymbol{\theta}) = p(\mathbf{z}_t|r_t,\boldsymbol{\theta})\prod_{i=1}^{d} g(y_{it})$$

where $p(\mathbf{z}_t|r_t,\boldsymbol{\theta})$ is the likelihood of observing $\mathbf{z}_t$ given the state, $r_t$, which is evaluated given the assumption that the state distributions are multivariate mixed truncated Normal. This is evaluated identically to (13) with the difference that $\bar{\mathbf{z}}_t = \boldsymbol{\mu}_{r_t}$, given there is no further a linear relationship between with $\bar{\mathbf{z}}_t$ and $\bar{\mathbf{z}}_{t-1}$.

## Specification of priors

The prior distribution $p(\boldsymbol{\theta})$ represents the modeller's subjective belief of the parameter values prior to model calibration. An experienced modeller may have a prior belief about the range in which the true parameters lie before fitting a model. The parameter prior allows formal incorporation of this belief. On the other hand, a modeller may have little knowledge of where the true parameters lie and therefore a prior with an equal density over all parameter values could represent the state of prior belief. Given the subjective nature of prior formulation, the introduction/specification of priors has been the subject of much controversy (Berger, 2000). Of course the formulation of the model itself is quite subjective. As Wikle (2003, p. 8) states 'One must simply recognize that a strength of the hierarchical (Bayesian) approach is quantification of such subjective judgment'. As, the posterior distribution can be sensitive to parameter prior specification (and hence can affect estimation) much literature regarding the choice of a prior distribution has been directed towards the formulation of non-informative priors, priors which it can be argued that there is 'no information' about the parameter vector (Carlin and Louis, 2000, p. 28–32), implying the resulting analysis is completely objective rather than subjective. Jeffreys (1961, p. 181) suggests a prior that is invariant under transformation (a desirable property – as the particular parameterization transformation which is chosen by the modeller is subjective). Subsequent work has moved towards the calculation of reference priors (Bernardo, 1979) (for single parameter models) and later modified for multi-parameter problems (Berger and Bernardo, 1992). However, it is recognized by Bernardo and Smith (2000, p. 298), that 'every prior has some informative posterior and predictive implications' and 'there is no ''objective'' prior that represents ignorance'.

Reference priors for the multivariate transformed AR(1)/HMM models used in this study have not been calculated (to the author's knowledge). It is also noted that some recommended non-informative priors are not necessarily proper. As in the study of Stephens (1997, p.12) we have used proper priors which attempt to be only ''weakly informative'' – representative of our subjective prior belief. We likewise complete the analysis with the warning that we feel further work is required on the appropriate specification of priors.

General notation is introduced for the complete set of parameters possible for all models used in this study is $\boldsymbol{\theta} \in (m_{ir}, s_{ir}, \lambda_i, \phi_i, p_{rj}, \omega : i = 1, \ldots, d \quad r, j = 1, \ldots, K \quad r \neq j)$ where $K$ denotes the number of HMM states. In some cases not all parameters are used (e.g. $\phi_i = 0$ for the HMM), while in others such general notation is not required (i.e. for one of the HMM variants it is assumed that the same standard deviation applies over all states for each site). Using such notation allows a wide range of model specifications, some of which will be used in the ensuing case studies.

Table 1 presents a summary of the priors used here, listing firstly the parameter of interest, then the model(s) that the parameter applies to. The prior distributions used for the parameters along with parameter bounds are given. As some parameters are used in the multiple models within this study, the same priors were placed on these shared parameters so as to not favour one model over another a priori.

*Mean $m_{ir}$ and standard deviation $s_{ir}$*: In the HMM study of Thyer and Kuczera (2003a), an Inverse-Wishart distribution was used for the prior on the covariance matrix. An Inverse-Wishart distribution is the multivariate generalization of the Inverse-$\chi^2$ distribution. In this study, a scaled inverse-$\chi^2$ distribution was used for the diagonals of the covariance matrices, and hence has the same prior on the variance

**Table 1** Parameter prior distributions

| Parameter | Prior distribution* | Lower bound | Upper bound | Prior parameters |
|---|---|---|---|---|
| $m_{ir}$ | $N(\mu_0, s_{ir}^2)$ | 0 | 10,000 | $\mu_0 = \bar{y}_i$ |
| $s_{ir}$ | Inverse $-\chi^2(\nu_0, \sigma_0^2)$ | 0 | $\infty$ | $\sigma_0^2 = \bar{s}_i^2, \nu_0 = 2$ |
| $\lambda_i$ | Uniform | −2 | 2 | Not applicable |
| $\phi_{ir}$ | $N(\mu_\phi, \sigma_\phi^2)$ | −1 | 1 | $\mu_\phi = 0.0, \sigma_\phi^2 = 0.5^2$ |
| $\omega$ | $\gamma(\alpha, \beta)$ | 0 | $\infty$ | $\alpha = 1, \beta = 500$ |
| $p_{WD}$, $p_{DW}$ | Uniform | 0 | 1 | Not applicable |

* Note: See Gelman et al. (1995, Table A.1) for functional forms.

parameter $s_{ir}^2$ as Thyer and Kuczera (2003a). An independent univariate normal distribution was used for each mean $m_{ir}$ component. Following Thyer and Kuczera (2003a) the site empirical means $\bar{y}_i$ and variance $\bar{s}_i^2$ were used to define hyper-parameters for each associated prior distribution. The prior degrees of freedom ($\kappa$ and $v_0$) were kept to a minimum to ensure that the priors remained diffuse. Identical priors were used for all states $r$ of the HMM.

*Box–Cox transformation parameter* $\lambda_i$: given no transformation and log transformations are modelled at parameter values of $\lambda_i = 1$ and $\lambda_i = 0$, respectively (transformations which have typically applied for annual streamflow/rainfall in the past) – a uniform prior with range $(-2,2)$ was considered adequate for this modelling exercise.

*Lag-one autocorrelation coefficient* $\phi_{ir}$: An independent Gaussian distribution with mean 0.0 and standard deviation 0.5 was used for this parameter. Based on the analysis of 44 rainfall stations across Australia, Srikanthan et al. (2001) report that the empirical lag-1 autocorrelation coefficient had an average value of 0.08 and standard deviation of 0.1. In comparison, the empirical 'worldwide' lag-1 autocorrelation coefficient for streamflow presented in McMahon and Mein (1986) as reported in Salas (1993) was 0.23 with a range of $-0.2$ to 0.8. In comparison the 95% probability limits for the prior used here was $(-0.98, 0.98)$. The variability is inflated here to be considerably greater than the average value from the previous studies to ensure the prior is diffuse. Thus the chosen prior includes the range of values empirically estimated from past studies. It is noted that the inclusion of the BC transformation results in the theoretical lag-1 autocorrelation in non-transformed space not being equivalent to that in transformed space, though Thyer et al. (2006) found the difference between the lag-1 autocorrelation in transformed and non-transformed space were practically negligible for synthetic case studies typical of rainfall data from the Sydney region.

*Exponential correlation decay parameter*: A Gamma distribution was used for the prior on $\omega$, following the approach of Sanso and Guenni (2000). Parameter values of $\alpha_\omega$ and $\beta_\omega$ were chosen using a visual technique based on plotting the empirical correlations versus distance and the 90% confidence limits of the prior distribution for the spatial correlation function. The hyperparameter values for the prior for $\alpha_\omega = 1$ and $\beta_\omega = 500$ were chosen such that the 90% confidence limits were very wide around the empirical correlations, thus ensuring a vague prior. It is noted that the inclusion of the BC transformation results in the theoretical spatial correlation in non-transformed space not being equivalent to that in transformed space. However, given the typical degree of transformation $\lambda$ used in such studies lies in the range [0,1], it is very unlikely that some combination of $\lambda$ values at each site could alter the observed values encountered here to lie outside the 90% prior confidence limits post-transformation.

## Hydrological data

Rainfall data from five sites located in and around the Warragamba catchment, the major water supply catchment for Sydney (Australia), was used in the case study presented here. These five rainfall site locations (Mt Victoria/Blackheath, Sydney, Moss Vale, Taralga and Yarra) are shown in Fig. 1. The monthly data span a period of 111 years from June 1883 to May 1994. Details of the data and composite record construction can be found within Thyer (2001). To aggregate to annual data a starting month is required. Thyer (2001) notes that for multiple site analysis the starting month did not have a significant effect on the persistence structure. In the single site analysis of over 40 sites spread across Australia, Srikanthan et al. (2001) could not identify any 'noticeable pattern in the starting months'. With little constraint on the water year starting month, the April-March water year was selected on the grounds that ENSO events tend to break by the end of the Austral Autumn. The annual rainfall statistics for each site are summarized in Table 2. A time series plot of the cumulative departures from the mean of the annual rainfall data are presented in Fig. 2a. These deviations were standardized by dividing these departures by the empirically estimated standard deviation. Such a plot is typically used to empirically diagnose persistence within time series (Pittock, 1975). Consistent periods of negative/positive gradient denote periods where the observed rainfall is consistently below/above the mean. There appears to be evidence of distinctly wet and dry periods, with some similarity in timing when changes in slope occur (e.g. during 1946–1948 all series slope changes from negative to positive). This provides some evidence for the hypothesis of a regional controlling climate, with marked switches between relatively wet and dry regimes. However, not all sites follow the overall trend (e.g. Moss Vale from 1910 to 1930). The degree to which the HMM and the AR(1) model can reproduce these characteristics will determine their applicability for use for simulation.

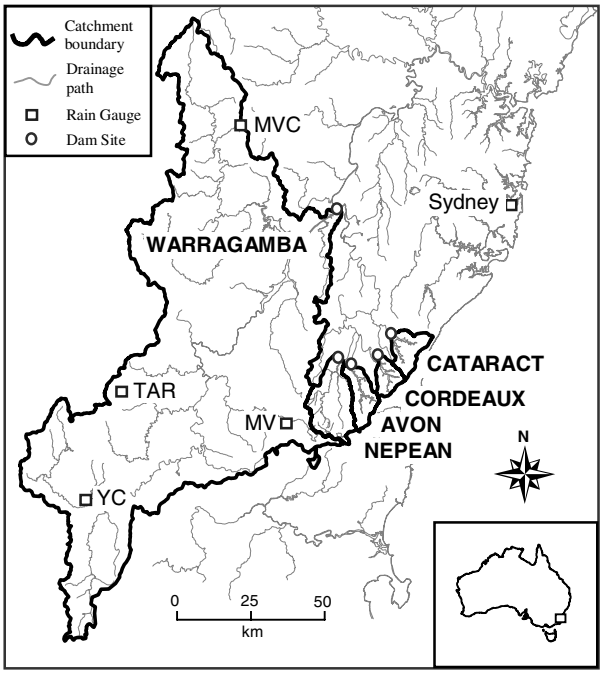


**Figure 1** Rainfall data site locations.

**Table 2** Rainfall data annual statistics

| Site | Mean (mm) | Standard deviation (mm) | Skew (−) | Lag-one autocorrelation (−) |
|---|---|---|---|---|
| Mt. Victoria/Blackheath (MVC) | 1070.0 | 318.8 | 0.61 | 0.18 |
| Sydney | 1210.2 | 334.1 | 0.68 | 0.06 |
| Moss Vale (MV) | 992.6 | 275.3 | 0.67 | 0.22 |
| Taralga (TAR) | 815.9 | 217.9 | 0.29 | 0.14 |
| Yarra (YC) | 670.6 | 186.9 | 0.50 | 0.31 |

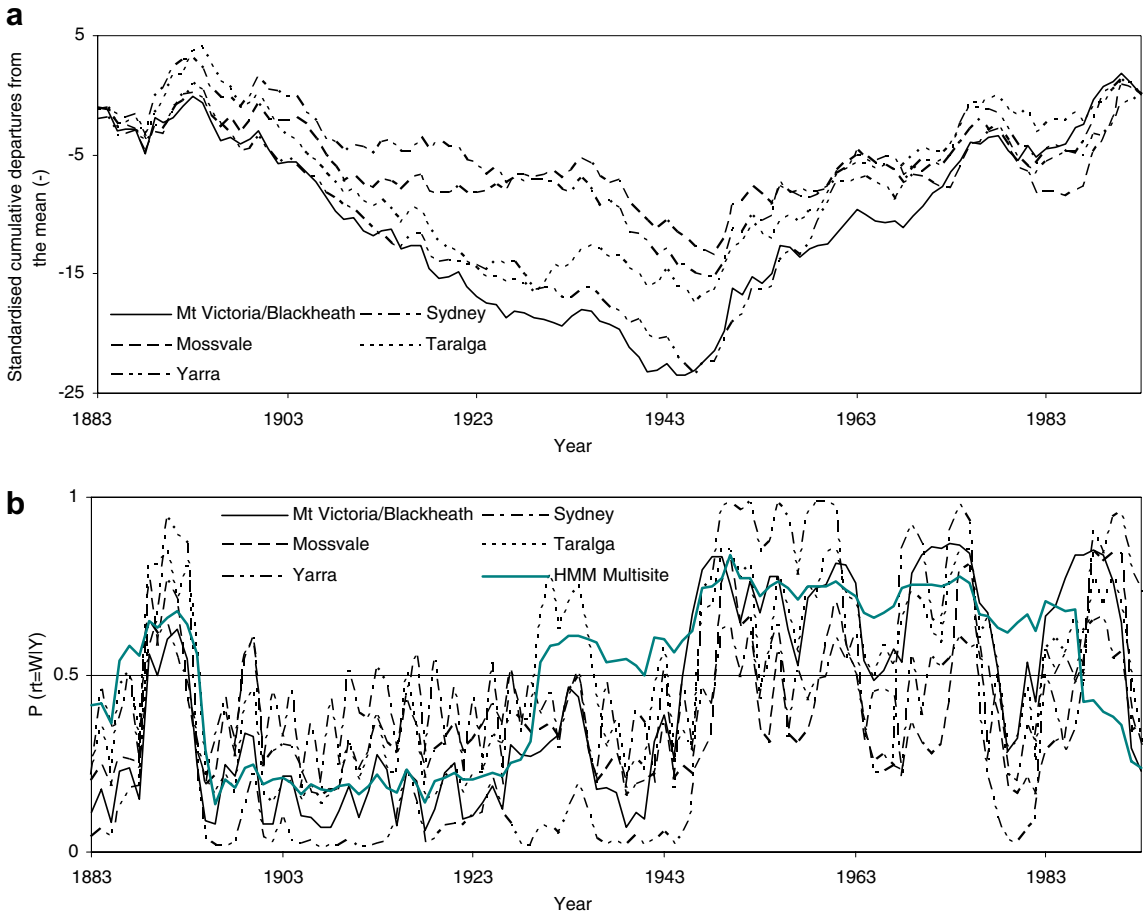


**Figure 2** Comparison of (a) standardized cumulative departures from the mean of annual rainfall and (b) the HMM posterior state probability $P(r_t = W|Y)$ series for the multi-site and individually estimated sites.

## Evaluation of model performance

### Observed rainfall data statistics

A crucial feature of any stochastic model is its ability to reproduce the observed statistics of the process it is endeavouring to replicate. The models will be checked to ensure they reproduce the observed mean, standard deviation and skew, using annual rainfall distribution plots for presentation. In addition, the long-term variability/persistence will be tested by examining the autocorrelation function and the multi-year accumulated rainfall distribution. Site-to-site spatial correlation was also validated.

### Bayesian model selection

Choosing the 'right' model, given a set of data, is not usually a trivial task. Of the many methods available, Bayesian model selection is attractive in that it directly provides the posterior probability of model $i$, $M_i$, being the correct model given the set, $\mathbf{M}$, of mutually exclusive models being compared:

$$p(M_i|\mathbf{Y}) = p(\mathbf{Y}|M_i)p(M_i)/\sum_{\mathbf{M}} p(\mathbf{Y}|M_i)p(M_i) \quad (15)$$

Given samples from the posterior for a particular model it is possible to calculate the normalizing constant from (10):

$$p(\mathbf{Y}|M_i) = \int_{\theta} f(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\,d\boldsymbol{\theta} \quad (16)$$

using the method suggested in Gelfand and Dey (1994) Eq. (27), with a multivariate Gaussian distribution with associated mean and covariance estimated from the parameter samples of the posterior used as for the importance sampling density. This procedure was used in Frost (2004, pp. 53–54). Kass and Raftery (1995) observed that this Gelfand and Dey (1994) estimator is stable provided the tails of the posterior density are sufficiently thin. It is recognized that there are other methods of estimation of the normalizing constant (e.g. Chib and Jeliazkov (2001)), which may, under model and data dependent circumstances provide differing estimates to those found here. However, testing the sensitivity and accuracy of this estimator is beyond the scope of this paper, and implementation of these methods are left as future work.

The estimate of $p(\mathbf{Y}|M_i)$ is then substituted back into (15) to calculate the probability of each model given the data, $p(M_i|\mathbf{Y})$. For this calculation, all models applied in this paper were given equal prior probability $p(M_i)$. That is, a uniform model prior was applied. This method takes into account both the model complexity and parameter uncertainty as it involves integration over all the parameters to calculate $p(\mathbf{Y}|M_i)$. It is noted that Bayesian model selection can be affected by a phenomenon known as Lindley's paradox (Lindley, 1957), where the use of noninformative unbounded priors on parameters in the case of nested model comparisons can bias model selection always in favour of the simpler nested model. As bounded priors over reasonable ranges have been specified here, this problem will not arise. Gelman et al. (1995, p.176) have questioned the use of Bayesian model selection in some applications, arguing that in cases where competing models are distinctly and legitimately different, Bayes factors coupled with proper noninformative model priors provide a worthwhile means for model selection. However, in cases where both competing models are special cases of a more general parametric model, they argue that it may be a hindrance because the true model is likely not to be one of the special cases. They recommend that the posterior distribution of the parameters characterizing the different models be studied. Other methods which address the issues of prior selection (both parameter and model) in hierarchical/nested model comparisons are discussed within Ntzoufras (1999) and Chipman et al. (2001). For a thorough discussion of the methods used here along with further detail on the parameter uncertainty methods see Frost (2004, Chapter 3).

## Simulated drought risk and yield estimates

To evaluate the practical implications a simple reservoir simulation was used to evaluate the impact on reservoir yield estimates. To provide the inputs for the reservoir simulation the multi-site simulations of rainfall from four sites in the Warragamba catchment (Mt Victoria/Blackheath, Taralga, Moss Vale, and Yarra) were first transformed into the catchment average rainfall using Thiessen polygons. The simulated catchment average rainfall was then transformed into streamflow based on a linear regression of catchment average rainfall and catchment runoff. This approach was used instead of directly simulating multi-site streamflow for two reasons: Firstly, the rainfall record was significantly longer than the streamflow record available. Secondly, preliminary analysis of streamflow records for the Hawkesbury-Nepean catchment revealed there was significant infilling. Such infilling has the potential to bias the stochastic modelling which treats every data point as independent information. In comparison the infilling of the rainfall records is known to be very minor (Thyer, 2001). Given the linear regression model of annual rainfall-runoff is imperfect ($r^2 = 0.72$) the drought probabilities should only be considered indicative of the possible effect of not including parameter uncertainty as input to such a process.

The simulated catchment inflow was used as an input to a simplified reservoir simulation. For each year in this reservoir simulation the inflow was added to the storage volume, and a constant average annual demand (AAD) was subtracted. The initial storage volume was set to the storage capacity, though given the long length of simulations used in this study (10,000 years) this is unlikely to have a major effect on drought probabilities. The storage capacity was set the same as the Warragamba dam total volume. Sydney catchment authority's (SCA) current restriction policy was implemented here, with a percentage reduction in AAD based on the storage levels. The drought probability was calculated as the percentage of years below a given storage level (% of capacity). The reservoir yield is the maximum AAD that can be supplied ensuring the reservoir does not drop below a given storage level with a given probability. SCA uses three system criteria to determine their system yield as part of their operating license (SCA, 2000):

(a) Reliability: Not less than 97% of months are to have no restrictions. (restrictions are enforced when storage <55%, hence this means not more than 3% of months are to have storage <55%)
(b) Robustness: Not less than 90% of years are to have no restrictions. (Not more than 10% of years are to have storage <55%)
(c) Security: Not less than 0.001% of months is the storage to fall below 5% of capacity (Not more than 0.001% of months are to have storage <5%)

Given that the reservoir simulation was undertaken using an annual time step the monthly criteria could not be directly applied. Thus for the purposes of this study the yield was determined based on criterion (b) and a modified criterion (a). Criterion (b) is that not more than 10% of years are to have storage <55% and will hereafter be referred to as the robustness criteria. Criterion (c) will be modified to the approximate annual equivalent of 0.01% (Not more than 0.01% of years to have storage <5%) and will hereafter be referred to as the security criteria. For a given AAD the drought probability will be calculated as the expected probability from 1000 replicates of length 10,000 years. This length of data are required to negate the effects of sampling variability and ensure accurate estimates of the expected drought probability. The yield will be calculated based on the lowest yield of either the robustness or security criteria outlined above. The results for different stochastic modelling parameterisations with and without parameter uncertainty will be compared to evaluate their impact on yield estimates.

**Table 3** Models tested and overall model probability

| Model | Model parameter[a] | | | | Model probability |
|---|---|---|---|---|---|
| | $\lambda$ | $\mu$ | $\sigma$ | $\phi$ | |
| AR(1)-noBC | 0 | 1 | 1 | 1 | 1.64E−09 |
| HMM-noBC | 0 | 2 | 2 | 0 | 7.55E−07 |
| HMM | 1 | 2 | 2 | 0 | 5.87E−06 |
| AR(1) | 1 | 1 | 1 | 1 | 7.45E−06 |
| HMM-mean | 1 | 2 | 1 | 0 | 1.50E−04 |
| AR(1)-onephi | 1 | 1 | 1 | 1/*D* | 6.47E−04 |
| HMM-oneBC | 1/*D* | 2 | 2 | 0 | 1.86E−03 |
| AR(1)-oneBC | 1/*D* | 1 | 1 | 1 | 0.011 |
| HMM-meanoneBC | 1/*D* | 2 | 1 | 0 | 0.076 |
| AR(1)-oneBCphi | 1/*D* | 1 | 1 | 1/*D* | 0.910 |

[a] Number indicates the number of model parameters per site. *D* indicates the total number of sites (Here *D* = 5).

## Model specification

The development of the generalised multi-site formulation outlined in Section ''Stochastic models for annual multi-site hydrologic data'' and a generalised likelihood function (Section ''Likelihood function'') provides the opportunity to evaluate various model parameterisations, including pooling parameters across regions. The model combinations chosen to be tested are aimed to achieve the study objectives. Table 3 lists the various models and their parameterisations — the model nomenclature provides guidance on the model parameterisation. The first four models will evaluate if the BC transformation is justified for either the AR(1) or HMM models, by comparing AR(1)-noBC and AR(1) (Objective A). Comparison of the AR(1) and HMM models will determine objective B. The remaining models will evaluate objective C — to determine if pooled parameters can improve model identifiability. The HMM-mean models have a common standard deviation in each state. The AR(1)-onephi models have a common $\phi$ for all sites, while the AR(1)/HMM-oneBC models have a common $\lambda$ for all sites.

## Results

The sampling distributions of selected statistics are compared with observed statistics. For all plotted statistics, the observed values are plotted as a point value, while simulated median is plotted as a thick line, and 90% confidence limits as dashed lines. Observed points lying outside of these bounds are attributable to either sampling variability or inadequacy of the model. Model probabilities are also presented for the tested models. Using these various measures of model performance several comparisons will be undertaken.

### No transformation versus BC transformation

The annual rainfall distributions are shown for the HMM and AR(1) models with and without the Box—Cox transformations in Fig. 3. For clarity only the results for two sites, Mt Victoria/Blackheath and Yarra, are shown. Similar trends are found for the other three sites. Fig. 3c and d demonstrate that both models incorporating the BC transformation perform better at capturing the skew (related to curvature when plotted on Normal probability axes — i.e. no curvature equates to zero skewness) in the data than without the BC transformation (Fig. 3a and b). The mean and standard deviation are reproduced well by all models, yet the skew is underestimated for the models not including the BC transformation, significantly for the AR(1) model and to a lesser extent for the HMM. The skew is reproduced well by the models incorporating the BC transformation. The minimum observed values are consistently above the simulated median for the no transformation results. The corresponding plots for the models with the BC transformations show no overall bias in terms of minimum, a result of reproducing the observed curvature of the observations.

The posterior model probabilities shown in Table 3 confirm these results. The models with BC transformation have a considerably higher probability than those without the BC transformation. All further models tested will include the BC transformation.

### AR(1) versus HMM

#### Temporal dependence

The key difference between the AR(1) and the HMM is the parameterization of the temporal dependence. This will be compared by considering the observed autocorrelation coefficient for lags 1—30 and the 30 year accumulated rainfall distributions. The 30 year accumulated rainfall distribution represents the distribution of the 30 year moving averages. The 30 year time step was chosen to examine whether these models capture the multi-decadal variability, which has a significant influence on Australia's climate (Kiem and Franks, 2004; Verdon et al., 2004).

The temporal correlation plot (Fig. 4 — shown for Mt Victoria/Blackheath and Yarra sites) illustrates the difference in correlation spectrum for the AR(1) and HMM models. The AR(1) model is quite strong at lag-1 but dampens quickly. For Mt Victoria/Blackheath (Fig. 4a), the observed values are within the 90% probability limits for all lags excluding lag-2, -3, -13, -14, -17 and -21. For Yarra (Fig. 4c), the observed temporal correlation is reproduced to a greater

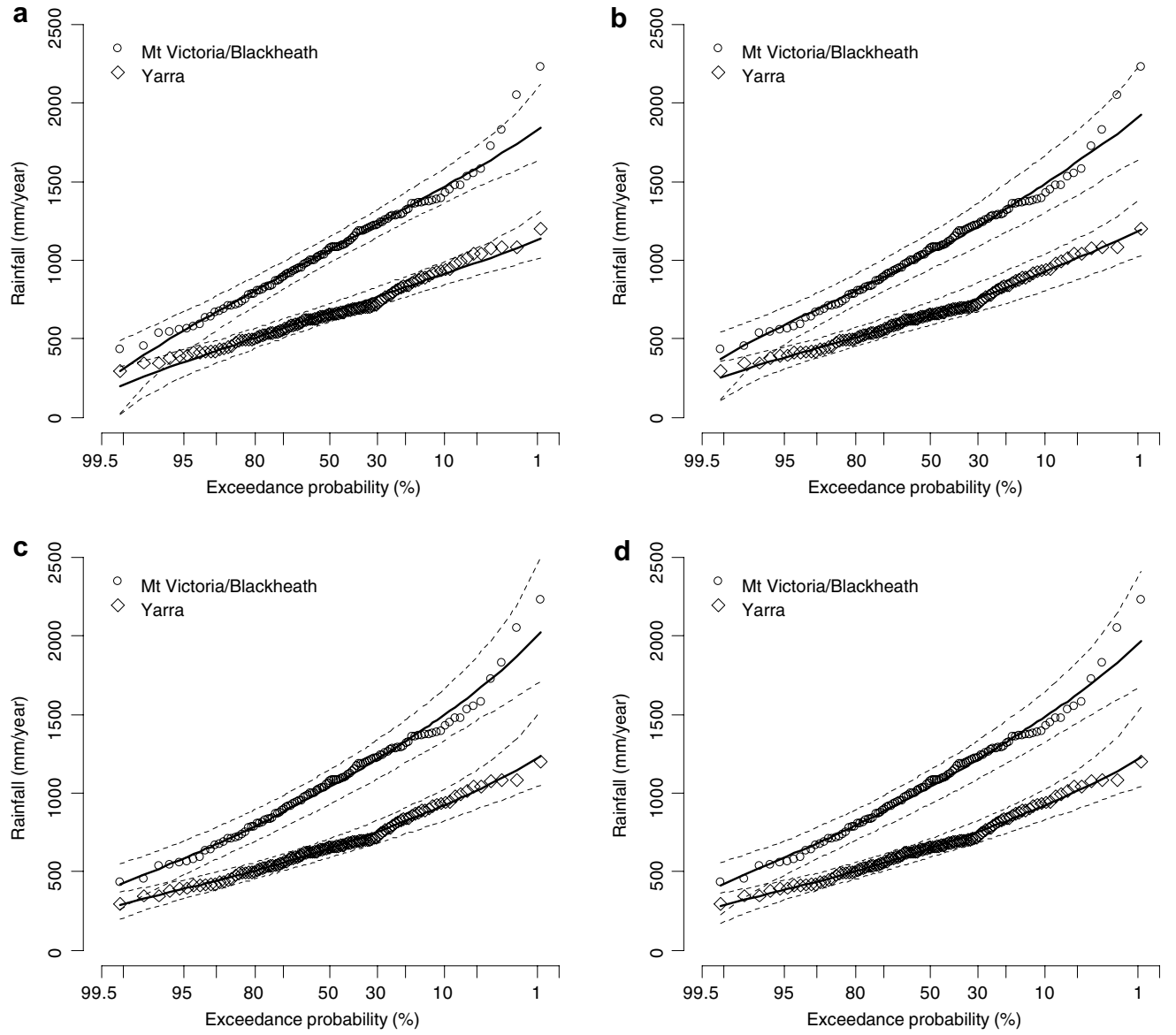


**Figure 3** Observed versus simulated annual distribution for the (a) AR(1)-noBC, (b) HMM-noBC, (c) AR(1) and (d) HMM. Solid represents the median (50% probability) and dashed line represents the 90% probability limits.

degree of satisfaction, with lag-2, -3, -13 and -14 observed autocorrelation lying on or just outside the simulated upper 90% confidence limits. The Taralga and Moss Vale autocorrelation pattern for observed and simulated values is similar to that of Yarra. Sydney (not shown) differs in that the only observed value lying outside the 90% confidence interval is at lag-2. Overall, the AR(1) model tends to be underestimating lag-2 and -3 autocorrelation. The HMM correlation spectrum (Fig. 4b and d) shows a slower decay in autocorrelation than the AR(1), though with a generally lower level at lag-1. In contrast to the AR(1), the Mt. Victoria/Blackheath lag-2 autocorrelation is reproduced by the HMM, although the lag-3 autocorrelation remains underestimated. For Yarra, both lag-1 and -2 autocorrelations are underestimated.

For the 30 year rainfall accumulations (Fig. 5), the majority of observed values are bounded by the 90% probability limits for the AR(1) and HMM. For the AR(1) the observed values for Mt. Victoria/Blackheath for the lower 30 year accumulation periods lie on, or just outside the 90% probability limits. The HMM has wider bounds for Mt. Victoria/Blackheath than the AR(1), with no observations outside the confidence limits. In contrast, the maximum values for Yarra are just outside the confidence bands for the HMM, whilst being within the bounds for the AR(1). These differences follow from the reproduction of autocorrelation presented in Fig. 4. For Mt. Victoria/Blackheath the HMM performs better than AR(1) in terms of reproduction of observed autocorrelation (especially at lag-2), whilst for Yarra the reverse is true in that the HMM underestimates the lag-1 and -2 autocorrelation, whilst the AR(1) performs satisfactorily.

The model probabilities comparing the HMM and AR(1) in Table 3 confirm the visual assessment of the various statistics that neither model is strongly favoured, yet the AR(1) model is preferred, apparently due to better reproduction of lag-1 autocorrelation. The AR(1) model cannot reproduce the temporal correlation structure such as that presented

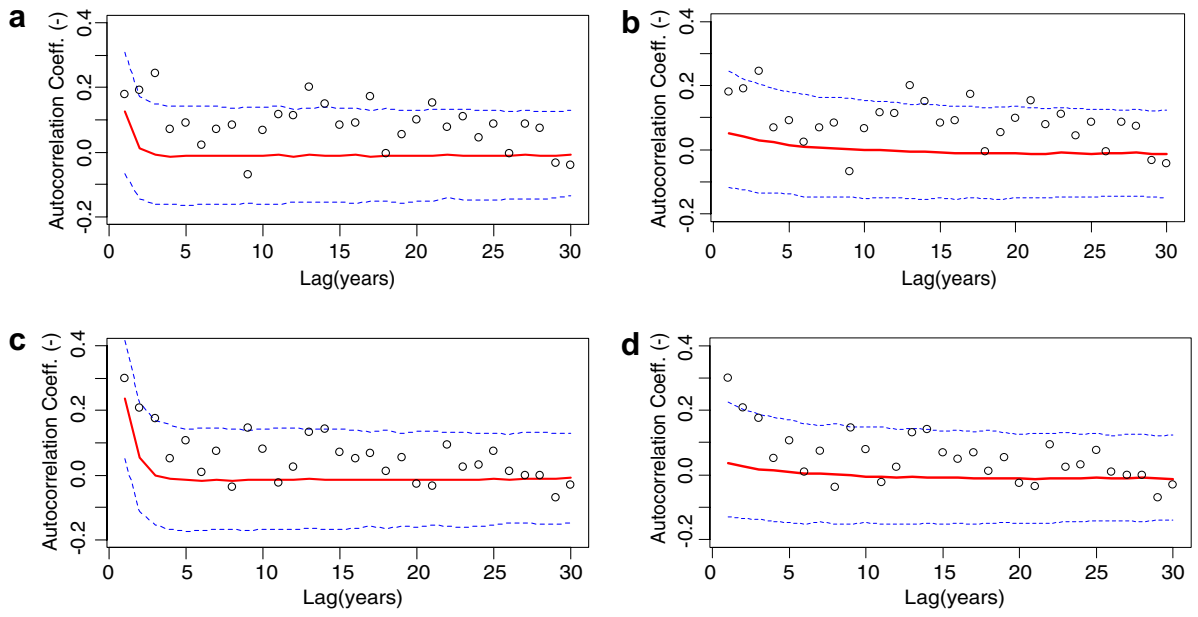


**Figure 4** Observed versus simulated autocorrelation function for AR(1) ((a) Mt. Victoria and (c) Yarra) and the HMM ((b) Mt. Victoria and (d) Yarra). Observed values are open circles. Solid represents the simulated median (50% probability) and dashed line represents the simulated 90% probability limits.

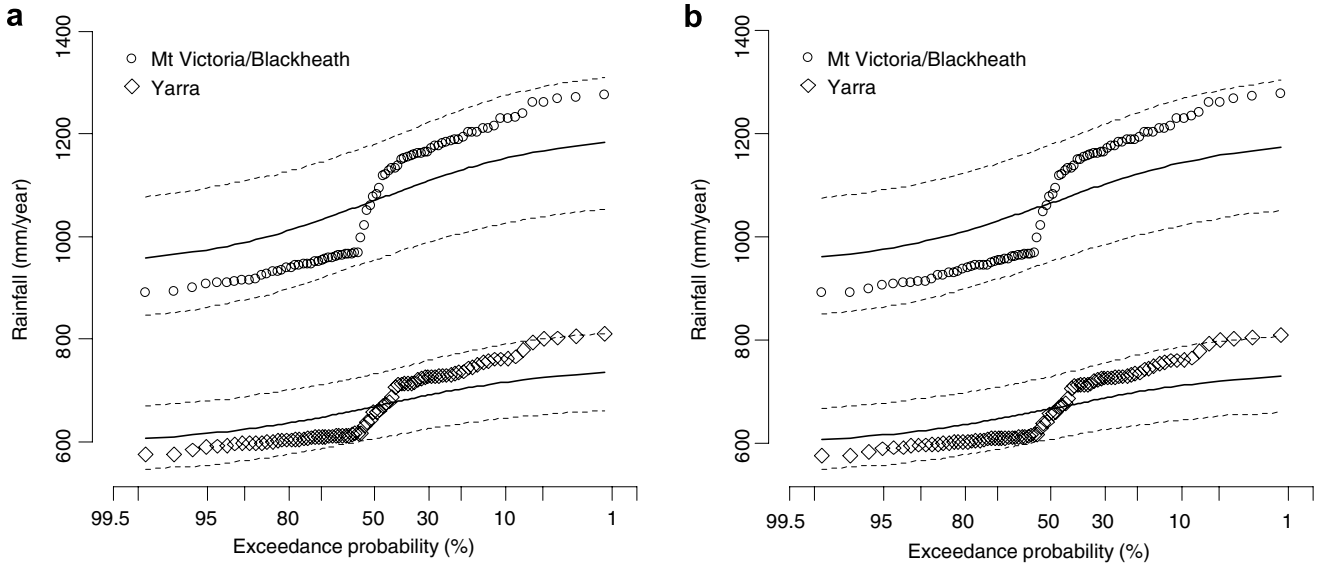


**Figure 5** Observed versus simulated 30 year accumulated rainfall distributions for the (a) AR(1) and (b) HMM.

for Mt. Victoria/Blackheath because the correlogram decays exponentially. As for the question of why the HMM was unable to reproduce the temporal correlation characteristics of the rainfall data at some sites (e.g. Yarra) — this is further investigated in Section ''Impact of the spatial correlation parameterisation on HMM identification''.

### Spatial correlation parameterization

Thyer and Kuczera (2003a) attempted fitting individual correlation coefficients between each pair of sites but were unable to achieve convergence of the MCMC algorithm. Here an alternate parameterization of the spatial correlation is trialled, the exponential decay correlation, with a single parameter describing the Gaussian correlation between all sites. When the exponential decay function was used, satisfactory convergence was achieved, illustrating the efficacy of this approach. Fig. 6a and b show the correlation plots versus distance for the AR(1) and HMM models, respectively. For the AR(1), (Fig. 6a), the observed spatial correlations are just outside the 90% probability limits for two of the 10 inter-site spatial correlations (Moss Vale-Taralga and Moss Vale-Sydney). For the HMM (Fig. 6b) all observed spatial correlations are within the 90% confidence limits. These differences are relatively minor and overall, the exponen-

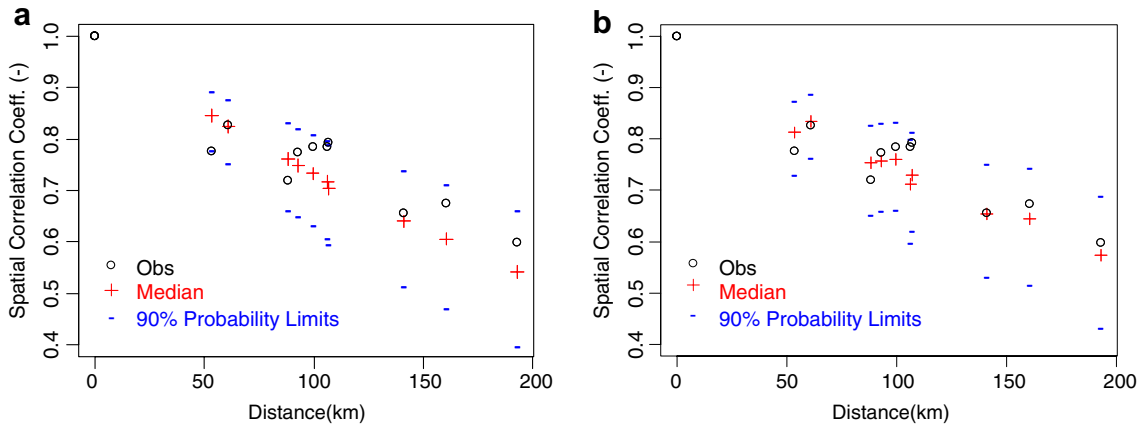


**Figure 6** Observed and simulated spatial correlation coefficient versus distance for the (a) AR(1) and (b) HMM.

tial decay correlation function is able to reproduce the observed spatial correlation for both models for the majority of spatial scales tested. The use of exponential decay correlation function does however influence the calibration of the HMM as discussed in Section ''Impact of the spatial correlation parameterisation on HMM identification''.

## AR(1) and HMM pooled parameter variants

The modelling framework used here enabled the calibration of various AR(1) and HMM pooled parameter variants. Of these variants there may be a more parsimonious representation of the models presented so far in that regional pooling of model parameters might be beneficial (e.g. using a single BC parameter across all sites). A range of models were compared, and their posterior model probabilities are shown within Table 3.

The AR(1) modelling regional pooling models tested (single and/or single across all sites) have a greater model probability than the AR(1) model, with the AR(1)-oneBCphi with single $\lambda$ and $\phi$ being strongly favoured over all models tested. For the HMM pooling, a single $\lambda$ across all sites and/or a single $\sigma$ for each site (as opposed to each state and site) was trialled. All trialled models had a greater posterior model probability than the original HMM. The most probable model (HMM-meanoneBC) pooled the BC parameter across all sites, and significantly only used a single $\sigma$ for each state.

## Impact on drought probabilities and yield estimates

Previously multi-site models only estimated single values for their parameters. The primary contribution of this paper is to develop a multi-site modelling framework for annual hydrological data that evaluates parameter uncertainty and model uncertainty. The practical impact of this contribution on the estimated drought probabilities and hence yield estimates will be evaluated by comparing:

(a) the drought probabilities estimated based on single parameter values;
(b) drought probabilities based on full evaluation of parameter uncertainty (for an individual model); and,
(c) drought probabilities based on full evaluation of parameter and model uncertainty.

For the single parameter value runs, the posterior modal estimate (most probable value) of the parameters will be used. The stochastic model parameterisations chosen are the AR(1) model, which, as the most commonly used model, provides a basis for comparison, the HMM to investigate the impact of differences in modelling temporal dependence, the pooled parameter AR(1)-oneBCphi model which had the highest model probability (Table 3), and finally a multi-model combination of the AR(1) and HMM models to illustrate the impact of model uncertainty. When comparing only the AR(1) and HMM models, their model probabilities (Table 3) are quite similar (0.559 for AR(1) c.f. 0.441 for HMM); thus one is uncertain which model to use for simulation. This model uncertainty can be taken into account when estimating drought probabilities using the following method: During the reservoir simulation each of the replicates was chosen from either the AR(1) or HMM models based on their model probabilities. Thus the estimated drought probabilities account for model uncertainty because the simulations from both the AR(1) and HMM models are weighted by their respective model probabilities. This demonstrates the power of evaluating the model probabilities: not only does it provide a statistical measure of a particular models goodness of fit, it can also be used to weight models during simulations when using an ensemble of models to evaluate key performance variables (e.g. reservoir yield).

Fig. 7 shows the impact of evaluating parameter uncertainty for the AR(1) and the HMM. For both models and both criteria the expected drought probability for a given annual supply is higher when parameter uncertainty is accounted for (posterior) compared to when single parameter estimates are used (mode). The difference is greater for the security criterion, which is a rare event, compared to the robustness criterion, which is a more frequent event. The security criterion produces lower yield estimates than the robustness criterion. For the AR(1) model without evaluating parameter uncertainty the yield is approximately 440 GL, while when parameter uncertainty is evaluated it decreases to approximately 410 GL, a decrease of 7.3%.

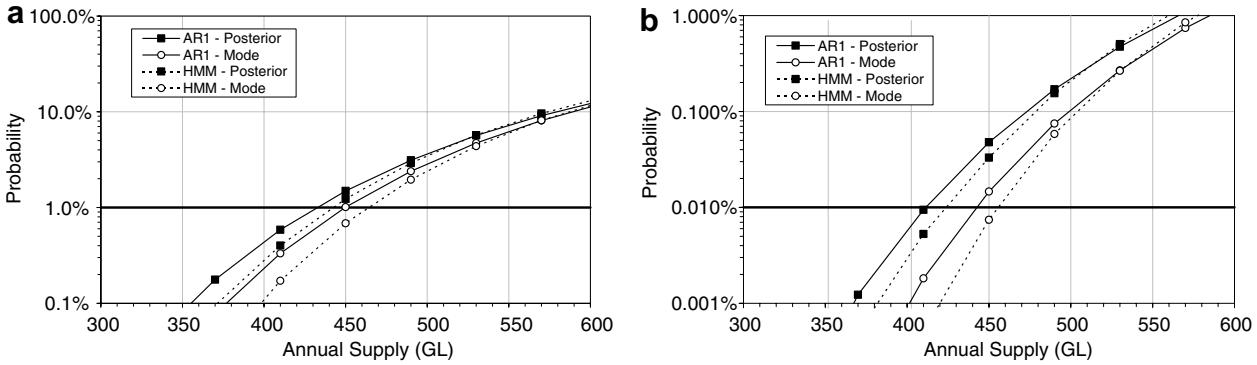


**Figure 7** Impact of parameter uncertainty on simulated drought probabilities for the AR(1) and HMM on (a) the robustness criterion (years <55% Storage) and (b) the security criterion (years <5% Storage). The SCA minimum threshold (robustness = 1.0% and security = 0.01%) is plotted as a bold horizontal line.

Importantly, also for the annual supply of 440 GL the drought probability estimate increases by half an order of magnitude (0.01% with no parameter uncertainty to 0.04% with parameter uncertainty). A similar magnitude decrease in the yield was found for the HMM; without parameter uncertainty the yield was approximately 455 GL whereas with parameter uncertainty it was 420 GL, a decrease of 7.6%. Thus it can be concluded evaluating parameter uncertainty produces higher drought probabilities and hence lower yield estimates. This has the practical implication that if only single value parameter estimates are used, which is the current practice, reservoir yield will be overestimated.

Fig. 8 shows the impact of using different models to estimate the drought probabilities for the security criterion. Of the four simulations shown the AR(1) model produced the lowest yield estimate (410 GL) while the HMM produced the highest yield (425 GL). When model uncertainty was taken into account, (denoted as AR and HMM in Fig. 8), the yield estimate was 415 GL. This is the same as the yield estimate for the AR(1)-oneBCphi model which had the highest model probability. The standard AR(1) model with no parameter uncertainty produced a yield estimate of 440 GL, 25 GL higher than the most realistic estimate. Although this is only a reduction of 6%, it is practically significant given that it is equivalent to 70% of the volume saved due to demand management measures implemented by Sydney Water Corporation (SWC) during 2004—2005 (SWC, 2005). If model uncertainty is ignored by using only the AR(1) or HMM (including parameter uncertainty), the yield is estimated with errors of −1.2% and 2.4%, respectively, compared to the best estimate.

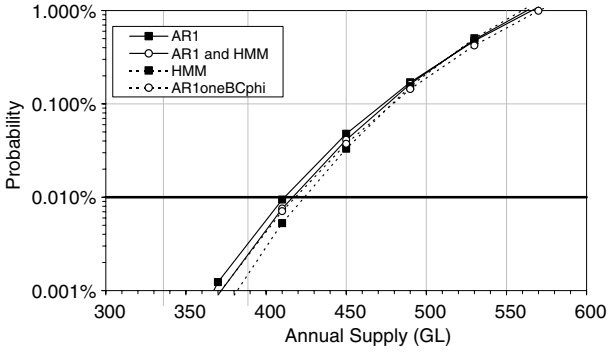


**Figure 8** Impact of model uncertainty on simulated drought probabilities for the security criterion (years < 5% storage). The SCA minimum threshold (security = 0.01%) is plotted as a bold horizontal line.

## Discussion

### Box—Cox transformation

Of the stochastic models tested here, models incorporating BC transformation are favoured using Bayes posterior model probability over those that do not. This result is supported by the distributional plots shown in Fig. 3 — the models without the BC transform less adequately reproduce the skew observed in the annual rainfall data. In terms of practical implications, the annual minima are consistently underestimated (for both models and all sites) by such a model, with consequent effects on estimation of drought risk. It is noted that the data are not either normally or log-normally distributed, rendering models that use such an assumption inadequate in this case.

### AR(1) versus HMM

Overall, the multi-site AR(1) model slightly outperformed the multi-site HMM in terms of model probability, though the difference between the two models was marginal. This is also shown in the reproduction of the observed statistics. The only major difference between the two models was that the AR(1) model was able to better reproduce the lag-1 temporal correlation than the HMM. This reproduction of temporal correlation by the HMM is related to the spatial parameterisation which is discussed in the following section.

### Impact of the spatial correlation parameterisation on HMM identification

The spatial correlation parameterization used in this study — the exponential correlation decay structure — was able to reduce the number of parameters and facilitate conver-

gence of the MCMC algorithm (c.f. Thyer and Kuczera, 2003b) and also able to reproduce the observed spatial statistics (Fig. 6). However, further inspection reveals this spatial correlation parameterisation does impact on the identification of the HMM.

The HMM involves estimation of latent (or hidden) state parameters $r_t \in (W,D)$ for each data timestep $t$. The posterior state probability series $p(r_t = W|\mathbf{Y},\boldsymbol{\theta})$, $t = 1, \ldots, n$ shown in Fig. 2b, given the estimated multi-site HMM parameters and entire data series $\mathbf{Y}$, provides an estimate of the probability that the rainfall is being generated from either the wet (or dry) state (see Bengio, 1999, for calculation details). For compactness, reference to the parameter set $\boldsymbol{\theta}$ is omitted in subsequent references. Although the state probability series is not strongly identified (that is the probability dithers around 0.5), clear wet and dry periods are identifiable by stratification into $p(r_t = W|\mathbf{Y}) > 0.5$ (higher probability of being in the wet HMM state) or $p(r_t = W|\mathbf{Y}) < 0.5$ (higher probability of being in the dry HMM state). A ''wet'' period is identified in 1886–1895, a ''dry'' period from 1895 to 1930, a wet period from 1930 to 1986, and a ''dry'' period from 1986 to 1992.

Fig. 2b shows the posterior state series for individually modelled sites compared to the multi-site HMM. The individual state series do show similarity to one another. However, there are some distinct differences between the single and multi-site state series. For example, in the period 1936–1946 the majority of the single site state series indicate the dry state has a higher probability, $p(r_t = W|\mathbf{Y}) < 0.5$, while the multi-site state series indicates the wet state has a higher probability, $p(r_t = W|\mathbf{Y}) > 0.5$. The multi-site result conflicts with Fig. 2a — the negative slope shows that almost all the sites have consistently below average rainfall in the period 1936–1946. It appears that identification of a temporal correlation structure consistent with the assumed HMM structure (of relatively wet and dry periods) is being confounded by the spatial parameterisation of the model.

Fig. 9 shows a plot of observed spatial correlations partitioned into the HMM wet period ($p(r_t = W|\mathbf{Y}) > 0.5$) — denoted as WO — and the period, ($p(r_t = W|\mathbf{Y}) < 0.5$) — denoted as DO. Fig. 9 also has an estimate based on the entire series ('all'). The data have been transformed according to the modal BC parameter set and the modelled exponential correlation decay function is also plotted. The partitioning of the observed data into the identified state epochs by the HMM state series provides a closer fit overall of the assumed exponential correlation decay function than was possible for the entire series. This is evident in particular for the site pairs Moss Vale–Taralga, Mt. Victoria/Blackheath–Taralga and Mt. Victoria/Blackheath–Yarra located 53 km, 100 km and 160 km apart, respectively. Although for some site pairings the DO correlation values lie further away from the exponential correlation line than using 'all' data, a greater overall fit (according to the likelihood) is achieved by identifying periods of differing correlation.

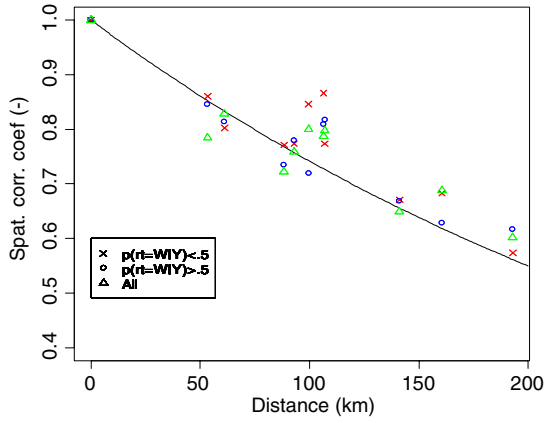


**Figure 9** Observed and simulated spatial correlation coefficient versus distance for the observed transformed data partitioned according to $P(r_t = W|Y) < 0.5$ and $P(r_t = W|Y) > 0.5$.

Fig. 10 illustrates the impact of this identification of state epochs on other parameters within the model. The WO and DO rainfall data are plotted for two site pairings corresponding to two of the previously mentioned pairings affected most by the imposition of the exponential correlation decay. The ellipses corresponding to the 95% probability limits of the HMM wet and dry state modal parameters are shown, with the axes being transformed according to the BC $\lambda$ parameter identified. Fig. 10a shows the Moss Vale–Taralga site pairing (the two most closely located sites) where a high correlation is enforced by the exponential correlation structure relative to that estimated empirically. To capture the lower correlation observed in the data, and still reproduce the observed variability, two periods differing significantly in only the Taralga mean are identified. Fig. 10b shows the Mt. Victoria/Blackheath–Taralga site pairing — the wet state has several points which are outside the 95% probability limits, while the dry state has none and the lower tail of the 95% probability ellipse is considerably lower than the lowest observed rainfall value. This is because the dry state variance parameter for Taralga is overestimated. If the correlation is underestimated by the exponential function, as it is for the DO correlation for this site pair, then the site variance must be overestimated to maintain the desired covariance.

The impact of this is that the HMM is slightly better than the AR(1) model at reproducing the observed spatial correlation (Fig. 6). This better reproduction of spatial correlation is however at the cost of reproduction of variance and temporal correlation. As the state series identified here (Fig. 2b) contains few state switches, negligible temporal correlation is induced. This explains the temporal correlation results for Moss Vale and Yarra which illustrate poor reproduction of the lag-1 autocorrelation.

This analysis has revealed that there are significant shifts in spatial correlation characteristics identified within the data. The AR(1) model, as specified here, is unable to reproduce such shifts in correlation. The HMM, whilst assuming a constant correlation structure, has extra flexibility compared to the AR(1) model in that it can produce two differing covariance structures, such that the overall spatial correlation is slightly better reproduced. Given the identification of these periods, the assumption of a state-independent correlation structure is put to question, perhaps suggesting that a correlation structure should be calibrated for each HMM state. Furthermore, observed data were found here to not be entirely consistent with the functional form used for the spatial correlation. It is suggested that the impact of assuming a particular spatial structure on state series identification should be considered carefully. We feel

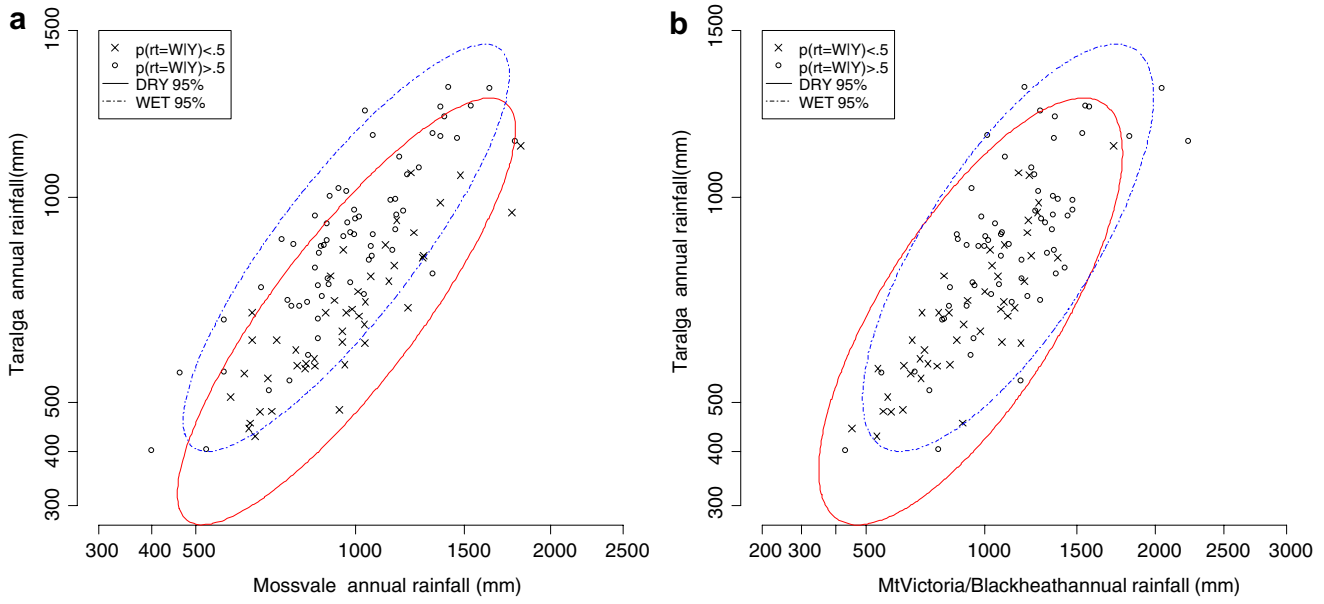


**Figure 10** Site versus site observed annual rainfall for (a) Moss Vale and Taralga (53 km apart) and (b) Mt. Victoria/Blackheath and Taralga (100 km apart). Observed data are partitioned according to $P(r_t = W|Y) < 0.5$ and $P(r_t = W|Y) > 0.5$. The HMM 95% Gaussian distribution ellipses for the wet and dry states are also plotted.

that further work on this correlation structure is required, with possible relaxation of the isotropic assumption and the introduction of a nugget term where appropriate.

As a final observation, it is suggested that implementation of the spatial correlation as function of distance between sites will prove challenging if this methodology were applied to multi-site streamflow data. Specifying the distance between sites when the streamflow data are from catchments of differing size and geometry presents an interesting challenge.

## Why is the AR(1)-oneBCphi strongly favoured?

To aid understanding of the model probabilities presented within Section ''AR(1) and HMM pooled parameter variants'', with the AR(1)-oneBCphi being strongly favoured compared to other models, the posterior distribution of key parameters are presented. The site specific posterior distributions of $\lambda$ and $\phi$ associated with the AR(1) model are presented in Figs. 11a and 11b, respectively. These distributions are compared to the associated singular regional parameter within the AR(1)-oneBCphi. The prior distribution of each parameter is also presented. Fig. 11a shows all posterior distributions of $\lambda$ having little or no density at value 1, while there is significant density at 0 whilst the bulk of the density lies between 0 and 1, further supporting the hypothesis that the BC transform is required (as discussed in Section ''Box–Cox transformation''). The site specific parameters of the AR(1) contain a large degree of overlap in terms of location of the distribution which is in agreement with the high model probabilities for the models containing a singular regional $\lambda$. The regional $\lambda$ shows a greater degree of identification (less variability) – with a distribution identified which is largely consistent with the individual site values. Fig. 11b shows the site specific posterior distributions of $\phi$ having little or no density for values at, or less than 0, with the exception of Sydney and Mt. Victoria/Blackheath. Although not as clearly as for $\lambda$, the site specific $\phi$ of the AR(1) contain a large degree of overlap in terms of location of the distribution which is in agreement with the high model probabilities for the models containing a singular regional $\phi$. The regional $\phi$ shows a greater degree of identification (less variability) – with a distribution identified which is largely consistent with the individual site values.

The posterior model probabilities favour AR(1)-oneBCphi as it is relatively parsimonious compared to the AR(1), and given there is a large region of overlap regarding the marginal distribution of $\lambda$ and $\phi$, allowing identification for the AR(1)-oneBCphi of a singular regional parameter in each case that is generally consistent with the AR(1) model. While it can be argued that the AR(1)-oneBCphi is restricting the distribution of $\phi$ in a possibly inappropriate way (ie. for Sydney and Mt. Victoria/Blackheath here), it can be conversely argued that by forcing a single regional value to be found, better identification of model parameters can be achieved by the use of a single parameter for $\lambda$ and $\phi$. In any case, the use of posterior model probabilities through model averaging ensures that both models are capable of contributing during simulation. An alternative to using model averaging and individual site versus regionally pooled model is by creating a further hierarchy in model parameters $\lambda$ and $\phi$, as advocated by Gelman et al. (1995, p. 176). While this is beyond the scope of this study, the formulation of such a model is an intended area of future research.

## Impact of parameter and model uncertainty

The reservoir simulations revealed that rainfall simulations that incorporate parameter uncertainty produced drought probabilities approximately a half an order of magnitude

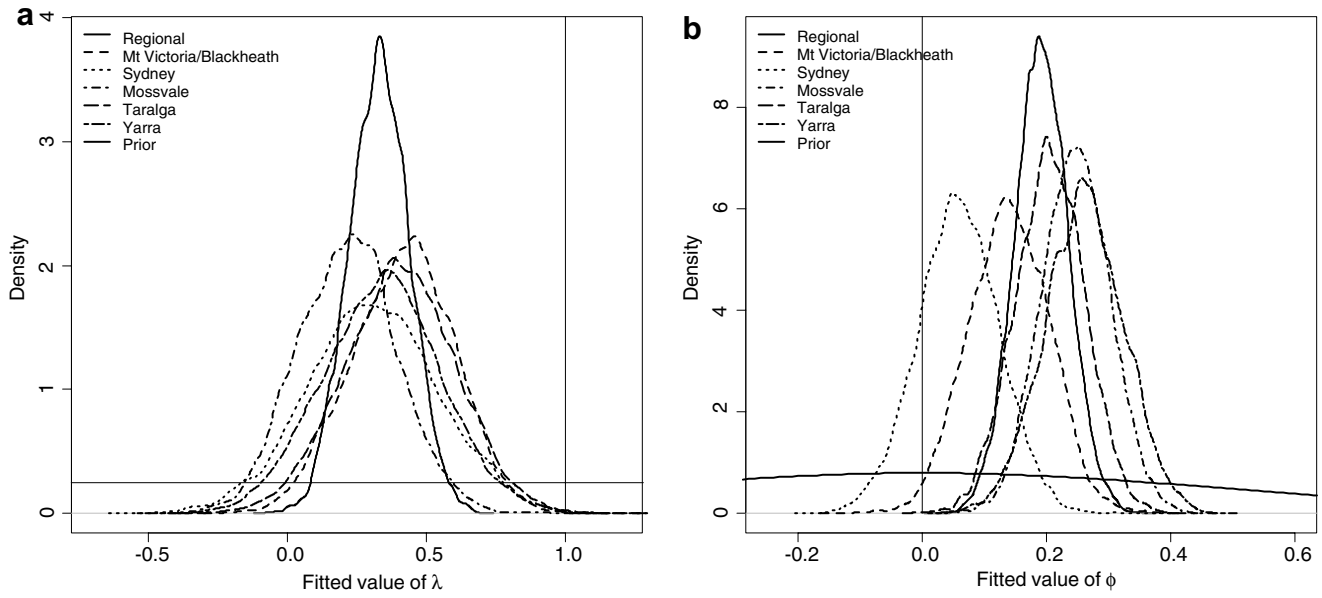


**Figure 11** Posterior marginal density plots of (a) $\lambda$ and (b) $\phi$ for the individual site AR(1) and regional AR(1)-oneBCphi parameters.

higher for severe drought events compared to rainfall simulations based on single value parameter estimates. To investigate why this occurred the relationship between the average catchment rainfall (based on simulated time series of 10,000 years) and drought probability for each of the 1000 replicates was examined for the two cases: with and without parameter uncertainty. Fig. 12 shows the drought probabilities for the security criterion with rainfall simulated by the AR(1) model and an annual supply of 450 GL. The diagram shows a boxplot of the drought probabilities when the average catchment rainfall is one of the four quartiles. With no parameter uncertainty the variability of the average catchment rainfall is very low (standard deviation = 2.7 mm) and the distribution of drought probabilities does not change for different average rainfall quartiles (Fig. 12a). In contrast, with parameter uncertainty the variability of the average catchment rainfall increases by an order of magnitude (standard deviation = 28 mm). Fig. 12b shows the impact that this increased variability has on drought probabilities. When the average catchment rainfall is in the 1st quartile, the expected drought probability is 0.2%, which is four times higher than the expected drought probability in 2nd quartile (0.05%). This is attributable to the greater variability of the generated rainfall (with relatively low rainfall being more likely to be generated under parameter uncertainty) and also to the nonlinear relationship between catchment runoff and reservoir yield. Thus the drought probabilities increase when parameter uncertainty is evaluated due to the increased rainfall variability and the nonlinear relationship between catchment rainfall and drought probabilities.

These results have important implications for water resource planners who rely on estimates of reservoir yield. For the case study, which had similar characteristics to the Sydney system, evaluating parameter uncertainty resulted in a decrease in the yield of the order of 10%. Importantly, if parameter uncertainty is ignored then reservoir yield will be overestimated.

Evaluating model uncertainty through simulation of the HMM and AR(1) showed significant differences in the estimated yield — with a range of −1.2% to 2.4% compared to the best estimate. Averaging these models within simulation

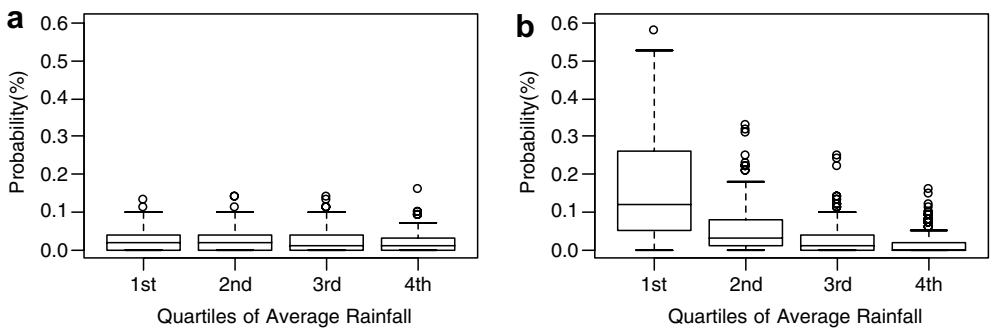


**Figure 12** Impact on drought probability (i.e. security criterion — years < 5% storage) of simulated rainfall variability for the AR(1) model (a) without parameter uncertainty and (b) with parameter uncertainty with annual supply = 450 GL.

provided a similar estimate to the best model identified through parameter pooling — indicating the benefit of model averaging. These results indicate that for this case study differences in yield due to model uncertainty are not as great compared to differences due to parameter uncertainty.

These sobering results are based on analysis of rainfall variability using stationary stochastic models and therefore do not take into account the possible impacts of climate change. Generalisation of the models to incorporate possible non-stationarity along with natural variability is an area of future research.

## Conclusion

A modelling framework has been developed for modelling multi-site hydrological data at annual timescales. Models included within this framework are the HMM (Thyer and Kuczera, 2003a) and the widely used lag-1 autoregressive (AR(1)) model (Srikanthan and McMahon, 1985). The aim was to evaluate and compare these multi-site models using a comprehensive and consistent Bayesian methodology.

For both models several extensions to the previous implementations were developed. For the HMM, the inclusion of Box–Cox transformation was undertaken to facilitate simulation of highly skewed data. For the AR(1) model, the full evaluation of parameter uncertainty and inclusion of the BC transformation in a multi-site framework also represents an advance on previous work. For both models, the application of MCMC techniques to evaluate parameter uncertainty required the development of a parameterization for the spatial correlation. The generalized framework for calibration of both models also facilitated the testing of pooled parameter AR(1) and HMM variant models. The evaluation of model performance included the ability to reproduce a wide range of important extreme statistics and the use of Bayesian model selection techniques to compare different models using estimates of the model probability.

For the five-site case study from the Sydney's major water supply catchments the following results were obtained. Firstly, the Box–Cox transformation is required for both models when modelling multi-site annual rainfall. In terms of model probability the difference between the AR(1) and HMM was marginal. The AR(1) was able to reproduce the short-term temporal correlation better than the HMM. On the other hand, the HMM was able to identify an apparent shift in observed spatial correlation structure between some of the sites.

Further research is needed on the parameterization of both spatial correlation and temporal dependence for multi-site models of long-term hydrological data. The advantage of the development of this general framework it that it allows alternative (current and future) models to be easily incorporated and their worth evaluated by a comprehensive range of techniques using a consistent Bayesian methodology.

Several pooled parameter variants were investigated. It was found that some strongly outperformed both the AR(1) and HMM demonstrating the utility of the general modelling approach coupled with a robust model selection method.

Finally, evaluation of parameter uncertainty produced significantly higher drought probabilities compared to simulations which ignored parameter uncertainty. This was due to the increased rainfall variability when incorporating parameter uncertainty and the nonlinear relationship between catchment rainfall and drought probabilities. The practical implication for water resource planners is that if parameter uncertainty is ignored the yield of a water supply system will be overestimated. For the case study undertaken, which had characteristics similar to Sydney's major water supply reservoir, this resulted in an overestimation of the reservoir yield of the order of 7%.

## Appendix A. Metropolis–Hastings algorithm

The basic idea of MCMC methods is to simulate a Markov chain iterative sequence, where for each iteration $i$, a sample of the model parameters $\boldsymbol{\theta}^{(i)}$ is generated according to a jump distribution $J_i(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i-1)})$ dependent only on the previous samples position $\boldsymbol{\theta}^{(i-1)}$. Given certain conditions the distribution of these samples converges to a stationary distribution (Mengersen and Tweedie, 1996; Roberts and Tweedie, 1996) the posterior distribution $p(\boldsymbol{\theta}|\mathbf{Y}_T)$. This distribution of parameters is used to evaluate posterior quantities of interest.

The Metropolis–Hastings algorithm (taken from Gelman et al., 2004) is described below:

1. Draw a starting point $\boldsymbol{\theta}^{(0)}$ that has a positive posterior probability.
2. For $i = 1,2,\ldots$
   (a) Sample a candidate point $\boldsymbol{\theta}^*$ from a jump distribution at iteration $i$, $J_i(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i-1)})$. The jump distribution is user chosen. A Gaussian distribution $J_i(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i-1)}) \sim N(\boldsymbol{\theta}^{(i-1)}, c^2\Sigma_J)$ with mean located at the current sample location was chosen here. The covariance matrix $\Sigma_J$ and scaling factor $c$ are user chosen and are described following this algorithm. This results in the random-walk Metropolis sampler noted in Chib and Greenberg (1995).
   (b) Calculate the ratio of densities $r = \frac{p(\boldsymbol{\theta}^*|\mathbf{Y})/J_i(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i-1)})}{p(\boldsymbol{\theta}^{(i-1)}|\mathbf{Y})/J_i(\boldsymbol{\theta}^{(i-1)}|\boldsymbol{\theta}^*)}$
   (c) Set $\boldsymbol{\theta} = \begin{cases} \boldsymbol{\theta}^* & \text{with probability } \min(r,1) \\ \boldsymbol{\theta}^{(i-1)} & \text{otherwise} \end{cases}$
   (d) Check convergence — if sufficient samples taken, stop. Otherwise continue.

Stated roughly, this algorithm will converge to the target distribution $p(\boldsymbol{\theta}|\mathbf{Y})$ (the posterior distribution) as $i \to \infty$. In practice, the sampling is stopped at a point that approximates the posterior to some degree of satisfaction. Multiple chain MCMC is used to assist determination of convergence. Using multiple chains allows testing of individual chains against all of the samples thereby allowing testing of whether each chain is yielding samples form the same distribution. The measure used in this study:

$$\sqrt{R_{GR}} = \sqrt{\text{var}(\theta_{all})/\text{var}(\theta_{chains})} \tag{17}$$

This 'scale reduction factor' introduced by Gelman and Rubin (1992) compares the estimated within-chain variance $\text{var}(\theta_{chains})$ to overall variance $\text{var}(\theta_{all})$ for each parameter. Values of $\sqrt{R_{GR}}$ below 1.2 for all parameters are recom-

mended as being acceptable (Gelman et al., 2004). Five chains of MCMC samples were used in this study.

The jump distribution determines the efficiency of the sampler. Generally, the better the jump distribution approximates the posterior, the more efficient the sampler will be. The following methods were used to increase the efficiency of the sampler in this study:

1. All chains were started at the mode of the distribution (Kuczera and Parent, 1998), thereby reducing the chance of some chains taking a very long time to reach areas of high posterior probability. The shuffled complex evolution algorithm of Duan et al. (1994) was used to find the modal estimate.
2. An initial estimate of the posterior covariance based on the Taylor series expansion of the log posterior density at the mode is used in this study. For each component of the Gaussian covariance matrix a finite difference scheme to is used to estimate $\Sigma_{ik}^{-1} = -\partial \log p(\boldsymbol{\theta} \mid \mathbf{Y}_T)/\partial\theta_i\partial\theta_k$. Gelman et al. (2004) recommend scaling the covariance matrix by the factor $c^2$ where $c = 2.4/\sqrt{n}$ where $n$ is the number of parameters based on empirical studies of Gaussian posteriors.
3. Four stages of sampling are used each containing 50,000 samples:
   (a) A heuristic updating scheme based on Gelman et-al. (2004) periodically updating the scaling factor $c^2$ of the jump distribution covariance to achieve an optimal jumping rate (average number of accepted jumps per iteration).
   (b) A constant covariance stage is then undertaken with no updating with the final scaling factor $c^2$ from the previous stage used throughout.
   (c) The covariance of the parameters is estimated from the previous stage samples. The scaling factor is reset to $c = 2.4/\sqrt{n}$. The adaptive scheme of Haario et al. (2001) is then employed. This is considered as the 'warm up' of the sampler.
   (d) The scheme of Haario et al. (2001) is further used to produce the final samples – with the $\sqrt{R_{GR}}$ statistic checked to be below 1.2 for all parameters.

## References


Akintug, B., Rasmussen, P.F., 2005. A Markov switching model for annual hydrologic time series. Water Resources Research 41 (9), W09424. doi:10.1029/2004WR00360.

Andrieu, C., Robert, C.P., 2001. Controlled MCMC for optimal sampling. Cahiers du Ceremade, 0125.

Bengio, Y., 1999. Markovian models for sequential data. Neural Computing Surveys 2, 129–162.

Beran, J., 1994. Statistics for long-memory processesMonographs on Statistics and Applied Probability, vol. 61. Chapman & Hall, New York, 315 pp.

Berger, J.O., 2000. Bayesian analysis: a look at today and thoughts of tomorrow. Journal of the American Statistical Association 95 (452), 1269–1276.

Berger, J.O., Bernardo, J.M., 1992. On the development of reference priors. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, vol. 4. Oxford University Press, Oxford.

Bernardo, J.M., 1979. Reference posterior distributions for Bayesian inference (with discussion). Journal of the Royal Statistical Society Series B-Methodological 36, 192–236.

Bernardo, J.M., Smith, A.F.M., 2000. Bayesian Theory Wiley Series in Probability and Statistics. Wiley Ltd., Chichester.

Box, G.E.P., Cox, D.R., 1964. An Analysis of Transformations. Royal Journal of Statistical Society Series B 26 (2), 211–252.

Carlin, B.P., Louis, T.A., 2000. Bayes and Empirical Bayes Methods for Data Analysis Texts in Statistical Science. Chapman & Hall, New York.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis–Hastings algorithm. American Statistician 49 (4), 327–335.

Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the Metropolis–Hastings output. Journal of the American Statistical Association 96 (453), 270–281.

Chipman, H., George, E.I., McCulloch, R.E., 2001. The practical implementation of Bayesian model selection. IMS Lecture Notes – Monograph Series 38, 67–107.

Duan, Q., Sorooshian, S., Gupta, V.K., 1994. Optimal use of the SCE-UA GLobal optimization method for calibrating watershed models. Journal of Hydrology 158, 265–284.

Fortin, V., Perreault, L., Salas, J.D., 2004. Retrospective analysis and forecasting of streamflows using a shifting level model. Journal of Hydrology 296 (1–4), 135–163.

Frost, A.J., 2004. Spatio-temporal hidden Markov models for incorporating inter-annual variability in rainfall. Ph.D. thesis, University of Newcastle, Newcastle, Australia, 240 pp. <www.newcastle.edu.au/service/library/adt/public/adt-NNCU20051126.010618>.

Gelfand, A.E., Dey, D.K., 1994. Bayesian model choice,asymptotics and exact calculations. Journal of the Royal Statistical Society Series B-Methodological 56 (3), 501–514.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7, 457–472.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. Bayesian Data Analysis Chapman & Hall Texts in Statistical Sciences. Chapman and Hall, London.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis Texts in Statistical Science. Chapman & Hall/CRC, New York.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721–741.

Genz, A., 1993. Comparison of methods for the computation of multivariate normal probabilities. Computing Science and Statistics 25, 400–405.

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.

Haario, H., Saksman, E., Tamminen, J., 1999. Adaptive proposal distribution for random walk Metropolis algorithm. Computational Statistics 14, 375–395.

Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive metropolis algorithm. Bernoulli 7 (2), 223–242.

Hipel, K.W., McLeod, A.I., 1994. Time Series Modelling of Water Resources and Environmental Systems. Elsevier, Amsterdam.

Hurst, H.E., 1951. Long term storage capacities of reservoirs. Transactions American Society of Civil Engineers 116, 770–799.

Jeffreys, H., 1961Theory of Probability, vol. viii. Clarendon Press, Oxford, 447 pp..

Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association 90 (430), 773–795.

Kiem, A.S., Franks, S.W., 2004. Multi-decadal variability of drought risk, eastern Australia. Hydrological Processes 18 (11), 2039–2050.

Klemes, V., 1974. The Hurst phenomena: a puzzle? Water Resources Research 10 (4), 675–688.

Kuczera, G., 1987. On maximum likelihood estimators for the multisite lag-1 streamflow model: complete and incomplete data cases. Water Resources Research 23 (4), 641–645.
Kuczera, G., Parent, E., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. Journal of Hydrology 211 (1–4), 69–85.
Lambert, M.F., Whiting, J.P., Metcalfe, A.V., 2003. A non-parametric hidden Markov model for climate state identification. Hydrology and Earth System Sciences 7 (5), 652–667.
Lee, P.M., 1989Bayesian Statistics: An Introduction, vol. X. Oxford University Press, New York, 294.
Lindley, D.V., 1957. A statistical paradox. Biometrika 44, 187–192.
Mandelbrot, B.B., Wallis, J.R., 1969. Computer experiments with fractional Gaussian noises, part 1, averages and variances. Water Resources Research 5 (1), 228–241.
Matalas, N., 1967. Mathematical assessment of synthetic hydrology. Water Resources Research 3 (4), 937–945.
McMahon, T.A., Mein, R.G., 1986River and Reservoir Yield, vol. ix. Water Resources Publications, Littleton, CO, USA, 368.
Mengersen, K.L., Tweedie, R.L., 1996. Rates of convergence of the hastings and metropolis algorithms. Annals of Statistics 24 (1), 101–121.
Montanari, A., Rosso, R., Taqqu, M.S., 1997. Fractionally differenced ARIMA models applied to hydrologic time series: identification, estimation and simulation. Water Resources Research 33, 1035–1044.
Ntzoufras, I., 1999. Aspects of Bayesian model and variable selection using MCMC. Ph.D. thesis, Athens University of Economics and Business, Athens.
Perreault, L., Bernier, J., Bobee, B., Parent, E., 2000a. Bayesian change-point analysis in hydrometeorological time series Part 1 The normal model revisited. Journal of Hydrology 235 (3–4), 221–241.
Perreault, L., Parent, E., Bernier, J., Bobee, B., Slivitzky, M., 2000b. Retrospective multivariate Bayesian change-point analysis: a simultaneous single change in the mean of several hydrological sequences. Stochastic Environmental Research and risk Assessment 14, 243–261.
Pittock, A.B., 1975. Climatic change and the patterns of variation in australian rainfall. Search 6, 498–504.
Potter, K.W., 1976. A stochastic model of the Hurst phenomenon: nonstationarity in hydrologic processes. Ph.D. thesis, John Hopkins University.
Roberts, G.O., Tweedie, R.L., 1996. Geometric convergence and central limit theorems for multidimensional hastings and metropolis algorithms. Biometrika 83 (1), 95–110.
Rodriguez-Iturbe, I., Mejia, J.M., Dawdy, D.R., 1972. Streamflow simulation 1, a new look at Markovian models, fractional Gaussian noise and crossing theory; 2, the broken line process as a potential model for hydrologic simulation. Water Resources Research, 921–941.
Salas, J., 1993. Analysis and modeling of hydrologic time series. In: Maidment, D. (Ed.), Handbook of Hydrology. McGraw-Hill, New York, USA, pp. 19.1–19.7.
Salas, J.D., Boes, D.C., 1980. Shifting level modeling of hydrologic series. Advances in Water Resources 3, 59–63.
Sanso, B., Guenni, L., 2000. A nonstationary multisite model for rainfall. Journal of the American Statistical Association 95 (452), 1089–1100.
SCA, 2000. Operating License. Sydney Catchment Authority, 61 pp.
Srikanthan, R., McMahon, T.A., 1985Stochastic Generation of Rainfall and Evaporation Data, vol. 84. Australian Water Resources Council, Canberra.
Srikanthan, R., McMahon, T.A., Pegram, G.G.S., Thyer, M., Kuczera, G., 2001. Generation of Annual Rainfall for Australian Stations. Cooperative Research Centre for Catchment Hydrology, Melbourne.
Stedinger, J.R., Taylor, M.R., 1982. Synthetic streamflow generation, 1: model verification and validation. Water Resources Research 18 (4), 909–918.
Stephens, M., 1997. Bayesian methods for mixtures of normal distributions. Ph.D. thesis, University of Oxford, Oxford.
Sveinsson, O.G.B., Salas, J.D., Boes, D.C., Pielke Sr., R.A., 2003. Modeling the dynamics of long term variability of hydroclimatic processes. Journal of Hydrometeorology 4 (2), 489–505.
SWC, 2005. Annual Report 2004–2005. Sydney Water Corporation, 68 pp.
Thyer, M.A., 2001. Modelling long-term persistence in hydrological time series. Ph.D. thesis, University of Newcastle, Newcastle, Australia. <www.newcastle.edu.au/services/library/adt/public/adt-NNCU20020531.035349/index.html>.
Thyer, M., Kuczera, G., 2000. Modeling long-term persistence in hydroclimatic time series using a hidden state Markov model. Water Resources Research 36 (11), 3301–3310.
Thyer, M., Kuczera, G., 2003a. A hidden Markov model for modelling long-term persistence in multi-site rainfall time series 1 Model calibration using a Bayesian approach. Journal of Hydrology 275, 12–26.
Thyer, M., Kuczera, G., 2003b. A hidden Markov model for modelling long-term persistence in multi-site rainfall time series 2 Real data analysis. Journal of Hydrology 275, 27–48.
Thyer, M., Kuczera, G., Wang, Q.J., 2002. Quantifying parameter uncertainty in stochastic models using the Box–Cox transformation. Journal of Hydrology 265 (1–4), 246–257.
Thyer, M., Frost, A.J., Kuczera, G., 2006. Parameter estimation and model identification for stochastic models of annual hydrological data: is the observed record long enough? Journal of Hydrology 330 (1–2), 313–328.
Verdon, D.C., Wyatt, A.M., Kiem, A.S., Franks, S.W., 2004. Multidecadal variability of rainfall and streamflow: Eastern Australia. Water Resources Research 40 (W10201). doi:10.1029/2004WR00323.
Vrugt, J., Gupta, H., Bouten, W., Sorooshian, S., 2003. A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resources Research 39 (8), 10.1029/2002WR001642..
Wikle, C.K., 2003. Hierarchical models in environmental science. International Statistical Review 71 (2), 181–200.