ELSEVIER

Journal of Hydrology 265 (2002) 246–257

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Quantifying parameter uncertainty in stochastic models using the Box–Cox transformation

Mark Thyer[a,1], George Kuczera[a,*], Q.J. Wang[b,2]

[a]*Department of Civil, Surveying and Environmental Engineering, University of Newcastle, Callaghan, NSW 2308, Australia*
[b]*Agriculture Victoria, Private Bag 1, Ferguson Road, Tatura, VIC 3616, Australia*

Received 6 July 2001; revised 3 May 2002; accepted 7 May 2002

Abstract

The Box–Cox transformation is widely used to transform hydrological data to make it approximately Gaussian. Bayesian evaluation of parameter uncertainty in stochastic models using the Box–Cox transformation is hindered by the fact that there is no analytical solution for the posterior distribution. However, the Markov chain Monte Carlo method known as the Metropolis algorithm can be used to simulate the posterior distribution. This method properly accounts for the nonnegativity constraint implicit in the Box–Cox transformation. Nonetheless, a case study using the AR(1) model uncovered a practical problem with the implementation of the Metropolis algorithm. The use of a multivariate Gaussian jump distribution resulted in unacceptable convergence behaviour. This was rectified by developing suitable parameter transformations for the mean and variance of the AR(1) process to remove the strong nonlinear dependencies with the Box–Cox transformation parameter. Applying this methodology to the Sydney annual rainfall data and the Burdekin River annual runoff data illustrates the efficacy of these parameter transformations and demonstrate the value of quantifying parameter uncertainty.

Keywords: Lag-one autoregressive models; Markov chain Monte Carlo methods; Metropolis algorithm; Parameter uncertainty; Box–Cox transformation

1. Introduction

The Box–Cox transformation (Box and Cox, 1964) is widely used to transform hydrological data to ensure the transformed data are approximately Gaussian (Chander et al., 1978; Hirsch, 1979; Jain and Singh, 1986; Salas, 1993). However, the proper quantification of parameter uncertainty in models using the Box–Cox transformation is rarely attempted. This technical note describes a robust method for quantifying uncertainty when using the Box–Cox transformation.

The technical note uses as a case study the lag-one autoregressive model (AR(1)) which is commonly recommended for simulating annual hydrological time series (Grayson et al., 1996; Salas, 1993; Srikanthan and McMahon, 2000). To apply this model the usual approach taken is to use maximum

* Corresponding author. Tel.: +61-2-49-216038; fax: +61-2-49-216991.
E-mail addresses: cegak@cc.newcastle.edu.au (G. Kuczera), thyer@interchange.ubc.ca (M. Thyer), qj.wang@nre.vic.gov.au (Q.J. Wang).
[1] Present address: Department of Forest Resource Management, University of British Columbia, Vancouver, BC, Canada V6T-1Z4. Fax: +1-604-822-9106.
[2] Fax: +61-3-5833-5337.

0022-1694/02/$ - see front matter
PII: S0022-1694(02)00113-0

likelihood estimates of the AR(1) model parameters to generate the synthetic hydrological time series. This approach ignores parameter uncertainty. Stedinger and Taylor (1982) have shown that incorporating parameter uncertainty increases the simulated drought risks and is as important as choosing the correct model.

The goal is to quantify the parameter uncertainty of the AR(1) model when used in conjunction with the Box–Cox transformation to normalise hydrological data. A Bayesian approach will be used to calculate the posterior distribution of the AR(1) model parameters. When the historical hydrological time series follows a Gaussian distribution it is possible to derive an analytical expression for the posterior distribution (Stedinger and Taylor, 1982). However, often hydrological data do not follow a Gaussian distribution and it is necessary to apply an appropriate transformation, such as the Box–Cox transformation (Box and Cox, 1964), to ensure approximate Gaussian behaviour. Unfortunately it is no longer possible to derive an analytical expression for the posterior distribution when a Box–Cox transformation is used. It is noted that Box and Tiao (1973) provide approximate analytical expressions for the posterior distribution but disregard the truncation constraint introduced by the transformation.

The contribution of this technical note will be to describe a method for calculating the posterior distribution of the AR(1) model parameters when a Box–Cox transformation is used. The approach will employ the Metropolis algorithm, a Markov chain Monte Carlo (MCMC) method, to simulate the posterior distribution of the model parameters. MCMC methods are commonly used in cases such as this one where it is not possible to derive analytical expressions for the posterior (Gelman et al., 1995). However, the implementation of the Metropolis algorithm is not necessarily a straightforward task. This was the case for the AR(1) model. This note describes the problems implementing the Metropolis algorithm and a robust resolution of the problems.

This technical note is organised as follows: after outlining the AR(1) model the methodology used to implement the Metropolis algorithm is described. Three synthetic data case studies will be used to verify that the MCMC procedure is able to recover the true synthetic parameter values. Following that the results of two case studies using real hydrological data, the Sydney annual rainfall data and the Burdekin River annual runoff data, will be presented to demonstrate the importance of assessing parameter uncertainty. These results will also be compared to the analytical expressions given by Box and Tiao (1973) to illustrate the effect of ignoring the truncation.

2. AR(1) model

The AR(1) model has the following form:

$$z_t = \mu + \phi_1(z_{t-1} - \mu) + \varepsilon_t \tag{1}$$

where z_t is the value of the time series at time step, t, μ is the mean of the time series, ϕ_1 is the lag-one autoregressive parameter and ε_t is an uncorrelated Gaussian random variable, with zero mean and variance σ_ε^2, such that $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

The first step in applying the AR(1) model is to calibrate its parameters using the hydrological time series, denoted as $Y_N = \{y_1, \ldots, y_n\}$. From Eq. (1) it can be seen that given the value for z_{t-1}, z_t must follow a Gaussian distribution, such that $z_t|z_{t-1} \sim N(\mu + \phi_1(z_{t-1} - \mu), \sigma_\varepsilon^2)$. Therefore, to calibrate the AR(1) model the hydrological data Y_N must also follow a Gaussian distribution. When this is not the case, a Box–Cox transformation (Box and Cox, 1964) is commonly applied such that

$$z_t = \begin{cases} \dfrac{y_t^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log y_t & \lambda = 0 \end{cases} \tag{2}$$

where λ is the transformation parameter chosen to ensure the transformed hydrological data Z are approximately Gaussian. It is important to note that this transformation places a constraint on the transformed values such that $z_t\lambda + 1 > 0$. Hence the z_t's actually follow a truncated Gaussian distribution (refer to Appendix A for more detatils). This truncation was ignored by Box and Tiao (1973) when deriving their approximate analytical expressions for the AR(1) posteriors. The transformation parameter is included in the vector of unknown

AR(1) model parameters, which is defined as

$$\theta' = (\mu, \sigma_\varepsilon, \phi_1, \lambda) \tag{3}$$

3. Model calibration—the Metropolis algorithm

The goal of model calibration is to infer the posterior distribution of the parameters conditioned on the observed data, $p(\theta|Y_N)$. In the case of the AR(1) model used together with the Box–Cox transformation it is not possible to derive an analytical expression for the posterior distribution—refer to Appendix A. Hence an MCMC method known as the Metropolis algorithm is used to simulate values from the posterior distribution.

Tierney (1994) provides a detailed description of the theoretical aspects of MCMC merthods. Chib and Greenberg (1995) provide a more intuitive tutorial of MCMC methods and the Metropolis algorithm, while Gelman et al. (1995) provide useful tips for their application. Here, a brief description will be given. The Metropolis algorithm produces a Markov chain sequence of samples that constitute a random walk in the parameter space. Each iteration of the algorithm proceeds by first generating candidate parameter samples using a suitable arbitrary probability distribution refered to as the jump distribution. These candidates are then either accepted or rejected using a criterion which ensures the algorithm is sampling from the posterior distribution when the Markov chain has become stationary.

The application of the Metropolis algorithm for drawing $\{\theta_1, \ldots, \theta_n\}$ samples from the posterior is outlined as follows:

Step 1. Initialise θ with arbitrary starting value θ^0.
Step 2. Repeat for $i = 1, 2, \ldots, n$
  Generate a candidate θ^* from $q(\cdot|\theta^i)$ and u from $U(0, 1)$
  If $u \leq \alpha(\theta^*|\theta^i)$
    set $\theta^{i+1} = \theta^*$
  else
    set $\theta^{i+1} = \theta^i$
Step 3 Return the values $\{\theta^1, \theta^2, \ldots, \theta^n\}$.

where $q(\cdot|\theta^i)$ is the jump density, which generates a candidate sample based on the previous sample θ^i, and $\alpha(\theta^*|\theta^i)$ represents the *jump probability*, which is the probability of jumping from θ^i to θ^*. When using the Metropolis algorithm to simulate Bayesian posteriors the jump probability is calculated using the ratio of the posterior density for candidate θ^* and the current sample θ^i. When the jump density is chosen as symmetric, such that $q(\theta^j|\theta^i) = q(\theta^i + \theta^j)$, then the probability of jump is calculated as

$$\alpha(\theta^*|\theta^i) = \min\left[\frac{p(Y_N|\theta^*)p(\theta^*)}{p(Y_N|\theta^i)p(\theta^i)}, 1\right] \tag{4}$$

where $p(Y_N|\theta)$ is the likelihood function and $p(\theta)$ is the prior density function. If uniform priors are assumed then the jump probability simplifies to

$$\alpha(\theta^*|\theta^i) = \min\left[\frac{p(Y_N|\theta^*)}{p(Y_N|\theta^i)}, 1\right] \tag{5}$$

Hence, all that is required is the ability to calculate the likelihood density for a particular set of parameter values. For the AR(1) model the constraint imposed by the Box–Cox transformation complicates the evaluation of the likelihood function—see Appendix A for derivation of the likelihood function.

3.1. Implementation

The Metropolis algorithm may seem simple enough but there are some issues that need to be addressed to ensure successful implementation.

The Metropolis algorithm requires that the parameter vector be first initialised with arbitrary starting values. To provide the arbitrary starting values for the AR(1) model an optimisation algorithm was used to estimate the parameters which correspond to the mode, θ_m, of the likelihood function. The Metropolis algorithm also requires the selection of a suitable jump density. In this application the multivariate Gaussian density was used. This jump distribution requires the specification of its location (defined by the mean vector $\boldsymbol{\mu}_\mathbf{J}$) and its scale (or spread), as defined by the covariance matrix $\boldsymbol{\Sigma}_\mathbf{J}$. In this algorithm the location was set to the current value of the parameter samples θ^i. This has the advantage that only the spread of the jump distribution needs to be tuned while the algorithm is proceeding. Hence the jump distribution used to generate candidates in this

algorithm was

$$\theta^* \sim N_r\left(\theta^i, \mathbf{\Sigma_J}\right) \tag{6}$$

where r is the dimension of the multivariate distribution which is equal to the number of elements in the parameter vector θ.

The spread of the jump distribution has important implications for the efficiency of the algorithm. It has a major influence on the acceptance rate, which is the percentage of times a candidate sample is accepted. If the spread of the jump distribution is too large relative to the posterior then the acceptance rate will be very low and the algorithm will converge slowly. On the other hand if the spread is too small then the acceptance rate will be higher. However, the algorithm will take a longer time to explore the full region of the posterior and hence the tails of the distribution may be undersampled. In practise the spread is tuned during the iterations of the Metropolis algorithm to ensure the acceptance rate remains within a suitable range.

The covariance matrix of the jump distribution is given starting values based on a Taylor series expansion of the log posterior density at the mode, where each element of the covariance matrix is calculated as

$$\Sigma_{ij}^{-1} = \left.\frac{-\partial \log p(\theta|Y_N)}{\partial\theta_i\,\partial\theta_j}\right|_{\theta=\theta_m} \tag{7}$$

The derivatives given above are approximated using a finite difference scheme. Gelman et al. (1995) recommend scaling this initial guess by the factor c^2 where $c = 2.4/\sqrt{r}$.

After a number of iterations the covariance matrix of the jump distribution is updated to be proportional to the estimated posterior covariance matrix calculated from the samples already completed. This produces an adaptive simulation algorithm. Gelman et al. (1995) state that the optimal acceptance rate for a multivariate Gaussian distribution is around 0.44 in one dimension reducing to about 0.23 in higher dimensions ($r > 5$). In a general context these optimal acceptance rates are only approximate because they are based on the assumption that the posterior is also a multivariate Gaussian distribution, which may not be the case. In this application the covariance matrix was scaled to bring the acceptance rate toward these approximate optimal values.

Once initialised the Metropolis algorithm is allowed to continuously sample until the Markov chain induced by the Metropolis algorithm has converged to a stationary distribution. Once converged the samples from the Metropolis algorithm can be considered to be samples from the posterior. The most critical issue in the implementing MCMC methods is how to determine whether convergence has been achieved. The question is "How many iterations, say b, should the Markov chain be allowed to 'warm-up' before the simulated output can be treated as samples from the posterior?" As there is no general technique for determining the number of iterations required for convergence, some form of analysis must be performed on the Metropolis algorithm to assess convergence. Such methods are collectively known as convergence diagnostic tools.

Cowles and Carlin (1996) provide an expository review of numerous convergence diagnostic tools. Their recommendations to use a variety of diagnostic tools and multiple independent parallel Markov chains were adopted in this study. Compared to a single Markov chain, multiple chains are able to more widely explore the parameter space. Ten chains with 1000 samples in each chain producing a total of 10,000 samples were found to provide enough samples for a good approximation to the posterior. Each of these multiple chains was started at a point close to the mode of posterior with the starting parameter values as derived above. This contradicts Gelman et al. (1995) and Cowles and Carlin (1996) who recommended the starting points be drawn from a distribution believed to be overdispersed with respect to the stationary distribution (the posterior). However, if the jump distribution is a poor approximation to the posterior, it is our experience that sampling from an overdispersed distribution is counterproductive and hinders convergence. Once initialised, each of the 10 chains was allowed to independently explore the parameter space. During sampling $\mathbf{\Sigma_J}$ was updated every time all of the ten chains had completed 1000 samples by estimating the covariance matrix of the 10,000 samples. Initially using a different covariance matrix for each chain was trialled. However, it was found that using the same covariance matrix for all the

chains increased the mixing rate and therefore increased the convergence rate.

To assess the convergence of these multiple chains the R statistic (as defined by Gelman et al. (1995)) was monitored for each of the parameters. The R statistic is a measure of the between-chain and within-chain variances. If the multiple chains are not mixing properly in the parameter space the R statistic will be relatively high. As the calculations for the R statistic are based on Gaussian theory approximations, the reliability of this diagnostic may be questionable for nonGaussian applications (Cowles and Carlin, 1996). In addition to the R statistic, time series plots of the percentiles of the sample distributions were monitored. These provide an indication when the sample distributions become stationary. The acceptance rate was also plotted for each of the chains to ensure that it remained in the vicinity of the optimal values as given above. If all these diagnostic plots did not show any signs of convergence failure then the output from the Metropolis algorithm was treated as samples from the posterior $p(\theta|Y_N)$.

3.2. Development of parameter transformations

During initial calibration runs it was found that the Metropolis algorithm had a very low acceptance rate (< 0.1), and continued to stay low no matter how the covariance matrix of the jump distribution was scaled. Investigations revealed there was a distinct curvilinear relationship between the Metropolis samples of the Box–Cox λ and the mean μ and the standard deviation σ_ε, as shown in Figs. 1(a) and 2(a), respectively, for the Sydney annual rainfall data (introduced in Section 4.2.1). The multivariate Gaussian jump distribution is unable to provide a good approximation to this kind of curvature. This is why the acceptance rate was very low. The cause of this problem is that the Metropolis algorithm is sampling the parameters μ and σ_ε which have a strong nonlinear dependence on the transformation parameter λ. Therefore, it was deemed necessary to develop suitable parameter transformations that removed this dependence. First-order approximations (refer to Appendix B) provided the motivation for using the following expressions as parameter transformations:

$$\mu = \begin{cases} \dfrac{m_y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log m_y & \lambda = 0 \end{cases} \tag{8a}$$

$$\sigma_\varepsilon = m_y^{\lambda-1} s_y \sqrt{1 - \phi_1^2} \tag{8b}$$

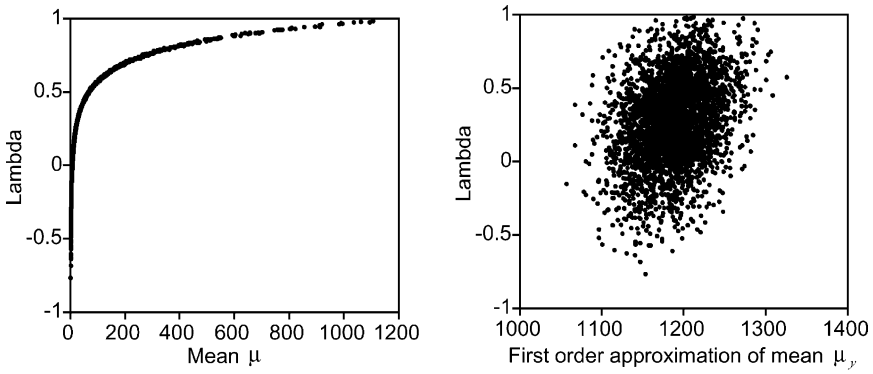
where m_y and s_y represent a first-order approximation to the expected value and the standard deviation of the untransformed rainfall data, Y. Using these expressions the following alternative scheme was implemented: use the Metropolis algorithm to sample the parameters m_y and s_y and then apply the parameter transformations given in Eqs. (8a) and (8b) to calculate the AR(1) likelihood function. This has the advantage that the parameters m_y and s_y are virtually independent of λ, as demonstrated in Figs. 1(b) and 2(b). Hence the multivariate Gaussian jump distribution is able to provide a much better approximation when sampling in the transformed parameter space of m_y and s_y. This is illustrated by the acceptance rates which increased to the optimal range of 0.2–0.4. Hence the parameter transformations given by Eqs. (8a) and (8b) provide a general approach for the implementation of the Metropolis algorithm for the simulation of posteriors of models using the Box–Cox transformation.

4. Case studies

4.1. Synthetic data

Synthetic data generated using the AR(1) model was analysed to verify that the MCMC procedure outlined previously was able to recover the known synthetic parameter values. Three sets of synthetic parameter values are used to test the MCMC procedure under a variety of conditions. For each set synthetic time series with $n = 100$, 1000 and 10,000 data points were generated because as the number of data points increases it would be expected that the posterior would converge to the true synthetic parameter value.

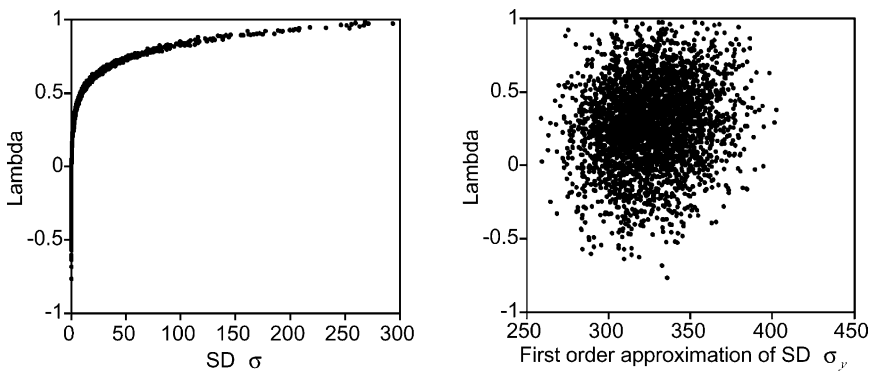
The true synthetic parameter values and the posterior results for each synthetic series are summarised in Table 1. For every parameter of all the series the true value is within one standard deviation



(a) Transformed space (b) Untransformed space

Fig. 1. Scatter plot of Metropolis algorithm samples showing relationship of the mean μ in transformed space and its first-order approximation μ_y in untransformed space to λ for the Sydney annual rainfall data.

of the posterior mean. This indicates that the MCMC procedure is able to recover the true synthetic parameter value. Also note that the decrease in the posterior standard deviation as the sample size of the synthetic data increases indicates that the posteriors do converges to the true parameter values. The difference between the series S1 and S2 is that the λ parameter was increased from 0.1 to 1.0. This illustrates that the MCMC procedure is robust across a range of values for λ. The MCMC posterior of λ for S1, $n = 100$ series is compared to the posterior calculated using the analytical expression given by Box and Tiao (1973) in Fig. 3(a). There is little difference between the two posteriors. A similar result is found for series S2 (not shown). This demonstrates that ignoring the truncation has little effect on the posteriors for series S1 and S2. This is to be expected as the proportion of the transformed data distribution that is truncated is negligible ($< 0.1\%$) for the S1 and S2 parameter sets. However, when σ_ε is increased to 5.0 in synthetic series S3 this proportion increases to approximately 11%. Fig. 3(b) compares Box and Tiao's analytical posterior to the MCMC samples for series S3 and shows there is a considerable difference between the two. In fact, the true synthetic parameter value is located only just $[P(\lambda < 0.5|Y_N) < 0.4\%]$ within the Box and Tiao's analytical posterior. This leads to the conclusion that Box and Tiao's (1973) analytical expressions may give misleading results for the posteriors when the truncation is significant.



(a) Transformed space (b) Untransformed pace

Fig. 2. Scatter plot of Metropolis algorithm samples showing relationship of the standard deviation (SD) σ in transformed space and its first-order approximation σ_y in untransformed space to lambda λ for the Sydney annual rainfall data.

Table 1
MCMC posterior results for the synthetic data

Parameter	True value	Posterior mean (SD)		
		$n = 100$	$n = 1000$	$n = 10{,}000$
Synthetic series S1				
μ	5.0	4.58 (1.16)	5.07 (0.39)	4.91 (0.11)
σ_ε	1.0	0.75 (0.41)	1.01 (0.14)	0.97 (0.04)
ϕ_1	0.5	0.43 (0.09)	0.49 (0.03)	0.50 (0.01)
λ	0.1	0.05 (0.11)	0.11 (0.03)	0.093 (0.01)
Synthetic series S2				
μ	5.0	3.98 (1.58)	5.16 (0.77)	4.82 (0.22)
σ_ε	1.0	0.70 (0.45)	1.04 (0.23)	0.95 (0.07)
ϕ_1	0.5	0.43 (0.10)	0.89 (0.07)	0.50 (0.01)
λ	1.0	0.73 (0.36)	0.48 (0.03)	0.97 (0.04)
Synthetic series S3				
μ	5.0	5.02 (1.83)	4.61 (0.98)	4.94 (0.17)
σ_ε	5.0	6.16 (2.43)	5.84 (1.04)	4.85 (0.26)
ϕ_1	0.5	0.47 (0.13)	0.47 (0.05)	0.50 (0.02)
λ	0.5	0.55 (0.11)	0.55 (0.05)	0.49 (0.01)

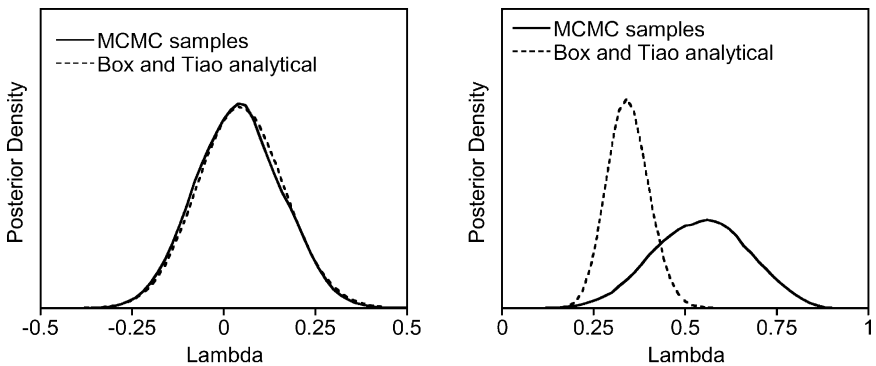
4.2. Real hydrological data

4.2.1. Sydney annual rainfall

The Sydney annual rainfall data was chosen for the first case study of real hydrological data. Sydney is a city located at a latitude of approximately 34°S on the east coast of Australia. Monthly rainfall data was available for the period January 1859 to April 1997. The monthly data set was aggregated to annual values using the September–August water year. This gauge has remained in basically the same position over the period of record, and although it may have been subject to changes in measurement technique, it is assumed to be a reasonably homogenous data set. Sample estimates for the statistical properties of the Sydney annual rainfall data are given in Table 2.

Fig. 4 shows the calibration results in the form of the posteriors for each of the AR(1) parameters. Of primary interest is the posterior of the autoregressive parameter, ϕ_1, which is a measure of the year-to-year persistence. If $\phi_1 = 0$ there is no persistence. The posterior of ϕ_1 (Fig. 4(c)) indicates there is some year-to-year persistence, although it is not considered very strong. The mode of the ϕ_1 posterior corresponds to a value of $\phi_1 = 0.17$. This compares well with the single value parameter estimate for ϕ_1 of 0.19 (Table 2), which is considered to be significant at the 95% level. This result for the single value estimate of ϕ_1 would lead to the conclusion that there is a significant lag-one correlation. In contrast, when the uncertainty of the ϕ_1 parameter is considered (Fig. 4(c)) the hypothesis that the data are uncorrelated cannot be outrightly rejected. The posterior probability that $\phi_1 < 0$, $P(\phi_1 < 0|Y_N) = 4\%$. This demonstrates the differences in the conclusions that can be drawn when parameter uncertainty is quantified.

Fig. 4(d) shows that the posterior of λ obtained from the MCMC samples is reasonably close to Box and Tiao's analytical posterior. This result is similar to synthetic series S1 and S2 and indicates that ignoring the truncation has only a minor effect on the



(a) Synthetic Series S1, $n = 100$ (b) Synthetic Series S3, $n = 100$

Fig. 3. Comparison of the posterior obtained from the MCMC samples and the analytical expression given by Box and Tiao (1973) for synthetic series S1 and S3.

Table 2
Sample estimates of the statistical properties for the real hydrological data

Series	No. of years	Mean	SD	Skew	Kurtosis	Lag-1 serial correlation coefficient
Sydney rainfall	137	1223.7	334.7	0.63	3.29	0.19
Burdekin runoff	99	65.4	62.6	2.31	11.80	0.08

posteriors for the Sydney annual rainfall. The posterior of λ also illustrates whether the application of the Box–Cox transformation to render the rainfall data approximately normal was justified. If $\lambda = 1$ then no transformation was required. The mode of the posterior corresponds to a λ value of approximately 0.3 and there is a very low posterior probability that $\lambda > 1$ ($< 0.1\%$). These results indicate that the use of the Box–Cox transformation is justified for the Sydney annual rainfall data. Fig. 6(a) shows that the observed data distribution is within the 90% confidence limits of the posterior predictive distribution (as defined by Gelman et al. (1995)). This indicates that the calibrated model is able to successfully reproduce the observed data.

4.2.2. Burdekin River annual runoff

The Burdekin River is a major tropical river system located in northern Queensland, on the east coast of Australia. Over 90% of the annual runoff occurs in the summer wet season (December–April). Annual runoff data aggregated to the October–September water year was available for the period 1894–1992. The sample estimates of the statistical properties (Table 2) indicate that the data is more highly skewed and nonGaussian than the Sydney annual rainfall data.

The posteriors for each of the AR(1) model parameters calibrated to the Burdekin River annual runoff are shown in Fig. 5. The posterior of the autoregressive parameter ϕ_1 (Fig. 5(c)) indicates that

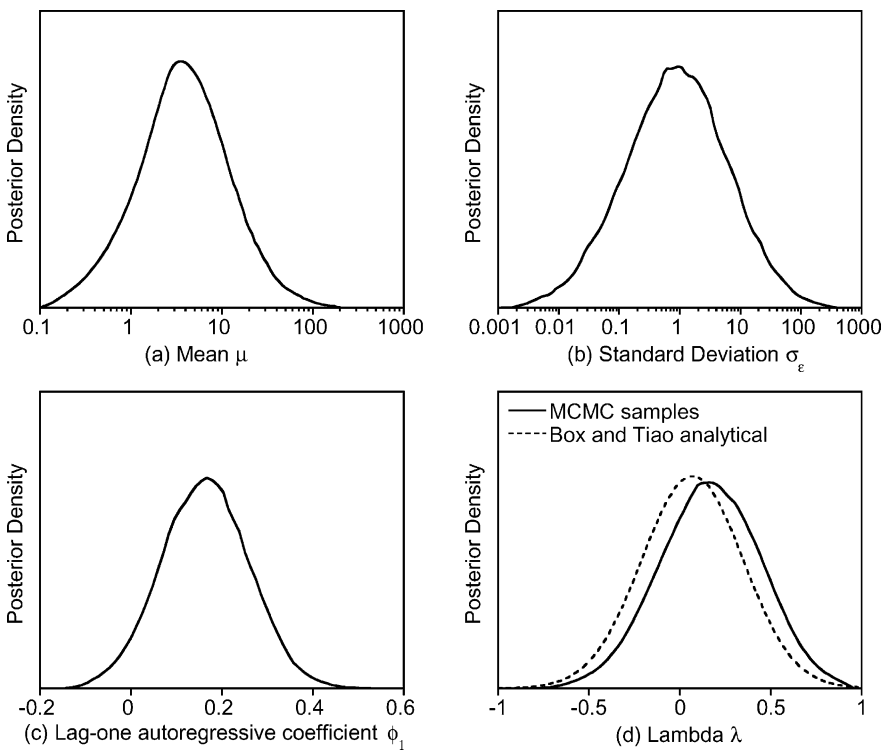


Fig. 4. Posterior distributions for the Sydney annual rainfall data (μ and σ_ε are shown a with log scale for ease of viewing).

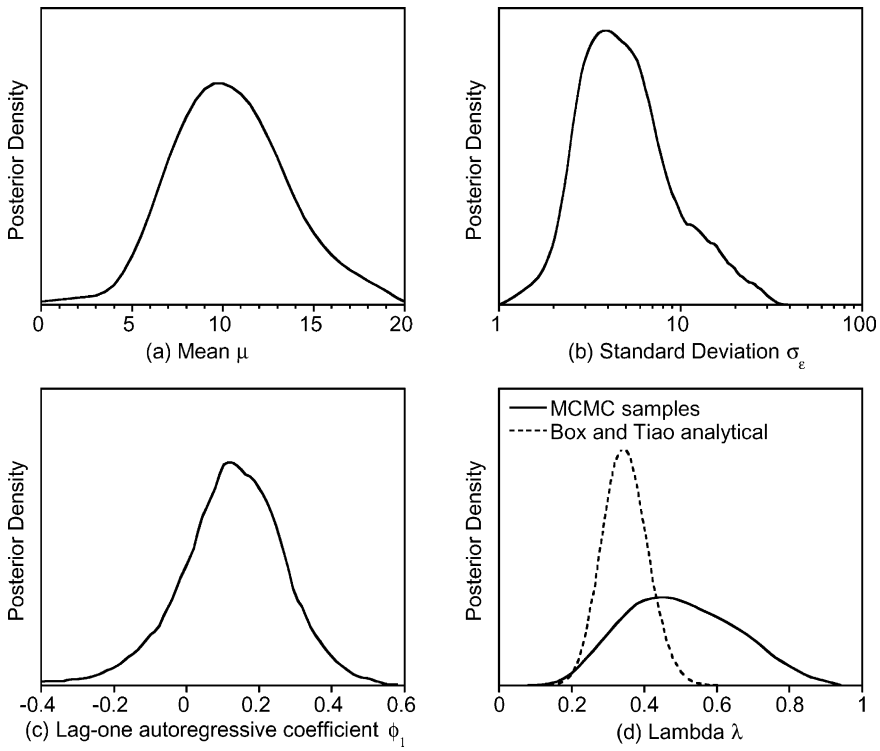


Fig. 5. Posterior distributions for the Burdekin River annual runoff data (σ_ε is shown with a log scale for ease of viewing).

there is some year-to-year persistence; the mode of the ϕ_1 posterior corresponds to a value of $\phi_1 = 0.12$. Furthermore, as the posterior probability that $\phi_1 < 0$, $P(\phi_1 < 0|Y_N) = 15\%$, there is a significant probability that the data are uncorrelated. Fig. 5(d) shows that there is a distinct difference between the posterior of λ obtained from the MCMC samples and Box and Tiao's analytical posterior. This indicates that ignoring the truncation has a significant effect on the λ posterior for the Burdekin River annual runoff. This result demonstrates the utility of the MCMC procedure outlined in this paper. The mode of the posterior corresponds to a λ value of approximately 0.45 and there is a very low posterior probability that $\lambda > 1$ ($< 0.1\%$) which indicates the use of the Box–Cox transformation is justified for the Burdekin River annual runoff data. Fig. 6(b) shows that the observed data is within the 90% confidence limits of the posterior predictive distribution and shows that the calibrated model is able to successfully reproduce the observed data, even when it is highly skewed.

5. Discussion and conclusions

A procedure was described for fully evaluating the uncertainty of the parameters in the AR(1) model when used with the Box–Cox transformation to normalise hydrological data. This procedure rigorously accounted for the complications introduced by the Box–Cox transformation. This appears to have been hitherto an unsolved problem.

The Metropolis algorithm, an MCMC method, was used to simulate the posterior distribution of the model parameters. This method only requires the ability to calculate the likelihood function for the AR(1) model. The derivation of the likelihood function for the AR(1) model with the Box–Cox transformation was given. Transformation of the parameter space was required to ensure successful implementation of the Metropolis algorithm. This transformation removed the strong nonlinear dependence of the AR(1) parameters on the Box–Cox transformation parameter λ which was found to severely degrade the ability of the Metropolis

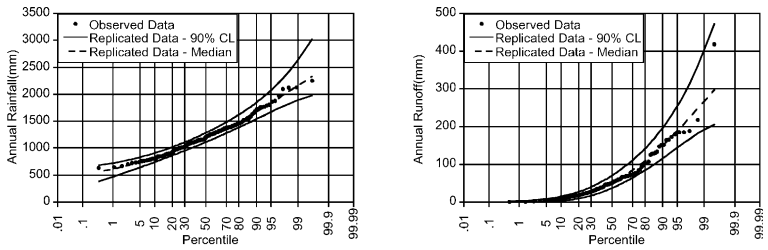


Fig. 6. Comparison of the observed data to the sampling distribution of the posterior predictive distribution for the hydrological data case studies. CL = confidence limits.

algorithm to explore the entire parameter space. The approach presented here can be applied to other stochastic models which make use of the Box–Cox transformation to render data approximately normally distributed. The extension of this methodology for cases where other transformations are more suitable (e.g. the two-step Box–Cox transformation (John and Draper, 1980)) is the subject of future research.

The technique was demonstrated by calibrating the AR(1) model to three synthetic case studies and two real hydrological data sets; the Sydney annual rainfall data and the Burdekin River annual runoff. The synthetic data studies verified that the MCMC procedure was robust across a range of values for the transformation parameter. For both the real data case studies the posterior of the lag-one autoregressive parameter indicated there was some year-to-year persistence, although the hypothesis that the data was uncorrelated could not be rejected outright. The posterior of the Box–Cox transformation parameter indicated that the application of the Box–Cox transformation was justified. Comparison to Box and Tiao's (1973) analytical expressions revealed that when the truncation caused by the Box–Cox transformation is significant their results can be misleading.

This study demonstrates the utility of Bayesian MCMC methods for evaluating parameter uncertainty in stochastic models. It is believed that the evaluation of parameter uncertainty has the potential to provide far greater hydrological insight than merely using single estimates of model parameter values. Despite this, the evaluation of parameter uncertainty is often sadly lacking in hydrological modelling endeavours. However, recently several studies have demonstrated the ability of MCMC methods to evaluate parameter uncertainty in conceptual hydrological models (Bates and Campbell, 2001; Kuczera and Parent, 1998). It is hoped that studies like these will motivate the increased use of techniques such as MCMC methods to investigate parameter uncertainty in hydrological modelling.

Acknowledgements

The authors would like to acknowledge Janice Lough from the Australian Institute of Marine Science who provided the Burdekin River runoff data and the helpful comments of the two anonymous reviewers which greatly improved this manuscript.

Appendix A. Likelihood function for AR(1) model

The likelihood function for the AR(1) model parameters for a time series of rainfall data $p(Y_N|\theta)$ will first be derived in terms of the transformed data $p(Z_N|\theta)$.

To derive the likelihood function for a single tranformed data point z_t it must be realised that there are complications caused by the Box–Cox transformation. If Eq. (2) is rearranged in terms of z_t, then

$$y_t = \begin{cases} (z_t\lambda + 1)^{1/\lambda} & \lambda \neq 0 \\ \exp(z_t) & \lambda = 0 \end{cases} \tag{A1}$$

This places a constraint on the transformed rainfall values, namely $z_t\lambda + 1 > 0$. Hence the z_t's actually

follow a truncated Gaussian distribution, such that

$$z_t|z_{t-1} \sim \mathrm{TN}\left(\tilde{z}_t, \sigma_\varepsilon^2\right) \tag{A2}$$

where $\mathrm{TN}(\mu, \sigma^2)$ denotes a truncated Gaussian distribution subject to some constraint, in this case $z_t\lambda + 1 > 0$, and $\tilde{z}_t = \mu + \phi_1(z_{t-1} - \mu)$. The corresponding probability density of a single observation z_t assumed to follow this truncated Gaussian distribution, is written as

$$p(z_t|z_{t-1}, \theta) \propto \begin{cases} \frac{1}{P_t^\lambda}\sigma_\varepsilon^{-1}\exp\left[-\frac{1}{2}\left(\frac{z_t - \tilde{z}_t}{\sigma_\varepsilon}\right)^2\right] & \text{if } z_t\lambda + 1 > 0 \\ 0 & \text{otherwise} \end{cases} \tag{A3}$$

where P_t^λ is a normalising probability which represents the region of the distribution that is not truncated, such that

$$P_t^\lambda = \begin{cases} \int_{-1/\lambda}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_\varepsilon}\exp\left[-\frac{1}{2}\left(\frac{x - \tilde{z}_t}{\sigma_\varepsilon}\right)^2\right]\mathrm{d}x & \text{if } \lambda > 0 \\ \int_{-\infty}^{-1/\lambda} \frac{1}{\sqrt{2\pi}\sigma_\varepsilon}\exp\left[-\frac{1}{2}\left(\frac{x - \tilde{z}_t}{\sigma_\varepsilon}\right)^2\right]\mathrm{d}x & \text{if } \lambda < 0 \end{cases} \tag{A4}$$

$$p(Y_N|\theta) \propto \begin{cases} \prod_{t=2}^{n}\left(\frac{y_t^{\lambda-1}}{P_t^\lambda\sigma_\varepsilon}\right)\exp\left[-\frac{1}{2}\sum_{t=2}^{n}\left(\frac{z_t - \tilde{z}_t}{\sigma_\varepsilon}\right)^2\right]p(y_1|\theta) & z_t\lambda + 1 > 0 \quad \forall t = 2, n \\ 0 & \text{otherwise} \end{cases} \tag{A8}$$

For a distribution to be a proper probability distribution the integral of its density must sum to 1. When a distribution is truncated its integral will not sum to 1. Therefore, P_t^λ is introduced to ensure the integral will sum to 1.

To determine the probability density of a single rainfall data point y_t the following change of variable transformation is applied to the density given in Eq. (A3)

$$p(y_t|y_{t-1}, \theta) = \left|\frac{\mathrm{d}z_t}{\mathrm{d}y_t}\right| p(z_t|z_{t-1}, \theta) = y_t^{\lambda-1}p(z_t|z_{t-1}, \theta)$$

$$\propto \begin{cases} (y_t)^{\lambda-1}\left(P_t^\lambda\sigma_\varepsilon\right)^{-1}\exp\left[-\frac{1}{2}\left(\frac{z_t - \tilde{z}_t}{\sigma_\varepsilon}\right)^2\right] & \text{if } z_t\lambda + 1 > 0 \\ 0 & \text{otherwise} \end{cases} \tag{A5}$$

Now to calculate the full likelihood function for a time series of rainfall data Y_N the following relationship is used (the notation $Y_t = \{y_1, \ldots, y_t\}$ as adopted by (Chib, 1996) is used)

$$p(Y_N|\theta) = p(y_n|Y_{n-1}, \theta)p(Y_{n-1}|\theta)$$

$$= p(y_n|y_{n-1}, \theta)p(Y_{n-1}|\theta) \tag{A6}$$

In the first line the conditional probability theorem is applied and in the second the assumed Markovian property of the data is used. By repeated application of Eq. (A6) we obtain the expression

$$p(Y_N|\theta) = p(y_n|y_{n-1}, \theta)\cdots p(y_t|y_{t-1}, \theta)\cdots p(y_2|y_1, \theta)p(y_1|\theta) \tag{A7}$$

The probability density of the typical term in this recursion $p(y_t|y_{t-1}, \theta)$ is given in Eq. (A5). When all the probability densities for each of the terms given in Eq. (A4) are multiplied together the following expression for the full likelihood results:

where P_t^λ and z_t are as given above. It is important to note that the normalising factor P_t^λ changes for each data point z_t because it is dependent on $\tilde{z}_t$. This expression does not include the likelihood density for the terminal point $p(y_1|\theta)$. In general terms, the likelihood for the terminal point can be calculated using

$$p(y_1|\theta) = \int p(y_1|y_0, \theta)p(y_0|\theta)\mathrm{d}y_0 \tag{A9}$$

However, this marginal density is not easily derived. Instead, the following expression is used:

$$p(y_1|\theta) = \left|\frac{\mathrm{d}z_1}{\mathrm{d}y_1}\right| p(z_1|\theta) = y_1^{\lambda-1}p(z_1|\theta) \tag{A10}$$

Ignoring the truncation, the marginal density for z_t, as

given by Box and Jenkins (1970), is

$$p(z_t|\theta) \sim N\left(\mu, \frac{\sigma_\varepsilon^2}{1-\phi^2}\right) \quad \text{(A11)}$$

Ignoring the truncation for only the terminal point is not expected to have a major impact on the inference.

Appendix B. First-order approximations of the AR(1) model parameters

Using first-order approximations it is possible to derive a relationship between the parameters of the AR(1) model, μ and σ_ε which are in transformed space to their first-order equivalents in untransformed space, m_y and s_y. This derivation begins by using the knowledge that $z_t = f(y_t)$, as defined in Eq. (2). Applying a Taylor series expansion to this function gives

$$z_t = f(m_y) + \left.\frac{\mathrm{d}z_t}{\mathrm{d}y_t}\right|_{y_t=m_y} (y_t - m_y) + \cdots \quad \text{(A12)}$$

Taking expectations and neglecting all the terms higher than the first-order gives

$$E[z_t] \approx E[f(m_y)] + \left.\frac{\mathrm{d}z_t}{\mathrm{d}y_t}\right|_{y_t=m_y} E[(y_t - m_y)] = E[f(m_y)] \quad \text{(A13)}$$

This implies

$$\mu = \begin{cases} \dfrac{m_y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log m_y & \lambda = 0 \end{cases} \quad \text{(A14)}$$

To derive a similar expression for σ_ε the Taylor series expansion is utilised again. Taking expectations and ignoring all the terms higher than the first order yields

$$E\left[(z_t - f(m_y))^2\right] \approx \left(\left.\frac{\mathrm{d}z_t}{\mathrm{d}y_t}\right|_{y_t=m_y}\right)^2 E\left[(y_t - m_y)^2\right]_t \quad \text{(A15)}$$

which gives

$$\sigma_z^2 = m_y^{2(\lambda-1)} s_y^2 \quad \text{(A16)}$$

Using the knowledge that $\sigma_z^2 = \sigma_\varepsilon^2/(1-\phi^2)$ (Box and Jenkins, 1970) then Eq. (A16) can be rearranged to give an expression in terms of σ_ε, where

$$\sigma_\varepsilon = m_y^{\lambda-1} s_y \sqrt{1-\phi^2} \quad \text{(A17)}$$

References

Bates, B.C., Campbell, E.P., 2001. A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. Water Resour. Res. 37 (4), 937–947.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. R Stat. Soc. Ser. B 26 (2), 211–252.

Box, G.E.P., Jenkins, G.M., 1970. Time Series Analysis: Forecasting and Control, Holden-Day, San Fransisco, pp. 553.

Box, G.E.P., Tiao, G.C., 1973. Bayesian Inference in Statistical Analysis, Addison-Wesley, Reading, MA, USA, pp. 588.

Chander, S., Kumar, A., Spolia, S.K., 1978. Flood frequency analysis by power transformation. J. Hyrdraul. Div., Proc. ASCE 104 (11), 1495–1504.

Chib, S., 1996. Calculating posterior distributions and modal estimates in Markov mixture models. J. Econom. 75, 79–97.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis–Hastings Algorithm. Am. Stat. 46 (4), 327–335.

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

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

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

Hirsch, R.M., 1979. Synthetic hydrology and water supply reliability. Water Resour. Res. 15 (6), 1603–1615.

Jain, D., Singh, V.P., 1986. A comparison of transformation methods for flood frequency analysis. Water Resour. Bull. 22 (6), 903–912.

John, J.A., Draper, N.R., 1980. An alternative family of transformations. J. Appl. Stat. 29, 190–197.

Kuczera, G., Parent, E., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models. J. Hydrol. 211, 69–85.

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

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

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

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