

BAYESIAN SELECTION OF THRESHOLD AUTOREGRESSIVE MODELS

BY EDWARD P. CAMPBELL

CSIRO Mathematical & Information Sciences

First Version received July 2000

Abstract. An approach to Bayesian model selection in self-exciting threshold autoregressive (SETAR) models is developed within a reversible jump Markov chain Monte Carlo (RJMCMC) framework. Our approach is examined via a simulation study and analysis of the Zurich monthly sunspots series. We find that the method converges rapidly to the optimal model, whilst efficiently exploring suboptimal models to quantify model uncertainty. A key finding is that the parsimony of the model selected is influenced by the specification of prior information, which can be examined and subjected to criticism. This is a strength of the Bayesian approach, allowing physical understanding to constrain the model selection algorithm.

Keywords. Model selection; reversible jump MCMC; SETAR model; sunspots.

1. INTRODUCTION

The modelling of physical systems is typically accomplished by building models incorporating the physics of relevant processes. For applications such as global climate models (e.g. Latif *et al.*, 1998), the level of complexity involved can be enormous because so many processes must be modelled to produce useful predictions. As pointed out by Tong *et al.* (1985), it may be worthwhile constructing relatively simple stochastic models that capture the essential features of the physics. These models will typically be easier to interpret, which can in turn provide valuable insights into the physical processes being modelled.

An important trend in statistics over the past 15 years or so has been a fusion of ideas from dynamical systems and nonlinear time series, expounded most clearly perhaps by Tong (1990). These tools provide a nonlinear framework for examining physical phenomena. We pay particular attention in this study to the self-exciting threshold autoregressive (SETAR) models developed by Tong and Lim (1980). However, the issue of model selection has not progressed so quickly, as noted by Wong and Li (1998). This is an important question in geophysical applications where some understanding of the important physical variables and their inter-relationships can provide important insights and new understanding. Existing methods for selecting nonlinear time series models are dominated by penalised likelihood methods, such as Akaike's information criterion (AIC). The AIC approach is described in some detail by Tong (1990, pp. 281–90).

Bayesian model selection is a rapidly expanding field of research, particularly with the development of reversible jump Markov chain Monte Carlo (RJMCMC) by Green (1995). This approach generalizes MCMC, which generates samples from posterior distributions of model parameters, to generate samples from a unified parameter space encompassing all models of interest. In this way a single simulation can be used to select the model having highest posterior support and obtain a sample from its posterior distribution.

The remainder of this paper is structured as follows. In Section 2 we introduce the class of SETAR models. In Section 3 we discuss the use of simulation methods in Bayesian model selection and in Section 4 we apply these ideas to SETAR models. In Section 5 we illustrate our results in a case study of the Zurich sunspots data. We finish with a discussion and some conclusions drawn from our work.

2. SELF-EXCITING THRESHOLD AUTOREGRESSIVE MODELS

Many physical systems exhibit complex behaviour that cannot be represented by classical autoregressive moving-average (ARMA) time-series models (Tong, 1990, pp. 7–12). One way to represent such systems is to model them as piecewise linear. A class of piecewise autoregressive models for a time series $\{Y_t\}$ is defined by breaking its state space R into a disjoint union of regimes $\{R_i\}$. The boundaries of the regimes are defined by an ordered set r_0, r_1, \dots, r_l where r_0 and r_l are taken to be $-\infty$ and $+\infty$ respectively. Thus $R = \bigcup_{i=1}^{l-1} R_i$, where $R_i = [r_{i-1}, r_i)$ and r_1, \dots, r_{n_r} ($n_r = l - 1$) are known as thresholds. We define a SETAR model as:

$$Y_t = \sum_{i=1}^{p_j} \phi_i^{(j)} Y_{t-i} + \varepsilon_t, \quad Y_{t-d} \in R_j, \quad j = 1, \dots, n_r + 1.$$

We depart from Tong's (1990, p. 101) general definition by explicitly allowing different orders in each regime and limiting the innovations to act at time t only. We also do not include an intercept term at this stage. The innovation variance in regime j is assumed to be σ_j^2 . The regime switch is triggered by the lag- d value of the series. In this study we assume $d = 1$ and examine the problem of obtaining posterior distributions for the model orders and parameters.

The model is appealing from a physical perspective as many physical systems are state dependent in the sense that the nature of their future evolution is dependent on their current state. A number of such examples are discussed by Tong (1990); Graham and Barnett (1987) found enhanced convection beyond a critical sea surface temperature of about 29°C. The notion of a physical switch causing a shift in regime is an appealing one of current interest in many areas, such as climatology (e.g. Berliner *et al.*, 2000) and hydrology (e.g. Lu and Berliner, 1999).

3. MODEL SELECTION BY SIMULATION

Modern computational approaches to Bayesian statistics use simulation to generate samples from the posterior distribution of interest. The most successful technique to date is MCMC (see Tierney, 1994 and Smith and Roberts, 1993, for a review of the theory and practice of MCMC). To use MCMC to select models from a collection of candidates $\Omega = \{\omega_i; i = 1, \dots, M\}$ the Markov random walk must be generalized to be over a combined model-parameter space $(\Omega, \Theta_1, \dots, \Theta_M)$ say. Here there are M possible models having parameters $\{\Theta_i; i = 1, \dots, M\}$.

Alternative approaches to this problem have been pursued by Carlin and Chib (1995) and Green (1995). The approach of Carlin and Chib (1995) essentially defines the unified space as the product of the parameter spaces, whereas Green's (1995) RJMCMC method defines the joint space as the union of the parameter spaces. As this latter approach seems more natural, we seek to develop this framework in a SETAR context.

Here we briefly review the fundamental concepts of the RJMCMC methodology. We may view RJMCMC as a generalization of the Metropolis-Hastings algorithm, for which the acceptance probability can be written as

$$\alpha(\Theta, \Theta') = \min \left(\frac{\text{likelihood}}{\text{ratio}} \times \frac{\text{prior}}{\text{ratio}} \times \frac{\text{proposal}}{\text{ratio}}, 1 \right),$$

where Θ is the current state of the chain and Θ' the proposed new state. For convenience, in this discussion we denote the posterior by $\pi(\cdot)$. A sufficient but not necessary condition for the convergence of an MCMC algorithm is reversibility or detailed balance defined by $\pi(\Theta')P(\Theta', \Theta) = \pi(\Theta)P(\Theta, \Theta')$, where $P(\Theta, \Theta')$ denotes the transition kernel from Θ to Θ' of the Markov chain defining the algorithm. Green's approach supplements the transition kernel by first selecting a move type (jump) between model subspaces defined by the elements of Ω . An independent proposal for the appropriate Θ' is then made. Thus the proposal ratio will be the product of at least two terms. The more general acceptance probability that satisfies the condition of reversibility is given by

$$\alpha(\Theta^{(1)}, \Theta^{(2)}) = \min \left(\frac{\text{likelihood}}{\text{ratio}} \times \frac{\text{prior}}{\text{ratio}} \times \frac{\text{move}}{\text{ratio}} \times \frac{\text{proposal}}{\text{ratio}} \times \text{Jacobian}, 1 \right). \tag{1}$$

This incorporates a Jacobian to account for any functional relationship between the parameters in making the proposal. In equation (1) $\Theta^{(1)}$ denotes the parameter vector before the jump and $\Theta^{(2)}$ the proposed parameter vector after the jump.

The key feature to ensuring reversibility is the structure of the proposal ratio. Moves that change the dimension of the parameter vector, and thus violate the assumptions of the Metropolis-Hastings algorithm, must be in detailed balance with the opposite move. Perhaps the most straightforward way to execute a

dimension-changing move is to augment $\Theta^{(1)}$ and $\Theta^{(2)}$ with random vectors $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(2)}$, drawn from densities $q_1(\cdot)$ and $q_2(\cdot)$ such that the resulting augmented vectors are of equal length. Thus $\Theta^{(1)} \rightarrow (\Theta^{(1)}, \mathbf{u}^{(1)})$ and $\Theta^{(2)} \rightarrow (\Theta^{(2)}, \mathbf{u}^{(2)})$.

Thus (1) becomes:

$$\alpha(\Theta^{(1)}, \Theta^{(2)}) = \min \left\{ \frac{P(\mathbf{Y}|\Theta^{(2)})}{P(\mathbf{Y}|\Theta^{(1)})} \frac{P(\Theta^{(2)})}{P(\Theta^{(1)})} \frac{j(2, \Theta^{(2)})}{j(1, \Theta^{(1)})} \frac{q_2(\mathbf{u}^{(2)})}{q_1(\mathbf{u}^{(1)})} \text{Jacobian}, 1 \right\}, \quad (2)$$

where $j(\cdot)$ denotes the move-type probability for a particular model and parameter combination.

Green developed the reversible jump theory by modelling $[\omega, \theta_\omega, \mathbf{Y}]$ using a natural hierarchical structure defined by

$$[\omega, \theta_\omega, \mathbf{Y}] = [\omega] [\theta_\omega | \omega] [\mathbf{Y} | \omega, \theta_\omega], \quad (3)$$

where we use $[\cdot]$ to denote 'distribution of' or 'density of'. Thus the specification of prior information also requires us to define a prior over Ω . In a time-series context, this will typically require a prior for the model's order, an approach also pursued by Heintel (1998), although in the context of conjugate inference. We refer the reader to Green (1995) for further technical details of RJMCMC.

A few applications on a variety of themes have appeared in the literature to date, including Dellaportas and Forster (1999), Denison *et al.* (1998), Heikkinen and Arjas (1999), Richardson and Green (1997) and Stephens and Fisch (1998). Relatively few applications to time-series models have been published (including De Jong, 1997; Chen, 1998; Lu and Berliner, 1999). Chen and Lee (1995) and Chen (1998) examine the use of Gibbs sampling to fit two-regime threshold models, the latter study incorporating exogenous predictors, but in both cases assuming the autoregressive orders to be known.

Barbieri and O'Hagan (1996) presented an application of RJMCMC to ARMA models, parameterizing the model using partial correlation coefficients. This allows stationarity of the model to be enforced in a straightforward manner. Troughton and Godsill (1997) developed a reversible jump sampler for autoregressive models using the full conditional density for the autoregressive coefficients directly, which is available explicitly. The trade-off is that they do not enforce stationarity. It could be argued that allowing the stationarity conditions to be relaxed allows a more thorough exploration of the structure of a particular time series to be examined. This idea has been explored by Naylor and Marriott (1996) in the context of autoregressive models.

Troughton and Godsill (1997) compare their complete update algorithm to only making proposals for new parameters. In some cases they found this latter procedure ('partial updates'), in which acceptance probabilities are much faster to compute, slow to converge due to a more highly correlated Markov chain. We examine an approach using partial updates for SETAR models. We also do not enforce stationarity, which is more complicated for SETAR models (Tong, 1990, pp. 139–86) than autoregressive models.

4. SELECTING SETAR MODELS

In terms of a reversible jump algorithm, we note that a method for selecting autoregressive orders combined with a move between regimes is sufficient to select general SETAR models. We therefore develop the algorithm by first considering autoregressive model selection within a regime and then incorporating moves between regimes. Note that we assume that the number of thresholds is first chosen by an appropriate diagnostic or by physical considerations, and is then assumed known. In the physical applications motivating this study we sought a small number of thresholds, so that the results are interpretable. If a large number of thresholds are indicated, then perhaps a more sophisticated approach is called for. We return to this topic in the discussion.

4.1. Order selection within a regime

In this section we develop a reversible jump MCMC approach to selecting the order of an autoregressive (AR) model defined by

$$(1 - \phi_1 B - \dots - \phi_p B^p) Y_t = \varepsilon_t,$$

for a time series $\{Y_t: t = 1, 2, \dots\}$ such that $E(Y_t) = 0$; $\varepsilon_t \sim N(0, \sigma^2)$ is an independent white-noise sequence and B denotes the backshift operator $BY_t = Y_{t-1}$. In practice, it is sensible to choose a maximum permissible order for the model, which we denote by p_{\max} . Thus the candidate models are indexed by the set $\{0, 1, \dots, p_{\max}\}$. Given that the current model subspace is defined by $p = k$, an important consideration in choosing suitable move types is to ensure hierarchical ordering of the model terms. One way to achieve this is to choose randomly from the move types ‘Birth’, ‘Remain’ and ‘Death’ in which $k \rightarrow k + 1$, $k \rightarrow k$ and $k \rightarrow k - 1$ respectively. We denote the ‘Birth’ and ‘Death’ move type probabilities by b_k and d_k respectively.

The ‘Remain’ move does not change the dimension of the current parameter subspace, and was included as an aid to efficient mixing. As this move can be accomplished using a conventional Metropolis-Hastings algorithm, we merely sketch the details. A single-site updating Metropolis-Hastings random walk algorithm was used. The time-series coefficients were updated using normal proposal distributions, whilst the error variance was updated using an inverse chi-squared distribution (Campbell *et al.*, 1999).

The ‘Birth’ and ‘Death’ moves do change the dimension of the current parameter subspace, and so must be in detailed balance with one another. Since there is only one pair of dimension changing moves it is sufficient to discuss the ‘Birth’ move in detail, the acceptance probability for the corresponding ‘Death’ move following by implication.

If $\phi^{(1)}$ and $\phi^{(2)}$ denote the time-series coefficients before and after the ‘Birth’ move respectively, then a straightforward way to accomplish the move is to

augment $\phi^{(1)}$ by a random observation u . Note that this is simpler than the general algorithm since we need only generate a univariate random number. We then set $\phi^{(2)} = \phi^{(2)}(\phi^{(1)}, u)'$, which will in general require the Jacobian term to be present in the acceptance probability (1). We adopt a simpler approach by generating a purely random proposal for $\phi_{k+1}|\phi^{(1)}$ and setting $\phi^{(2)} = (\phi^{(1)}, \phi_{k+1})'$, so there is no need for a Jacobian.

Note that as an aid to efficient mixing we adopt the same approach as Stephens and Fisch (1998) and update σ^2 after each update of time-series coefficients.

4.1.1. Choice of prior distributions

In keeping with the hierarchical framework (3), we specify priors for the model order and $[\phi_p, \sigma^2|p]$ as follows. We used an inverse chi-squared distribution for σ^2 , having mean 1 and variance 1. This choice should be a satisfactory representation of vague prior knowledge for data standardized to have variance 1. We chose normal priors for the time series coefficients, each having mean 0 and variance 1. All parameters were assumed to be *a priori* independent.

It remains to select $[p]$, the prior for the model order. In classical time-series analysis, AIC penalises overly complex models. In a Bayesian framework we seek to do this via our prior beliefs, which become an explicit modelling assumption. As noted above, we select a maximum order (p_{\max}) for the model so that the prior for p is truncated above. We have investigated both a uniform prior and a Poisson prior to examine the impact on model parsimony.

4.1.2. Move-type probabilities

We follow Green (1995) in using $[p]$ to construct the move-type probabilities, and set

$$\begin{aligned} b_k &= c \min\{1, [k+1]/[k]\} \\ d_{k+1} &= c \min\{1, [k]/[k+1]\} \end{aligned}$$

and we choose c such that $b_k + d_k \leq 0.9$; the 'Remain' probability is set to $1 - b_k - d_k$. In this formulation, the move ratio for a 'Birth' step is b_k/d_{k+1} , whilst for a 'Death' step the move ratio is d_k/b_{k-1} .

4.1.3. Choice of proposal distributions

To ensure that the proposal process is efficient, it is important that the 'Birth' proposal produces sensible values. We have used a normal proposal distribution, the variance of which can be tuned according to the generally accepted rules described by Weir (1997), i.e., we seek to set the proposal variance to be a little more than the marginal posterior variance whilst achieving an acceptance rate in the range of 30–70%. It is not entirely obvious that these rules apply directly to RJMCMC, but in our experience they provide a reasonable guide. We have calculated a sensible mean for the proposal as follows. In obvious notation, a conditional least-squares (Tong, 1990, pp. 296–302) estimate can be found by minimizing

$$\begin{aligned}
 S(\phi_{k+1}) &= \sum_{t=k+2}^n \left(Y_t - \sum_{i=1}^{k+1} \phi_i Y_{t-i} \right)^2 \\
 &= S(\phi_k) - 2\phi_{k+1} \sum_{t=k+2}^n Y_{t-k-1} \left(Y_t - \sum_{i=1}^k \phi_i Y_{t-i} \right) + \phi_{k+1}^2 \sum_{t=k+2}^n Y_{t-k-1}^2.
 \end{aligned}$$

If we substitute the current values for ϕ_k and solve for ϕ_{k+1} we obtain

$$\hat{\phi}_{k+1} = \frac{\sum_{t=k+2}^n Y_{t-k-1} Z_t}{\sum_{t=k+2}^n Y_{t-k-1}^2}, \tag{4}$$

where $Z_t = Y_t - \sum_{i=1}^k \phi_i Y_{t-i}$. We use $\hat{\phi}_{k+1}$ as the mean of a proposed birth move in the work reported below.

The proposal ratio for a birth step therefore becomes $1/q(\phi_{k+1}|\phi^{(1)})$, where $q(\cdot)$ denotes a normal distribution having mean given by (4) and an appropriately chosen variance. The proposal ratio for the corresponding death step is 1 because we always kill the highest order term when a death move is proposed.

4.2. Moves between regimes

Note that, in general, we allow separate innovation variances and orders for each regime. The parameter vector in this context may therefore be written as $\Theta = \{r_i: i = 1, \dots, n_r; \phi_{ij}, \sigma_i^2: i = 1, \dots, n_r + 1, j = 1, \dots, p_i\}$. The autoregressive algorithm is easily adapted to this situation by selecting a regime at random as part of the move-type selection. In obvious notation, our move-type algorithm for selecting SETAR models becomes:

- (1) Select a regime at random with probability $(n_r + 1)^{-1}$;
- (2) Calculate ‘Birth’ and ‘Death’ probabilities $b_k^{(r)}$ and $d_k^{(r)}$, and set the ‘Remain’ probability to $1 - b_k^{(r)} - d_k^{(r)}$ and select the move-type according to these probabilities.

The procedure for selecting AR(p) models then applies directly. A notable difference in this algorithm is the calculation of the ‘Remain’ probability, which now varies with regime. However, since we choose $b_k^{(r)} + d_k^{(r)} \leq 0.9$ this implies that the ‘Remain’ probability is at least 0.1, as before. Note that the move-type probabilities do not depend on the model parameters, so the move-type probabilities are not required in the acceptance probability for moves that do not bring about a change in dimension. A single-site updating Metropolis algorithm with normal proposal distributions was used for the threshold parameters $\{r_i: i = 1, \dots, n_r\}$. The algorithm was more stable and efficient using independence chains (Tierney, 1994) than a random walk, which was found to wander between thresholds.

4.3. *Simulation study*

We first examined the autoregressive algorithm. Our main finding is that there is some sensitivity to the specification of the prior for the model order. When using a uniform prior the algorithm tends to choose large models, depending on the size of p_{\max} , because there is no penalty for lack of parsimony. We found that provided p_{\max} was not chosen too large the algorithm worked well, but high-order models were selected otherwise. A Poisson distribution provides a more informative prior, and was found to work well. In particular, there was no dependence on the value of p_{\max} .

To examine the performance of the SETAR model selection algorithm we simulated a time series of length 1000 from the model

$$Y_t = \begin{cases} 0.5Y_{t-1} - 0.2Y_{t-2} + 0.4Y_{t-3} + \varepsilon_t^{(1)}, & Y_{t-1} \leq 0 \\ 0.5Y_{t-1} - 0.6Y_{t-2} + \varepsilon_t^{(2)}, & Y_{t-1} > 0, \end{cases}$$

where $\varepsilon_t^{(1)}$ and $\varepsilon_t^{(2)}$ are both distributed as $N(0,1)$. The independence chain mean used for the threshold parameter was the known value of 0. The prior for the model orders is a Poisson distribution having mean 3 and truncated above at $p_{\max} = 6$ in each case.

The model occupation probabilities after a burn-in of 10,000 iterations followed by 10,000 iterations that were collected for analysis are shown in Table I. We see that the algorithm has selected the correct model orders with a substantial degree of certainty, although there is some evidence that the first regime could be of order 4. Summary statistics for the corresponding posterior distribution are given in Table II. We see that the true parameter values lie within the 95% credibility interval

TABLE I
MODEL OCCUPATION PROBABILITIES FOR THE SIMULATED SETAR DATA SET

Regime 2 model order	Regime 1 model order			
	3	4	5	6
2	0.7535	0.2331	0.0132	0.0002

TABLE II
SUMMARY OF POSTERIOR DISTRIBUTION FOR THE OPTIMAL SETAR MODEL FOR THE SIMULATED DATA SET. ALL ESTIMATES ARE QUOTED CORRECT TO 2 sf* EXCEPT FOR VARIANCE PARAMETERS, WHICH ARE CORRECT TO 4 sf

Parameter	Posterior mean	95% credibility interval
r	-0.0096	-0.055 to 0.0068
ϕ_{11}	0.44	0.38 to 0.52
ϕ_{12}	-0.13	-0.21 to -0.048
ϕ_{13}	0.39	0.33 to 0.45
σ_1^2	0.9322	0.8266 to 1.058
ϕ_{21}	0.55	0.47 to 0.63
ϕ_{22}	-0.62	-0.69 to -0.54
σ_2^2	1.091	0.9693 to 1.246

*Significant figures.

in each case, although ϕ_{12} is somewhat marginal. These results are quite impressive and suggest that RJMCMC has much to offer in identifying nonlinear time-series models. To assess the sensitivity of these results to prior assumptions about model order in each regime, a sensitivity analysis was conducted. A number of combinations of prior means were drawn from the set $\{1, 2, 3, 5, 7\}$ for each regime, and $p_{\max} \in \{6, 7, 8, 9, 10\}$. In general, it appears to be better to allow the prior to have a generous mean, since a choice that is too small seems to inhibit the random walk. A too large choice will show up as an optimal model having poorly defined higher order terms. In this case, the prior mean can be amended appropriately. We may prefer to deal with this by assigning a hyper-prior to the prior mean. Note that as the algorithm was found to be most stable when the threshold parameter was updated last, this approach was adopted for the case study.

5. CASE STUDY

We now apply our SETAR model selection method to an analysis of the Zurich monthly mean sunspot numbers (1749–1984) described and listed up to 1983 by Brillinger (1985, pp. 67–74). These data are also available via the world-wide web at <http://lib.stat.cmu.edu/datasets/Andrews/>. The data were standardized prior to analysis to have mean 0 and variance 1, and are shown in Figure 1.

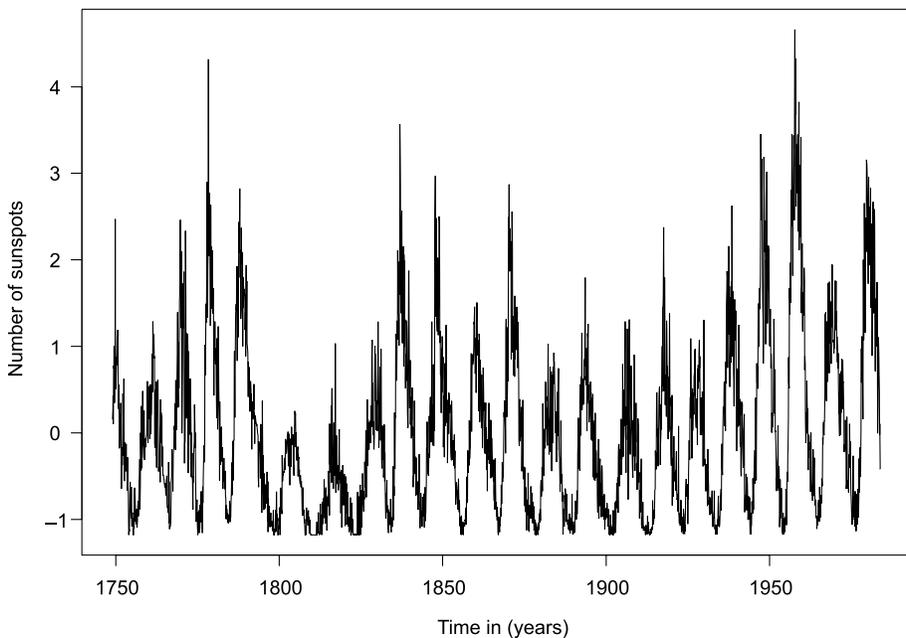


FIGURE 1. Zurich monthly mean sunspot numbers 1749–1984, standardized to have mean 0 and variance 1.

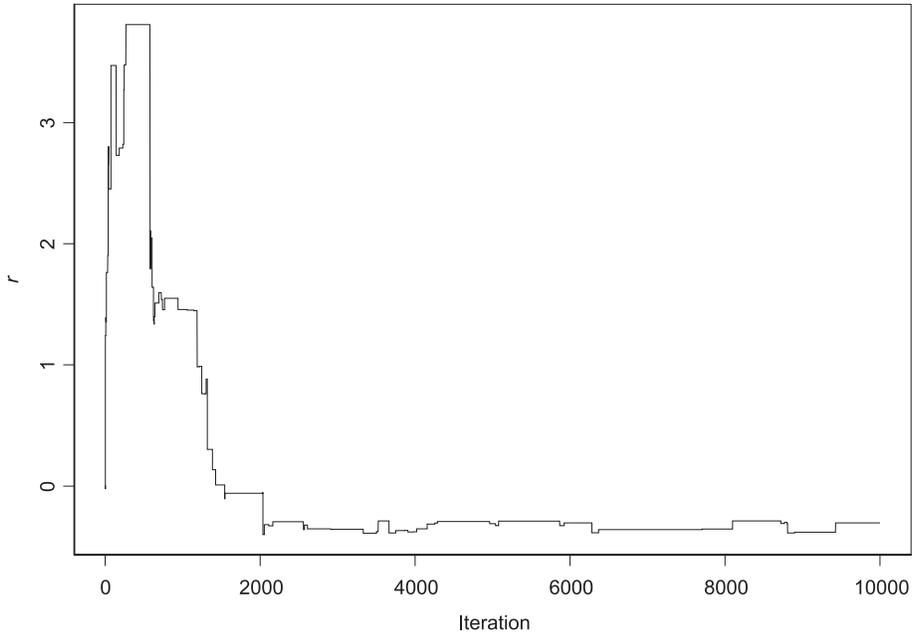


FIGURE 2. Results of applying an over-dispersed independence chain for the threshold parameter assuming a single threshold.

An important consideration in modelling using SETAR models is choosing the number of thresholds. Some preliminary runs of our algorithm assuming $n_r = 1$, but with a relatively large proposal variance, will provide an indication both of the appropriate number of thresholds and their approximate location. The results of a run of 10,000 iterations for the sunspot series are shown in Figure 2, where the independence chain mean and variance were 0 and 2 respectively. This threshold appears to be well defined, at about -0.4 , which is supported by the likelihood function shown in Figure 3. Note that this likelihood function is for the optimal model orders found in the tuning run, presented here for clarity in preference to the likelihood trace for the entire set of iterations. Further runs of the algorithm with a range of means and proposal variances provided further support for this approximate location, with no evidence of any further thresholds, such as the threshold parameter residing in more than one part of its state space for extensive periods. Note that the algorithm in some early runs placed the threshold below the range of the data. This corresponds essentially to a degenerate autoregressive model; in likelihood terms, the fit was relatively poor and so not of interest.

Note that the initial tuning run resulted in optimal regime model orders (11, 5). However, the coefficients for the six highest order parameters in regime 1 were not well defined, suggesting that the initial prior mean of 10 was too large. In subsequent runs, the prior means were set to 5 in each case.

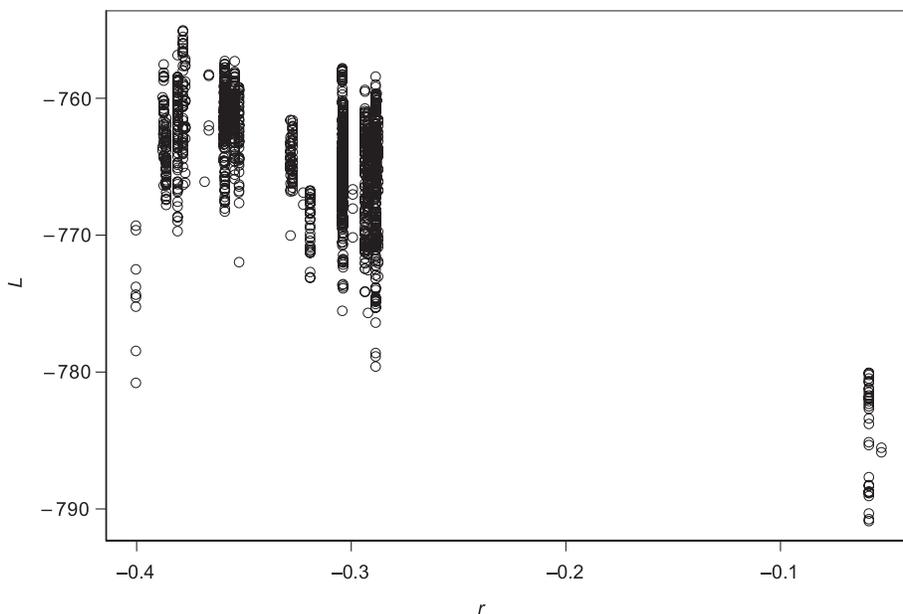


FIGURE 3. Log-likelihood trace for optimal model found in tuning run.

Proposal variances for each of the parameters were chosen as rounded-up values of the marginal variances found in the tuning runs, based on the suggested optimal model. The algorithm was then re-run with a burn-in of 10,000 iterations, followed by 10,000 iterations collected for analysis. The optimal model selected has regime orders (4, 5), with model occupation probabilities shown in Table III. For reference, AIC selects an autoregressive model of order 28, so the SETAR model uses 17 fewer parameters. The posterior distribution is summarized in Table IV. The threshold of -0.4 on the standardized scale corresponds to 33.9 on the original scale. Not surprisingly the error variance in the upper regime is considerably larger than the error variance for the lower regime. Each of the time-series coefficients is well defined.

The observed data are reproduced together with the one-step-ahead predictions in Figure 4, and the match appears to be excellent. This is largely confirmed by

TABLE III
MODEL OCCUPATION PROBABILITIES FOR SUNSPOTS CASE STUDY

Regime 1 model order	Regime 2 model order			
	4	5	6	7
4	0.1405	0.4900	0.3265	0.0031
5	0.0192	0.0138	0.0067	0.0000
6	0.0000	0.0001	0.0001	0.0000

TABLE IV

SUMMARY OF POSTERIOR DISTRIBUTION FOR THE OPTIMAL SETAR MODEL FOR THE ZURICH MEAN MONTHLY SUNSPOT NUMBERS DATA SET. ALL ESTIMATES ARE QUOTED CORRECT TO 2 sf EXCEPT FOR VARIANCE PARAMETERS, WHICH ARE CORRECT TO 4 sf

Parameter	Posterior mean	95% credibility interval
r	-0.40	-0.41 to -0.39
ϕ_{11}	0.53	0.47 to 0.58
ϕ_{12}	0.18	0.12 to 0.22
ϕ_{13}	0.17	0.13 to 0.23
ϕ_{14}	0.095	0.051 to 0.16
σ_1^2	0.04492	0.04145 to 0.04872
ϕ_{21}	0.59	0.54 to 0.67
ϕ_{22}	0.11	0.012 to 0.17
ϕ_{23}	0.089	0.032 to 0.17
ϕ_{24}	0.095	0.048 to 0.14
ϕ_{25}	0.071	0.015 to 0.11
σ_2^2	0.2022	0.1888 to 0.2180

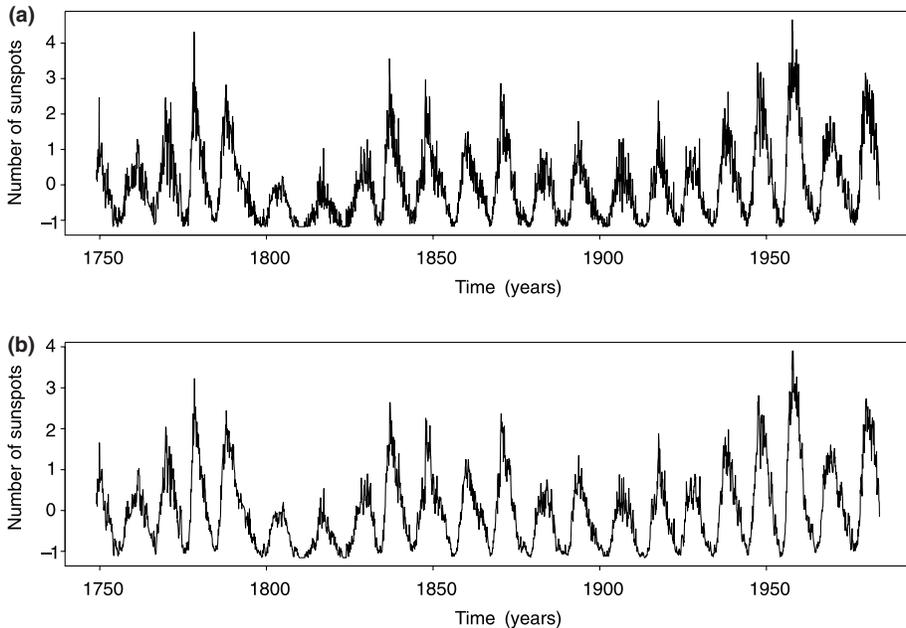


FIGURE 4. Observed values (a) and one-step-ahead predictions (b) for the Zurich monthly mean sunspot series.

Figure 5, which shows the one-step-ahead predictions plotted against the observed values. The correlation between observed values and one-step-ahead predictions is 0.93. Note that as these one-step-ahead predictions are used as a quick check on the fitted model, they simply use the posterior means of the parameters for the optimal model. There is some evidence that the variance

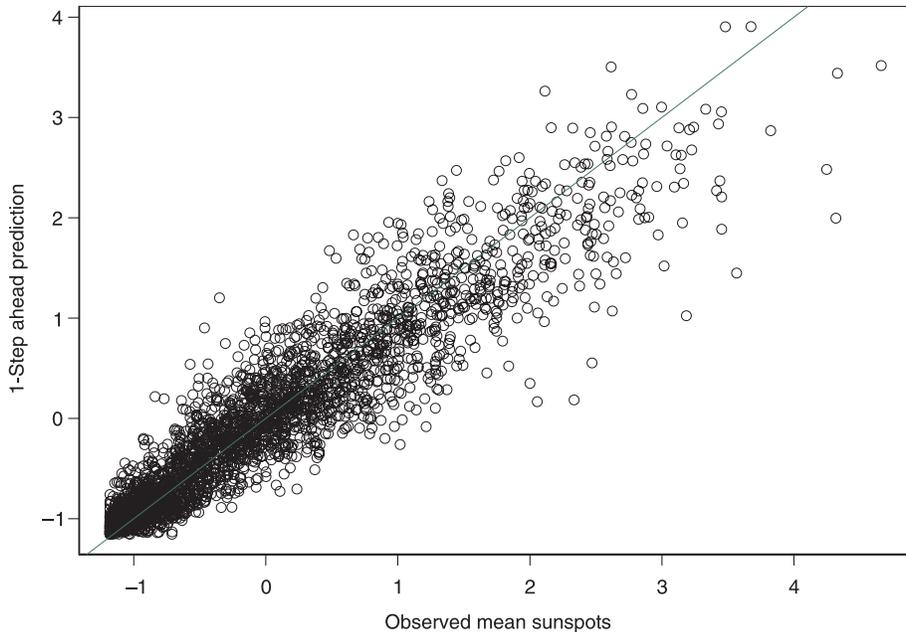


FIGURE 5. One-step-ahead predictions against observed mean sunspot numbers.

increases with the mean, which has been found by other researchers working on the Wolf sunspot series (e.g. Tong, 1990, pp. 419–29) who examined a square-root transformation). We do not explore this issue here.

6. DISCUSSION AND CONCLUSIONS

We have developed a Bayesian approach to the selection of nonlinear time series models using RJMCMC methodology. It has been shown that this approach can reliably identify known models, and has performed well with a real data set. The method is able to quickly identify models having strong posterior support, whilst also exploring less well supported models. The method is fast because it only updates new parameters, an approach termed ‘partial updates’ by Troughton and Godsill (1997). These authors raised the concern that such an approach may suffer from poor mixing as a result of high correlations between successive steps. This has not been found to be so in our applications to date, but could become an issue in more complicated models, such as in addition estimating the number of thresholds and the delay parameter.

An important feature of the Bayesian approach is that model parsimony is determined by explicit prior assumptions about model order. These assumptions can be checked and amended in the light of available data, a feature that is not

available using AIC. This is especially valuable when no strong prior knowledge is available, but does require a nonrigid approach to the specification of prior knowledge in relation to model order. It could be argued that neither the uniform nor Poisson priors are ideal, although the Poisson prior was clearly superior. One suggestion made in this study is to place a hyper-prior on the prior mean; given a suitable choice, such as a gamma distribution, this should prove to be an aid to rapid mixing by inducing an over-dispersed prior for model order.

The method at present assumes the number of thresholds to be known, although their location is not. The algorithm can be used with a large proposal variance to indicate the appropriate number of thresholds and their approximate location, and this approach worked well in our case study. However, the uncertainty in not knowing the number of thresholds is unaccounted for by this procedure. The approach developed in this study could readily be extended to estimate the number of thresholds by incorporating an additional move type. This move type would represent the birth of an additional threshold or the death of an existing threshold. These moves change the dimension of the parameter space, and so would have to be reversible moves in detailed balance. In this way a full accounting for uncertainty could be made. Note also that a move-type to increase or decrease the delay (assumed to be 1 in this paper) could also be incorporated in a very natural way. These suggestions are the topic of ongoing research, along with the incorporation of exogenous predictors. As noted above though, the use of partial updates may be less efficient in these contexts.

Rigorous diagnosis of MCMC convergence is elusive, and therefore especially so for RJMCMC. One approach is to use RJMCMC to identify an optimal model, which is then fitted using a conventional MCMC algorithm. The sample collected may then be analysed using existing convergence diagnostics. In our work convergence was essentially obvious, in that the algorithm settled on a small subset of optimal or near-optimal models relatively quickly. A reasonable burn-in therefore seemed sufficient to guarantee convergence.

A separate study of monthly rainfall data found that the one-step-ahead predictions did not reproduce the volatility of the observed series. The inclusion of intercept terms in each autoregressive regime was found to improve the fit in this regard, with no change required to the methodology.

As in the case study presented in this paper, there is typically support for a number of candidate models. This suggests that posterior summary statistics should in general be model-averaged to incorporate the uncertainty due to model selection. These are readily available by retaining samples from sub-optimal models, but model averaging is a big area in itself and not the purpose of this study. In this study we have presented model occupation probabilities as a guide to the performance of the algorithm.

The software and simulated data sets used in this study will be available from the author's web page at <http://www.cmis.csiro.au/Eddy.Campbell/>.

ACKNOWLEDGEMENTS

The Western Australian State Government provided funding for this research through the Indian Ocean Climate Initiative. The data for the case study were obtained from STATLIB. The author gratefully acknowledges Quanxi Shao for his suggestion to use conditional least squares in the birth proposal; Dave Denison for helpful discussions on reversible jump MCMC; Peter Green for general advice and pointers on reversible jump MCMC; Peter Toscas and Yun Li for their thoughtful reviews of an earlier manuscript. I am grateful for the comments of an anonymous reviewer, which led to a much-improved version. I thank Bryson Bates for his insights and so much more.

NOTES

Corresponding author: Edward P. Campbell, CSIRO Mathematical & Information Sciences, Private Bag No. 5, Wembley, Western Australia 6913, Australia. E-mail: eddy.campbell@csiro.au

REFERENCES

- BARBIERI, M. M. and O'HAGAN, A. (1996) *Probabilità e Statistiche Applicate*. Dipartimento di Statistica, Università "La Sapienza" and Department of Mathematics, University of Nottingham, Nottingham.
- BERLINER, L. M., WIKLE, C. K. and CRESSIE, N. (2000) Long-lead prediction of Pacific SST via Bayesian dynamic modeling. *J. Climatol.* 13, 3953–68.
- BRILLINGER, D. R. (1985) Monthly mean sunspot numbers. In *Data* (eds D. F. ANDREWS, M. A. HERZBERG). New York: Springer-Verlag.
- CAMPBELL, E. P., FOX, D. R. and BATES, B. C. (1999) A Bayesian approach to parameter estimation and pooling in nonlinear flood event models. *Water Resources Res.* 35, 211–20.
- CARLIN, B. P. and CHIB, S. (1995) Bayesian model choice via Markov chain Monte Carlo methods. *J. R. Statist. Soc. B* 57, 473–84.
- CHEN, C. W. S. (1998) A Bayesian analysis of generalized threshold autoregressive models. *Statist. Probab. Lett.* 40, 15–22.
- CHEN, C. W. S. and LEE, J. C. (1995) Bayesian inference of threshold autoregressive models. *J. Time Series Anal.* 16, 483–92.
- DE JONG, P. (1997) The scan sampler for time series models. *Biometrika* 84, 929–37.
- DELLAPORTAS, P. and FORSTER, J. J. (1999) Markov chain Monte Carlo model determination for hierarchical and graphical log-linear models. *Biometrika* 86, 615–33.
- DENISON, D. G. T., MALLICK, B. K. and SMITH, A. F. M. (1998) Automatic Bayesian curve fitting. *J. R. Statist. Soc. B* 60, 333–50.
- GRAHAM, N. E. and BARNETT, T. P. (1987) Sea surface temperature, surface wind divergence, and convection over tropical oceans. *Science* 238, 657–9.
- GREEN, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82, 711–32.
- HEIKKINEN, J. and ARJAS, E. (1999) Modeling a Poisson forest in variable elevations: a nonparametric Bayesian approach. *Biometrics* 55, 738–45.
- HEINTEL, M. (1998) A note on a Bayesian order determination procedure for vector autoregressive processes. *Statistical Papers* 39, 213–21.

- LATIF, M., ANDERSON, D., BARNETT, T., CANE, M., KLEEMAN, R., LEETMAA, A., O'BRIEN, J., ROSATI, A. and SCHNEIDER, E. (1998) A review of the predictability and prediction of ENSO. *J. Geophys. Res.* 103, 14375–93.
- LU, Z. Q. and BERLINER, L. M. (1999) Markov switching time series models with application to a daily runoff series. *Water Resources Res.* 35, 523–34.
- NAYLOR, J. C. and MARRIOTT, J. M. (1996) A Bayesian analysis of non-stationary AR series. In *Bayesian Statistics*, Vol. 5 (eds J. M. BERNARDO, J. O. BERGER, A. P. DAWID and A. F. M. SMITH). Oxford: Oxford University Press, pp. 705–12.
- RICHARDSON, S. and GREEN, P. J. (1997) On Bayesian analysis of mixtures with an unknown number of components. *J. R. Statist. Soc. B* 59, 731–92.
- SMITH, A. F. M. and ROBERTS, G. O. (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. R. Statist. Soc. B* 55, 3–23.
- STEPHENS, D. A. and FISCH, R. D. (1998) Bayesian analysis of quantitative trait locus data using reversible jump Markov chain Monte Carlo. *Biometrics* 54, 1334–47.
- TIERNEY, L. (1994) Markov chains for exploring posterior distributions. *Ann. Statist.* 22, 1701–62.
- TONG, H. (1990) *Non-linear Time Series. A Dynamical Systems Approach*. New York: Oxford University Press.
- TONG, H. and LIM, K. S. (1980) Threshold autoregression, limit cycles and cyclical data. *J. R. Statist. Soc. B* 42, 245–92.
- TONG, H., THANOON, B. and GUDMUNDSSON, G. (1985) Threshold time series modeling of two Icelandic riverflow systems. *Water Resources Bull.* 21, 651–61.
- TROUGHTON, P. J. and GODSILL, S. J. (1997). A reversible jump sampler for autoregressive time series, employing full conditionals to achieve efficient model space moves. Technical Report CUED/F-INFENG/TR.304, Department of Engineering, University of Cambridge.
- WEIR, I. S. (1997) Fully Bayesian reconstructions from single-photon emission computed tomography data. *J. Am. Statist. Assoc.* 92, 49–60.
- WONG, C. S. and LI, W. K. (1998) A note on the corrected Akaike information criterion for threshold autoregressive models. *J. Time Series Anal.* 19, 113–24.

Copyright of Journal of Time Series Analysis is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.