REVERSIBLE JUMP MARKOV CHAIN MONTE CARLO STRATEGIES FOR BAYESIAN MODEL SELECTION IN AUTOREGRESSIVE PROCESSES

By J. Vermaak*, C. Andrieu†, A. Doucet* and S. J. Godsill*

*University of Cambridge and †University of Bristol

First Version received February 2001

Abstract. This paper addresses the problem of Bayesian inference in autoregressive (AR) processes in the case where the correct model order is unknown. Original hierarchical prior models that allow the stationarity of the model to be enforced are proposed. Obtaining the quantities of interest, such as parameter estimates, predictions of future values of the time series, posterior model-order probabilities, etc., requires integration with respect to the full posterior distribution, an operation which is analytically intractable. Reversible jump Markov chain Monte Carlo (MCMC) algorithms are developed to perform the required integration implicitly by simulating from the posterior distribution. The methods developed are evaluated in simulation studies on a number of synthetic and real data sets.

Keywords. Autoregressive process; Bayesian estimation; Markov chain Monte Carlo; model selection.

1. introduction

As a result of its simplicity and ease of use, the autoregressive (AR) process remains a popular choice for modelling time series data (Box *et al.*, 1994; West and Harrison, 1997). However, when fitting an AR process to real data the correct model order is often unknown, and needs to be estimated from the data alongside the other unknown quantities. Classical methods to estimate the model order include Akaike's information criterion (AIC) (Akaike, 1974) and minimum description length (MDL) (Rissanen, 1978), which rely on information theoretic criteria, and Bayes' information criterion (BIC) (Schwarz, 1985), which is approximately an asymptotic Bayes factor. These methods often fail, however, when the data sets are small, and for AR processes AIC is not asymptotically consistent. They are also not easily modified to accommodate constraints, such as enforcing the stationarity of the model.

In this paper, the inference problem is posed within a Bayesian framework. Original hierarchical prior models are proposed that allow the stationarity of the model to be enforced. The prior models facilitate the design of efficient estimation strategies, and appear to be robust for a wide range of choices for their fixed hyperparameters. Similar hierarchical Bayesian models have been

developed in different contexts in (Richardson and Green, 1997; Robert *et al.*, 2000).

Obtaining the quantities of interest, such as parameter estimates, predictions of future values of the time series, posterior model-order probabilities, etc., requires integration with respect to the full posterior distribution, an operation which is analytically intractable. This intractable integration problem is commonly solved by applying asymptotic approximations (in the size of the data set). Quite often, however, the approximations obtained in this way may be poor, leading to incorrect results, especially if the data sets are small.

An attractive alternative to solve the integration problem is the use of Markov chain Monte Carlo (MCMC) techniques (Robert and Casella, 1999). Given the importance of AR modelling, several authors have already proposed Bayesian models and MCMC-based computations. Fixed-dimensional MCMC algorithms are proposed in (Barnett *et al.*, 1996; Smith *et al.*, 1998; Huerta and West, 1999a,b), whereas reversible jump MCMC algorithms can be found in (Barbieri and O'Hagon, 1996; Troughton and Godsill, 1997). In (Barnett *et al.*, 1996; Smith *et al.*, 1998), the AR coefficients are reparameterized in terms of the reflection coefficients, also known as the partial correlation (PARCOR) coefficients. Model-order selection is performed by associating a binary indicator variable with each coefficient, and using these to perform subset selection. In (Huerta and West, 1999a,b), a prior structure is defined directly on the roots of the AR characteristic polynomial. Model uncertainty is implicitly accounted for by allowing the roots to have zero moduli. The reversible jump MCMC strategy in (Barbieri and O'Hagon, 1996) utilizes the fact that under the parameterization in terms of the reflection coefficients the AR model can be treated as a nested model, and employs model moves that increase or decrease the model dimension by one, with the coefficients common to the current and proposed models remaining unchanged. In contrast with this, the strategy in (Troughton and Godsill, 1997; Godsill, 2001) allows model jumps of arbitrary magnitude and proposes the new coefficient values from the corresponding posterior conditional distribution. Stationarity is, however, not enforced.

In this paper, reversible jump MCMC (Green, 1995) algorithms are developed to solve the Bayesian inference problem. The algorithms make the best use of the statistical structure of the Bayesian models, and are robust and computationally efficient. They are shown to perform well on a variety of problems using synthetic and real data sets.

The remainder of the paper is organized as follows. In Section 2 the signal model is specified and a parameterization in terms of the reflection coefficients is proposed for the case where stationarity is to be enforced. Section 3 presents the Bayesian models and estimation objectives. Reversible jump MCMC algorithms to simulate from the resulting posterior distributions are developed in Section 4. Section 5 reports the results of simulation studies on synthetic and real data sets. Finally, in Section 6, some conclusions are reached.

2. STATISTICAL MODELS

Let $\mathbf{x}_{1:T} \triangleq (x_1, \ldots, x_T)$ denote an observed vector of T real samples. The elements of $\mathbf{x}_{1:T}$ are here assumed to be represented by one of the models $\mathcal{M}_k$, $k \in \mathcal{K} \triangleq \{0, \ldots, k_{\max}\}$, with $\mathcal{M}_k$ corresponding to a kth order AR process, i.e.

$$\mathcal{M}_0 : x_t = \sigma_0 e_t, \quad k = 0$$

$$\mathcal{M}_k : x_t = \sum_{i=1}^{k} a_{i,k} x_{t-i} + \sigma_k e_t, \quad k \in \{1, \ldots, k_{\max}\},$$

where $\mathbf{a}_k \triangleq (a_{1,k}, \ldots, a_{k,k})$ are the coefficients of the kth order process, σ_k^2 is the excitation variance, and $e_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. The AR equation above can be rewritten in vector–matrix form as

$$\sigma_k \mathbf{e}_{1:T} = \mathbf{x}_{1:T} - \mathbf{X}_k \mathbf{a}_k = \mathbf{A}_k \mathbf{x}_{-k+1:T},$$

where the matrices $\mathbf{X}_k \in \mathbb{R}^{T \times k}$ and $\mathbf{A}_k \in \mathbb{R}^{T \times (T+k)}$ are defined as

$$\mathbf{X}_k \triangleq \begin{bmatrix} x_0 & x_{-1} & \ldots & x_{-k+1} \\ x_1 & x_0 & \ldots & x_{-k+2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{T-1} & x_{T-2} & \ldots & x_{T-k} \end{bmatrix}$$

$$\mathbf{A}_k \triangleq \begin{bmatrix} -a_{k,k} & \ldots & -a_{1,k} & 1 & 0 & \ldots & 0 \\ 0 & -a_{k,k} & \ldots & -a_{1,k} & 1 & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & -a_{k,k} & \ldots & -a_{1,k} & 1 \end{bmatrix}.$$

In addition the constraint $k_{\max} < T$ is imposed here. If this is not the case, the columns of $\mathbf{X}_k$ are necessarily linearly dependent and in the case of a flat prior the model parameters are not uniquely determined by the data. The likelihood function for the AR model parameters follows from a straightforward random variable transformation from $\mathbf{e}_{1:T}$ to $\mathbf{x}_{1:T}$, and is given by

$$p(\mathbf{x}_{1:T}|k, \mathbf{a}_k, \sigma_k^2, \mathbf{x}_{0,k}) = \mathcal{N}(\mathbf{x}_{1:T}; \mathbf{X}_k \mathbf{a}_k, \sigma_k^2 \mathbf{I}_T), \tag{1}$$

where $\mathbf{x}_{0,k} \triangleq (x_0, \ldots, x_{-k+1})$ denotes the initial state of the AR process. In many practical applications, the initial state is known. Even in cases where the initial state is unknown, practical data sets are often large enough so that their initial samples can be taken as the initial state of the AR process. However, in a few important applications, the data sets are too small to sacrifice any of the available samples in this manner. For optimal performance in these cases the initial state should be considered as an unknown random quantity to be estimated alongside the other unknown parameters. This is the approach taken here.

2.1. *Stationarity issues*

Stationarity of the AR process is often easier to check and enforce by reparameterising the AR coefficients in terms of the reflection coefficients, denoted by $\boldsymbol{\rho}_k \overset{\triangle}{=} (\rho_1, \ldots, \rho_k)$. For the process to be stationary, each of the reflection coefficients has to lie in $(-1, 1)$. The AR and reflection coefficients are related through the well-known Levinson recursions (Friedlander, 1982), denoted by $\mathbf{a}_k = \mathcal{L}(\boldsymbol{\rho}_k)$. These recursions define a non-linear invertible mapping between the AR and reflection coefficients. This mapping can be expressed in vector–matrix form as (see Friedlander, 1982, for details)

$$\boldsymbol{\rho}_k = \mathbf{U}_k^{-1} \mathbf{a}_k, \tag{2}$$

with $\mathbf{U}_k$ an upper triangular matrix, defined as

$$\mathbf{U}_k \overset{\triangle}{=} \begin{bmatrix} 1 & a_{1,1} & \ldots & a_{k-1,k-1} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{k-1,1} \\ 0 & \ldots & 0 & 1 \end{bmatrix}.$$

The matrix $\mathbf{U}_k$ diagonalizes the data covariance matrix, i.e.

$$\boldsymbol{\Sigma}_k^{-1} = \mathbf{U}_k \boldsymbol{\Delta}_k^2 \mathbf{U}_k^{\mathrm{T}},$$

where $\boldsymbol{\Sigma}_k$ is the (symmetric) data covariance matrix for which the entry in the ith row and jth column is given by $\mathbb{E}[x_{t+i} x_{t+j}]$, and $\boldsymbol{\Delta}_k^2 \overset{\triangle}{=} \mathrm{diag}(\delta_{1,k}^2, \ldots, \delta_{k,k}^2)$ is a diagonal matrix with the ith diagonal entry corresponding to the inverse prediction error of the $(i-1)$th order predictor (see Friedlander, 1982 for details). An approximate linearization of the nonlinear mapping defined by (2) can be obtained by fixing the AR coefficients in the matrix $\mathbf{U}_k$ to correspond to a particular autocorrelation sequence, for example the sample autocorrelation sequence of the data itself. Cholesky factorization of the Toeplitz sample autocorrelation matrix, denoted by $\hat{\mathbf{R}}_{xx}$, yields

$$\hat{\mathbf{R}}_{xx}^{-1} = \tilde{\mathbf{U}}_k \tilde{\boldsymbol{\Delta}}_k^2 \tilde{\mathbf{U}}_k^{\mathrm{T}}.$$

In this case $\tilde{\mathbf{U}}_k$ is independent of the model parameters, so that the approximate mapping

$$\boldsymbol{\rho}_k \approx \tilde{\mathbf{U}}_k^{-1} \mathbf{a}_k \tag{3}$$

is linear and invertible. Extensive use is made of this approximation later when designing an MCMC algorithm for the case where stationarity is to be enforced. Note that a similar formulation is commonly used in linear prediction coding of speech signals (Friedlander, 1982).

3. BAYESIAN MODELS

Denote by $\boldsymbol{\theta}_k \in \boldsymbol{\Theta}_k$ the parameter vector associated with the model indexed by $k \in \mathcal{K}$, here defined as

$$\boldsymbol{\theta}_0 \triangleq (\sigma_0^2, \delta^2, \Lambda, \zeta^2) \quad \text{if } k = 0, \quad \text{and} \quad \boldsymbol{\theta}_k \triangleq (\boldsymbol{\phi}_k, \sigma_k^2, \delta^2, \Lambda, \zeta^2, \mathbf{x}_{0,k}) \quad \text{if } k \in \{1, \ldots, k_{\max}\},$$

where $\boldsymbol{\phi}_k = \boldsymbol{\rho}_k$ if stationarity is to be enforced and $\boldsymbol{\phi}_k = \mathbf{a}_k$ otherwise, and $(\delta^2, \Lambda, \zeta^2)$ are hyperparameters whose meaning will be defined later. In cases where the initial state is known, the quantities $(\zeta^2, \mathbf{x}_{0,k})$ disappear from the parameter vector. Thus, the Bayesian model is defined in the reflection coefficient domain if stationarity is to be enforced, and in the standard AR coefficient domain, if not. This is a choice which must be made from considerations of the data generating mechanism, for example physically generated data cannot be non-stationary if finite power considerations are incorporated. In many circumstances a stationary model is the most realistic choice. The overall parameter space can be written as a countable union of subspaces, i.e.

$$\boldsymbol{\Theta} = \bigcup_{k=0}^{k_{\max}} \{k\} \times \boldsymbol{\Theta}_k, \quad \text{where } \boldsymbol{\Theta}_0 \triangleq (\mathbb{R}^+)^4 \quad \text{if } k = 0,$$

and

$$\boldsymbol{\Theta}_k \triangleq \boldsymbol{\Phi}_k \times (\mathbb{R}^+)^4 \times \mathbb{R}^k, \quad \text{if } k \in \{1, \ldots, k_{\max}\},$$

with $\boldsymbol{\Phi}_k = (-1,1)^k$ for the stationary reflection coefficient model and $\boldsymbol{\Phi}_k = \mathbb{R}^k$ otherwise.

Within a Bayesian framework all the information of interest concerning the unknown parameters is contained in the posterior distribution $p(k, \boldsymbol{\theta}|\mathbf{x}_{1:T})$, which may be expressed as

$$p(k, \boldsymbol{\theta}|\mathbf{x}_{1:T}) = \sum_{i=0}^{k_{\max}} p(i, \boldsymbol{\theta}_i|\mathbf{x}_{1:T}) \mathbb{I}_{\{i\} \times \boldsymbol{\Theta}_i}(k, \boldsymbol{\theta}), \tag{4}$$

where $\mathbb{I}_S(x)$ is the indicator function that is unity if $x \in S$ and zero otherwise. Thus, only one term in (4) is ever non-zero, corresponding to the kth model subspace. The distributions on the right hand side of (4) follow from Bayes' rule as

$$p(k, \boldsymbol{\theta}_k|\mathbf{x}_{1:T}) \propto p(\mathbf{x}_{1:T}|k, \boldsymbol{\theta}_k) p(k, \boldsymbol{\theta}_k), \quad k \in \mathcal{K}.$$

In the above $p(\mathbf{x}_{1:T}|k, \boldsymbol{\theta}_k)$ is the likelihood function, here given by (1), and $p(k, \boldsymbol{\theta}_k)$ is the prior distribution of the unknown parameters of the kth model. A hierarchical prior structure of the form

$$p(k, \boldsymbol{\theta}_k) \triangleq p(\mathbf{x}_{0,k}|k, \sigma_k^2, \zeta^2) p(\boldsymbol{\phi}_k, \sigma_k^2|k, \delta^2) p(k|\Lambda) p(\delta^2) p(\Lambda) p(\zeta^2), \tag{5}$$

is adopted here, where

$$p(\mathbf{x}_{0,k}|k, \sigma_k^2, \zeta^2) \triangleq \mathcal{N}(\mathbf{x}_{0,k}; \mathbf{0}_{k \times 1}, \zeta^2 \sigma_k^2 \boldsymbol{\Sigma}_{\mathbf{x}_{0,k}})$$

$$p(\boldsymbol{\phi}_k, \sigma_k^2|k, \delta^2) \triangleq p(\boldsymbol{\phi}_k|k, \sigma_k^2, \delta^2) p(\sigma_k^2)$$

$$\triangleq c_k \mathcal{N}(\boldsymbol{\phi}_k; \mathbf{0}_{k \times 1}, \delta^2 \sigma_k^2 \boldsymbol{\Sigma}_{\boldsymbol{\phi}_k}) \mathbb{I}_{\boldsymbol{\Phi}_k}(\boldsymbol{\phi}_k) \mathcal{IG}(\sigma_k^2; \alpha_0, \beta_0)$$

$$p(k|\Lambda) \propto \frac{\Lambda^k}{k!} \mathbb{I}_{\mathcal{K}}(k), \quad p(\delta^2) \triangleq \mathcal{IG}(\delta^2; \alpha_{\delta^2}, \beta_{\delta^2})$$

$$p(\Lambda) \triangleq \mathcal{G}a(\Lambda; \alpha_\Lambda, \beta_\Lambda), \quad p(\zeta^2) \triangleq \mathcal{IG}(\zeta^2; \alpha_{\zeta^2}, \beta_{\zeta^2}).$$

In the above $\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, and the gamma and inverted-gamma distributions are defined as $\mathcal{G}a(x; a, b) \triangleq cx^{a-1} \exp(-bx)$ and $\mathcal{IG}(x; a, b) \triangleq cx^{-(a+1)} \exp(-b/x)$, respectively, where the normalizing constant is given by $c = b^a/\Gamma(a)$, with $\Gamma(\cdot)$ denoting the gamma function.

The covariance of the prior on the AR parameters $\boldsymbol{\phi}_k$ is set to $\boldsymbol{\Sigma}_{\boldsymbol{\phi}_k} = \mathbf{I}_k$, resulting in an isotropic prior for these parameters. The degree of uncertainty is controlled by the hyperparameter δ^2, which is assumed to follow an inverted-gamma prior. If stationarity is ignored, the normalizing constant c_k becomes unity. However, if stationarity is to be enforced, this normalizing constant follows straightforwardly as

$$c_k^{-1} = \prod_{i=1}^k \text{erf}\left(\frac{1}{\sqrt{2\delta^2 \sigma_k^2}}\right),$$

since the covariance matrix of the prior distribution is diagonal. In the above, the Gaussian error function is defined as

$$\text{erf}(x) \triangleq 2\pi^{-1/2} \int_0^x \exp(-s^2) \mathrm{d}s.$$

When stationarity is assumed, the prior distribution on the initial state is normally taken to be the initial state likelihood (see e.g. Box *et al.*, 1994). To obtain this distribution, the hyperparameter ζ^2 is set to one, and the matrix $\boldsymbol{\Sigma}_{\mathbf{x}_{0,k}}$ is taken to be the covariance matrix of k samples drawn from an AR process with coefficients $\mathbf{a}_k$ and unity excitation variance. The elements of $\boldsymbol{\Sigma}_{\mathbf{x}_{0,k}}$ are then complex nonlinear functions of the AR coefficients. Realising that the initial state can be quite arbitrary in practice, i.e. it cannot be assumed that the data is drawn from a steady-state regime, no such structure is imposed here. Instead, $\boldsymbol{\Sigma}_{\mathbf{x}_{0,k}}$ is taken to be the identity matrix, i.e. $\boldsymbol{\Sigma}_{\mathbf{x}_{0,k}} = \mathbf{I}_k$, and the hyperparameter controlling the variance of the prior on the initial state ζ^2 is considered an inverted-gamma distributed random variable to be estimated alongside the other parameters.

REMARK 1.   The structure of the initial state covariance matrix can be retained if the exact likelihood is required. However, in this case, marginalization of the AR coefficients is not possible, so that a conditional sampling strategy is required.

Alternatively, the Markov transition kernel implied by the sampling strategy developed in Section 4 can be used as a proposal distribution for a Metropolis–Hastings (MH) sampler to simulate from the posterior formed by combining the prior with the exact likelihood.

The prior on the model-order is a truncated Poisson distribution whose hyperparameter Λ can be interpreted as the mean expected model order. This hyperparameter is assumed to follow a gamma prior, thus facilitating a wide range of prior beliefs about the model-order. Other model-order priors, such as those associated with AIC and BIC (Poskitt and Tremayne, 1983), can be straightforwardly incorporated in the framework presented here.

The structure and form of the prior distribution in (5) leads to conjugacy with the likelihood for several parameters. This allows these variables to be marginalized, leading to efficient sampling strategies. Similar hierarchical prior models were developed before in (Richardson and Green, 1997; Robert, 2000). In the prior specification, the quantities $(\alpha_0, \beta_0, \alpha_{\delta^2}, \beta_{\delta^2}, \alpha_\Lambda, \beta_\Lambda, \alpha_{\zeta^2}, \beta_{\zeta^2})$ are assumed to be fixed and known, and can be set to reflect vague or informative prior information concerning the corresponding unknown parameters. The effects of different settings for these parameters are investigated in Section 5.1.

Prior distributions reflecting different beliefs, e.g. nearly non-stationary models, can also be accommodated within the MCMC framework. However, not all such priors facilitate analytical marginalization, leading to less efficient conditional samplers (see e.g. Troughton and Godsill, 1997). The hierarchical prior adopted here provides a trade-off between modelling accuracy and estimation efficiency. It is flexible enough to capture a wide range of prior beliefs, while still facilitating efficient sampling strategies (see e.g. Section 5.3 for the estimation of a nearly non-stationary model).

REMARK 2. In cases where the prior beliefs cannot be compromised, the Markov transition kernels implied by the sampling strategies developed in Section 4 can still be used, but now in the capacity of an efficient proposal distribution in a MH sampler for the posterior resulting from the true prior. A similar strategy is employed in Section 4.3 where the MCMC kernel to sample from the approximate posterior distribution based on the linear approximation for the reflection coefficients in (3) is used as a proposal for the intractable true posterior distribution in the case where stationarity is enforced. Another alternative is to post-process the samples from the Markov chain using importance sampling to obtain estimates under different prior distributions.

Within a Bayesian framework, all inference is with respect to the posterior distribution in (4). Typical quantities of interest are, for example,

$$\hat{k}^{\text{MMAP}} \stackrel{\triangle}{=} \arg\max_{k \in \mathcal{K}} \{ p(k|\mathbf{x}_{1:T}) \} \tag{6}$$

for model selection purposes,

$$\hat{\boldsymbol{\theta}}_{k^*}^{\text{MMSE}} \triangleq \mathbb{E}_{p(\boldsymbol{\theta}_k | k = k^*, \mathbf{x}_{1:T})}[\boldsymbol{\theta}_k]$$

for minimum mean square error (MMSE) parameter estimation of the model indexed by $k = k^*$, and

$$\hat{\mathbf{x}}_{T+1:T+K}^{\text{MMSE}} \triangleq \mathbb{E}_{p(\mathbf{x}_{T+1:T+K} | \mathbf{x}_{1:T})}[\mathbf{x}_{T+1:T+K}]$$

for the MMSE prediction of the K future values of the time series, based on the observed values $\mathbf{x}_{1:T}$. For the models under investigation here, the integrations to compute the above quantities are analytically intractable. Section 4 develops reversible jump MCMC algorithms to solve this problem.

4. MCMC COMPUTATION FOR BAYESIAN INFERENCE

If it is possible to simulate from the full posterior distribution in (4) the required estimates can be obtained by Monte Carlo integration. For the model presented here direct sampling from the posterior distribution is impossible. This section develops reversible jump MCMC (Green, 1995) algorithms to sample from this distribution.

The reversible jump MCMC algorithm is a generalization of the MH sampler for distributions defined over a union of spaces of (possibly) different dimensions, such as those encountered in model selection problems. It achieves model space moves by employing a proposal distribution and acceptance probability designed to preserve detailed balance, and hence ensure convergence to the correct invariant distribution. For a move from model k with parameters $\boldsymbol{\theta}_k$ to model k' with parameters $\boldsymbol{\theta}'_{k'}$, proposed from $q(k', \boldsymbol{\theta}'_{k'} | k, \boldsymbol{\theta}_k)$, the acceptance probability to preserve detailed balance is given by

$$\alpha(k', \boldsymbol{\theta}'_{k'} | k, \boldsymbol{\theta}_k) \triangleq \min\left\{1, \frac{p(k', \boldsymbol{\theta}'_{k'} | \mathbf{x}_{1:T}) q(k, \boldsymbol{\theta}_k | k', \boldsymbol{\theta}'_{k'})}{p(k, \boldsymbol{\theta}_k | \mathbf{x}_{1:T}) q(k', \boldsymbol{\theta}'_{k'} | k, \boldsymbol{\theta}_k)}\right\}.$$

Note that this form for the acceptance probability is slightly different from that in (Green, 1995) in that it explicitly includes the probability of proposing the move from k to k', and the proposal for the new parameters is made directly in the associated model space.

4.1. An MCMC algorithm

A general MCMC algorithm employing Gibbs-like steps to sample from the posterior distribution $p(k, \boldsymbol{\theta} | \mathbf{x}_{1:T})$ proceeds as follows.

ALGORITHM 1: MCMC ALGORITHM TO SIMULATE FROM $p(k, \boldsymbol{\theta}|\mathbf{x}_{1:T})$

1. *Initialize* $(k^{(0)}, \boldsymbol{\phi}_{k^{(0)}}^{(0)}, \sigma_{k^{(0)}}^{2(0)}, \delta^{2(0)}, \Lambda^{(0)}, \zeta^{2(0)}, \mathbf{x}_{0,k^{(0)}}^{(0)}) \in \boldsymbol{\Theta}$ *randomly or deterministically.*
2. *Iteration* $i \geq 1$:
   (*a*) *Simulate the AR model parameters:*

   • $(k^{(i)}, \boldsymbol{\phi}_{k^{(i)}}^{(i)}, \sigma_{k^{(i)}}^{2(i)}, \mathbf{x}_{0,k^{(i)}}^{(i)}) \sim p(k, \boldsymbol{\phi}_k, \sigma_k^2, \mathbf{x}_{0,k}|\delta^{2(i-1)}, \Lambda^{(i-1)}, \zeta^{2(i-1)}, \mathbf{x}_{1:T})$

   (*b*) *Simulate the hyperparameters:*

   • $\delta^{2(i)} \sim p(\delta^2|k^{(i)}, \boldsymbol{\phi}_{k^{(i)}}^{(i)}, \sigma_{k^{(i)}}^{2(i)})$

   • $\Lambda^{(i)} \sim p(\Lambda|k^{(i)})$

   • $\zeta^{2(i)} \sim p(\zeta^2|k^{(i)}, \sigma_{k^{(i)}}^{2(i)}, \mathbf{x}_{0,k^{(i)}}^{(i)})$

In what follows, the superscript (*i*) is dropped from all variables when there is no danger of any ambiguity. Simulating from $p(k, \boldsymbol{\phi}_k, \sigma_k^2, \mathbf{x}_{0,k}|\delta^2, \Lambda, \zeta^2, \mathbf{x}_{1:T})$ proceeds differently for the two cases where stationarity is ignored or explicitly enforced. The next section considers the case where stationarity is not enforced. Following this, the case where stationarity is explicitly enforced is considered. Finally, sampling strategies are proposed for the hyperparameters, and some convergence results are given.

4.2. *Stationarity not enforced*

Reversible jump MCMC algorithms for the case where stationarity is not enforced have been previously developed (Troughton and Godsill, 1997; Godsill, 2001). The algorithm described here adopts a more sophisticated Bayesian model which allows for exact marginalization of both the AR coefficients and excitation variance. The reversible jump strategy then operates in the marginal space, allowing a more efficient exploration of the state space than the earlier methods. Moreover, Godsill (2001) and Troughton and Godsill (1997) do not consider the case where the initial state is unknown and random.

Recall that for the case where stationarity is not enforced, $\boldsymbol{\phi}_k = \mathbf{a}_k$ and $\boldsymbol{\Phi}_k = \mathbb{R}^k$. The posterior conditional for the AR parameters can be factorized as

$$
\begin{aligned}
p(k, \mathbf{a}_k, \sigma_k^2, \mathbf{x}_{0,k}|\delta^2, \Lambda, \zeta^2, \mathbf{x}_{1:T}) = {} & p(\mathbf{a}_k|k, \sigma_k^2, \delta^2, \mathbf{x}_{0,k}, \mathbf{x}_{1:T}) \\
& \times p(\sigma_k^2|k, \delta^2, \zeta^2, \mathbf{x}_{0,k}, \mathbf{x}_{1:T}) p(k, \mathbf{x}_{0,k}|\delta^2, \Lambda, \zeta^2, \mathbf{x}_{1:T}),
\end{aligned}
\tag{7}
$$

where the distributions on the right-hand side are given by

$$
p(\mathbf{a}_k|k, \sigma_k^2, \delta^2, \mathbf{x}_{0,k}, \mathbf{x}_{1:T}) = \mathcal{N}(\mathbf{a}_k; \hat{\mathbf{a}}_k, \sigma_k^2 \mathbf{M}_k) \tag{8}
$$

$$
p(\sigma_k^2|k, \delta^2, \zeta^2, \mathbf{x}_{0,k}, \mathbf{x}_{1:T}) = \mathcal{IG}(\sigma_k^2; \alpha_k, \beta_k) \tag{9}
$$

$$p\left(k, \mathbf{x}_{0,k} | \delta^2, \Lambda, \zeta^2, \mathbf{x}_{1:T}\right) \propto \frac{\Gamma(\alpha_k) \Lambda^k \beta_k^{-\alpha_k} \mathbb{I}_{\mathcal{K}}(k)}{\left(2\pi\delta^2\zeta^2\right)^{k/2} k!} \left(\frac{|\mathbf{M}_k|}{|\mathbf{\Sigma}_{\mathbf{a}_k}||\mathbf{\Sigma}_{\mathbf{x}_{0,k}}|}\right)^{1/2}, \qquad (10)$$

with

$$\mathbf{M}_k \triangleq \left(\mathbf{X}_k^{\mathrm{T}}\mathbf{X}_k + \frac{1}{\delta^2}\mathbf{\Sigma}_{\mathbf{a}_k}^{-1}\right)^{-1}, \quad \hat{\mathbf{a}}_k \triangleq \mathbf{M}_k\mathbf{X}_k^{\mathrm{T}}\mathbf{x}_{1:T}$$

$$\mathbf{P}_k \triangleq \mathbf{I}_T - \mathbf{X}_k\mathbf{M}_k\mathbf{X}_k^{\mathrm{T}}, \quad \alpha_k \triangleq \alpha_0 + \frac{1}{2}(T + k)$$

$$\beta_k \triangleq \beta_0 + \frac{1}{2}\left(\mathbf{x}_{1:T}^{\mathrm{T}}\mathbf{P}_k\mathbf{x}_{1:T} + \frac{1}{\zeta^2}\mathbf{x}_{0,k}^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{x}_{0,k}}^{-1}\mathbf{x}_{0,k}\right).$$

Simulating from the posterior distribution in (7) requires simulation from each of the distributions on the right-hand side. This is straightforward for the posterior conditionals of the AR coefficients and excitation variance. Because of the difficulty associated with constructing efficient proposal distributions for the initial state for arbitrary model moves, the joint posterior conditional for the model order and initial state is here sampled from using a reversible jump MCMC step employing only three model moves, viz.

1 A birth move, selected with probability b_k, in which the dimension is increased from k to $k + 1$.
2 A death move, selected with probability d_k, in which the dimension is decreased from k to $k - 1$.
3 An update move, selected with probability u_k, in which the dimension remains fixed to k.

Thus, the reversible jump MCMC step is a mixture of the moves described above, i.e. at each iteration one of the candidate moves, viz. birth, death and update, is randomly attempted with corresponding probabilities b_k, d_k and u_k, such that $b_k + d_k + u_k = 1$, for all $k \in \mathcal{K}$. For $k = 0$ the death move is impossible, so that $d_0 = 0$, whereas for $k = k_{\max}$ the birth move is impossible, so that $b_{k_{\max}} = 0$. For all other values of k the birth and death probabilities are taken to be

$$b_k \triangleq c \min\left\{1, \frac{p(k+1|\Lambda)}{p(k|\Lambda)}\right\} \quad d_{k+1} \triangleq c \min\left\{1, \frac{p(k|\Lambda)}{p(k+1|\Lambda)}\right\},$$

where c is a constant which tunes the proportion of dimension change moves vs. update moves. As pointed out by Green (1995), this choice ensures that

$$\frac{b_k p(k|\Lambda)}{d_{k+1} p(k+1|\Lambda)} = 1,$$

which means that an MH algorithm on the dimension alone would have a unity acceptance probability. The outcome of a move is subjected to the appropriate MH acceptance probability. Each of the moves is now considered in more detail.

4.2.1. *Update move*

For the update move, the model-order remains fixed, so that the posterior conditional in (10) simplifies to

$$p\big(\mathbf{x}_{0,k}|k, \delta^2, \Lambda, \zeta^2, \mathbf{x}_{1:T}\big) \propto \beta_k^{-\alpha_k}\left(\frac{|\mathbf{M}_k|}{|\mathbf{\Sigma}_{\mathbf{a}_k}||\mathbf{\Sigma}_{\mathbf{x}_{0,k}}|}\right)^{1/2}. \tag{11}$$

Assuming the model-order to be fixed to k, a valid strategy to simulate from this distribution at iteration i is summarized below.

ALGORITHM 2: UPDATE MOVE

1. *Set* $\bar{\mathbf{x}}_{0,k}^{(0)} = \mathbf{x}_{0,k}^{(i-1)}$.
2. *For* $j = 1,\dots,k$:
   If $(u \sim \mathcal{U}_{[0,1]}) \le \lambda_u$:
   * *Simulate* $\bar{\mathbf{x}}_{0,k}^{(j)}$ *in a MH step with* (11) *as invariant distribution and* $q_{1,j}(\mathbf{x}_{0,k}|\bar{\mathbf{x}}_{0,k}^{(j-1)})$ *as proposal distribution.*

   Else:
   * *Simulate* $\bar{\mathbf{x}}_{0,k}^{(j)}$ *in a MH step with* (11) *as invariant distribution and* $q_{2,j}(\mathbf{x}_{0,k}|\bar{\mathbf{x}}_{0,k}^{(j-1)})$ *as proposal distribution.*
3. *Set* $\mathbf{x}_{0,k}^{(i)} = \bar{\mathbf{x}}_{0,k}^{(k)}$.

The proposal distributions in the above algorithm are defined as

$$q_{1,j}\big(\mathbf{x}'_{0,k}|\mathbf{x}_{0,k}\big) \triangleq \delta_{\mathbf{x}_{j,0,k}}(d\mathbf{x}'_{j,0,k})q_1\big(x'_{-j+1}|\mathbf{x}'_{j,0,k}\big) \tag{12}$$

$$q_{2,j}\big(\mathbf{x}'_{0,k}|\mathbf{x}_{0,k}\big) \triangleq \delta_{\mathbf{x}_{j,0,k}}(d\mathbf{x}'_{j,0,k})q_2\big(x'_{-j+1}|x_{-j+1}\big), \tag{13}$$

where $\mathbf{x}_{j,0,k}$ is $\mathbf{x}_{0,k}$ with x_{-j+1} removed. Thus, each element of the initial state is sampled individually using a mixture of two MH steps with proposal distributions as defined above, and mixture weights λ_u and $1 - \lambda_u$, $0 \le \lambda_u \le 1$, respectively.

The proposal distribution for the first MH step in (12) is constructed by making use of the reversible character of the AR process, meaning that the initial state values can be simulated according to

$$x_t = \sum_{i=1}^{k} a_{i,k}x_{t+i} + \sigma_k e_t, \quad e_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \ t = 0, -1, \dots, -k+1,$$

should the AR coefficients and excitation variance be known. As this is not the case, these parameters are substituted with their maximum likelihood (ML) estimates, obtained by assuming the initial state to be zero, i.e.

$$\tilde{\mathbf{a}}_k^{\text{ML}} \triangleq \left(\tilde{\mathbf{X}}_k^{\text{T}} \tilde{\mathbf{X}}_k\right)^{-1} \tilde{\mathbf{X}}_k^{\text{T}} \mathbf{x}_{1:T}, \quad \tilde{\sigma}_k^{2\text{ML}} \triangleq \frac{1}{T} \mathbf{x}_{1:T}^{\text{T}} \tilde{\mathbf{P}}_k \mathbf{x}_{1:T}, \tag{14}$$

with $\tilde{\mathbf{P}}_k \triangleq \mathbf{I}_T - \tilde{\mathbf{X}}_k \left(\tilde{\mathbf{X}}_k^{\text{T}} \tilde{\mathbf{X}}_k\right)^{-1} \tilde{\mathbf{X}}_k^{\text{T}}$. The matrix $\tilde{\mathbf{X}}_k$ is identical to $\mathbf{X}_k$, but with the initial state values set to zero. The implied joint proposal distribution for all the elements of the initial state now follows as

$$q_1(\mathbf{x}'_{0,k}) \triangleq \mathcal{N}(\mathbf{x}'_{0,k}; \mathbf{m}_{0,k}, \tilde{\sigma}_k^{2\text{ML}} \mathbf{M}_{0,k}), \tag{15}$$

with

$$\mathbf{M}_{0,k} \triangleq \left(\tilde{\mathbf{A}}_{0,k}^{\text{T}} \tilde{\mathbf{A}}_{0,k}\right)^{-1}, \quad \mathbf{m}_{0,k} \triangleq -\mathbf{M}_{0,k} \tilde{\mathbf{A}}_{0,k}^{\text{T}} \tilde{\mathbf{A}}_{1,k} \mathbf{x}_{k:1},$$

where $\tilde{\mathbf{A}}_k \triangleq \left[\tilde{\mathbf{A}}_{1,k} \ \tilde{\mathbf{A}}_{0,k}\right] \in \mathbb{R}^{k \times 2k}$ is of the same form as $\mathbf{A}_k$, but constructed with the ML estimate of the AR coefficients in (14). The matrices $\tilde{\mathbf{A}}_{1,k}$ and $\tilde{\mathbf{A}}_{0,k}$ are the first k and final k columns of $\tilde{\mathbf{A}}_k$, respectively. The required conditional proposal distributions $q_1(x'_{-j+1}|\mathbf{x}'_{j,0,k})$, $j = 1,\ldots,k$, are easily obtained from (16), as it is Gaussian.

The proposal distribution for the second MH step in (13) is taken to be a Gaussian random walk, i.e.

$$q_2\left(x'_{-j+1}|x_{-j+1}\right) \triangleq \mathcal{N}(x'_{-j+1}; x_{-j+1}, \sigma_{\text{RW}}^2),$$

where σ_{RW}^2 is the variance of the random walk. This proposal distribution is introduced to perform local exploration of the posterior distribution.

For both MH steps, the corresponding acceptance probability becomes

$$\alpha_{i,j}\left(\mathbf{x}'_{0,k}|\mathbf{x}_{0,k}\right) = \min\left\{1, r_{i,j}\left(\mathbf{x}'_{0,k}|\mathbf{x}_{0,k}\right)\right\}, \quad i = 1, 2,$$

with the acceptance ratio given by

$$r_{i,j}\left(\mathbf{x}'_{0,k}|\mathbf{x}_{0,k}\right) = \left(\frac{\beta_k}{\beta'_k}\right)^{\alpha_k} \left(\frac{|\mathbf{M}'_k||\mathbf{\Sigma}_{\mathbf{A}_k}||\mathbf{\Sigma}_{\mathbf{x}_{0,k}}|}{|\mathbf{M}_k||\mathbf{\Sigma}'_{\mathbf{A}_k}||\mathbf{\Sigma}'_{\mathbf{x}_{0,k}}|}\right)^{1/2} \frac{q_{i,j}\left(\mathbf{x}_{0,k}|\mathbf{x}'_{0,k}\right)}{q_{i,j}\left(\mathbf{x}'_{0,k}|\mathbf{x}_{0,k}\right)},$$

where β'_k, $\mathbf{M}'_k$, $\mathbf{\Sigma}'_{\mathbf{A}_k}$ and $\mathbf{\Sigma}'_{\mathbf{x}_{0,k}}$ are similar to β_k, $\mathbf{M}_k$, $\mathbf{\Sigma}_{\mathbf{A}_k}$ and $\mathbf{\Sigma}_{\mathbf{x}_{0,k}}$, respectively, with $\mathbf{x}_{0,k}$ replaced by $\mathbf{x}'_{0,k}$. Several other proposal distributions may be constructed, but the combination presented here proved to be very efficient in simulations.

4.2.2. Birth and death moves

Numerous sampling strategies can be devised for the birth and death moves. The one presented below makes use of the proposal distributions developed for the update move, and was found to work well in practice. Assuming the current state of the Markov chain to be in $\{k\} \times \mathbf{\Theta}_k$, the birth move at iteration i is summarized below.

ALGORITHM 3: BIRTH MOVE

1. *Set* $\mathbf{x}_{k+1,0,k+1} = (x_0^{(i-1)}, \ldots, x_{-k+1}^{(i-1)})$.
2. *Propose a value for the new element of the initial state* $x_{-k} \sim q_1$ $(x_{-k}|\mathbf{x}_{k+1,0,k+1})$ [see (12)] *and set* $\mathbf{x}_{0,k+1} = (\mathbf{x}_{k+1,0,k+1}, x_{-k})$.
3. *If* $(u \sim \mathcal{U}[0,1]) \leq \alpha_{\mathrm{birth}}\left(k+1, \mathbf{x}_{0,k+1}|k, \mathbf{x}_{0,k}^{(i-1)}\right)$ [see (16)]:
   - *Set* $k^{(i)} = k+1$, $\mathbf{x}_{0,k+1}^{(i)} = \mathbf{x}_{0,k+1}$.
   Else:
   - *Set* $k^{(i)} = k$, $\mathbf{x}_{0,k}^{(i)} = \mathbf{x}_{0,k}^{(i-1)}$.

Similarly, assuming the current state of the Markov chain to be in $\{k+1\} \times \mathbf{\Theta}_{k+1}$, the death move at iteration i is summarized below.

ALGORITHM 4: DEATH MOVE

1. *Set* $\mathbf{x}_{0,k} = (x_0^{(i-1)}, \ldots, x_{-k+1}^{(i-1)})$.
2. *If* $(u \sim \mathcal{U}[0,1]) \leq \alpha_{\mathrm{death}}(k, \mathbf{x}_{0,k}|k+1, \mathbf{x}_{0,k+1}^{(i-1)})$ [see (17)]:
   - *Set* $k^{(i)} = k$, $\mathbf{x}_{0,k}^{(i)} = \mathbf{x}_{0,k}$.
   Else :
   - *Set* $k^{(i)} = k+1$, $\mathbf{x}_{0,k+1}^{(i)} = \mathbf{x}_{0,k+1}^{(i-1)}$.

For both the birth and death moves, the model-order and initial state are sampled in a MH step that admits (10) as invariant distribution. For the birth move, a candidate for the initial state is obtained by setting the first k elements to the corresponding elements of the current initial state, and proposing a candidate for the additional element from $q_1(x_{-k}|\mathbf{x}_{k+1,0,k+1})$, conditional on the current values of the other elements. Similarly, a candidate for the initial state for the death move is obtained as the first k elements of the current initial state. The corresponding MH acceptance probabilities are given by

$$\alpha_{\mathrm{birth}}\left(k+1, \mathbf{x}_{0,k+1}|k, \mathbf{x}_{0,k}\right) = \min\left\{1, r_{\mathrm{birth}}\left(k+1, \mathbf{x}_{0,k+1}|k, \mathbf{x}_{0,k}\right)\right\} \tag{16}$$

$$\alpha_{\mathrm{death}}\left(k, \mathbf{x}_{0,k}|k+1, \mathbf{x}_{0,k+1}\right) = \min\left\{1, r_{\mathrm{birth}}^{-1}\left(k+1, \mathbf{x}_{0,k+1}|k, \mathbf{x}_{0,k}\right)\right\}, \tag{17}$$

with the acceptance ratio given by

$$r_{\mathrm{birth}}\left(k+1, \mathbf{x}_{0,k+1}|k, \mathbf{x}_{0,k}\right)$$
$$= \frac{\Gamma(\alpha_{k+1})\beta_k^{\alpha_k}}{\left(2\pi\delta^2\zeta^2\right)^{1/2}\Gamma(\alpha_k)\beta_{k+1}^{\alpha_{k+1}}} \left(\frac{|\mathbf{M}_{k+1}||\mathbf{\Sigma}_{\mathbf{a}_k}||\mathbf{\Sigma}_{\mathbf{x}_{0,k}}|}{|\mathbf{M}_k||\mathbf{\Sigma}_{\mathbf{a}_{k+1}}||\mathbf{\Sigma}_{\mathbf{x}_{0,k+1}}|}\right)^{1/2} \frac{1}{q_1\left(x_{-k}|\mathbf{x}_{k+1,0,k+1}\right)}.$$

To summarize, the sampling step for the AR parameters at iteration i is presented below.

ALGORITHM 5: SAMPLING THE AR PARAMETERS – STATIONARITY NOT ENFORCED

1. *If* $(u \sim \mathcal{U}[0,1]) \leq b_{k^{(i-1)}}$:
   - *Perform birth move (see Algorithm* 3).

   Else if $u \leq b_{k^{(i-1)}} + d_{k^{(i-1)}}$:
   - *Perform death move (see Algorithm* 4).

   Else:
   - *Perform update move (see Algorithm* 2).

2. *Simulate* $\sigma_{k^{(i)}}^{2(i)} \sim p(\sigma_k^2 | k^{(i)}, \delta^{2(i-1)}, \zeta^{2(i-1)}, \mathbf{x}_{0,k^{(i)}}^{(i)}, \mathbf{x}_{1:T})$ [*see* (9)].

3. *Simulate* $\mathbf{a}_{k^{(i)}}^{(i)} \sim p(\mathbf{a}_k | k^{(i)}, \sigma_{k^{(i)}}^{2(i)}, \delta^{2(i-1)}, \mathbf{x}_{0,k^{(i)}}^{(i)}, \mathbf{x}_{1:T})$ [*see* (8)].

4.3. *Stationarity enforced*

Many MCMC strategies have been developed for the case where stationarity is explicitly enforced [see e.g. (Barnett *et al.*, 1996; Huerta and West, 1999b; Barbieri and O'Hagan, 1996)]. The approach developed here is novel in that all the reflection coefficients are simulated jointly, and not conditional on each other, as in most other strategies. This leads to an algorithm that converges rapidly. For the sake of brevity the initial state is assumed to be fixed and known in what follows. A strategy similar to the one described in Section 4.2 can be adopted if the initial state should be considered as unknown and random.

Recall that for the case where stationarity is enforced, $\boldsymbol{\phi}_k = \boldsymbol{\rho}_k$ and $\boldsymbol{\Phi}_k = (-1,1)^k$. The posterior conditional for the AR parameters follows from Bayes' rule as

$$p(k, \boldsymbol{\rho}_k, \sigma_k^2 | \delta^2, \Lambda, \mathbf{x}_{1:T}) \propto p(\mathbf{x}_{1:T} | k, \mathbf{a}_k = \mathcal{L}(\boldsymbol{\rho}_k), \sigma_k^2) p(\boldsymbol{\rho}_k, \sigma_k^2 | k, \delta^2) p(k | \Lambda). \quad (18)$$

Sampling from this distribution using a reversible jump MCMC step similar to the one developed for the case where stationarity is not enforced is made difficult by the nonlinear nature of the transformation between the AR and reflection coefficients. However, if the true AR coefficients $\mathbf{a}_k = \mathcal{L}(\boldsymbol{\rho}_{kt})$ are replaced by the linear approximation $\mathbf{a}_k = \tilde{\mathbf{U}}_k \boldsymbol{\rho}_k$, a reversible jump MCMC strategy can be developed for the resulting approximate posterior distribution. The Markov chain transition kernel for this step can then be used as a proposal for the true posterior distribution in (18). The details of this approach are now presented.

The approximate posterior distribution, defined as

$$p'(k, \boldsymbol{\rho}_k, \sigma_k^2 | \delta^2, \Lambda, \mathbf{x}_{1:T}) \propto p(\mathbf{x}_{1:T} | k, \mathbf{a}_k = \tilde{\mathbf{U}}_k \boldsymbol{\rho}_k, \sigma_k^2) p(\boldsymbol{\rho}_k, \sigma_k^2 | k, \delta^2) p(k | \Lambda).$$

can be marginalized over the reflection coefficients, and can thus be factorized analytically as

$$p'(k, \boldsymbol{\rho}_k, \sigma_k^2 | \delta^2, \Lambda, \mathbf{x}_{1:T}) = p'(\boldsymbol{\rho}_k | k, \sigma_k^2, \delta^2, \mathbf{x}_{1:T}) p'(k, \sigma_k^2 | \delta^2, \Lambda, \mathbf{x}_{1:T}), \quad (19)$$

where the distributions on the right-hand side are given by

$$p'\left(\boldsymbol{\rho}_k|k,\sigma_k^2,\delta^2,\mathbf{x}_{1:T}\right) = \bar{c}_k\mathcal{N}(\boldsymbol{\rho}_k;\hat{\boldsymbol{\rho}}_k,\sigma_k^2\mathbf{Q}_k)\mathbb{I}_{(-1,1)^k}(\boldsymbol{\rho}_k) \tag{20}$$

$$p'\left(k,\sigma_k^2|\delta^2,\Lambda,\mathbf{x}_{1:T}\right) \propto \frac{c_k\Lambda^k\beta_k^{-(\alpha_0+T/2)}\mathbb{I}_{\mathcal{K}}(k)}{\bar{c}_k\left(\delta^2\right)^{k/2}k!}\left(\frac{|\mathbf{Q}_k|}{|\boldsymbol{\Sigma}_{\boldsymbol{\rho}_k}|}\right)^{1/2}\mathcal{IG}(\sigma_k^2;\alpha_0+T/2,\beta_k), \tag{21}$$

with

$$\mathbf{Q}_k \triangleq \left(\Delta_k^{-2}+\frac{1}{\delta^2}\boldsymbol{\Sigma}_{\boldsymbol{\rho}_k}^{-1}\right)^{-1} \triangleq \mathrm{diag}(q_{1,k}^2,\ldots,q_{k,k}^2)$$

$$\hat{\boldsymbol{\rho}}_k \triangleq \mathbf{Q}_k\tilde{\mathbf{U}}_k^{\mathrm{T}}\mathbf{X}_k^{\mathrm{T}}\mathbf{x}_{1:T} \triangleq (\hat{\rho}_1,\ldots,\hat{\rho}_k)$$

$$\beta_k \triangleq \beta_0+\frac{1}{2}\mathbf{x}_{1:T}^{\mathrm{T}}\mathbf{P}_k\mathbf{x}_{1:T}$$

$$\mathbf{P}_k \triangleq \mathbf{I}_T-\mathbf{X}_k\tilde{\mathbf{U}}_k\mathbf{Q}_k\tilde{\mathbf{U}}_k^{\mathrm{T}}\mathbf{X}_k^{\mathrm{T}}$$

$$\bar{c}_k^{-1} \triangleq 2^{-k}\prod_{i=1}^k\left(\mathrm{erf}\left(\frac{1-\hat{\rho}_i}{\sqrt{2\sigma_k^2q_{i,k}^2}}\right)+\mathrm{erf}\left(\frac{1+\hat{\rho}_i}{\sqrt{2\sigma_k^2q_{i,k}^2}}\right)\right)$$

Note that the normalizing constants c_k and $\bar{c}_k$ are complex non-linear functions of σ_k^2 and δ^2, making further integration of (21) over the excitation variance impossible.

Samples can be generated from the approximate posterior distribution in (19) by using a reversible jump MH step with a proposal distribution of the form

$$q'\left(k',\boldsymbol{\rho}_{k'}',\sigma_{k'}^{2\prime}|k,\boldsymbol{\rho}_k,\sigma_k^2\right) \triangleq q'(k'|k)q'\left(\sigma_{k'}^{2\prime}|k'\right)q'\left(\boldsymbol{\rho}_{k'}'|k',\sigma_{k'}^{2\prime}\right), \tag{22}$$

where the component proposal distributions are given by

$$q'(k'|k) \triangleq s_k\exp(-\lambda|k'-k|)\mathbb{I}_{\mathcal{K}}(k') \tag{23}$$

$$q'\left(\sigma_k^2|k\right) \triangleq \mathcal{IG}(\sigma_k^2;\alpha_0+T/2,\beta_k) \tag{24}$$

$$q'\left(\boldsymbol{\rho}_k|k,\sigma_k^2\right) \triangleq p'\left(\boldsymbol{\rho}_k|k,\sigma_k^2,\delta^2,\mathbf{x}_{1:T}\right). \tag{25}$$

In the proposal for the model move $\lambda > 0$ is a scale parameter, and the normalizing constant is given by

$$s_k^{-1} = \sum_{k'\in\mathcal{K}}\exp(-\lambda|k'-k|).$$

Such a proposal was used before by e.g. Troughton and Godsill (1998), and ensures that most of the proposed model moves are small, without eliminating the possibility of the occasional large one. This proposal can be substituted by any other convenient model move proposal distribution.

The proposal distribution for the reflection coefficients is a truncated Gaussian distribution, and can be sampled from using an accept/reject procedure that simulates from the untruncated distribution until a sample that falls within the stable region is obtained. Note that as the matrix $\mathbf{Q}_k$ is diagonal, this sampling step can be performed for each reflection coefficient independently, utilising the corresponding marginal distributions of (20). For the proposal distribution in (22) the corresponding MH acceptance probability becomes

$$\alpha_c'\left(k', \sigma_{k'}^{2'}|k\right) = \min\left\{1, r_c'\left(k', \sigma_{k'}^{2'}|k\right)\right\}, \tag{26}$$

with the acceptance ratio given by

$$r_c'\left(k', \sigma_{k'}^{2'}|k\right) = \frac{c_{k'}\bar{c}_k \Lambda^{(k'-k)} k!}{c_k \bar{c}_{k'} \left(\delta^2\right)^{(k'-k)/2} k'!} \left(\frac{\beta_k}{\beta_{k'}}\right)^{\alpha_0 + T/2} \left(\frac{|\mathbf{Q}_{k'}||\mathbf{\Sigma}_{\boldsymbol{\rho}_k}|}{|\mathbf{Q}_k||\mathbf{\Sigma}_{\boldsymbol{\rho}_{k'}}|}\right)^{1/2} \frac{q'(k|k')}{q'(k'|k)}$$

Note that this expression is independent of the reflection coefficients, so that these quantities only need to be sampled once a proposed model move is accepted.

The sampling step for the approximate posterior distribution described above defines a Markov chain transition kernel $\mathcal{K}'\left(k', \mathrm{d}\boldsymbol{\rho}_{k'}', \mathrm{d}\sigma_{k'}^{2'}|k, \boldsymbol{\rho}_k, \sigma_k^2\right)$ that has $p'\left(k, \boldsymbol{\rho}_k, \sigma_k^2|\delta^2, \Lambda, \mathbf{x}_{1:T}\right)$ in (19) as invariant distribution, and is reversible, i.e.

$$\int p'\left(k, \mathrm{d}\boldsymbol{\rho}_k, \mathrm{d}\sigma_k^2|\delta^2, \Lambda, \mathbf{x}_{1:T}\right) \mathcal{K}'\left(k', \mathrm{d}\boldsymbol{\rho}_{k'}', \mathrm{d}\sigma_{k'}^{2'}|k, \boldsymbol{\rho}_k, \sigma_k^2\right)$$
$$= \int p'\left(k', \mathrm{d}\boldsymbol{\rho}_{k'}', \mathrm{d}\sigma_{k'}^{2'}|\delta^2, \Lambda, \mathbf{x}_{1:T}\right) \mathcal{K}'\left(k, \mathrm{d}\boldsymbol{\rho}_k, \mathrm{d}\sigma_k^2|k', \boldsymbol{\rho}_{k'}', \sigma_{k'}^{2'}\right).$$

This kernel can be used as a proposal to sample from the true posterior distribution in (18) in a MH step, i.e.

$$q\left(k', \mathrm{d}\boldsymbol{\rho}_{k'}', \mathrm{d}\sigma_{k'}^{2'}|k, \boldsymbol{\rho}_k, \sigma_k^2\right) = \mathcal{K}'\left(k', \mathrm{d}\boldsymbol{\rho}_{k'}', \mathrm{d}\sigma_{k'}^{2'}|k, \boldsymbol{\rho}_k, \sigma_k^2\right).$$

The corresponding acceptance probability for this MH step becomes

$$\alpha_c\left(k', \boldsymbol{\rho}_{k'}', \sigma_{k'}^{2'}|k, \boldsymbol{\rho}_k, \sigma_k^2\right) = \min\left\{1, \frac{\mathrm{LLR}\left(k', \boldsymbol{\rho}_{k'}', \sigma_{k'}^{2'}\right)}{\mathrm{LLR}\left(k, \boldsymbol{\rho}_k, \sigma_k^2\right)}\right\}, \tag{27}$$

where $\mathrm{LLR}\left(k, \boldsymbol{\rho}_k, \sigma_k^2\right)$ is the ratio of the true and approximate likelihood functions, i.e.

$$\mathrm{LLR}\left(k, \boldsymbol{\rho}_k, \sigma_k^2\right) \triangleq \frac{p\left(\mathbf{x}_{1:T}|k, \mathbf{a}_k = \mathcal{L}(\boldsymbol{\rho}_k), \sigma_k^2\right)}{p\left(\mathbf{x}_{1:T}|k, \mathbf{a}_k = \tilde{\mathbf{U}}_k \boldsymbol{\rho}_k, \sigma_k^2\right)}.$$

REMARK 3. The method described above uses the Markov chain transition kernel for the approximate distribution in (19) as a proposal to sample from the true distribution in (18) in an MH step. Using this methodology many different

choices are possible for the approximate distribution and the corresponding sampling strategy, as long as the resulting Markov chain transition kernel is reversible, and the acceptance probability is correctly specified. For example, alternative formulations of $\tilde{\mathbf{U}}_k$ may be considered, based on e.g. the current estimate of the reflection coefficients. The choices adopted here were found to give good results in practice.

To summarize, the sampling step for the AR parameters at iteration i is presented below.

Algorithm 6: Sampling the AR Parameters – Stationarity Enforced

1. *Propose a new value for the model-order* $k' \sim q'(k|k^{(i-1)})$.
2. *Propose a new value for the excitation variance* $\sigma_{k'}^{2'} \sim q'(\sigma_k^2|k')$ [see (24)].
3. *If* $(u \sim \mathcal{U}[0,1]) \leq \alpha_c'(k', \sigma_{k'}^{2'}|k^{(i-1)})$ [see (26)]:
   - *Simulate* $\boldsymbol{\rho}_{k'}' \sim q'(\boldsymbol{\rho}_k|k', \sigma_{k'}^{2'})$ [see (25)].
   Else:
   - *Set* $k' = k^{(i-1)}$, $\sigma_{k'}^{2'} = \sigma_{k^{(i-1)}}^{2(i-1)}$, $\boldsymbol{\rho}_{k'}' = \boldsymbol{\rho}_{k^{(i-1)}}^{(i-1)}$.
4. *If* $(u \sim \mathcal{U}[0,1]) \leq \alpha_c(k', \boldsymbol{\rho}_{k'}', \sigma_{k'}^{2'}|k^{(i-1)}, \sigma_{k^{(i-1)}}^{2(i-1)}, \boldsymbol{\rho}_{k^{(i-1)}}^{(i-1)})$ [see (27)]:
   - *Set* $k^{(i)} = k'$, $\sigma_{k^{(i)}}^{2(i)} = \sigma_{k'}^{2'}$, $\boldsymbol{\rho}_{k^{(i)}}^{(i)} = \boldsymbol{\rho}_{k'}'$.
   Else:
   - *Set* $k^{(i)} = k^{(i-1)}$, $\sigma_{k^{(i)}}^{2(i)} = \sigma_{k^{(i-1)}}^{2(i-1)}$, $\boldsymbol{\rho}_{k^{(i)}}^{(i)} = \boldsymbol{\rho}_{k^{(i-1)}}^{(i-1)}$.

Remark 4. In experimental evaluations, the rejection rate for the second MH step, given that the proposal for the first MH step is accepted (the second MH step is always accepted if the first is rejected), ranged between 30% and 50%, depending on the data and the values chosen for the fixed parameters of the algorithm, showing this to be an efficient sampling strategy.

4.4. Sampling the hyperparameters

The posterior conditionals for the hyperparameters $(\delta^2, \Lambda, \zeta^2)$ follow straightforwardly from Bayes' rule, and are given by

$$p(\delta^2|k, \boldsymbol{\phi}_k, \sigma_k^2) \propto c_k \mathcal{IG}\left(\delta^2; \alpha_{\delta^2} + k/2, \beta_{\delta^2} + \frac{1}{2\sigma_k^2}\boldsymbol{\phi}_k^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\phi}_k}^{-1}\boldsymbol{\phi}_k\right) \qquad (28)$$

$$p(\Lambda|k) \propto c_\Lambda \mathcal{G}a(\Lambda; \alpha_\Lambda + k, \beta_\Lambda) \qquad (29)$$

$$p(\zeta^2|k, \sigma_k^2, \mathbf{x}_{0,k}) = \mathcal{IG}\left(\zeta^2; \alpha_{\zeta^2} + k/2, \beta_{\zeta^2} + \frac{1}{2\sigma_k^2}\mathbf{x}_{0,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{x}_{0,k}}^{-1}\mathbf{x}_{0,k}\right), \qquad (30)$$

where c_Λ is the normalizing constant of the truncated Poisson prior on the model order. If stationarity is not enforced, $c_k = 1$, and the proportionality in (28) becomes an equality. Sampling for δ^2 is then by standard procedures. If stationarity is enforced, however, c_k is a complex non-linear function of δ^2. Sampling for δ^2 can then be performed in a MH step with proposal distribution

$$q(\delta^2) \triangleq \mathcal{IG}\left(\delta^2; \alpha_{\delta^2} + k/2, \beta_{\delta^2} + \frac{1}{2\sigma_k^2}\phi_k^T\Sigma_{\phi_k}^{-1}\phi_k\right)$$

and corresponding MH acceptance probability

$$\alpha_{\delta^2}\left(\delta^{2\prime}|\delta^2\right) = \min\left\{1, \frac{c_k'}{c_k}\right\},$$

where c_k' is similar to c_k, with δ^2 replaced by $\delta^{2\prime}$.

For large values of $k_{\max}$, the posterior conditional for Λ becomes approximately $p(\Lambda|k) \approx \mathcal{G}a(\Lambda; \alpha_\Lambda + k, \beta_\Lambda + 1)$. Thus, Λ is here sampled using a mixture of two MH steps with proposal distributions $q_1(\Lambda|k) \triangleq \mathcal{G}a(\Lambda; \alpha_\Lambda + k, \beta_\Lambda)$ and $q_2(\Lambda|k) \triangleq \mathcal{G}a(\Lambda; \alpha_\Lambda + k; \beta_\Lambda + 1)$, and mixture weights λ_Λ and $1 - \lambda_\Lambda$, $0 < \lambda_\Lambda < 1$, respectively. The first proposal ensures the geometric ergodicity of the Markov chain, whereas the second provides a good approximation to the true posterior conditional in (30). For both these proposals, the corresponding MH acceptance probability becomes

$$\alpha_{\Lambda,i}(\Lambda'|\Lambda) = \min\left\{1, \frac{p(\Lambda'|k)q_i(\Lambda)}{p(\Lambda|k)q_i(\Lambda')}\right\}, \quad i = 1, 2.$$

Finally, the posterior conditional for ζ^2 is an inverted-gamma distribution, sampling from which is by standard procedures.

4.5. Convergence results

It is straightforward to prove that the algorithm converges, i.e. that the Markov chain is ergodic, by proving its irreducibility and aperiodicity. However, a stronger result can be proved, showing that the Markov chain converges geometrically to the required posterior distribution. The proof is suppressed due to space constraints, but can be found in (Vermaak *et al.*, 2000).

5. SIMULATION STUDIES

5.1. Effect of the fixed parameters of the prior distributions

Recall from Section 3 that the Bayesian models are specified by the fixed parameters $(\alpha_0, \beta_0, \alpha_{\delta^2}, \beta_{\delta^2}, \alpha_\Lambda, \beta_\Lambda, \alpha_{\zeta^2}, \beta_{\zeta^2})$. The purpose of this experiment is to

determine suitable values for these parameters and to investigate the sensitivity of the estimation results to changes in their values. Recall that the mean and variance for a gamma distribution $\mathcal{G}a(a,b)$ are given by ab^{-1} and ab^{-2}, respectively, and those for an inverted-gamma distribution $\mathcal{IG}(a, b)$, by $(a-1)^{-1}b$ and $(a-1)^{-2}(a-2)^{-1}b^2$, respectively.

In all the experiments, the parameters of the inverted-gamma prior on the excitation variance are set to $\alpha_0 = \beta_0 = 0$, so that $p(\sigma_k^2) \propto 1/\sigma_k^2$ is the uninformative Jeffreys' prior. An uninformative prior is also assigned to Λ by setting $\alpha_\Lambda = \varepsilon_1 + \frac{1}{2}$, $\beta_\Lambda = \varepsilon_2$, with $\varepsilon_1, \varepsilon_2 \ll 1$. More specifically, for the results reported here these values were set to $\varepsilon_1 = 0.001$, $\varepsilon_2 = 0.0001$.

As δ^2 and ζ^2 are scale parameters, they are assigned slightly informative priors by setting $\alpha_{\delta^2} = \alpha_{\zeta^2} = 2$ and choosing β_{δ^2}, $\beta_{\zeta^2} > 0$. This choice for α_{δ^2} and α_{ζ^2} ensures an infinite variance, so that the crucial parameters affecting the priors are β_{δ^2} and β_{ζ^2}. Here the weak influence of these parameters on the marginal posterior distributions $p(\delta^2|\mathbf{x}_{1:T})$, $p(\zeta^2|\mathbf{x}_{1:T})$ and $p(k|\mathbf{x}_{1:T})$ is experimentally demonstrated.

For the data $T = 100$ samples were simulated from a sixth order AR process with roots at $0.99\angle \pm 0.1\pi$, $0.9\angle \pm 0.3\pi$ and $0.85\angle \pm 0.7\pi$, and excitation variance $\sigma_6^2 = 10$. The sampler was run without enforcing stationarity for a total of $N_b + N_s = 500 + 5000$ iterations, with the fixed parameters of the algorithm set to $k_{\max} = 30$, $\lambda = 0.25$, $\lambda_\Lambda = 0.1$, $c = 0.5$, $\lambda_u = 0.5$ and $\sigma_{RW}^2 = 0.1$. N_b is the burn-in period of the Markov chain, and samples generated during this period were discarded in subsequent posterior analyses. The algorithm was initialized with a zero model-order and random values for the other parameters. The estimates of $p(\delta^2|\mathbf{x}_{1:T})$ and $p(k|\mathbf{x}_{1:T})$ for $\beta_{\delta^2}=1,10,100$ are depicted in Figure 1, and show that the estimation results are stable over a wide range of choices for β_{δ^2}. Similar results were obtained for ζ^2 in the case where the initial state is assumed to be unknown.

5.2. Detection

In this experiment, the model-order detection performance of the methods developed in this paper is compared with that of standard criteria, such as AIC, BIC and MDL. The model selection rule for the standard criteria is of the form

$$\hat{k} = \underset{k \in \mathcal{K}}{\arg\max}\left\{ - \log\ p\left(\mathbf{x}_{1:T}|k, \hat{\boldsymbol{\theta}}_k^{\text{ML}}\right) + P(k, T)\right\},$$

where $\hat{\boldsymbol{\theta}}_k^{\text{ML}}$ is the ML estimate of the parameters, and $P(k, T)$ is a penalty term that increases with the model-order. For an AR process with a fixed and known initial state and for which stationarity is not enforced, the model selection rules corresponding to AIC, BIC and MDL can be evaluated analytically. BIC and MDL yield the same rule, which is asymptotically consistent. However, AIC is not asymptotically consistent (see Gustafsson and Hjalmarsson, 1995, for details).

For the comparison, 100 realizations of T samples each were simulated from a third-order AR process with roots at 0.9 and $0.5\angle \pm 0.85\pi$, and excitation

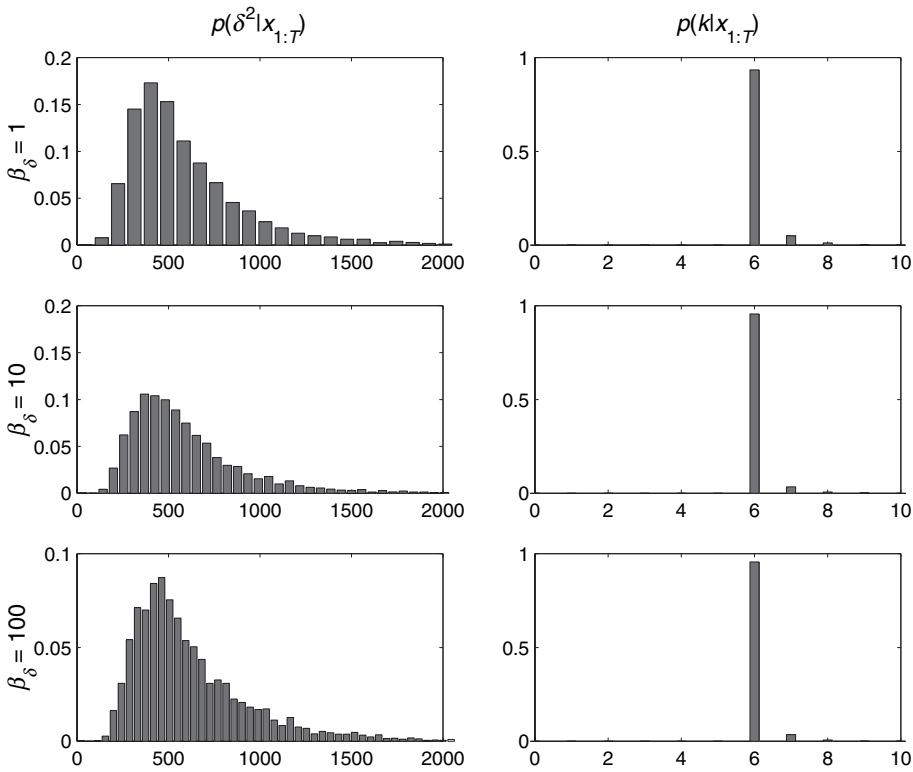FIGURE 1. Estimates of $p(\delta^2|\mathbf{x}_{1:T})$ (left) and $p(k|\mathbf{x}_{1:T})$ (right) for various values of β_{δ^2}.

TABLE 1

DETECTION PERFORMANCE RESULTS (IN ACCURACY PERCENTAGE)

T:	35	50	75	100	200	300
AIC	20	31	49	59	74	76
BIC/MDL	19	30	46	57	76	94
MMAP	23	33	49	64	78	95

variance $\sigma_3^2 = 10$. For each realization, $\hat{k}^{AIC}$, $\hat{k}^{BIC}$ and $\hat{k}^{MMAP}$ were evaluated. Stationarity was not enforced, and the MCMC sampler to compute $\hat{k}^{MMAP}$ was run for a total number of $N_b + N_s = 500 + 5000$ iterations, using values for the fixed parameters of the algorithm and the prior distribution similar to those in Section 5.1.

To investigate the effect of the size of the data set, the experiment was performed for a wide range of values for T. The results are summarized in Table I, and show that the Bayesian method consistently outperforms the standard criteria. The detection accuracy for the Bayesian method and BIC/MDL

approaches 100% as the data sets become large, confirming that these strategies are indeed asymptotically consistent. This is not the case for AIC, for which the detection accuracy stabilizes at around 75%. The Bayesian method also gives more information than the classical techniques in the sense that the samples from the Markov chain can be used to construct empirical approximations to $p(k, \boldsymbol{\theta}_k | \mathbf{x}_{1:T})$ and its marginals for all $k \in \mathcal{K}$.

5.3. *Enforcing stationarity*

This experiment compares the two estimation strategies where stationarity is either ignored or explicitly enforced. For the results reported here the fixed parameters of the algorithm and prior distributions were set to values similar to those in Section 5.1. For the data a realization of $T = 100$ samples was simulated from the AR process specified in Section 5.1, such that the ML estimates of the poles for this realization lie at $1.0026\angle \pm 0.1004\pi$, $0.9046\angle \pm 0.2965\pi$ and $0.8431\angle \pm 0.6972\pi$. The sampler was run on this realization for a total number of $N_b + N_s = 500 + 5000$ iterations for both cases.
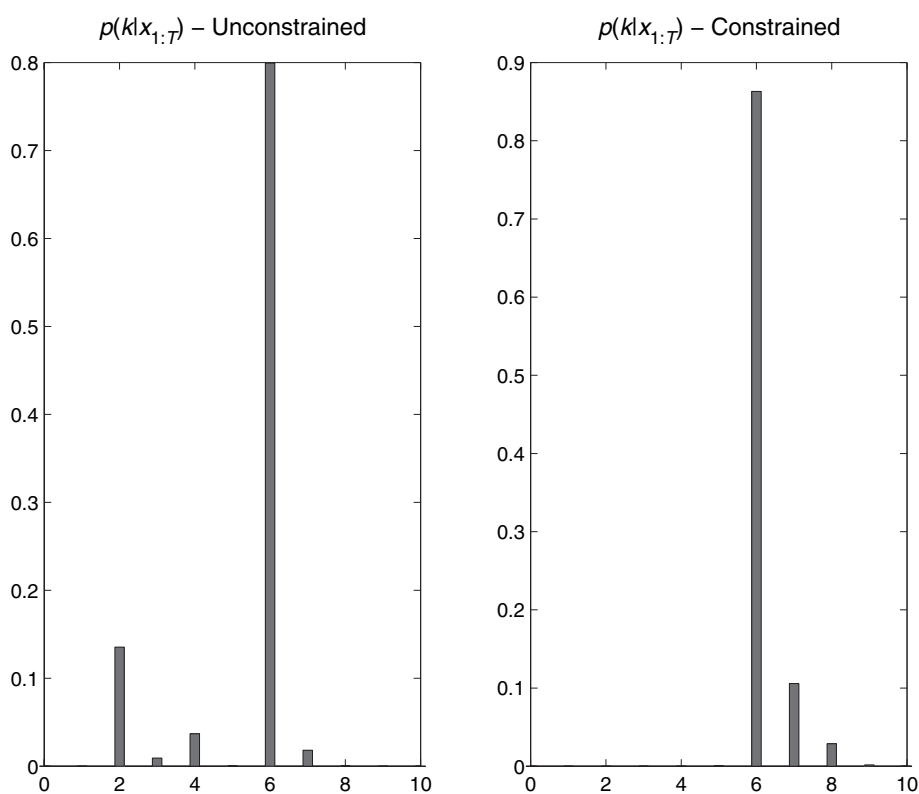

FIGURE 2. Estimates of $p(k|\mathbf{x}_{1:T})$ when ignoring (left) and enforcing (right) stationarity.

The estimates of $p(k|\mathbf{x}_{1:T})$ for both cases are depicted in Figure 2, and indicate that both algorithms correctly estimated the true model-order (note that the histograms are quite different, though). Approximations to the MMSE estimates of the poles were also computed for both cases, and are given by $0.9881\angle \pm 0.10059\pi$, $0.8642\angle \pm 0.3063\pi$ and $0.7822\angle \pm 0.6846\pi$ when enforcing stationarity, and $1.0057\angle \pm 0.1002\pi$, $0.8956\angle \pm 0.3039\pi$ and $0.8471\angle \pm 0.6906\pi$ when not.

5.4. Southern oscillation index data

In this experiment the southern oscillation index (SOI) data is analysed. The El Niño–Southern Oscillation (SOI) is a climatological phenomenon that has been of some interest in climate change studies in recent decades. The SOI series, depicted in the top graph of Figure 3, has 540 observations, computed as the difference of the departure from the long-term monthly mean sea level pressures at Tahiti in the South Pacific and Darwin in Northern Australia. The monthly index spans the period from 1950 to 1995, and is related to sea surface temperatures.
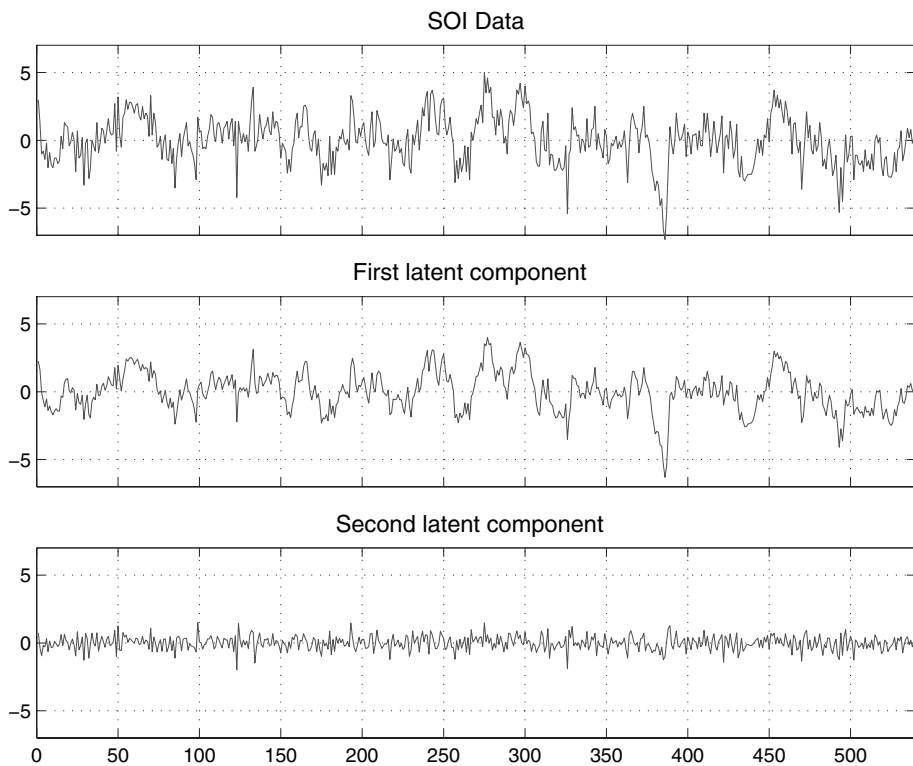

FIGURE 3. The SOI series (top) and its decomposition into two latent components (middle and bottom).

© Blackwell Publishing Ltd 2004

This data set was analysed using AR processes within a Bayesian setting before in (Huerta and West, 1999b). The Bayesian model adopted there specifies a prior structure directly on the roots of the AR characteristic polynomial. This prior structure places upper bounds on the maximum number of real roots and complex conjugate root pairs, and explicitly allows for zero and unit roots. Stationarity is enforced by setting the prior to be zero for roots with moduli greater than one. Model uncertainty is implicitly accounted for as a consequence of allowing the roots to have zero moduli.

The analysis performed here is based on a different Bayesian model that yielded an interesting and more parsimonious interpretation of the SOI data than by Huerta and West (1999b). One of the main differences between the strategy developed here and that reported by Huerta and West (1999b) is the fact that the prior distribution in (5) allows for arbitrary combinations of real roots and complex conjugate root pairs. The model developed here does not, however, explicitly allow for unit roots as the model by Huerta and West (1999b) does. The form of the Bayesian model adopted here also facilitates block parameter sampling strategies that lead to fast convergence.
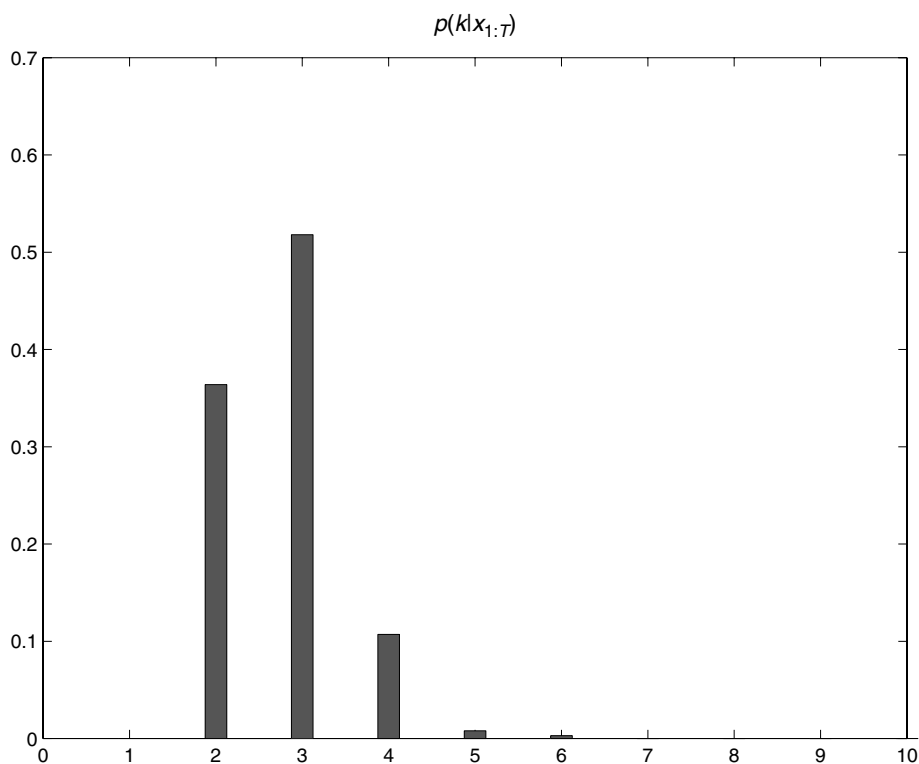

$p(k|x_{1:T})$

FIGURE 4. An estimates of $p(k|\mathbf{x}_{1:T})$ for the SOI series.

The results presented by Huerta and West, (1999b) favour model-orders ranging from 8 to 17 with 3 to 5 complex components. The mode of the posterior distribution of the model-order is at 11. However, in the decomposition of the time series only one latent component was found to be clearly dominant, suggesting that the data may be described by a more parsimonious model. This is indeed confirmed by the results of the analysis performed here, which favour a model-order of 3 (the order chosen by BIC is 2), as depicted by the estimate of $p(k|\mathbf{x}_{1:T})$ in Figure 4. These results were obtained by running the sampler for a total number of $N_b + N_s = 500 + 5000$ iterations. The initial state was assumed to be unknown, and stationarity was not enforced. Values similar to those in Section 5.1 were used for the fixed parameters of the algorithm and the prior distribution, except for k_{max}, which was set to 40, as is the case of Huerta and West (1999b). Similar results were obtained when explicitly enforcing stationarity.

The approximate MMSE roots for the model with $k = 3$ were calculated, and yielded one real root with modulus 0.8361 and one complex conjugate root pair at $0.3622\angle \pm 0.6838\pi$. The corresponding latent components are depicted in the bottom and middle graphs of Figure 3, and confirm the fact that there is only one clearly dominant latent component.

6. CONCLUSIONS

In this paper, the Bayesian inference problem for AR processes was addressed. Original hierarchical prior models were proposed that allow the stationarity of the model to be enforced. These models were shown to be robust over a wide range of values for their fixed hyperparameters. Reversible jump MCMC algorithms were developed to perform the analytically intractable Bayesian computations. The simulation results presented in Section 5 indicate that the algorithms are accurate and computationally efficient.

NOTES

Corresponding author: C. Andrieu, Department of Mathematics, Statistics Group, University of Bristol, University Walk, Bristol BS81 TW, UK. E-mail: c.andrieu@bris.ac.uk

REFERENCES

AKAIKE, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* AC-19, 716–23.

BARBIERI, M. M. and O'HAGAN, A. (1996) *A reversible jump MCMC sampler for Bayesian analysis of ARMA time series*. Dipartimento di Statistica Probabilità e Statistiche Applicate Università "La Sapienza" Roma, Italy and Department of Mathematics, University of Nottingham, Nottingham, UK.

BARNETT, G., KOHN, R. and SHEATHER, S. (1996) Bayesian estimation of an autoregressive model using Markov chain Monte Carlo. *Journal of Econometrics* 74, 237–54.

BOX, G. E. P., JENKINS, G. M. and REINSEL, G. C. (1994) *Time Series Analysis, Forecasting and Control* (3rd edn). Englewood Cliffs, NJ: Prentice-Hall. ISBN 0-13-060774-6.

FRIEDLANDER, B. (1982) Lattice filters for adaptive processing. *Proceedings of the IEEE* 70, 829–67.

GODSILL, S. J. (2001) On the relationship between MCMC model uncertainty methods. *Journal of Computational and Graphical Statistics* 10:2, 230–48.

GREEN, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82, 711–32.

GUSTAFSSON, F. and HJALMARSSON, H. (1995) Twenty-one ML estimators for model selection. *Automatica* 31, 1377–92.

HUERTA, G. and WEST, M. (1999a) Bayesian inference on periodicities and component spectral structure in time series. *Journal of Time Series Analysis* 20, 401–16.

HUERTA, G. and WEST, M. (1999b) Priors and component structures in autoregressive time series models. *Journal of the Royal Statistical Society Series* B 61, 881–99.

POSKITT, D. S. and TREMAYNE, A. R. (1983) On the posterior odds of time series models. *Biometrika* 70, 157–62.

RICHARDSON, S. and GREEN, P. J. (1997) On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society Series* B 59, 731–58.

RISSANEN, J. (1978) Modeling by shortest data description. *Automatica* 14, 465–78.

ROBERT, C. P. and CASELLA, G. (1999) *Monte Carlo Statistical Methods*. New York: Springer-Verlag.

ROBERT, C. P., RYDÉN, T. and TITTERINGTON, D. M. (2000) Bayesian inference in hidden Markov models through jump Markov chain Monte Carlo. *Journal of the Royal Statistical Society Series* B 62, 57–75.

SCHWARZ, G. (1985) Estimating the dimension of a model. *The Annals of Statistics* 6, 461–4.

SMITH, M., WONG, C. and KOHN, R. (1998) Additive nonparametric regression with autocorrelated errors. *Journal of the Royal Statistical Society Series* B 60, 311–31.

TROUGHTON, P. T. and GODSILL, S. J. (1997) *A reversible jump sampler for autoregressive time series. Employing Full conditionals to achieve efficient model space moves*. CUED/F-INFENG/TR.304 Signal Processing Group. Cambridge: Cambridge University Engineering Department.

TROUGHTON, P. T. and GODSILL, S. J. (1998) A reversible jump sampler for autoregressive time series. *Proceedings of the IEEE International Conference on Acoustic Speech and Signal Processing* 4, 2257–60.

VERMAAK, J., ANDRIEU, C., DOUCET, A. and GODSILL, S. J. (2000) *Bayesian model selection of autoregressive processes*. CUED/F-INFENG/TR.360 Signal Processing Group. Cambridge: Cambridge University Engineering Department.

WEST, M. and HARRISON, J. (1997) *Bayesian Forecasting and Dynamic Models* (2nd edn). New York: Springer-Verlag ISBN 0-387-94725-6.