Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Forecasting 20 (2004) 713–730
international journal of forecasting
www.elsevier.com/locate/ijforecast

Bayesian time series analysis of periodic behaviour and spectral structure

E.J. McCoy, D.A. Stephens*

Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2BZ, UK

Abstract

Analysis, model selection and forecasting in univariate time series models can be routinely carried out for models in which the model order is relatively small. Under an ARMA assumption, classical estimation, model selection and forecasting can be routinely implemented with the Box–Jenkins time domain representation. However, this approach becomes at best prohibitive and at worst impossible when the model order is high. In particular, the standard assumption of stationarity imposes constraints on the parameter space that are increasingly complex. One solution within the pure AR domain is the latent root factorization in which the characteristic polynomial of the AR model is factorized in the complex domain, and where inference questions of interest and their solution are expressed in terms of the implied (reciprocal) complex roots; by allowing for unit roots, this factorization can identify any sustained periodic components. In this paper, as an alternative to identifying periodic behaviour, we concentrate on frequency domain inference and parameterize the spectrum in terms of the reciprocal roots, and, in addition, incorporate Gegenbauer components. We discuss a Bayesian solution to the various inference problems associated with model selection involving a Markov chain Monte Carlo (MCMC) analysis. One key development presented is a new approach to forecasting that utilizes a Metropolis step to obtain predictions in the time domain even though inference is being carried out in the frequency domain. This approach provides a more complete Bayesian solution to forecasting for ARMA models than the traditional approach that truncates the infinite AR representation, and extends naturally to Gegenbauer ARMA and fractionally differenced models.

Keywords: Gegenbauer ARMA models; Long memory; Latent root; Spectral representation; Markov chain Monte Carlo; Forecasting

1. Introduction

Recent developments in Bayesian time series analysis have been facilitated by computational methods such as Markov chain Monte Carlo (MCMC). Parametric modelling of time series has thus far been restricted to autoregressive (AR) models (see, for example, Barnett, Kohn, & Sheather, 1996). Huerta and West (1999a, 1999b) have developed an MCMC scheme based on the characteristic root structure of AR processes.

We extend this methodology here by allowing for autoregressive moving average (ARMA) components and adopting a frequency domain approach. The Wold decomposition theorem states that any indeterministic stationary process can be expressed as an infinite MA process, and any such process can be well approxi-

* Corresponding author.
E-mail addresses: e.mccoy@imperial.ac.uk (E.J. McCoy), d.stephens@imperial.ac.uk (D.A. Stephens).

0169-2070/$ - see front matter © 2004 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2004.02.001

mated by an ARMA model. By allowing for MA components, the approximating model is generally parsimonious, so avoiding the potential need for very high order autoregressions.

A simple extension to k-factor GARMA (Gegenbauer ARMA) models of Woodward, Cheng, and Gray (1998) allows for the modelling of persistent autocorrelations associated with several frequencies. The spectrum associated with these models becomes unbounded at the specified frequencies making them particularly appropriate for data with slowly damping autocorrelations, which also exhibit cyclical behaviour. Time domain inference for these more general models tends to be cumbersome, in contrast, due to the fact that, for stationary Gaussian processes, the Fourier transform converts serial correlation into heteroscedasticity, the Whittle based frequency domain approach we employ is computationally more attractive.

In this paper, we utilize a variable dimension Markov chain Monte Carlo algorithm which caters for prior uncertainty about model order in both the autoregressive and moving average constituents, as well as prior uncertainty about the number and location of the Gegenbauer components. However, our emphasis is on reconstruction of the spectral density/autocovariance sequence, rather on model order identification. This emphasis is due to issues of practical identifiability (many ARMA models of different orders have very similar spectral densities, and thus distinction of one model from another for a finite sample may be difficult, and in any case not relevant), and also because the spectral density is a quantity that is interpretable for, and comparable across, all stationary processes, whereas ARMA coefficients are not.

In this paper, our approach concentrates on the spectral structure of the GARMA time-series model. However, the methodology could easily be extended to incorporate non-parametric or semi-parametric estimates of the spectrum as in Carter and Kohn (1997), or by allowing for non-symmetry in the spectrum around the poles as in Arteche and Robinson (2000).

The basic structure of the model in both time and frequency domain is developed in Section 2. Section 3 concentrates on prior modelling and posterior inference, while Section 4 describes how forecasting and imputation of missing values can easily be incorporated into the scheme. In Section 5, we discuss model checking, and time and frequency domain residual analysis. In Section 6, we demonstrate the methodology through the analysis of the variable star *S. Carinæ* data, analyzed by Huerta and West (1999b) in which periodic components are evident.

2. Time series models

2.1. Time domain representation

The process $\{X_t\}$ with zero mean follows an ARMA(p,d,q) model if it satisfies the following equation:

$$\Phi(B)\Delta^d X_t = \Theta(B)\epsilon_t, \tag{1}$$

where B is the backward shift operator, $\Delta^d=(1-B)^d$, where d is the positive integer valued level of differencing required to achieve stationarity and $\{\epsilon_t\}$ is a white noise process with zero mean and variance σ_ϵ^2. The AR and MA polynomials are

$$\begin{aligned}\Phi(B) &= 1-\phi_1 B-\ldots-\phi_p B^p\\ \Theta(B) &= 1+\theta_1 B+\ldots+\theta_q B^q\end{aligned} \tag{2}$$

with $\phi_p \neq 0$, $\theta_q \neq 0$. Sufficient conditions for stationarity and invertibility of the differenced series $Y_t=\Delta^d X_t$ are that the roots of the two polynomial equations

$$\Phi(z)=0,\ \ \Theta(z)=0$$

lie outside the unit circle.

Following Huerta and West (1999a), we use factorizations of Φ and Θ in terms of reciprocal roots, that is, we write

$$\Phi(z)=\prod_{j=1}^{p}(1-\zeta_{1j}z) \qquad \Theta(z)=\prod_{j=1}^{q}(1-\zeta_{2j}z),$$

where $(\zeta_{11},\ldots,\zeta_{1p})$ and $(\zeta_{21},\ldots,\zeta_{2q})$ are potentially complex-valued parameters; for both Φ and Θ, the reciprocal roots are either real, or appear in complex conjugate pairs. We denote by p_1 and p_2, and q_1 and q_2 the number of real and complex conjugate pairs that appear in the factorizations of Φ and Θ, respectively (so that $p=p_1+2p_2$ and $q=q_1+2q_2$). For the

complex roots, we further reparameterize so that, for $r=1, 2$,

$$\zeta_{rj} = \alpha_{rj}e^{i2\pi\omega_{rj}} = \alpha_{rj}\cos 2\pi\omega_{rj} + i\alpha_{rj}\sin 2\pi\omega_{rj};$$
$$\zeta_{rj+1} = \alpha_{rj}e^{-i2\pi\omega_{rj}} = \alpha_{rj}\cos 2\pi\omega_{rj} - i\alpha_{rj}\sin 2\pi\omega_{rj}.$$

For stationarity and invertibility, we merely require that the Moduli/Arguments (α_{rj}, ω_{rj}) satisfy

$$0 \le \alpha_{1j}, \alpha_{2j} < 1, \quad 0 \le \omega_{1j}, \omega_{2j} < \frac{1}{2}.$$

For the real roots, we write $\zeta_{1j}=\alpha_{1j}$ and $\zeta_{2j}=\alpha_{2j}$. Finally, we label the reciprocal roots in increasing order of Argument, and then Modulus, so that, for Φ we have

$$0 = \omega_{11} = \omega_{12} = \ldots = \omega_{1p_1} < \omega_{1p_1+1} = \omega_{1p_1+2} < \ldots < \omega_{1p-1} = \omega_{1p} < \tfrac{1}{2};$$
$$0 \le \alpha_{11} < \alpha_{12} < \ldots < \alpha_{1p_1} < \alpha_{1p_1+1} = \alpha_{1p_1+2} < \ldots < \alpha_{1p-1} = \alpha_{1p} < 1, \tag{3}$$

and similarly for Θ.

We note here that, if inference about the individual roots is required, then the labelling issue cannot be completely ignored. Furthermore, if there is interest in the components of the AR and MA vectors, then that is possible simply by reparameterization (straightforward in a simulation-based setting). However, as we prefer to concentrate on reconstruction of the spectral density, the labelling and reparameterization issues do not arise significantly. The coefficients/roots of the AR and MA polynomials are regarded merely as artifacts of our modelling framework that facilitate the reconstruction and prediction components of our analysis.

2.2. Frequency domain representation

The stationary process $\{Y_t\}$ follows an ARMA(p,q) model whose spectrum can be written as,

$$S_Y(f) = \sigma_\epsilon^2 \frac{\left|\Theta(e^{i2\pi f})\right|^2}{\left|\Phi(e^{i2\pi f})\right|^2} = \sigma_\epsilon^2 \frac{\prod_{j=1}^{q}\left|1-\zeta_{2j}e^{i2\pi f}\right|^2}{\prod_{j=1}^{p}\left|1-\zeta_{1j}e^{i2\pi f}\right|^2}$$
$$= \sigma_\epsilon^2 \frac{\prod_{j=1}^{q_1}\left|(1-\alpha_{2j}e^{i2\pi f})\right|^2 \prod_{j=q_1+1}^{q_2}\left|\left(1-\alpha_{2j}e^{i2\pi(f-\omega_{2j})}\right)\left(1-\alpha_{2j}e^{i2\pi(f+\omega_{2j})}\right)\right|^2}{\prod_{j=1}^{p_1}\left|(1-\alpha_{1j}e^{i2\pi f})\right|^2 \prod_{j=p_1+1}^{p_2}\left|\left(1-\alpha_{1j}e^{i2\pi(f-\omega_{1j})}\right)\left(1-\alpha_{1j}e^{i2\pi(f+\omega_{1j})}\right)\right|^2}.$$

It is clear from the spectral representation that the characteristic roots of the AR polynomial determine the periodic behaviour of the data, with bounded peaks in the spectral density near to the ω_{1j} at frequency

$$\bar{\omega}_{1j} = \frac{1}{2\pi}\cos^{-1}\left(\frac{1+\alpha_{1j}^2}{2\alpha_{1j}}\cos 2\pi\omega_{1j}\right).$$

The peaks become more pronounced as $\alpha_{1j}\to 1$ (so that $\bar{\omega}\to\omega$), implying that those roots with Moduli near the unit circle tend to dominate the behaviour of the process. Other features of the ARMA spectrum are determined, for example, by the complex roots of the MA polynomial; the root (α_{2j}, ω_{2j}) corresponds to a local minimum at

$$\bar{\omega}_{2j} = \frac{1}{2\pi}\cos^{-1}\left(\frac{1+\alpha_{2j}^2}{2\alpha_{2j}}\cos 2\pi\omega_{2j}\right).$$

2.3. Seasonally varying long memory models

Granger and Joyeux (1980) and Hosking (1981) defined a generalization of the ARIMA model by allowing d in Eq. (1) to take non-integer values. For $0<d<1/2$, such models are stationary, but exhibit long memory or persistence. The decay of the autocovariance sequence for such processes is hyperbolic rather than the exponential decay implicit in stationary ARMA models. Such models associate persistence with frequency zero. This long-memory model has been investigated extensively; for a review, see Beran (1994), Bisaglia and Guegan (1998), or Robinson (1995a, 1995b).

A further generalization introduced by Giraitis and Leipus (1995) and Woodward et al. (1998) allow for seasonal persistence to be associated with any of k frequencies, $f_j \in [0, 1/2)$, $j=1,\ldots,k$. These so called k-factor GARMA models are written

$$\Phi(B)\prod_{j=1}^{k}(1-2\lambda_j B+B^2)^{\delta_j}Y_t = \Theta(B)\epsilon_t, \tag{4}$$

where $\{\lambda_j\}$, $\{\delta_j\}$ model the k seasonally persistent periodicities. The process $\{Y_t\}$ is stationary, invertible and long-memory whenever the roots of $\Phi(z)$ and $\Theta(z)$ lie outside the unit circle, and the λ_i are distinct and $\delta_i<1/2$ whenever $|\lambda_i|\neq 1$ and $\delta_i<1/4$ whenever $|\lambda_i|=1$.

The spectrum of a k-factor GARMA model can be written as,

$$S_Y(f) = \sigma_\epsilon^2 \frac{\prod_{j=1}^{q} \left|(1-\zeta_{2j}e^{i2\pi f})\right|^2}{\prod_{j=1}^{p} \left|(1-\zeta_{1j}e^{i2\pi f})\right|^2} \times \frac{1}{\prod_{j=1}^{k}[4\{\cos(2\pi f)-\lambda_j\}^2]^{\delta_j}},$$

where the long-memory periodic behaviour is associated with frequencies $f_j = 1/(2\pi)\cos^{-1}\lambda_j$, while the associated δ_j determine the rate of the decay of the autocorrelations. Hassler (1994), Ooms (1995), Ooms and Franses (2001), and Porter-Hudak (1990) propose seasonal long memory models, which are special cases of the above k-factor GARMA model.

The interpretation of the role of the λ parameters in the k-factor GARMA model is similar to the interpretation of the ω_1 parameters in an ARMA model, that is, they correspond to peaks in the spectral density at the specified frequency, the only distinction being in fact that the GARMA frequencies introduce long memory whereas the AR frequencies do not. Therefore, we will tailor the computational methods used to reflect this common interpretation. Note also that, theoretically, we can distinguish between a GARMA and an AR spectral peak, but in reality it may be difficult to obtain practical identifiability from a finite sample.

2.4. Frequency domain analysis

The periodogram, $\hat{S}^{(p)}(f)$, given by

$$\hat{S}^{(p)}(f) = \frac{1}{N}\left|\sum_{t=1}^{N}(Y_t-\mu)e^{-i2\pi ft}\right|^2 \quad -\frac{1}{2}\le f\le\frac{1}{2},$$

is an asymptotically unbiased, but inconsistent estimator of $S(f)$.

For N large enough, the periodogram ordinates $\hat{S}^{(p)}(f_j)$ satisfy the following model

$$Z_j = \hat{S}^{(p)}(f_j) = S(j/N)u_j, \quad j = 0,\ldots,J = \frac{N}{2};\ f_j = \frac{j}{N}, \tag{5}$$

where the $\{u_j, j=1,\ldots,J\}$ are independently distributed as $\chi^2_2/2$ random variables, while u_0 and u_J are χ^2_1 random variables. For a full discussion of the properties of the periodogram, see, for example, Percival and Walden (1993). Asymptotic theory for the Whittle function has been investigated for long memory models including those with cyclical behaviour (see, for example, Arteche & Robinson, 2000; Taqqu & Teverovsky, 1997; Walker, 2000). In particular, Walker (2000) showed that all standard theory for periodogram values of stationary processes with continuity at the Fourier frequencies is valid under very general conditions, relaxing the usually stronger conditions on the spectrum (for example, Anderson, 1971). In addition, an empirical study by Montanari, Taqqu, and Teverovsky (1999) demonstrated the suitability of Whittle-type estimators in the presence of periodicity. The only issue that remains to be addressed is the behaviour and statistical properties of the periodogram near to the spectral poles that arise in GARMA models. This problem has been studied, for example, by Hurvich and Beltrao (1993) for fractionally differenced processes with poles at zero, and Robinson (1995a) gives practical recommendations as to how to estimate the fractional difference parameter when the pole is at zero. As noted in Arteche and Robinson (2000), the distributional assumption of periodogram ordinates will break down for frequencies very close to spectral poles.

Hurvich and Beltrao (1993) show that, for long-memory time series, it is not true that the periodogram ordinates may be treated as i.i.d. exponential random variables. They show that, asymptotically, the periodogram values follow that of an unequally weighted linear combination of two independent χ^2_1 random variables. They also show that the asymptotic relative bias in the periodogram is positive for most values of the long-memory parameter. These results, at first sight, appear to contradict the work of Yajima (1989) and Walker (2000), who state that the usual distributional results stated in Eq. (5) hold, even for long-memory processes. However, Hurvich and Beltrao are looking at the asymptotic distributions at the Fourier frequencies (which change with n), while Yajima and Walker consider the periodogram ordinates at *fixed* frequencies, which do not depend on n.

Some results concerning the asymptotic properties of the maximum likelihood estimators for GARMA models have been established (Giraitis, Hidalgo, &

Robinson, 2001; Ferrara & Guegan, 2001). Olhede, McCoy, and Stephens (2003) have recently extended the Hurvich and Beltrao (1993) results for the distribution of the periodogram ordinates in the case of an underlying GARMA model. The crucial factor determining the divergence of the distribution from standard theory is the distance between the periodogram ordinates and the Gegenbauer pole. These theoretical results have been confirmed with an extensive simulation study, in which empirically the distributional assumptions we make are realistic at the grid of Fourier frequencies, except at those frequencies close to the Gegenbauer frequencies, where indeed the periodogram ordinates are biased.

Hurvich and Beltrao (1993) note that tapering dramatically improves the asymptotic bias at the Fourier frequencies, such tapering could be employed in the GARMA framework and the effect of such tapering is again investigated in Olhede et al. (2003), where it is shown that tapering does not necessarily improve the bias. If this approach is undertaken, then an additional consequence is that the values of the tapered periodogram evaluated at the Fourier frequencies are no longer independent, so this would need to be taken into account.

2.5. Inference via the Whittle-type likelihood

The vector of parameters η, in the simplest ARMA model, is given by the $2p+2q+4$ parameter vector η

$$\eta = \{p, q, \alpha_{11}, \ldots, \alpha_{1p}, \alpha_{21}, \ldots, \alpha_{2q}, \omega_{11}, \ldots, \omega_{1p}, \omega_{21}, \ldots, \omega_{2q}, \mu, \sigma_\epsilon^2\}, \tag{6}$$

(the AR and MA orders, p and q modulus and argument pairs for the complex AR and MA reciprocal roots, respectively—note that, for the real roots, some of the argument parameters will be identically zero—and unconditional mean and conditional variance). The associated Whittle-type log-likelihood is given by

$$\begin{aligned} l(\eta \mid Z) = &-\sum_{j=1}^{J-1}\left[\log\{S_Y(f_j)\} + \frac{Z_j}{S_Y(f_j)}\right] \\ &-\frac{1}{2}\left[\frac{Z_0}{S_Y(f_0)} + \frac{Z_J}{S_Y(f_J)} + \log\{Z_0 S_Y(f_0)\}\right. \\ &\left. + \log\{Z_J S_Y(f_J)\}\right] \end{aligned} \tag{7}$$

where Z is the vector of periodogram values evaluated at the Fourier frequencies given in Eq. (5). As outlined in Section 2.4, periodogram ordinates close to the poles must be treated with care. We have several options: these ordinates may be excluded from the likelihood, as they are likely to be affected by smoothing bias; or alternatively the bias can be incorporated with the inclusion of a bias correction which can be calculated from the results of Olhede et al. (2003). Further discussion of the effect of such choices can be found in the examples of Section 6.

In the k-factor GARMA model, the parameter η in Eq. (6) is augmented by the $2k$ parameters $((\delta_j, \lambda_j): j=1,\ldots,k)$ that appear in Eq. (4), giving a parameter vector of length $2(p+q+k+2)$.

3. Bayesian analysis

Inference about the unknown spectral density or, equivalently, about the many parameters in the ARMA-type models that are used as basis functions over which we are mixing in order to obtain the spectral representations is required. Here, we concentrate on Bayesian inference, and thus interest centres on the posterior distribution of the parameters in the ARMA models, and of course on model order.

Generically, the quantity of interest is the full posterior distribution is given up to proportionality by

$$p(\eta|Z) \propto l(\eta|Z)p(\eta) \tag{8}$$

where $l(\eta|Z)$ is the Whittle likelihood as in Eq. (7) and where the parameter vector η incorporates all ARMA process parameters, the Gegenbauer process parameters, and also the model order of each process, plus process mean and variance parameters μ and σ_ϵ^2.

Although we will compute the full posterior distribution, we may be interested in posterior marginals or posterior summaries for functions of the parameters in our parametric representation. In fact, our principal inferential interest will lie with computing posterior summaries for the spectral density function itself, which is a function of the various ARMA/Gegenbauer parameters that appear in the likelihood. Furthermore, we will be interested in *predictive* rather than *inferential* performance of the models, and principally predictive performance in the time domain. However,

for periodic data, we will also be interested in posterior inference about the locations (in the frequency domain) of the periodicities.

For this non-linear and non-conjugate modelling specification, posterior results are not analytically available, and our only recourse is simulation-based inference, specifically MCMC.

A note on identifiability: The mixture (across GARMA models) representation that we use is globally richly parameterized, each ARMA (p,q) model has $2(p+q)$ parameters, each GARMA component has $2k$, and recall that we are allowing p,q and k to vary. It is therefore questionable whether the parameters will be practically identifiable within the model, although all parameters will be technically identifiable if a sensible prior specification is used. This phenomenon is entirely expected, but not a negative feature; essentially, many GARMA models of different orders are effectively indistinguishable in terms of spectral density/autocovariance function, and it is really these quantities (rather than individual model parameters) that are being inferred about via the likelihood function. In a sense, the mixture of GARMA models representation is a form of non-linear (extended Fourier) basis representation that is applicable to stationary processes. Note that the quantity of primary interest, the spectral density, has a common interpretation in all models irrespective of model order.

As an example, one possible form of practical non-identifiability is associated with a Gegenbauer factor that places an integrable pole in the spectral density at frequency f_j that may be effectively represented by a complex AR root that places a peak (but not a pole) at the same frequency. For any finite sample, the likelihood may not distinguish between the two alternative representations. In terms of model fit, there may well be little to distinguish the two models, and in fact the only difference may be in terms of predicted behavior, and only then for distant time horizons.

3.1. Prior modelling

For the full Bayesian specification, the prior distribution for the various parameters that appear in the many GARMA mixture components should be specified. Broadly, we will adopt a pragmatic prior modelling strategy and choose default priors for the parameters. For example, we could have uniform priors on all of the continuous parameters in a model of fixed order, that is, for the ARMA parameters

$$\alpha \sim U(0,1) \qquad \omega \sim U(0,1/2)$$

and for the Gegenbauer parameters

$$\lambda \sim U(-1,1),\ \delta \sim U(0,1/4) \text{ when } |\lambda| = 1,$$
$$\delta \sim U(0,1/2) \text{ when } |\lambda| \neq 1$$

with an (improper) flat prior for μ and a (proper but diffuse) Inverse Gamma prior for σ_ϵ^2. For the discrete parameters in the model, that is the numbers of real and complex AR and MA roots and the Gegenbauer factor order, we assume identical uniform priors over the integers on the range from zero up to some maximum value (typically 30 in our analysis).

This specification differs slightly from that of Huerta and West (1999a), but not significantly. One natural prior extension would be to have a prior on the complex latent roots that is uniform on the interior of the unit circle, that is, where the prior for modulus α is changed to

$$\alpha \sim \text{Beta}(3,1).$$

We have experimented with other prior specifications, usually by relaxing the uniform prior assumption to an assumption of a more general Beta distribution prior, but have found that inferences, particularly about the spectral density itself, and predictions, are disappointingly robust to the prior specification. The only possible concern regarding the prior specifications is that they too strongly favour particular spectral shapes that we would regard a priori to be unlikely. This is fairly easy to check (by simulation for example) in models of fixed order and also in the mixture representation that we utilize. All elicitation experiments that we have carried out (forward simulation of spectral shapes from the prior model) indicate that the default uniform specification supports a sufficiently broad range of (but does not overly favour particular) spectral shapes. Finally, we note that, technically, the Wold decomposition theorem says that we can give support to all spectral shapes (albeit perhaps not uniformly) using a mixture of ARMA representations.

The Bayesian prior can be viewed as a means of introducing a subjective element in the modelling process; for example, if prior opinion favours an AR(1) parameter near 1 then a non-uniform beta prior can be used (as in, for example, the modelling of stationary financial time series via stochastic volatility; see, for example, Shephard & Pitt, 1997). Furthermore, seasonalities that are believed to be present a priori can be modelled using a Gegenbauer factor, or an appropriately chosen (complex) AR root. It is clear that, whereas prior modelling with reference to time domain behaviour is reasonably complicated, the equivalent modelling process with reference to the frequency domain is much more straightforward. For example, as stated above, complex AR roots are associated with peaks in the spectral density, and in our parameterization it can be shown that the local peak associated with a complex root pair with parameters (α, ω) occurs at frequency

$$\varpi = \hat{f}_{\alpha,\omega} = \frac{1}{2\pi}\cos^{-1}\left(\frac{1+\alpha^2}{2\alpha}\cos 2\pi\omega\right)$$

which is near to ω if α is large enough. This representation suggests an alternative parameterization in terms of α and the modal frequency ϖ, and independent uniform priors on (α, ϖ) induces a non-uniform joint prior on ω of the form

$$p(\alpha,\omega) = \frac{\sin(2\pi\omega)}{\sqrt{4\alpha^2 - (1+\alpha^2)^2\cos^2(2\pi\omega)}}$$
$$0 < \alpha < 1, 0 < \omega < 1/2.$$

Informative priors on the numbers of real/complex roots and on the GARMA factor parameters are a little harder to construct; we again merely point out that forward simulation from the prior models produces an acceptable degree of support for a wide range of spectral shapes.

3.2. Bayesian computation via MCMC

Bayesian inference for the posterior distribution in Eq. (8) is reasonably straightforward using MCMC methods. A key technical reference for MCMC is Tierney (1994), and for a survey of Bayesian inference methods using MCMC; see, for example, Robert and Casella (1999). Essentially, we use a combination of Gibbs sampler and Metropolis-within-Gibbs to sample the full conditional posterior distributions within a model, and augment these moves with dimension-changing proposals that allow the inclusion of an extra term, or exclusion of a current term, in the GARMA representation. MCMC methods are well established, but variable dimension MCMC is less well known, and hence we include a brief summary below in the context of our GARMA mixture representation.

Due to space constraints, we omit precise details of the full conditional distributions, which are of similar but non-standard form; the likelihood in Eq. (7) is complicated and non-linear in the parameters due to the form of the spectral density function. The full posterior distribution in Eq. (8) yields the set of full conditionals easily, as, for component s of η, by standard theory

$$p(\eta_s|Z,\eta_{(s)}) \propto p(\eta|Z) \propto l(\eta|Z)p(\eta)$$
$$\propto l(\eta|Z)p(\eta_s|\eta_{(s)})$$

where $\eta_{(s)}$ is η with component s removed—the simplification of $l(\eta|Z)$ to $l(\eta_s|Z, \eta_{(s)})$ is of little utility.

3.2.1. Variable dimension MCMC

Essentially, variable dimension MCMC (specifically Metropolis–Hastings, MH) is identical to fixed dimension MCMC with additional moves to allow changes of dimension. In this analysis, the variable dimension nature of the problem arises because we do not wish to fix ARMA or GARMA model order. The simplest dimension changing MCMC moves (that make proposals in models with one more or one fewer (G)ARMA root)—*Birth* and *Death* moves, for K to $K+1$ or $K-1$ total parameters say, are easy to formulate.

Full and quite general details of the dimension-changing move were given by Green (1995); we reproduce the central argument here for a generic move between two real spaces. Consider MCMC sampling of a target posterior distribution π defined (at least) on the product of subspaces of d_1 and d_1+d_2 dimensions respectively. From current point $x \in \mathbb{R}^{d_1}$, a move is proposed move to point $y \in \mathbb{R}^{d_1+d_2}$ as a function of x and new random variables $u \sim Q(\cdot)$ where $u \in \mathbb{R}^{d_2}$ as

$y=g(x,u)$ where g is $1-1$ and Q is some proposal density. The move to y is accepted with probability

$$\alpha(x,y)=\min\left\{1,\frac{\pi(y)}{\pi(x)}\frac{P_{d_1+d_2,d_1}}{P_{d_1,d_1+d_2}}\frac{1}{Q(u)}\left|\frac{\partial y}{\partial(x,u)}\right|\right\} \tag{9}$$

where $P_{k,l}$ is the probability of selecting the move from k to l dimensions, and the final term is the Jacobian of the transformation from (x,u) to y. The final combination of terms

$$\frac{P_{d_1+d_2,d_1}}{P_{d_1,d_1+d_2}}\frac{1}{Q(u)}\left|\frac{\partial y}{\partial(x,u)}\right| \tag{10}$$

is the ratio of transition densities that appears in the calculation of an MH acceptance probability. To move from the lower to the higher dimensional space, we need to first decide to make that type of move (the first term in Eq. (10)), augment the current parameter vector x with some new variables u sampled from some probability density (term $Q(u)$ in Eq. (10)), and then reparameterize (x,u) to y, requiring the inclusion of a Jacobian term.

Dimension changing moves often amount to choosing a move type (listed in Section 3.2.2) uniformly on those available, sampling the parameters for augmentation from their prior distribution. This ensures that some terms from $Q(u)$ in Eq. (10) cancel with corresponding terms from posterior π, so that the Jacobian term is one. We give further details in Section 3.2.2.

3.2.2. MCMC moves for the spectral density analysis

The MH move types of both fixed and variable dimension that we use are listed below; we give details of a specific dimension changing move, but omit details of the other moves as they use standard Metropolis within Gibbs transition mechanics.

Fixed dimension moves

- Change of AR or MA term (real or complex) or Gegenbauer factor
- Change of mean or innovations variance.

Variable dimension/model changing moves

- Birth/Death of AR or MA term (real or complex) or Gegenbauer factor
- Change of a Gegenbauer factor to the equivalent complex AR root.

At the start of each iteration, one of these move types is selected; where necessary, terms are selected to be removed at random and uniformly from the terms available. Whenever a dimension-changing Birth move is proposed, the extra parameters required for the new AR/MA root or Gegenbauer factor are sampled from their prior distribution; this ensures that terms in the acceptance probability that relate to these parameters cancel, and the remaining terms involve only the likelihood ratio, the prior on the model order, and the (transition) probabilities of proposing a Birth move or its complementary death move.

The dimension-changing moves are the most complicated and we discuss here an illustrative example. Consider birth moves in the AR component. If the current model is AR(p), where $p=p_1+2p_2$, the sum of the number of real roots and complex-paired roots, let the current parameter vector be denoted $x=(x_1,\ldots,x_{p_1},x_{p_1}+1,\ldots,x_{p_1}+2p_2)$. With probability 1/2, we select either to birth a real root (move type 1) or birth a complex-pair root (move type 2). These moves involve adding either one or two parameters. Consider first move type 1; we generate single new variable u from the prior density for real AR parameters, say $p_R^{(\mathrm{AR})}(.)$ and augment x to $y=(x_1,\ldots,x_{p_1},u,x_{p_1}+1,\ldots,x_{p_1}+2_{P2})$. The acceptance probability from Eq. (9) becomes

$$\alpha(x,y)=\min\left\{1,\frac{l(y\,|\,\eta_{(y)},Z)}{l(x\,|\,\eta_{(x)},Z)}\frac{p(y\,|\,\eta_{(y)})}{p(x\,|\,\eta_{(x)})}\frac{\mathrm{Prob}[p_1+1\text{ real AR roots}]}{\mathrm{Prob}[p_1\text{ real AR roots}]}\frac{Q(y\to x)}{Q(x\to y)}\right\}, \tag{11}$$

where $\eta_{(y)}$ is the (conditioning) parameter vector, that is, η with the elements corresponding to y omitted, $p(\cdot|\cdot)$ is the conditional prior for the parameter concerned and Q is a generic proposal density. The likelihood ratio must be computed, but the prior ratio simplifies (under the assumption of prior independence) as

$$\frac{p(y|\eta_{(y)})}{p(x|\eta_{(x)})}=p_R^{(\mathrm{AR})}(u). \tag{12}$$

The complementary reverse death move, in which one of an existing set of p_1+1 real roots is to be deleted, involves first proposing the elimination of a real root (with probability 1/2), and then removing one of the

available p_1+1 roots (which we do by first arbitrarily labelling and then selecting label at random with equal probability). This gives the transition ratio as

$$\frac{Q(y \to x)}{Q(x \to y)} = \frac{\frac{1}{2} \times \frac{1}{p_1+1}}{\frac{1}{2} \times q(u)} = \frac{\frac{1}{p_1+1}}{p_R^{(\mathrm{AR})}(u)} \tag{13}$$

and the acceptance probability as

$$\alpha(x,y) = \min\left\{1, \frac{l(y \mid \eta_{(y)}, Z)}{l(x \mid \eta_{(x)}, Z)} \frac{\mathrm{Prob}[p_1+1 \text{ real AR roots}]}{\mathrm{Prob}[p_1 \text{ real AR roots}]}\right\} \tag{14}$$

because the other terms cancel; the numerator in Eq. (13) cancels with a term induced in the prior by introducing the AR root labels for selection of the root to be deleted, the denominator cancels with the remaining term in Eq. (11).

The calculation for move type 2 proceeds identically, except that two new variables u_1 and u_2 must be generated from the corresponding prior $p_C^{(\mathrm{AR})}(.,.)$; the acceptance probability becomes

$$\alpha(x,y) = \min\left\{1, \frac{l(y \mid \eta_{(y)}, Z)}{l(x \mid \eta_{(x)}, Z)} \frac{\mathrm{Prob}[p_2+1 \text{ complex} - \text{paired AR roots}]}{\mathrm{Prob}[p_2 \text{ complex} - \text{paired AR roots}]}\right\}$$

An alternative AR model move would involve replacing all current real or complex AR roots by a new set or roots with one more (or one fewer) root or pair of roots. The acceptance probability is identical to that above.

For fixed dimension moves, a uniform prior/proposal for many of the model parameters proved to be rather efficient, although other moves (local or reflecting boundary Metropolis on the bounded parameter space) were also implemented with success. Although there did not seem to be any great improvement in convergence, our final ''hybrid'' algorithm utilized a mixture of Metropolis–Hastings kernels. Parameter μ is easily updated using a local symmetric Metropolis step and σ_ϵ^2 is updated by a hybrid Metropolis–Hastings approach, utilizing both local and independent proposal mechanisms.

4. Forecasting and imputation of missing values

In theory, prediction for time series data is straightforward in a Bayesian setting. Formally, the *posterior predictive* distribution for future data $Y^\dagger$ in light of data Y and the model is defined by

$$p(Y^\dagger \mid Y) = \int p(Y^\dagger \mid \eta, Y) p(\eta \mid Y) \mathrm{d}\eta \tag{15}$$

which can then be used as the basis for prediction via its mean or median, with a measure of uncertainty automatically available via the posterior predictive variance. In practice, however, the integration in Eq. (15) is generally not analytically tractable. In an MCMC setting, the integral may be computed numerically, or a sample from the posterior predictive obtained using a Gibbs sampler step, where the sample is obtained from the full conditional $p(Y^\dagger \mid \eta, Y)$ given the current values of the model parameters η.

For an AR model, the full conditional for $Y^\dagger$, given η and Y, is Gaussian and the sample may be obtained routinely. The most common approach for ARMA models, for which the full conditional is not directly available, is to truncate the infinite AR representation to give a Gaussian full conditional. For GARMA and fractionally differenced models, a similar truncation approach may be used, but the adequacy of the approximation decreases. Hence, for the more complex models, the truncation method is somewhat unsatisfactory. Below, we propose an MCMC approach to forecasting future values and imputing missing values that obviates the need for this truncation.

To recap, imputation and prediction of missing and future values can be incorporated into the MCMC scheme using the method detailed, for example, by McCulloch and Tsay (1994), where the missing values are treated as a set of additional latent variables. In most cases, in a time series context, we require predictions in the time domain. Initially, this appears problematic, as our model and inferences are formulated in the frequency domain. Generating a realization of a stochastic process given its spectral representation is not straightforward; approximate methods have been proposed by, for example, Percival (1992).

Our novel approach uses the MCMC algorithm that has been used for the posterior computations. Specifically, a simple Metropolis step is used to accept or reject proposed values according to the spectrum-based likelihood (7), conditional on system parame-

ters. Specifically, for missing data Y^* (that is, data missing from the range 1,. . .,N), we have

$$p(Y^*|Y) = \int p(Y^*|\eta, Y)p(\eta|Y)\mathrm{d}\eta,$$

which involves an awkward, high-dimensional integral over the parameter space. In a Gibbs sampler/MCMC setting, the full conditional for the (unobserved) Y^* is given by

$$p(Y^*|\eta, Y) \propto p(Y, Y^*|\eta) = l(Y, Y^*|\eta),$$

which provides a route to incorporating Y^* in the inference; we merely include Y^* as a parameter, and update it as part of the MCMC analysis via a (possibly symmetric) Metropolis step in which the acceptance probability is

$$\alpha(Y^*_{\text{current}}, Y^*_{\text{candidate}}) = \min\left\{1, \frac{l(Y, Y^*_{\text{candidate}}|\eta)}{l(Y, Y^*_{\text{current}}|\eta)} \frac{q(Y^*_{\text{candidate}}, Y^*_{\text{current}})}{q(Y^*_{\text{current}}, Y^*_{\text{candidate}})}\right\}.$$

Note that the augmented likelihood $l(Y,Y^*|\eta)$ appears in all full conditionals for the parameters and model order, that is, it contributes to the inference. Similarly, we can sample data as yet unobserved but to be predicted in the future $Y^\dagger$. In the full conditional, we have

$$p(Y^\dagger|\eta, Y, Y^*) \propto p(Y, Y^*, Y^\dagger|\eta) = l(Y, Y^*, Y^\dagger|\eta).$$

The acceptance probability now is based on the doubly augmented likelihood, that is

$$\alpha(Y^\dagger_{\text{current}}, Y^\dagger_{\text{candidate}}) = \min\left\{1, \frac{l(Y, Y^*, Y^\dagger_{\text{candidate}}|\eta)}{l(Y, Y^*, Y^\dagger_{\text{current}}|\eta)} \frac{q(Y^\dagger_{\text{candidate}}, Y^\dagger_{\text{current}})}{q(Y^\dagger_{\text{current}}, Y^\dagger_{\text{candidate}})}\right\}.$$

Note here that the predicted values $Y^\dagger$ do not play any further role in the inference. For convenience, to maintain the same periodogram ordinates in the case of forecasting, the system parameters are determined from the original data, but the proposal likelihood is determined from $\{Y_j\}$ for $j = N+1,\ldots,N+M$, where M is the required number of forecasts and $\{Y_{N+1},\ldots, Y_{N+M}\}$ are the latent variables. Typically, we will select a symmetric proposal mechanism q, but could also use an independent proposal based on predictions from a fit in the time domain of an appropriate (and low order) AR model.

The highest frequency about which we can obtain spectral information is the Nyquist frequency which corresponds to a period of two time units, the consequences of this for imputation is a correlation between adjacent residuals, that is, the model smooths the data. To counter this problem we implement a data augmentation technique, which samples at twice the Nyquist rate but produces fitted values determined from the spectrum fitted to the original data, thus producing uncorrelated residuals at the original sampling rate.

Our proposed forecasting approach is quite straightforward to implement and provides a more complete method for prediction than many currently proposed methods. The extra computation required is not negligible, but neither is it excessive with careful programming. In practice, we only carry out the prediction from good, but fairly automatically chosen starting predictions—say those obtained from an AR approximation—after convergence has been reached for the model parameters. For the imputation of missing values, of course, we have to iteratively update as part of the general MCMC implementation, so this does induce more of a computational burden. Finally, we note that this approach to forecasting and imputation could be applied for any parametric or semiparametric

Table 1
S. Carinæ data posterior probabilities on marginal/joint model order: single MCMC run with points near poles not removed

	AR order	MA order	Gegenbauer order
3	0.105	0.023	0.029
4	0.724	0.085	0.044
5	0.167	0.556	0.104
6	0.003	0.001	0.453
7	0.001	0.023	0.293
8	0.001	0.149	0.069
9	0.000	0.158	0.008
10	0.000	0.004	0.000

Model order (AR,MA,G)	Probability
(4,5,6)	0.258
(4,8,7)	0.123
(4,5,5)	0.073
(4,5,7)	0.066

Illustration of posterior summaries available.

modelling approach to spectral density reconstruction, for example, that of Carter and Kohn (1997) and Mallick, Denison, and Gangopadhyay (2002).

5. Model validation and selection

We employ classical model validation tools (specifically residual analysis) applied in a Bayesian context, that is, we look at the distributions of random quantities derived from sampled parameters and spectra, and the observed data. The frequency domain residuals are straightforward to compute and analyze; it follows from Eq. (5) that the frequency domain residuals on a decibel scale

$$10\log10(\hat{S}^{(p)}(f_j)) - 10\log10(S_Y(f_j)); \quad j = 1,\ldots,J-1,$$

should be i.i.d. with the same distribution as 10log10 of a r.v. with a $\chi^2_2/2$ distribution. Inspection of these residuals gives an impression of whether the spectral shape as indicated by the periodogram is reflected in the reconstructed (posterior) spectral density.

For previous frequency domain analyses, inspection of fit in the *time domain* has not been readily possible, but in our MCMC approach the analysis is relatively

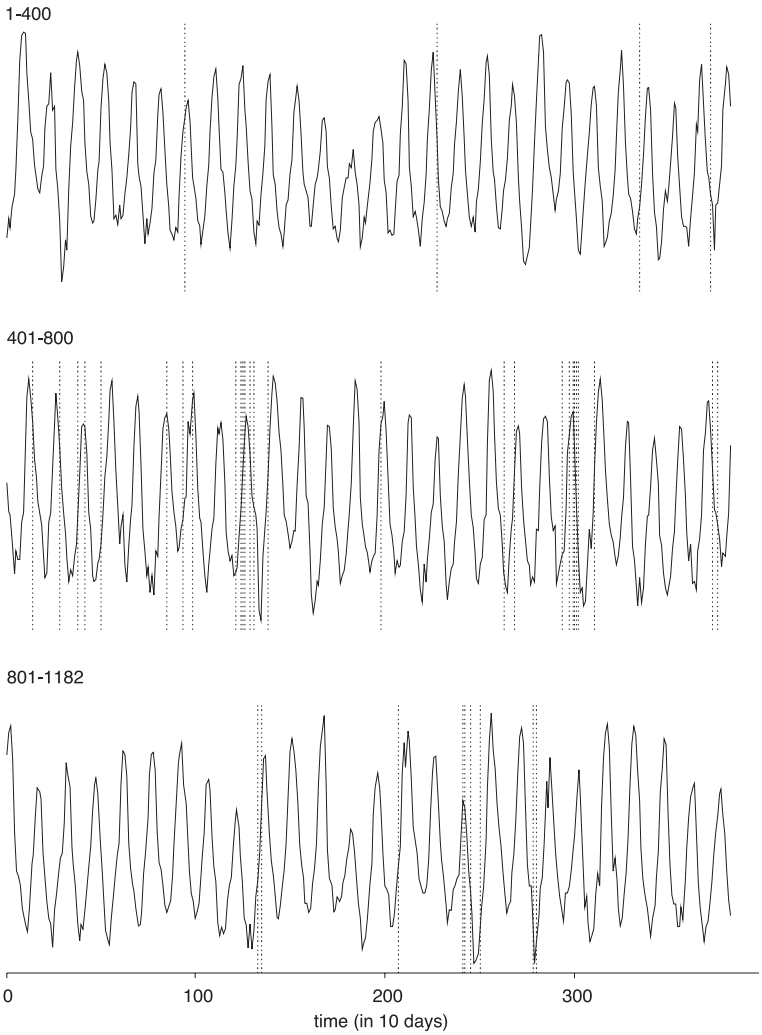

Fig. 1. *S. Carinæ* data: raw data with the positions of missing values shown by the vertical dotted lines.

straightforward; we merely use the Metropolis–Hastings mechanism described in Section 4. To analyze time domain residuals, we employ the same algorithm as for prediction of missing values, that is, samples from the predictive posterior distribution can be obtained by leaving out each point in turn and using a Metropolis rejection step to sample from the posterior distribution of the missing point under the current model. In the notation of Section 3, the full conditional distribution for the prediction at time point t is given by

$$p(Y_t^{\dagger}|\eta, Y_{(t)}) \propto p(Y_t^{\dagger}, Y_{(t)}|\eta) = l(Y_t^{\dagger}, Y_{(t)}|\eta),$$

from which we sample (using Metropolis) a value $y_t^{\dagger}$, and then construct $r_t = y_t^{\dagger} - y_t$.

For quantitative model adequacy assessment, we could utilize either prior predictive (marginal likelihood) or posterior predictive model selection criteria, again either in the time of frequency domains, to assess further whether the model (or more precisely the model class used) provide an adequate fit to the data. This presents only some further computational expense as outlined above; essentially, we produce a Monte Carlo estimate of the required log marginal likelihood/predictive density. If required, it is straightforward to provide model comparisons (on whatever basis) for fixed order AR models, or mixtures of AR models, or ARMA models, or GARMA models, etc. Our emphasis throughout this paper has, however, been on providing a means of implementing a flexible class of GARMA type models, the members of which we can mix over in order to produce a fit that is satisfactory overall, but more importantly provides reliable forecasts.

6. Example: variable star *S. Carinæ* data

To illustrate the method of reconstruction, model validation and selection, and prediction as outlined in

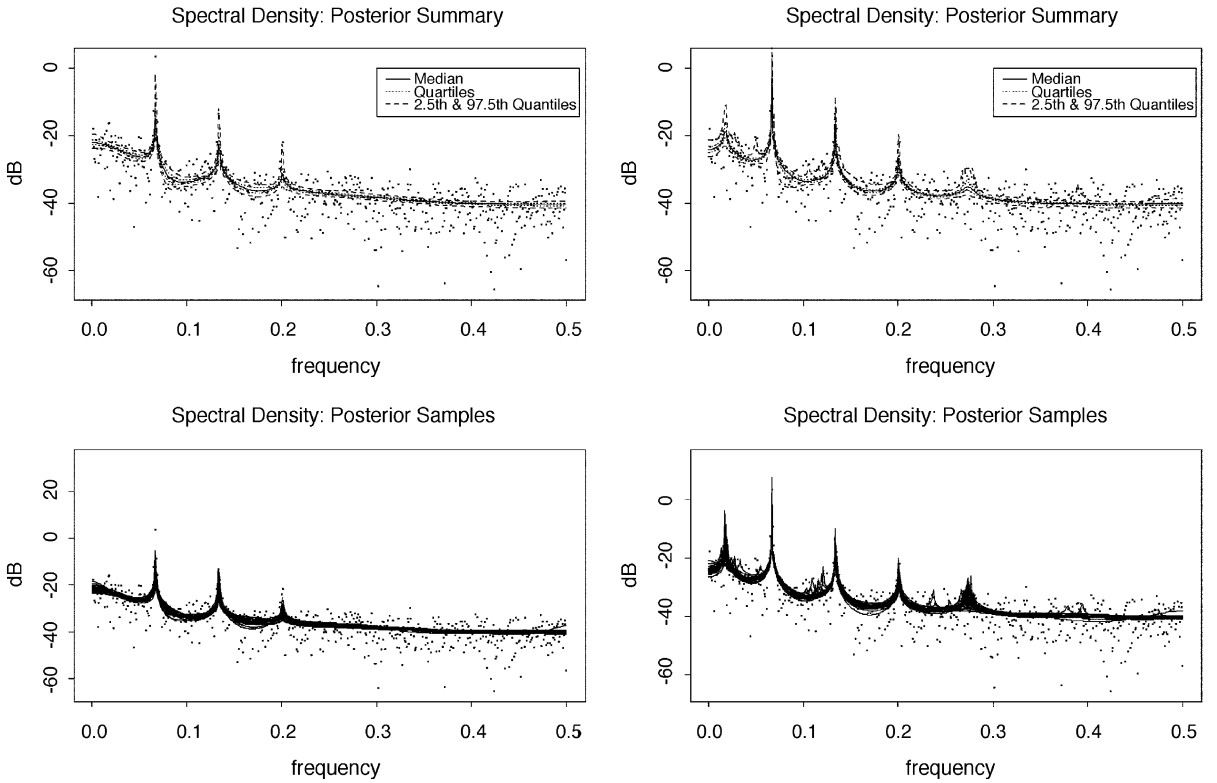

Fig. 2. *S. Carinæ* data: posterior samples of the GARMA spectrum with data points near spectral poles removed (left-hand column) and not removed (right-hand column).

our discussion above, we consider the *S. Carinæ* data which has an interesting periodic structure and missing data. For this data, we present spectral reconstructions a posteriori, an investigation of model fit, and predictions for a range of *k*-step ahead predictions.

The posterior results presented are illustrative of the kinds of summaries available, and include posterior probabilities on models and model orders (Table 1), posterior spectral density reconstructions (quantile summaries and samples), and boxplots of predicted values. The results presented correspond to a single MCMC run (of 2.5 million iterations, including a burn-in of 0.5 million iterations) that have been validated over several repeat runs.

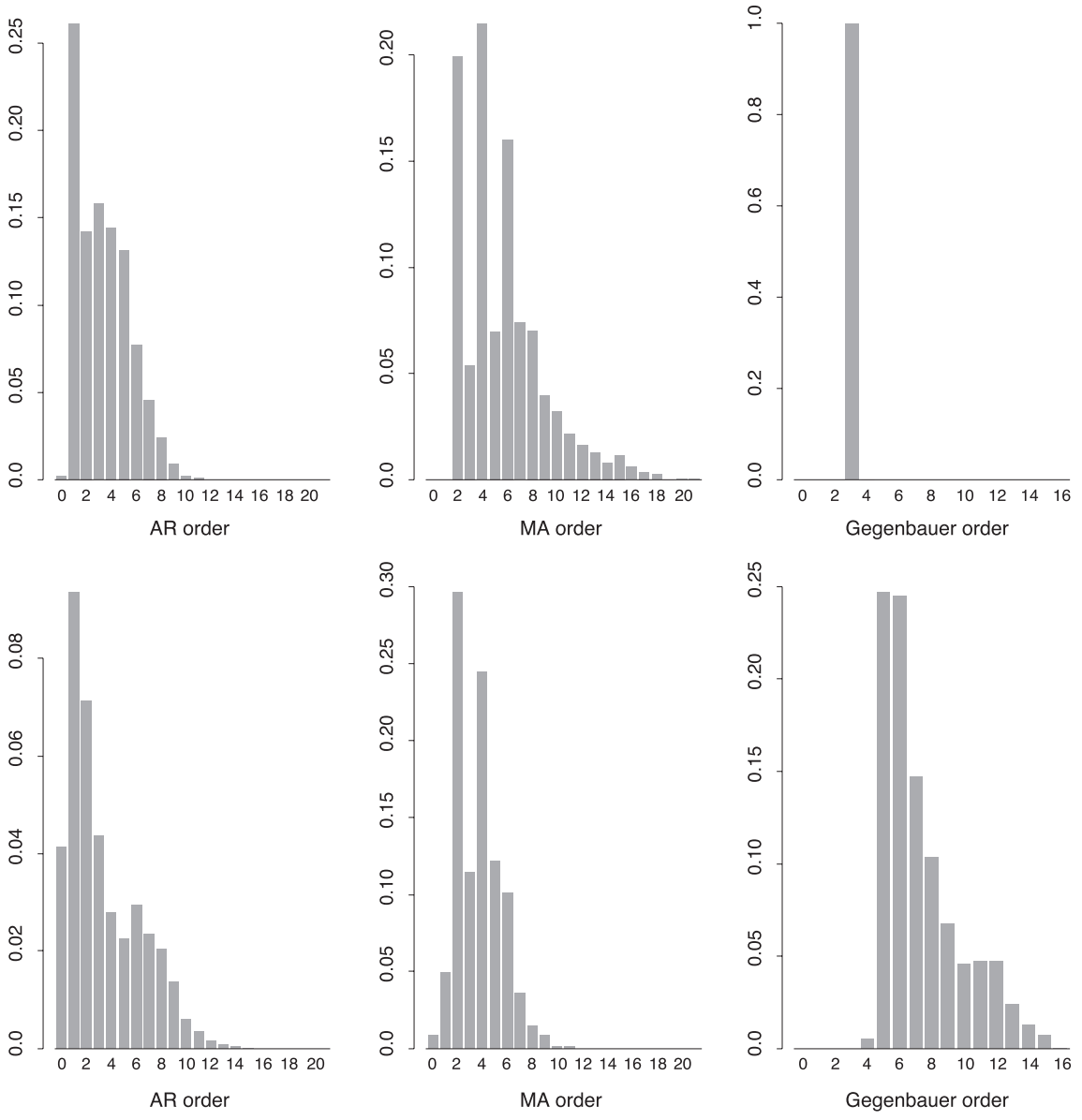

Fig. 3. *S. Carinæ* data: posterior distributions for *k*, *p* and *q*, with data points near spectral poles removed (top row) and not removed (bottom row).

Fig. 1 shows 1182 consecutive 10-day averages of light intensities of the *S. Carinæ* variable star in the southern hemisphere sky. These data, which contain 40 missing values, are from Baldwin and Thomson (1978), and have recently been analyzed by Carter and Kohn (1997) and Huerta and West (1999b). As noted in Huerta and West, a visual inspection of the data displays a period of around 150 days. In order to facilitate direct comparisons, we employ the double log transformation adopted by Huerta and West.

To compare the results obtained by the two alternative means of treating the periodogram values at

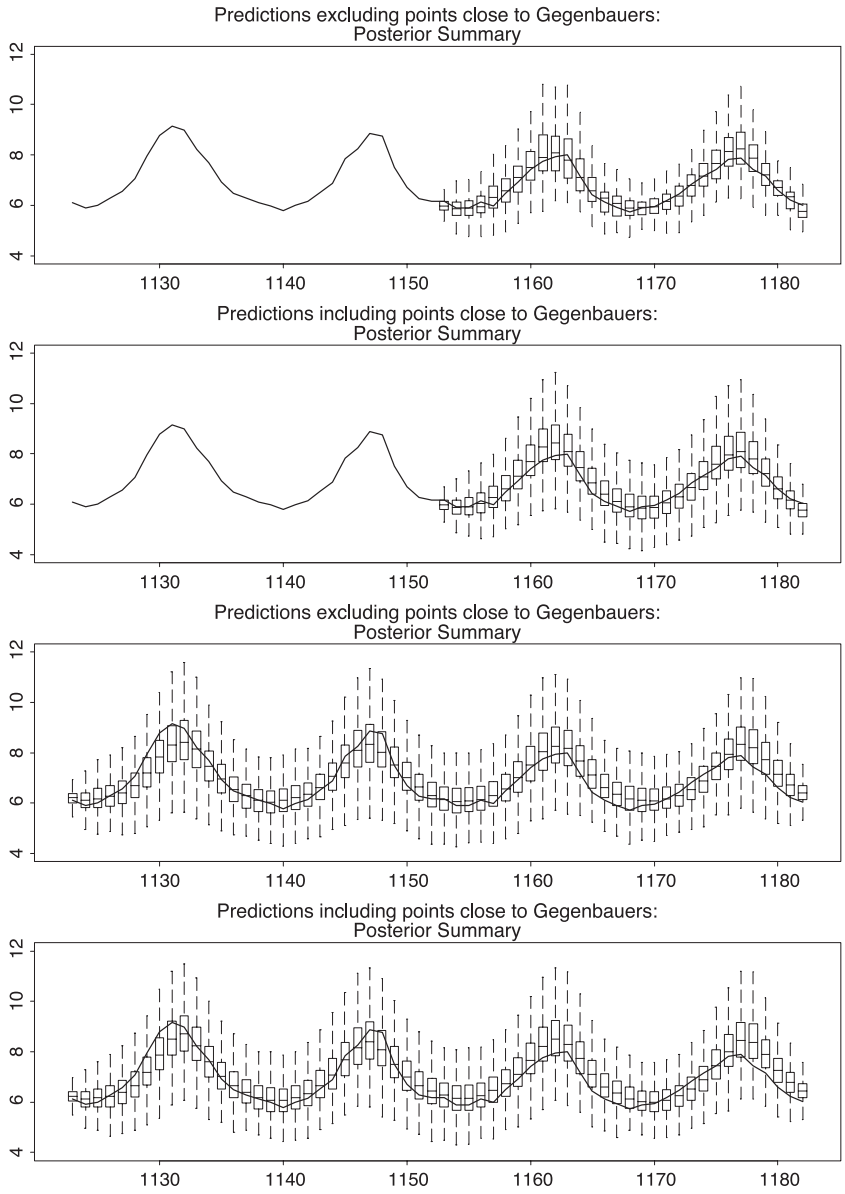

Fig. 4. *S. Carinæ* data: predictions for 30 and 60 steps ahead, with data points near spectral poles removed (top in pair) and not removed (bottom). Boxplots of medians and 5%, 25%, 75%, and 95% quantiles of the posterior predictive distributions; the solid line shows the actual values.

ordinates near to the spectral GARMA poles, the MCMC algorithm was used to compute the posterior and predictive quantities of interest in both cases. Specifically, we analyse the data including all values of the periodogram and additionally, by excluding those points that lie within $1/N$ of the Gegenbauer poles, as this corresponds to the width of Fejer's kernel associated with the periodogram. We note that the posterior inference is quite sensitive to the method used to treat these periodogram values, but that prediction is generally robust. The height of the sidelobes of Fejer's kernel actually implies that, for high values of the persistence parameter, the bias can extend to ordinates even further than $1/N$ away from the pole (see Olhede et al., 2003), and a full methodology for determining criteria for excluding points, and indeed, the full implications of the choice of methodology, is yet to be developed, and is an area of current research.

Fig. 2 shows 100 samples from the posterior distribution of the theoretical spectrum, as well as the posterior spectral mean on the periodogram of the data. The peaks in the spectrum are associated with the periodic behaviour present in the data. When points near to spectral poles are included, a periodicity near frequency 0.025 is detected with relatively high posterior probability, whereas that GARMA pole is

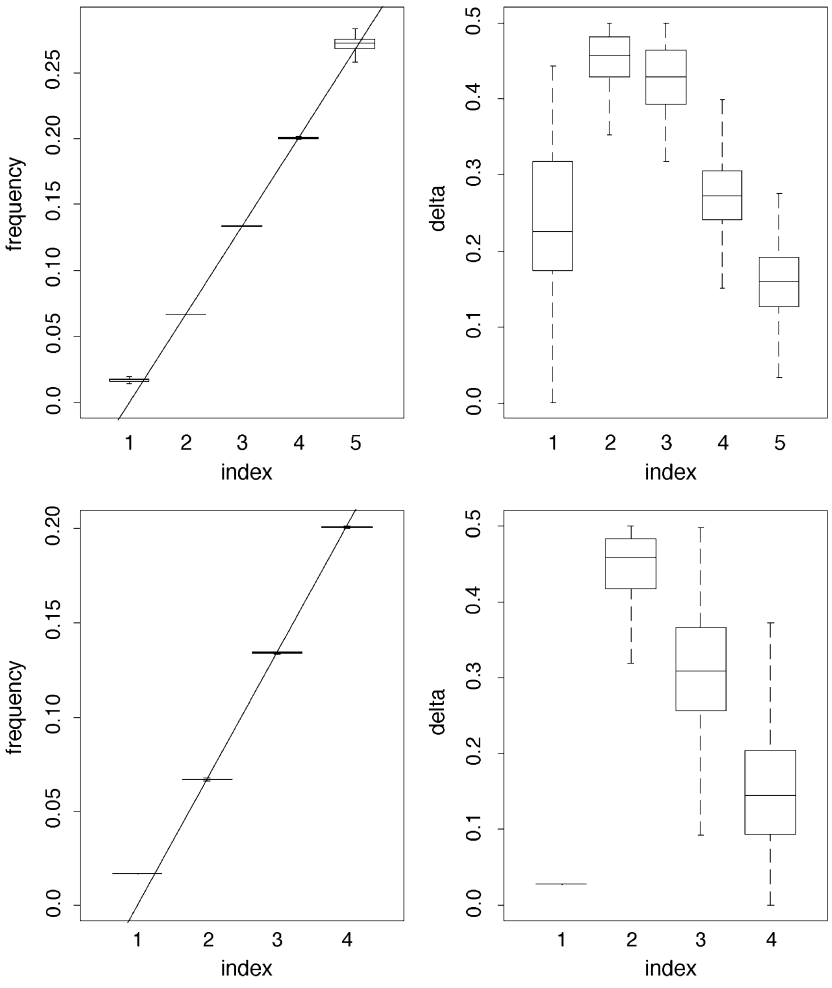

Fig. 5. *S. Carinæ* data: positions of Gegenbauer frequencies and the associated values δ, with data points near spectral poles not removed (top row) and removed (bottom row).

only rarely included in the MCMC scheme when those ordinates are excluded.

The posterior distributions of model order are shown in Fig. 3; we see that the range for Gegenbauer, AR and MA orders vary according to whether the points near spectral poles are excluded or not. In both cases, however, the ARMA model order is considerably lower than that reported in Huerta and West's, where the fit of a pure AR model favours model orders in the range 20–25, implying that the introduction of MA parameters and Gegenbauer components can provide different, and potentially more parsimonious representation.

Fig. 4 shows summaries of the posterior distributions for 30 and 60 step-ahead predictions, and demonstrate the ability to follow the periodicities inherent in the data. The estimation procedure of missing values and predictions is outline fully in Section 4. Due to space limitations, we do not include the posterior summaries for the missing values here, but note that the imputation of missing values is more robust to algorithmic specification than predictions.

The Gegenbauer components identify the periodic behaviour of the data. Fig. 5 shows boxplots of δ and f associated with the ordered Gegenbauer components. In the left-hand column of plots, the higher frequencies poles lie very close to the diagonal lines which give the locations of exact multiples of the dominant component, and, as noted in Huerta and West, very likely correspond to harmonics of that component. The dominant component, with posterior frequency mean of $f = 0.0668$, corresponds to a period of 149.7, in agreement with previous analyses. As in Huerta and West, our analysis also identifies the very low frequency periodic component, this is identified either with a Gegenbauer component (posterior probability 0.29) or alternatively with a complex AR root with an overall mean period of about 570 days.

A residual analysis showed that there is evidence of slightly heavier tails than normality, this is in agreement with Huerta and West. The time domain correlogram does not give any indication of residual correlation evident in the innovations of Huerta and West, possibly due to the fact that we have replaced their unit root with a Gegenbauer component, and this is better able to model pseudo-cyclical behaviour. The frequency domain residuals display no evidence of either departure from the theoretical distribution (as described in Section 5) or serial correlation.

7. Conclusions

By introducing MA and Gegenbauer components to the AR time series representation, we can model periodicities and high frequency behaviour in a parsimonious manner. By shifting inference to the frequency domain, the MCMC scheme we propose is computationally feasible and provides the natural setting in which to identify periodicities. The methodology could easily be extended to incorporate forecasting by following the same procedure outlined for the imputation of missing values.

The two most important features of the analysis presented are, first, that the MCMC algorithm is largely automatic in terms of implementation, and secondly, that the MCMC algorithm facilitates time domain predictions and model validation. Although we have used a variable dimension MCMC approach, we have concentrated on reconstructions of the spectral density, and thus difficulties associated with model comparisons for models of different order are avoided; we reconstruct the spectral density using a mixture of models having different GARMA order, but concentrate on a posterior functional that has interpretation across all models. Our reconstruction method is akin to wavelet and other basis function reconstructions, implemented in a fully Bayesian framework.

We note the fact that our analysis is based on the asymptotic distribution of the periodogram and that, while the periodogram is asymptotically unbiased, this does not necessarily mean that its bias is small for any particular sample size. Hausser (1999) provides empirical evidence that tapered versions of the Whittle likelihood provide more reliable estimators for ARMA and ARFIMA models. We believe that while this may be crucial if the aim of the investigation is purely determination of the parameters, it is less important if the aim is prediction. However, there are examples where the periodogram fails to detect multiple periodicities, or where bias can exist in the estimate of the frequency of a single periodicity (see Percival & Walden, 1993, p. 486), it has been shown (Chen, Wu, & Dahlhaus, 2000) that both these problems can be alleviated with the use of tapers. Previous

determination of periodicities from the periodogram has implicitly implied that any spectral lines occur at Fourier frequencies, this may be problematic for small sample sizes where the Fourier frequencies may not necessarily be close to any true spectral lines. Our methodology avoids this problem as Fourier frequencies close to spectral lines are omitted from the likelihood. Potential solutions to these problems are the aims of current research.

References

Anderson, T. W. (1971). *The statistical analysis of time series*. New York: Wiley.

Arteche, J., & Robinson, P. M. (2000). Semiparametric inference in seasonal and cyclical long memory processes. *Journal of Time Series Analysis*, *21*(1), 1–25.

Baldwin, A. J., & Thomson, P. J. (1978). Periodogram analysis of S. Carinæ, variable star section, royal astronomical. *Society of New Zealand*, *6*, 31–38.

Barnett, G., Kohn, R., & Sheather, S. (1996). Bayesian estimation of an autoregressive model using Markov chain Monte Carlo. *Journal of Econometrics*, *74*, 237–254.

Beran, J. (1994). *Statistics for long-memory processes*. London: Chapman and Hall.

Bisaglia, L., & Guegan, D. (1998). A comparison of techniques of estimation in long-memory processes. *Computational Statistics & Data Analysis*, *27*, 61–81.

Carter, C. K., & Kohn, R. (1997). Semiparametric Bayesian inference for time series with mixed spectra. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *59*(1), 255–268.

Chen, Z. -G., Wu, K. H., & Dahlhaus, R. (2000). Hidden frequency estimation with data tapers. *Journal of Time Series Analysis*, *21*(2), 113–142.

Ferrara, L., & Guegan, D. (2001). Forecasting with *k*-factor Gegenbauer processes: Theory and applications. *Journal of Forecasting*, *20*(8), 581–601.

Giraitis, L., Hidalgo, J., & Robinson, P. M. (2001). Gaussian estimation of parametric spectral density with unknown pole. *Annals of Statistics*, *29*(4), 987–1023.

Giraitis, L., & Leipus, R. (1995). A generalized fractionally differencing approach in long-memory modelling. *Lithuanian Mathematical Journal*, *35*, 65–81.

Granger, C. W. J., & Joyeux, R. (1980). An introduction to long-memory time series and fractional differencing. *Journal of Time Series Analysis*, *1*, 15–30.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*(4), 711–732.

Hassler, U. (1994). (Mis)specification of long-memory in seasonal time series. *Journal of Time Series Analysis*, *15*, 19–30.

Hauser, M. A. (1999). Maximum likelihood estimators for ARMA and ARFIMA models: A Monte Carlo study. *Journal of Statistical Planning and Inference*, *80*(1–2), 229–255.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, *68*, 165–176.

Huerta, G., & West, M. (1999a). Priors and component structures in autoregressive time series models. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *61*(4), 881–899.

Huerta, G., & West, M. (1999b). Bayesian inference on periodicities and component spectral structure in time series. *Journal of Time Series Analysis*, *20*, 401–416.

Hurvich, C. M., & Beltrao, K. I. (1993). Asymptotics for the low-frequency ordinates of the periodogram of a long memory time series. *Journal of Time Series Analysis*, *14*, 455–472.

Mallick, B. K., Denison, D. G. T., & Gangopadhyay, A. (2002). A Bayesian curve fitting approach to power spectrum estimation. *Journal of Nonparametric Statistics*, *14*(1–2), 141–153.

McCulloch, R. E., & Tsay, R. S. (1994). Bayesian analysis of autoregressive time series via Gibbs sampling. *Journal of Time Series Analysis*, *15*, 235–250.

Montanari, A., Taqqu, M. S., & Teverovsky, V. (1999). Estimating long-range dependence in the presence of periodicity: An empirical study. *Mathematical and Computer Modelling*, *29*(10–12), 217–228.

Olhede, S. C., McCoy, E. J., Stephens, D. A., (2003). On the distribution of the periodogram of seasonally persistant processes, To appear in *Biometrika*.

Ooms, M. (1995). Flexible seasonal long-memory and economic time series, Discussion paper 9515/A, Econometric Institute, Erasmus University, Rotterdam.

Ooms, M., & Franses, P. H. (2001). A seasonal periodic long-memory model for monthly river flows. *Environmental Modelling and Software*, *16*, 559–569.

Percival, D. B. (1992). Simulating Gaussian random processes with specified spectra. *Computing Science and Statistics*, *24*, 534–538.

Percival, D. B., & Walden, A. T. (1993). *Spectral analysis for physical applications*. Cambridge, UK: Cambridge University Press.

Porter-Hudak, S. (1990). An application of seasonal fractionally difference model to the monetary aggregates. *Journal of the American Statistical Association*, *85*, 338–344.

Robert, C. P., & Casella, G. (1999). *Monte Carlo statistical methods*. New York: Springer-Verlag.

Robinson, P. M. (1995a). Log-periodogram regression of time series with long-range dependence. *Annals of Statistics*, *23*, 1048–1072.

Robinson, P. M. (1995b). Gaussian semiparametric estimation of long-range dependence. *Annals of Statistics*, *23*, 1630–1661.

Shephard, N., & Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, *84*(3), 653–667.

Taqqu, M. S., & Teverovsky, V. (1997). Robustness of Whittle-type estimators for time series with long-range dependence. *Stochastic Models*, *13*, 723–757.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, *22*(4), 1701–1762.

Walker, A. M. (2000). Some results concerning the asymptotic distribution of sample Fourier transforms and periodograms for a discrete-time stationary process with a continuous spectrum. *Journal of Time Series Analysis*, *21*(1), 95–109.

Woodward, W. A., Cheng, Q. C., & Gray, H. L. (1998). A *k* factor GARMA long-memory model. *Journal of Time Series Analysis*, *19*(4), 485–504.

Yajima, Y. (1989). A central limit theorem of Fourier transforms of strongly dependent stationary processes. *Journal of Time Series Analysis*, *10*, 375–383.