Time series decomposition and analysis in a study of oxygen isotope records

MIKE WEST
Institute of Statistics & Decision Sciences,
Duke University, Durham NC 27708-0251

October 1995

Research partially supported by NSF grant DMS-9311071.
Abstract. Some methodological developments in Bayesian time series analysis are discussed and illustrated in analyses of an oxygen isotope record of some geological interest. The developments include the implementation of Bayesian simulation analyses of auto-regressions, and state-space auto-regressions, with a particular focus on spectral characteristics; various such analyses are summarised. A useful time series decomposition result is developed and illustrated in connection with exploring latent cyclical components in the time series; though almost transparent and evidently very useful in practical time series analysis, this decomposition method does not seem to have surfaced significantly in the literature to date. Analyses incorporating additive time series outlier components are discussed, as are questions of time variation in time series parameters and structure.

Keywords: Auto-regressive time series; Bayesian computations; Quasi-periodic components; State space models; Time series outliers; Time-varying cycles
1 INTRODUCTION

In the study of geological time series of proxy records of variations in climatic conditions, identification of the forms of periodic and quasi-periodic components of recorded time series is of key interest (Park and Maasch 1993; West 1995b). This paper discusses some features of various analysis of a single deep ocean core oxygen isotope series, with a view to exploring key quasi-periodic components using more or less standard models. These include harmonic regression, auto-regressive models, and state-space auto-regressions. In connection with exploration of the latent sub-structure of a time series, and, in particular, to explore quasi-cyclical sub-components of the isotope series, a simple result on the decomposition of time series is developed and illustrated. Though evidently rather interesting and useful in isolating sub-components of a series, this result seems new to the literature. It is developed from state space representations of certain classes of dynamic linear models, and applies particularly to the important special case of state-space auto-regressions, i.e. auto-regressive processes observed with additive noise, hence includes traditional, measurement error free auto-regressions. This is developed and explored in analyses of the isotope series, as are questions of additive outlier modelling in the state space context. Throughout, Bayesian analyses are summarised by representing simulations from posterior distributions for model parameters in various ways, including direct posterior simulation, where feasible, and using iterative methods otherwise. Though by no means standard in the time series literature, simulation methods are becoming established in the Bayesian time series arena, and here we illustrate some of their uses and benefits.

Geological time variations in oxygen, and other, isotope measurements from deep ocean cores relate to patterns of variation in global ice volume and ocean temperature (Shackleton and Hall 1989; Park and Maasch 1993). The single series graphed in Figure 1 is representative of several oxygen isotope series from cores from various geographical locations, and is explored here in various ways. This data derives from original $\delta^{18}O$ Site 677 measurements presented in Shackleton and Hall (1989) and was provided by J Park of Yale University. The values estimate relative abundance of $\delta^{18}O$ and are timed on an equally spaced, 3kyear scale. The time scale calibration is based on that of Ruddiman, McIntyre and Raymo (1989), and discussed in Park (1992), and Park and Maasch (1993). This latter reference also discusses the process of interpolation of original, unequally spaced measurements to this equally spaced scale. This series stretches back roughly 2.5 million year, and is plotted with a reverse of sign, by convention, so that the apparent increase in levels in modern times reflects generally warmer average global temperatures and smaller average ice masses. Time-varying periodicities are evident; the nature and structure of theses periodicities of some geological importance, and one focus for our analyses here.

Useful discussion of the relationships between climatic indicators, such as the $\delta^{18}O$ measures
here, and cyclical patterns of changes in the Earth’s orbital dynamics appears in Park (1992). The precession and obliquity of the Earth’s orbit impact on solar insolation received by the Earth and cause substantial variations in climate characteristics as a result. Periodicities in eccentricity are generally believed to be associated with periods around 96kyears, 126kyears and, perhaps more speculatively, longer periods of around 413 years; each of these figures is subjective to uncertainty. Shorter term cycles include that associated with precession, with period around 19-23kyears, and obliquity of the Earth’s orbit, with period around 41kyears. The 100kyear “ice-age cycle”, so-called, is of major interest and has been the subject of intensive investigation in recent years. Identifying the nature and structure of quasi-periodic components of period around 100kyears is of importance in contributing to debates over the genesis of the ice-age cycles, roughly a million years ago, and, particularly, to questions of whether or not the onset was gradual and inherent or the result of a significant structural climatic change. (e.g. Ruddiman et al 1989; Park 1992, and references therein).

The following section begins an exploration of the δ\(^{18}\)O series and highlights some of the key spectral features. This is followed by further, model-based exploration, with graphical investigations of quasi-periodicities based on a novel time series decomposition method. Further analysis accounts for corrupting measurement errors and additive outliers in the series, and the paper concludes with summary comments and discussion.

2 INITIAL DATA ANALYSES

Initial data exploration uses Bayesian periodogram analysis, following Bretthorst (1988) and as used in West (1995a and b), for example. After possibly detrending the original data series is some way, this involves a simple reference Bayesian analysis of a single sinusoid regression model

\[
x_t = a \cos(\omega t) + b \sin(\omega t) + \epsilon_t\]

under the usual i.i.d. normal assumption for the error terms, involving a constant error variance \(v\). Integrating the likelihood function with respect to the usual reference prior \(p(a, b, v) \propto v^{-1}\), leads to a marginal likelihood function for the frequency \(\omega\) which, when evaluated at the usual Fourier frequencies, is closely related to the traditional periodogram. This function may, however, be evaluated across a much finer range of frequencies, and is then termed a Bayesian periodogram. In practice the grid of values for evaluation is important. Evaluation at only the Fourier frequencies can, and often does, “miss” frequency ranges with high likelihood that will show up in the Bayesian periodogram evaluated across a finer range. Hence, in preference to traditional periodogram displays, Bayesians and others should refer to displays of this likelihood function across a suitably fine grid. We do this here, displaying the function on natural log scale, and plotting as a function of the natural wavelength parameter \(\lambda = 2\pi/\omega\).

Figure 2 provides plots of the Bayesian periodograms for four cases: (a) based on the original

2
data series; (b) based on a series of residuals after subtracting the least squares straight line fit to the raw data; (c) based on a series as in (b), but now the fitted smooth trend is based on the lowess smoother as implemented in S-Plus, \( \text{lowess}(. , f = 0.2) \); and (d) based on the first differences of the oxygen series. In each case, the spike at around 41kyears is evident, suggesting a sustained and persistent cyclical feature of that period and corresponding to the period of the obliquity cycle of the Earth’s orbit. In the first three frames, there are broadband features evident around periods of 90-100, 110 and 120-130kyears, mapping directly to the approximate periods of the orbital forcing mechanisms and the ice-age cycle. In terms of the basic qualitative features of these figures, the two versions of detrending in (b) and (c) do not obviously distort the display based on the original series. In the fourth figure, the relatively deleterious effects of differencing in masking the lower frequency characteristics are apparent.

Further analyses are vested in AR models and their variants. Here a specific auto-regressive model is analysed in some detail, the oxygen series analysed having been detrended, as above, via lowess. On the basis of exploring ARMA models of various orders, \( AR(p) \) models with \( p \) around 16-20 are suggested. This is supported by traditional AIC measures of fit, as computed in S-Plus, for example, and related Bayes’ factor calculations in reference conditional analyses. Here \( p = 20 \) is chosen for illustration. An \( AR(p) \) model is capable of exhibiting quasi-cyclical behaviour at various distinct frequencies; in fitting such models to real data, values of \( p \) required to adequately capture periodicities up to a maximum period of, say, \( \lambda \) time units are of the order of \( \lambda/2 \), or more. Thus \( p \) around 16 – 20 will in principle “find” periods up to 32 – 40; on the 3kyear time scale of the oxygen series, this corresponds to periods of up to 120 – 130kyears, as we expect to be evidenced based on the periodogram explorations.

A reference conditional analysis is summarised here, and used as the basis for an initial time series decomposition in the next section. Conditioning on the first \( p \) values of the series, the \( AR(p) \) model is fitted as a reference linear regression \( x_t = \sum_{j=1}^{p} \phi_j x_{t-j} + \epsilon_t, \ (t = p + 1, \ldots, n) \), with assumedly independent innovations distributed as \( N(\epsilon_t|0, v) \). In vector form, \( x_t = f_t^\prime \phi + \epsilon_t \), with \( AR \) parameter vector \( \phi = (\phi_1, \ldots, \phi_p)^\prime \) and regressor \( f_t^\prime = (x_{t-1}, x_{t-2}, \ldots, x_{t-p})^\prime \). Stacking these \( n - p \) equations gives \( x = F\phi + \epsilon \) in an obvious notation. The usual reference prior \( p(\phi, v) \propto v^{-1} \) gives the posterior \( p(\phi|v, x) = N(\phi|\hat{\phi}, vA) \) and \( p(s/v|x) = \chi^2_{n-2p}(s/v) \) where \( A = (F'F)^{-1} \), \( \hat{\phi} = AF'x \) and \( s = (x - F'\hat{\phi})^\prime(x - F'\hat{\phi}) \). Note that \( \hat{\phi} \) is the reference posterior mode and mean, and the maximum likelihood estimate of \( \phi \) conditional on the first \( p \) observations. For the oxygen data analysis here, \( p = 20 \) and \( n = 866 \).

This model reasonably represents the oxygen series, in spite of the apparent change in form in the central part of the time interval of measurement; this point is revisited in further analyses in section 3. Residual analyses indicate a small fraction of residuals larger than consistent with normality in the tails, perhaps suggestive of a heavy-tailed structure, though the departure from
approximate normality is not dramatic; we proceed with further exploration of this model analysis here, reserving further comments on the error structure and its impact to section 3.

Of primary interest is the nature and form of the periodic features of the series, and so we consider some inferential implications of the fitted model in respect of such features. This exploration, below, also illustrates simple and now routine posterior inferences based on Monte Carlo posterior simulations.

Any \( \phi \) vector determines the auto-regressive characteristic polynomial \( \phi(u) = 0 \) where \( \phi(u) = 1 - \phi_1 u - \ldots - \phi_p u^p \); denote the reciprocal roots of this equation by \( \alpha_1, \ldots, \alpha_p \). In the analysis, the reciprocal roots corresponding to \( \phi = \hat{\phi} \) all lie inside the unit circle, the maximum modulus being 0.96. This supports stationarity. A Monte Carlo sample of 10,000 independent draws from the posterior \( p(\phi, v|x) \) was made, and the reciprocal roots computed for each sampled vector. In every case all 20 moduli lay inside the unit circle, so that the Monte Carlo estimate of the posterior probability of stationarity is unity. We know, of course, that this probability of non-stationarity is non-zero, as the posterior for \( \phi \) is unconstrained; the posterior simulation indicates that it is very small. Complex roots occur in conjugate pairs and correspond to damped, quasi-periodic behaviour in the series modelled. A root pair \( r e^{i \omega} \) has period \( \lambda = 2\pi/\omega \); low values of \( \lambda \) usually represent high frequency correlated noise characteristics rather than more sustained, though damped, periodicities, and this is supported by low moduli \( r \) in many cases. Of the 10,000 posterior samples, in 55.4% of cases all roots are complex, and in 44.4% of cases all but two are complex. We can explore inferences about the dominant roots by ordering each set of 20 roots according to either moduli or wavelengths; we do this in terms of wavelengths here. The two longer wavelengths are apparently quite widely separate, so this ordering has little impact on them directly, while their moduli are apparently similar, and correspond with the highest values in any case. Thus the two components of higher wavelengths are also the most persistent or sustained over time, the damping factors being closest to unity. At \( \hat{\phi} \), the first three roots, in order, have periods, after multiplying by 3 to move to the kyear scale from the 3kyear scale of the data, of 111.6, 41.3 and 22.8 kyears, with corresponding moduli 0.96, 0.96, and 0.93, respectively. The approximate posterior mean of \( \phi \), computed by averaging the Monte Carlo sample of 10,000 draws, produces values of 11.6, 41.3 and 22.8 for the periods, and 0.96, 0.96 and 0.93, respectively, for the moduli; so, to this level of accuracy, the posterior estimates are in very close agreement. Figures 3 and 4 graph density estimates of the bivariate posterior margins for first two the periods and moduli separately. Note that the period of the first, larger component is rather uncertain relative to the second; the latter is quite precisely isolated to the 40 – 42kyear range, spanning the traditional and “nominal” 41kyears of the forcing obliquity cycle in the Earth’s orbit. Other, less apparent patterns driven by Earth-orbital dynamics include periods of around 96kyears, reflected in the earlier Bayesian periodogram analyses. These do not show up directly in this analysis, the longer
periods around 110 ky years dominating at the higher levels. This may be due, in part, by masking issues that arise in fitting time domain parametric models in cases where data exhibit quasi-periodic features with perhaps distinct but close periods. Note, however, that the subsidiary root structure may provide alternative explanation. At \( \hat{\phi} \) the third to sixth complex roots, in order of decreasing moduli, have period/moduli pairs of approximately 22.8/0.93, 12.4/0.88, 16.5/0.88 and 8.5/0.88, respectively. Notice that these are all close to harmonics of a wavelength in the 96-98kyr range; the moduli are also reasonably large. Hence a damped component, periodic of period around 96−98kyr but quite non-sinusoidal, might be evidenced here through the superposition of these higher frequency terms.

3 TIME SERIES DECOMPOSITION

Having identified some key features of the fitted model, related to the main patterns of quasi-cyclical behaviour in the series assuming the model is adequate, it is of interest to explore just how the separate features contribute to the overall data description and how they vary in time. Questions of geological interest focus on the nature and form of time-variation in the dominant cyclical components. To explore this, graphically and visually here, we make use of a very useful time series decomposition that arises from the state space representation of the time series model. This result is not specific to AR models, such as used here initially, and so we explain the construction in a more general context of a class of dynamic linear models (West and Harrison 1989).

Consider the dynamic linear model, or state space model, given by

\[
\begin{align*}
x_t & = F'z_t + \nu_t, \\
z_t & = Gz_{t-1} + \omega_t
\end{align*}
\]  

with the following components and assumptions: \( z_t = (z_{t,1}, \ldots, z_{t,p})' \), the \( p \times 1 \) state vector at time \( t \); \( F \) is a column \( p \)-vector of known constants or regressors, fixed and constant for all \( t \); \( \nu_t \) is the observation noise, distributed as \( N(\nu_t|0,v_t) \) for known \( v_t \); \( G \) is a known \( p \times p \) matrix, the state evolution matrix, fixed and constant for all \( t \); \( \omega_t \) is the state evolution noise, or innovation, at time \( t \), distributed as \( N(\omega_t|0,W_t) \) for a known variance matrix \( W_t \); and, finally, the noise sequences \( \nu_s \) and \( \omega_t \) are independent and mutually independent.

While more general dynamic models have time-varying \( F \) and \( G \) elements, the case of constant elements is quite far ranging. Of interest here is the usual AR\((p)\) process which, in the notation of section 1 above, is an important special case with the following useful state space representation.
The additive observation errors are zero, \( \nu_t = 0 \) for all \( t \), and we have

\[
F = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad G = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_{p-1} & \phi_p \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & \cdots & 1 & 0 \end{pmatrix}, \quad \omega_t = \begin{pmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix};
\]

(2)

the form of the evolution error \( \omega_t \) implies that \( W_t = W \), for all \( t \), having entries all zero but for \( W_{1,1} = w > 0 \).

Note the trivial extension to time dependent innovations variances, \( w_t \) at time \( t \), or to models with purely additive measurement errors, i.e. non-degenerate \( \nu_t \) terms, in which case the AR process is latent, the observation process being subject to contamination; this extension is considered further below. Other important extensions add underlying trend terms, possibly non-stationary, to the AR process, extending the model via superposition (West and Harrison 1989, chapter 5). Such extensions are also used in further analysis below.

A useful decomposition of the series is obtained in models in which, as in this case, \( F \) is the vector of zeroes but for the leading one, and \( \nu_t = 0 \) for all \( t \) so that \( x_t \) is the first element of the state vector \( z_t \). In cases with non-degenerate observation errors \( \nu_t \), the same decomposition applies to the expected value of the series \( x_t - \nu_t \), and this extension is used below.

To begin, consider the conditional predictive mean \( E(x_{t+k}|z_t) = F^t G^k z_t \) as a function of the lead-time \( k \geq 0 \). Using standard theory (e.g. West and Harrison 1989, chapter 6) the form in \( k \) is simplified by exploiting the eigen-structure of \( G \). We restrict attention cases in which \( G \) has distinct, non-zero eigenvalues, such as is usual in the case of the AR model here and various other models; denote the eigenvalues by \( \alpha_1, \ldots, \alpha_p \).

Suppose that the \( p \) eigenvalues occur as \( q \) pairs of complex conjugate pairs and \( s = p - 2q \) real and distinct values; write the complex eigenvalues as \( r_j \exp(\pm i\omega_j) \) for \( j = 1, \ldots, q \), noting the non-zero arguments \( \omega_j \) correspond to periods \( \lambda_j = 2\pi/\omega_j \), and write the real values as simply \( r_j \) for \( j = 2q + 1, \ldots, p \). Then \( G = EAE^{-1} \) where \( A \) is the diagonal matrix of eigenvalues, in the arbitrary order specified, and \( E \) is the matrix whose columns are the corresponding normalised eigenvectors. Define \( p \)-vectors \( a = E^t F \) and \( b_t = E z_t \) for each \( t \).

Then

\[
E(x_{t+k}|z_t) = F^t E A^k E z_t = a^t A^k b_t = \sum_{j=1}^{p} c_{tj} \alpha_j^k
\]

where \( c_{tj} = a_j b_{tj} \). At \( k = 0 \), we know that \( E(x_t|z_t) = x_t \), so the above expression implies the direct decomposition

\[
x_t = \sum_{j=1}^{p} c_{tj}.
\]

(3)
Thus the entire time series is decomposed into the sum of time-varying components corresponding to the auto-regressive roots. The complex factors \(c_{tj}\) are easily computable given any specified \(G\) matrix. For real roots \(\alpha_j\), the corresponding \(c_{tj}\) values are real. A pair of complex conjugate roots leads to the corresponding two \(c_{tj}\) coefficients being complex conjugates, so that their sum will be real and computed as twice the common real part.

There are various ways of using this result to explore the derived components as functions over time \(t\) to examine their relative contributions to the decomposition of the series. In the special case of an \(AR\) model, any specified \(AR\) parameter vector \(\phi\) determines \(G\) and hence the decomposition. In this case the eigenvalues of \(G\) are precisely the reciprocal roots of the auto-regressive characteristic polynomial \(\phi(x) = 0\). So posterior inference about the latent subseries \(c_{tj}\) in the decomposition is derived from inference on \(\phi\). For example, given a sample of \(\phi\) vectors from a posterior distribution in a specific analysis of the series, the corresponding posterior samples of the \(c_{tj}\) terms may be directly computed and then averaged to produce approximate posterior means of the components. In addition or alternatively, direct overall estimates of the \(c_{tj}\) may be computed simply by plugging-in a posterior mean or modal estimate of \(\phi\).

To explore this in the earlier reference \(AR(20)\) analysis of the oxygen isotope series, first note that the reference posterior mean, or conditional MLE, of \(\phi\) corresponds to a full set of \(q=10\) complex roots, no real roots, and so the estimated decomposition at \(\phi = \hat{\phi}\) is the sum of the corresponding 10 real components. Due to the arbitrary labelling of the roots, we must impose identification somehow; as before, we can do this by simply ordering components according to the estimated periods or moduli of the corresponding roots. These calculations are represented in Figure 5. As we have seen above, the first three most persistent roots of the \(AR\) polynomial evaluation at posterior estimates of \(\phi\) have arguments corresponding to the periods of 111.6, 41.3 and 22.8 kyears, with corresponding moduli 0.96, 0.96, and 0.93, respectively. On this basis, the subseries with damped cyclical behaviour of periods around 110 and 41 kyears appear comparable in persistence, the shorter period components are less persistent. The figure displays part of the corresponding data decomposition, plotting the three, longer period components only; this is based on evaluation at the mode \(\hat{\phi}\) and is graphically indistinguishable from that based on the approximate posterior mean. The 7 subsidiary components have progressively smaller estimated periods and, generally, decreasing moduli; the time plots of these components confirm that they are negligible in amplitudes by comparison with those displayed. These higher order components represent increasingly negligible features of short-term correlation in the data.

This decomposition is relevant in the geological context of this data series. From the figure we see clearly that the component of 110kyears or so in period has a form that changes significantly somewhere back in time, probably between 1.2Myears and 1.3Myears. The other components, in contrast, have consistent forms across the time scale. In addition, the estimated innovations
sequence $\epsilon_t$ has a consistent form over the full stretch of the data, with no clear indication of changes in variance. These facts suggest that the clear change in form of the original data is due to essentially entirely to a change in the nature of the major, longer term oscillations of period between 100-120 kyears. This ties in with the onset of the ice-age cycle of period in this range, and the visual impression gained from the graph of the decomposition is that the ice-age cycle does indeed change in form in the region of 1.2 – 1.3 million years ago. Notice that the model neither accommodates nor predicts real structural change in the series, assuming a fixed auto-regression over the entire time period; nevertheless, the changes in one, and essentially only one, of the key quasi-cyclical components are evident from this decomposition and graphical display.

4 FURTHER DATA ANALYSIS

The above analysis makes no allowance for the possibilities of routine measurement, sampling and laboratory errors contaminating the oxygen recordings, nor for the almost inevitable occurrence of occasional gross errors, or outliers. The kinds of inferences just explored can be expected to be reasonably insensitive to very low levels of purely additive measurement errors, particularly with such a long series in which the major components have periods much longer than the sampling interval. However, observational errors of meaningful magnitudes induce an errors-in-variables structure with the consequence that estimation of the auto-regressive parameters is subject to damping, i.e. the absolute values of elements of $\hat{\phi}$ tend to be erroneously shrunk towards zero. Furthermore, the occurrence of occasional larger additive errors, either singly or in batches, can much more dramatically impact the model analysis and subsequent inferences (Kleiner, Martin Thompson 1979; West 1995c). For these reasons, and taking a lead from the suggestion of somewhat heavy-tailed residual deviations in the previous analysis, mentioned above, further analysis extends the model to include both routine measurement error and outlier components. This is done via a traditional normal mixture model, as follows.

Assuming independent measurement errors $\nu_t$ corrupting the underlying $\delta^{18}$O levels $x_t$, the previous $AR$ model is modified to the full state space auto-regression, mentioned earlier. We then have observations $x_t$ given by equations (1) and (2), where the errors $\nu_t$ are now non-zero, and the leading element of $z_t$ represents the value of the latent auto-regression at time $t$. Underlying trend in the series may also be modelled as a latent component, simply added to the latent $AR$ process via superposition (West and Harrison 1989). Further analysis of the oxygen isotope series adopts this approach with a simple, first-order polynomial trend, represented by $\mu_t = \mu_{t-1} + \eta_t$ over time $t$ where the changes $\eta_t$ are assumedly independent $N(\eta_t|0, u)$. The state space model now looks like
the form of the evolution error \( \omega_t \) implies that the evolution variance matrix \( V(\omega_t) = W \) has zero entries apart from \( W_{1,1} = u > 0 \), the variance of the level changes \( \eta_t \), and \( W_{2,2} = w > 0 \), the variance of the AR innovations. Now the state vector is \( z_t = (\mu_t, x_t^*, x_{t-1}^*, \ldots, x_{t-p+1}^*)' \) where \( \{x_t^*\} \) is the unobserved AR\((p)\) process.

One traditional measurement error model adopts a normal mixture distribution of the form \( \nu_t \sim (1 - \pi)N(\nu_t|0, v) + \pi N(\nu_t|0, k^2v) \); this admits a background level of routine measurement error, with variance \( v \), together with occasional extremes, or outliers, generated by the inflated variance component with \( k > 1 \) with some small probability \( \pi \) (e.g. West and Harrison 1989, chapter 12). Though these kinds of models have been around for some time, it is only through recent developments in simulation methods of Bayesian analysis that we can begin to explore them fully in practice; indeed, even the basic normal model with \( \pi = 0 \) is intractable as the AR process \( x_t^* \) is unobserved. Various authors have been exploring simulation methods in time series analysis quite recently. Key references include Carter and Kohn (1994), Frühwirth-Schnatter (1994), McCulloch and Tsay (1994), and some applications appear in West (1995a,b and c). The first two references develop basic simulation methods for state space models, and are partly motivated by interest in mixture models for error distributions. Direct development of the algorithms in these papers, and with extensions of further developments in West (1995a), lead to easily implemented algorithms for full posterior simulations in models (1) and (2) subject to mixture error distributions. Full details are omitted, though a sketch of the algorithm used here is given in an appendix, and the interested reader is referred to the above papers. Using these methods, it is possible to sample the full joint posterior distribution for the entire latent state variable \( z_t \) over all times of observations, the AR model parameters \( \phi \) appearing in the evolution matrix \( G \), the variance components \( v, u \) and \( w \), and the mixture parameters \( (k, \pi) \), though, in many cases, exploratory analysis is carried out with the latter two parameters fixed, simply investigating the effects of moving a little away from the baseline normal measurement error model. Posterior samples of the elements of \( z_t \) translate directly into posterior samples of the underlying smooth trend \( \mu_t \) together with samples of the latent AR\((p)\) process \( x_t^* \), over all \( t \).

The earlier auto-regressive model for the \( \delta^{18}O \) series is now extended this way, using an outlier model with \( k = 10 \) and \( \pi = 0.05 \). As above, the traditional reference prior for \( (\phi, w) \) is adopted.
The prior for the measurement error variance $v$ is based on a finite range uniform prior for the standard deviation. Based on a posterior sample of 5,000 draws, we compute the approximate posterior mean of $\phi$ and use this value to evaluate the decomposition (3), now for the latent series $x^*_t$. Parts of the decomposition appear in Figure 6; here, in addition to the trend and key quasi-cyclical components, we display approximate posterior means of the errors $\nu_t$ over time. There are several marked outliers apparent, though the level of baseline measurement error appears very small, almost negligible on the scale of the raw data. This is reflected in the approximate posterior inferences for variance components; from the Monte Carlo sample, the posterior quartiles for $\sqrt{v}$ are approximately $(0.086, 0.092, 0.098)$, those for $\sqrt{w}$ are $(0.172, 0.187, 0.211)$, and those for $\sqrt{u}$ are $(0.008, 0.009, 0.011)$. A small fraction of the measurement errors are very large, essential outliers, but their impact on inferences is apparently slight; the time trajectories of the two key components are quite similar to those in Figure 5. This graphical impression is confirmed in examination of numerical summaries of the posterior samples for the periods and moduli of the latent $AR$ components.

These kinds of analyses, though by no means standard, are now accessible computationally and should become routine. Though inferences in this particular case appear relatively unaffected by the low level of background measurement error, other cases, with smaller sample sizes and with higher levels of observation error, can be more severely impacted (West 1995c).

Further analyses adding mixture/outlier models for the innovations $\epsilon_t$ can be similarly developed. More importantly in the scientific context of this, and related, isotope records is the question of more incisive inferences on the changing structure of the key latent component representing the ice-age cycle. Simple exploratory methods can provide direction. For example, Figure 7 displays the results of exploratory Bayesian spectral analysis, as illustrated in section 2, based on simply splitting the data into two and analysing the two shorter series separately. A central section of the series was ignored here, in the hope that the change in structure is clearly isolated. In the figure it is clear that the major, longer terms periodicities are much less apparent in the earlier section of the isotope record, the lower frame (b). The sustained nature of the 41kyear cycle, identified in the decompositions in Figure 5 and 6, is also highlighted. Further analyses of the two sub-series using the state space auto-regressions confirm that the longer term periodicities are simply not evident in the older segment, whereas inferences about component structure and periodicities based on the more recent data are closely comparable with the results already described based on the entire series. Furthermore, though posterior inferences about the auto-regressive coefficients differ between the two subseries, inferences about the variance components, particularly the key variance $w$ of the auto-regressive innovations series, are comparable; thus the change in structure is essentially determined by changes in $\phi$ or some function of $\phi$ related more directly to the quasi-cyclical components. Further analysis with models permitting time-varying auto-regressive coefficients, by
extending the state space forms (1), (4) to include time-variation in \( G \), may be useful in developing more refined inferences about this changing structure, though this is beyond the current scope and will be reported elsewhere.

**APPENDIX**

An outline of the posterior sampling algorithm for state space auto-regressions is given here. The model is a specific example of a wider class in which elements of the evolution matrix \( G \) and the variance components \( v \) and \( W \) are treated as uncertain parameters to be estimated along with the sequence of state vectors over time. Posterior simulation is iterative, using extensions of basic Gibbs sampling for state space models. The techniques build on basic DLM model theory, following West and Harrison (1989, chapter 4), and early simulation work of Carter and Kohn (1994), and Frühwirth-Schnatter (1994), with extensions to include uncertain parameters that stem from work in West (1995a) in a related context.

For concreteness, we focus on the specific model superposing a smooth, first-order polynomial trend \( \{ \mu_t \} \) with the latent AR\((p)\) process \( \{ x_t^* \} \), and subject to additive measurement errors \( \{ \nu_t \} \).

With state vector 
\[
z_t = (\mu_t, x_t^*, x_{t-1}^*, \ldots, x_{t-p+1}^*)',
\]
the model for observations \( \{ x_t \} \) is as earlier described, namely
\[
x_t = F'z_t + \nu_t, \quad \text{and} \quad z_t = Gz_{t-1} + \omega_t
\]
with
\[
F = \begin{pmatrix}
1 \\
1 \\
0 \\
\vdots \\
0
\end{pmatrix}, \quad G = \begin{pmatrix}
1 & \phi_1 & \phi_2 & \phi_3 & \ldots & \phi_{p-1} & \phi_p \\
0 & 1 & 0 & 0 & \ldots & 0 & 0 \\
0 & 0 & 1 & 0 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \ldots & 1 & 0
\end{pmatrix}, \quad \omega_t = \begin{pmatrix}
\eta \\
\epsilon_t \\
0 \\
\vdots \\
0
\end{pmatrix}.
\]

Also, \( W = V(\omega_t) \) has zero entries apart from \( W_{1,1} = u > 0 \), the variance of the level changes \( \eta_t \), and \( W_{2,2} = w > 0 \), the variance of the AR innovations. The observational errors are assumedly independent normal, \( N(\nu_t|0, v h_t) \), with a constant scale factor \( v \) and individual variance weights \( h_t \).

Various mixture error structures are modelled by assuming the \( h_t \) to be drawn independently from a common, specified prior \( p(h_t) \), which may be discrete, continuous or mixed; for example, the usual additive outlier model has \( p(h_t) = (1 - \pi)\delta_1(h_t) + \pi\delta_{k2}(h_t) \), and a model with \( T_k \) distributed errors has \( p(h_t) \propto h_t^{-(k+1)/2}\exp(-k/2h_t) \). Initial information is summarised through additional marginal priors \( p(u) \), \( p(v) \) and \( p(w) \) for the variance components, with an implicit independence assumption; independence is not necessary though will often be assumed, and is taken here for definiteness.
Additionally, assume an initial normal prior \((z_0|X_0) \sim N(z_0|\alpha_0, C_0)\), where \(X_0\) represents initial prior information. Then, for each \(n > 0\), define \(Z_n = \{z_0, \ldots, z_n\}\), the complete set of state vectors over times \(t = 0, \ldots, n\), and \(X_n = \{X_0, x_1, \ldots, x_n\}\), the complete set of observations up to time \(t = n\).

The uncertain quantities of interest are now the AR parameters \(\phi\), the full set of state vectors \(Z_n\), the variances \(v, u, w\), the variance weights \(\{h_t\}\). The various conditional distributions used in iterative simulations of this full set of quantities are now summarised. For any subsets of these quantities, generally denoted \(\theta\), write \(\theta^-\) for all remaining parameters and variables; e.g. \(\phi^- = \{Z_n, v, u, w, \{h_t\}\}\).

(a) Sampling \((Z_n|X_n, Z_n^-)\)

With fixed parameters and variances, the standard normal DLM theory applies. This is used to compute various useful normal distributions, as follows. Here the conditioning parameters \(Z_n^- = \{\phi, v, u, w, \{h_t\}\}\) are omitted from the notation for clarity.

- Filter forward through time to compute the usual sequence of Kalman filter-based moments of conditional distributions \((z_t|X_t) = N(z_t|m_t, C_t)\) for each \(t\).
- At \(t = n\), simulate a value of \(z_n\) from \(p(z_n|X_n)\); save this as the new value.
- For each \(t = n-1, n-2, \ldots, 0\), note that elements \(2, \ldots, p-1\) of the state vector \(z_t\) are known if \(z_{t+1}\) is known; they are simply the elements \(3, \ldots, p\) of the latter. Hence, given \(z_{t+1}\), replace entries \(2, \ldots, p-1\) of \(z_t\) accordingly. Sample the remaining two elements \((\mu_t, x^*_{t-p+1})\) as follows.
  
  - Compute the moments of \(p(\mu_t, x^*_{t-p+1}|X_t, x^*_{t}, \ldots, x^*_{t-p+2})\); this is just the bivariate conditional for the first and final elements of \(z_t\) from the full \(N(z_t|m_t, C_t)\) distribution. This is most efficiently done by sequentially conditioning on the elements \(x^*_{t-i}\) for \(i = 0, \ldots, p-2\) in turn, reducing the dimension of the normal distribution by 1 at each stage, eventually reducing from the full \(p+1\) dimensions to 2.
  
  - Compute \(y_{t+1} = x^*_{t+1} - \sum_{j=1}^{p-1} \phi_j x^*_{t+1-j}\) based on the previously sampled values of the elements of \(z_{t+1}\). We then have a pair of independent “observations” \(N(\mu_{t+1}|\mu_t, v)\) and \(N(y_{t+1}|\phi_p x^*_{t-p+1}, w)\), on the two parameters \((\mu_t, x^*_{t-p+1})\). Use the corresponding likelihood to update the bivariate normal “prior” \(p(\mu_t, x^*_{t-p+1}|X_t, x^*_{t}, \ldots, x^*_{t-p+2})\), to the corresponding “posterior” to result in the bivariate margin of the first and last elements of \(z_t\) under the distribution \(p(z_t|z_{t+1}, X_t)\). Sample this bivariate normal distribution, so producing a sampled value.
This results in a sequence $z_n, z_{n-1}, \ldots, z_0$ that represents a sample from the posterior distribution $p(Z_n|X_n, Z_n)$, as required.

(b) **Sampling** ($\phi|X_n, \phi^-$)

Given $Z_n$ and $w$ we have a standard AR framework, observing the normally distributed values $N(x_t^*|\sum_{j=1}^{p} \phi_j x_{t-j}^*, w)$ for $t = 1, \ldots, n$. Under a standard reference prior $p(\phi) \propto \text{constant}$ the conditional posterior is the usual reference normal posterior, and sampling is straightforward. Other priors may be used, with consequent changes to the sampling algorithm.

(c) **Sampling** ($v|X_n, v^-$)

For each $t = 1, \ldots, n$, compute the residuals $e_t = x_t - F'z_t$, and sample the posterior proportional to $p(v)v^{-n/2}\exp(-\sum_{t=1}^{n} e_t^2/2v)$. Inverse gamma priors for $v$ are conjugate, though sampling is straightforward with other forms.

(d) **Sampling** ($u|X_n, u^-$)

Similar to the above, compute $r_t = \mu_t - \mu_{t-1}$ and sample the posterior proportional to

$$p(u)u^{-n/2}\exp(-\sum_{t=1}^{n} r_t^2/2u).$$

(e) **Sampling** ($w|X_n, w^-$)

Again as above, compute $c_t = x_t^* - \sum_{j=1}^{p} \phi_j x_{t-j}^*$ and sample the posterior proportional to

$$p(w)w^{-n/2}\exp(-\sum_{t=1}^{n} c_t^2/2w).$$

(f) **Sampling** ($h_t|X_n, h^-_t$)

Given $\{h_t\}^-$, the $h_t$ are conditionally independent and so sampled from individual posteriors proportional to $p(h_t)h_t^{1/2}\exp(-e_t^2/2vh_t)$ with $e_t = x_t - F'z_t$, as above.

REFERENCES


Figure 1: Time series plot of oxygen isotope series.
Figure 2: Four Bayesian periodograms. Figure (a) is the log-likelihood function for the period $\lambda$ of a single sine wave model for the original oxygen series; (b) provides a similar plot for the series after subtracting a linear trend, (c) is that after subtracting a smooth trend based on lowess, and (d) that after differencing the series. The four functions are plotted on the same vertical scale for direct comparison.
Figure 3: Approximate posterior density contours for the periods of the two dominant quasi-cyclical components of the oxygen series in the fitted $AR(20)$ model.
Figure 4: Approximate posterior density contours for the modulii of the two dominant quasi-cyclical components of the oxygen series in the fitted $AR(20)$ model.
Figure 5: Estimated data decomposition based on conditional posterior mean of $AR(20)$ model parameter. The components displayed are the smooth, lowess trend, and the three major quasi-cyclical components of periods approximately 110kyr, 41kr and 23kyr respectively.
Figure 6: Estimated data decomposition based \( AR(20) \) analysis incorporating the measurement error/outlier mixture model. The components displayed are estimates of the smooth underlying trend, the measurement error/contaminations, together with the two major quasi-cyclical components.
Figure 7: Two Bayesian periodograms. Figure (a) is the log-likelihood function for the period $\lambda$ of a single sine wave model for a smooth detrended version of the first 400 oxygen observations, and (b) is that for the final 450 observations.