Bayesian Inference in Cyclic Component Dynamic Linear Models

Mike West

1. INTRODUCTION

Applications of traditional dynamic linear models (DLM’s), or state-space models, with cyclical components have to date been broadly restricted to contexts in which periods of cyclical components are known and integer, the typical contexts of seasonal modeling and time-varying harmonic analysis (West and Harrison 1989, chap. 8). Problems of inference about uncertain wavelengths of time-varying cyclical components in such models have been untouched, in part due to analytic complications. This article begins to address such problems.

Consider a real-valued, zero-mean time series $y_t$ observed at times $t = 1, 2, \ldots$. We deal with models of the form

$$y_t = x_{t,0} + \sum_{j=1}^{k} x_{t,j} + \nu_t,$$  \hspace{1cm} (1)

where the $k$ distinct, latent processes $x_{t,j}, t = 1, 2, \ldots,$ represent persistent periodic components of distinct frequencies and with, quite generally, possibly time-varying amplitude and phase characteristics. The component $x_{t,0}$ represents additional time series structure, including possibly nonstationary trend or growth and regression effects, and $\nu_t$ represents observation noise, typically combining measurement, sampling, and residual misspecification errors. Our interest lies specifically in models in which each of the cyclical processes has the form $x_{t,j} = r_{jt}\cos(\lambda_j t + \phi_{jt})$, a sinusoid of fixed period $\lambda_j = 2\pi / \alpha_j$, and stochastically time-varying amplitude $\alpha_j t$ and phase $\phi_{jt}$. We are interested in inference for the entire model to identify acyclic trends and, simultaneously, the wavelengths $\lambda_j$ and the patterns of behavior of the cyclical subsseries $x_{t,j}$ over the period of the observed data. One applied focus for such models is in evaluating trends in the presence of periodicities of uncertain wavelengths; another, more usual context is when interest lies in identifying persistent subcycles, possibly of generally low amplitude, in noisy series. The latter problem is typified by climatological studies in which the existence of subtle periodic components may influence discussions of global climate change (see, for example, West 1995).

A complete model specification requires descriptions of the precise forms of time-variation anticipated in amplitudes and phases, and that is done here through the use of cyclical form DLM’s (West and Harrison 1989, chap. 8 and sec. 9.4). These models provide generalizations and alternatives to more traditional parametric models, such as harmonic regression (Bretthorst 1988) and stationary models, with features common to all state-space models. In particular, the key issue of time variation in cyclical components naturally extends traditional harmonic regression models. DLM’s explicitly partition of sources of variation into purely observation noise versus noise components affecting signals, cyclical or otherwise. Thus the signals $x_{t,j}$ have characteristics that change over time, to a greater or lesser extent, and these changes are determined by noise terms identified and distinct from the noise terms $\nu_t$ that affect only the observations. In some problems, measurement and sampling errors and possible additive outliers account for a major component of observed data variation, and so this explicit representation, which is absent from traditional stationary time series models, is critical. Potential future extensions to model outliers and abrupt changes in signal form can build on existing modeling and intervention ideas in DLM’s (West and Harrison, chaps. 10 and 11). Further, DLM representations allow for routine handling of missing values and irregularly spaced data in time series (West and Harrison, sec. 10.5), in contrast to standard stationary time series and spectral methods. Moreover, the models include traditional harmonic regressions (Bretthorst 1988) as special cases, and so provide for the assessment of stable cyclical forms within the more general dynamic framework, and computational tools to fit essentially constant harmonic components.

Mike West is Professor and Director, Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708-0251. Partial support for this work was provided by National Science Foundation Grants DMS-9024792 and DMS-9311071.

© 1995 American Statistical Association
Journal of the American Statistical Association
December 1995, Vol. 90, No. 432, Theory and Methods
In addition to the specific focus on cyclical component models, this article develops, for the first time, simulation analyses for state evolution matrix parameters and variance components in DLM’s. This builds on and extends recent work on Gibbs sampling and related methods for posterior distributions for state vectors themselves (Carter and Kohn 1993; Frühwirth-Schnatter 1994) and provides the basis for wider application of state-space models uncertain defining evolution and variance parameters, in some generality.

2. DYNAMIC LINEAR MODEL AUTOREGRESSIONS

2.1 Basic Model Structure Using Cyclical Autoregressive Component Dynamic Linear Models

Of the several possible representations of time-varying cycles, the simple cyclical autoregressive form (West and Harrison 1989, sec. 9.4) is the focus here, for reasons that will become apparent. Consider each of the latent model components \( x_{t,j} \) in (1). The basic representation of interest is

\[
x_{t,j} = \beta_j x_{t-1,j} - x_{t-2,j} + \omega_t,j,
\]

where \( \beta_j \) is a model parameter to be discussed later and \( \omega_t,j \) is an innovation term at time \( t \). As usual in dynamic linear modeling, the \( \omega_t,j \), \((t = 1, 2, \ldots)\), are assumed to be independent and mutually independent sequences, with \( \omega_t,j \sim N(\omega_t,j|0, w_j) \).

This model for \( x_{t,j} \) is a standard AR(2) form with a state-space representation (West and Harrison 1989, sec. 9.4),

\[
x_{t,j} = (1, 0)z_{t,j} \quad \text{and} \quad z_{t,j} = G_j z_{t-1,j} + (\omega_t,j, 0)'
\]

where \( z_{t,j} = (x_{t,j}, x_{t-1,j})' \) for all \( t \). Conditional on an arbitrary starting value \( z_{1,j} \), the implied forecast function for the future of the process is \( E(x_{t,j}|z_{1,j}) = (1, 0)G_j^{t-1}z_{1,j} \), and this has a form, as a function of \( t \), determined by the eigenstructure of \( G_j \) using standard theory (West and Harrison 1989, chap. 5). It easily follows that \( E(x_{t,j}|z_{1,j}) = a_{1,j} \cos(2\pi t/\lambda_j + b_{1,j}), \) where \( \lambda_j = 2\pi/\cos(\beta_j/2), \) or \( \beta_j = 2 \cos(2\pi/\lambda_j) = 2 \cos(\alpha_j), \) and the amplitude \( a_{1,j} \) and phase \( b_{1,j} \) are constants determined by the starting value \( z_{1,j} \) and \( \beta_j \). Thus we have a model component defining a cyclical process of constant period \( \lambda_j \). Over time, the innovations \( \omega_t,j \) impact the process, inducing stochastic time variation in the amplitude and phase characteristics, with such variations related directly to the innovations variance \( w_j \) (Priestley 1981, sec. 3.5.3; West and Harrison 1989, p. 253).

Thus (1) and (2) combine to give the overall model

\[
y_t = x_{t,0} + \sum_{j=1}^{k} x_{t,j} + \nu_t
\]
and

\[ x_{t,j} = \beta_j x_{t-1,j} - x_{t-2,j} + \omega_{t,j}, \quad (j = 1, \ldots, k), \quad (3) \]

with \( \beta_j = 2 \cos(2\pi/\lambda_j) \) for \( j = 1, \ldots, k \). In addition to the earlier assumptions about the innovations process, it is further assumed that the observational errors form an independent sequence and are independent of the \( \omega_{t,j} \), with \( \omega_t \sim N(\nu_t|0, D) \).

It remains to model the acyclic trend process \( x_{t,0} \). To take a simple and specific model, though one of much utility and used in an illustration that follows, we suppose here a first-order polynomial. Other DLM forms are possible, of course, but the first-order model provides for simple but flexible local smoothing of underlying, acyclical, and non-stationary trends in the absence of more informed trend models such as regression DLM’s. Then (following West and Harrison 1989, chap. 2),

\[ x_{t,0} = x_{t-1,0} + \omega_{t,0}, \]

where \( \omega_{t,0} \sim N(\omega_{t,0}|0, \nu_0) \) independently over time and independently of the innovations \( \omega_{t,j} \) in (3). This trend model combines with (3) to give the general DLM representation

\[ y_t = F'z_t + \nu_t \]

and

\[ z_t = Gz_{t-1} + \delta_t, \quad (4) \]

with the following ingredients:

\[
\begin{pmatrix}
  x_{t,0} \\
  x_{t,1} \\
  x_{t-1,1} \\
  x_{t,2} \\
  x_{t-1,2} \\
  \vdots \\
  x_{t,k} \\
  x_{t-1,k}
\end{pmatrix}, \quad
\begin{pmatrix}
  1 \\
  1 \\
  0 \\
  1 \\
  0 \\
  \vdots \\
  1 \\
  0
\end{pmatrix},
\]

\[
G = \begin{pmatrix}
  1 & \beta_1 & \beta_2 & \cdots & \beta_k & -1 \\
  1 & 0 & \beta_2 & \cdots & \beta_k & -1 \\
  \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
  1 & 0 & \cdots & \cdots & 1 & 0
\end{pmatrix},
\]

\[
\delta_t = \begin{pmatrix}
  \omega_{t,0} \\
  \omega_{t,1} \\
  \omega_{t,2} \\
  \vdots \\
  \omega_{t,k}
\end{pmatrix}.
\quad (5)
\]

Here \( G \) is block-diagonal, so the “empty” entries in (5) are all zero. Note further that \( \delta_t \sim N(\delta_t|0, W) \) with diagonal variance matrix \( W = \text{diag}(w_0; w_1, 0; w_2, 0; \ldots; w_k, 0) \).

2.2 Posterior Structure and Framework of Simulation Analysis

Assume the model (4) for the series of \( n \) observations \( \{y_t; t = 1, \ldots, n\} \). Inference is required on the wavelengths \( \lambda \), equivalently the state parameters \( \beta \equiv \{\beta_1, \ldots, \beta_k\} \), the variance components \( \nu \) and \( W \), the individual “latent” stochastic subcycles \( X_j \equiv \{x_{t,j}; t = 0, \ldots, n\} \) for each \( j = 1, \ldots, k \), and the trend component \( X_0 \equiv \{x_{t,0}; t = 0, \ldots, n\} \). In standard notation, define \( D_t = \{y_t, D_{t-1}\} \) for the information set available at time \( t \), having observed the time series value at that time; at \( t = 0 \), \( D_0 \) represents all initial information including the model specification and assumptions. Assume the analysis to be closed to external inputs and interventions, so that \( D_t \) is a sufficient summary of the history at time \( t \). Hence analysis involves specifying
classes of priors for \( \{ \beta, v, W, z_0 \} \), then computing posterior distributions and appropriate summaries for \( \{ \beta, v, W, Z \} \), where \( Z \) is the complete set of state variables,

\[
Z \overset{\text{def}}{=} \{ X_0, X_1, \ldots, X_k \} \equiv \{ z_0, \ldots, z_n \}.
\]

Analyses here are based on priors of the form

\[
p(\beta, v, W, z_0) = p(\beta)p(v)p(z_0) \prod_{j=0}^{k} p(w_j),
\]

where \( p(\cdot) \) is generic notation for density function. Extensions to incorporate further prior dependencies are possible. On this basis, the following structure may be deduced:

1. Assume that \( \beta, v, \) and \( W \) are specified. Then (4) is a standard normal DLM whose analysis is fully defined in closed analytic form (West and Harrison 1989, chap. 4) if it is assumed that the prior for the initial state vector \( z_0 \) is a specified normal distribution, \( z_0 \sim N(z_0|m_0, C_0) \). Note that this prior may depend on the conditioning values of \( \beta, v, \) and \( W \), though this is of little practical relevance, and so no such dependence is assumed throughout. Standard updating results (West and Harrison 1989, sec. 4.3) apply to sequentially compute moments \( m_t \) and \( C_t \) of the posterior normal distributions for \( (z_t|D_t) \) at all times \( t \). Because the values of \( \beta, v, \) and \( W \) are conditioned on, make this explicit in the notation; thus it is simple to compute the moments defining the “on-line” posteriors,

\[
(z_t|D_t, \beta, v, W) \sim N(z_t|m_t, C_t), \quad (t = 1, \ldots, n). \tag{6}
\]

2. Assume the complete set of subseries \( Z = \{ X_0, X_1, \ldots, X_k \} \) to be specified in addition to \( \beta \), and focus on inference about the variance components \( v \) and \( W \). Under the assumed conditionally independent prior,

\[
p(v, W|D_n, \beta, Z) = p(v|D_n, Z)p(w_0|X_0) \prod_{j=1}^{k} p(w_j|\beta_j, X_j), \tag{7}
\]

with the following components:

(a) \( p(v|D_n, Z) \propto p(v)v^{-n/2}\exp(-s/(2v)) \), where \( s = \sum_{t=1}^{n}(y_t - F'x_t)^2 \);

(b) \( p(w_0|X_0) \propto p(w_0)w_0^{-n/2}\exp(-r_0/(2w_0)) \), where \( r_0 = \sum_{t=1}^{n}(x_{t,0} - x_{t-1,0})^2 \); and

(c) for each \( j = 1, \ldots, k, p(w_j|\beta_j, X_j) \propto p(w_j)w_j^{-(n-1)/2} \times \exp(-r_j/(2w_j)) \), and with sum of squares \( r_j = \sum_{t=2}^{n}(x_{t,j} - \beta_j x_{t-1,j} + x_{t-2,j})^2 \).

It is evident that independent inverse gamma priors for all \( k + 2 \) scalar variance components here are conditionally conjugate, the sets of posteriors described then being inverse gamma forms with updated shape and scale parameters in each case. Other priors may be used. One recommended alternative is rather uninformative priors in which the standard deviations \( \sqrt{v} \) and \( \sqrt{w_j} \) for each \( j \) are assumed to be uniformly distributed over finite and specified ranges. Then the foregoing conditional posteriors are truncated gamma distributions.

3. Assume the complete set of subseries \( Z = \{ X_0, X_1, \ldots, X_k \} \) to be specified in addition to the variance component \( v \) and \( W \), and focus on inference about \( \beta \). For each \( j = 1, \ldots, k, \) define scalar quantities \( b_j \) and \( B_j \) by

\[
B_j^{-1} = \sum_{t=2}^{n} x_{t-1,j}^2 \quad \text{and} \quad b_j = B_j \sum_{t=2}^{n} (x_{t,j} + x_{t-2,j})x_{t-1,j}.
\]

Then

\[
p(\beta|D_n, v, W, Z) \propto p(\beta) \prod_{j=1}^{k} \exp(-b_j^2/(2B_jw_j^2)). \tag{8}
\]

It is now easy to see that simulation-based analysis using Markov chain Monte Carlo is straightforward to implement. Iterative sampling of conditional posterior distributions de-
fined earlier may proceed by sampling through the sequence
\[
\begin{align*}
\cdots & \rightarrow p(Z|D_n, \beta, v, W) \\
& \rightarrow p(v|D_n, Z) p(w_0|X_0) \prod_{j=1}^{k} p(w_j|\beta_j, X_j) \\
& \rightarrow p(\beta|W, Z) \rightarrow \cdots
\end{align*}
\]

at each iteration.

The second stage here requires little further comment; the conditional posteriors for all variance components are outlined under Step 2. Steps 1 and 3 are considered in more detail in the next two sections. Before proceeding, note that (as in Carter and Kohn 1994 and Frühwirth-Schnatter 1994), successive draws under this simulation scheme will eventually resemble draws from the full joint posterior distribution for \(\{\beta, v, W, Z\}\). The very general results provided by Smith and Roberts (1993, thm. 1 and lem. 1 of appendix), for example, are applicable.

### 2.3 Simulation of Model Component Subseries

The first stage of each simulation iteration (9) involves sampling the conditional posterior of the full set \(Z\) of subseries, \(p(Z|D_n, \beta, v, W)\). Recent work by Carter and Kohn (1993) and Frühwirth-Schnatter (1994) provides the basis for this step; earlier ideas appear in work of Carlin, Polson, and Stoffer (1992), though the specific algorithms of the first two references are more appropriate, as was made explicit by Carter and Kohn. A value of \(Z = \{z_0, \ldots, z_n\}\) may be drawn from the rather complicated and high-dimensional full posterior \(p(Z|D_n, \beta, v, W)\) as follows:

- Simulate a value of \(z_n\) from \(p(z_n|D_n, \beta, v, W) \sim N(z_n|m_n, C_n)\), simply (6) at \(t = n\).
- For each \(t = n-1, n-2, \ldots, 0\), reducing the time index by one at each step, simulate a value of \(z_t\) from the conditional posterior \(p(z_t|D_n, \beta, v, W, z_{t+1}, z_{t+2}, \ldots, z_n)\); at each stage the conditioning values of \(z_n, z_{n-1}, \ldots\) are those just generated.

This produces a sequence \(z_n, z_{n-1}, \ldots, z_0\), equivalently the set \(Z\), as required. The computations are quite straightforward when it is recognized that on the basis of the strong conditional independence structure in the model (4),
\[
p(z_t|D_n, \beta, v, W, z_{t+1}) \equiv p(z_t|D_n, \beta, v, W, z_{t+1}).
\]

Furthermore, combining the second equation in (4) with the on-line posterior in (6) leads to
\[
p(z_t|D_n, \beta, v, W, z_{t+1}) \propto N(z_t|m_t, C_t)N(z_{t+1}|Gz_t, W),
\]

so that \(p(z_t|D_n, \beta, v, W, z_{t+1})\) is normal with moments easily found; in detail, the mean and variance matrix are \(m_t = A_t(z_{t+1} - Gm_t)\) and \(C_t = A_tQ_tA_t'\), where \(Q_t = GC_tG' + W\) and \(A_t = C_tGQ_t^{-1}\). This normal distribution is easily sampled, and then sequencing down through \(t = n-1, n-2, \ldots, 0\) produces a full draw, \(Z = \{z_n, \ldots, z_0\}\), as required. Further details have been given by Carter and Kohn (1994) and Frühwirth-Schnatter (1994).

A slight modification of the foregoing development that produces a more easily implemented algorithm involves simulating \(Z\) by drawing each of the model component subseries sequentially conditional on values of the other subseries. Specifically, for each \(j = 0, \ldots, k\), introduce \(Z^{(j)} = Z\backslash X_j\), the set \(Z\) with \(X_j = \{x_{j,0}, \ldots, x_{j,n}\}\) removed; write \(J\) for the corresponding index set \(J = \{0, \ldots, j-1, j+1, \ldots, k\}\). Fix the index \(j\) and suppose that \(Z^{(j)}\) is set at the previously sampled values. Compute the adjusted time series values \(y_{t,j} = y_t - \sum_{h \in J} x_{t,h}\), and write \(z_{t,j} = (x_{t,j} - x_{t-1,j})'\) for all \(t\). Then, based on a previously simulated full set of subseries \(Z\), a new set is generated by the following procedure:

![Figure 9. Approximate Posterior Density for the Smaller of the Two Wavelengths in the Two-Component Model for the EEG Series.](image-url)
Now the foregoing development in terms of the full state vector $\mathbf{z}_t$ can be reworked in terms of the two-dimensional vectors $\mathbf{z}_{t,j}$. Routine DLM updating is trivially performed to provide the collection of “on-line” posteriors $(\mathbf{z}_{t,j}\mid D_{t,j}, Z^{(j)})$, whence trivial calculations lead to $(\mathbf{z}_{t,j}\mid D_{t,j}, Z^{(j)})$, using the theory as in (9); it produces the normal with moments $\mathbf{m}_{t,j} + \mathbf{A}_{t,j}(\mathbf{z}_{t+1,j} - \mathbf{G}_j \mathbf{m}_{t,j})$ and $\mathbf{C}_{t,j} - \mathbf{A}_{t,j} \mathbf{Q}_{t,j} \mathbf{A}_{t,j}^\top$, where $\mathbf{Q}_{t,j} = \mathbf{G}_j \mathbf{C}_{t,j} \mathbf{G}_j^\top + \mathbf{W}_j$ and $\mathbf{A}_{t,j} = \mathbf{C}_{t,j} \mathbf{G}_j \mathbf{Q}_{t,j}^{-1}$. Here

$$G_j = \begin{pmatrix} \beta_j & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad W_j = \begin{pmatrix} w_j & 0 \\ 0 & 0 \end{pmatrix}.$$  

The original algorithm that samples each complete vector $\mathbf{z}_t$ at each iteration theoretically converges faster; breaking $\mathbf{z}_t$ into constituents and sampling these conditionally as in the second algorithm theoretically slows convergence by inducing higher correlations between successive draws. But the latter algorithm is very much faster in models with even very moderate values of $k$; the matrix inversions $\mathbf{Q}_{t,j}^{-1}$ required at each $t$ in the former algorithm lead to significant overheads. Some approximations to avoid repeat matrix inversions are possible, though their effectiveness is unexplored. Experience with both algorithms has suggested that the faster, conditional approach is in fact typically adequate in terms of producing simulations in close agreement with the former algorithm; further investigations are needed to compare and assess the two approaches for both statistical and computational efficiencies.

2.4 Simulation of Wavelengths

Restricting wavelengths to $\lambda_j > 2$, the Nyquist limit and practical limit for identification of cycles, the map between $\lambda_j > 2$ and $\beta_j$ is monotonically decreasing, with $\beta_j = 2 \cos(2\pi/\lambda_j)$ lying between $\pm 2$ (indeed, rather often, relevant wavelengths will exceed 4, so that $\beta_j > 0$.) Hence we can work interchangeably with either the wavelengths or the

$$G_j = \begin{pmatrix} \beta_j & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad W_j = \begin{pmatrix} w_j & 0 \\ 0 & 0 \end{pmatrix}.$$

Figure 11. Approximate Posterior Density for the Trend Innovation Standard Deviation $\sqrt{\mathbf{w}_0}$ in the EEG Analysis.
autoregressive coefficients; simulation of the $\beta$ parameters from conditional posteriors $p(\beta|W, Z)$ in (9) produces sampled wavelengths via the identities $\lambda_j = 2\pi/\arccos(\beta_j/2)$ for each $j$, and posterior inferences may be explored directly from the samples. Conversely, it is natural to think in terms of priors for the wavelengths and then transform to deduce priors for the $\beta_j$.

Strict constraints on wavelengths may be enforced through the prior distribution. Some form of constraint is necessary for identification—note that the model remains the same under arbitrary permutation of the wavelength indices. The constraint here is simply the ordering $\lambda_1 < \lambda_2 < \cdots < \lambda_k$ under the prior, and this provides parameter identification. Other strict constraints, such as enforcing a minimum separation between any two wavelengths, may be desirable in some applications and can be similarly introduced.

In an illustration that follows we use conditional priors that are effectively reference priors for the $\beta_j$ conditional on all other parameters, simply conditional uniforms. This provides a benchmark analysis. Alternative analyses based on plausible, informed priors for the $\lambda_j$ that induce specific forms of priors for the $\beta_j$ parameters may be similarly implemented within this framework. Write $U(a, b)$ for the uniform distribution on $(a, b)$ and suppose a specified minimum wavelength separation of $\delta > 0$ time units (e.g., $\delta = 1$). The class of priors used is defined by the set of complete conditionals $(\beta_j|\beta(j))$, where $\beta(j) = \beta \setminus \beta_j \equiv \{\beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_k\}$, for all $j$. Extend notation to include fixed and specified limits on wavelengths $\lambda_0 < \lambda_{k+1}$; often $\lambda_0 = 2$ and $\lambda_{k+1}$ is some appropriately large value. Then take conditionals $(\beta_j|\beta(j)) \sim U(\beta_j|\beta(j)_1, \beta(j)_2)$ for all $j$, where $\beta(j)_1 = 2\cos(2\pi/(\lambda_{j-1} + \delta))$ and $\beta(j)_2 = 2\cos(2\pi/(\lambda_{j+1} - \delta))$.

Under the joint prior so specified, the conditional posterior (8) implies

$$p(\beta_j|D_n, \beta(j), V, W, Z) \propto N(\beta_j|b_j, B_j w_j),$$

$$(\beta(j)_1 < \beta_j < \beta(j)_2).$$

So sampling may proceed conditionally, cycling through the $\beta_j$ parameters given previously sampled values of $\beta(j)$ (among other things), with new values of the $\beta_j$ trivially generated from a truncated normals.

As noted, other priors might be used, as context demands, based on transforming from plausible priors for the $\lambda_j$. Working in terms of conditional priors is natural, and the foregoing development is modified only through conditionals $p(\beta_j|\beta(j))$ that will be other than uniforms. The conditional posteriors are then $p(\beta_j|D_n, \beta(j), V, W, Z) \propto p(\beta_j|\beta(j))N(\beta_j|b_j, B_j w_j)$ so the analysis differs only technically; sampling these conditional posteriors will gener-

![Figure 12. Approximate Posterior Density for the Component Innovation Standard Deviation $\sqrt{w_1}$ in the EEG Analysis.](image1)

![Figure 13. Approximate Posterior Density for the Component Innovation Standard Deviation $\sqrt{w_2}$ in the EEG Analysis.](image2)

![Figure 14. Approximate Posterior Density for the Observation Error Standard Deviation $\sqrt{v}$ in the EEG Analysis.](image3)
generally be more complicated than in the case of truncated normals, and rejection methods will be needed.

3. ILLUSTRATIONS

3.1 An EEG Time Series

Figures 1–14 display some features of analysis of a series of \( n = 110 \) electroencephalogram (EEG) recordings from the scalp of an individual undergoing electroconvulsive therapy (ECT). These data were provided by Dr Andrew Krystal in Psychiatry at Duke University and arise in studies of waveform characteristics in multichannel EEG signal analyses; these studies are germane to assessments of differing ECT protocols (see, for example, Krystal et al. 1992). Comparison of two or more such time series underlies part of the study, and appropriate modeling of individual time series represents a starting position for comparative analyses. Here attention focuses on the single series appearing in Figure 1 and the identification and estimation of time-varying periodicities in the data (further substantive developments of this study will be reported elsewhere). The actual series shown is a short segment of a much longer series from just one of several channels, and the numbers are raw digitized values generated prior to conversion to electrical potentials. There are \( n = 110 \) observations over a total time span of just over 1 second, corresponding to the very start of an ECT-induced seizure.

Figure 1 plots the raw data over time; periodicities are evident, as are variations in amplitude over time, and there are noticeable nonlinear but apparently acyclic trends underlying the evolution of the series. Some initial exploratory graphics geared to identifying periodicities appear in Figures 2 and 3. These are based on a detrended series computed by subtracting an estimate of an underlying trend over time. (This estimated trend, from lowess in S-Plus, is similar to that in the DLM analysis reported here, and the details of how it was computed are not important here.) Figure 2 shows the smoothed periodogram estimate of the spectral density function (using the default spec.pgram in S-Plus) for the detrended data and plotted as functions of wavelength rather than frequency. Figure 3 displays the likelihood/Bayesian analog, namely the (integrated) log-likelihood function for wavelength \( \lambda \) in a constant, single-cycle harmonic regression model (see, for example, Bretthorst 1988, chap. 3); this represents the log-posterior under a uniform prior for \( \lambda \). (Note that, as is theoretically the case whatever the data, the log-likelihood here is bounded below, so that a proper posterior distribution is obtained only under a proper prior for \( \lambda \).) Qualitatively, these figures indicate high power at wavelengths around \( \lambda = 6 \), with some possible contributions in the wavelength range 10–15. Note that qualitatively similar conclusions are arrived at using the raw data, rather than the specific detrended version used here. The differences are that additional and apparently important peaks arise in the periodogram and log-likelihood functions at much higher wavelengths; these peaks are induced by the underlying acyclic trends in the data rather than real low-frequency periodicities.

On use of these exploratory graphical methods is to identify suitable starting values for wavelengths in simulation analysis of DLM autoregression, as described in Section 2. Such was done in various analyses of these data; results from a model with \( k = 2 \) component are summarized. This analysis assumed priors as follows:

- Uninformative, independent priors for the initial level of the series, \( x_{1,0} \), and the initial values of the component subcycles, \( x_{1,j} \) and \( x_{0,j} \) for \( j = 1, 2 \); specifically, \( N(x_{1,0}|0,1,000) \) and \( N(z_{1}|0,1,000I) \), where \( I \) is the identity matrix.
- Conditionally uniform priors for the wavelength coefficients \( \beta \) induced by constraints on wavelengths of \( 2.0 \leq \lambda_j \leq 25.0 \), and a minimum separation of one wavelength, \( \lambda_2 > \lambda_1 + 1.0 \) (though this turns out to be nonbinding, as the likelihood function gives little support to close values of the two wavelengths.)
Proper priors for the variance components induced by uniform priors for the corresponding standard deviations: $p(\nu) \propto \nu^{-1/2}$ over $0 < \nu < 10^4$; similarly, for each of the evolution variances the prior is $p(w_j) \propto w_j^{-1/2}$ over $0 < w_j < 625$. Other priors might be used; these uniform (for standard deviations) priors are assumed to provide a baseline or reference analysis.

The priors here are proper and rather uninformative. The simulation analysis ran through 10,000 iterations of the Gibbs sampling scheme before any draws were saved; this period of “burn-in” to convergence is quite conservative, as repeat runs with differing starting values confirm. Following this, 1 million iterations were run; the graphs here display summaries of 10,000 of these samples obtained at every 100th iteration. This substantially reduces correlations between successive draws. All approximate posterior inferences here are based directly on the Monte Carlo draws, in terms of histograms of the 10,000 sampled values for all parameters, simple evolution means and standard deviations, and so forth. Figure 4 provides a scatterplot of the data over time with the estimated trajectory of the underlying trend superimposed; the full line represents $E(x_{t,0}|D_n)$ over $t = 1, \ldots, n$, and the dashed lines provide uncertainties in terms of $\pm \sqrt{V(x_{t,0}|D_n)}$. Similar trajectories for the $k = 2$ subcycles appear in Figures 7 and 8, plotted with the same range as the data. Figures 5 and 6 display the fitted values and residuals. Figures 9–14 display posteriors for the wavelengths and variance components. Figures 7 and 9 confirm the subcycle with a wavelength around $\lambda_1 \approx 6.3 \pm .2$. The former frame indicates the amplitude over time of this component, apparently negligible during the first 20 or 30 time intervals but quite prominent and persistent later on. The relevance of a dynamic model allowing for time variation in amplitudes is clear here. The second component picks up on the periodogram-based exploratory analyses that indicated some minor residual periodicities at higher wavelengths between 8 and 15; the posterior in frame 1.10 has mass around $\lambda_2 \approx 12 \pm 3$; note the high degree of uncertainty. Figure 8 indicates that although the estimated amplitude is apparently substantial and important during the latter half of the observation period, uncertainty about this subcycle is relatively high throughout the period. Because of the increased contribution from this component in the later stages, future predictions are affected, and so the component is indeed quite relevant despite the high degree of uncertainty. It was noted earlier that the data arose at the very start of an ECT-induced seizure, and these summary inferences are consistent with initial transient movement in underlying trend plus initial variation in the cyclical be-
behavior as the seizure begins; later on, the trend variation typically dissipates and the cyclical components stabilize.

Some minor residual, negative correlation is evident at lag two in the fitted residuals; though very small, such residual structure could be modeled by including a residual noise component, such as an assumedly stationary residual autoregressive process. This has been explored using straightforward extensions of the current modeling framework, though with little net gain, as residual variation is slight. Alternatively, rather than simply describing residual structure in terms of a noise model, some explanation might be sought; for example, it may be attributable to a nonlinear cyclical process, the assumedly linear (though time-varying) model providing a good but not perfect approximation. Another, closely related possibility is that the waveforms might vary in frequency over time, so that then residual structure would emerge in this analysis that assumes constant frequencies. It is in fact plausible that this is the case here, and this can be addressed in extensions of the models to allow for time-varying wavelengths, to be reported elsewhere.

A repeat analysis assuming \( k = 3 \) cyclical components essentially confirms these results with some minor differences. First, the estimated underlying trend is rather smoother; the residuals are essentially unchanged, the additional component contributing negligibly to the data description. Correspondingly, the third component has an insignificant amplitude over time, essentially contributing to the very weak indication of an additional wavelength in the 10–16 range, with the more evident, though still subtle, second cycle of wavelength around 8–11. Note the correspondence with the mass distribution under the Bayesian “periodogram” in Figure 3.

### 3.2 A Paleoclimatological Time Series

Figures 15–25 display some features of analysis of a series of \( n = 177 \) observations constructed from raw measurements of a geochemical indicator of climatic conditions taken from sedimentary cores taken from the bed of Lake Turkana in eastern Africa (Halfman and Johnson 1988; West 1995). The data are measures of calcium carbonate (as a percentage of dry weight) in the sediments, made at various locations down the core; following Halfman, Johnson, and Finney (1994), the data have been approximately timed via a linear map from depth in the core to nominal calendar time, indicated in the graphs. Assuming this timing to be accurate, trends and cycles in the carbonate series are taken as indicative of variations in ambient climatic conditions, and so of interest in connection with climate change issues; in particular, the existence and nature of cycles, and their implications for the near term future, are of central interest. The displayed data are just those used in analysis by the aforementioned authors and are of interest here for comparison with their analysis. They are obtained from original, unequally spaced data as interpolated values obtained by fitting a cubic spline to the raw depth:carbonate...
data, and then evaluating the interpolant at equally spaced points down the core prior to mapping to nominal real time.

The graphs display summary inferences, in a format analogous to that of the EEG analysis, for a two-component model with a locally linear trend. As with the EEG analysis, we use diffuse uniform priors for the square roots of variance components. To summarize the outputs, there is some evidence of one subtle subcycle of wavelength in the 140–160-year range, and this echoes results of related data in this field (Halfman et al. 1994). Additional cyclical structure, if present, lies in the 100–115-year range, with some suggestion of 80–100 years too; however, the estimated amplitude of such secondary features is essentially negligible, these data being as well described by a one-cycle model as by a two-cycle model. And then the period of 150± is evidenced only very weakly, the observation noise dominating the data variation.

This example contrasts with the EEG analysis in that the estimated trajectories of cyclical components are rather stable over time. In this sense it is apparent that the models and approach essentially encompass the more usual context of harmonic regression with fixed sinusoids (Bretthorst 1988) in cases of low levels of evolution variance in the model components.

Further and more penetrating analyses of these and related data has been provided by West (1995).

4. CONCLUSIONS

Some immediate issues arising in applications have to do with simulation sample sizes and convergence checks, standard considerations in Bayesian analyses via simulation, and, particularly, the choices of prior distributions for DLM variance components. The priors used in the foregoing examples are proper uniform distributions for \(v\) and each of the \(w_j\), with ranges that turn out not to conflict with the likelihood function. But as a result, as is clear from the figures, the underlying trend component of the model is inferred to be possibly rather more variable than perhaps expected; this may be viewed as tending to overfit the data (partly supported by the observed though minor degree of negative residual correlation). More informed priors on the variance components to control to smaller values and hence smooth out some of the irregularity might be desirable—indeed, common practice with DLM’s is to directly control evolution variances to smaller values, perhaps through the use of discount factors (Pole, West, and Harrison 1994; West and Harrison 1989).

The formulation in DLM terms permits application in contexts where the time series is irregularly spaced over time or subject to missing values. Missing observations are directly subsumed by the usual sequential DLM analysis; prior to posterior update is vacuous if no data are observed. This leads neatly into the treatment of irregularly spaced data, as developed by West and Harrison (1989, chap. 10). Unequally spaced data can usually be treated by

![Figure 23. Approximate Posterior Density for the Component Innovation Standard Deviation \(\sqrt{w_i}\) in the CaCO3 Series.](image1)

![Figure 24. Approximate Posterior Density for the Component Innovation Standard Deviation \(\sqrt{w_2}\) in the CaCO3 Series.](image2)

![Figure 25. Approximate Posterior Density for the Observation Error Standard Deviation \(\sqrt{v}\) in the CaCO3 Series.](image3)
identifying a base minimum time interval underlying the observed times of observations and developing the equally spaced DLM on that time scale. Then the times with no observations are treated as points of missing data. This is important for application. For example, these models are currently in development as part of a broad study of the geochemical indicators of climatic change used in the example, in which raw data series are irregularly spaced (West 1995). Indeed, the timings of observations there are subject to uncertainties that introduce additional complications that can be accommodated with further modeling extensions of obvious utility in other applications (West 1995, sec. 4). The DLM formulation also permits access to other model consequences, including predictive aspects. If relevant and desired in an application, predictions may be routinely generated from the simulation analysis. For each sampled set of parameters and subcycles, insert the values into model (4) and simply project into the future to produce a set of sampled “futures” of the series. Averaging produces approximate marginal predictive means and other features of marginal forecast distributions. But though they are useful (and traditional), marginal inferences represent crude summaries of full joint predictive distributions, and other ways of exploring and summarizing and sets of sequences of such possible “futures” are needed. This represents an open area to investigation.

The specific models here are clearly extensible to unrestricted autoregressive components (Pole and West 1990; West and Harrison 1989, sec. 9.4) to model stationary cyclical behavior and additional residual time series structure in the data, as mentioned in the EEG example. The simulation framework directly extends to such models, as additive autoregressive components simply introduce further uncertain parameters into the system matrix $G$ in (4). Such models are currently under investigation in specific application contexts.

REFERENCES


