1

BAYESIAN DYNAMIC MODELLING

Mike West
Department of Statistical Science
Duke University

Chapter 8 of
Bayesian Theory and Applications
published in honour of Professor Sir Adrian F.M. Smith

(eds: P. Damien and P. Dellaportas and N. G. Polson and D. A. Stephens)
Clarendon: Oxford University Press, 2013

1.1 Introduction

Bayesian time series and forecasting is a very broad field and any attempt at other than a very selective and personal overview of core and recent areas would be foolhardy. This chapter therefore selectively notes some key models and ideas, leavened with extracts from a few time series analysis and forecasting examples. For definitive development of core theory and methodology of Bayesian state-space models, readers are referred to [74, 46] and might usefully read this chapter with one or both of the texts at hand for delving much further and deeper. The latter parts of the chapter link into and discuss a range of recent developments on specific modelling and applied topics in exciting and challenging areas of Bayesian time series analysis.

1.2 Core Model Context: Dynamic Linear Model

1.2.1 *Introduction*

Much of the theory and methodology of all dynamic modelling for time series analysis and forecasting builds on the theoretical core of linear, Gaussian model structures: the class of univariate normal dynamic linear models (DLMs or NDLMs). Here we extract some key elements, ideas and highlights of the detailed modelling approach, theory of model structure and specification, methodology and application.

Over a period of equally-spaced discrete time, a univariate time series $y_{1:n}$ is a sample from a DLM with $p-$vector state θ_t when

$$y_t = x_t + \nu_t, \quad x_t = F_t'\theta_t, \quad \theta_t = G_t\theta_{t-1} + \omega_t, \qquad t = 1, 2, \ldots, \tag{1.1}$$

where: each F_t is a known regression $p-$vector; each G_t a $p \times p$ state transition matrix; ν_t is univariate normal with zero mean; ω_t is a zero-mean $p-$vector representing evolution noise, or innovations; the pre-initial state θ_0 has a normal prior; the sequences ν_t, ω_t are independent and mutually independent, and also independent of θ_0. DLMs are hidden Markov models; the state vector θ_t is a latent or hidden state, often containing values of underlying latent processes as well as time-varying parameters (chapter 4 of [74]).

1.2.2 *Core Example DLMs*

Key special cases are distinguished by the choice of elements F_t, G_t. This covers effectively all relevant dynamic linear models of fundamental theoretical and practical importance. Some key examples that underlie much of what is applied in forecasting and time series analysis are as follows.

Random Walk in Noise (chapter 2 of [74]): $p = 1,\ F_t = 1, G_t = 1$ gives this first-order polynomial model in which the state $x_t \equiv \theta_{t1} \equiv \theta_t$ is the scalar local level of the time series, varying as a random walk itself.

Local Trend/Polynomial DLMs (chapter 7 of [74]): $F_t = E_p = (1, 0, \cdots, 0)'$ and $G_t = J_p$, the $p \times p$ matrix with 1s on the diagonal and super-diagonal, and zeros elsewhere, define "locally smooth trend" DLMs; elements of θ_t are the local level of the underlying mean of the series, local gradient and change in gradient etc., each undergoing stochastic changes in time as a random walk.

Dynamic Regression (chapter 9 of [74]): When $G_t = I_p$, the DLM is a time-varying regression parameter model in which regression parameters in θ_t evolve in time as a random walk.

Seasonal DLMs (chapter 8 of [74]): $F_t = E_2$ and $G_t = rH(a)$ where $r \in (0, 1)$ and

$$H(a) = \begin{pmatrix} \cos(a) & \sin(a) \\ -\sin(a) & \cos(a) \end{pmatrix}$$

for any angle $a \in (0, 2\pi)$ defines a dynamic damped seasonal, or cyclical, DLM of period $2\pi/a$, with damping factor r per unit time.

Autoregressive and Time-varying Autoregressive DLMs (chapter 5 of [46]): Here $F_t = E_p$ and G_t depends on a $p-$vector $\phi_t = (\phi_{t1}, \dots, \phi_{tp})'$ as

$$G_t = \begin{pmatrix} \phi_{t1} & \phi_{t2} & \phi_{t3} & \cdots & \phi_{tp} \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

with, typically, the evolution noise constrained as $\omega_t = (\omega_{t1}, 0, \dots, 0)'$. Now $y_t = x_t + \nu_t$ where $x_t \equiv \theta_{t1}$ and $x_t = \sum_{j=1:p} \phi_{tj} x_{t-j} + \omega_{t1}$, a time-varying autoregressive process of order p, or TVAR(p). The data arise through additive noisy observations on this hidden or latent process.

If the ϕ_{tj} are constant over time, x_t is a standard $\mathrm{AR}(p)$ process; in this sense, the main class of traditional linear time series models is a special case of the class of DLMs.

1.2.3 *Time Series Model Composition*

Fundamental to structuring applied models is the use of building blocks as components of an overall model– the principal of composition or *superposition* (chapter 6 of [74]). DLMs do this naturally by collecting together components: given a set of individual DLMs, the larger model is composed by concatenating the individual component θ_t vectors into a longer state vector, correspondingly concatenating the individual F_t vectors, and building the associated state evolution matrix as the block diagonal of those of the component models. For example,

$$\begin{aligned} F' &= (1,\, f_t,\, E_2',\, E_2',\, E_2'), \\ G &= \text{block diag}\left\{1,\ 1,\ H(a_1),\ H(a_2),\ \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}\right\} \end{aligned} \tag{1.2}$$

defines the model for the signal as

$$x_t = \theta_{t1} + \theta_{t2} f_t + \rho_{t1} + \rho_{t2} + z_t$$

where:

- θ_{t1} is a local level/random walk intercept varying in time;
- θ_{t2} is a dynamic regression parameter in the regression on the univariate predictor/independent variable time series f_t;
- ρ_{tj} is a seasonal/periodic component of wavelength $2\pi/a_j$ for $j = 1, 2$, with time-varying amplitudes and phases– often an overall annual pattern in weekly or monthly data, for example, can be represented in terms of a set of harmonics of the fundamental frequency, such as would arise in the example here with $a_1 = \pi/6, a_2 = \pi/3$ yielding an annual cycle and a semi-annual (six month) cycle;
- z_t is an $\mathrm{AR}(2)$ process– a short-term correlated underlying latent process– that represents residual structure in the time series signal not already captured by the other components.

1.2.4 *Sequential Learning*

Sequential model specification is inherent in time series, and Bayesian learning naturally proceeds with a sequential perspective (chapter 4 of [46]). Under a specified normal prior for the latent initial state θ_0, the standard normal/linear sequential updates apply: at each time $t-1$ a "current" normal posterior evolves via the evolution equation to a 1-step ahead prior distribution for the next state θ_t; observing the data y_t then updates that to the time t posterior, and we progress further in time sequentially. Missing data in the time series is trivially dealt with: the prior-to-posterior update at any time point where the observation is missing involves no change. From the early days– in the 1950s– of so-called

Kalman filtering in engineering and early applications of Bayesian forecasting in commercial settings (chapter 1 of [74]), this framework of closed-form sequential updating analysis– or *forward filtering* of the time series– has been the centerpiece of the computational machinery. Though far more complex, elaborate, nonlinear and non-normal models are routinely used nowadays based on advances in simulation-based computational methods, this normal/linear theory still plays central and critical roles in applied work and as components of more elaborate computational methods.

1.2.5 *Forecasting*

Forecasting follows from the sequential model specification via computation of predictive distributions. At any time t with the current normal posterior for the state θ_t based on data $y_{1:t}$, and any other information integrated into the analysis, we simply extrapolate by evolving the state through the state evolution equation into the future, with implied normal predictive distributions for sequences $\theta_{t+1:t+k}, y_{t+1:t+k}$ into the future any $k > 0$ steps ahead. Forecasting via simulation is also key to applied work: simulating the process into the future– to generate "synthetic realities"– is often a useful adjunct to the theory, as visual inspection (and perhaps formal statistical summaries) of simulated futures can often aid in understanding aspects of model fit/misfit as well as formally elaborating on the predictive expectations defined by the model and fit to historical data; see Figures 1.2 and 1.3 for some aspects of this in the analysis of the climatological Southern Oscillation Index (SOI) time series discussed later in Section 1.3.2. The concept is also illustrated in Figure 1.4 in a multivariate DLM analysis of a financial time series discussed later in Section 1.4.1.

1.2.6 *Retrospective Time Series Analysis*

Time series analysis– investigating posterior inferences and aspects of model assessment based on a model fitted to a fixed set of data– relies on the theory of *smoothing* or *retrospective filtering* that overlays the forward-filtering, sequential analysis. Looking back over time from a current time t, this theory defines the revised posterior distributions for historical sequences of state vectors $\theta_{t-1:t-k}$ for $k > 0$ that complement the forward analysis (chapter 4 of [74]).

1.2.7 *Completing Model Specification: Variance Components*

The Bayesian analysis of the DLM for applied work is enabled by extensions of the normal theory-based sequential analysis to incorporate learning on the observational variance parameters $V(\nu_t)$ and specification of the evolution variance matrices $V(\omega_t)$. For the former, analytic tractability is maintained in models where $V(\nu_t) = k_t v_t$, with known variance multipliers k_t, and $V(\omega_t) = v_t W_t$ with two variants: (i) constant, unknown $v_t = v$ (section 4.3.2 of [46]) and (ii) time-varying observational variances in which v_t follows a stochastic volatility model based on variance discounting– a random walk-like model that underlies many applications where variances are expected be be locally stable but globally varying (section 4.3.7 of [46]). Genesis and further developments are given in chapter 10 of [74] and, with recent updates and new extensions, in chapters 4,7 and 10 of [46].

The use of discount factors to structure evolution variance matrices has been and remains central to many applications (chapter 6 of [74]). In models with non-trivial state vector dimension $p,$ we must maintain control over specification of the W_t to avoid exploding the numbers of free parameters. In many cases, we are using W_t to reflect low levels of stochastic change in elements of the state. When the model is structured in terms of block components via superposition as described above, the W_t matrix is naturally structured in a corresponding block diagonal form; then the strategy of specifying these blocks in W_t using the discount factor approach is natural (section 4.3.6 of [46]). This strategy describes the innovations for each component of the state vector as contributing a constant stochastic "rate of loss of information" per time point, and these rates may be chosen as different for different components. In our example above, a dynamic regression parameter might be expected to vary less rapidly over time than, perhaps, the underlying local trend.

Central to many applications of Bayesian forecasting, especially in commercial and economic studies, is the role of "open modelling". That is, a model is often one of multiple ways of describing a problem, and as such should be open to modification over time as well as integration with other formal descriptions of a forecasting problem (chapter 1 of [74]). The role of statistical theory in guiding changes– interventions to adapt a model at any given time based on additional information– that maintain consistency with the model is then key. Formal sequential analysis in a DLM framework can often manage this via appropriate changes in the variance components. For example, treating a single observation as of poorer quality, or a likely outlier, can be done via an inflated variance multiplier $k_t;$ feeding into the model new/external information that suggests increased chances of more abrupt change in one or more components of a state vector can be done via larger values of the corresponding elements of W_t, typically using a lower discount factor in the specification for just that time, or times, when larger changes are anticipated. Detailed development of a range of subjective monitoring and model adaptation methods of these forms, with examples, are given in chapters 10-12 of [74] and throughout [43]; see also chapter 4 of [46] and earlier relevant papers [72, 62, 73].

1.2.8 *Time Series Decomposition*

Complementing the strategy of model construction by superposition of component DLMs is the theory and methodology of model decomposition that is far-reaching in its utility for retrospective time series analysis (chapter 9 of [74]). Originally derived for the class of *time series DLMs* in which $F_t = F, G_t = G$ are constant for all time [66–70], the theory of decompositions applies also to time-varying models [45, 44, 47]. The context of DLM AR(p) and TVAR(p) models– alone or as components of a larger model– is key in terms of the interest in applications in engineering and the sciences, in particular (chapter 5 of [46]).

Consider a DLM where one model component z_t follows a TVAR(p) model. The main idea comes from the central theoretical results that a DLM implies a decomposition of the form

$$z_t = \sum_{j=1:C} z^c_{tj} + \sum_{j=1:R} z^r_{tj}$$

where each z^*_{tj} is an underlying latent process: each z^r_{tj} is a TVAR(1) process and each z^c_{tj} is a quasi-cyclical time-varying process whose characteristics are effectively those of a TVAR(2) overlaid with low levels of additive noise, and that exhibits time-varying periodicities with stochastically varying amplitude, phase and period. In the special case of constant AR parameters, the periods of these quasi-cyclical z^c_{tj} processes are also constant.

This DLM decomposition theory underlies the use of these models– state-space models/DLMs with AR and TVAR components– for problems in which we are interested in a potentially very complicated and dynamic autocorrelation structure, and aim to explore underlying contributions to the overall signal that may exhibit periodicities of a time-varying nature. Many examples appear in [74, 46] and references there as well as the core papers referenced above. Figures 1.1, 1.2 and 1.3 exemplify some aspects of this in the analysis of the climatological Southern Oscillation Index (SOI) time series of Section 1.3.2.

1.3 Computation and Model Enrichment

1.3.1 *Parameter Learning and Batch Analysis via MCMC*

Over the last couple of decades, methodology and applications of Bayesian time series analysis have massively expanded in non-Gaussian, nonlinear and more intricate conditionally linear models. The modelling concepts and features discussed above are all central to this increasingly rich field, while much has been driven by enabling computational methods.

Consider the example DLM of equation (1.2) and now suppose that $V(\nu_t) = k_t v$ with known weights k_t but uncertain v to be estimated, and the evolution variance matrix is

$$W_t \equiv W = \text{block diag}\left\{\tau_1,\ \tau_2,\ \tau_3 I_2,\ \tau_4 I_2,\ \begin{pmatrix} w & 0 \\ 0 & 0 \end{pmatrix}\right\}. \tag{1.3}$$

Also, write $\phi = (\phi_1, \phi_2)'$ for the AR parameters of the latent AR(2) model component. The DLM can be fitted using standard theory assuming the full set of model parameters $\mu = \{v, \phi, w, \tau_{1:4}\}$ to be known. Given these parameters, the forward-filtering and smoothing based on normal/linear theory applies.

Markov chain Monte Carlo methods naturally open the path to a complete Bayesian analysis under any specified prior $p(\mu)$; see chapter 15 of [74] and section 4.5 of [46] for full details and copious references, as well as challenging applications in chapter 7 of [46]. Given an observed data sequence $y_{1:n}$, MCMC iteratively re-simulates parameters and states from appropriate conditional distributions. This involves conditional simulations of elements of μ conditioning on current values of other parameters and a current set of states $\theta_{0:n}$ that often break down into tractable parallel simulators. The example above is a case in point under independent priors on ϕ and the variances v, τ_j, w, for example.

Central to application is the *forward filtering, backward sampling* (FFBS– [6,18]) algorithm that arises naturally from the normal/linear theory of the DLM conditional on parameters μ. This builds on the sequential, forward filtering theory to run through the data, updating posterior distributions for states over time, and then steps back in time: at each point $t = n, t = n-1, \ldots, t = 1, t = 0$ in turn, the retrospective distributional theory of this *conditionally linear, normal* model provides normal distributions for the states that are simulated. This builds up a sequence $\{\theta_n, \theta_{n-1}, \ldots, \theta_1, \theta_0\}$ that represents a draw– sampled via composition backwards in time– from the relevant conditional posterior $p(\theta_{0:n}|\mu, y_{1:n})$. The use of MCMC methods also naturally deals with missing data in a time series; missing values are, by definition, latent variables that can be simulated via appropriate conditional posteriors each step of the MCMC.

1.3.2 *Example: SOI Time Series*

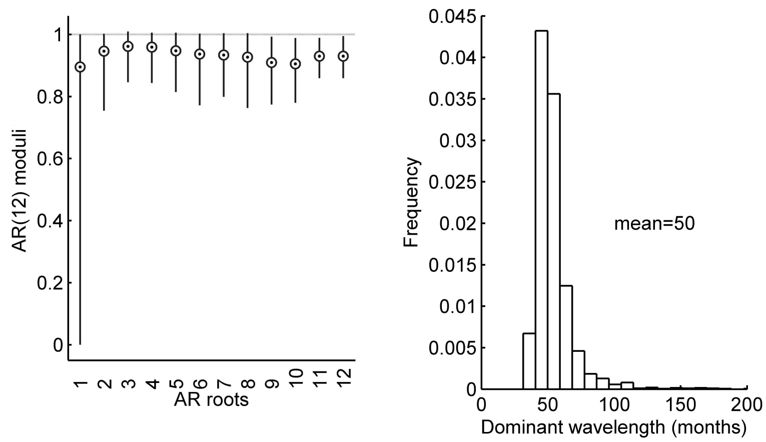


FIG. 1.1. *Left frame:* Approximate posterior 95% credible intervals for the moduli of the 12 latent AR roots in the AR component of the model fitted to the SOI time series. *Right frame:* Approximate posterior for the wavelength of the latent process component z^c_{tj} with largest wavelength, indicating a dominant quasi-periodicity in the range of 40-70 months.

Figures 1.1, 1.2 and 1.3 show aspects of an analysis of the climatological Southern Oscillation Index (SOI) time series. This is a series of 540 monthly observations computed as the “difference of the departure from the long-term monthly mean sea level pressures” at Tahiti in the South Pacific and Darwin in

Northern Australia. The index is one measure of the so-called "El Nino-Southern Oscillation"– an event of critical importance and interest in climatological studies in recent decades and that is generally understood to vary periodically with a very noisy 3-6 year period of quasi-cyclic pattern. As discussed in [24]– which also details the history of the data and prior analyses– one of several applied interests in this data is in improved understanding of these quasi-periodicities and also potential non-stationary trends, in the context of substantial levels of observational noise.

The DLM chosen here is $y_t = \theta_{t1} + z_t + \nu_t$ where θ_{t1} is a first-order polynomial local level/trend and z_t is an AR(12) process. The data is monthly data over the year, so the AR component provides opportunities to identify even quite subtle longer term (multi-year) periodicities that may show quite high levels of stochastic variation over time in amplitude and phase. Extensions to TVAR components

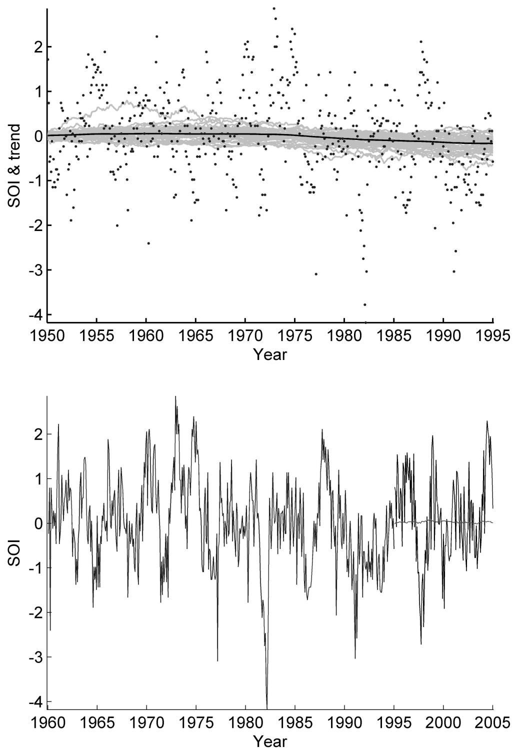


FIG. 1.2. *Upper frame:* Scatter plot of the monthly SOI index time series superimposed on the trajectories of the posterior mean and a few posterior samples of the underlying trend. *Lower frame:* SOI time series followed by a single synthetic future– a sample from the posterior predictive distribution over the three or fours years following the end of the data in 1995; the corresponding sample of the predicted underlying trend is also shown.

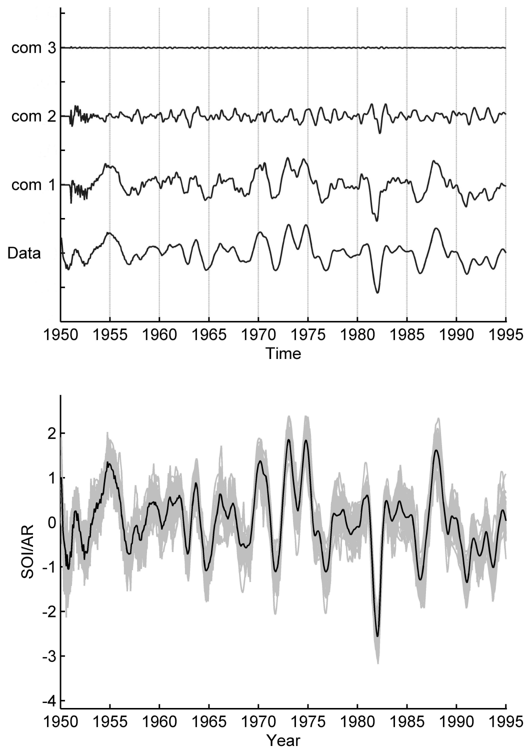


FIG. 1.3. Aspects of decomposition analysis of the SOI series. *Upper frame:* Posterior means of (from the bottom up) the latent AR(12) component z_t (labelled as "data"), followed by the three extracted component z^c_{tj} for $j = 1, 2, 3$, ordered in terms of decreasing estimated periods; all are plotted on the same vertical scale, and the AR(12) process is the direct sum of these three and subsidiary components. *Lower frame:* A few posterior samples (in grey) of the latent AR(12) process underlying the SOI series, with the approximate posterior mean superimposed.

would allow the associated periods to also vary as discussed and referenced above. Here the model parameters include the 12-dimensional AR parameter ϕ that can be converted to autoregressive roots (section 9.5 of [74]) to explore whether the AR component appears to be consistent with an underlying stationary process or not as well as to make inferences on the periods/wavelengths of any identified quasi-periodic components. The analysis also defines posterior inferences for the time trajectories of all latent components z^r_{tj} and z^c_{tj} by applying the decomposition theory to each of the posterior simulation samples of the state vector sequence $\theta_{0:n}$.

Figures 1.1 shows approximate posteriors for the moduli of the 12 latent AR roots, all very likely positive and almost surely less than 1, indicating stationarity of z_t in this model description. The figure also shows the corresponding posterior for the wavelength of the latent process component z_{tj}^c having highest wavelength, indicating a dominant quasi-periodicity in the data with wavelength between 40-70 months– a noisy "4-year" phenomenon, consistent with expectations and prior studies. Figure 1.2 shows a few posterior samples of the time trajectory of the latent trend θ_{t1} together with its approximate posterior mean, superimposed on the data. The inference is that of very limited change over time in the trend in the context of other model components. This figure also shows the data plotted together with a "synthetic future" over the next three years: that is, a single draw from the posterior predictive distribution into the future. From the viewpoint of model fit, exploring such synthetic futures via repeat simulations studied by eye in comparison with the data can be most informative; they also feed into formal predictive evaluations for excursions away from (above/below) the mean, for example [24].

Additional aspects of the decomposition analysis are represented by Figures 1.3. The first frame shows the posterior mean of the fitted AR(12) component plotted over time (labelled as "data" in the upper figure), together with the corresponding posterior mean trajectories of the three latent quasi-cyclical components having largest inferred periods, all plotted on the same vertical scale. Evidently, the dominant period component explains much of the structure in the AR(12) process, the second contributing much of the additional variation at a lower wavelength (a few months). The remaining components contribute to partitioning the noise in the series and have much lower amplitudes. The figure also shows several posterior draws for the z_t processes to give some indication of the levels of uncertainty about its form over the years.

1.3.3 *Mixture Model Enrichment of DLMs*

Mixture models have been widely used in dynamic modelling and remain a central theme in analyses of structural change, approaches to modelling non-Gaussian distributions via discrete mixture approximations, dealing with outlying observations, and others. Chapter 12 of [74] develops extensive theory and methodology of two classes of dynamic mixture models, building on seminal work by P.J. Harrison and others [23]. The first class relates to model uncertainty and learning model structure that has its roots in both commercial forecasting and engineering control systems applications of DLMs from the 1960s. Here a set of DLMs are analysed sequentially in parallel, being regarded as competing models, and sequentially updated "model probabilities" track the data-based evidence for each relative to the others in what is nowadays a familiar model comparison and Bayesian model-averaging framework.

The second framework– adaptive *multi-process* models– entertains multiple possible models at each time point and aims to adaptively reweight sequentially over time; key examples are modelling outliers and change-points in subsets of the state vector as in applications in medical monitoring, for example [56,55]. In the DLM of equation (1.2) with a "standard" model having $V(\nu_t) = v$ and evolution variance matrix as in equation (1.3), a multi-process extension for outlier accommodation would consider a mixture prior induced by $V(\nu_t) = k_t v$ where, *at each time* t, k_t may take the value 1 or, say, 100, with some probability. Similarly, allowing for a larger stochastic change in the underlying latent AR(2) component z_t of the model would involve an extension so that the innovations variance w in equation (1.3) is replaced by $h_t w$, where now h_t may take the value 1 or 100, with some probability. These multi-process models clearly lead to a combinatorial explosion of the numbers of possible "model states" as time progresses, and much attention has historically been placed on approximating the implied unwieldy sequential analysis. In the context of MCMC methods and batch analysis, this is resolved with simulation-based numerical approximations where the introduction of indicators of mixture component membership naturally and trivially opens the path to computation: models are reduced to *conditionally linear, normal DLMs* for conditional posterior simulations of states and parameters, and then the mixture component indicators are themselves re-simulated each step of the MCMC. Many more elaborate developments and applications appear in, and are referenced by, [19] and chapter 7 of [46].

Another use of mixtures in DLMs is to define direct approximations to non-normal distributions, so enabling MCMC analysis based on conditionally normal models they imply. One key example is the univariate stochastic volatility model pioneered pioneered by [53, 26] and that is nowadays in routine use to define components of more elaborate dynamic models for multivariate stochastic volatility time series approaches [1, 41, 2, 12, 36, 37]; see also chapter 7 of [46].

1.3.4 *Sequential Simulation Methods of Analysis*

A further related use of mixtures is as numerical approximations to the sequentially updated posterior distributions for states in non-linear dynamic models when the conditionally linear strategy is not available. This use of mixtures of DLMs to define adaptive sequential approximations to the filtering analysis by "mixing Kalman filters" [3, 11] has multiple forms, recently revisited with some recent extensions in [38]. Mixture models as direct posterior approximations, and as sequences of sequentially updated *importance sampling* distributions for non-linear dynamic models were pioneered in [63,65,64] and some of the recent developments build on this.

The adaptive, sequential importance sampling methods of [65] represented an approach to sequential simulation-based analysis developed at the same time as the approach that became known as *particle filtering* [21]. Bayesian sequential analysis in state-space models using "clouds of particles" in states and model parameters, evolving the particles through evolution equations that may be highly non-linear and non-Gaussian, and appropriately updating weights associated with particles to define approximate posteriors, has defined a fundamental change in numerical methodology for time series. Particle filtering and related methods of sequential Monte Carlo (SMC) [14, 7], including problems of parameter learning combined with filtering on dynamic states [31], are reviewed in this book: see the chapter by H.F. Lopes and C.M. Carvalho, on *Online*

Bayesian learning ....

Recent methods have used variants and extensions of the so-called technique of *approximate Bayesian computation* [34, 54]. Combined with other SMC methods, this seems likely to emerge in coming years as a central approach to computational approximation for sequential analysis in increasingly complex dynamic models; some recent studies in dynamic modelling in systems biology [35, 4, 58] provide some initial examples using such approaches.

1.4 Multivariate Time Series

The basic DLM framework generalizes to multivariate time series in a number of ways, including multivariate non-Gaussian models for time series of counts, for example [5], as well as a range of model classes based on multi- and matrix-variate normal models (chapter 10 of [46]). Financial and econometric applications have been key motivating areas, as touched on below, while multivariate DLMs are applied in many other fields– as diverse as experimental neuroscience [1,27,28,47], computer model emulation in engineering [30] and traffic flow forecasting [57]. Some specific model classes that are in mainstream application and underlie recent and current developments– especially to increasingly high-dimensional times series– are keyed out here.

1.4.1 *Multivariate Normal DLMs: Exchangeable Time Series*

In modelling and forecasting a $q \times 1$ vector times series, a so-called *exchangeable time series DLM* has the form

$$\begin{array}{ll} y_t' = F_t'\Theta_t + \nu_t', & \nu_t \sim N(0, \Sigma_t) \\ \Theta_t = G_t\Theta_{t-1} + \Omega_t, & \Omega_t \sim N(0, W_t, \Sigma_t) \end{array} \tag{1.4}$$

where $N(\cdot,\cdot,\cdot)$ denotes a *matrix normal distribution* (section 10.6 of [46]). Here the row vector y_t' follows a DLM with a matrix state Θ_t. The $q \times q$ time-varying variance matrix Σ_t determines patterns of co-changes in observation and the latent matrix state over time. These models are building blocks of larger (factor, hierarchical) models of increasing use in financial time series and econometrics; see, for example, [49, 48], chapter 16 of [74] and chapter 10 of [46].

Modelling multivariate stochastic volatility– the evolution over time of the variance matrix series Σ_t– is central to these multivariate extensions of DLMs. The first multivariate stochastic volatility models based on variance matrix discount learning [50, 51], later developed via matrix-beta evolution models [59, 60], remain central to many implementations of Bayesian forecasting in finance. Here Σ_t evolves over one time interval via a non-linear stochastic process model involving a matrix beta random innovation inducing priors and posteriors of conditional inverse Wishart forms. The conditionally conjugate structure of the exchangeable model form for $\{\Theta_t, \Sigma_t\}$, coupled with discount factor-based specification of the W_t evolution variance matrices, leads to a direct extension of the closed form sequential learning and retrospective sampling analysis of the univariate case (chapter 10 of [46]). In multiple studies, these models have proven their value in adapting to short-term stochastic volatility fluctuations and leading to improved portfolio decisions as a result [48].

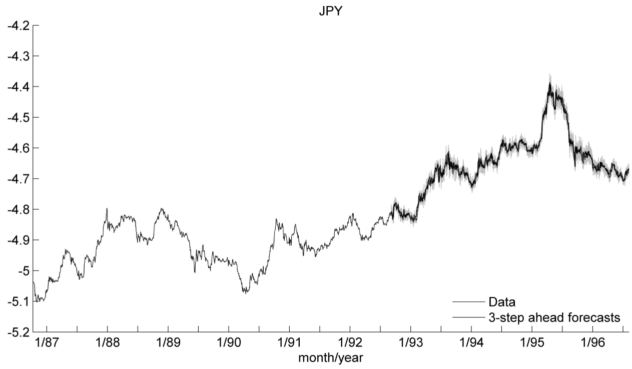
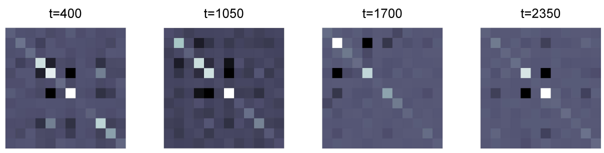


FIG. 1.4. Daily prices of the Japanese Yen in US dollars over several years in the 1980s-1990s, followed by plots of forecasts from the multivariate TV-TVAR model with stochastic volatility fitted to a 12-dimensional FX time series of which the Yen is one element. The shaded forecast region is made up of 75 sets of 3-day ahead forecasts based on the sequential analysis: on each day, the "current" posterior for model states and volatility matrices is simulated to generate forecasts over the next 3 days.

An example analysis of a time series of $q = 12$ daily closing prices (FX data) of international currencies relative to the US dollar, previously analyzed using different models (chapter 10 of [46]), generates some summaries including those in Figures 1.4 and 1.5. The model used here incorporates time-varying vector autoregressive (TV-VAR) models into the exchangeable time series structure. With y_t the logged values of the $12-$vector of currency prices at time $t,$ we take F_t to be the $37-$dimensional vector having a leading 1 followed by the lagged values of all currencies over the last three days. The dynamic autoregression naturally anticipates the lag-1 prices to be the prime predictors of next time prices, while considering 3 day lags leads to the opportunity to integrate "market momentum". Figure 1.4 selects one currency, the Japanese Yen, and plots the data together with forecasts over the last several years. As the sequential updating analysis proceeds, forecasts on day t for day $t + 1$ are made by direct simulation of the 1-step ahead predictive distribution; each forecast vector y_{t+1} is then used in the model in order to use the same simulation strategy to sample the future at time $t + 2$ from the current day $t,$ and this is repeated to simulate day $t + 3.$ Thus we predict via the strategy of generating synthetic realities, and the figure shows a few sets of these $3-$day ahead forecasts made every day over three or four years, giving some indication of forecast uncertainty as well as accuracy.

Figure 1.5 displays some aspects of multivariate volatility over time as inferred by the analysis. Four images of the posterior mean of the *precision matrix*



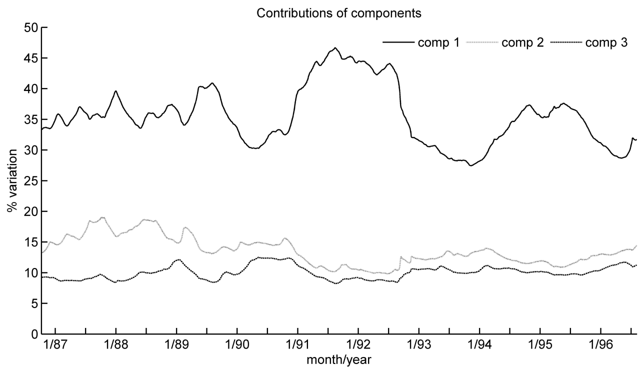


FIG. 1.5. *Upper frames:* Images of posterior estimates of the 12×12 precision matrices Σ_t^{-1} in the analysis of the multivariate currency prices time series. The differences in patterns visually evident reflect the extent and nature of changes in the volatility structure across time as represented by the four selected time points spaced apart by a few hundred days. *Lower frame:* Percent variation explained by the first three dominant principal components of the posterior mean of Σ_t for each time $t = 1 : n$ over the FX time series, illustrating the nature of variation in the contribution of the main underlying "common components" of volatility in the 12 currency price series over the ten year period.

Σ_t^{-1} at 4 selected time points capture some flavour of time variation, while the percent of the total variation in the posterior mean of Σ_t explained by the first three dominant principal components at each t captures additional aspects.

1.4.2 *Multivariate Normal DLMs: Dynamic Latent Factor and TV-VAR Models*

Time-varying vector autoregressive (TV-VAR) models define a rich and flexible approach to modelling multivariate structure that allows the predictive relationships among individual scalar elements of the time series to evolve over time. The above section already described the use of such a model within the exchangeable time series framework. Another way in which TV-VAR models are used is to represent the dynamic evolution of a vector of *latent factors* underlying structure in a higher-dimensional data series. One set of such *dynamic latent factor TV-VAR models* has the form

$$\begin{aligned} y_t &= F_t\theta_t + B_t x_t + \nu_t, && \nu_t \sim N(0,\Psi), \\ x_t &= \textstyle\sum_{i=1:p} A_{ti} x_{t-i} + \omega_t, && \omega_t \sim N(0,\Sigma_t). \end{aligned} \tag{1.5}$$

Here x_t is a latent $k-$vector state process following a TV-VAR(p) model with, typically, $k << q$ so that the common structure among the elements of y_t is heavily driven by a far lower-dimensional dynamic state. The set of $p,\ q \times q$ autoregressive coefficient matrices A_{ti} are often time-varying with elements modelled via, for example, sets of univariate $AR(1)$ processes or random walks. The factor loadings matrix B_t maps factors to data; in some models, including prior Bayesian factor analysis approaches [2], this will be taken as constant. The $k \times k$ dynamic covariance matrix Σ_t drives the innovations of the state evolution, and allowing for stochastic volatility here defines an enrichment of the TV-VAR structure. The additional component $F_t\theta_t$ superimposed may include dynamic regression and other terms, with relevant state evolution equations for θ_t. Such models are receiving increasing use in natural science and engineering applications [45, 75, 15, 16, 20, 46] as well as in econometrics and finance [36, 37]. Chapters 8 and 9 of [46] describe aspects of the theory and methodology of vector AR and TVAR models, connections with latent factor modelling in studies of multiple time series in the neurosciences, and discussion of multiple other creative applications of specialized variants of this rich class of models. The model contains traditional DLMs, VAR models, latent factor models as previously developed as well as the more elaborate TV-VAR factor forms.

1.5 Some Recent and Current Developments

Among a large number of recent and currently active research areas in Bayesian time series analysis and forecasting, a few specific modelling innovations that relate directly to the goals of addressing analysis of increasingly high-dimensional time series and non-linear models are keyed out.

1.5.1 *Dynamic Graphical and Matrix Models*

A focus on inducing parsimony in increasingly high-dimensional, time-varying variance matrices in dynamic models led to the integration of Bayesian *graphical modelling* ideas into exchangeable time series DLMs [10, 9]. The standard theory of Gaussian graphical models using hyper-inverse Wishart distributions– the conjugate priors for variance matrices whose inverses Σ_t^{-1} have some off-diagonal elements at zero corresponding to an underlying conditional independence graph [29]– rather surprisingly extends directly to the time-varying case. The multivariate volatility model based on variance matrix discounting generalizes to define sequential analysis in which the posterior distributions for the $\{\Theta_t, \Sigma_t\}$ sequences are updated in closed multivariate normal, hyper-inverse Wishart forms. These theoretical innovations led to the development of dynamic graphical models, coupled with learning about graphical model structures based on existing model search methods [13, 25]. Applications in financial time series for predictive portfolio analysis show improvements in portfolio outcomes that illustrate the practical benefits of the parsimony induced via appropriate graphical model structuring in multivariate dynamic modelling [10, 9].

These developments have extended to contexts of *matrix time series* [61] for applications in econometrics and related areas. Building on Bayesian analyses of matrix-variate normal distributions, conditional independence graphical structuring of the characterizing variance matrix parameters of such distributions again opens the path to parsimonious structuring of models for increasingly high-dimensional problems. This is complemented by the development of a broad class of dynamic models for matrix-variate time series within which stochastic elements defining time series errors and structural changes over time are subject to graphical model structuring.

1.5.2 *Dynamic Matrix Models for Stochastic Volatility*

A number of recent innovations have aimed to define more highly structured, predictive stochastic process models for multivariate volatility matrices Σ_t, aiming to go beyond the neutral, random walk-like model that underlies the discounting approach. Among such approaches are multivariate extensions of the univariate construction method inspired by MCMC [42]; the first such extension yields a class of stationary AR(1) stochastic process models for Σ_t that are reversible in time and in which the transition distributions give conditional means of the attractive form $E(\Sigma_t|\Sigma_{t-1}) = S + a(\Sigma_{t-1} - S)$ where $a \in (0, 1)$ is scalar and S an underlying mean variance matrix. This construction is, however, inherently limited in that there is no notion of multiple AR coefficients for flexible autocorrelation structures and the models do not allow time irreversibilty. Related approaches directly build transition distributions $p(\Sigma_t|\Sigma_{t-1})$ as inverse-Wisharts [40, 39] or define more empirical models representing Σ_t as an explicit function of sample covariance matrices of latent vector AR processes [22]. These are very interesting approaches but are somewhat difficult to work with theoretically and model fitting is a challenge.

Recently, [32] used linear, normal AR(1) models for off-diagonal elements of the Cholesky of Σ_t and for the log-diagonal elements. This is a natural parallel of Bayesian factor models for multivariate volatility and defines an approach to building highly structured stochastic process models for time series of dynamic variance matrices with short-term predictive potential.

A related approach builds on theoretical properties of the family of inverse Wishart distributions to define new classes of stationary, *inverse Wishart autoregressive (IW-AR)* models for the series of $q \times q$ volatility matrices Σ_t [17]. One motivating goal is to maintain a defined inverse Wishart marginal distribution for the process for interpretation. Restricting discussion to the (practically most interesting) special case of a first-order model, the basic idea is to define an IW-AR(1) Markov process directly via transition densities $p(\Sigma_t|\Sigma_{t-1})$ that are the conditionals of a joint inverse Wishart on an augmented $2q \times 2q$ variance matrix and whose block diagonals are Σ_{t-1} and Σ_t. This yields

$$\Sigma_t = \Psi_t + \Upsilon_t \Sigma_{t-1} \Upsilon_t'$$

where the $q \times q$ random *innovations matrices* Υ_t and Ψ_t have joint matrix normal, inverse Wishart distributions independently over time. Conditional means have the form

$$E(\Sigma_t | \Sigma_{t-1}) = S + R(\Sigma_{t-1} - S)R' + C_t(\Sigma_{t-1})$$

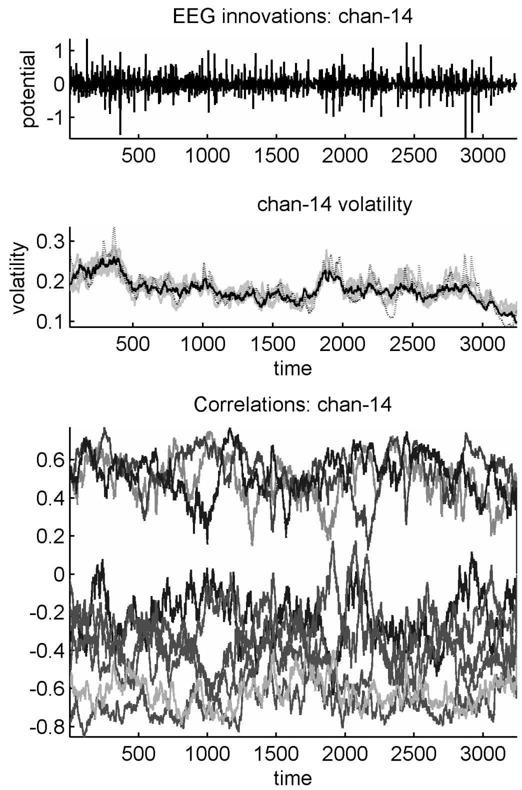


FIG. 1.6. Aspects of results of approximate fitting a $q \times q$ dimensional IW-AR(1) model to $q = 10$ EEG series from [27, 47]; here Σ_t is the volatility matrix of innovations driving a TV-VAR model for the potential fluctuations the EEG signals represent. *Upper frame:* Estimated innovations time series for one EEG series/channel, labeled *chan-14*. *Centre frame:* Several posterior sample trajectories (grey) and approximate mean (black) for the standard deviation of channel 14. *Lower frame:* Corresponding estimates of time-varying correlations of *chan-14* with the other channels. Part of the applied interest is in patterns of change over time in these measures as the EEG channels are related spatially on the scalp of the test individual.

where S is the mean variance matrix parameter of the stationary process, R is a $q \times q$ autoregressive parameter matrix and $C_t(\cdot)$ is a matrix naturally related to the skewness of the inverse Wishart model. This model has the potential to embody multiple aspects of conditional dependence through R as well as defining both reversible and irreversible special cases [17]. Some initial studies have explored use of special cases as models for volatility matrices of the innovations process driving a TV-TVAR model for multiple EEG time series from studies in experimental neuroscience. One small extract from an analysis of multi-channel EEG data [27] appears in Figure 1.6, showing aspects of the estimated time trajectories of volatility for one channel along with those of time-varying correlations from Σ_t across multiple channels. As with other models above, computational issues for model filtering, smoothing and posterior simulation analysis require customized MCMC and SMC methods, and represent some of the key current research challenges. The potential is clear, however, for these approaches to define improved representations of multivariate volatility processes of benefit when integrated into time series state space analysis.

1.5.3 *Time-Varying Sparsity Modelling*

As time series dimension increases, the dimension of latent factor processes, time-varying parameter processes and volatility matrix processes in realistic dynamic models– such as special cases or variants of models of equation (1.5)– evidently increase very substantially. Much current interest then rests on modelling ideas that engender parsimonious structure and, in particular, on approaches to inducing data-informed *sparsity* via full shrinkage to zero of (many) parameters. Bayesian sparsity modelling ideas are well-developed in "static" models, such as sparse latent factor and regression models [71, 8], but mapping over to time series raises new challenges of defining general approaches to *dynamic sparsity*. For example, with a dynamic latent factor component $B_t f_t$ of equation (1.5), a zero element $B_{t,(i,j)}$ in the factor loadings matrix B_t reflects lack of association of the i^{th} series in y_t with the j^{th} latent factor in f_t. The overall sparsity pattern of B_t– with potentially many zeros– reflects a model context in which each of the individual, univariate factor processes impacts on a subset of the output time series, but not all, and allows for complex patterns of cross-talk. The concept of dynamic sparsity is that these sparsity patterns will typically vary over time, so models are needed to allow time-variation in the values of elements of B_t that can dynamically shrink completely to zero for some epochs, then reappear and evolve according to a specific stochastic model at others. A general approach has been introduced by [36, 37], referred to as *latent threshold modelling* (LTM).

The basic idea of LTM for time series is to embed traditional time series model components into a larger model that thresholds the time trajectories, setting their realized values strictly to zero when they appear "small". For example, take one scalar coefficient process β_t, such as one element of the factor loadings matrix or a single dynamic regression parameter, and begin with a traditional evolution model of the AR(1) form $\beta_t = \mu + \rho(\beta_{t-1} - \mu) + \epsilon_t$. The LTM

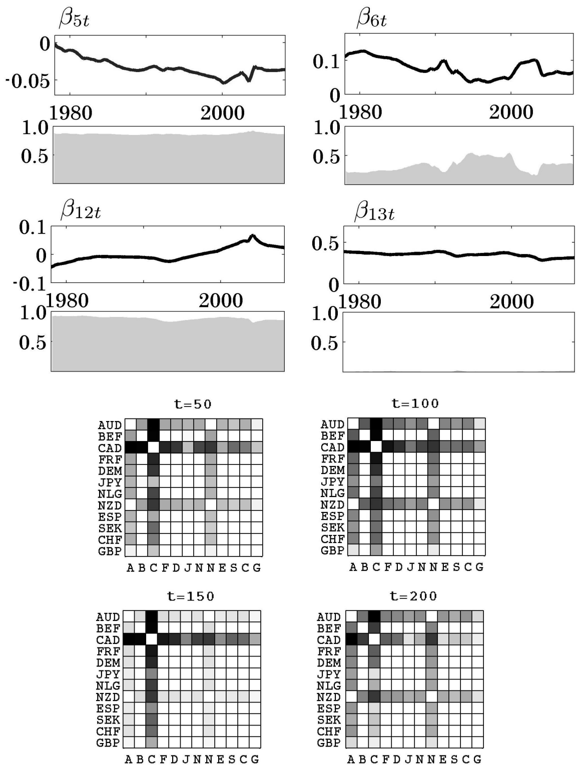


FIG. 1.7. Examples of latent thresholding from an analysis of a 3-dimensional economic time series using a TV-VAR(2) model (data from [36]). *Upper four frames:* Trajectories of approximate posterior means of 4 of the (18) *latent* time-varying autoregressive coefficients; the grey shading indicates estimated time-varying posterior probabilities of zero coefficients from the LTM construction. *Lower four images:* Images showing estimates of posterior probabilities (white = high, black = low) of non-zero entries in dynamic precision matrices Σ_t^{-1} modelled using an LTM extension of the Cholesky AR(1) model [32]. The data are time series on $q = 12$ daily international exchange rates (data from [46]) and the images show posterior sparsity probabilities for the 12×12 matrices at four selected time points, indicating both the ability of the LTM to identify zeros as well as how the sparsity pattern changes over time based on latent thresholding.

approach replaces the sequence $\beta_{1:n}$ with the thresholded version $b_{1:n}$ where $b_t = \beta_t I(|\beta_t| < \tau)$ for each t and based on some threshold τ. The concept is simple: the coefficient process is relevant, taking non-zero values, only when it beats the threshold, otherwise it is deemed insignificant and shrunk to zero. Extending MCMC analyses of multivariate DLMs to integrate the implied hierarchical model components, now embedding mutliple latent processes underlying the actual thresholded parameter processes, states and factors, requires substantial computational development, as detailed in [36]. The payoffs can be meaningful, as demonstrated in a series of financial time series and portfolio decision making examples in [37] where improved fit and parsimony feeds through to improvements in short-term forecasting and realized portfolio returns.

Figure 1.7 gives some flavour of the approach in extracts from two time series analyses: a TV-VAR(2) model of a $q = 3$-dimensional economic time series, and a multivariate stochastic volatility model analysis of a $q = 12$-dimensional financial time series. One key attraction of the LTM approach is its generality. Some of the model contexts addressed via LTM ideas in [36] include the following: *(i)* dynamic latent factor models; *(ii)* TV-VAR models, where the dynamic sparsity arises in collections of TV-VAR coefficient matrices A_{ti} of equation (1.5); *(iii)* dynamic regressions, where the approach can be regarded as a model for dynamic variable selection as well as a parsimony inducing strategy; and *(iv)* dynamic volatility modelling using extensions of the Cholesky volatility models of [32].

1.5.4 *Nonlinear Dynamical Systems*

The recent advances in Bayesian computational methods for dynamic models have come at a time when biotechnology and computing are also promoting significant advances in formal modelling in systems biology at molecular and cellular levels. In models of temporal development of components of gene networks and in studies of systems of cells evolving over time (and space, e.g. [33]), increasingly complex, multivariate non-linear *mechanistic* models are being expanded and explored; these come from both the inherently stochastic biochemical modelling perspective and from the applied mathematical side using systems of coupled (ordinary or stochastic) differential equations [76, 38, 77, 78].

Statistical model development naturally involves discrete-time representations with components that realistically reflect stochastic noise and measurement error and inherently involve multiple underlying latent processes representing unobserved states that influence the network or cellular system. A specific class of models has a multivariate time series y_* modelled as

$$\begin{array}{ll|l} y_j & = x_{t_j} + \nu_j & p(y_j|x_{t_j}, \Theta) \\ x_{t+h} & = x_t + G_h(x_t, \Theta)x_t + g_h(\Theta) + \omega_{t,h} & p(x_{t+h}|x_t, \Theta) \end{array}$$

where the j^{th} observation comes at real-time t_j and the spacings between consecutive observations are typically far greater than the fine time step h. Here x_t represents the underlying state vector of the systems (levels of gene or protein expression, numbers of cells, etc.), Θ all model parameters, ν_* measurement error and ω_* state evolution noise. The density forms to the right indicate more general model forms in which errors and noise may not be additive, when only partial observation is made on the state, and so forth.

The forefront research challenges in this area include development of efficient and effective computations for posterior inference and model comparison. Increasingly, SMC methods including SMC/importance sampling and ABC-SMC are being explored and evaluated, with models of increasing dimension and complexity [35, 52, 58, 4, 77]. In coming years, complex, multivariate dynamical systems studies in biology are sure to define a major growth area for Bayesian dynamic models and time series analysis.

Acknowledgement

The author is grateful to Ioanna Manolopoulou and Emily Fox for comments on this chapter. This work was partly supported by grant DMS-1106516 from the U.S. National Science Foundation (NSF), and grants P50-GM081883 and RC1-AI086032 of the U.S National Institutes of Health. Any opinions, findings and conclusions or recommendations expressed in this work are those of the author and do not necessarily reflect the views of the NSF or the NIH.

REFERENCES

[1] Aguilar, O., Prado, R., Huerta, G., and West, M. (1999). Bayesian inference on latent structure in time series (with discussion). In *Bayesian Statistics 6* (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith), pp. 3–26. Oxford University Press, Oxford.

[2] Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics*, **18**, 338–357.

[3] Alspach, D. L. and Sorenson, H. W. (1972). Non-linear Bayesian estimation using Gaussian sum approximations. *IEEE Transactions on Automatic Control*, **AC-17**, 439–448.

[4] Bonassi, F. V., You, L., and West, M. (2011). Bayesian learning from marginal data in bionetwork models. *Statistical Applications in Genetics & Molecular Biology*, **10**, Art 49.

[5] Cargnoni, C., Müller, P., and West, M. (1997). Bayesian forecasting of multinational time series through conditionally Gaussian dynamic models. *Journal of the American Statistical Association*, **92**, 640–647.

[6] Carter, C. K. and Kohn, R. (1994). Gibbs sampling for state space models. *Biometrika*, **81**, 541–553.

[7] Carvalho, C. M., Johannes, M., Lopes, H. F., and Polson, N. G. (2010). Particle learning and smoothing. *Statistical Science*, **25**, 88106.

[8] Carvalho, C. M., Lucas, J. E., Wang, Q., Chang, J., Nevins, J. R., and West, M. (2008). High-dimensional sparse factor modelling - Applications in gene expression genomics. *Journal of the American Statistical Association*, **103**, 1438–1456.

[9] Carvalho, C. M. and West, M. (2007). Dynamic matrix-variate graphical models. *Bayesian Analysis*, **2**, 69–98.

[10] Carvalho, C. M. and West., M. (2007). Dynamic matrix-variate graphical models - A synopsis. In *Bayesian Statistics 8* (ed. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West), pp. 585–590. Oxford University Press.

[11] Chen, R. and Liu, J. S. (2000). Mixture Kalman filters. *Journal of the Royal Statistical Society B*, **62**(3), 493–508.

[12] Chib, S., Omori, Y., and Asai, M. (2009). Multivariate stochastic volatility. In *Handbook of Financial Time Series* (ed. T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch), New York, pp. 365–400. Springer-Verlag.

[13] Dobra, A., Jones, B., Hans, C., Nevins, J. R., and West, M. (2004). Sparse graphical models for exploring gene expression data. *Journal of Multivariate Analysis*, **90**, 196–212.

[14] Doucet, A., de Freitas, N., and Gordon, N. J. (2001). *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York.

[15] Doucet, A., Godsill, S. J., and West, M. (2000). Monte Carlo filtering and smoothing with application to time-varying spectral estimation. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, **II**, 701–704.

[16] Fong, W., Godsill, S. J., Doucet, A., and West, M. (2002). Monte Carlo smoothing with application to speech enhancement. *IEEE Trans. Signal Processing*, **50**, 438–449.

[17] Fox, E. B. and West, M. (2011). Autoregressive models for variance matrices: Stationary inverse Wishart processes. Discussion Paper 11-15, Department of Statistical Science, Duke University. Submitted for publication.

[18] Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15**, 183–202.

[19] Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer-Verlag, New York.

[20] Godsill, S. J., Doucet, A., and West, M. (2004). Monte Carlo smoothing for non-linear time series. *Journal of the American Statistical Association*, **99**, 156–168.

[21] Gordon, N. J., Salmond, D., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Proc. IEE F*, **140**, 107–113.

[22] Gourieroux, C., Jasiak, J., and Sufana, R. (2009). The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, **150**, 167–181.

[23] Harrison, P. J. and Stevens, C. (1976). Bayesian forecasting (with discussion). *Journal of the Royal Statististical Society, Series B*, **38**, 205–247.

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

[25] Jones, B., Dobra, A., Carvalho, C. M., Hans, C., Carter, C., and West, M. (2005). Experiments in stochastic computation for high-dimensional graphical models. *Statistical Science*, **20**, 388–400.

[26] Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, **65**, 361–393.

[27] Krystal, A. D., Prado, R., and West, M. (1999). New methods of time series analysis for non-stationary EEG data: Eigenstructure decompositions of time varying autoregressions. *Clinical Neurophysiology*, **110**, 1–10.

[28] Krystal, A. D., Zoldi, S., Prado, R., Greenside, H. S., and West, M. (1999). The spatiotemporal dynamics of generalized tonic-clonic seizure EEG data: Relevance to the climinal practice of electroconvulsive therapy. In *Nonlinear Dynamics and Brain Functioning* (ed. N. Pradhan, P. Rapp, and R. Sreenivasan). New York: Novascience.

[29] Lauritzen, S. L. (1996). *Graphical Models*. Clarendon Press, Oxford.

[30] Liu, F. and West, M. (2009). A dynamic modelling strategy for Bayesian

computer model emulation. *Bayesian Analysis*, **4**, 393–412.

[31] Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo Methods in Practice* (ed. A. Doucet, J. D. Freitas, and N. Gordon), pp. 197–217. New York: Springer-Verlag.

[32] Lopes, H. F., McCulloch, R. E., and Tsay, R. (2010). Cholesky stochastic volatility. Technical report, University of Chicago, Booth Business School.

[33] Manolopolou, I., Matheu, M. P., Cahalan, M. D., West, M., and Kepler, T. B. (2012). Bayesian spatio-dynamic modelling in cell motility studies: Learning nonlinear taxic fields guiding the immune response (with invited discussion). *Journal of the American Statistical Association*, **107**. 10.1080/01621459.2012.655995.

[34] Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences USA*, **100**, 15324–15328.

[35] Mukherjee, C. and West, M. (2009). Sequential Monte Carlo in model comparison: Example in cellular dynamics in systems biology. In *JSM Proceedings,Section on Bayesian Statistical Science. Alexandria, VA: American Statistical Association*, pp. 1274–1287.

[36] Nakajima, J. and West, M. (2012). Bayesian analysis of latent threshold dynamic models. *Journal of Business & Economic Statistics*. To appear.

[37] Nakajima, J. and West, M. (2012). Bayesian dynamic factor models: Latent threshold approach. *Journal of Financial Econometrics*. 10.1093/jjfinec/nbs013.

[38] Niemi, J. B. and West, M. (2010). Adaptive mixture modelling Metropolis methods for Bayesian analysis of non-linear state-space models. *Journal of Computational and Graphical Statistics*, **19**, 260–280.

[39] Philipov, A. and Glickman, M. E. (2006). Factor multivariate stochastic volatility via Wishart processes. *Econometric Reviews*, **25**, 311–334.

[40] Philipov, A. and Glickman, M. E. (2006). Multivariate stochastic volatility via Wishart processes. *Journal of Business and Economic Statistics*, **24**, 313–328.

[41] Pitt, M. K. and Shephard, N. (1999). Time varying covariances: A factor stochastic volatility approach (with discussion). In *Bayesian Statistics VI* (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith), Oxford, pp. 547–570. Oxford University Press.

[42] Pitt, M. K. and Walker, S. G. (2005). Constructing stationary time series models using auxilliary variables with applications. *Journal of the American Statistical Association*, **100**, 554–564.

[43] Pole, A., West, M., and Harrison, P. J. (1994). *Applied Bayesian Forecasting & Time Series Analysis*. Chapman-Hall.

[44] Prado, R., Huerta, G., and West, M. (2001). Bayesian time-varying autoregressions: Theory, methods and applications. *Resenhas*, **4**, 405–422.

[45] Prado, R. and West, M. (1997). Exploratory modelling of multiple non-

stationary time series: Latent process structure and decompositions. In *Modelling Longitudinal and Spatially Correlated Data* (ed. T. Gregoire), pp. 349–362. Springer-Verlag.

[46] Prado, R. and West, M. (2010). *Time Series: Modeling, Computation & Inference*. Chapman & Hall/CRC Press.

[47] Prado, R., West, M., and Krystal, A. D. (2001). Multi-channel EEG analyses via dynamic regression models with time-varying lag/lead structure. *Journal of the Royal Statistical Society (Ser. C)*, **50**, 95–110.

[48] Quintana, J. M., Carvalho, C. M., Scott, J., and Costigliola, T. (2010). Futures markets, Bayesian forecasting and risk modeling. In *The Handbook of Applied Bayesian Analysis* (ed. A. O'Hagan and M. West), pp. 343–365. Oxford University Press, Oxford.

[49] Quintana, J. M., Lourdes, V., Aguilar, O., and Liu, J. (2003). Global gambling. In *Bayesian Statistics VII* (ed. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West), Oxford, pp. 349–368. Oxford University Press.

[50] Quintana, J. M. and West, M. (1987). An analysis of international exchange rates using multivariate DLMs. *The Statistician*, **36**, 275–281.

[51] Quintana, J. M. and West, M. (1988). Time series analysis of compositional data. In *Bayesian Statistics 3* (ed. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith), pp. 747–756. Oxford University press.

[52] Secrier, M., Toni, T., and Stumpf, M. P. H. (2009). The ABC of reverse engineering biological signalling systems. *Molecular Biosystems*, **5(12)**, 1925–1935.

[53] Shephard, N. (1994). Local scale models: State-space alternative to integrated GARCH models. *Journal of Econometrics*, **60**, 181–202.

[54] Sisson, S. A., Fan, Y., and Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences USA*, **104**, 1760–1765.

[55] Smith, A. F. M. and West, M. (1983). Monitoring renal transplants: An application of the multi-process Kalman filter. *Biometrics*, **39**, 867–878.

[56] Smith, A. F. M., West, M., Gordon, K., Knapp, M. S., and Trimble, I. (1983). Monitoring kidney transplant patients. *The Statistician*, **32**, 46–54.

[57] Tebaldi, C., West, M., and Karr, A. F. (2002). Statistical analyses of freeway traffic flows. *Journal of Forecasting*, **21**, 39–68.

[58] Toni, T. and Stumpf, M. P. H. (2010). Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics*, **26**, 104–110.

[59] Uhlig, H. (1994). On singular Wishart and singular multivariate beta distributions. *Annals of Statistics*, **22**, 395–405.

[60] Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica*, **1**, 59–73.

[61] Wang, H. and West, M. (2009). Bayesian analysis of matrix normal graphical models. *Biometrika*, **96**, 821–834.

[62] West, M. (1986). Bayesian model monitoring. *Journal of the Royal Statistical Society (Ser. B)*, **48**, 70–78.

[63] West, M. (1992). Modelling with mixtures (with discussion). In *Bayesian Statistics 4* (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith), pp. 503–524. Oxford University Press.

[64] West, M. (1993). Approximating posterior distributions by mixtures. *Journal of the Royal Statistical Society (Ser. B)*, **54**, 553–568.

[65] West, M. (1993). Mixture models, Monte Carlo, Bayesian updating and dynamic models. *Computing Science and Statistics*, **24**, 325–333.

[66] West, M. (1995). Bayesian inference in cyclical component dynamic linear models. *Journal of the American Statistical Association*, **90**, 1301–1312.

[67] West, M. (1996). Bayesian time series: Models and computations for the analysisd of time series in the physical sciences. In *Maximum Entropy and Bayesian Methods 15* (ed. K. Hanson and R. Silver), pp. 23–34. Kluwer.

[68] West, M. (1996). Some statistical issues in Palæoclimatology (with discussion). In *Bayesian Statistics 5* (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith), pp. 461–486. Oxford University Press.

[69] West, M. (1997). Modelling and robustness issues in Bayesian time series analysis (with discussion). In *Bayesian Robustness* (ed. J. O. Berger, B. Betrò, E. Moreno, L. R. Pericchi, F. Ruggeri, G. Salinetti, and L. Wasserman), IMS Monographs, pp. 231–252. Institute for Mathematical Statistics.

[70] West, M. (1997). Time series decomposition. *Biometrika*, **84**, 489–494.

[71] West, M. (2003). Bayesian factor regression models in the "large p, small n" paradigm. In *Bayesian Statistics 7* (ed. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. David, D. Heckerman, A. F. M. Smith, and M. West), pp. 723–732. Oxford University Press.

[72] West, M. and Harrison, P. J. (1986). Monitoring and adaptation in Bayesian forecasting models. *Journal of the American Statistical Association*, **81**, 741–750.

[73] West, M. and Harrison, P. J. (1989). Subjective intervention in formal models. *Journal of Forecasting*, **8**, 33–53.

[74] West, M. and Harrison, P. J. (1997). *Bayesian Forecasting & Dynamic Models* (2nd edn). Springer Verlag.

[75] West, M., Prado, R., and Krystal, A. D. (1999). Evaluation and comparison of EEG traces: Latent structure in non-stationary time series. *Journal of the American Statistical Association*, **94**, 1083–1095.

[76] Wilkinson, D. J. (2006). *Stochastic Modelling for Systems Biology*. London: Chapman & Hall/CRC.

[77] Wilkinson, D. J. (2011). Parameter inference for stochastic kinetic models of bacterial gene regulation: a Bayesian approach to systems biology (with discussion). In *Bayesian Statistics 9* (ed. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. David, D. Heckerman, A. F. M. Smith, and M. West), pp. 679–700. Oxford University Press.

[78] Yao, G., Tan, C., West, M., Nevins, J. R., and You, L. (2011). Origin of bista-

bility underlying mammalian cell cycle entry. *Molecular Systems Biology*, **7**, 485.