An Approximation Algorithm for the Updating of Non-Gaussian Dynamic Processes

A. S. Gargoum
Department of Statistics, College of Business & Economics, UAE University, United Arab Emirates (alig@uaeu.ac.ae)

Abstract: In the environmental issue of forecasting the geographical spread of the release of toxic gases in the event of a chemical or a nuclear accident, puffs of contaminated masses (uncertain quantities) emitted from a source, dispersed by a wind field and fragment into other puffs over time. The number of the fragmented puffs grows enormously. The problem here is to produce realistic estimates of contamination concentration in space and time. The atmospheric dispersion follows the Markovian property, that is the distribution of future puff fragments will depend only on the joint distribution of puffs currently existing. In such complex high-dimensional environments, that change dynamically, the vector of puff masses existing on or before time T can be set as a states vector $\boldsymbol{\Theta_T} = (\boldsymbol{\theta_1}, \ldots \boldsymbol{\theta_T})$. These states can be represented on an undirected acyclic graph whose cliques, that are formed by joining nodes with their neighbors if these neighbors have an edge pointing to them, contain components of $\boldsymbol{\Theta}_T$. If we assume that the observations are Gaussian, then well established fast data propagation algorithms for Bayesian networks can be used. In this paper I suggest an approximation methodology to accommodate non-Gaussian distributions using a slight generalization of the class of dynamic generalized linear models. The algorithm is very fast and updating is achieved in a closed form.

Keywords: Puff Models; Bayesian networks; Dynamic Generalized Linear Models

1 INTRODUCTION

The last decade has witnessed efficient applications of Bayesian probabilistic networks in complex high-dimensional problems provided that the underlying relationships between variables in the problem can be described by a fixed conditional independence structure. A detailed description may be found in Dawid [1992] and Lauritzen [1992]. However, there are many dynamic situations where such relationships evolve with time [see, e.g., Boyen and Koller, 1998]. In such situations an evolution of structure is inevitable. Learning in dynamic contexts is different from those where relationships among variables are static [Kjaerulff, 1992]. To model such dynamic processes, Bayesian probabilistic networks need to be defined over state spaces. Here we consider dynamic processes which can be described in terms of parametric models. The model parameters or states (uncertain quantities) are denoted by the vector $\boldsymbol{\Theta_T} = (\boldsymbol{\theta_1}, \ldots \boldsymbol{\theta_T})$. The defining parameter $\boldsymbol{\Theta}_T$ is indexed by T to indicate that the parameterizations may be dynamic. One of the problems that we might face is that $\boldsymbol{\Theta}_T$ may expand with the passage of time where new variables may be added and new conditional independencies are created. Moreover, in many applications, it is very important to produce forecast distributions at each time step. The model states $\boldsymbol{\Theta_T} = (\boldsymbol{\theta_1}, \ldots \boldsymbol{\theta_T})$ provide the means by which information relevant to forecasting future is summarized and used in forming the forecast distribution. The learning process sequentially revises the uncertainty about the parameters, by adjusting the probability distribution of the states. So the probability distributions on the state space are updated sequentially in the light of new observations. These constrains indicate that we need a graphical representation of the dependence structure that should change dynamically according to the evolution of the state space. Such representation provides an effective means for performing efficient and quick propagation algorithm (computational efficiency) that can help in calculating the posterior distribution of the states in the light of new data. Smith and Papamichail [1999] formalized how probabilistic propagation algorithms can be applied to high dimensional dynamic processes. Settimi et al. [1999] proposed a general framework for fast approximate Bayesian algorithms for learning in

complex high-dimensional environments, where the relationships between the explanatory state space variables can be described by a Bayesian network evolving dynamically over time, and where the observations taken are not Gaussian. These algorithms are based on generalization of a time series, first introduced by West, Harrison, and Migon [1985] and then adapted in Smith [1992]. Settimi and Smith [2000] used an MCMC-based approach to study the efficiency, in predicting future observations, of these algorithms developed by Settimi et al. [1999]. Their results–using an example on the Poisson dynamic system– show that the approximate algorithm can be extremely efficient and have the practical advantage of being quick in updating probabilities. An important practical example of the implementation of the proposed approximate algorithm is the forecasting of the geographical spread of the release of toxic gases in the event of a nuclear or chemical accident [see, e.g., Smith and French, 1993; Gargoum, 2001], in the light of various measurements that might be taken. Puffs of contaminated masses (uncertain quantities) are emitted from a source dispersed by a wind field and fragment into other puffs over time. The wind field, mass release and fragmentation process follow a complicated physical model. For detailed descriptions of puff models [see Mikkelsen et al., 1984; Thykier-Nielsen and Mikkelsen, 1991]. The structure of the problem is inherently high-dimensional and solutions based on stochastic numerical techniques such as MCMC [see, e.g., Smith and Roberts, 1993; George et el., 1994] would not be appropriate because probability prediction need to be continually updated very fast and the prior information need to be coded in the system on-line.

In this work, I investigate the use of non-normal distributions, in particular the lognormal distributions to describe the updating process. I implement the approximate algebraic algorithm in the lognormal case using the above environmental example as a high-dimensional complex dynamic process. Lognormal distributions have been used extensively in atmospheric sciences to describe phenomena that take on non-negative values such as particle size distribution and pollutant concentrations [Row, 1988]. It is often natural to assume that the distribution of observations conditional on their states is lognormal. In section 2 of the paper I give a brief background about the dynamic systems, how they can be represented by a directed acyclic graph DAG $\mathcal{G}$ and consequently how to construct a junction tree from DAG $\mathcal{G}$. Section 3 introduces the non-Gaussian dynamic systems. In section 4, I discuss the approximate propagation procedure using the lognormal dynamic models. Section 5 presents a metric for testing the validity of the approximation.

2 BACKGROUND MATERIAL

One way to address the computational efficiency problem at any given time T is to store the distribution of puff masses on a graph $\mathcal{G}$. Thus at time T, the probability distribution over the states (puff masses) $\boldsymbol{\Theta_T} = (\boldsymbol{\theta_1}, \ldots \boldsymbol{\theta_T})$, specifying the dynamic system till time T, is represented as a directed acyclic graph DAG $\mathcal{G}$ whose nodes are the parameters (states) in $\boldsymbol{\Theta}_T$. The well known propagation algorithms defined over junction trees $\mathcal{T}$ can be employed. [see, e.g., Lauritzen and Spiegelhalter, 1988; Spiegelhalter et al., 1993 and Jensen et al., 1994]. A junction tree can be constructed from a DAG $\mathcal{G}$ by firstly connecting all parents (nodes) in $\mathcal{G}$ that have the same child (node). Secondly identifying a sequence of cliques in $\mathcal{G}$, where a clique is a maximally connected subset of nodes. The cliques $C[1], \ldots, C[m]$, say, can be ordered [see, e.g., Tarjan and Yannakaskis, 1984], so that they satisfy the so called running intersection property [Lauritzen et al., 1984]. This property states that: given the cliques $C[1], \ldots, C[m]$, there exists an ordering of the cliques such that for all $2 \leq i \leq m$ a separator $S(i, b_i)$ is defined by

$$S(i, b_i) = C[i] \cap [\cup_{j=1}^{i-1} C[j]] \subseteq C[i] \cap C[b_i]$$

where $C[b_i]$ is any clique listed before $C[i]$ i.e. b_i is an index such that $1 \leq b_i \leq i$. Thirdly constructing the undirected graph (junction tree) whose nodes are the cliques $C[1], \ldots, C[m]$ where the clique $C[i]$ for $1 \leq i \leq m$, is connected to the clique $C[b_i]$ with an undirected edge if $S[i, b_i] = \phi$. In the dynamic scenario, at each time T the joint distribution $p(\boldsymbol{\Theta}_T)$ is stored in terms of a junction tree $\mathcal{T}$ associated with a Bayesian network $\mathcal{G}$ over the states $\boldsymbol{\Theta}_T$ and the set of cliques $\mathcal{C}_T = \{C_T[1], \ldots, C_T[m_T]\}$. The density $p(\boldsymbol{\Theta}_T)$ is said to be decomposable if it can be written as

$$p(\boldsymbol{\Theta}_T) = \frac{\prod_{i=1}^{m_T} p_i(\boldsymbol{\Theta}_T[i])}{\prod_{i=2}^{m_T} q_i(\boldsymbol{\Theta}_T[i, b_i])}$$

where p_i and q_i are densities of $\boldsymbol{\Theta}_T[i]$ and $\boldsymbol{\Theta}_T[i, b_i]$ respectively such that $\boldsymbol{\Theta}_T[i]$ is a subvector of $\boldsymbol{\Theta}_T$ whose components lie in the clique $C[i]$, $1 \leq i \leq m_T$. The vector $\boldsymbol{\Theta}_T[i, b_i]$ is a subvector of $\boldsymbol{\Theta}_T[i]$ and $\boldsymbol{\Theta}_T[b_i]$ for $1 \leq b_i \leq i$ and its components are in the separator $S_T(i, b_i)$ contained in the cliques $C_T[i]$ and $C_T[b_i]$.

Now, in this complex system, if the sampling distribution of the observations is Gaussian, the well

known fast propagation algorithms over dynamic junction trees can be used. In this case information can be transmitted through the junction tree by updating the probabilities of each clique sequentially very fast and in closed form Smith et al.[1995]. However, in the non-linear case, when the sampling distribution of the observations given the states is not Gaussian, then the posterior distribution of the states cannot be determined in closed form but the conditional independence relationships among variables still hold and hence the junction tree remains valid. In the following sections, I shall present an approximate algebraic propagation algorithm which is related to the dynamic generalized linear model and is useful when the sampling distribution given the states is non-Gaussian. I will consider the case where the sampling distribution is lognormal.

3 NON-GAUSSIAN DYNAMIC SYSTEMS

Dynamic Linear Models (DLM), also known as state space models have been widely used to analyze time series. They provide a useful statistical framework to study the behaviour of dynamic systems. Dynamic systems are characterized as permitting smooth and abrupt changes in the time series generating process and as well as permitting the accommodation of subjective information [West and Harrison 1997]. Let $\mathbf{X_t}$ be a vector of observations taken at time t for $t = 1, 2, \ldots$ and assume that the time series is a realization of the dynamic system whose state space at time t is denoted by $\mathbf{\Theta}_t = (\theta_1, \ldots \theta_t)$. Generally we assume normal observations, but a generalization to the exponential family called Dynamic Generalized Linear Model (DGLM) was formalized by West, Harrison and Migon [1985]. This class is a generalization of the generalized linear models with the parameters (states) changing through time. Dynamic generalized linear models involve an observational model and an evolution equation on the states. The extension of the dynamic linear model to nonlinear state space models has been widely used. In these models the sampling distribution of $\mathbf{X_t}$ given a random variable λ_t, belongs to the exponential family where $\lambda_\mathbf{t}$ is a function of a linear combination of the states $\mathbf{\Theta_t}$, that is

$$\lambda_t = g(\eta_t), \quad \eta_t = \mathbf{F}_t^T \mathbf{\Theta}_t \tag{1}$$

for a known regression vector $\mathbf{F}_t^T$. In practice it is common that $g(.)$ is the identity map. The evolution equation is

$$\mathbf{\Theta}_t = G_t \mathbf{\Theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N[\mathbf{0}, \mathbf{W_t}] \tag{2}$$

where G_t is a known evolution matrix and the error terms are assumed to be mutually independent. At any time the joint density of the states $\mathbf{\Theta}_t$ is Markov with respect to the DAG $\mathcal{G}$. The states are assumed to lie in a clique. When the system is Gaussian i.e. the states are normally distributed and the observations $\mathbf{X}_t|\eta_t$ have Gaussian density with mean $\eta_t = \mathbf{F}_t^T \mathbf{\Theta}_t$, then the posterior distribution of $\mathbf{\Theta}_t|\mathbf{X}_t$ can be obtained in closed form [see Lauritzen and Spiegelhalter, 1988; Jensen, 1996]. However, unless the distribution of $\mathbf{X}_t|\eta_t$ is discrete or Gaussian, there is no exact sequential analysis and some approximations are needed. Here is a brief outline of the dynamic generalized linear models. More details can be found in West and Harrison [1997].

Let $\eta_t = g^{-1}(\lambda_t)$ belong to some parameterized family of densities ϕ which is closed under the sampling of an observation $\mathbf{X}_t$. Note from (1) that the sampling distribution of $\mathbf{X}_t$. depends on $\mathbf{\Theta}_t$, only through η_t. Suppose that $\mathbf{X}_t|\eta_t$ lies in an exponential family. Let η_t be a linear function of normally distributed states $\mathbf{\Theta}_t$, as in (1) above. Now the actual density function of λ_t is approximated by a distribution in ϕ, the exponential family, which is closed under sampling to $\mathbf{X}_t|\eta_t$ with mean and variance derived from the mean and variance of the distribution of $g(\eta_t)$. The approximate density function of $\lambda_t|\mathbf{X}_t$, say $\hat{p}(\lambda_t|\mathbf{X}_t)$ is calculated in closed form using conjugate analysis. The posterior distribution of $\eta_t = g^{-1}(\lambda_t)$ can be calculated directly from $\hat{p}(\lambda_t|\mathbf{X}_t.)$ Finally the posterior mean and variance of the states are estimated from the mean and variance of η_t. In this context we assume $g(.)$ to be a simple function like the identity map. This implies that λ_t is linearly related to the states that belong to a single clique. Also assume simple ways of approximating the density of $\lambda_t|\mathbf{X}_t$ by $\hat{p}(\lambda_t|\mathbf{X}_t)$ like equating moments. This approach is adopted as a dynamic approximation technique to update the cliques marginal probabilities in the light of new observations. We start by updating the marginal probability in the clique using the dynamic generalized linear model approximate analysis. The approximate posterior distribution $\hat{p}(\lambda_t|\mathbf{X}_t)$ is computed by conjugate analysis and finally the updated clique distribution is approximated as Gaussian distribution with moments obtained from $\hat{p}(\lambda_t|\mathbf{X}_t)$. As the states are normally distributed, the data assimilation process continues through the junction tree by applying the standard Gaussian propagation al-

gorithms.

4 LOGNORMAL DYNAMIC MODELS

The primary development in this section concerns non-Gaussian observational distributions. Consider the time series scalar observations $X_t, (t = 1, 2, \ldots)$ where $X_t|\lambda_t(\mathbf{\Theta}_t(i))$ is lognormally distributed with median $\lambda_t(\mathbf{\Theta}_t(i))$ where $(\mathbf{\Theta}_t(i))$ is the states vector of variables lying in a clique $C(i)$. Now $\log X_t|\lambda_t(\mathbf{\Theta}_t(i))$ is normally distributed with mean $\log \lambda_t(\mathbf{\Theta}_t(i))$ and variance V.
Let D_t denote the information until time t. Then D_0 is the prior information. If there is no information out of the sample at time t, then $D_t = D_{t-1} \bigcup \{x_t\}$. At time t, historical information D_{t-1} is summarized through a prior distribution for $\mathbf{\Theta}_t(i)|D_{t-1}$ as normal distribution with mean $\boldsymbol{\mu}_t(i)$ and variance $\Sigma_t(i)$, and the random variable $\log \lambda_t(\mathbf{\Theta}_t(i))$ is normally distributed with mean α_0 and variance τ_0^2. The updating procedure develops as follows.
Approximate the distribution of $\log \lambda_t = \eta_t$ which is a univariate normal by a variable λ_t which is lognormally distributed with mean and variance

$$\begin{aligned} \mu_0 &= e^{\alpha_0 + \frac{1}{2}\tau_0^2}, \\ \sigma_0^2 &= (e^{\tau_0^2} - 1)\mu_0^2. \end{aligned}$$

From the joint normal distribution of η_t and $\mathbf{\Theta}_t(i)$, calculate the conditional normal distribution of $\mathbf{\Theta}_t(i)|\eta_t$ with mean and variance

$$\begin{aligned} \boldsymbol{\mu}_t^*(i) &= \boldsymbol{\mu}_t(i) + \frac{\Sigma_t(i)\mathbf{F}_t(\eta_t - \mathbf{F}_t^T\boldsymbol{\mu}_t(i))}{\mathbf{F}_t^T\Sigma_t(i)\mathbf{F}_t}, \\ \Sigma_t^*(i) &= \Sigma_t(i) - \frac{\Sigma_t(i)\mathbf{F}_t\mathbf{F}_t^T\Sigma_t(i)}{\mathbf{F}_t^T\Sigma_t(i)\mathbf{F}_t}. \end{aligned}$$

After observing X_t, find the posterior of $\log \lambda_t$ which is normal with mean and variance

$$\begin{aligned} \alpha_1 &= (1 - \frac{\tau_0^2}{\tau_0^2 + V})\alpha_0 + (\frac{\tau_0^2}{\tau_0^2 + V}) \log X_t \\ \tau_1^2 &= (\frac{\tau_0^2}{\tau_0^2 + V})V, \\ &= (1 - \frac{\tau_0^2}{\tau_0^2 + V})\tau_0^2. \end{aligned}$$

Consequently the posterior of λ_t is lognormal with mean and variance

$$\begin{aligned} \mu_1 &= e^{\alpha_1 + \frac{1}{2}\tau_1^2}, \\ \sigma_1^2 &= (e^{\tau_1^2} - 1)\mu_1^2. \end{aligned}$$

To update η_t we approximate the posterior of λ_t by a normal distribution by equating means and variances to obtain the distribution of η_t as Gaussian with mean and variance

$$\begin{aligned} \mu_1 &= X_t^{\frac{\tau_0^2}{\tau_0^2+V}} \mu_0^{(1-\frac{\tau_0^2}{\tau_0^2+V})}, \\ \sigma_1^2 &= (e^{(\frac{\tau_0^2}{\tau_0^2+V})V} - 1)\mu_1^2. \end{aligned}$$

From the conditional distribution of $\mathbf{\Theta}_t(i)|\eta_t$ and the posterior of η_t above, calculate the posterior distribution of $\mathbf{\Theta}_t(i)$ as Gaussian with mean and covariance matrix

$$\begin{aligned} \boldsymbol{\mu}_t'(i) &= \boldsymbol{\mu}_t(i) + \Sigma_t(i)\frac{\mathbf{F}_t(\mu_1 - \mathbf{F}_t^T\boldsymbol{\mu}_t(i))}{\mathbf{F}_t^T\Sigma_t(i)\mathbf{F}_t} \\ \Sigma_t'(i) &= \Sigma_t(i) - \Sigma_t(i)\mathbf{F}_t\mathbf{F}_t^T\Sigma_t(i)\frac{1 - \frac{\sigma_1^2}{\mathbf{F}_t^T\Sigma_t(i)\mathbf{F}_t}}{\mathbf{F}_t^T\Sigma_t(i)\mathbf{F}_t} \end{aligned}$$

Finally, as a result of obtaining the Gaussian marginal for the vector of variables in the clique $C(i)$, we order the cliques of the junction tree to update the nodes sequentially.

5 VALIDATION

The validity of the approximation algorithm proposed above can be checked by introducing a metric over the distribution of the states at a given time and use it to judge the approximation. The Hellinger distance is a useful one to use for this purpose. It is a measure of the closeness of two densities, defined by

$$H(p, \hat{p}) = \int (p^{1/2} - \hat{p}^{1/2})^2$$

where p and $\hat{p}$ are, respectively, a density and its approximating density. A nice property of the Hellinger distance is that $H^2(p, \hat{p})$ can be calculated in closed form for densities in most standard families. As discussed in the previous section, the algorithm approximates only the distribution of η.

It is not difficult to show that the closeness of the joint density over all states Θ_t depends only on the closeness of our approximation of the one dimensional normal distribution of η. So the closeness of the true distribution of η to its approximation ensures the appropriateness of our approximation.

6 CONCLUSION

In an iterative high-dimensional dynamic systems, where the states (parameters) are allowed to change with time, computational efficiency is essential. As an example of such situation is the environmental problem of forecasting the spread of a release of toxic gases after a nuclear or chemical accident where puffs of contaminated masses (states) are emitted from a source. These puff fragment into other puffs and the number of puffs and puffs fragments becomes very large with the passage of time. To model such scenarios Bayesian networks were defined over state spaces. When the system is Gaussian, i.e. the states are normally distributed and the observations have Gaussian density, quick exact propagation algorithms that calculate the posterior distribution-in closed form- in the light of incoming data are well known.

This paper provides an approximate algorithm of propagation and probability updating for non-Gaussian dynamic systems, in particular the lognormal incoming data. This algorithm is based on the dynamic generalized linear model of West and Harrison [1997] The validity of the dynamic approximation is checked by using the Hellinger metric.

7 REFERENCES

Boyen, X., and D. Koller, Tractable inference for complex stochastic processes. *In Proceedings of the Fourteenth Conference in Uncertainty in Artificial Intelligence*, Morgan Kauffman, San Francisco, CA, 33-42, 1998.

Dawid, A.P., Applications of general propagation algorithm for probabilistic expert systems, *Statistics and Computing*, **2**, 25-36, 1992.

Gargoum, A.S., Use of Bayesian dynamic models for updating estimates of contaminated material, *Environmetrics*, **12**, 1-9, 2001.

George, E.I., U.E. Markov, and A.F.M. Smith, Fully Bayesian hierarchical analysis. In Freeman, P.R., and Smith, A.F.M.,(eds), *Aspects of Uncertainty*. Wiley, 191-8, 1994.

Jensen, F., F.V. Jensen, and S.L. Ditter, From influence diagrams to junction trees. *In Proceedings of the Tenth Conference in Uncertainty in Artificial Intelligence*, Morgan Kauffman, San Francisco, CA, 367-373, 1994.

Jensen, F.V., An introduction to Bayesian networks, UCL Press, London, 1996.

Kjaerulff, U., A computational scheme for reasoning in dynamic probabilistic networks. *In Proceedings of the Eighth Conference in Uncertainty in Artificial Intelligence*, Morgan Kauffman, San Francisco, CA, 121-129, 1992.

Lauritzen, S.L., Propagation of probabilities, means and variances in mixed graphical association models. *J. Amer. Statist. Assoc.* **87**, 1098-1108, 1992

Lauritzen, S.L., and D.J. Spiegelhalter, Local computations with probabilities on graphical structures and their application to expert systems (with discussion) *J. Roy. Statist. Soc. (Ser B)* **50**, 157-224, 1988.

Lauritzen, S.L., T.P. Speed, and K. Vijayan, Decomposable graphs and hypergraphs, *J. Austral. Math. Soc. (Ser.A)*, **36**, 12-29, 1984.

Mikkelsen, T., S.E. Larsen, and S.Thykier-Neilsen, Description of the Riso puff diffusion model, *Nuclear Safety* **67**, 56-65, 1984.

Row, E., Application in atmospheric sciences, In Grow, E.L., Shimizu, K., (eds)*Lognormal Distributions: Theory and Applications*, Dekker, New York, 1988.

Settimi, R., and J.Q. Smith, A Comparison of Approximate Bayesian Forecasting Methods for Non-Gaussian Time series, *Journal of Forecasting*. **19**, 135-148, 2000.

Settimi, R., J.Q. Smith and A.S. Gargoum, Approximate learning in complex dynamic Bayesian networks, *In Proceedings of the Fifteenth Conference in Uncertainty in Artificial Intelligence*, Morgan Kauffman, San Francisco, CA, 585-593, 1999.

Smith, A.F.M., and G.O. Roberts, Bayesian computation via Gibbs and related Markov Chain Mont Carlo methods (with discussion). *J. Roy. Statist. Soc. (Ser.B)*. **55**, 3-24, 1993.

Smith, J.Q., A comparison of the characteristics of some Bayesian forecasting models. *International Statistical Review*, **60**, 75-87, 1992.

Smith, J.Q., and S. French, Bayesian updating of atmospheric dispersion models for use after an accidental release of radioactivity, *The Statistician*, **42**, 501-511, 1993.

Smith, J.Q., and K.N. Papamichail, Fast Bayes and the dynamic junction forest. *Artificial Intelligence*, **107**, 99-124, 1999.

Smith, J.Q., S. French and D.C. Ranyard, An efficient graphical algorithm for updating estimates of the dispersal of gaseous waste after an accidental release. In Gammerman A. (ed.) *probabilistic reasoning and Bayesian belief networks.* Henley-on-Thames: Alfred Waller, 124-144, 1995.

Spiegelhalter, D.J., A.P. Dawid, S.L. Lauritzen and R.G. Cowell, Bayesian, analysis in expert systems, *Statistical Science*, **8**, No.3, 219-283, 1993.

Tarjan, R.E., and M. Yannakakis, Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM. J. Comput*, **13**, 566-579, 1984.

Thykier-Nielsen, S., and T. Mikkelsen, Rimpuff User Guide: version 30, *Riso National Laboratory*, Denmark, 1991.

West, M., and P.J. Harrison, Bayesian forecasting and dynamic models, 2nd edition, New York, *Springer-Verlag*, 1997.

West, M., P.J. Harrison and H.S. Migon, Dynamic generalized linear models and Bayesian forecasting (with discussion), *J. Amer. Statist. Assoc*. **80**, 73-97, 1985.