BAYESIAN STATISTICS 7, pp. 535–542
J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid,
D. Heckerman, A. F. M. Smith and M. West (Eds.)

Dynamic Lattice-Markov Spatio-Temporal Models for Environmental Data

LINDA M. GARSIDE and DARREN J. WILKINSON
Newcastle University, U.K.
l.m.garside@ncl.ac.uk d.j.wilkinson@ncl.ac.uk

SUMMARY

This paper is concerned with the modelling of the latent structure of a Bayesian spatio-temporal model with a view to improving parameter inference, smoothing and prediction. The equilibrium distribution of a time stationary system will be examined, paying particular attention to edge-effects and the effect of grid-coarsening. In order to develop an effective MCMC algorithm, the latent process is integrated out of the model. These techniques will be illustrated using North Atlantic ocean temperature data.

Keywords: COARSE AND FINE MODELS; EDGE-EFFECTS; MARGINAL UPDATING; MCMC.

1. INTRODUCTION

This paper concerns Bayesian inference for dynamic lattice-Markov models with a latent process of Spatio-Temporal Auto-Regressive (STAR) form (Cressie 1991, Pfeifer and Deutsch 1980), observed irregularly, and with error. Attention focuses on the 2+1-D case. A point at location s and time $t+1$ is influenced by the point s and 8 neighbours at time t. Well-known "edge-effects" destroy the spatial stationarity of finite lattices, and this can lead to biased inferences. In order to reduce bias, the autoregressive weights at the edges are adjusted using multivariate Normal methods since the "usual" solution of embedding the area of interest into a much larger lattice is computationally prohibitive.

The problem of estimating the model parameters from data is explored. Gibbs and block-Gibbs sampling methods suffer from poor mixing due to the high dependence between the parameters and the latent process. Integrating out the latent process leads to an MCMC scheme with good mixing properties. Other approaches to hierarchical modelling in a similar context can be found in Wikle *et al.* (1998). Our technique is applied to Atlantic sea temperatures measured at a depth of 100m, within the region 20 to 30 degrees latitude and -80 to -20 degrees longitude, for 1978–1988. The data is collected by ships trawling the area and thus occurs irregularly within space and time.

The method described above is computationally intensive. One method for reducing the computation time is to consider modelling on a coarser grid (Higdon *et al.* 2003); here a grid based on one in four spatial locations is considered. The coarse model has second-order time dependence, but the parameters have the same interpretation as for the fine model.

2. THE MODEL

Consider a Dynamic Linear Model (DLM) which can evolve through space and time (West and Harrison 1997). The model is described using two equations:

(i) System equation

$$\theta^{(t+1)} \mid \theta^{(t)} \sim \mathrm{N}(G\theta^{(t)}, W), \qquad W = \tau_s^{-1} I$$

(ii) Observation equation

$$X^{(t)} \mid \theta^{(t)} \sim \mathrm{N}(F_t\theta^{(t)}, V), \qquad V = \tau_o^{-1} I$$

Where $\theta^{(t)} = (\theta_1^t, \theta_2^t, \ldots, \theta_n^t)$ is the vector of latent points in space at time t. Point $\theta_i^{(t)}$ represents location i at time t and G is the matrix which captures the evolution of the the points in space through time. Corresponding to each vector $\theta^{(t)}$ we have a vector of observations $X^{(t)}$, related by the observation matrix F_t. Typically F_t will consist of a subset of rows of the identity matrix corresponding to the points in space for which data is available at time t.

2.1. *2+1-D Model*

Consider the case where we have data on a 2-dimensional regular square lattice in space evolving through discrete time. Initialise the latent structure as a grid of independent spatial observations at time $t = 0$. For $t > 0$, the latent points at time $t + 1$ depend on a linear combination of the points at time t. In this model, the vector of latent points $\theta^{(t)}$ is defined to be

$$\theta^{(t)} = \left(\theta_{1,1}^{(t)}, \ldots, \theta_{1,n}^{(t)}, \theta_{2,1}^{(t)}, \ldots, \theta_{2,n}^{(t)}, \ldots, \theta_{m,n}^{(t)}\right).$$

The obvious first order (five-neighbour) model suffers badly from lattice co-ordinate system artifacts. These can be reduced by considering a second order (nine-neighbour) model with carefully chosen weights. This corresponds to a dynamic space-time version of the "second order" neighbour system described by Besag (1974). The model is defined by

$$\begin{aligned}\theta_{i,j}^{(t+1)} \mid \theta^{(t)} \sim N\Big(\alpha\Big\{ & p_{11}^2\theta_{i-1,j-1}^{(t)} + 2p_{11}p_{12}\theta_{i-1,j}^{(t)} + p_{12}^2\theta_{i-1,j+1}^{(t)} \\ & + 2p_{11}p_{12}\theta_{i,j-1}^{(t)} + (2p_{11}p_{22} + 2p_{12}p_{21})\theta_{i,j}^{(t)} + 2p_{12}p_{22}\theta_{i,j+1}^{(t)} \\ & + p_{21}^2\theta_{i+1,j-1}^{(t)} + 2p_{21}p_{22}\theta_{i+1,j}^{(t)} + p_{22}^2\theta_{i+1,j+1}^{(t)}\Big\}, \tau_s^{-1}\Big),\end{aligned} \tag{1}$$

where $\sum_{i,j} p_{ij} = 1$. This model therefore has five free parameters. The three free p_{ij} parameters control the degree of anisotropy, α is an autoregressive parameter, and τ_s controls the degree of innovation noise. Stationarity of this system for $\alpha \in [0, 1)$ is established in Pfeifer and Deutsch (1980). The precise form of the weight structure is motivated by considering an additional "hidden" layer at time $t + 1/2$ which is also a square lattice, but perfectly offset from the lattices at times t and $t + 1$. If a first order (four-neighbour) model with edge weights p_{ij} is imposed on the full lattice structure, marginalising out the hidden layer leads to model (1).

2.2. *Correcting for Edge-effects*

This paper considers only the isotropic case of the model defined by (1) with weights $p_{ij} = 1/4$ giving

$$\begin{aligned}\theta_{ij}^{(t+1)} \mid \theta^{(t)} \sim N\Big(\alpha\Big\{ & \frac{1}{16}\left[\theta_{i-1,j-1}^{(t)} + \theta_{i-1,j+1}^{(t)} + \theta_{i+1,j-1}^{(t)} + \theta_{i+1,j+1}^{(t)}\right] \\ & + \frac{1}{8}\left[\theta_{i-1,j}^{(t)} + \theta_{i,j-1}^{(t)} + \theta_{i,j+1}^{(t)} + \theta_{i+1,j}^{(t)}\right] + \frac{1}{4}\theta_{i,j}^{(t)}\Big\}, \tau_s^{-1}\Big).\end{aligned}$$

The above equation only defines the distribution of the internal nodes. Along the edges and at the corners of the system some neighbours are missing. Standard approaches either simply drop terms corresponding to missing neighbours, or drop terms and then re-scale the remaining weights so that they sum to one. Both of these strategies lead to prominent "edge-effects" which affect the time stationary distribution of the process in space. In the purely spatial context, such effects are often reduced by embedding the region of interest into a much larger lattice. However, in the spatio-temporal context, such an approach is typically computationally infeasible. In the context of a purely spatial Gaussian Markov random fields, more sophisticated approaches to edge weight adjustments have been examined in detail by Besag and Kooperberg (1995). Here we adopt a related approach, better suited to the context of dynamic spatio-temporal models. For a given value of α, the process is linear Gaussian and the time-stationary joint distribution may be computed using numerical linear algebra. From this, the joint distribution of any edge point and its neighbours may be computed, leading to the conditional distribution of the point given available neighbours. This process only needs to be carried out for one typical edge node and one typical corner node, but must be re-computed for each possible α. In practice, the conditional distribution is computed for 100 different values of α in the range $[0, 1)$, and linear interpolation is used for other values of α. Note that the conditional distribution does not need to be re-computed for every possible (α, τ_s) pair as the conditional distributions re-scale in the obvious way for varying τ_s. Figure 1 shows an example of the time-stationary spatial variance surface for the three described modelling strategies with $\alpha = 0.99$ on a 10×10 spatial grid. The first plot represents the model with no edge adjustments (dropping terms corresponding to non-existent neighbours), the second plot has had the edge weights adjusted in a simple way (drop terms and re-scale edge weights to sum to one), whereas the third plot shows the results for adjusting both the edge weights and correcting the variance according to the time stationary distribution. This latter plot is much "flatter" than the previous two, and hence closer to the desired spatial stationarity.

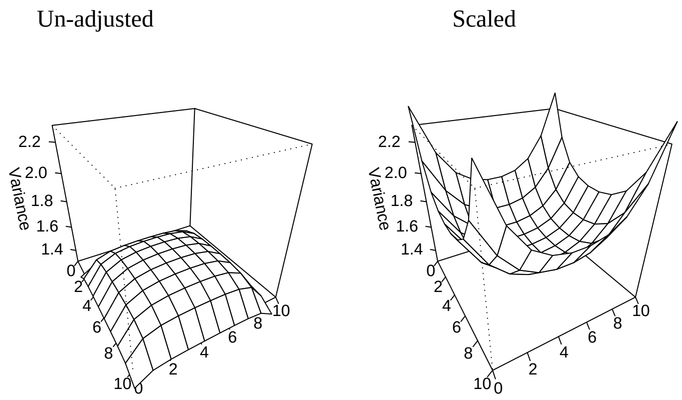

Figure 1. *Spatial variance surfaces for time-stationary distributions based on un-adjusted, scaled and adjusted edge weights.*

2.3. *Coarsening the Model*

For given parameter values, the model described in the previous subsection performs well for spatio-temporal smoothing and interpolation on reasonably fine lattices. However, as the model underlying a good MCMC scheme for parameter inference, it is computationally demanding. One way to reduce the computing time required is to consider a coarser model resulting from considering only one in every four points in space. Unfortunately, if the original fine model is marginalised to the coarser lattice the resulting process is not Markovian, of any order, but can be well approximated by a Markov process of order 2. Each point in space and time $t+2$ is regarded as depending on the nearest neighbour at time $t+1$ and the nine neighbours at time t, as shown in Figure 2.

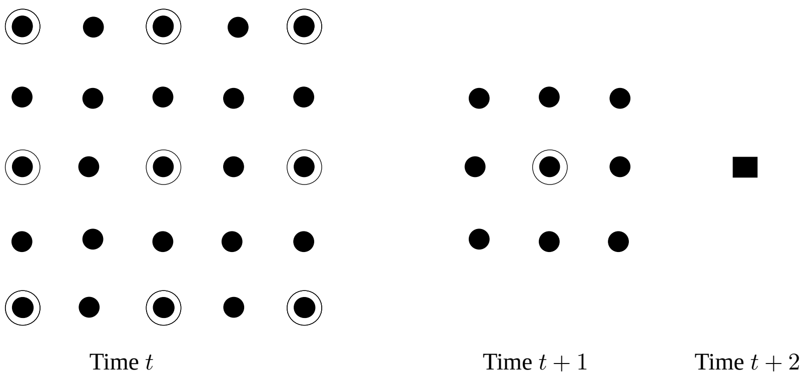

Figure 2. *Circled nodes represent the neighbours in a 2+1-D Coarse Model*

As is the case for the edge-effect corrections, analysis of the stationary distribution can uncover the appropriate weights and conditional variances for the coarse model in a manner which makes parameters consistent and share interpretations between the coarse and fine scales.

3. PARAMETER ESTIMATION

In this model we have three unknown parameters, τ_s, τ_o, and α, and latent process $\theta = \{\theta^{(t)} \mid t = 1, 2, \ldots, T\}$. Naive MCMC techniques (such as univariate Gibbs samplers) perform particularly poorly for problems of this type due to the very high-dimensional nature of the missing data. A two-block data augmentation scheme which alternately samples $\sigma \mid \theta, X$ and $\theta \mid \sigma, X$, where $\sigma = (\tau_s, \tau_o, \alpha)$, performs much better than the naive techniques (Wilkinson and Yeung 2002a), but still suffers from poor mixing due to the high posterior dependence between σ and θ (Papaspiliopoulos *et al.* 2003). Fortunately a sampler with good mixing properties can be constructed by integrating θ out of problem and directly constructing a Metropolis-Hastings scheme for $\sigma \mid X$.

The following independent priors are adopted for the parameters:

$$\tau_i \sim \text{Ga}(b_i, d_i),\ i \in \{s, o\}, \qquad \alpha \sim \text{Be}(\gamma, \delta).$$

The beta prior ensures that α remains between zero and one (the intuitively plausible range for temporal dependence). Using these priors, the Metropolis-Hastings acceptance probability for the proposed update $\sigma^* = (\tau_s^*, \tau_o^*, \alpha^*)$ is $\min\{1, A\}$, where

$$A = \frac{[\tau_s^*][\tau_o^*][\alpha^*][X \mid \sigma^*] f(\sigma \mid \sigma^*)}{[\tau_s][\tau_o][\alpha][X \mid \sigma] f(\sigma^* \mid \sigma)}, \tag{2}$$

and $f(\sigma^* \mid \sigma)$ is the proposal density for the parameters. If a symmetric density is used (2) becomes

$$A = \frac{[\tau_s^*][\tau_o^*][\alpha^*][X \mid \sigma^*]}{[\tau_s][\tau_o][\alpha][X \mid \sigma]}.$$

To obtain a rapidly mixing chain, normal random walk updates on the log-precisions work well, and so the density of the log of a gamma is used in the the acceptance probability. The marginal log-likelihood ratio can be calculated in a variety of ways (for example, using the Kalman filter or causal tree propagation), but is computationally demanding whatever strategy is adopted.

The authors recommend using sparse matrix techniques, such as those implemented in the C software library GDAGsim (Wilkinson 2001). The proposal is constructed by independently sampling

$$\log \tau_o^* \,|\, \tau_o \sim \mathrm{N}(\log \tau_o, g^{-1}),$$
$$\log \tau_s^* \,|\, \tau_s \sim \mathrm{N}(\log \tau_s, g^{-1}),$$
$$\alpha^* \,|\, \alpha \sim \mathrm{N}(\alpha, e),$$

where e and g are Metropolis-Hastings tuning parameters. For a more detailed discussion of the MCMC techniques which can be used for large linear models, see Wilkinson and Yeung (2002b).

4. EXAMPLE: ATLANTIC OCEAN TEMPERATURE DATA

We illustrate these techniques with spatio-temporal data obtained from the National Oceanographic Data Center (`http://www.nodc.noaa.gov/`). The data used in this example were taken from a rectangular region of the Atlantic ocean 20 to 30 degrees latitude and -80 to -20 degrees longitude. Ocean station data (NANSEN) was considered for the time period 1979-1988 inclusive. Observations were indexed according to their time and location for depths of approximately 100m. The location of the data in space and time is illustrated graphically in Figure 3. Similar data have been analysed previously by Lavine and Loizier (1999) and examined at one location over time. They also looked at data from a different spatial location between 1905 and 1988, as in Higdon (1998) and Stroud *et al.* (2001).

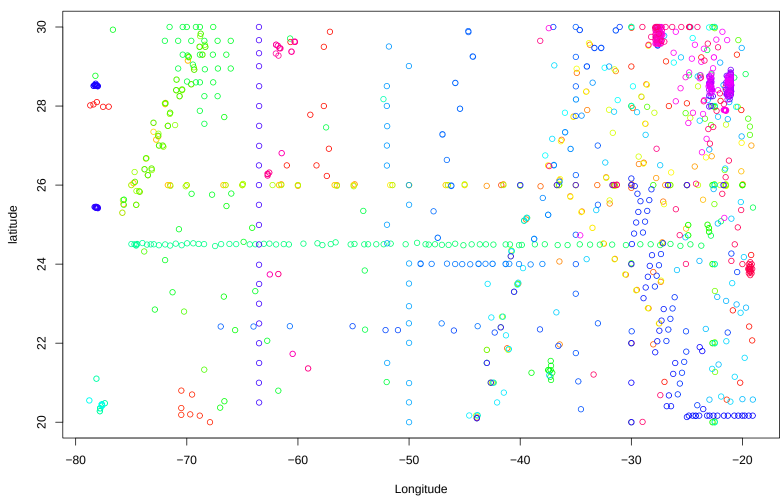

Figure 3. *Sites in space for which data is available (shade indicates time)*

In this example we consider a fixed α of 0.9 as there is very little information regarding α in this data set. Expert subjective priors for all three parameters would improve on this analysis, and should reduce the confounding between α and τ_s. For illustrative purposes, we adopt vague proper prior distributions for the precisions

$$\tau_s \sim \mathrm{Ga}(1, 0.001), \quad \tau_o \sim \mathrm{Ga}(1, 0.001)$$

and consider a $\mathrm{N}(0, 400^{-1})$ random walk update on each component independently. We model the data on a $12 \times 60 \times 10$ grid, but carry out MCMC on the coarsened version of the model,

Table 1. *Posterior means and associated standard errors from an MCMC run of 10,000 with 1,000 iterations discarded as burn-in and no thinning (conditional on $\alpha = 0.9$).*

	Mean	SD	Naive SE	Time-series SE
$\log(\tau_s)$	-0.5573	0.09848	0.0010381	0.001836
$\log(\tau_o)$	-0.8682	0.04005	0.0004222	0.000759

Table 2. *Posterior quartiles for the parameters from an MCMC run of 10,000 iterations with 1,000 iterations discarded as burn-in and no thinning (conditional on $\alpha = 0.9$).*

	2.5%	25%	50%	75%	97.5%
$\log(\tau_s)$	-0.7477	-0.6243	-0.5582	-0.4911	-0.3613
$\log(\tau_o)$	-0.9499	-0.8941	-0.8677	-0.8424	-0.7886

which uses a $6 \times 30 \times 10$ grid. Point estimates (posterior means) of the parameters are computed from the MCMC run, and these are then input into the fine model for smoothing/interpolation.

4.1. *Results*

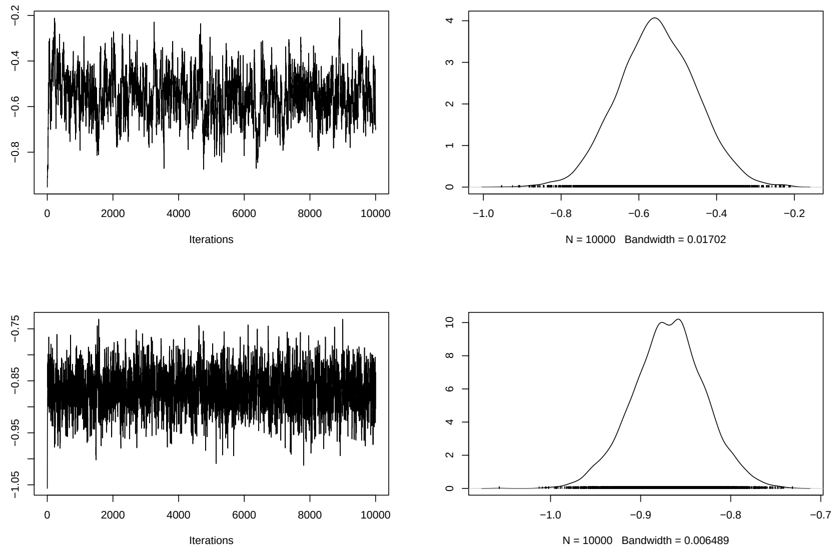

Figure 4. *Trace and density plots for $\log(\tau_s)$ and $\log(\tau_o)$ with α fixed at 0.9, based on 10,000 iterations and no thinning.*

The results from a typical MCMC run are given in Tables 1 and 2 and Figures 4 and 5. Note, in particular, the rapid decay in autocorrelations and convergence to stationarity signifying a rapidly mixing chain. The posterior mean parameter estimates are $\log(\tau_s) = -0.56$, and $\log(\tau_o) = -0.87$, based on 9,000 iterations. Due to the consistency between the coarse and fine formulations, these parameter estimates can be used as plug-in estimates for the parameters of a fine model for the purposes of smoothing and interpolation. This allows the computation of posterior mean temperature surfaces on the fine scale, as shown in Figures 6, 7 and 8. The height represents the posterior mean temperature conditional on the plug-in parameter estimates from the coarse model MCMC run.

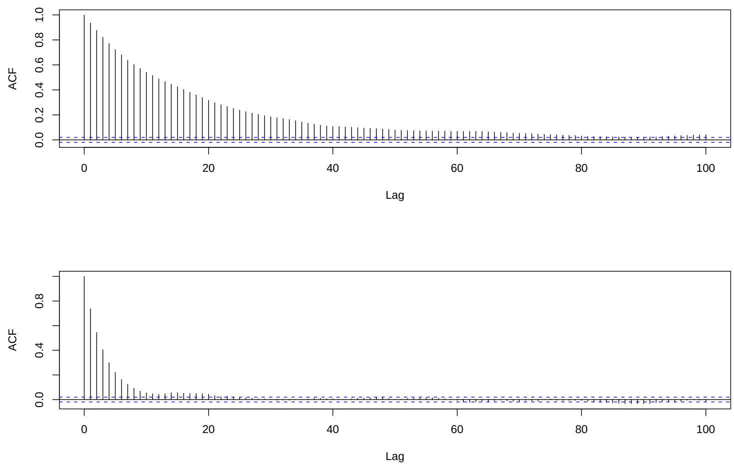

Figure 5. *Autocorrelation plots for for* $\log(\tau_s)$ *and* $\log(\tau_o)$ *with with* α*=0.9*

The surface plots below show how the temperature field varies both in space and in time. The original data set had far more data points associated with the earlier years in the study, and this could account for the increased smoothness in the plots for later time periods.

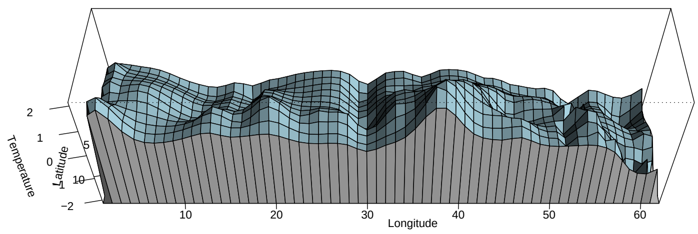

Figure 6. *Surface plot for 1984.*

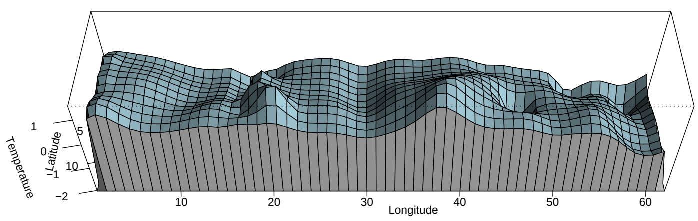

Figure 7. *Surface plot for 1986.*

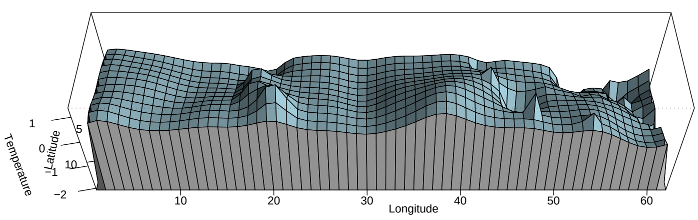

Figure 8. *Surface plot for 1988.*

5. CONCLUSIONS

Using a STAR model to capture the dynamics of a stationary spatio-temporal process and then using the DLM formulation to relate the process to available data provides a powerful and flexible framework for inference. Because of the computationally intensive nature of spatio-temporal models it is desirable to reduce the edge-effects associated with lattice-Markov systems, and this can be carried out via numerical analysis of the time-stationary distribution of the STAR process. Similar computational considerations motivate the use of coarsened models constructed so as to remain consistent with the original fine formulation, and this can be tackled in a similar way to the edge-effects.

REFERENCES

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. B* **36**, 192–236.

Besag, J., Kooperberg, C. (1995) On conditional and intrinsic autoregressions. *Biometrika* **82**, 733–746.

Cressie, N.A.C. (1991). *Statistics for Spatial Data.* Chichester: Wiley.

Higdon, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic ocean. *Environmental and Ecological Statistics* **5**, 173–190.

Higdon, D., Lee, H. and Holloman, C. (2003). Markov chain Monte Carlo-based approaches for inference in computationally intensive inverse problems. *In this volume*.

Lavine, M. and Loizier, S. (1999). A Markov random field spatio-temporal analysis of ocean temperature. *Environmental and Ecological Statistics* **6**, 249-273.

Papaspiliopoulos, O., Roberts, G. O. and Sköld, M. (2003). Non-centered parameterisations for hierarchical models and data augmentation. *In this volume*.

Pfeifer, P. E. and Deutsch, S. J. (1980). Stationarity and invertibility regions for low order STARMA models. *Comm. Statist. Simulation and Computation* **9**, 551–562.

Stroud, J. R., Muller, P. and Sanso, B. (2001). Dynamic models for spatio-temporal data. *J. Roy. Statist. Soc. B* **63**, 673–689.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models (2nd ed.)* New York: Springer.

Wikle, C. K., Berliner, M. L. and Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics* **5**, 117–154.

Wilkinson, D. J. (2001). GDAGsim 0.2 User Guide. *Tech. Rep.*, University of Newcastle. UK. Available from `http:// www.staff.ncl.ac.uk/d.j.wilkinson/software/gdagsim.`

Wilkinson, D. J. and Yeung, S. K. H. (2002a). Conditional simulation from highly structured Gaussian systems, with application to blocking-MCMC for the Bayesian analysis of very large linear models. *Statist. Computing* **12**, 287–300.

Wilkinson, D. J. and Yeung, S. K. H. (2002b). A sparse matrix approach to Bayesian computation in large linear models. *Comput. Statist. and Data Analysis* (to appear).