

# Modeling Hourly Ozone Concentration Fields

Yiping Dou

*University of British Columbia, Vancouver, Canada.*

Nhu D Le

*BC Cancer Research Centre, Vancouver, Canada*

James V Zidek

*University of British Columbia, Vancouver, Canada.*

**Summary.** This paper presents a dynamic linear model for modeling hourly ozone concentrations over the eastern United States. That model, which is developed within an Bayesian hierarchical framework, inherits the important feature of such models that its coefficients, treated as states of the process, can change with time. Thus the model includes a time-varying site invariant mean field as well as time varying coefficients for 24 and 12 diurnal cycle components. This cost of this model's great flexibility comes at the cost of computational complexity, forcing us to use an MCMC approach and to restrict application of our model domain to a small number of monitoring sites. We critically assess this model and discover some of its weaknesses in this type of application.

**Keywords:** Dynamic linear model, hierarchical Bayes, ozone, random field, space-time fields.

## 1. Introduction

This paper presents a model for the spatio-temporal field of hourly ozone concentrations for subregions of the eastern United States, one that can in principle be used for both spatial and temporal prediction. It goes on to critically assess that model and the approach used for its construction, with mixed results.

Such models are needed for a variety of purposes described in Ozone (2005) where a comprehensive survey of the literature on such methods is given, along with their strengths and weaknesses. In particular, they can be used to help characterize population levels of exposures to ozone in outdoor environments, based on measurements taken at often remote ambient monitors.

These interpolated concentrations can also be used as input to computer models that incorporate indoor environments to more accurately predict population levels of exposure to an air pollutant. Such models can reduce the deleterious effects of errors resulting from the use of ambient monitoring measurements to represent exposure. For example, on hot summer days the ambient levels will overestimate exposure since people tend to stay in air conditioned indoor environments where exposures are lower. To address that problem, the US Environmental Protection Agency (EPA) developed APEX. It is being used by policy-makers to set air quality standards under hypothetical emission reduction scenarios (Ozone, 2005). Interpolated ozone fields could well be used as input to APEX to further reduce that measurement error although that application has not been made to date for ozone. However, it has been made for particulate air pollution through an exposure model called SHEDS (Burke et al., 2001) as well as a simplified version of SHEDS (Calder et al., 2003).

Interest in predicting human exposure and hence in mapping ozone space-time fields, stems from concern about the adverse human health effects of ozone. Ozone (2005) reviews an extensive literature on that subject. Exposure chamber studies show that inhaling high concentrations of ozone compromises lung function quite dramatically in healthy individuals (and presumably

E-mail: ydou@stat.ubc.ca

to an even greater degree in unhealthy individuals such as those suffering from asthma). Moreover, epidemiology studies show strong associations between adverse health effects such as exposures. Consequently, the US Clean Air Act mandates that National Ambient Air Quality Standards are necessary for ozone to protect human health and welfare. Thus, spatio-temporal models can have a role in setting these NAAQS.

Ozone concentrations over a geographic region vary randomly over time, and therefore constitute a spatio-temporal field. In both rural and urban areas such fields are customarily monitored, the latter to ensure compliance with the NAAQS amongst other things. In fact, failure can result in substantial financial penalties.

A number of approaches can be taken to modelling such space time fields. Here we investigate a promising one that involves selecting a member of a very large class of so-called state space models. Section 2 describes our choice, a dynamic linear model (DLM), a variation of those proposed by Huerta et al. (2004) and Stroud et al. (2001). Here “dynamic”, refers to the DLM’s capability of systematically modifying its parameters over time, a seemingly attractive feature since the processes it models will themselves generally evolve “due to the passage of time as a fundamental motive force” (West and Harrison, 1997). However, other approaches are possible and in a companion report currently in preparation, the DLM selected here will be compared with other possibilities.

Section 2 introduces the hourly concentration field that is to be modeled in this report. There consideration of measurements made at fixed site monitors and reported in the AIRS dataset leads to the construction of our DLM. [The EPA (Environmental Protection Agency) changed the AIRS (Aerometric Information Retrieved System) to the AFS (Air Facility Subsystem) in 2001.] That model becomes the object of our assessment in subsequent sections. To illustrate how to select some of the model parameters in the DLM, we use the simple first-order polynomial DLM in Section 3 to shed some light on this problem. Moreover, we prove there in a simple but representative case, that under the type of model constructed here and by Huerta et al. (2004), the predictive variances for successive time points conditional on all the data must be monotonically increasing, an undesirable property. Theoretical results and algorithms on the DLM are represented in Sections 4 and 5. The MCMC sampling scheme is outlined in Section 4.1. The forward-filtering-backward-sampling (FFBS) method is demonstrated in Section 4.1 to estimate the state parameters in the DLM. Moreover, we outline the MCMC sampling scheme to obtain samples for other model parameters from their posterior conditional distributions with a Metropolis-Hasting step. Section 5 gives theoretical results for prediction and interpolation at unmonitored (ungauged) sites from their predictive posterior distributions. Section 6 shows the results of MCMC sampling along with interpolation results on the ozone study. Section 7 describes problems with the DLM process revealed by our assessment. We summarize our findings and draw conclusions from our assessment in Section 8.

As an added note, we have developed software, written in C and R and available online (<http://enviro.stat.ubc.ca>) that may be used to reproduce our findings or to use the model for modeling hourly pollution in other settings.

## 2. Model development

Although we believe the methods described in this paper apply quite generally to hourly pollution concentration space-time fields, it focuses on an hourly ozone concentrations (ppb) over part of the eastern United States during the summer of 1995. In all, 375 irregularly located sites (or “stations”) monitor that field. To enable a focused assessment of the DLM approach and make computations feasible, we consider just one cluster of ten stations (Cluster 2), in close proximity to one another. However, in work not reported here for brevity, two other such clusters led to similar findings. Note that Cluster 2 has the same number of stations as the one in Mexico City studied by Huerta et al.(2004).

The initial exploratory data analysis followed that of Huerta et al. (2004) with a similar result,

a square-root transformation of the data is needed to validate the normality assumption for the DLM residuals. [Note that a small amount of randomly missing data were filled in by the spatial regression method (SRM), before we began.] The plot of a Bayesian periodogram (Dou et al., 2007) for the transformed data at the sites in our cluster reveals a peak between 1 pm and 3 pm each day with a significant 24-hour cycle for the stations in Cluster 2. We also found a slightly significant 12-hour cycle. However, no obvious weekly cycles or nightly peaks were seen. Thus, the DLM suggested by our analysis turns out to be a variant of the one in Huerta et al. (2004); it has states for both local trends as well as periodicity across sites.

To define the model, let  $y_{it}$  denote the square-root of the observable ozone concentration, at site  $\mathbf{s}_i$ ,  $i = 1, \dots, n$ , and time  $t$ ,  $t = 1, \dots, T$ ,  $n$  being the total number of gauged (that is, monitoring) sites in the geographical subregion of interest and  $T$ , the total number of time points. Furthermore, let  $\mathbf{y}_t = (y_{1t}, \dots, y_{nt})' : n \times 1$ . Then the DLM for the field is

$$\mathbf{y}_t = \mathbf{1}'_n \beta_t + S_{1t}(a_1) \alpha_{1t} + S_{2t}(a_2) \alpha_{2t} + \nu_t \quad (1)$$

$$\beta_t = \beta_{t-1} + w_t \quad (2)$$

$$\alpha_{jt} = \alpha_{j,t-1} + \omega_t^{\alpha_j}, \quad (3)$$

where  $\nu_t \sim N[\mathbf{0}, \sigma^2 V_\lambda]$ ,  $w_t \sim N[0, \sigma^2 \tau_y^2]$ ,  $\omega_t^{\alpha_j} \sim N[\mathbf{0}, \sigma^2 \tau_j^2 V_{\lambda_j}]$ ,  $V_\lambda = \exp(-V/\lambda)$ ,  $V_{\lambda_j} = \exp(-V/\lambda_j)$ ,  $j = 1, 2$ , and  $\alpha_{jt} = (\alpha_{j1t}, \dots, \alpha_{jnt})' : n \times 1$ ,  $j = 1, 2$ . Here  $\beta_t$  denotes a canonical spatial trend and  $\alpha_{jit}$ , a seasonal coefficient for site  $i$  at time  $t$  corresponding to a periodic component,  $S_{jt}(a_j)$ , where  $S_{jt}(a_j) = \cos(\pi t j / 12) + a_j \sin(\pi t j / 12)$ ,  $j = 1, 2$ . Note that  $V = (v_{ij}) : n \times n$  represents the distance matrix for the gauged sites  $\mathbf{s}_1, \dots, \mathbf{s}_n$ , that is,  $v_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$  for  $i, j = 1, \dots, n$ , where  $\|\mathbf{s}_i - \mathbf{s}_j\|$  denotes the Euclidean distance (km) between sites  $\mathbf{s}_i$  and  $\mathbf{s}_j$ .

Models (1)–(3) can also written in the form of a state space model with the observation and state equations

$$\mathbf{y}_t = \mathbf{F}'_t \mathbf{x}_t + \nu_t \quad (4)$$

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \omega_t, \quad (5)$$

where  $\mathbf{x}'_t = (\beta_t, \alpha_{1t}', \alpha_{2t}')$ ,  $\omega'_t = (\omega_t, \omega_t^{\alpha_1'}, \omega_t^{\alpha_2'})'$ , and  $\mathbf{F}'_t$  is given by

$$\begin{bmatrix} 1 & S_{1t}(a_1) & 0 & \dots & 0 & S_{2t}(a_2) & 0 & \dots & 0 \\ 1 & 0 & S_{1t}(a_1) & \dots & 0 & 0 & S_{2t}(a_2) & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & S_{1t}(a_1) & 0 & 0 & \dots & S_{2t}(a_2) \end{bmatrix}.$$

Note that  $\omega_t \sim N[\mathbf{0}, \sigma^2 W]$ ,  $W$  being the block diagonal matrix with diagonal entries  $\tau_y^2$ ,  $\tau_1^2 \exp(-V/\lambda_1)$ , and  $\tau_1^2 \exp(-V/\lambda_2)$ .

Let  $y_{1:T} = (y_{1:T}^m, y_{1:T}^o)'$ , where  $y_{1:T}^m = (y_1^m, \dots, y_T^m)$  represents all the missing values and  $y_{1:T}^o$ , all the observed values in Cluster 2 sites for  $t = 1, \dots, T$ . The model unknowns are therefore the coordinates of the vector  $(\lambda, \sigma^2, x_{1:T}, y_{1:T}^m, a_1, a_2)$ , in which the vector of state parameters up to time  $T$  is  $x_{1:T} = (\mathbf{x}_1, \dots, \mathbf{x}_T)$ , the range parameter is  $\lambda$ , the variance parameter is  $\sigma^2$  and finally the vector of phase parameters is  $\mathbf{a} = (a_1, a_2)$ . Let  $\gamma = (\tau_y^2, \tau_1^2, \lambda_1, \tau_2^2, \lambda_2)$  be the vector of parameters fixed in the DLM to render computation feasible.

Specification of the DLM is completed by prescribing the hyperpriors for the distributions of some of the model parameters:

$$\begin{aligned} \lambda &\sim IG(\alpha_\lambda, \beta_\lambda) \\ \sigma^2 &\sim IG(\alpha_{\sigma^2}, \beta_{\sigma^2}) \\ \mathbf{a} &\sim N(\mu_a^o, \Sigma_a^o). \end{aligned}$$

Notice that  $\lambda$  and  $\sigma^2$  have inverse Gamma distributions for computational convenience.‡ The choice of the hyperpriors is discussed in Section 6.

‡  $X \sim IG(\alpha, \beta)$  if  $Y = 1/X \sim G(\alpha, \beta)$ , where  $p(y) \propto y^{\alpha-1} \exp(-\beta y)$  for  $\alpha, \beta > 0$ .

We express the state–space model in two different ways because of our dual objectives of parameter inference and interpolation. For simplicity, we use models (4)–(5) for inference about the range, variance and state parameters (see Section 4.1), and use models (1)–(3) for inference on the phase parameters (see Section 4.1) and interpolation (see Section 5).

### 3. Parameter specification

Before turning to the implementation of the approach in the next section, we explore theoretically, albeit in a tractable special case, some features of the model. That exploration leads to insight about how the model’s parameters should be specified as well as undesirable consequences of inappropriate choices. Our assessment will focus on the accuracy of the model’s predictions.

This simple model we consider is a special case of the so–called “first–order polynomial model”, a mathematically tractable, commonly used model. It captures many important features and properties of the DLM we have adopted.

For  $i = 0, 1, \dots, n$  and  $t = 1, \dots, T$ , the first–order polynomial DLM is given by

$$y_{it} = \beta_t + \varepsilon_{it} \quad (6)$$

$$\beta_t = \beta_{t-1} + \delta_t, \quad (7)$$

where  $\varepsilon_{\mathbf{t}} = (\varepsilon_{0t}, \dots, \varepsilon_{nt})' \sim N(\mathbf{0}, \sigma_\varepsilon^2 \exp(-V/\lambda))$ , and  $\delta_t \sim N(0, \sigma_\delta^2)$ . Assume  $\beta_0 \sim N(0, \sigma_\beta^2)$  and  $\lambda, \sigma_\varepsilon^2, \sigma_\delta^2$  and  $\sigma_\beta^2$  are all currently known.

The first–order polynomial DLM is particularly useful for short–term prediction since then the underlying evolution  $\beta_t$  is roughly constant. Observe that the zero–mean evolution error  $\delta_t$  process is independent over time, so that the underlying process is merely a random walk; the model does not anticipate long–term variation. At any fixed time  $t$ :

$$\beta_t = \beta_0 + \sum_{k=1}^t \delta_k \quad (8)$$

$$y_{it} = \beta_0 + \sum_{k=1}^t \delta_k + \varepsilon_{ik}. \quad (9)$$

Consequently, the first–order polynomial DLM has the following covariance structure:

$$\text{Var}(y_{it}) = \sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2 \quad (10)$$

$$\text{Cov}(y_{it}, y_{jt}) = \sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{ij}/\lambda) \quad (i \neq j) \quad (11)$$

$$\text{Cov}(y_{it}, y_{js}) = \sigma_\beta^2 + \min\{t, s\}\sigma_\delta^2 \quad (s \neq t), \quad (12)$$

where  $d_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$ , for  $i, j = 0, 1, \dots, n$  and  $t, s = 1, \dots, T$ .

This DLM defines a non–stationary spatio–temporal process since for the first–order polynomial model to be stationary, the eigenvalues of state transfer matrix,  $G = G_t$  in the notation of West and Harrison (1997), must lie inside of the unit circle. However,  $G_t = 1$  so that this process is not a stationary Gaussian DLM. Furthermore, the DLM defined in Section 2 is non–stationary because  $G_t = I_{2n+1}$  given all the model parameters in (4)–(5). The DLM in (6)–(7) has an important property that the covariance functions in (11)–(12) depends on the time point of  $\min\{t, s\}$ , not on  $|t - s|$  thus confirming our observation of non–stationary.

We readily find the correlation between  $y_{it}$  and  $y_{js}$  to be

$$\text{Cor}(y_{it}, y_{jt}) = \frac{\sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{ij}/\lambda)}{\sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2} \quad (i \neq j) \quad (13)$$

$$\text{Cor}(y_{it}, y_{js}) = \frac{\sigma_\beta^2 + \min\{t, s\}\sigma_\delta^2}{\sqrt{\sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2} \sqrt{\sigma_\beta^2 + s\sigma_\delta^2 + \sigma_\varepsilon^2}} \quad (s \neq t) \quad (14)$$

where  $i, j = 0, \dots, n$  and  $s, t = 1, \dots, T$ .

**Remarks.**

1. The correlations in (13) and (14) have the following properties when  $i \neq j$ :

(i)

$$\text{Cor}(y_{it}, y_{jt}) > \text{Cor}(y_{it}, y_{js}) \quad (15)$$

for  $s \neq t, s, t = 1, \dots, T$  and

(ii)

$$\text{Cor}(y_{it}, y_{jt}) - \text{Cor}(y_{it}, y_{js}) \quad (16)$$

is a monotone increasing function of  $|t - s|$ .

Thus for any fixed time point  $t$ ,  $\text{Cor}(y_{it}, y_{js})$  as a function of  $s$  attains its maximum at  $s = t$  and decreases as  $|s - t|$  increases.

2. By (13),  $\text{Cor}(y_{it}, y_{jt}) \rightarrow 1$  as  $t \rightarrow \infty$  for  $i \neq j, i, j \in \{0, \dots, n\}$ . That property seems unreasonable; the degree of association between two fixed monitors should not increase as an artifact of a larger time  $t$ . That suggests a need to make some of the model parameters, say  $\sigma_\delta^2$ , depend on time. More specifically, (13) suggests making  $t\sigma_\delta^2 = O(1)$  stabilize  $\text{Cor}(y_{it}, y_{jt})$ . Carrying this assessment further, for any two sites in close proximity, i.e. for  $d_{ij} \simeq 0$ ,

$$\text{Cor}(y_{it}, y_{jt}) \simeq \frac{\sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2}{\sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2} = 1,$$

a result that seems quite reasonable. For two sites very far apart so that  $d_{ij} \rightarrow \infty$ ,

$$\text{Cor}(y_{it}, y_{jt}) \rightarrow \frac{\sigma_\beta^2 + t\sigma_\delta^2}{\sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2} = \frac{\sigma_\beta^2 + O(1)}{\sigma_\beta^2 + O(1) + \sigma_\varepsilon^2}.$$

This correlation should be close to 0. In other words, we should have  $\sigma_\beta^2 + O(1) \ll \sigma_\varepsilon^2$ . A sufficient condition for this to hold is  $\sigma_\beta^2 \ll \sigma_\varepsilon^2$  and  $t\sigma_\delta^2 = O(1) \ll \sigma_\varepsilon^2$ .

The key result, (13), suggests a simple but straightforward way to adjust the model parameter  $\sigma_\delta^2$  according to the size of  $T$ , namely, to replace it by  $\sigma_\delta^2/T$ . That choice is empirically validated in Section 8.

We turn now to study the behavior of the predictive variances in the first-order polynomial DLM that helps us understand the interpolation results. To that end consider the correlations of responses at an ungauged site  $\mathbf{s}_0$  with those at the gauged site  $\mathbf{s}_j, j \in \{1, \dots, n\}$ , respectively. Note that both (15) and (16) hold for  $i = 0$ . The properties of the correlation structure in (15)–(16), lead us to the conjecture that the model's predictive bands should increase monotonically over time as more data become available, in the absence of restrictions on  $t\sigma_\delta^2 = O(1)$  suggested above. Furthermore, even conditioning on all the data, the predictive bands should also increase over time. In support of these conjectures, we prove that they hold in a simple case where  $n = 1$  and  $T = 2$  in (6)–(7).

**THEOREM 1.** *For the first-order polynomial DLM in (6)–(7) with  $n = 1$  and  $T = 2$ , assume the prior for  $\beta_0$  to be  $N(0, \sigma_\beta^2)$ . The joint distribution of  $\mathbf{y} = (y_{01}, y_{11}, y_{02}, y_{12})'$  is  $N(\mathbf{0}, \Sigma)$ , where*

$$\Sigma = (\sigma_\beta^2 + \sigma_\delta^2)\mathbf{1}_4'\mathbf{1}_4 + \text{block-diagonal}\{\sigma_\varepsilon^2 \exp(-V/\lambda), \sigma_\delta^2 \mathbf{1}_2'\mathbf{1}_2 + \sigma_\varepsilon^2 \exp(-V/\lambda)\},$$

$\mathbf{1}'_k$  being the  $k \times 1$  vector of 1s ( $k = 1, 2, \dots$ ). Then we have the following predictive conditional variances:

$$\text{Var}(y_{01}|y_{11}) = \frac{(\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)^2 - (\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{01}/\lambda))^2}{\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2}; \quad (17)$$

$$\text{Var}(y_{02}|y_{12}) = \frac{(\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)^2 - (\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{01}/\lambda))^2}{\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2}; \quad (18)$$

$$\text{Var}(y_{01}|y_{11}, y_{12}) = \frac{M_1}{\Delta}; \quad (19)$$

$$\text{Var}(y_{02}|y_{11}, y_{12}) = \frac{M_2}{\Delta}; \quad (20)$$

where

$$\Delta = (\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)(\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2) - (\sigma_\beta^2 + \sigma_\delta^2)^2, \quad (21)$$

$$M_1 = (\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)\{(\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)^2 - (\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{01}/\lambda))^2\} - 2(\sigma_\beta^2 + \sigma_\delta^2)^2(\sigma_\varepsilon^2 - \sigma_\varepsilon^2 \exp(-d_{01}/\lambda)), \quad (22)$$

and

$$M_2 = (\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)\{(\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)^2 - (\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{01}/\lambda))^2\} - 2(\sigma_\beta^2 + \sigma_\delta^2)^2(\sigma_\varepsilon^2 - \sigma_\varepsilon^2 \exp(-d_{01}/\lambda)). \quad (23)$$

For this simple case, we would expect the predictive variance of  $y_{01}$  based on more data collected over time to be no greater than that of  $y_{01}$  based on less, that is,

$$\text{Var}(y_{01}|y_{11}) \geq \text{Var}(y_{01}|y_{11}, y_{12})$$

and

$$\text{Var}(y_{02}|y_{12}) \geq \text{Var}(y_{02}|y_{11}, y_{12}).$$

Moreover, we would expect that, based on the same amount of data, the predictive variance of  $y_{01}$  would be no greater than that of  $y_{02}$ , that is,

$$\text{Var}(y_{01}|y_{11}, y_{12}) \leq \text{Var}(y_{02}|y_{11}, y_{12}).$$

Dou et al. (2007) prove these conjectures and provide other comparisons of these predictive variances. We conclude that the predictive variance function is a monotonic increasing function of time  $t$  based on the same set of data. It decreases when more data or equivalently, more time is involved. Furthermore, the difference between these predictive variances decreases as  $t$  increases. It increases with time even when conditioning on the same dataset.

**THEOREM 2.** *For the first-order polynomial DLM in Theorem 1, we have the following properties of the predictive conditional variances:*

$$\text{Var}(y_{01}|y_{11}) - \text{Var}(y_{01}|y_{11}, y_{12}) = \frac{\sigma_\varepsilon^4(\sigma_\beta^2 + \sigma_\delta^2)^2(1 - \exp(-d_{01}/\lambda))^2}{\Delta(\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)} \geq 0; \quad (24)$$

$$\text{Var}(y_{02}|y_{12}) - \text{Var}(y_{02}|y_{11}, y_{12}) = \frac{\sigma_\varepsilon^4(\sigma_\beta^2 + \sigma_\delta^2)^2(1 - \exp(-d_{01}/\lambda))^2}{\Delta(\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)} \geq 0; \quad (25)$$

$$\text{Var}(y_{02}|y_{11}, y_{12}) - \text{Var}(y_{01}|y_{11}, y_{12}) = \frac{\sigma_\varepsilon^4\sigma_\delta^2(1 - \exp(-d_{01}/\lambda))^2}{\Delta} \geq 0; \quad (26)$$

$$\text{Var}(y_{02}|y_{12}) - \text{Var}(y_{01}|y_{11}) = \frac{\sigma_\varepsilon^4 \sigma_\delta^2 (1 - \exp(-d_{01}/\lambda))^2}{(\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)(\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)} \geq 0; \quad (27)$$

$$\text{Var}(y_{01}|y_{11}) - \text{Var}(y_{01}|y_{11}, y_{12}) \geq \text{Var}(y_{02}|y_{12}) - \text{Var}(y_{02}|y_{11}, y_{12}); \quad (28)$$

$$\text{Var}(y_{02}|y_{12}) - \text{Var}(y_{01}|y_{11}) \leq \text{Var}(y_{02}|y_{11}, y_{12}) - \text{Var}(y_{01}|y_{11}, y_{12}). \quad (29)$$

As an immediate consequence of (26), the predictive variances increase monotonically at successive time points conditional on all the data. That leads to monotonically increasing coverage probabilities at the ungauged sites, an interesting phenomenon discussed in Section 7. There we will also discuss the lessons learned in this section in relation to our empirical findings.

Next, we present a curious result about the properties of the above predictive variances that may explain some of their key features. This result concerns these predictive variances as functions of  $\lambda$ ,  $d_{01}$  or  $\sigma_\varepsilon^2$ . Part of its proof is included in Appendix A.1.

**COROLLARY 1.** *The predictive conditional variances in (17)–(23) increase as  $d_{01}$  increases, or  $\sigma_\varepsilon^2$  increases, or  $\lambda$  decreases.*

Thus, keeping two parameters fixed, these predictive conditional variances are monotone functions of the remaining one. Therefore, the DLM can paradoxically lead to larger predictive variances when conditioning on more data. For example, in the case  $n = 1$  and  $T = 2$ , applying the DLM model with only the data at  $T = 2$  yields the predictive variance  $\text{Var}^*(y_{02}|y_{12})$ , which is exactly the same as  $\text{Var}(y_{01}|y_{11})$  in (17). This predictive variance is smaller than  $\text{Var}(y_{02}|y_{11}, y_{12})$  in (20), which is based on more data, under certain condition specified in the next corollary.

**COROLLARY 2.** *For the first-order polynomial DLM in Theorem 1,*

$$\text{Var}^*(y_{02}|y_{12}) < \text{Var}(y_{02}|y_{11}, y_{12}) \quad \text{if and only if} \quad \sigma_\varepsilon^2 > \sigma_\beta^2 \left( 1 + \frac{\sigma_\beta^2}{\sigma_\delta^2} \right). \quad (30)$$

The behavior suggested by Corollary 2 is actually observed in our application (see Section 7).

## 4. Implementation

This section very briefly describes how to implement our model using the MCMC method, more specifically, the forward-filtering-backward-sampling algorithm of Carter and Kohn (1994). The details are given by Dou et al. (2007).

### 4.1. Metropolis-within-Gibbs algorithm

The joint distribution,  $p(\lambda, \sigma^2, x_{1:T}, y_{1:T}^m, a_1, a_2 | y_{1:T}^o)$ , is the object of interest. Here  $y_{1:T}^o = (\mathbf{y}_1^o, \dots, \mathbf{y}_T^o)$  represents the observation matrix at the  $n$  gauged sites up to time  $T$ . Moreover,  $x_{1:T} = (x_1, \dots, x_T) : (2n+1) \times T$  is the vector of state parameters at the  $n$  gauged sites until time  $T$ . For simplicity, the values of  $\gamma$  are fixed but the problem of setting them will be addressed below. Additional detail can be found in Appendix A.2.

Since that joint distribution does not have a closed form, direct sampling methods fail, leading to the use of the Markov Chain Monte Carlo (MCMC) method. A *blocking MCMC* scheme increases iterative sampling efficiency, three blocks being chosen for reasons given in Dou et al. (2007):  $(\lambda, \sigma^2, x_{1:T})$ ,  $y_{1:T}^m$  and  $(a_1, a_2)$ . More precisely we can:

- (i) sample from  $p(x_{1:T}, \lambda, \sigma^2 | a_1, a_2, y_{1:T})$
- (ii) sample from  $p(y_{1:T}^m | \lambda, \sigma^2, x_{1:T}, a_1, a_2, y_{1:T}^o)$  and
- (ii) sample from  $p(a_1, a_2 | x_{1:T}, \lambda, \sigma^2, y_{1:T})$ .

Since  $p(\lambda, \sigma^2, x_{1:T}|a_1, a_2, y_{1:T})$  has no closed form, the full conditional posterior distribution of  $x_{1:T}$  is obtained by Kalman filtering and smoothing, in other words, by the FFBS algorithm. Assuming an inverse Gamma hyperprior for  $\sigma^2$ , the conditional posterior distribution of  $\sigma^2$  given the range and phase parameters is also inverse Gamma distributed with new shape and scale parameters. Note that

$$p(\lambda, \sigma^2, x_{1:T}|a_1, a_2, y_{1:T}) = p(\lambda|a_1, a_2, y_{1:T})p(\sigma^2|\lambda, a_1, a_2, y_{1:T}) \times p(x_{1:T}|\lambda, \sigma^2, a_1, a_2, y_{1:T}), \quad (31)$$

indicating that we can sample iteratively from the three conditional posterior distributions on the right-hand-side of (31) to obtain samples from  $p(\lambda, \sigma^2, x_{1:T}|a_1, a_2, y_{1:T})$ . However,  $p(\lambda|a_1, a_2, y_{1:T})$  has no closed form, leading us to sample  $\lambda$  by a *Metropolis–Hasting* chain within a Gibbs sampling cycle, an algorithm as described in the next three subsections.

### Sampling from $p(\lambda, \sigma^2, x_{1:T}|a_1, a_2, y_{1:T})$

To sample  $(\lambda, \sigma^2, x_{1:T})$  from  $p(\lambda, \sigma^2, x_{1:T}|a_1, a_2, y_{1:T})$ , we use the block MCMC scheme. Because of (31), we could ideally iteratively sample  $\lambda$  from  $p(\lambda|a_1, a_2, y_{1:T})$ ,  $\sigma^2$  from  $p(\sigma^2|\lambda, a_1, a_2, y_{1:T})$  and  $x_{1:T}$  from  $p(x_{1:T}|\lambda, \sigma^2, a_1, a_2, y_{1:T})$ . However, because we do not have a closed form for the posterior density of  $p(\lambda|a_1, a_2, y_{1:T})$ , we use instead the *Metropolis–Hasting algorithm* to sample  $\lambda$ , given the data, from the following a quantity that is proportional to its posterior density, that is,

$$p(\lambda|a_1, a_2, y_{1:T}) \propto p(\lambda) \prod_{t=1}^T |Q_t|^{-\frac{1}{2}} \left[ \beta + \frac{1}{2} \sum_{t=1}^T \mathbf{e}_t' Q_t^{-1} \mathbf{e}_t \right]^{-(nT/2+\alpha)}. \quad (32)$$

Since we cannot compute the normalization constant for  $p(\lambda|a_1, a_2, y_{1:T})$ , the Metropolis–Hasting algorithm is used. The proposal density,  $q(., .)$ , is selected to be a lognormal distribution, because the parameter space is bounded below by 0, making the Gaussian distribution inappropriate. As Moller (2003) points out, this alternative to a random walk Metropolis considers the proposal move to be a random multiple of the current state. From the current state  $\lambda^{(j-1)}$  ( $j > 1$ ), the proposed move is  $\lambda^* = \lambda^{(j-1)} e^Z$ , where  $Z$  is drawn from a symmetric density, such as normal. In other words, at iteration ( $j$ ), we sample a new  $\lambda^*$  from this proposal distribution, centered at the previously sampled  $\lambda^{(j-1)}$  with a tuning parameter,  $\tau^2$ , as the variance for the distribution of  $Z$ . Gamerman (2006) suggests the acceptance rate, that is, the ratio of accepted  $\lambda^*$  to the total number of iterations, be around 50%. We tune  $\tau^2$  to attain that rate. If the acceptance rate were too high, for example, 70% to 100%, we would increase  $\tau^2$ . If too low, for example, 0 to 20%, we would decrease  $\tau^2$ , to narrow down the search area for  $\lambda^*$ .

The Metropolis–Hasting algorithm proceeds as follows. Given  $\lambda^{(j-1)}$ ,  $a_1^{(j-1)}$ ,  $a_2^{(j-1)}$  and  $y_{1:T}^{(j-1)}$ , where  $j > 1$ :

- Draw  $\lambda^*$  from  $LN(\lambda^{(j-1)}, \tau^2)$ .
- Compute the acceptance probability:

$$\alpha(\lambda^{(j-1)}, \lambda^*) = \min \left\{ 1, \frac{p(\lambda^*|a_1^{(j-1)}, a_2^{(j-1)}, y_{1:T}^{(j-1)})/q(\lambda^{(j)}, \lambda^*)}{p(\lambda^{(j-1)}|a_1^{(j-1)}, a_2^{(j-1)}, y_{1:T}^{(j-1)})/q(\lambda^*, \lambda^{(j-1)})} \right\}.$$

- Accept  $\lambda^*$  with probability  $\alpha(\lambda^{(j-1)}, \lambda^*)$ . In other words, sample  $u \sim U[0, 1]$  and let  $\lambda^{(j)} = \lambda^*$  if  $\lambda^* < u$  and  $\lambda^{(j)} = \lambda^{(j-1)}$  otherwise.

We run this algorithm iteratively until convergence is reached.

Next, we sample  $\sigma^2$  given the accepted  $\lambda$ 's,  $a_1$ 's,  $a_2$ 's and  $y_{1:T}$ . The prior for  $\sigma^2$  is chosen to be an inverse gamma distribution with shape parameter  $\alpha$  and scale parameter  $\beta$ . The posterior



distribution for  $\sigma^2$  is also an inverse gamma distribution, but with a shape parameter  $\alpha + \frac{nT}{2}$  and a scale parameter  $\beta + \frac{1}{2} \sum_{t=1}^T \mathbf{e}_t' Q_t^{-1} \mathbf{e}_t$ .

We now sample  $x_{1:T}$  given the accepted  $\lambda$ 's,  $\sigma^2$ 's, phase parameters and  $y_{1:T}$ , using FFBS. West and Harrison (1997) propose a general theorem for inference about the parameters in the DLM framework. For time series data, the usual method for updating and predicting is the Kalman filter. Dou et al. (2007) present a FFBS algorithm (similar to the Kalman filter algorithm) to resample the state parameters conditional on all the other parameters and observations as part of the MCMC method for sampling  $x_{1:T} = (\mathbf{x}_1, \dots, \mathbf{x}_T)$  from the smoothing distribution  $p(\mathbf{x}_t | \lambda, \sigma^2, a_1, a_2, y_{1:T})$ .

The initial state parameter is given by

$$(\mathbf{x}_0 | \mathbf{y}_0, \theta) \sim N[\mathbf{m}_0, \sigma^2 C_0], \quad (33)$$

where  $\mathbf{y}_0$  being the initial information, with  $\mathbf{m}_0$  and  $C_0$  known. Later in Section 6, we consider how to set them for Cluster 2 AIRS dataset (1995). Let  $\theta = (\lambda, \sigma^2, a_1, a_2, \gamma)$ . Now suppose for expository purposes, that all the prior information has been given and  $\theta$ 's coordinates are mutually independent.

**Sampling from**  $p(y_{1:T}^m | \lambda, \sigma^2, x_{1:T}, a_1, a_2, y_{1:T}^o)$

MCMC can be used to fill in missing values at each iteration. To see how, note that at any fixed time point  $t$ , after appropriately defining a scale matrix  $R_t$ , we can rewrite the observation vector  $\mathbf{y}_t$  as follows:

$$R_t \mathbf{y}_t = \begin{pmatrix} y_t^m \\ y_t^o \end{pmatrix},$$

where  $y_t^m : n_t \times 1$  denotes the missing response(s) at time  $t$  and  $y_t^o : (n - n_t) \times 1$  the observed response(s) at  $t$ . Notice that ‘‘o’’ represents ‘‘observed’’ and ‘‘m’’, ‘‘missing’’.

Let  $R_t = (\mathbf{e}_{n_1}, \dots, \mathbf{e}_{n_t}, \mathbf{e}_{k_1}, \dots, \mathbf{e}_{k_{n-n_t}})'$ , where  $\{\mathbf{s}_{n_j} : j = 1, \dots, t\}$  represents the set of gauged sites containing missing values at time point  $t$ ,  $\{\mathbf{s}_{k_j} : j = 1, \dots, n - n_t\}$  the set of gauged sites containing observed values at time  $t$ , for all  $t = 1, \dots, T$ ; and  $\mathbf{e}_j = (e_{j1}, \dots, e_{jn})' : n \times 1$  such that  $e_{jk} = I_{j=k}$  for  $k = 1, \dots, j$  and  $j \in \mathcal{Z}^+$ .

We already know that

$$(\mathbf{y}_t | \lambda, \sigma^2, \mathbf{x}_t, \mathbf{a}) \sim N[F_t' \mathbf{x}_t, \sigma^2 \exp\{-V/\lambda\}],$$

so that  $R_t \mathbf{y}_t$  is also multivariate normally distributed, that is,

$$(R_t \mathbf{y}_t | \lambda, \sigma^2, \mathbf{x}_t, \mathbf{a}) = ((\mathbf{y}_t^m, \mathbf{y}_t^o)' | \lambda, \sigma^2, \mathbf{x}_t, \mathbf{a}) \sim N[\tilde{\mu}_t, \tilde{\Sigma}_t],$$

where

$$\begin{aligned} \tilde{\mu}_t &= R_t F_t' \mathbf{x}_t \\ \tilde{\Sigma}_t &= \sigma^2 R_t \exp\{-V/\lambda\} R_t' \end{aligned}$$

We can also partition  $\tilde{\mu}_t$  as  $\tilde{\mu}_t = (\tilde{\mu}_t^{m'}, \tilde{\mu}_t^{o'})'$ , where  $\tilde{\mu}_t^{m'} : n_t \times 1$  and  $\tilde{\mu}_t^{o'} : (n - n_t) \times 1$ . Similarly, we have

$$\tilde{\Sigma}_t = \begin{pmatrix} \tilde{\Sigma}_t^{mm} & \tilde{\Sigma}_t^{mo} \\ \tilde{\Sigma}_t^{om} & \tilde{\Sigma}_t^{oo} \end{pmatrix},$$

where  $\tilde{\Sigma}_t^{mm} : n_t \times n_t$ ,  $\tilde{\Sigma}_t^{mo} : n_t \times (n - n_t)$  and  $\tilde{\Sigma}_t^{oo} : (n - n_t) \times (n - n_t)$ .

By a standard property of the multivariate normal distribution, we have

$$(\mathbf{y}_t^m | \lambda, \sigma^2, \mathbf{x}_t, \mathbf{a}, \mathbf{y}_t^o) \sim N[\mu_t^{**}, \Sigma_t^{**}], \quad (34)$$

where

$$\mu^{**} = \tilde{\mu}_t^m + \tilde{\Sigma}_t^{mo}(\tilde{\Sigma}_t^{oo})^{-1}(\mathbf{y}_t^o - \tilde{\mu}_t^o), \quad (35)$$

and

$$\Sigma_t^{**} = \tilde{\Sigma}_t^{mm} - \tilde{\Sigma}_t^{mo}(\tilde{\Sigma}_t^{oo})^{-1}\tilde{\Sigma}_t^{om}, \quad (36)$$

for  $t = 1, \dots, T$ .

At each iteration, we draw  $\mathbf{y}_t^m$  from the corresponding distribution (34) at each time point  $t$  and then we can write the response variables as  $y_{1:T} = (y_{1:T}^m, y_{1:T}^o)$  where  $y_{1:T}^m = (y_1^m, \dots, y_T^m)$  and  $y_{1:T}^o = (y_1^o, \dots, y_T^o)$ .

### Sampling from $p(a_1, a_2 | \lambda, \sigma^2, x_{1:T}, y_{1:T})$

We now present our method for sampling the phase parameters  $\mathbf{a} = (a_1, a_2)'$  from its full conditional posterior distribution, that is,  $p(\mathbf{a} | \lambda, \sigma^2, x_{1:T}, y_{1:T})$ , by using the samples of  $\lambda$ ,  $\sigma^2$  and  $x_{1:T}$ . For simplicity, we use the notation for models (1)–(3) in this section.

We then sample the constant phase parameters conditional on all the other parameters and observations. Suppose  $\mathbf{a} = (a_1, a_2)'$  has a conjugate bivariate normal prior with mean vector  $\mu^o = (\mu_{1o}, \mu_{2o})'$  and covariance matrix  $\Sigma^o$ . Then the posterior conditional distribution for  $\mathbf{a}$  is normal with mean vector  $\mu^*$  and covariance matrix  $\Sigma^*$ , where  $\mu^*$  and  $\Sigma^*$  can be obtained from equations given in Dou et al. (2007).

We will not use a non-informative prior such as  $p(\mathbf{a}) \propto 1$  for  $\mathbf{a}$  since that choice can lead to non-identified posterior means or posterior variances. In fact for that choice we find the posterior conditional distribution of  $\mathbf{a}$  to be normal with mean vector  $\mu = (\mu_1, \mu_2)'$  and covariance matrix  $\Sigma$  from equations given in Dou et al. (2007) along with the elements of  $\Sigma$ , where  $\Sigma$  can be singular for any  $t = 12k$  ( $k \in \mathbb{Z}$ ). Hence, we obtain the extreme values at times 12, 24,  $\dots$ , 2880, that invalidates the assumption of constant phase parameters across all the time scales when we sample from its full conditional posterior distribution.

For fixed values of  $\lambda$ ,  $\sigma^2$  and  $x_{1:T}$ , we sample the model parameter  $\mathbf{a} = (a_1, a_2)$  from  $N(\mu^*, \Sigma^*)$  at each time point, and then obtain the “sample” of  $\mathbf{a}$  at this iteration by the median of these samples across all the time points, under the assumption that  $(a_1, a_2)$  are constant phase parameters in the models (4)–(5).

## 4.2. Summary

The MCMC algorithm we use here resembles that of Huerta et al. (2004), one difference being that we unlike them, use all the samples after the burn-in period, not just the chain containing the accepted samples. We believe the Markov chains of only accepted results will lead to biased samples, thereby changing the detailed balance equation of the Metropolis–Hasting algorithm.

The above algorithm we use for Cluster 2 AIRS dataset is summarized as follows:

1. Initialization: sample

$$\begin{aligned} \lambda^{(1)} &\sim IG(\alpha_\lambda, \beta_\lambda) \\ \sigma^{2(1)} &\sim IG(\alpha_{\sigma^2}, \beta_{\sigma^2}) \\ x_{1:T}^{(1)} &\sim N(m_0, \sigma^{2(1)} C_0). \end{aligned}$$

2. Given the  $(j-1)$ <sup>th</sup> value  $\lambda^{(j-1)}$ ,  $\sigma^{2(j-1)}$ ,  $x_{1:T}^{(j-1)}$ ,  $y_{1:T}^m{}^{(j-1)}$ ,  $a_1^{(j-1)}$ ,  $a_2^{(j-1)}$  and the observations  $y_{1:T}^o$ :

- (1) Sample  $(\lambda^{(j)}, \sigma^{2(j)}, x_{1:T}^{(j)})$  from  $p(\lambda, \sigma^2, x_{1:T} | a_1^{(j-1)}, a_2^{(j-1)}, y_{1:T}^{(j-1)})$ , where

$$y_{1:T}^{(j-1)} = (y_{1:T}^m{}^{(j-1)}, y_{1:T}^o).$$

- (i) • Generate a candidate value  $\lambda^*$  from a logarithm proposal distribution  $q(\lambda^{(j-1)}, \lambda)$ , that is,  $LN(\lambda^{(j-1)}, \tau^2)$  for some suitable tuning parameter  $\tau^2$ .  
 • Compute the acceptance ratio  $\alpha(\lambda^{(j-1)}, \lambda^*)$  where

$$\alpha(\lambda^{(j-1)}, \lambda^*) = \min \left\{ 1, \frac{p(\lambda^* | a_1^{(j-1)}, a_2^{(j-1)}, y_{1:T}^{(j-1)}) \lambda^*}{p(\lambda^{(j-1)} | a_1^{(j-1)}, a_2^{(j-1)}, y_{1:T}^{(j-1)}) \lambda^{(j-1)}} \right\}$$

- With probability  $\alpha(\lambda^{(j-1)}, \lambda^*)$  accept the candidate value and set  $\lambda^{(j)} = \lambda^*$ ; otherwise reject and set  $\lambda^{(j)} = \lambda^{(j-1)}$ .  
 (ii) Sample  $\sigma^{2(j)}$  from  $p(\sigma^2 | \lambda^{(j)}, a_1^{(j-1)}, a_2^{(j-1)}, y_{1:T}^{(j-1)})$ .  
 (iii) Sample  $x_{1:T}^{(j)}$  from  $p(x_{1:T} | \lambda^{(j)}, \sigma^{2(j)}, a_1^{(j-1)}, a_2^{(j-1)}, y_{1:T}^{(j-1)})$ .

(2) Sample  $y_{1:T}^m^{(j)}$  from  $p(y_{1:T}^m | \lambda^{(j)}, \sigma^{2(j)}, x_{1:T}^{(j)}, a_1^{(j-1)}, a_2^{(j-1)}, y_{1:T}^o)$ .

(3) Sample  $(a_1^{(j)}, a_2^{(j)})$  from  $p(a_1, a_2 | \lambda^{(j)}, \sigma^{2(j)}, x_{1:T}^{(j)}, y_{1:T}^{(j)})$ , where  $y_{1:T}^{(j)} = (y_{1:T}^m^{(j)}, y_{1:T}^o)$ .

3. Repeat until convergence.

We have developed software to implement the DLM approach of this section. To enhance the Metropolis–within–Gibbs algorithm, we augment the R code with C to speed up the computation. The current version, *GDLM.1.0*, is freely available at <http://enviro.stat.ubc.ca> for different platforms such as Windows, Unix and Linux.

## 5. Interpolation and prediction

This section describes how to interpolate hourly ozone concentrations at ungauged sites using the DLM and the simulated Markov chains for the model parameters (see Section 4). In other words, suppose  $\mathbf{s}_1, \dots, \mathbf{s}_u$  are  $u$  ungauged sites of interest within the geographical region of Cluster 2 sites (excluding the possibility of extrapolation). The objective is to draw samples from

$$p(y_{1:T}^s | \lambda, \sigma^2, x_{1:T}, a_1, a_2, y_{1:T}),$$

where  $y_{1:T}^s = (\mathbf{y}_1^s, \dots, \mathbf{y}_T^s) : 1 \times T$  and  $y_t^s$  denotes the unobserved square-root of ozone concentrations at the ungauged site  $\mathbf{s}$  and time  $t$ , for  $t = 1, \dots, T$  and for  $\mathbf{s} \in \{\mathbf{s}_1, \dots, \mathbf{s}_u\}$ . Let  $(\alpha_{1t}^s, \alpha_{2t}^s)$  denote the unobserved state parameters at site  $\mathbf{s}$  and time  $t$ . The DLM is given by

$$\mathbf{y}_t^{\text{new}} = \mathbf{1}_{n+1}' \beta_t + S_{1t}(a_1) \alpha_{1t}^{\text{new}} + S_{2t}(a_2) \alpha_{2t}^{\text{new}} + \nu_t^{\text{new}}, \quad (37)$$

where  $\mathbf{y}_t^{\text{new}} = (y_t^s, \mathbf{y}_t^o)'$ ,  $\alpha_{\mathbf{t}}^{\text{new}} = (\alpha_{1t}^s, \alpha_{1t}^o, \alpha_{2t}^s, \alpha_{2t}^o)'$ , and  $\nu_t^{\text{new}} \sim N(0, \sigma^2 \exp(-V^{\text{new}}/\lambda))$ .

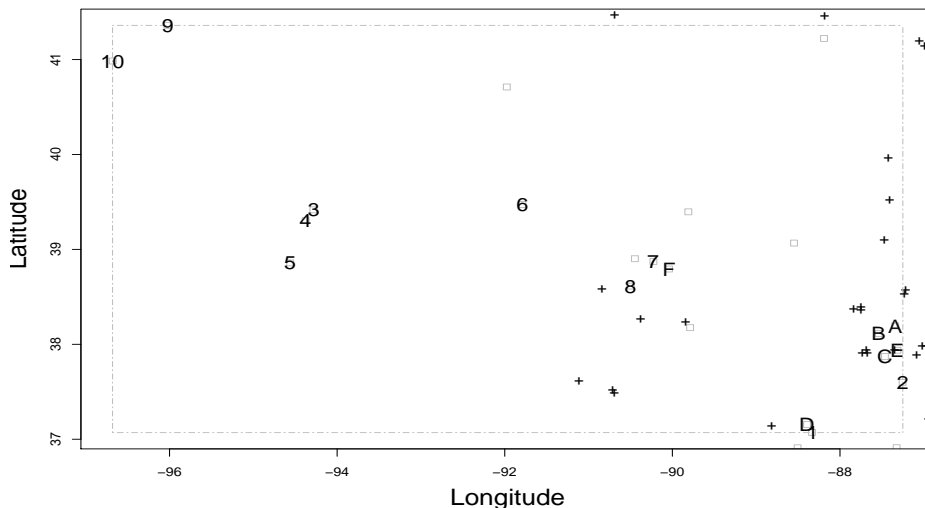
In the following two subsections, we illustrate how to sample the unobserved state parameters  $\{(\alpha_{1t}^s, \alpha_{2t}^s) : t = 1, \dots, T\}$  from the corresponding conditional posterior distribution, and demonstrate the spatial interpolation at the ungauged site  $\mathbf{s}$ .

### Sampling the unobserved state parameters

We first sample  $\alpha_{jt}^s$  given  $\alpha_{j,t-1}^s$ ,  $\alpha_{\mathbf{j}t}$  and  $\alpha_{\mathbf{j},t-1}$ ,  $j = 1, 2$ . From the state equation (5) for  $\alpha_{\mathbf{j}t}^{\text{new}}$ , we know that the joint density of  $\alpha_{jt}^s$  and  $\alpha_{\mathbf{j}t}$  follows a normal distribution, with covariance matrix  $\sigma^2 \tau_j^2 \exp(-V^{\text{new}}/\lambda_j)$ , where  $V^{\text{new}}$  denotes the distance matrix for the unobserved station and the monitoring stations. The conditional posterior distribution,

$$p(\alpha_{jt}^s | \alpha_{j,t-1}^s, \lambda, \sigma^2, \beta_t, \alpha_{1t}, \alpha_{2t}, a_1, a_2, y_{1:T}),$$

is derived in Appendix A.3.



**Fig. 1.** Schematic representation of the locations of ten gauged sites in Cluster 2 and the randomly chosen six ungauged sites. (Number = Cluster 2 sites and letter = ungauged sites.)

### *Spatial interpolation at ungauged sites*

We interpolate the square-root of ozone concentration at the ungauged sites by conditioning on all the other parameters and observations at the gauged sites. As above,  $y_t^s$  and  $\mathbf{y}_t$  are jointly normally distributed as a consequence of the observation equation. The predictive conditional distribution for  $y_t^s$ , that is,  $p(y_t^s | \alpha_{1t}^s, \alpha_{2t}^s, \lambda, \sigma^2, \beta_t, \alpha_{1t}, \alpha_{2t}, a_1, a_2, y_{1:T})$ , is given in Appendix A.3.

## 6. Application

This section applies our model to the hourly ozone concentration field described above. Six ungauged sites were randomly selected from those available within the range of the sites in Cluster 2 to play the role of “unmonitored sites” and help us assess the performance of the DLM. The geographical locations of these six ungauged sites, represented by the alphabetic letters,  $A, \dots, F$ , are shown in Figure 1, along with the sites in Cluster 2.

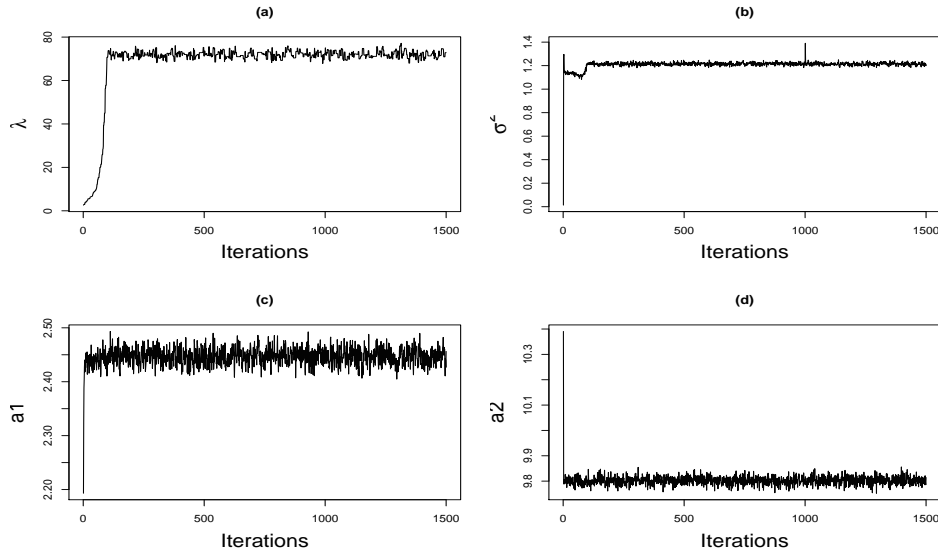
### 6.1. MCMC sampling

This subsection presents a MCMC simulation study in which samples are drawn sequentially from the joint posterior distribution of the model parameters in the DLM.

#### *Initial settings*

Following Huerta et al. (2004), we use the following initial settings for the starting values, hyperpriors and fixed model parameters in the DLM:

- The hyperprior for  $\lambda$  is  $IG(1, 5)$  and for  $\sigma^2$ ,  $IG(2, 0.01)$ . The expected value of  $IG(1, 5)$  is  $\infty$  and so are both of the variances of  $p(\lambda)$  and  $p(\sigma^2)$ . These vague priors for  $\lambda$  and  $\sigma^2$  are selected to reflect our lack of prior knowledge about their distributions.
- The initial information for  $x_0$ , the initial state parameter, is assumed to be normally distributed with mean vector  $\mathbf{m}_0 = (2.85, -0.75\mathbf{1}'_n, -0.08\mathbf{1}'_n)'$  and covariance matrix  $\sigma_1^2 C_0$ , where  $\sigma_1^2 \sim IG(2, 0.01)$  and  $C_0$  is a block diagonal matrix with diagonal entries 1,  $0.01\mathbf{1}'_n$  and  $0.01\mathbf{1}'_n$ .



**Fig. 2.** Traces of model parameters with the number of iterations of the Markov chains. The parameters are: (a)  $\lambda$ , the range parameter; (b)  $-\sigma^2$ , the variance parameter; (c)  $-a_1$ , the phase parameter with respect to the 24-hour periodicity; and (d)  $-a_2$ , the phase parameter with respect to the 12-hour periodicity.

**Table 1.** Posterior summaries for  $\lambda$ ,  $\sigma^2$ ,  $a_1$  and  $a_2$ .

Quantile	$\lambda$	$\sigma^2$	$a_1$	$a_2$
2.5%	69.29	1.19	2.42	9.77
Median	71.83	1.21	2.45	9.80
97.5%	75.37	1.24	2.48	9.84

- The hyperprior for  $\mathbf{a}$  is a bivariate normal distribution with mean vector  $\mu^o = (2.5, 9.8)'$  and a diagonal matrix  $\Sigma^o$  with diagonal entries 0.5 and 0.5.
- Some of the model parameters in the DLM are fixed as follows:  $\tau_y^2 = 0.02$ ,  $\tau_1^2 = 0.0002$ ,  $\tau_2^2 = 0.0004$ ,  $\lambda_1 = 25$  and  $\lambda_2 = 25$ .

### Monitoring the convergence of the Markov chains

Figure 2 shows the trace plots of model parameters  $\lambda$ ,  $\sigma^2$ ,  $a_1$  and  $a_2$  with the number of iterations of the simulated Markov chains where the total number of iterations is 4,268. The burn-in period is chosen to be 2,269 and all the remaining Markov samples are collected for posterior inference. The acceptance rate is approximately 62%. We observe that the Markov Chain converges after a run of less than five hundreds iterations.

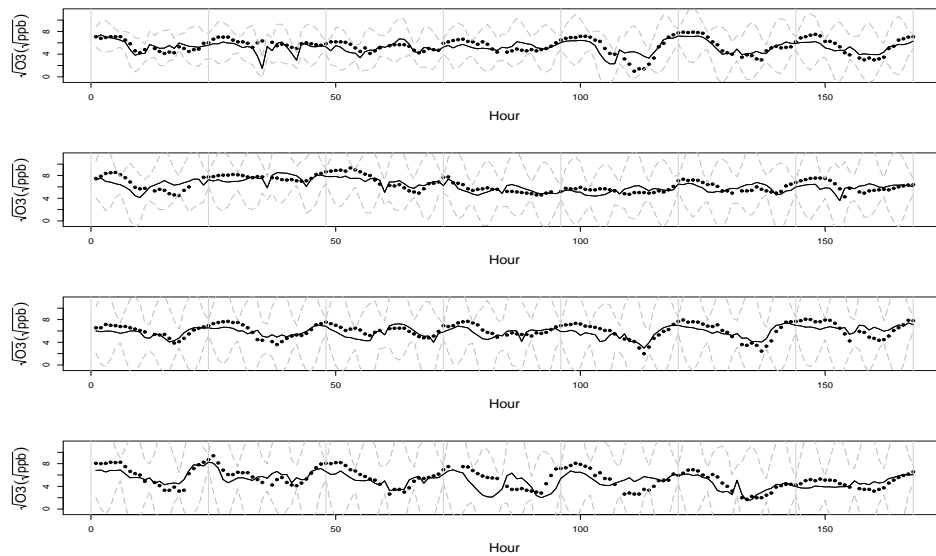
Table 1 displays the median and 95% quantile from the simulated Markov chains for the model parameters  $\lambda$ ,  $\sigma^2$ ,  $a_1$  and  $a_2$ .

### 6.2. Spatial interpolation

This subsection assesses the model's performance by comparing the interpolated values at the ungauged sites,  $A, \dots, F$ , with the measurements made there. We use the entire dataset to assess the performance of the interpolation results. Table 2 shows the coverage probabilities of the credibility intervals (or "credible intervals" for short) for these six ungauged sites at various nominal levels.

**Table 2.** Comparisons between the empirical credible probability and the nominal levels at the ungauged sites  $A, \dots, F$ .

Nominal Prob (%)	Coverage Prob.s (%)					
	$A$	$B$	$C$	$D$	$E$	$F$
95	94.9	96.9	96.5	99.7	96.1	98.1
90	91.9	93.7	93.5	99.4	93.6	96.8
80	84.8	88.5	88.2	97.7	89.6	94.3
70	78.7	83.5	83.3	94.0	85.8	90.6
60	73.0	78.5	77.1	89.7	81.6	86.6
50	65.2	71.5	70.4	85.6	76.1	81.4
40	55.2	61.4	61.0	79.2	67.9	74.7
30	42.2	47.6	47.5	69.6	54.9	64.4



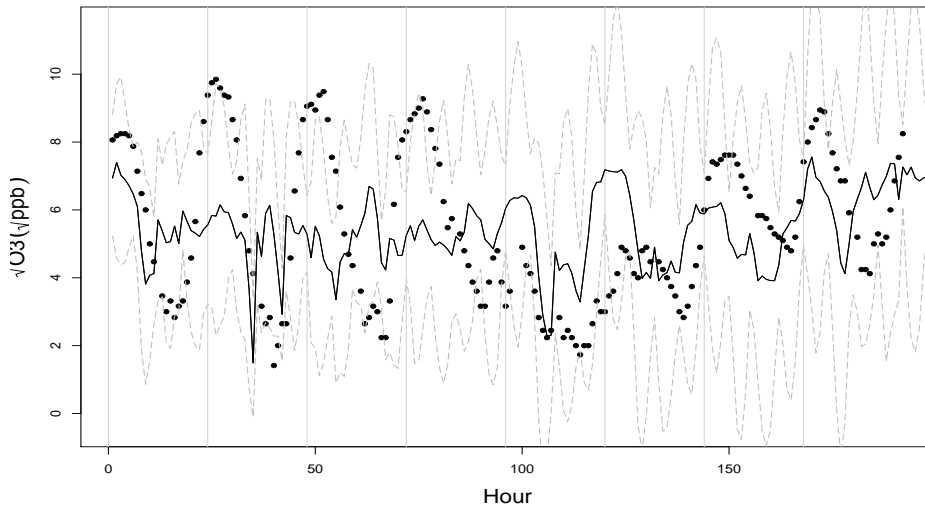
**Fig. 3.** Interpolation at Ungauged Site  $D$  for four successive weeks beginning from May 14, 1995. The square-root of hourly ozone concentrations are plotted on the vertical axes, hours on the horizontal axes. The solid lines represent the predicted median of the responses, the dashed lines represent the 95% predictive intervals for the predicted square-root of ozone concentrations and the solid dots represent the observations at Ungauged Site  $D$ .

Generally, the coverage probabilities at the ungauged sites exceed their nominal levels indicating that the error bands are too wide.

Among these six ungauged sites, Site  $D$  has the highest coverage probability seen in Table 2. This may be because of  $D$ 's nearness to a close “relative” among the gauged sites, namely, Site 1. That would be consistent with our assumption that the spatial correlation is inversely proportional to the intersite distance. At the same time, these unsatisfactory large coverage probabilities point to a deficiency of the DLM.

To explore this issue further, we compared the values predicted for Ungauged Site  $D$  from May 14 to September 11, 1995 and the measurements made there. Figures 3 and in more detail 4, which exemplify results reported in more detail by Dou et al. (2007), depict the results for the first four weeks and the last week of that period, respectively.

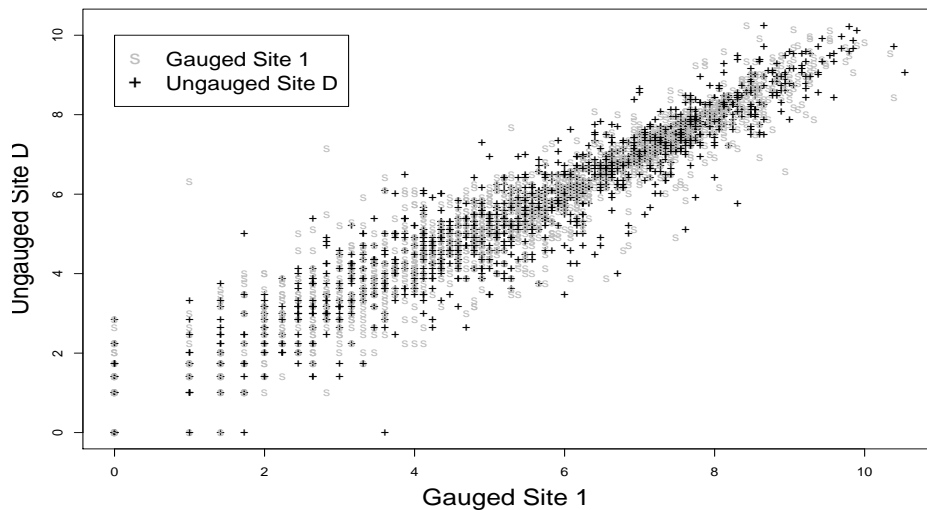
Furthermore, Table 3 shows for all the ungauged sites, the close relatives they have among the Cluster 2 sites that lie within a radius of 100 km, the corresponding global circle distance (GCD) in km, and along with the average of their correlations. This table confirms that indeed  $D$  does enjoy the highest correlation with its relative. That relationship is further explored in Figure 5 where



**Fig. 4.** Interpolation at Ungauged Site D from the 17th week to the 120th day. The square-root of hourly ozone concentrations are plotted in the vertical axes, hours on the horizontal axes.

**Table 3.** Close “relatives” of the ungauged sites, their global circle distance (km) and the average of their correlation with their associated gauged sites.

Ungauged Site	Relative(s)	GCD (km)	Pearson's r
<i>A</i>	2	66.6	0.73
<i>B</i>	2	62.5	0.74
<i>C</i>	2	35.5	0.84
<i>D</i>	1	11.0	0.95
<i>E</i>	2	38.0	0.70
<i>F</i>	(7, 8)	(18.6, 44.9)	(0.84, 0.82)



**Fig. 5.** Scatterplot for the square-root of ozone concentrations at Ungauged Site  $D$  and its close relative, Gauged Site 1. The square-root of hourly ozone concentrations are plotted in both vertical and horizontal axes.

we see a strong linear relationship between Sites  $D$  and 1 as our coverage probability assessment had suggested.

In spite of its reliance on the relatives, the DLM does not predict responses at the ungauged sites very accurately as illustrated in Figure 4. That points to problems with this model which will be discussed in the next section.

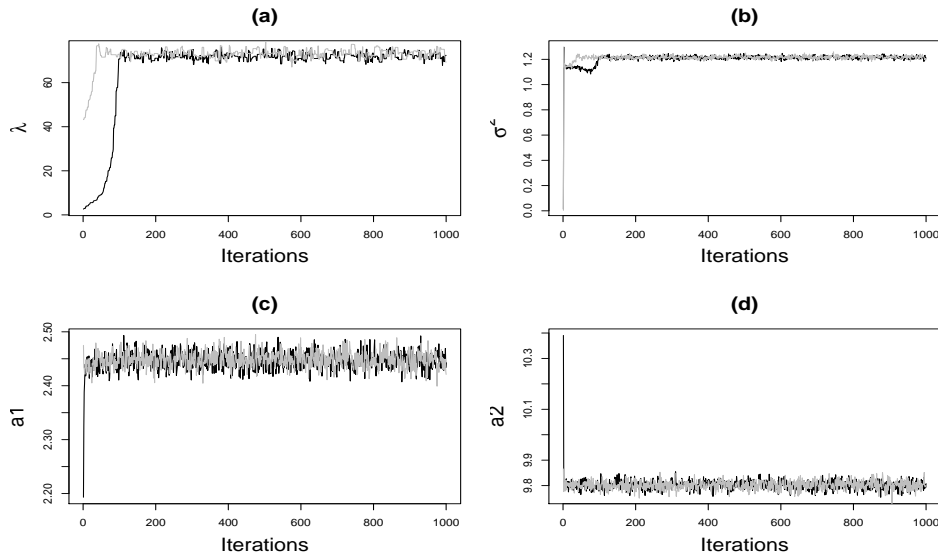
## 7. Discussion

In general, the DLM provides a remarkably powerful modelling tool, made practical by advances in statistical computing. However, its substantial computational requirements still limits its applicability. Moreover, the very flexibility that makes it so powerful also imposes an immense burden of choice on the model. This section summarizes critical issues and includes some suggestions for improvement.

### *Monitoring MCMC convergence*

Figure 6 represents the trace plots of model parameters  $\lambda$ ,  $\sigma^2$ ,  $a_1$  and  $a_2$  of two chains from the initial settings in Section 6.1. These two chains seem to mix well after several hundreds iterations, suggesting at first glance the Markov chains have converged.





**Fig. 6.** Traces of model parameters for a number of iterations of two chains. The parameters are: (a)  $-\lambda$ , the range parameter; (b)  $-\sigma^2$ , the variance parameter; (c)  $-a_1$ , the phase parameter with respect to the 24-hour periodicity; and (d)  $-a_2$ , the phase parameter with respect to the 12-hour periodicity.

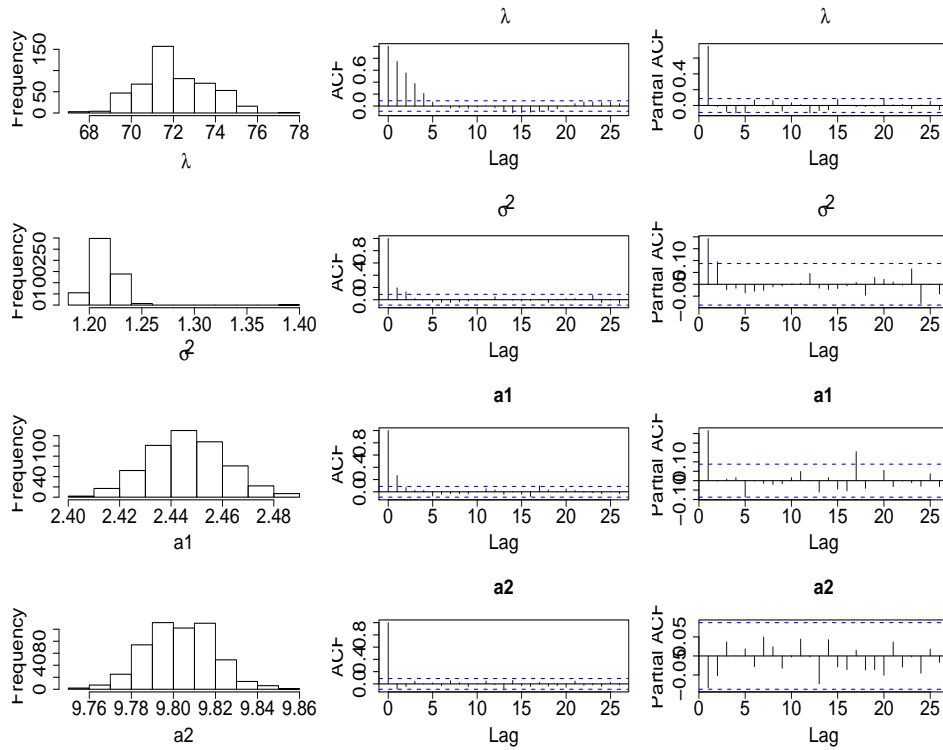
### *Autocorrelation and partial autocorrelation of the simulated Markov chains*

However, we know that the autocorrelation, as measured by the autocorrelation function (ACF), is very important when considering the length of the chain. A highly auto-correlated chain needs a long run to yield accurate estimates. Moreover, the partial autocorrelation function (PACF) is also an important index for assessing a Markov chain since large values of the PACF at lag  $h$  indicates that the next value in the chain is dependent on past values, not just on the most recent ones.

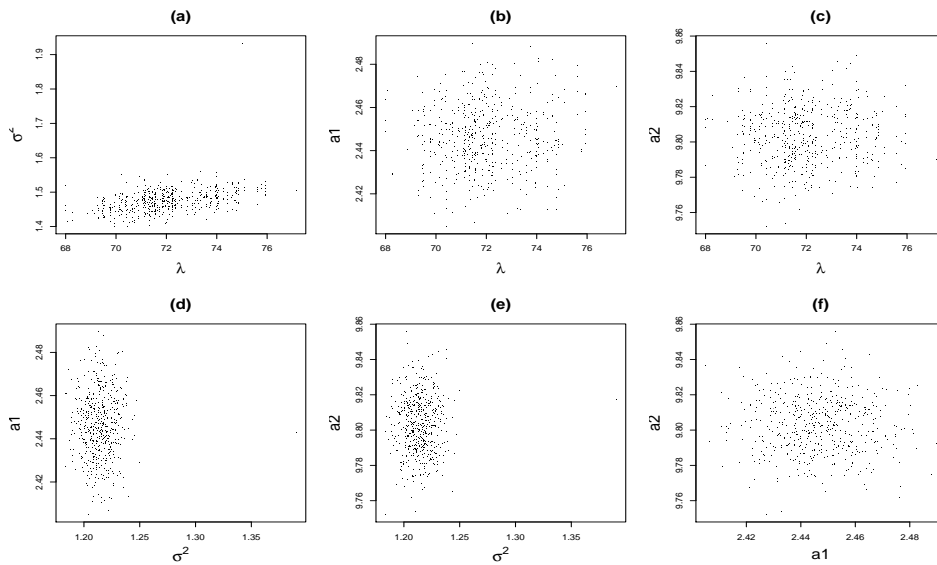
Figure 7 shows the histogram, ACF and PACF plots for the Markov chains used in Section 6.2, after a burn-in period of 1,000. The ACF plots show the  $\lambda$ -chain to be highly autocorrelated, in other words that the  $\lambda$ -chain does not mix well, potentially leading to biased estimates in Section 6.2. Thinning the chain might reduce that autocorrelation. In other words, using every  $k^{\text{th}}$  ( $k > 1, k \in \mathcal{Z}^+$ )  $\lambda$  generated by the chain could be used to produce the estimates. However, computational challenges make that strategy impractical; we need to use the entire chain.

### *Relationship between pairs of $\lambda$ , $\sigma^2$ , $a_1$ and $a_2$*

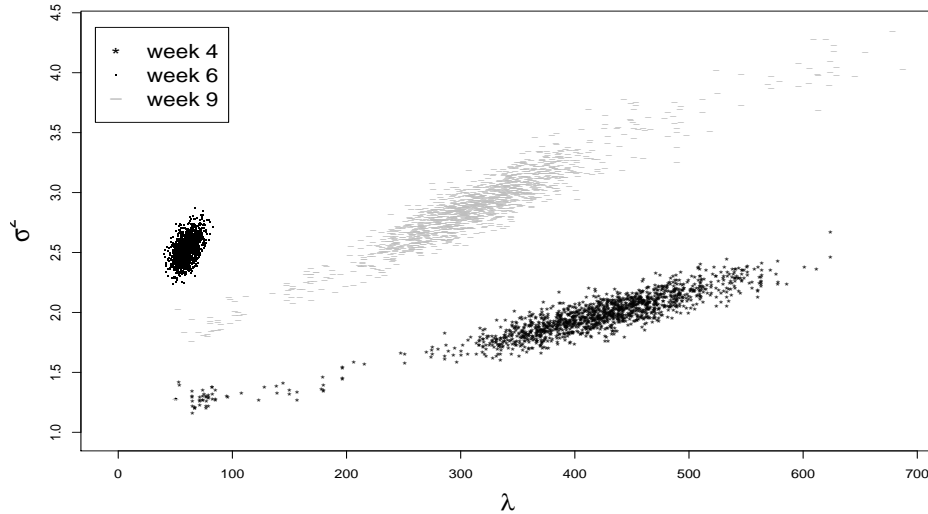
Our prior assumptions make the model parameters  $\lambda$ ,  $\sigma^2$ ,  $a_1$  and  $a_2$  uncorrelated. Figure 8 shows the relationship between the pairs of these parameters as a way of investigating that assumption. It seems valid except for the  $\lambda$ - $\sigma^2$  pair in graph (a). That graph shows a weak linear association between  $\lambda$  and  $\sigma^2$ , thus pointing to a failure of that assumption for that pair. Since  $\sigma^2$  determines spatial variability while  $\lambda$  determines correlation this relationship seems intriguing. Larger values of  $\sigma^2$  tend to go with larger  $\lambda$ s, i.e., diminished spatial correlation. Why they are



**Fig. 7.** Histogram (left panel), ACF (middle panel) and PACF (right panel) of model parameters of the Markov chains after a burn-in period of 1,000. The parameters are: (i) first row:  $-\lambda$ , the range parameter; (ii) second row:  $-\sigma^2$ , the variance parameter; (iii) third row:  $-a_1$ , the phase parameter with respect to the 24-hour periodicity; and (iv) last row:  $-a_2$ , the phase parameter with respect to the 12-hour periodicity.



**Fig. 8.** Scatterplots for the pairs of model parameters: (a)  $\lambda$  v.s.  $\sigma^2$ ; (b)  $\lambda$  v.s.  $a_1$ ; (c)  $\lambda$  v.s.  $a_2$ ; (d)  $\sigma^2$  v.s.  $a_1$ ; (e)  $\sigma^2$  v.s.  $a_2$ ; and (f)  $a_1$  v.s.  $a_2$ .



**Fig. 9.** Scatterplots for  $\lambda$  against  $\sigma^2$  for various weeks, based on the MCMC samples using one week’s data, that is, weeks 4, 6 and 9, but starting from the same initial values as those in Section 6.1.

coupled in this way is unknown but it should be accounted for in future applications of this model.

*Time varying  $\lambda$ s and  $\sigma^2$ s: empirical coverage probabilities versus nominal credible probabilities*

Although we follow Huerta et al. (2004) in assuming the temporal constancy of  $\lambda$  and  $\sigma^2$ , it is natural to ask if those generated by the MCMC method change over time. A variant of this issue concerns the time domain of the application. Would the results for these parameters change if we switched from one time span to a longer one containing it? A “yes” to this question would pose a challenge to anyone intending to apply the model, knowing that the choice would have implications for the size of  $\sigma^2$  and  $\lambda$ .

To address these concerns we carried out the following studies:

- (i) Study  $\tilde{A}$  : Implement the DLM at ungauged sites using weekly data ( $W_k : k = 1, \dots, 17$ ). Generate Markov chains for  $\lambda$ ,  $\sigma^2$ ,  $a_1$  and  $a_2$ . Obtain the coverage probabilities at each ungauged site and week for fixed credibility interval probabilities.
- (ii) Study  $\tilde{B}$  : Implement the DLM at ungauged sites using week 1 to week 17 data ( $W_{1:17} = \{W_1, \dots, W_{17}\}$ ). Estimate model parameters and interpolate the results at those ungauged sites. Obtain the coverage probabilities at each ungauged site and week for fixed credibility interval probabilities using each week’s data.
- (iii) Study  $\tilde{C}$  : Fix  $\lambda_k^*$  at week  $k$  ( $k = 1, \dots, 17$ ) using values suggested by the Markov chains generated in Study  $\tilde{A}$ . Then use these  $\lambda^* = \{\lambda_1^*, \dots, \lambda_{17}^*\}$  as fixed values in the DLM to reduce computation time. In other words, go through all the steps in the algorithm of Section 4.2 but now using only fixed  $\lambda^*$ s instead of generating them by a Metropolis–Hasting step. (Note that we are then only using Gibbs sampling and an MCMC blocking scheme.) Compute the corresponding coverage probabilities using  $W_{1:17}$  at each ungauged site and week for fixed credibility interval probabilities.

Studies  $\tilde{A}$  and  $\tilde{B}$  are intended to explore the effect of data and time propagation on the interpolation results. Study  $\tilde{C}$  aims to pick out any significant difference in the interpolation results

**Table 4.** Fixed values of  $\lambda^*$  in Study  $\tilde{C}$ .

<b>Week</b>	1	2	3	4	5	6	7	8	9
$\lambda^*$	54.2	178.5	83.7	405.4	86.6	59.7	199.3	144.1	322.7
<b>Week</b>	10	11	12	13	14	15	16	17	
$\lambda^*$	142.2	172.7	187.9	315.8	419.0	99.8	260.3	284.8	

**Table 5.** Summary for the computational time in Studies  $\tilde{A}$ ,  $\tilde{B}$  and  $\tilde{C}$ . Time is measured in seconds. The total is for a complete summer long MCMC run without spatial prediction.

<b>Study</b>	<b>Data</b>	<b>Iteration total</b>	<b>Accept(%)</b>	<b>Time (seconds)</b>	
				<b>Total</b>	<b>/Iteration</b>
$\tilde{A}$	$W_k$	1,500	0.82	17018	13.8
$\tilde{B}$	$W_{1:17}$	1,000	0.35	326782	932.3
$\tilde{C}$	$W_{1:17}$	1,000	1.00	329349	329.3

when using the fixed  $\lambda^*$  rather than using the Markov samples of  $\lambda$ s. It is also aimed at finding how much time would be saved by avoiding the inefficient Metropolis step. Table 4 shows these fixed  $\lambda^*$ s used in Study  $\tilde{C}$ . Table 5 shows the time saved using fixed  $\lambda^*$ s against the one using the Metropolis–Hastings algorithm.

Figure 9 illustrates the MCMC estimation results obtained in Study  $\tilde{A}$ . It plots the Markov chains of  $\lambda$  and  $\sigma^2$  using weekly data. It is obvious that  $\lambda$  and  $\sigma^2$  vary from week to week, which implies that the constant  $\lambda$ – $\sigma^2$  model is not tenable over a whole summer for this dataset.

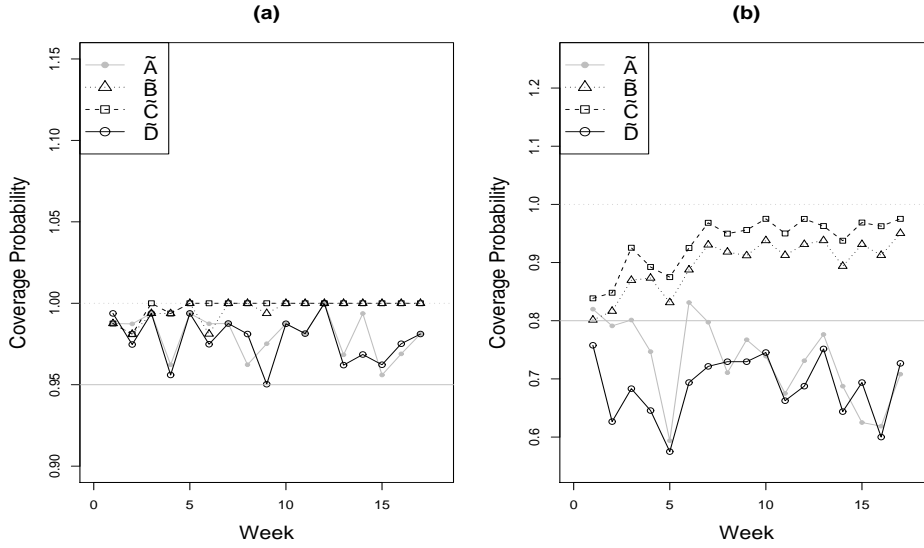
Figure 10 typifies figures in Dou et al. (2007) showing the coverage probabilities for various predictive intervals associated with the interpolators in these three studies. The solid line with bullets represents the results for Study  $\tilde{A}$ , the dotted line with up-triangles for Study  $\tilde{B}$ , and the dashed line with squares for Study  $\tilde{C}$ . These graphs show that the coverage probabilities of Study  $\tilde{B}$  are similar to that of Study  $\tilde{C}$ . This suggests that we could use the entries in Table 4 as fixed  $\lambda^*$ s in the DLM to obtain interpolation results similar to those obtained using the Metropolis–within–Gibbs algorithm.

We have studied the prediction accuracy of the simplest DLM, namely, the first–order polynomial model, in Section 3. As a result, the predictive variances should increase monotonically at successive time points conditional on all the 17 weeks’ data, in the general DLM setting (see Section 3). The plots exhibit a monotonic increasing trend in the coverage probabilities of both Studies  $\tilde{B}$  and  $\tilde{C}$ . This trend agrees with the graph of the coverage probabilities in Figure 10. Nevertheless, those coverage probabilities of both studies deviate slightly from the expected monotonically increasing trend at some time points because of the time varying effect of  $\lambda$ – $\sigma^2$  monitored in Figure 9.

On the other hand, Study  $\tilde{C}$  enjoys significant computational time savings compared with  $\tilde{B}$ . Table 5 suggests that the computation time of the former is almost 2.8 times faster than the latter.

Study  $\tilde{B}$  shows an intuitively unappealing increase in the uncertainty of interpolation results as time increases; coverage probabilities get larger over time as we see in Table 6. This increase may be interpreted as saying that for the DLM models, the  $\lambda$ s and  $\sigma^2$ s collected from the data should vary over the entire time span of the study, while the prior postulates that they do not vary over that time span. The observed phenomenon may also be due to mis–specification of the model parameter values  $\gamma = (\tau_y^2, \tau_1^2, \dots, \lambda_2)$  (See the initial settings for  $\gamma$  in Section 6.1.).

Comparing the results of these studies, we find that sometimes, paradoxically, the model gives better results using only one week’s data rather than all. However, Corollary 2 in Section 3 predicts this finding. Because the prior for  $\sigma_1^2$  is  $IG(2, 0.01)$  the expectation of  $\sigma_1^2$  is 0.01, implying that  $\sigma_\beta^2 \simeq 0.01$  and  $\sigma_\delta^2 \simeq 0.01 \times 0.02$ . Hence,  $\sigma_\beta^2 \left(1 + \frac{\sigma_\beta^2}{\sigma_\delta^2}\right) \simeq 0.51$ , which is less than  $\sigma_\varepsilon^2$  (for example, the median of  $\sigma^2$  is around 1.21 in Study  $\tilde{B}$  and even larger in Study  $\tilde{A}$ ). By the sufficient and necessary condition in Corollary 2, the predictive variance of Study  $\tilde{A}$  is less than that of Study  $\tilde{B}$ .



**Fig. 10.** Coverage probability versus: (a) 95% nominal level for Ungauged Site D, and (b) 80% nominal level for Ungauged Site C. These coverage probabilities are computed for *Study A-tilde*: weekly data (solid bullet with solid line); *Study B-tilde*:  $W_{1:17}$  (up-triangle with dotted line); *Study C-tilde*:  $W_{1:17}$  but with fixed  $\lambda^*$  (square with dashed line); and *Study D-tilde*:  $W_{1:17}$  but with fixed  $\lambda^*$  and modified  $\tau_y^2$ ,  $\tau_1^2$ , and  $\tau_2^2$  (empty circle with solid line).

**Table 6.** Coverage probabilities (%) for studies  $\tilde{A}$ ,  $\tilde{B}$  and  $\tilde{C}$  at Ungauged Sites A, B, and C, at 80% nominal level.

Ungauged Site	A			B			C		
	$\tilde{A}$	$\tilde{B}$	$\tilde{C}$	$\tilde{A}$	$\tilde{B}$	$\tilde{C}$	$\tilde{A}$	$\tilde{B}$	$\tilde{C}$
Week 1	66	65	72	80	78	89	82	80	84
Week 2	73	71	80	76	78	83	79	81	85
Week 3	63	73	82	82	86	91	80	87	93
Week 4	57	74	81	66	83	88	75	87	89
Week 5	53	70	82	68	83	90	59	83	88
Week 6	73	80	88	75	83	89	83	89	93
Week 7	69	88	90	80	92	94	80	93	97
Week 8	66	89	93	66	90	93	71	92	95
Week 9	63	82	88	84	90	94	77	91	96
Week 10	61	87	92	75	93	96	74	94	98
Week 11	58	86	89	77	93	94	68	91	95
Week 12	69	90	92	69	97	96	73	93	98
Week 13	60	87	90	74	91	94	77	94	96
Week 14	67	87	89	81	92	95	69	89	94
Week 15	66	91	95	65	93	96	63	93	97
Week 16	65	91	93	79	94	97	62	91	96
Week 17	68	90	95	81	93	98	71	95	98

However, notice that  $\sigma^2$  and  $\lambda$  vary from week to week in  $\tilde{A}$ , which may also lead to the paradox observed in the empirical findings of this section. For example, in (b) of Figure 10, the coverage probability of  $\tilde{B}$  at the 4<sup>th</sup> week is larger than that of  $\tilde{A}$ . From the above discussion, we know that the predictive variance of  $\tilde{A}$  should be less than that of  $\tilde{B}$ . However,  $\sigma^2$  of  $\tilde{A}$  is larger than that of  $\tilde{B}$ , leading an inflated predictive variance of  $\tilde{A}$ . This feature makes it difficult to compare these two predictive variances, but explains the paradox we see in those figures.

## 8. Summary and Conclusions

To assess the dynamic linear modelling approach to modelling space–time fields, we have applied it to an hourly ozone concentration field over a geographical spatial domain covering most of the eastern United States. To focus that assessment we consider just one cluster of spatial sites we call Cluster 2 during a single ozone season. Moreover, we have used a variant of the dynamic linear modelling approach of Huerta et al. (2004) implemented through MCMC sampling.

Our assessment reveals some difficulties with that very flexible approach and practical challenges that it presents. We also have made some recommendations on improvement.

A curious finding is the posterior dependence of  $\lambda$  and  $\sigma^2$ , in contradiction to our prior assumption. Although the very efficient method Huerta et al. (2004) propose to sampling these parameters is biased, that bias does not appear large enough to account for that phenomenon. We also discovered that the assumption of their constancy over time is untenable.

The coverage probabilities of the model’s posterior predictive credibility intervals over successive weeks, conditional on all 17 weeks of data, increase monotonically. Counter to intuition, that would imply more and more uncertainty as time evolves, an artifact of the modelling that seems hard to explain. A pragmatic way around this undesirable property involves incorporating the length of the time span of the temporal domain  $T$  into the selection of the values of the model parameters, such as  $\tau_y^2$ ,  $\tau_1^2$  and  $\tau_2^2$ . Section 3 studies the correlation structure of the simplest first–order polynomial DLM and finds reasonable conditions to impose on those parameters.

One further Study  $\tilde{D}$  tests the proposed constraints on the data. The settings are identical with those in Study  $\tilde{C}$  except that  $\tau_y^2$ ,  $\tau_1^2$  and  $\tau_2^2$  are replaced by  $\tau_y^2/17$ ,  $\tau_1^2/17$  and  $\tau_2^2/17$ , respectively, to take account of the longer 17 week time span of our study compared to the one week time span of the application in Huerta et al. (2004). Figure 10 compares Study  $\tilde{D}$  with the others. Observe that its coverage probabilities behave like those of Study  $\tilde{A}$ . This adjustment does seem to eliminate the undesirable property of increasing credibility bands of Studies  $\tilde{B}$  and  $\tilde{C}$ .

Another possible approach to dealing with the unsuitability of fixed model parameters uses the composition of Metropolis–Hasting kernels. In other words, we could include these parameters in the Metropolis–Hasting algorithm as in Section 4.1. We can use six Metropolis–Hasting kernels to sample from the target distribution  $\pi(\gamma|y_{1:T})$ , updating each component of  $\gamma$  iteratively, where  $\gamma$  has defined in Section 2. But, not surprisingly that approach fails because of the extreme computational burden it entails. However, that alternative is the subject of current work along with an approach that admits time varying  $\lambda$ s and  $\sigma^2$ s.

The greatest difficulty involved in the use of the DLM in modelling air pollution space–time fields lies in the computational burden it entails. For that reason, we have not been able to address the geographical domain of real interest, one that embraces 274 sites in the eastern United States, with 120 days of hourly ozone concentrations. In a manuscript under preparation, an alternative hierarchical Bayesian method that can cope with that larger domain will be compared with the DLM where the latter can practically be applied.

## Acknowledgements

We thank Prasad Kasibhatla of Nicholas School of the Environment of Duke University for providing the dataset used in this paper and helping with its installation. The funding for the work was provided by the Natural Sciences and Engineering Research Council of Canada.

## A. Supplementary results

### A.1. Results for Section 3

Only the results about the predictive variances of  $y_{01}|y_{11}$  and  $y_{01}|y_{11}, y_{12}$  are shown in this appendix. The other two cases can be obtained by the same method. Refer to Theorem 1, the predictive variance of  $y_{01}|y_{11}$  can also be written as follows:

$$\text{Var}(y_{01}|y_{11}) = (1 - \exp(-\frac{d_{01}}{\lambda}))\sigma_\varepsilon^2 \left\{ 2 - \frac{1 - \exp(-\frac{d_{01}}{\lambda})}{1 + \frac{\sigma_\beta^2 + \sigma_\delta^2}{\sigma_\varepsilon^2}} \right\}.$$

The first partial derivatives of this predictive variances regarding to  $d_{01}$ ,  $\lambda$  and  $\sigma_\varepsilon^2$  are given by:

$$\frac{\partial}{\partial d_{01}} \text{Var}(y_{01}|y_{11}) = \frac{2d_{01}}{\lambda} \exp(-\frac{d_{01}}{\lambda}) \sigma_\varepsilon^2 \frac{\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-\frac{d_{01}}{\lambda})}{\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2},$$

$$\frac{\partial}{\partial \lambda} \text{Var}(y_{01}|y_{11}) = -\frac{2d_{01}}{\lambda^2} \exp(-\frac{d_{01}}{\lambda}) \sigma_\varepsilon^2 \frac{\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-\frac{d_{01}}{\lambda})}{\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2},$$

and

$$\begin{aligned} \frac{\partial}{\partial \sigma_\varepsilon^2} \text{Var}(y_{01}|y_{11}) &= (1 - \exp(-\frac{d_{01}}{\lambda})) \left\{ 2 - (1 - \exp(-\frac{d_{01}}{\lambda})) \sigma_\varepsilon^2 \frac{\sigma_\varepsilon^2 + 2\sigma_\beta^2 + 2\sigma_\delta^2}{(\sigma_\varepsilon^2 + \sigma_\beta^2 + \sigma_\delta^2)^2} \right\} \\ &> (1 - \exp(-\frac{d_{01}}{\lambda})) \left\{ 2 - \frac{\sigma_\varepsilon^2(2\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)}{(\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)^2} \right\} \\ &= \frac{1 - \exp(-\frac{d_{01}}{\lambda})}{(\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)^2} \{2(\sigma_\beta^2 + \sigma_\delta^2)^2 + \sigma_\varepsilon^4 + 2\sigma_\varepsilon^2(\sigma_\beta^2 + \sigma_\delta^2)\}, \end{aligned}$$

respectively. It is straightforward to obtain that  $\text{Var}(y_{01}|y_{11})$  is increasing when  $d_{01}$  increases, or  $\lambda$  decreases, or  $\sigma_\varepsilon^2$  increases. We next show these properties also hold for  $\text{Var}(y_{01}|y_{11}, y_{12})$ . By Theorem 1,  $\text{Var}(y_{01}|y_{11}, y_{12})$  can also be written as:

$$\text{Var}(y_{01}|y_{11}, y_{12}) = (1 - \exp(-\frac{d_{01}}{\lambda}))\sigma_\varepsilon^2 \left\{ 2 - \frac{1 - \exp(-\frac{d_{01}}{\lambda})}{1 + \frac{(\sigma_\beta^2 + \sigma_\delta^2)(\sigma_\delta^2 + \sigma_\varepsilon^2)}{\sigma_\varepsilon^2(\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)}} \right\}.$$

The corresponding first partial derivatives are given as follows:

$$\frac{\partial}{\partial d_{01}} \text{Var}(y_{01}|y_{11}, y_{12}) = \frac{2}{\lambda} \exp(-\frac{d_{01}}{\lambda}) \sigma_\varepsilon^2 \frac{A + \exp(-\frac{d_{01}}{\lambda})}{1 + A},$$

$$\frac{\partial}{\partial \lambda} \text{Var}(y_{01}|y_{11}, y_{12}) = -\frac{2d_{01}}{\lambda^2} \exp(-\frac{d_{01}}{\lambda}) \sigma_\varepsilon^2 \frac{A + \exp(-\frac{d_{01}}{\lambda})}{1 + A},$$

and

$$\begin{aligned} \frac{\partial}{\partial \sigma_\varepsilon^2} \text{Var}(y_{01}|y_{11}, y_{12}) &= (1 - \exp(-\frac{d_{01}}{\lambda})) \left\{ 2 - (1 - \exp(-\frac{d_{01}}{\lambda})) \frac{\sigma_\varepsilon^2}{A^2} (c_1 A - c_2 c_3) \right\} \\ &> \frac{1 - \exp(-\frac{d_{01}}{\lambda})}{A^2} c_4, \end{aligned}$$

respectively, where  $A = \frac{(\sigma_\beta^2 + \sigma_\delta^2)(\sigma_\delta^2 + \sigma_\varepsilon^2)}{\sigma_\varepsilon^2(\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)}$ ,  $c_1 = \sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2$ ,  $c_2 = \sigma_\beta^2 + \sigma_\delta^2$ ,  $c_3 = \sigma_\delta^2 c_1 + \sigma_\varepsilon^2(\sigma_\delta^2 + \sigma_\varepsilon^2)$ , and  $c_4 = \sigma_\varepsilon^2 c_1(2\sigma_\beta^2 + 3\sigma_\delta^2 + \sigma_\varepsilon^2) + \sigma_\varepsilon^2 c_2(\sigma_\delta^2 + \sigma_\varepsilon^2)(3\sigma_\beta^2 + 6\sigma_\delta^2 + 4\sigma_\varepsilon^2) + c_2^2(\sigma_\delta^2 + \sigma_\varepsilon^2)^2$ .

### A.2. Results for Section 4.1

The joint posterior distribution for  $x_{1:T}$ ,  $\lambda$  and  $\sigma^2$  is given by

$$\begin{aligned} p(x_{1:T}, \lambda, \sigma^2 | y_{1:T}) &= p(\lambda, \sigma^2) p(\mathbf{x}_T | \lambda, \sigma^2, y_{1:T}) \prod_{t=1}^T p(\mathbf{x}_{T-t} | \mathbf{x}_{T-t+1}, \lambda, \sigma^2, y_{1:T}) \\ &\quad \prod_{t=1}^T p(\mathbf{y}_t | \lambda, \sigma^2, y_{1:t-1}) \\ &= p(x_{1:T} | \lambda, \sigma^2, y_{1:T}) p(\sigma^2 | \lambda, y_{1:T}) p(\lambda | \mathbf{y}_T). \end{aligned}$$

Suppose  $p(\lambda, \sigma^2) = p(\lambda)p(\sigma^2)$ , that is, the priors for  $\lambda$  and  $\sigma^2$  are independent of other.

The joint posterior distribution for  $\lambda$  and  $\sigma^2$  can be written as follows:

$$p(\lambda, \sigma^2 | y_{1:T}) \propto p(\lambda) p(\sigma^2) (\sigma^2)^{-nT/2} \prod_{t=1}^T |Q_t|^{-1/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^T \mathbf{e}_t' Q_t^{-1} \mathbf{e}_t \right\}.$$

If the prior for  $\sigma^2$  is an inverse gamma distribution with shape parameter  $\alpha$  and scale parameter  $\beta$ , then the posterior distribution for  $\sigma^2$  is also an inverse gamma distribution with shape parameter  $\alpha + \frac{nT}{2}$  and scale parameter  $\beta + \frac{1}{2} \sum_{t=1}^T \mathbf{e}_t' Q_t^{-1} \mathbf{e}_t$ .

Hence, the posterior density for  $\lambda$  can be written as follows:

$$\begin{aligned} p(\lambda | y_{1:T}) &= \frac{p(\lambda, \sigma^2 | y_{1:T})}{p(\sigma^2 | \lambda, y_{1:T})} \\ &\propto p(\lambda) \prod_{t=1}^T |Q_t|^{-1/2} \left[ \beta + \frac{1}{2} \sum_{t=1}^T \mathbf{e}_t' Q_t^{-1} \mathbf{e}_t \right]^{-(\alpha + nT/2)}. \end{aligned}$$

Therefore, the posterior density for  $\mathbf{x}_{1:T}$  is given by

$$p(x_{1:T} | \lambda, \sigma^2, y_{1:T}) = p(\mathbf{x}_T | \lambda, \sigma^2, y_{1:T}) \prod_{t=1}^T p(\mathbf{x}_{T-t} | \mathbf{x}_{T-t+1}, \lambda, \sigma^2, y_{1:T}).$$

### A.3. Results for Section 5

Given the values of the phase parameters, range and variance parameters and the observations until time  $t$ , the joint distribution of  $\alpha_{1t}^s, \alpha_{1t}$  is

$$\begin{pmatrix} \alpha_t^s \\ \alpha_{1t} \end{pmatrix} \sim N \left[ \begin{pmatrix} \alpha_{1,t-1}^s \\ \alpha_{1,t-1} \end{pmatrix}, \sigma^2 \tau_1^2 \Sigma^*(\lambda_1) \right],$$

where

$$\Sigma^*(\theta) = \exp\{-V^*/\theta\} = \begin{bmatrix} \Sigma_{11}^*(\theta) & \Sigma_{12}^*(\theta) \\ \Sigma_{21}^*(\theta) & \Sigma_{22}^*(\theta) \end{bmatrix},$$

with  $\Sigma_{11}^*(\theta)$  a scalar,  $\Sigma_{12}^*(\theta)$  a 1 by  $n$  vector, and  $\Sigma_{22}^*(\theta)$  a  $n$  by  $n$  matrix. We use  $V^*$  to denote the new distance matrix for the unknown site  $s$  and the monitoring stations  $s_1, \dots, s_n$ .

We then have the conditional posterior distribution of  $\alpha_{1t}^s$  as follows:

$$(\alpha_{1t}^s | \alpha_{1,t-1}^s, \alpha_{1t}, \alpha_{1,t-1}, \mathbf{y}_t, \lambda, \sigma^2) \sim N[\alpha_{1,t-1}^s + \Sigma_{12}^*(\lambda_1) \Sigma_{22}^*(\lambda_1)^{-1} (\alpha_{1t} - \alpha_{1,t-1}), \sigma^2 \tau_1^2 (\Sigma_{11}^*(\lambda_1) - \Sigma_{12}^*(\lambda_1) \Sigma_{22}^*(\lambda_1)^{-1} \Sigma_{21}^*(\lambda_1))]. \quad (38)$$

Similarly, the conditional posterior distribution for  $\alpha_{2t}^s$  is

$$(\alpha_{2t}^s | \alpha_{2,t-1}^s, \alpha_{2t}, \alpha_{2,t-1}, \mathbf{y}_t, \lambda, \sigma^2) \sim N[\alpha_{2,t-1}^s + \Sigma_{12}^*(\lambda_2) \Sigma_{22}^*(\lambda_2)^{-1} (\alpha_{2t} - \alpha_{2,t-1}), \sigma^2 \tau_2^2 (\Sigma_{11}^*(\lambda_2) - \Sigma_{12}^*(\lambda_2) \Sigma_{22}^*(\lambda_2)^{-1} \Sigma_{21}^*(\lambda_2))]. \quad (39)$$



Using the observation equation as in Model (1), we have the conditional predictive distribution for  $y_t^s$  as follows:

$$\begin{aligned} (y_t^s | \mathbf{y}_t, \alpha_{1t}^s, \alpha_{2t}^s, \alpha_{1t}, \alpha_{2t}, \beta_t, \lambda, \sigma^2) \sim & N[\beta_t + S_{1t}(a_1)\alpha_{1t}^s + S_{2t}(a_2)\alpha_{2t}^s \\ & + \Sigma_{12}^*(\lambda)\Sigma_{22}^*(\lambda)^{-1}(\mathbf{y}_t - \mathbf{1}_n\beta_t \\ & - S_{1t}(a_1)\alpha_{1t} - S_{2t}(a_2)\alpha_{2t}), \\ & \sigma^2(\Sigma_{11}^*(\lambda) - \Sigma_{12}^*(\lambda)\Sigma_{22}^*(\lambda)^{-1}\Sigma_{21}^*(\lambda))]. \end{aligned} \quad (40)$$

The software is in <http://enviro.stat.ubc.ca>.

## References

- Burke, J.M. and Zufall, M.J. and Ozkaynak, H. (2001). A population exposure model for particulate matter: case study results for PM<sub>2.5</sub> in Philadelphia, PA, *J. Exposure Anal. Environ. Epidemiol.*, 11, 470–489.
- Calder, C.A. and Holloman, C.H. and Bortnick, S.M. and Strauss, W.J. and Morara, M. (2003). Relating ambient particulate matter concentration levels to mortality using an exposure simulator, Preprint #725, Dept. Statist., Ohio State U.
- Carter, C.K. and Kohn, R. (1994). On Gibbs sampling for state space models, *Biometrika*, 81, 541–553.
- Dou, Y., Le, N.D. and Zidek, J.V. (2007). A Dynamic Linear Model for Hourly Ozone Concentrations, TR # 228, Dept. Statist., U of British Columbia.
- Gamerman, D. and Lopes H.F. (2006). *Markov Chain Monte Carlo: stochastic simulation for Bayesian inference, 2nd Edition*, London: Chapman & Hall.
- Huerta, G. and Sanso, B. and Stroud, J.R. (2004). A spatio-temporal model for Mexico city ozone levels, *JRRSC*, 53, 231–248.
- Moller, J. (2003). *Spatial Statistics and Computational Methods*, New York: Springer-Verlag.
- Ozone (2005). US EPA's Air Quality Criteria for Ozone and Related Photochemical Oxidants (First External Review Draft), <http://cfpub.epa.gov~ncea/cfm/recordisplay.cfm?deid=149923>
- Stroud, J.R. and Muller, P. and Sanso, B. (2001). Dynamic models for spatio-temporal data, *JRSSB*, 63, 673–689.
- West, M. and Harrison, J. (1997). *Bayesian forecasting and dynamic models, 2nd Edition*, New York: Springer.