

THE UNIVERSITY OF BRITISH COLUMBIA
DEPARTMENT OF STATISTICS
TECHNICAL REPORT # 228

A DYNAMIC LINEAR MODEL FOR
HOURLY OZONE CONCENTRATIONS

By

YIPING DOU
NHU D LE
JAMES V ZIDEK

March 2007

A Dynamic Linear Model for Hourly Ozone Concentrations *

Yiping Dou, Nhu D Le and James V Zidek

March 12, 2007

Abstract

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.

*The research reported in this paper was partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

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 as 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 (Caldor 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 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 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, the 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 among 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 two other possibilities.

For completeness, Section 2 also introduces some basic features of space-time theory along with the a review of the celebrated Kalman filter out of which the state space model of statistics evolved.

Section 3 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 AIRS¹ dataset leads to the construction of our DLM. 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 4 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 5 and 6. The MCMC sampling scheme is outlined in Section 5.1. The forward-filtering-backward-sampling (FFBS) method is demonstrated in Section 5.2 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 6 gives theoretical results for prediction and interpolation at unmonitored (ungauged) sites from their predictive posterior distributions. Section 7 shows the results of MCMC sampling along with interpolation results on the ozone study. Section 8 describes problems with the DLM process revealed by our assessment. We summarize our findings and draw conclusions from our assessment in Section 9.

As an added note, we have developed software, written by C and R (see Appendix B), that alleviates some of the immense computational burden associated with the MCMC method. That software may be downloaded from <http://enviro.stat.ubc.ca> for others wishing to reproduce our findings or apply the model for modelling hourly ozone in other settings.

¹Notice that *EPA* (Environmental Protection Agency) changed the *AIRS* (Aerometric Information Retrieved System) to the *AFS* (Air Facility Subsystem) in 2001.

2 Background

2.1 Space–time process

Let \mathcal{D} be the spatial domain of a space–time field, that can in practice be taken to be a finite set of spatial locations which is monitored at sites, \mathbf{s}_i , $i = 1, \dots, n$, some of which are monitored. Let $Y(\mathbf{s}_i, t)$ denote the random response at time $t \in \mathcal{R}$ and site $\mathbf{s}_i \in \mathcal{D}$ and $Z(\mathbf{s}_i, t)$, its measured value.

A very general model is given by

$$Z(\mathbf{s}_i, t) = Y(\mathbf{s}_i, t) + \epsilon(\mathbf{s}_i, t), \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (1)$$

where $\epsilon(\mathbf{s}_i, t)$ is a white noise process.

The space–time process $Y(\mathbf{s}_i, t)$ can be expressed as

$$Y(\mathbf{s}_i, t) = \mu(\mathbf{s}_i, t) + w(\mathbf{s}_i, t), \quad (2)$$

where $\mu(\mathbf{s}_i, t)$ is the mean process that may depend on the covariate $x(\mathbf{s}_i, t)$ while $w(\mathbf{s}_i, t)$ is a spatio–temporal process with mean $\mathbf{0}$.

Since we will be assuming below that our fields have conditional Gaussian distributions, the covariance defined next plays a fundamental role.

Definition 1 *The space–time covariance function is defined as*

$$C(\mathbf{s}_1, \mathbf{s}_2; t_1, t_2) = \text{Cov}[w(\mathbf{s}_1, t_1), w(\mathbf{s}_2, t_2)], \quad (3)$$

where \mathbf{s}_i is the location, and t_i is the time point, for $i = 1, 2$.

To simplify inference about this fields, the covariance is assumed to have the simple form defined next.

Definition 2 *The zero mean spatio–temporal process $w(\mathbf{s}, t)$ is covariance stationary if*

$$C(\mathbf{s}_1, \mathbf{s}_2; t_1, t_2) = C(\mathbf{s}_1 - \mathbf{s}_2; t_1 - t_2) = C(\mathbf{h}; \eta), \quad (4)$$

where $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$ and $\eta = t_1 - t_2$.

Note that \mathbf{h} in (4) denotes the vector distance between the two sites. If $w(\mathbf{s}, t)$ does not satisfy (4), it is called a non–stationary spatio–temporal process.

Sometimes an even more specialized form of the covariance is assumed.

Definition 3 *The zero mean stationary spatio–temporal process $w(\mathbf{s}, t)$ is isotropic if*

$$C(\mathbf{h}; \eta) = C(\|\mathbf{h}\|; |\eta|). \quad (5)$$

Covariances with this property depend on the difference (or separation) vectors only through the length of their difference vectors, $\|\mathbf{h}\|$ and $|\eta|$. If the spatio–temporal process is not isotropic, then it is called anisotropic.

The covariance structure for the spatio–temporal process can be greatly simplified by assuming separability defined as follows.

Definition 4 *The zero mean stationary and isotropic spatio–temporal process $w(\mathbf{s}, t)$ is separable if*

$$C(\|\mathbf{h}\|; |\eta|) = C_s(\|\mathbf{h}\|)C_t(|\eta|), \quad (6)$$

that is, the covariance function for the spatio–temporal process can be decomposed as the product of an isotropic spatial and an isotropic temporal covariance function.

If the covariance for spatio–temporal process is not separable, it is called non–separable and the process is then non–separable. A fairly general approach to modelling space–time fields entails the construction of a valid separable covariance function (Gelfand et al., 2004). Brown, Le and Zidek (1994) and Le, Sun and Zidek (1997) use a separable covariance structure to deal with non-stationary multivariate spatial models in a hierarchical Bayesian framework. This approach has two advantages: (i) it reduces the number of parameters to be estimated in the model; (ii) it guarantees a positive definite covariance matrix.

Because of the dynamic features of spatio–temporal processes, we specialize the model above and adopt with a dynamic modelling approach. In the following sections, we follow that approach, introduce that general state space model and present some basic results.

2.2 The general state–space model

Suppose $\mathbf{y}_t : n \times 1$ is a vector of observable random responses for $t = 1, 2, \dots$. The state space model that is sometimes referred to as a dynamic linear model (DLM) involves both an “observation equation” and an “evolution equation”. The former is given by

$$\mathbf{y}_t = F_t' \mathbf{x}_t + \nu_t, \quad \nu_t \sim N[\mathbf{0}, V_t], \quad (7)$$

the latter by

$$\mathbf{x}_t = G_t \mathbf{x}_{t-1} + \omega_t, \quad \omega_t \sim N[\mathbf{0}, W_t], \quad (8)$$

$F_t : p \times n$, $G_t : p \times p$, $V_t : n \times n$, and $W_t : p \times p$ being known matrices. In Equation (7), F_t is called the design matrix, \mathbf{x}_t is the state (or system) vector and ν_t is the observational error. Equation (8) is also called the state or system equation. G_t is the system or state matrix and ω_t , the system or evolution error with evolution matrix W_t .

The DLM’s specification is completed with the initial information for the state parameter given by

$$(\mathbf{x}_0 | \mathbf{y}_0) \sim N[\mathbf{m}_0, C_0]. \quad (9)$$

2.3 Kalman filter and smoother

Kalman recursion, or Kalman filtering as it is sometimes referred to, is used to update and forecast the state parameters. One version of that method is derived by West and Harrison (1997, Chapter 4) for updating and forecasting state parameters in the DLM framework. This result can also be derived by means of Bayesian inference. Carter and Kohn (1994) label this approach forward-filtering-backward-sampling. It is particularly powerful and will be used in the following sections. It is summarized in the following theorem which will be applied to the filtering and smoothing processes.

Theorem 1 *Let $x_{1:t} = (\mathbf{x}_1, \dots, \mathbf{x}_t)$ and $y_{1:t} = (\mathbf{y}_1, \dots, \mathbf{y}_t)$, $t \geq 1$, denote all the state parameters and the responses observed until time t , respectively. Under models (7)–(8), together with the initial information (9), we have*

(i)

$$\begin{aligned} (\mathbf{x}_{t-1} | y_{1:t-1}, \theta) &\sim N[\mathbf{m}_{t-1}, C_{t-1}] \\ (\mathbf{x}_t | y_{1:t-1}, \theta) &\sim N[\mathbf{a}_t, R_t] \\ (\mathbf{y}_t | y_{1:t-1}, \theta) &\sim N[\mathbf{f}_t, Q_t] \\ (\mathbf{x}_t | y_{1:t}, \theta) &\sim N[\mathbf{m}_t, C_t], \end{aligned}$$

where

$$\begin{aligned} \mathbf{a}_t &= G_t \mathbf{m}_{t-1} & R_t &= G_t C_{t-1} G_t' + W_t \\ \mathbf{f}_t &= F_t' \mathbf{a}_t & Q_t &= F_t' R_t F_t + V_t \\ \mathbf{e}_t &= \mathbf{y}_t - \mathbf{f}_t & A_t &= R_t F_t Q_t^{-1} \\ \mathbf{m}_t &= \mathbf{a}_t + A_t \mathbf{e}_t & C_t &= R_t - A_t Q_t A_t'. \end{aligned}$$

(ii) Let $B_t = C_t G_{t+1}' R_{t+1}^{-1}$. For $0 \leq k \leq T-1$,

$$(\mathbf{x}_{T-k} | y_{1:T}, \theta) \sim N[\mathbf{a}_T(-k), R_T(-k)], \quad (10)$$

where

$$\begin{aligned} \mathbf{a}_T(-k) &= \mathbf{m}_{T-k} + B_{T-k} [\mathbf{a}_T(-k+1) - \mathbf{a}_{T-k+1}] \\ R_T(-k) &= C_{T-k} + B_{T-k} [R_T(-k+1) - R_{T-k+1}] B_{T-k}' \end{aligned}$$

with $\mathbf{a}_T(0) = \mathbf{m}_T$, $R_T(0) = C_T$, $\mathbf{a}_{T-k}(1) = \mathbf{a}_{T-k+1}$, and $R_{T-k}(1) = R_{T-k+1}$.

This approach has the advantage that the smoothing process does not require any more information about the observations $y_{1:T}$. Instead, all the information about $y_{1:T}$ has been assimilated into $x_{1:T}$ through the forward filtering process. That fact allows us to discard the past observations. Note that the distributions for the smoothing and filtering processes are obtained conditionally on the parameter vector θ . In practice, it is often exceedingly

difficult to obtain the posterior distribution of θ because it may not have a closed form. In this case, the Markov Chain Monte Carlo (MCMC) method is often used to sample θ from its posterior conditional distribution. After obtaining the θ s, we can use Theorem 1 to get samples of state parameters from their corresponding distributions. If we can obtain all the samples of θ s and state parameters, we can then work out methods for prediction and interpolation, the general goal of the well known kriging method in a purely spatial setting. Instead of the space–time setting we confront, the DLM will be developed for interpolation and prediction. That application will be shown in Section 7. But first we need to describe more precisely in the next section, the specific context in which that model is needed.

3 Model specification

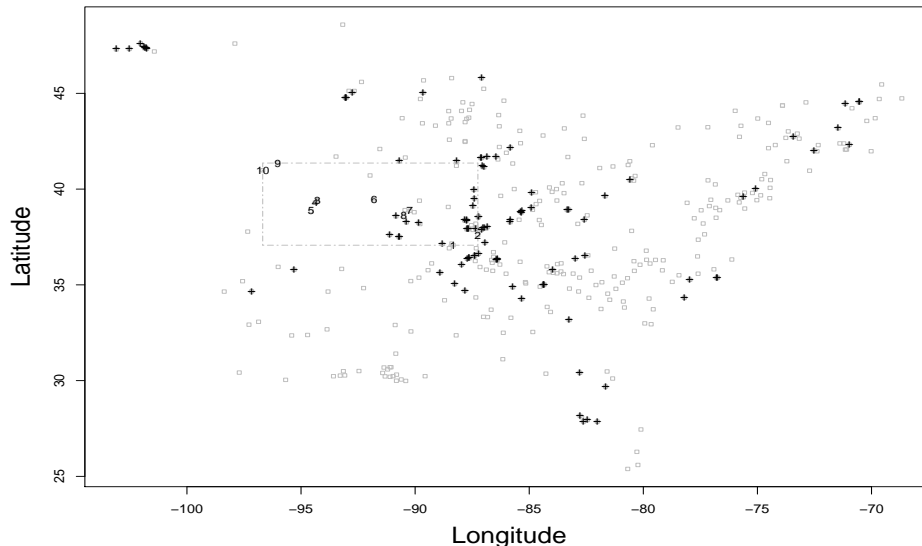


Figure 1: Geographic locations for AIRS dataset, where the latitude and longitude are measured by degrees. (Those of our cluster appear in the square.)

The space–time field addressed in this paper represents hourly ozone concentrations in the unit of ppb (parts per billion) over part of the eastern United States during the summer of 1995. That field is monitored by 375 irregularly located monitoring sites (or stations). To enable a focused assessment of DLM approach as well as overcome computational barriers, we consider just one cluster of ten sites (Cluster 2), in close proximity to one another. However, in unreported work (not reported here for brevity), two other such clusters have also been investigated with similar results. The geographical locations specified by latitude and longitude of the monitoring sites, including those of our

cluster, are depicted in Figure 1. Incidentally, Cluster 2 has the same number of sites that for Mexico City studied by Huerta et al. (2004).

To develop a suitable model for this space–time field, and construct a DLM (which of necessity must have normally distributed residuals), we conducted an exploratory data analysis (EDA) like the one used by Huerta et al. (2004) to construct their model. For convenience, a small amount of randomly missing data were filled in by the spatial regression method (SRM), before we began.

The analysis revealed an asymmetric empirical distribution for the hourly ozone concentrations, suggesting a data transformation to validate the assumed normality of DLM residuals. The square–root transformation has been proved best among various options considered, in line with Huerta et al. (2004).

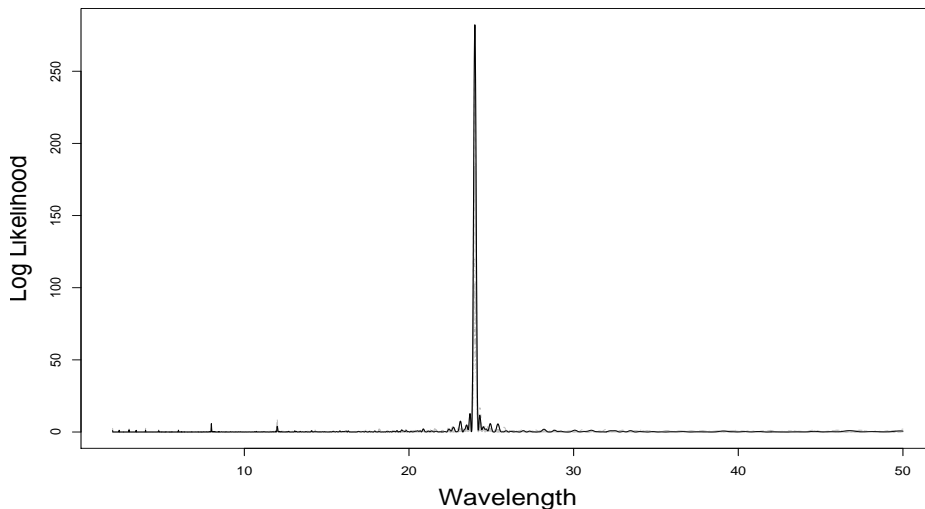


Figure 2: Bayesian periodograms for square–root transformed hourly ozone concentrations at the Cluster 2 sites under the study from May 15 to September 11, 1995 (data taken from the AIRS database).

To look for periodicity in the ozone data, we plotted the Bayesian periodograms for the square–root transformed ozone levels measured during the summer of 1995 at the sites in our cluster (Figure 2). A high peak was seen between 1 pm and 3 pm each day for 120 days. In fact, a significant 24–hour cycle was obtained for all these gauged (that is, monitoring) sites. We also found a slightly significant 12–hour cycle by plotting the periodicities that contribute the most variation judging from the spectrum for each site in our cluster (Figure 3). However, no obvious weekly cycles or nightly peaks were seen.

The DLM thus suggested by our analysis is a variant of that of Huerta et

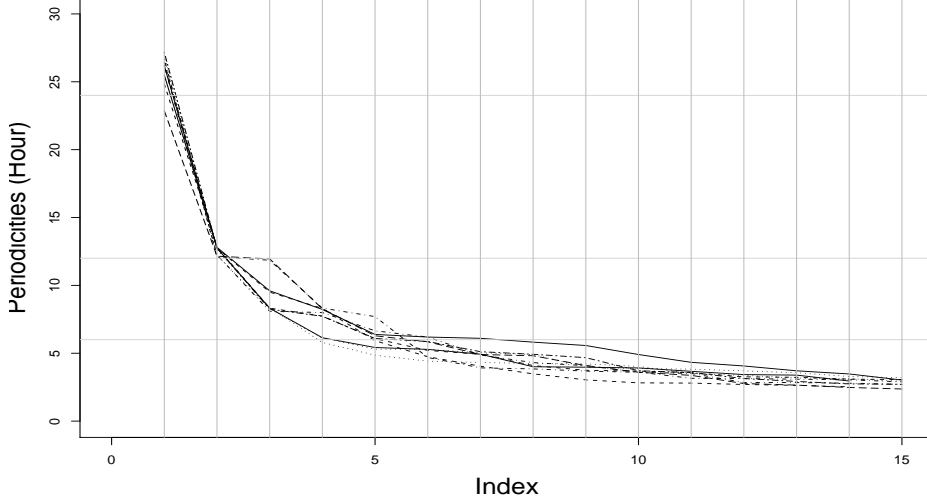


Figure 3: Periodicity plots for the Cluster 2 sites. The horizontal axes represents the index of the descending order at which the spectrum density for each site is estimated.

al. (2004), with a state vector equation accounting for trend and periodicity across sites.

To describe the model suggested by the DLM, let y_{it} denote the square-root of the observable ozone concentration, at site \mathbf{s}_i , $i = 1, \dots, n$, and time t , $t = 1, \dots, T$, where n being the total number of gauged 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

$$\mathbf{y}_t = \mathbf{1}'_n \beta_t + S_{1t}(a_1)\alpha_{1t} + S_{2t}(a_2)\alpha_{2t} + \nu_t \quad (11)$$

$$\beta_t = \beta_{t-1} + w_t \quad (12)$$

$$\alpha_{jt} = \alpha_{j,t-1} + \omega_t^{\alpha_j}, \quad (13)$$

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 β_t denotes a canonical spatial trend and α_{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 (11)–(13) can also be written as in (7)–(8) by letting $G_t = I$, $V_t = \sigma^2 V_\lambda$ and $W_t = \sigma^2 W$, where W being a block diagonal matrix with diagonal entries τ_y^2 , $\tau_1^2 \exp(-V/\lambda_1)$, and $\tau_2^2 \exp(-V/\lambda_2)$. In other words, the observation and state equations can also be written as

$$\mathbf{y}_t = \mathbf{F}'_t \mathbf{x}_t + \nu_t \quad (14)$$

and
$$\mathbf{x}_t = \mathbf{x}_{t-1} + \omega_t, \tag{15}$$

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

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

Let $y_{1:T} = (y_{1:T}^m, y_{1:T}^o)'$, where $y_{1:T}^m = (y_1^m, \dots, y_T^m)$ representing all the missing values and $y_{1:T}^o$, all the observed values in Cluster 2 sites for $t = 1, \dots, T$. The model parameters are $(\lambda, \sigma^2, x_{1:T}, y_{1:T}^m, a_1, a_2)$, in which the vector of state parameters up to time T is given by $x_{1:T} = (\mathbf{x}_1, \dots, \mathbf{x}_T)$, the range parameter by λ , the variance parameter by σ^2 and the phase parameters by $\mathbf{a} = (a_1, a_2)$. 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 make computation feasible.

Specification of the DLM is then completed with the following hyperpriors for 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 the hyperpriors for λ and σ^2 whose choice is discussed in Section 7 have inverse Gamma distributions for computational convenience.²

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 (14)–(15) for the purpose of inference on the range, variance and state parameters (see Section 5.2), and use models (11)–(13) for inference on the phase parameters (see Section 5.4) and interpolation (see Section 6).

The state–space models in (14)–(15) capture some important features of field as indicated by data from the AIRS database. In particular, they reflect the time–dependent structure of the field and capture the diurnal cycles of the hourly ozone concentrations over the geographical subdomains of concern. In the next section, the DLM is implemented in this field.

4 Parameter specification considerations

Before proceeding 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 on how the model’s parameters should be specified as well as undesirable consequences of inappropriately choices. Our assessment will focus on the accuracy of the model’s predictions.

² $X \sim IG(\alpha, \beta)$ if $Y = 1/X \sim G(\alpha, \beta)$, where $p(y) \propto y^{\alpha-1} \exp(-\beta y)$ for $\alpha, \beta > 0$.

The simple model we consider is a special case of the so-called “first-order polynomial model”, a mathematically tractable, commonly used model that captures many important features and properties of the DLM we use.

Emulating the DLM setting described above, let $p = 1$, $F_t' = \mathbf{1}_n'$: $n \times 1$, $G_t = 1$, and $\mathbf{x}_t = \beta_t$ in (7) and (8). For $i = 0, 1, \dots, n$ ($i = 0$ indicates the ungauged location) and $t = 1, \dots, T$, the first-order polynomial DLM is given by

$$y_{it} = \beta_t + \varepsilon_{it} \quad (16)$$

$$\beta_t = \beta_{t-1} + \delta_t, \quad (17)$$

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 σ_β^2 are all known.

The first-order polynomial DLM is particularly useful for short-term prediction. For a small time change the underlying evolution β_t is roughly constant. Moreover, the evolution error δ_t has zero-mean and is independent over time, so that the underlying process is merely a random walk, meaning the model does not anticipate long-term variation. At any fixed time t ,

$$\beta_t = \beta_0 + \sum_{k=1}^t \delta_k \quad (18)$$

$$y_{it} = \beta_0 + \sum_{k=1}^t \delta_k + \varepsilon_{ik}. \quad (19)$$

Consequently, the covariance structure of the first-order polynomial DLM is as follows:

$$\text{Var}(y_{it}) = \sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2 \quad (20)$$

$$\text{Cov}(y_{it}, y_{jt}) = \sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{ij}/\lambda) \quad (21)$$

$$\text{Cov}(y_{it}, y_{js}) = \sigma_\beta^2 + \min\{t, s\}\sigma_\delta^2 \quad (s \neq t), \quad (22)$$

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

The first-order polynomial DLM in (16)–(17) has the important property of non-stationarity because the covariance functions in (21)–(22) depend on $\min\{t, s\}$, rather than $|t - s|$. This DLM thus defines a non-stationary spatio-temporal process since for the first-order polynomial model to be stationary, the eigenvalues of $G_t = G$ must lie inside of the unit circle (West and Harrison, 1997), whereas $G_t = 1$ in our model. Furthermore, the general DLM defined in Section 3 is non-stationary because $G_t = I_{2n+1}$ given all the model parameters in (14)–(15).

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

$$\text{Cor}(y_{it}, y_{jt}) = \frac{\sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{ij}/\lambda)}{\sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2} \quad (23)$$

$$\text{Cor}(y_{it}, y_{js}) = \frac{\sigma_\beta^2 + \min\{t, s\}\sigma_\delta^2}{\sqrt{\sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2} \sqrt{\sigma_\beta^2 + s\sigma_\delta^2 + \sigma_\varepsilon^2}} \quad (s \neq t) \quad (24)$$

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

Remark 1 *The correlations in (23) and (24) have the following properties when $i \neq j$:*

(i)

$$\text{Cor}(y_{it}, y_{jt}) > \text{Cor}(y_{it}, y_{js}) \quad (25)$$

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

(ii)

$$\text{Cor}(y_{it}, y_{jt}) - \text{Cor}(y_{it}, y_{js}) \quad (26)$$

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

Thus for any fixed time point t , $\text{Cor}(y_{it}, y_{js})$ is maximized at $s = t$ and decreases as $|s - t|$ increases.

Remark 2 *By (23), $\text{Cor}(y_{it}, y_{jt}) \rightarrow 1$ as $t \rightarrow \infty$ for $i, j \in \{0, \dots, n\}$. 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 σ_δ^2 , depend on time. More specifically, (23) suggests making $t\sigma_\delta^2 = O(1)$ ³ to stabilize $\text{Cor}(y_{it}, y_{jt})$.*

Carrying this assessment further,

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

for any two sites in close proximity, i.e., $d_{ij} \simeq 0$, a result that seems quite reasonable. For two disjoint sites very far from each other 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 $t\sigma_\delta^2 = O(1) \ll \sigma_\varepsilon^2$.

³ $O(1)$ represents a constant.

Now 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. Both (25) and (26) hold for $i = 0$. Furthermore, we have restrictions on selecting the model parameters σ_β^2 and σ_δ^2 , namely, $\sigma_\beta^2 \ll \sigma_\varepsilon^2$ and $t\sigma_\delta^2 = O(1) \ll \sigma_\varepsilon^2$. The key result in (23) suggests a simple but straightforward way to adjust the model parameter σ_δ^2 according to the size of T , namely, to replace σ_δ^2 by σ_δ^2/T . That choice is empirically validated in Section 9.

We now turn to the behavior of the predictive variances in the first-order polynomial DLM that helps us understand the interpolation results. In fact, the properties of the correlation structure in (25)–(26), 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 the 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 (16)–(17).

Result 1 *For the first-order polynomial DLM in (16)–(17) with $n = 1$ and $T = 2$, assume the prior for β_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*

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

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

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

$$\text{Var}(y_{02}|y_{12}) = \frac{(\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)^2 - (\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{01}/\lambda))^2}{\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2}; \quad (28)$$

$$\text{Var}(y_{01}|y_{11}, y_{12}) = \frac{M_1}{\Delta}; \quad (29)$$

$$\text{Var}(y_{02}|y_{11}, y_{12}) = \frac{M_2}{\Delta}; \quad (30)$$

where

$$\Delta = (\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)(\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2) - (\sigma_\beta^2 + \sigma_\delta^2)^2; \quad (31)$$

$$\begin{aligned} M_1 &= (\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)\{(\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)^2 - (\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{01}/\lambda))^2\} \\ &\quad - 2(\sigma_\beta^2 + \sigma_\delta^2)^2(\sigma_\varepsilon^2 - \sigma_\varepsilon^2 \exp(-d_{01}/\lambda)); \end{aligned} \quad (32)$$

and

$$\begin{aligned} M_2 &= (\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)\{(\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)^2 - (\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{01}/\lambda))^2\} \\ &\quad - 2(\sigma_\beta^2 + \sigma_\delta^2)^2(\sigma_\varepsilon^2 - \sigma_\varepsilon^2 \exp(-d_{01}/\lambda)). \end{aligned} \quad (33)$$

For our simple model, 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,

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

The next result, whose proof is straightforward and hence omitted, proves these conjectures and provides other comparisons of these predictive variances.

Result 2 *Under the particular first-order polynomial DLM in Result 1, we have the following properties of the predictive conditional variances:*

$$\text{Var}(y_{01}|y_{11}) - \text{Var}(y_{01}|y_{11}, y_{12}) = \frac{\sigma_\varepsilon^4(\sigma_\beta^2 + \sigma_\delta^2)^2(1 - \exp(-d_{01}/\lambda))^2}{\Delta(\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)} \geq 0; \quad (34)$$

$$\text{Var}(y_{02}|y_{12}) - \text{Var}(y_{02}|y_{11}, y_{12}) = \frac{\sigma_\varepsilon^4(\sigma_\beta^2 + \sigma_\delta^2)^2(1 - \exp(-d_{01}/\lambda))^2}{\Delta(\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)} \geq 0; \quad (35)$$

$$\text{Var}(y_{02}|y_{11}, y_{12}) - \text{Var}(y_{01}|y_{11}, y_{12}) = \frac{\sigma_\varepsilon^4\sigma_\delta^2(1 - \exp(-d_{01}/\lambda))^2}{\Delta} \geq 0; \quad (36)$$

$$\text{Var}(y_{02}|y_{12}) - \text{Var}(y_{01}|y_{11}) = \frac{\sigma_\varepsilon^4\sigma_\delta^2(1 - \exp(-d_{01}/\lambda))^2}{(\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)(\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)} \geq 0; \quad (37)$$

$$\text{Var}(y_{01}|y_{11}) - \text{Var}(y_{01}|y_{11}, y_{12}) \geq \text{Var}(y_{02}|y_{12}) - \text{Var}(y_{02}|y_{11}, y_{12}); \quad (38)$$

$$\text{Var}(y_{02}|y_{12}) - \text{Var}(y_{01}|y_{11}) \leq \text{Var}(y_{02}|y_{11}, y_{12}) - \text{Var}(y_{01}|y_{11}, y_{12}). \quad (39)$$

Remark 3 *In Result 2, (34)–(36) verify our conjectures and we then conclude that the predictive variance function is a monotonic increasing function of time t given the same set of data. It decreases when more data and equivalently, more time is involved. Furthermore, the difference between these predictive variances decrease as t increases. It increases with time even when conditioning on the same amount of data.*

As an immediate consequence of (36), 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 8. There we will also discuss the lessons learned in this section in relation to our empirical findings.

Next, an interesting finding on the properties of the above predictive variances may explain their key features in the next result. This result considers these predictive variances changing as a function of λ , d_{01} or σ_ε^2 . Part of its proof is included in Appendix A.1.

Result 3 *The predictive conditional variances in (27)–(33) increase as d_{01} increases, or σ_ε^2 increases, or λ decreases.*

Thus, keeping two variables as fixed, these predictive conditional variances are monotone as a function of the remaining variable. It is interesting to note that applying the DLM could paradoxically lead to larger predictive variances when conditioning on more data. For example, for the case that $n = 1$ and $T = 2$, applying the DLM model with only data at $T = 2$ yields a predictive variance $Var^*(y_{02}|y_{12})$, which is exactly the same as $Var(y_{01}|y_{11})$ in (27). This predictive variance is smaller than $Var(y_{02}|y_{11}, y_{12})$ in (30), which is based on more data, under certain condition as specified in the result below.

Result 4 *Under the particular first-order polynomial DLM in Result 1,*

$$Var^*(y_{02}|y_{12}) < Var(y_{02}|y_{11}, y_{12}) \quad \text{if and only if} \quad \sigma_\varepsilon^2 > \sigma_\beta^2 \left(1 + \frac{\sigma_\beta^2}{\sigma_\delta^2} \right) \quad (40)$$

This proof is straightforward and so omitted here. This behavior is also seen in our application as demonstrated by the empirical results in Section 8.

5 Estimation algorithms

This section describes how to sample from the parameters' posterior distributions using the MCMC method, more specifically the forward-filtering-backward-sampling algorithm of Carter and Kohn (1994).

Section 5.1 introduces the Metropolis-within-Gibbs method to sample from the target joint posterior distribution of the model parameters. Sections 5.2, 5.3 and 5.4 show how to carry out that method in the DLM. Section 5.5 summarizes the algorithm to estimate the model parameters.

5.1 Metropolis-within-Gibbs algorithm

Recall the state space model defined in Equations (7) and (8). Let $y_{1:T}^o = (\mathbf{y}_1^o, \dots, \mathbf{y}_T^o) : n \times T$ be the observation matrix at the n gauged sites up to time T . Let $x_{1:T} = (x_1, \dots, x_T) : (2n + 1) \times T$ be the state parameters at the n gauged sites until time T . For simplicity, the values of γ are fixed but the problem of setting them will be addressed below.

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. Since it does not have a closed form, direct sampling methods encounter difficulties in drawing samples from it. The Markov chain Monte Carlo (MCMC) method is a popular way to sequentially sample parameters from their posterior distributions depending on the last value drawn, until the chain converges. In fact, the MCMC method can be used to obtain the posterior, predictive and interpolated results for the DLM (Huerta et al., 2004).

Here a *blocking MCMC* scheme is used to sample iteratively each component of the posterior distribution. Three blocks are chosen: $(\lambda, \sigma^2, x_{1:T})$, $y_{1:T}^m$ and (a_1, a_2) . Two reasons lead us to this choice. Firstly, it is natural to select blocks in which the parameters are highly correlated but relatively conditionally independent between the blocks. The phase parameters are assumed independent in time and location and so are not much correlated with the other model parameters. Another reason stems from the fact that the full conditional posterior distribution of the phase parameters can be obtained by assuming a bivariate normal hyperprior. These considerations lead to the method of inference presented in Appendix A.2.

Having no direct way of sampling from the posterior distribution, *Gibbs sampling* is used to sample these two blocks iteratively from their full conditional posterior distribution. In other words, iteratively, 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 $p(a_1, a_2 | x_{1:T}, \lambda, \sigma^2, y_{1:T})$.

However, $p(\lambda, \sigma^2, x_{1:T} | a_1, a_2, y_{1:T})$ has no closed form either. Instead, the full conditional posterior distribution of $x_{1:T}$ can be obtained explicitly by Kalman filtering and smoothing (Section 2.3), in other words, by the FFBS algorithm. Assuming an inverse Gamma hyperprior for σ^2 , the conditional posterior distribution of σ^2 given the range and phase parameters is also inverse Gamma distributed with new shape and scale parameters. Note that

$$p(\lambda, \sigma^2, x_{1:T} | a_1, a_2, y_{1:T}) = p(\lambda | a_1, a_2, y_{1:T}) p(\sigma^2 | \lambda, a_1, a_2, y_{1:T}) \times p(x_{1:T} | \lambda, \sigma^2, a_1, a_2, y_{1:T}), \quad (41)$$

indicating that we can sample iteratively from the three conditional posterior distributions on the right-hand-side of (41) 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 λ by a *Metropolis–Hasting* chain within a Gibbs sampling cycle, an algorithm often referred to as *Metropolis–within–Gibbs*. The next three subsections selectively present the details.

5.2 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 (41), we could ideally iteratively sample λ from $p(\lambda | a_1, a_2, y_{1:T})$, σ^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 λ , given the data, from the following a quantity that is proportional to its posterior density, that

is,

$$p(\lambda|a_1, a_2, y_{1:T}) \propto p(\lambda) \prod_{t=1}^T |Q_t|^{-\frac{1}{2}} \left[\beta + \frac{1}{2} \sum_{t=1}^T \mathbf{e}_t' Q_t^{-1} \mathbf{e}_t \right]^{-(nT/2+\alpha)}. \quad (42)$$

Appendix A.2 gives the details. Since we cannot compute the normalization constant for $p(\lambda|a_1, a_2, y_{1:T})$, the Metropolis–Hasting algorithm is used here; it is impossible to sample λ directly from its posterior distribution. The main reason for developing the Metropolis–Hasting algorithm was the need for an equilibrium distribution for the Markov Chain, since computation and simulation are easier for reversible chains where the transition probabilities and stationary density of the chain satisfy the detailed balance equations. In the Metropolis–Hasting algorithm, the transition kernel is a mixed distribution for the new state of the chain: $q(., .)$, the proposal density and $\alpha(., .)$, the acceptance probability.

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 λ^* from this proposal distribution, centered at the previously sampled $\lambda^{(j-1)}$ with a tuning parameter, τ^2 , as the variance for the distribution of Z . Gamerman (2006) suggests the acceptance rate, that is, the ratio of accepted λ^* to the total number of iterations, be around 50%. We tune τ^2 to attain that rate. If the acceptance rate were too high, for example, 70% to 100%, we would increase τ^2 . If too low, for example, 0 to 20%, we would decrease τ^2 , to narrow down the search area for λ^* .

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 λ^* 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 λ^* 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 σ^2 given the accepted λ 's, a_1 's, a_2 's and $y_{1:T}$. The prior for σ^2 is chosen to be an inverse gamma distribution with shape parameter α and scale parameter β . The posterior distribution for σ^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 λ 's, σ^2 's, phase parameters and $y_{1:T}$, using the forward-filtering-backward-sampling (FFBS) method as described in Section 2.3.

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. We present Theorem 2 in Appendix A.3 as the forward-filtering-backward-sampling (FFBS) algorithm (similar to the Kalman filter algorithm) to resample the state parameters conditional on all the other parameters and observations. FFBS is used as part of a MCMC method to sample $x_{1:T} = (\mathbf{x}_1, \dots, \mathbf{x}_T)$ from the smoothing distribution $p(\mathbf{x}_t | \lambda, \sigma^2, a_1, a_2, y_{1:T})$. It is called FFBS because recent data are used to update the state parameters, \mathbf{x}_t 's, recursively from $t = 1, \dots, T$, as well as to sample each element of the $\{\mathbf{x}_t\}$ using all the information recursively from $t = T, \dots, 1$.

Theoretical results for models (14) and (15) are presented to give some idea on how to resample $x_{1:T}$ conditionally on the other parameters and observations. One can also obtain it from Theorem 1 in Section 2.3, by letting $G_t = I$, $V_t = \sigma^2 V_\lambda$ and $W_t = \sigma^2 W$.

The initial state parameter is given by

$$(\mathbf{x}_0 | \mathbf{y}_0, \theta) \sim N[\mathbf{m}_0, \sigma^2 C_0], \quad (43)$$

where \mathbf{y}_0 being the initial information, with \mathbf{m}_0 and C_0 known. Later in Section 7, we consider how to set them for Cluster 2 AIRS dataset (1995). Let $\theta = (\lambda, \sigma^2, a_1, a_2, \gamma)$. Assume all the prior information has been given and θ 's coordinates are mutually independent. The details on this algorithm are included in Appendix A.3.

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

MCMC enjoys the advantage of being able to fill in missing values at each iteration. In other words, missing values can be treated as if they were model “parameters”. It then avoids using the roughly fitted missing values by period means or spatial regression method, a possible reason leading to the inaccurate interpolation results.

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 $\mathbf{e}_j = (e_{j1}, \dots, e_{jn})' : 1 \times n$ 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

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

where

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

We can also decompose $\tilde{\boldsymbol{\mu}}_t$ as

$$\tilde{\boldsymbol{\mu}}_t = \begin{pmatrix} \tilde{\boldsymbol{\mu}}_t^{\mathbf{m}} \\ \tilde{\boldsymbol{\mu}}_t^{\mathbf{o}} \end{pmatrix}$$

where $\tilde{\boldsymbol{\mu}}_t^{\mathbf{m}} : n_t \times 1$ and $\tilde{\boldsymbol{\mu}}_t^{\mathbf{o}} : (n - n_t) \times 1$.

Similarly, we have

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

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

By a standard property of multivariate normal distribution, we have

$$(\mathbf{y}_t^{\mathbf{m}} | \lambda, \sigma^2, \mathbf{x}_t, \mathbf{a}, \mathbf{y}_t^{\mathbf{o}}) \sim N[\boldsymbol{\mu}_t^{**}, \boldsymbol{\Sigma}_t^{**}], \quad (44)$$

where

$$\boldsymbol{\mu}_t^{**} = \tilde{\boldsymbol{\mu}}_t^{\mathbf{m}} + \tilde{\boldsymbol{\Sigma}}_t^{mo} (\tilde{\boldsymbol{\Sigma}}_t^{oo})^{-1} (\mathbf{y}_t^{\mathbf{o}} - \tilde{\boldsymbol{\mu}}_t^{\mathbf{o}}), \quad (45)$$

and

$$\boldsymbol{\Sigma}_t^{**} = \tilde{\boldsymbol{\Sigma}}_t^{mm} - \tilde{\boldsymbol{\Sigma}}_t^{mo} (\tilde{\boldsymbol{\Sigma}}_t^{oo})^{-1} \tilde{\boldsymbol{\Sigma}}_t^{om}, \quad (46)$$

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

At each iteration, we draw $\mathbf{y}_t^{\mathbf{m}}$ from the corresponding distribution (44) at each time point t and then we can write the response variables as $y_{1:T} = (y_{1:T}^{\mathbf{m}}, y_{1:T}^{\mathbf{o}})$ without loss of generality, where $y_{1:T}^{\mathbf{m}} = (y_1^{\mathbf{m}}, \dots, y_T^{\mathbf{m}})$ and $y_{1:T}^{\mathbf{o}} = (y_1^{\mathbf{o}}, \dots, y_T^{\mathbf{o}})$.

5.4 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 λ , σ^2 and $x_{1:T}$ as illustrated in Sections 5.2 and 5.3. For simplicity, we use the notation for models (11)–(13) 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 Σ^o . Then the posterior conditional distribution for \mathbf{a} is normal with mean vector μ^* and covariance matrix Σ^* , where μ^* and Σ^* can be obtained from Equations (60) and (61), respectively. This result is shown in Appendix A.4.

We will not use a non-informative prior for \mathbf{a} such as $p(\mathbf{a}) \propto 1$, a problematic choice. It is due to the fact that we need to avoid non-identified posterior means or posterior variances. To be more specific, assume $p(\mathbf{a}) \propto 1$. Using the above approach, we find that the posterior conditional distribution of \mathbf{a} is normal with mean vector $\mu = (\mu_1, \mu_2)'$ and covariance matrix Σ from equations (49) and (50), respectively. The elements of Σ are also given in Appendix A.4, where Σ can be singular for any $t = 12k$, where k is an integer. 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 λ , σ^2 and $x_{1:T}$, we sample \mathbf{a} from $N(\mu^*, \Sigma^*)$ and then obtain the median as the estimator for a for each fixed iteration due to the assumption that (a_1, a_2) are constant phase parameters in the models (14)–(15).

5.5 Summary

The MCMC methods 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:

Algorithm	The Metropolis-within-Gibbs method
-----------	------------------------------------

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 $(j-1)$ th 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 λ^* from a logarithm proposal distribution $q(\lambda^{(j-1)}, \lambda)$, that is, $LN(\lambda^{(j-1)}, \tau^2)$ for some suitable tuning parameter τ^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 develop the software into implementing the DLM approach illustrated in this section. To carry out the Metropolis-within-Gibbs algorithm, this software is written on R and C language to speed up the computation. Current version, *GDLM.1.0*, is available for free download at <http://enviro.stat.ubc.ca> for different platforms such as Windows, Unix and Linux.

6 Algorithms for 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 5). 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

$$\mathbf{y}_t^{\text{new}} = \mathbf{1}_{n+1}' \beta_t + S_{1t}(a_1) \alpha_{1t}^{\text{new}} + S_{2t}(a_2) \alpha_{2t}^{\text{new}} + \nu_t^{\text{new}}, \quad (47)$$

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

In Section 6.1, 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. Spatial interpolation at the ungauged site \mathbf{s} is demonstrated in Section 6.2.

6.1 Sampling the unobserved state parameters

We first sample α_{jt}^s given $\alpha_{j,t-1}^s$, α_{jt} and $\alpha_{j,t-1}$, $j = 1, 2$. From the state equation (15) for α_{jt}^{new} , we know that the joint density of α_{jt}^s and α_{jt} follows a normal distribution, with covariance matrix $\sigma^2 \tau_j^2 \exp(-V^{\text{new}}/\lambda_j)$, where V^{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.5.

6.2 Spatial interpolation at the 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. Similarly as above, y_t^s and \mathbf{y}_t are jointly normally distributed from 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.5.

7 Application

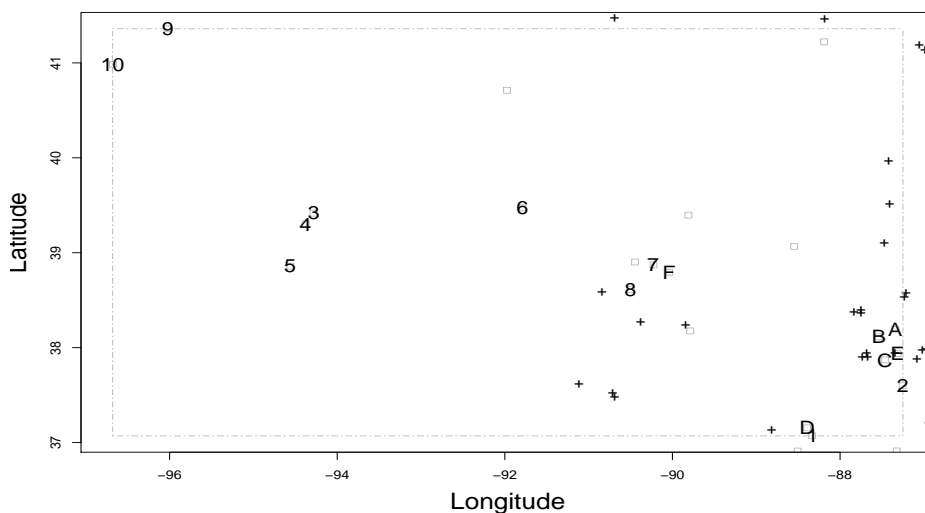


Figure 4: Geographical locations of ten gauged sites in Cluster 2 and the randomly chosen six ungauged sites. (Number = Cluster 2 sites and letter = ungauged sites.)

This section applies the model to the hourly ozone concentration field described above. Within the range of Cluster 2, six additional monitoring sites are randomly selected to play the role of “unmonitored sites” from those available. They enable us to assess the spatial predictor based on our DLM and hence the DLM itself. The geographical locations of these six ungauged sites, represented by the alphabetic letters, A, \dots, F , are shown in Figure 4, along with the sites in Cluster 2.

7.1 MCMC simulation

This subsection presents a MCMC simulation study in which we sample from the posterior distributions 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 λ is $IG(1, 5)$ and for σ^2 , $IG(2, 0.01)$. The expected value of $IG(1, 5)$ is ∞ and so are both of the variances of $p(\lambda)$ and $p(\sigma^2)$. These vague priors for λ and σ^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 $0.01\mathbf{1}'_n$.
- The hyperprior for \mathbf{a} is a bivariate normal distribution with mean vector $\mu^o = (2.5, 9.8)'$ and a diagonal matrix Σ^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

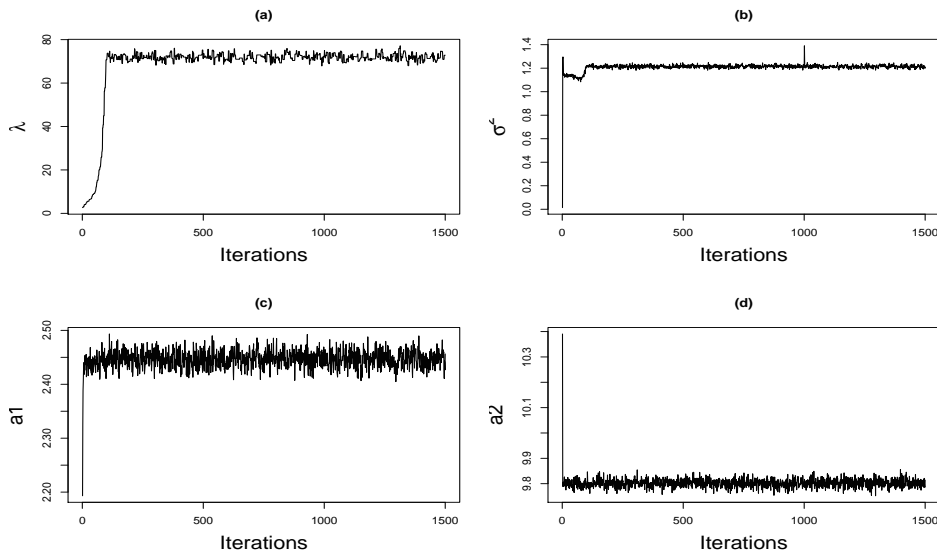


Figure 5: Traces of model parameters with the number of iterations of the Markov chains. The parameters are: (a) $-\lambda$, the range parameter; (b) $-\sigma^2$, the variance parameter; (c) $-a_1$, the phase parameter with respect to the 24-hour periodicity; and (d) $-a_2$, the phase parameter with respect to the 12-hour periodicity.

Figure 5 shows the trace plots of model parameters λ , σ^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 λ , σ^2 , a_1 and a_2 .

Quantile	λ	σ^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

Table 1: Posterior summaries for λ , σ^2 , a_1 and a_2 .

7.2 Spatial interpolation

This subsection assesses the model’s performance by comparing the interpolation values at the ungauged sites, A, \dots, F , with the measurements made there. We use the entire data 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. In general, 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 being in close proximity to a “friend” 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 unsatisfactorily large coverage probabilities point to a deficiency of the DLM.

Nominal Credible 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

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

To explore this issue further, Figures 6–10 show the interpolation results at Ungauged Site D from May 14 to September 11, 1995, where 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 this “ungauged” site.

Furthermore, Table 3 shows for all the ungauged sites, the friends they have among the Cluster 2 sites that lie within a radius of 100 km, the corre-

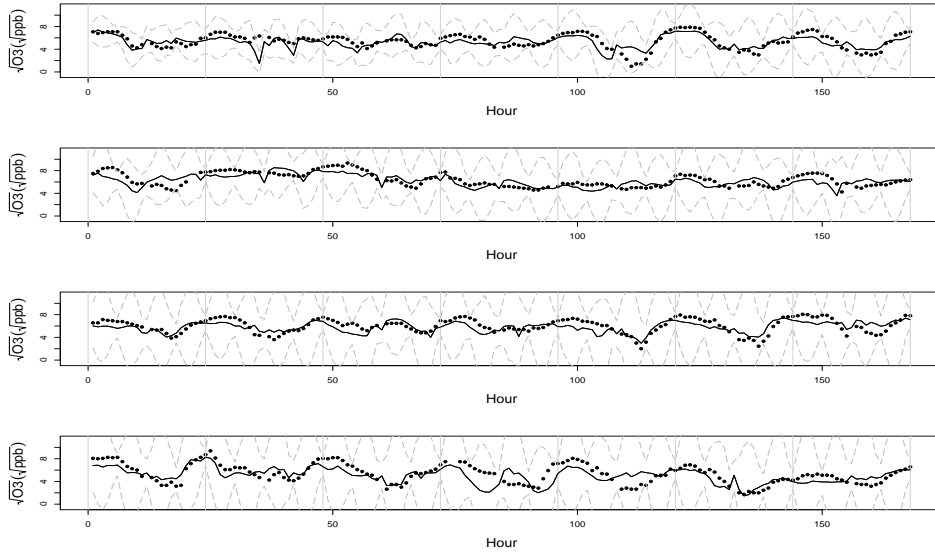


Figure 6: Interpolation at Ungauged Site D from the 1st week to the 4th week. The square-root of hourly ozone concentrations are plotted in the vertical axes, hours on the horizontal axes.

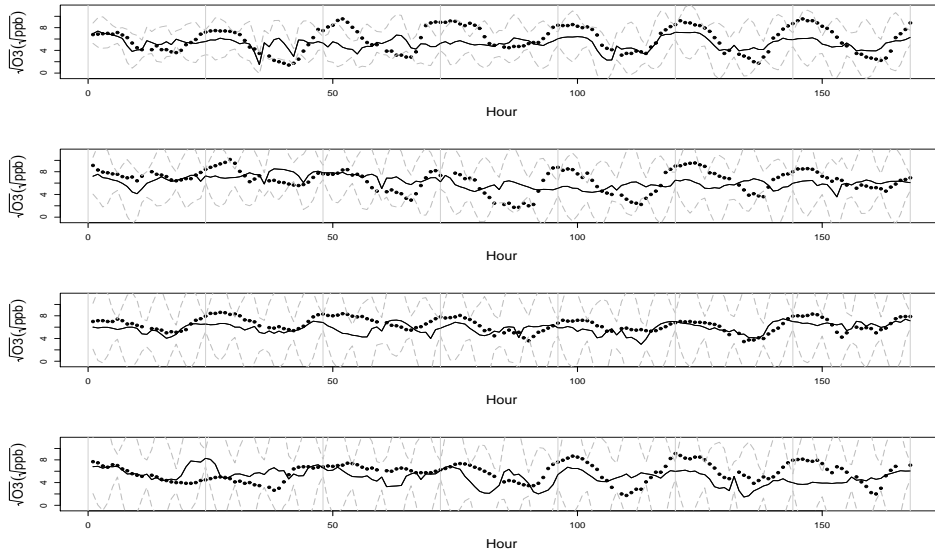


Figure 7: Interpolation at Ungauged Site D from the 5th week to the 8th week. The square-root of hourly ozone concentrations are plotted in the vertical axes, hours on the horizontal axes.

sponding 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 friend. That relationship is further explored in Figure

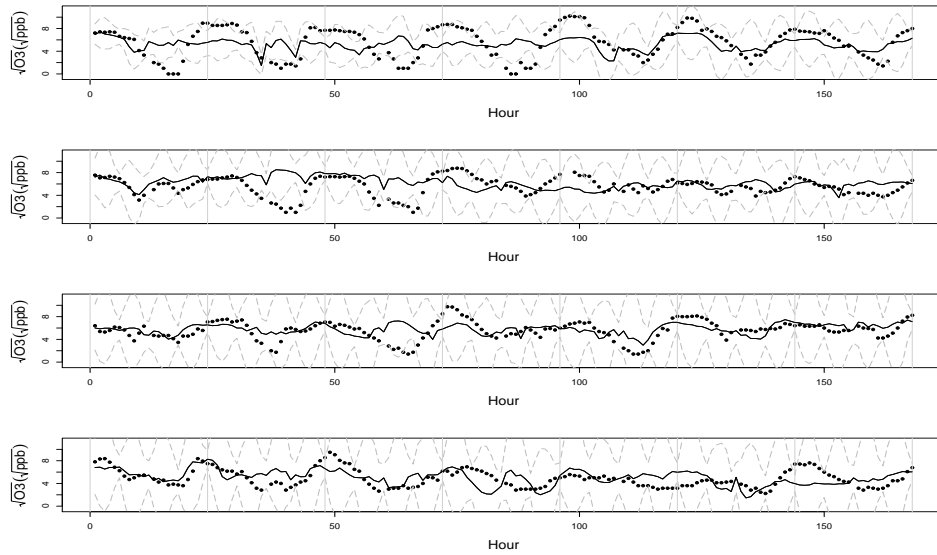


Figure 8: Interpolation at Ungauged Site D from the 9th week to the 12th week. The square-root of hourly ozone concentrations are plotted in the vertical axes, hours on the horizontal axes.

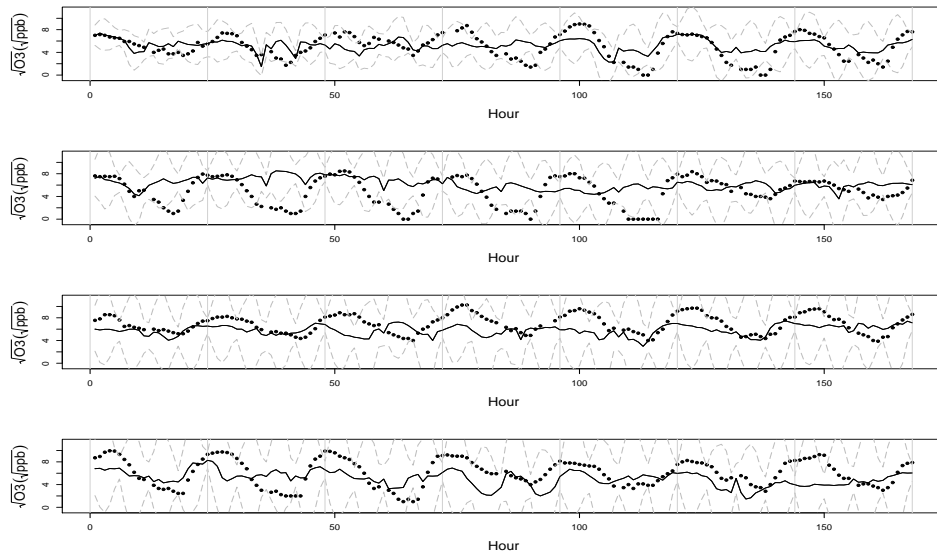


Figure 9: Interpolation at Ungauged Site D from the 13th week to the 16th week. The square-root of hourly ozone concentrations are plotted in the vertical axes, hours on the horizontal axes.

11 where 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 friends, the DLM does not predict responses

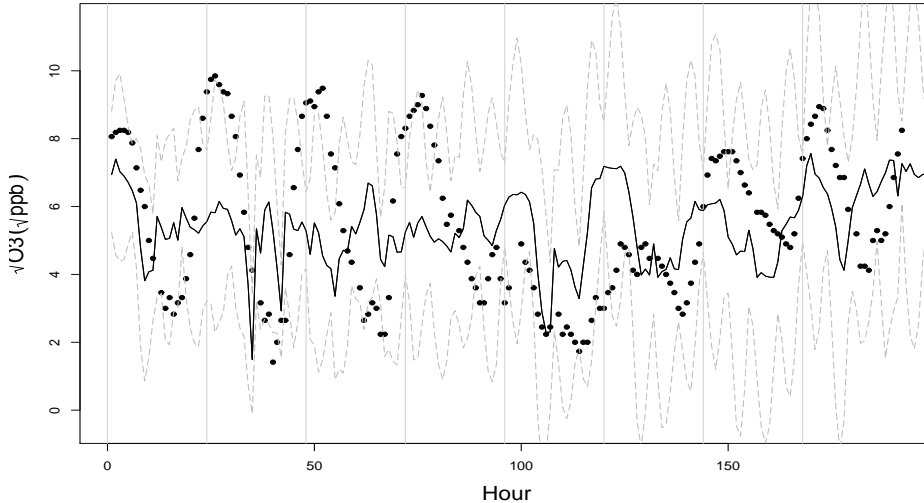


Figure 10: Interpolation at Ungauged Site D from the 17th week to the 120th day. The square-root of hourly ozone concentrations are plotted in the vertical axes, hours on the horizontal axes.

Ungauged Site	Friend(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)

Table 3: The “friends” of the ungauged sites, their GCDs in km, and the average of their correlations with their associated gauged sites.

at the ungauged sites very accurately for instance as illustrated in Figure 10. That points to problems with this model which will be discussed in the next section.

8 Discussion

In general, the DLM is a remarkably powerful modelling tool that has been made practical by advances in statistical computing. However, computational burden limits its domain of applicability. Moreover, the very flexibility that makes it so powerful also imposes an immense burden of choice on the mod-

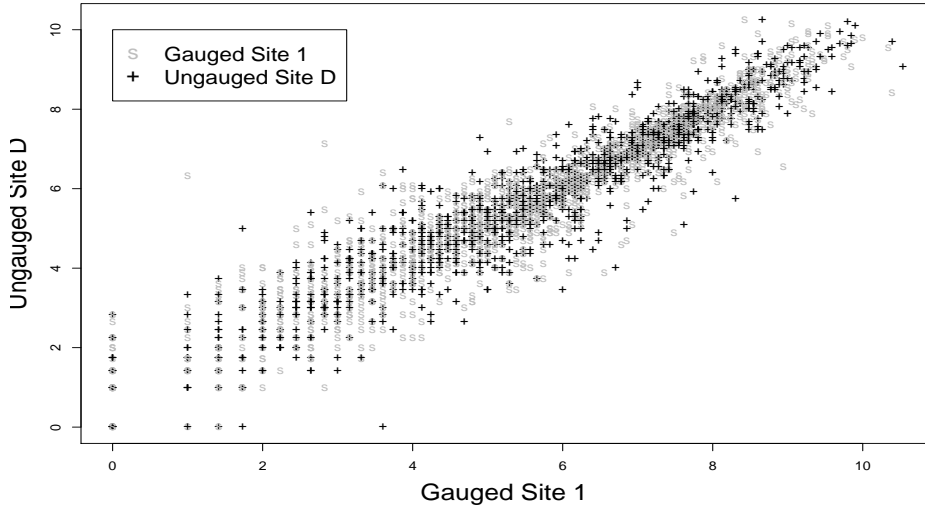


Figure 11: Scatterplot for the square-root of ozone concentrations at the ungauged site D and its “friend”, gauged site 1. The square-root of hourly ozone concentrations are plotted in both vertical and horizontal axes.

eller. This section summarizes critical issues and suggestions for improvement.

Monitoring MCMC convergence

Figure 12 represents the trace plots of model parameters λ , σ^2 , a_1 and a_2 of two chains from the initial settings in Section 7.1. These two chains seem to mix well after several hundreds iterations, suggesting at the first glance the convergence of the Markov chains.

Autocorrelation and partial autocorrelation of the simulated Markov chains

However, we know that the autocorrelation, which is 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 measure in assessing the Markov chain since large values of the PACF at lag h indicate that the next value in the chain is dependent on past values but not only the most recent one.

Figure 13 shows the histogram, ACF and PACF plots for the Markov chains used in Section 7.2, after a burn-in period of 1,000. It shows the λ s to be highly autocorrelated, judging from its ACF plots. In fact, it reveals that

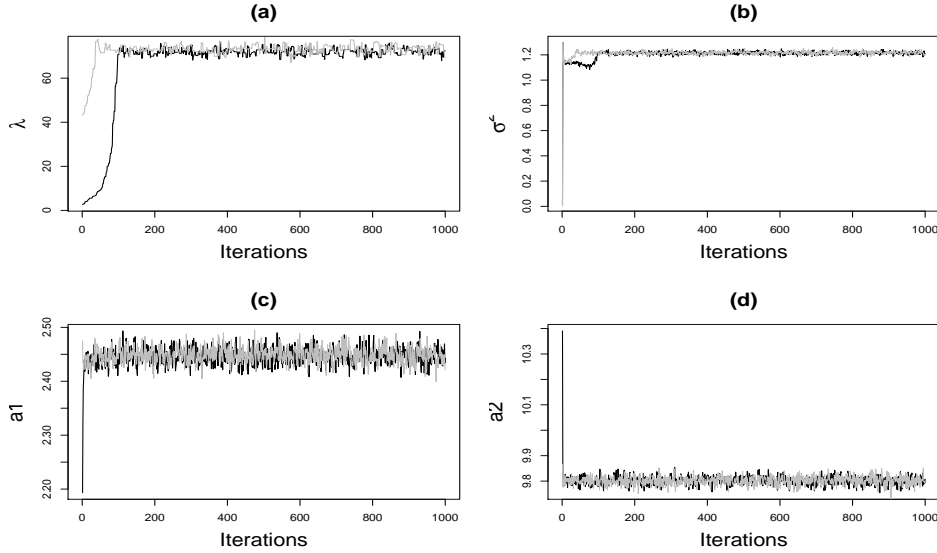


Figure 12: Traces of model parameters for a number of iterations of two chains. The parameters are: (a) $-\lambda$, the range parameter; (b) $-\sigma^2$, the variance parameter; (c) $-a_1$, the phase parameter with respect to the 24-hour periodicity; and (d) $-a_2$, the phase parameter with respect to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

the chain for λ is not mixing very well, leading to biased estimates in Section 7.2. A possible way to reduce the autocorrelation between these λ s is to thin the Markov chain. That is, we can use every k^{th} ($k > 1, k \in \mathcal{Z}^+$) λ generated by the chain to obtain the estimates. However, computational limitations have forced us to use the whole chain for estimation and interpolation without thinning.

Relationship between pairs of λ , σ^2 , a_1 and a_2

Our prior assumptions make the model parameters λ , σ^2 , a_1 and a_2 uncorrelated. Figure 14 investigates the relationship between the pairs of these parameters to explore that assumption. It seems valid except for the λ - σ^2 pair in graph (a). That graph shows a weak linear association between λ and σ^2 , thus pointing to a failure of that assumption for that pair. Since σ^2 determines spatial variability while λ determines correlation this relationship seems intriguing. Larger values of σ^2 tend to go with larger λ s, i.e., diminished spatial correlation. Why they are coupled in this way is unknown but it should be accounted for in future applications of this model.

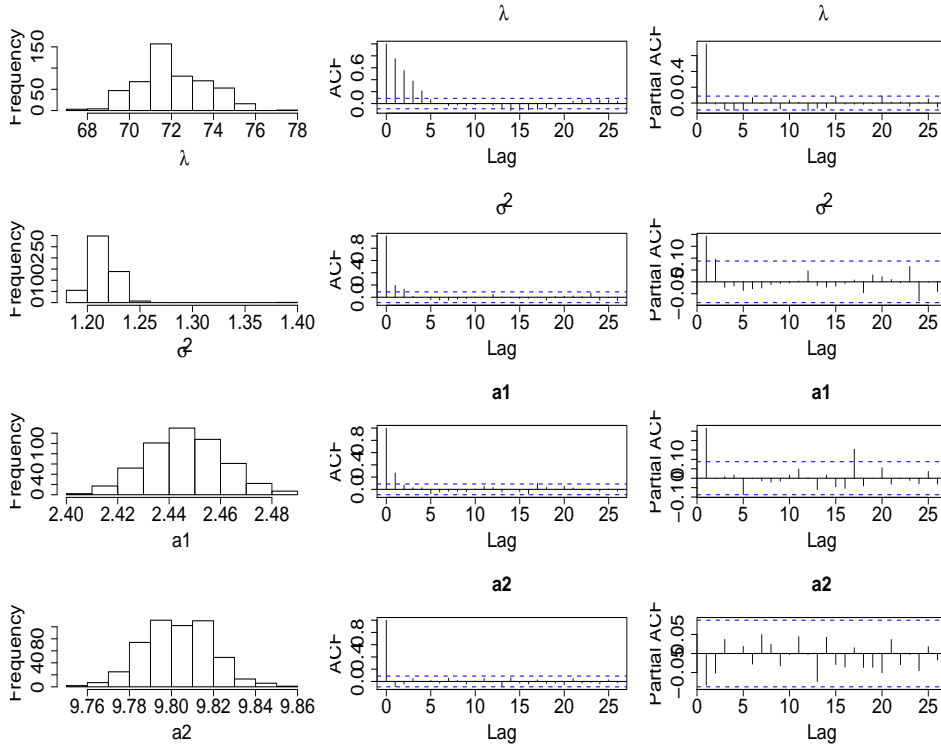


Figure 13: Histogram (left panel), ACF (middle panel) and PACF (right panel) of model parameters of the Markov chains after a burn-in period of 1,000. The parameters are: (i) first row: $-\lambda$, the range parameter; (ii) second row: $-\sigma^2$, the variance parameter; (iii) third row: $-a_1$, the 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.

Time varying effect of λ - σ^2 : empirical coverage probabilities versus nominal credible probabilities

Although we follow Huerta et al. (2004) in assuming the temporal constancy of λ and σ^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 σ^2 and λ with the freedom to choose the time span.

To answer these questions 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 λ , σ^2 , a_1 and a_2 . Obtain the coverage probabilities at each ungauged site and week for fixed credibility interval probabilities.

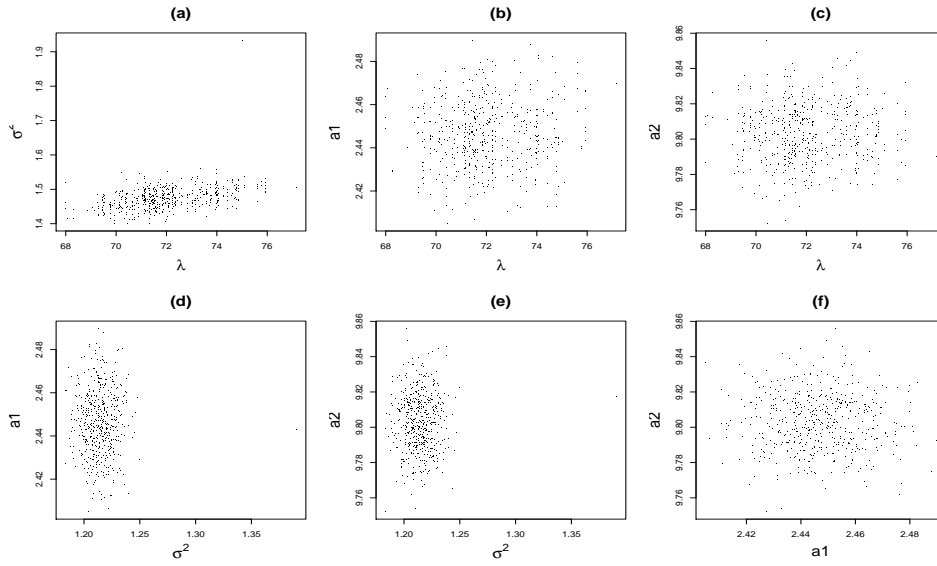


Figure 14: Scatterplots for the pairs of model parameters: (a) λ v.s. σ^2 ; (b) λ v.s. a_1 ; (c) λ v.s. a_2 ; (d) σ^2 v.s. a_1 ; (e) σ^2 v.s. a_2 ; and (f) a_1 v.s. a_2 .

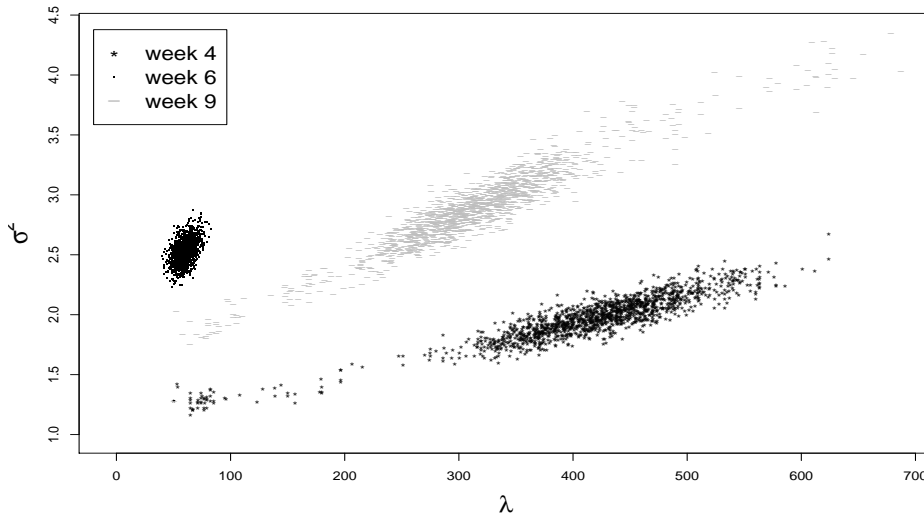


Figure 15: Scatterplots for λ against σ^2 for various weeks, based on the MCMC samples using one week's data, that is, week 4, 6 and 9, but starting from the same initial values as those in Section 7.1.

- (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 λ_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 5.5 but now using only fixed λ^* 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 differences in the interpolation results when using the fixed λ^* rather than using the Markov samples of λ s. It is also aimed at finding how much time would be saved by avoiding the inefficient Metropolis step. Table 4 shows these fixed λ^* s used in Study \tilde{C} . Table 5 shows the time saved using fixed λ^* s against the one using the Metropolis–Hasting algorithm.

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

Table 4: Fixed values of λ^* in Study \tilde{C} .

Study	Data	Iteration total	Accept(%)	Time (seconds)	
				Total	/Iteration
\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

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 run without spatial prediction.

Figure 15 illustrates the MCMC estimation results obtained in Study \tilde{A} . It plots the Markov chains of λ and σ^2 using weekly data. Obviously, λ and σ^2 vary from week to week, which implies that the constant λ – σ^2 model is not tenable over a whole summer for this dataset.

Figures 16–22 depict 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

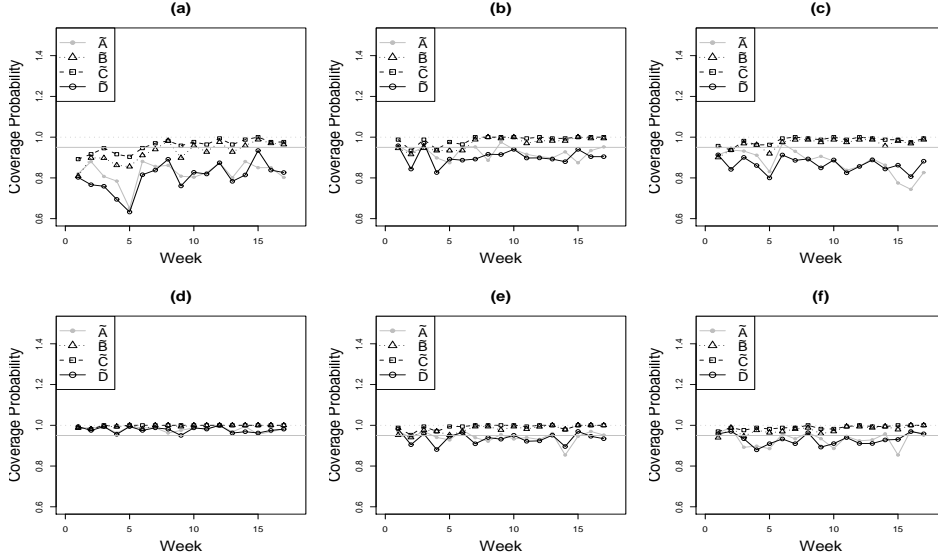


Figure 16: Coverage probability versus 95% nominal level for ungauged sites: (a) – site A; (b) – site B; (c) – site C; (d) – site D; (e) – site E; and (f) – site F. These coverage probabilities are computed for *Study \tilde{A}* : weekly data (solid bullet with solid line); *Study \tilde{B}* : $W_{1:17}$ (up-triangle with dotted line); *Study \tilde{C}* : $W_{1:17}$ but with fixed λ^* (square with dashed line); and *Study \tilde{D}* : $W_{1:17}$ but with fixed λ^* and modified τ_y^2 , τ_1^2 and τ_2^2 (empty circle with solid line).

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 λ^* s in the DLM to obtain similarly interpolating results as using the Metropolis–within–Gibbs algorithm.

We have studied the prediction accuracy of the simplest DLM, namely, the first–order polynomial model, in Section 4. 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 4). The plots exhibit a monotonic increasing trend in the coverage probabilities of both Studies \tilde{B} and \tilde{C} . This trend agrees with the graphs of the coverage probabilities in Figures 16–22. 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 $\lambda\text{--}\sigma^2$ monitored in Figure 15.

On the other hand, Study \tilde{C} involves significant computational time savings as compared with \tilde{B} . Table 5 suggests that the computation time–saving of the former study 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

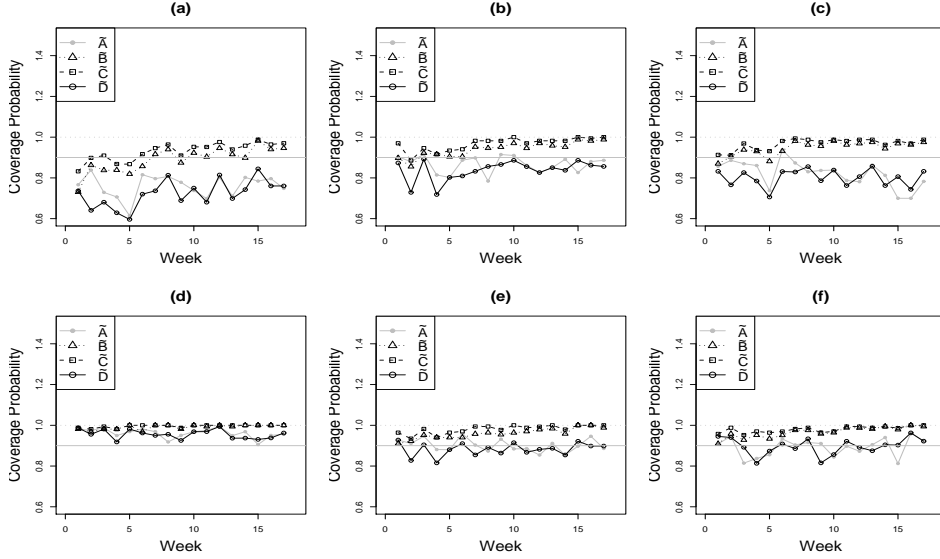


Figure 17: Coverage probability versus 90% nominal level for ungauged sites: (a) – site A; (b) – site B; (c) – site C; (d) – site D; (e) – site E; and (f) – site F. These coverage probabilities are computed for *Study \tilde{A}* : weekly data (solid bullet with solid line); *Study \tilde{B}* : $W_{1:17}$ (up-triangle with dotted line); *Study \tilde{C}* : $W_{1:17}$ but with fixed λ^* (square with dashed line); and *Study \tilde{D}* : $W_{1:17}$ but with fixed λ^* and modified τ_y^2 , τ_1^2 and τ_2^2 (empty circle with solid line).

for the DLM models, the λ s and σ^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 γ in Section 7.1.)

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, Result 4 in Section 4 predicts this finding. Because the prior for σ_1^2 is $IG(2, 0.01)$ the expectation of σ_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 σ_ε^2 (for example, the median of σ^2 is around 1.21 in Study \tilde{B} and even larger in Study \tilde{A}). By the sufficient and necessary condition in Result 4, the predictive variance of Study \tilde{A} is less than that of Study \tilde{B} . However, notice that σ^2 and λ 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 (d) of Figure 18, the coverage probability of \tilde{B} at the 4th 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, σ^2 of \tilde{A} is larger than that of \tilde{B} , leading an inflated predictive variance of \tilde{A} . This produces the difficulty to compare these two

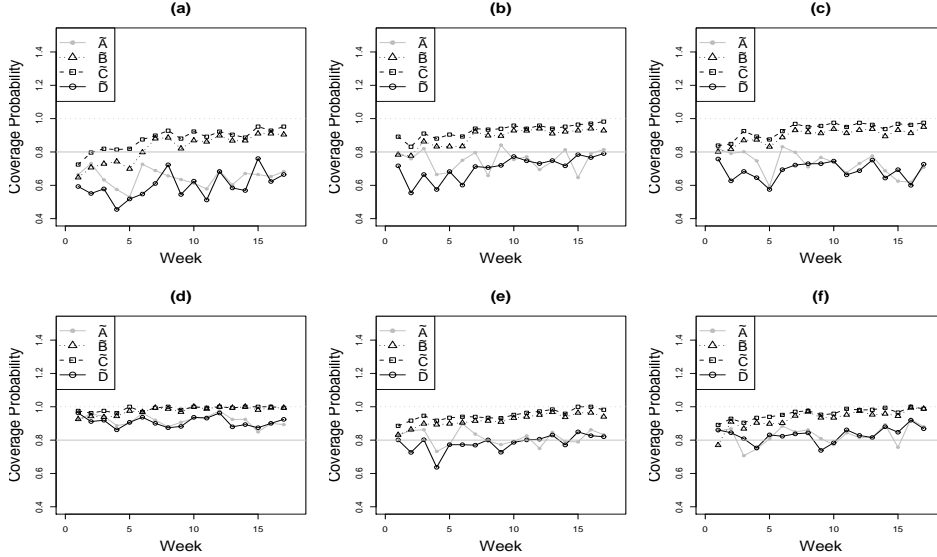


Figure 18: Coverage probability versus 80% nominal level for ungauged sites: (a) – site A; (b) – site B; (c) – site C; (d) – site D; (e) – site E; and (f) – site F. These coverage probabilities are computed for *Study \tilde{A}* : weekly data (solid bullet with solid line); *Study \tilde{B}* : $W_{1:17}$ (up-triangle with dotted line); *Study \tilde{C}* : $W_{1:17}$ but with fixed λ^* (square with dashed line); and *Study \tilde{D}* : $W_{1:17}$ but with fixed λ^* and modified τ_y^2 , τ_1^2 and τ_2^2 (empty circle with solid line).

predictive variances, but explains that paradox we saw in those figures.

9 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 λ and σ^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

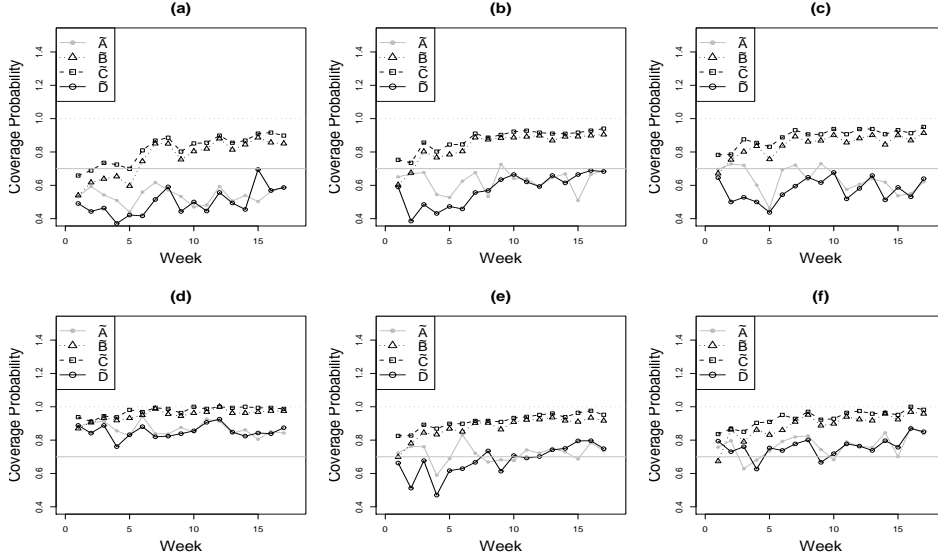


Figure 19: Coverage probability versus 70% nominal level for ungauged sites: (a) – site A; (b) – site B; (c) – site C; (d) – site D; (e) – site E; and (f) – site F. These coverage probabilities are computed for *Study A-tilde*: weekly data (solid bullet with solid line); *Study B-tilde*: $W_{1:17}$ (up-triangle with dotted line); *Study C-tilde*: $W_{1:17}$ but with fixed λ^* (square with dashed line); and *Study D-tilde*: $W_{1:17}$ but with fixed λ^* and modified τ_y^2 , τ_1^2 and τ_2^2 (empty circle with solid line).

intervals over successive weeks, conditional on all 17 week 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 τ_y^2 , τ_1^2 and τ_2^2 . Section 4 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 τ_y^2 , τ_1^2 and τ_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). Figures 16–22 compare 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 5.2. Suppose $\gamma = (\lambda, \tau_y^2, \tau_1^2, \lambda_1, \tau_2^2, \lambda_2)$. We can use six

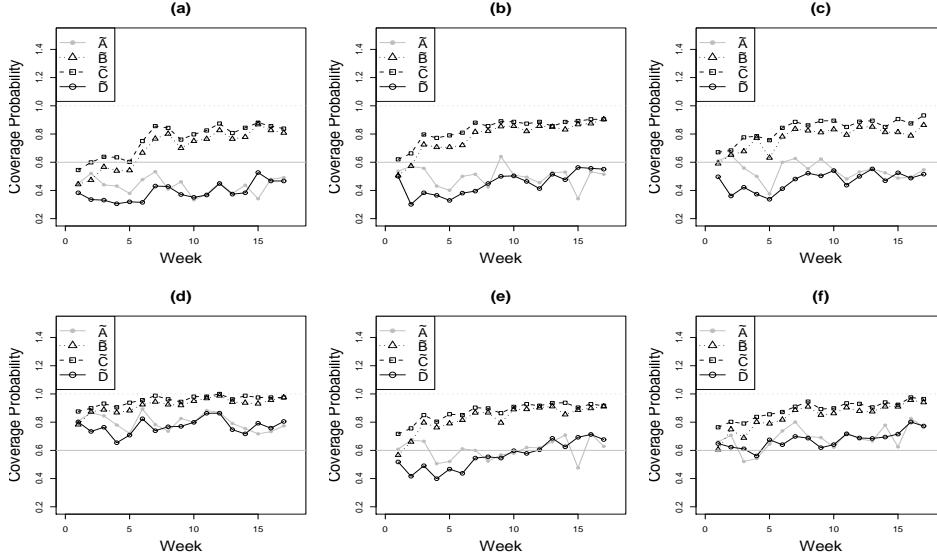


Figure 20: Coverage probability versus 60% nominal level for ungauged sites: (a) – site A; (b) – site B; (c) – site C; (d) – site D; (e) – site E; and (f) – site F. These coverage probabilities are computed for *Study A*: weekly data (solid bullet with solid line); *Study B*: $W_{1:17}$ (up-triangle with dotted line); *Study C*: $W_{1:17}$ but with fixed λ^* (square with dashed line); and *Study D*: $W_{1:17}$ but with fixed λ^* and modified τ_y^2 , τ_1^2 and τ_2^2 (empty circle with solid line).

Metropolis–Hasting kernels to sample from the target distribution $\pi(\gamma|y_{1:T})$, updating each component of γ iteratively. 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 λ s and σ^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 use 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.

A Supplementary results

A.1 Results for Section 4

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

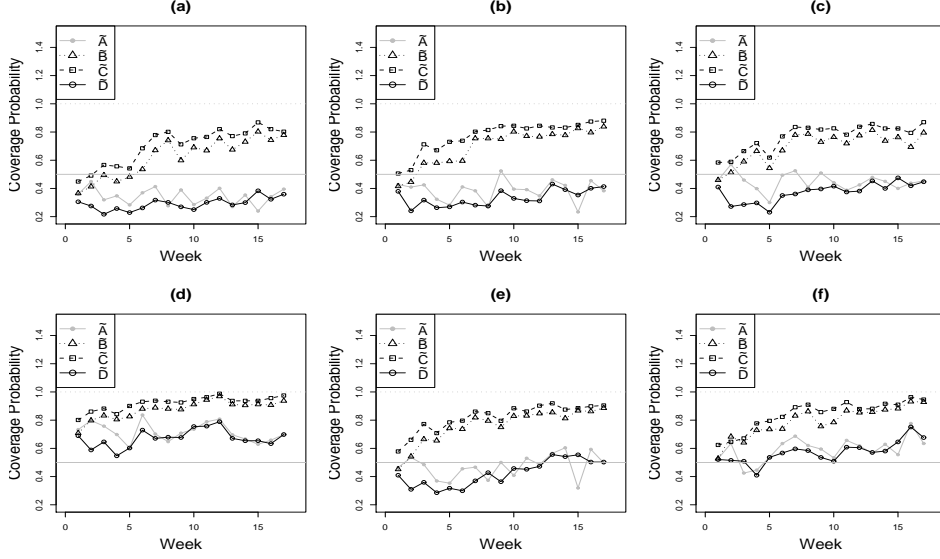


Figure 21: Coverage probability versus 50% nominal level for ungauged sites: (a) – site A; (b) – site B; (c) – site C; (d) – site D; (e) – site E; and (f) – site F. These coverage probabilities are computed for *Study A*: weekly data (solid bullet with solid line); *Study B*: $W_{1:17}$ (up-triangle with dotted line); *Study C*: $W_{1:17}$ but with fixed λ^* (square with dashed line); and *Study D*: $W_{1:17}$ but with fixed λ^* and modified τ_y^2 , τ_1^2 and τ_2^2 (empty circle with solid line).

method. Refer to Result 1, the predictive variance of $y_{01}|y_{11}$ can also be written as follows:

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

The first partial derivatives of this predictive variances regarding to d_{01} , λ and σ_ε^2 are given by:

(i)

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

(ii)

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

and

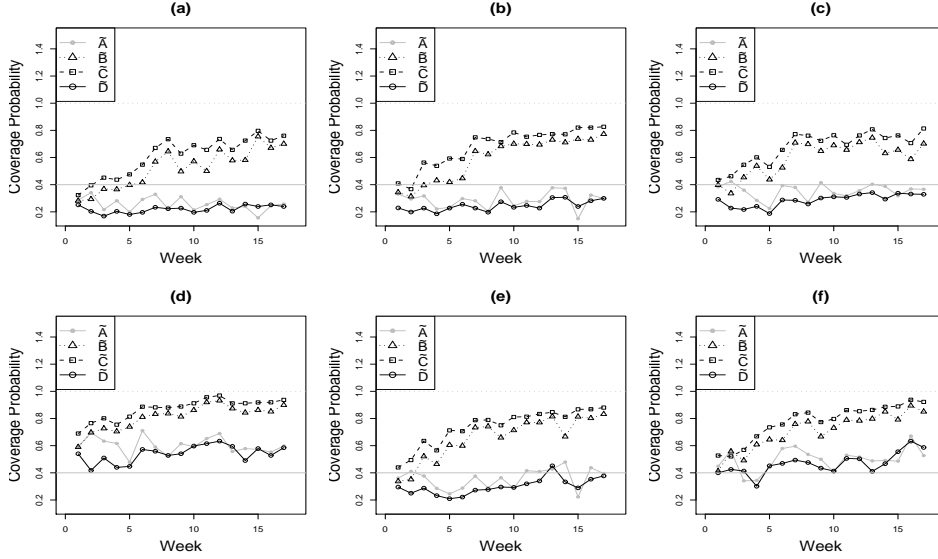


Figure 22: Coverage probability versus 40% nominal level for ungauged sites: (a) – site A; (b) – site B; (c) – site C; (d) – site D; (e) – site E; and (f) – site F. These coverage probabilities are computed for *Study A*: weekly data (solid bullet with solid line); *Study B*: $W_{1:17}$ (up-triangle with dotted line); *Study C*: $W_{1:17}$ but with fixed λ^* (square with dashed line); and *Study D*: $W_{1:17}$ but with fixed λ^* and modified τ_y^2 , τ_1^2 and τ_2^2 (empty circle with solid line).

(iii)

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

respectively. It is straightforward to obtain that $\text{Var}(y_{01}|y_{11})$ is increasing when d_{01} increases, or λ decreases, or σ_ε^2 increases. We next show these properties also hold for $\text{Var}(y_{01}|y_{11}, y_{12})$. By Result 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:

Ungauged Site	A			B			C		
Study	\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

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

(i)

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

(ii)

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

and

(iii)

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

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 5.1

The joint posterior distribution for $x_{1:T}$, λ and σ^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 λ and σ^2 are independent with each other.

The joint posterior distribution for λ and σ^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 σ^2 is an inverse gamma distribution with shape parameter α and scale parameter β , then the posterior distribution for σ^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 λ 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 5.2

Theorem 2 Under Models (14)–(15) and initial prior (43), for any $1 \leq t \leq T$, conditional on θ , we have

(i)

$$\begin{aligned} (\mathbf{x}_{t-1} | y_{1:t-1}, \theta) &\sim N[\mathbf{m}_{t-1}, \sigma^2 C_{t-1}] \\ (\mathbf{x}_t | y_{1:t-1}, \theta) &\sim N[\mathbf{a}_t, \sigma^2 R_t] \\ (\mathbf{y}_t | y_{1:t-1}, \theta) &\sim N[\mathbf{f}_t, \sigma^2 Q_t] \\ (\mathbf{x}_t | y_{1:t}, \theta) &\sim N[\mathbf{m}_t, \sigma^2 C_t], \end{aligned}$$

where

$$\begin{aligned}
\mathbf{a}_t &= \mathbf{m}_{t-1} & R_t &= C_{t-1} + W \\
\mathbf{f}_t &= F'_t \mathbf{a}_t & Q_t &= F'_t R_t F_t + V_\lambda \\
\mathbf{e}_t &= \mathbf{y}_t - \mathbf{f}_t & A_t &= R_t F_t Q_t^{-1} \\
\mathbf{m}_t &= \mathbf{a}_t + A_t \mathbf{e}_t & C_t &= R_t - A_t Q_t A'_t.
\end{aligned}$$

(ii) Define $B_t = C_t R_{t+1}^{-1}$. For $0 \leq k \leq T-1$,

$$(\mathbf{x}_{\mathbf{T}-k} | y_{1:T}, \theta) \sim N[\mathbf{a}_{\mathbf{T}}(-k), \sigma^2 R_{\mathbf{T}}(-k)], \quad (48)$$

where

$$\begin{aligned}
\mathbf{a}_{\mathbf{T}}(-k) &= \mathbf{m}_{\mathbf{T}-k} + B_{T-k}[\mathbf{a}_{\mathbf{T}}(-k+1) - \mathbf{a}_{\mathbf{T}-k+1}] \\
R_{\mathbf{T}}(-k) &= C_{T-k} + B_{T-k}[R_{\mathbf{T}}(-k+1) - R_{\mathbf{T}-k+1}]B'_{T-k},
\end{aligned}$$

with $\mathbf{a}_{\mathbf{T}}(0) = \mathbf{m}_{\mathbf{T}}$, $R_{\mathbf{T}}(0) = C_{\mathbf{T}}$, $\mathbf{a}_{\mathbf{T}-k}(1) = \mathbf{a}_{\mathbf{T}-k+1}$, and $R_{\mathbf{T}-k}(1) = R_{\mathbf{T}-k+1}$.

A.4 Results for Section 5.4

The observation equation is given by

$$\mathbf{y}_t = \mathbf{1}'_n \beta_t + S_{1t}(a_1)\alpha_{1t} + S_{2t}(a_2)\alpha_{2t} + \nu_t, \quad \nu_t \sim N[\mathbf{0}, \sigma^2 \Sigma(\lambda)],$$

where $\Sigma(\lambda) = \exp(-V/\lambda)$ and V denotes the distance matrix for the monitoring sites $\mathbf{s}_1, \dots, \mathbf{s}_n$.

Given $\lambda, \sigma^2, \mathbf{x}_t$ (that is, β_t, α_{1t} and α_{2t}) and \mathbf{y}_t , the posterior conditional distribution for the constant phase parameters, a_1 and a_2 , is given by

$$\begin{aligned}
p(a_1, a_2 | \lambda, \sigma^2, \mathbf{x}_t, \mathbf{y}_t) &\propto p(\mathbf{y}_t | \lambda, \sigma^2, \mathbf{x}_t, a_1, a_2) p(a_1, a_2) \\
&\propto p(\mathbf{M}_t | \lambda, \sigma^2, \beta_t, \alpha_{1t}, \alpha_{2t}, a_1, a_2) p(a_1, a_2),
\end{aligned}$$

where $\mathbf{M}_t = \mathbf{y}_t - \mathbf{1}'_n \beta_t - \cos(\frac{\pi t}{12})\alpha_{1t} - \cos(\frac{\pi t}{6})\alpha_{2t}$ for $t = 1, \dots, T$.

We consider the following two cases for the prior of $\mathbf{a} = (a_1, a_2)'$:

- *Case (i)* a standard reference prior: $p(a) \propto 1$;
- *Case (ii)* a bivariate normal prior: $\mathbf{a} \sim N(\mu, \Sigma)$, where $\mu = (\mu_1, \mu_2)'$ and Σ is a 2 by 2 covariance matrix.

For a fixed $t = 1, \dots, T$, in Case (i), let $\mathbf{l}_1 = \sin(\frac{\pi t}{12})\alpha_{1t}$, $\mathbf{l}_2 = \sin(\frac{\pi t}{6})\alpha_{2t}$, $\mathbf{m} = \mathbf{M}_t$ and $S = \sigma^2 \Sigma(\lambda)$. Given $\lambda, \sigma^2, \mathbf{x}_t, \mathbf{y}_t$, the posterior conditional distribution for the phase parameter vector \mathbf{a} is now given by the following expression where for simplicity we have deleted the conditioning values:

$$\begin{aligned}
p(a_1, a_2 | \lambda, \sigma^2, \mathbf{x}_t, \mathbf{y}_t) &\propto p(\mathbf{M}_t | \lambda, \sigma^2, \mathbf{x}_t, a_1, a_2, x_t, \lambda, \sigma^2) p(a_1, a_2) \\
&\propto \exp\{-1/2[\mathbf{M}_t - a_1 \sin(\frac{\pi t}{12})\alpha_{1t} - a_2 \sin(\frac{\pi t}{6})\alpha_{2t}]'\} \\
&\quad \times (\sigma^2 \Sigma(\lambda))^{-1} [\mathbf{M}_t - a_1 \sin(\frac{\pi t}{12})\alpha_{1t} - a_2 \sin(\frac{\pi t}{6})\alpha_{2t}] \\
&= \exp\{-\frac{1}{2}(\mathbf{m} - a_1 \mathbf{l}_1 - a_2 \mathbf{l}_2)' S^{-1} (\mathbf{m} - a_1 \mathbf{l}_1 - a_2 \mathbf{l}_2)\} \\
&\propto \exp\{-\frac{1}{2}(a \Sigma^{-1} a' - 2a \Sigma^{-1} \mu')\} \\
&\propto \exp\{-\frac{1}{2}[\mathbf{a}'(\mathbf{l}_1, \mathbf{l}_2)' S^{-1} (\mathbf{l}_1, \mathbf{l}_2) \mathbf{a} - \mathbf{a}'(\mathbf{l}_1, \mathbf{l}_2)' S^{-1} \mathbf{m} - \mathbf{m} S^{-1} (\mathbf{l}_1, \mathbf{l}_2) \mathbf{a}]\} \\
&\propto \exp\{-\frac{1}{2}(\mathbf{a} - \mu)' \Sigma^{-1} (\mathbf{a} - \mu)\},
\end{aligned}$$

where

$$\Sigma^{-1} = (\mathbf{l}_1, \mathbf{l}_2)' S^{-1} (\mathbf{l}_1, \mathbf{l}_2), \quad (49)$$

$$\mu = \Sigma (\mathbf{l}_1, \mathbf{l}_2)' S^{-1} \mathbf{m}. \quad (50)$$

Note that equation (50) is equivalent to

$$\Sigma^{-1} \mu = (\mathbf{l}_1, \mathbf{l}_2)' S^{-1} \mathbf{m}. \quad (51)$$

More specifically, we obtain the following elements of the mean vector and covariance matrix for the posterior conditional distribution of the phase parameter vector \mathbf{a} :

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}, \quad (52)$$

$$\Delta^{-1} = (\sigma_{11} \sigma_{22} - \sigma_{12}^2)^{-1} \quad (53)$$

$$= (\mathbf{l}_1' S^{-1} \mathbf{l}_1)(\mathbf{l}_2' S^{-1} \mathbf{l}_2) - (\mathbf{l}_1' S^{-1} \mathbf{l}_2)^2 \quad (54)$$

$$\sigma_{11} = \Delta (\mathbf{l}_2' S^{-1} \mathbf{l}_2) \quad (55)$$

$$\sigma_{12} = -\Delta (\mathbf{l}_1' S^{-1} \mathbf{l}_2) \quad (56)$$

$$\sigma_{22} = \Delta (\mathbf{l}_1' S^{-1} \mathbf{l}_1) \quad (57)$$

$$\mu_1 = \sigma_{11} (\mathbf{l}_1' S^{-1} \mathbf{m}) + \sigma_{12} (\mathbf{l}_2' S^{-1} \mathbf{m}) \quad (58)$$

$$\mu_2 = \sigma_{12} (\mathbf{l}_1' S^{-1} \mathbf{m}) + \sigma_{22} (\mathbf{l}_2' S^{-1} \mathbf{m}) \quad (59)$$

Therefore, we have the following results for the conditional posterior distribution of the phase parameter vector \mathbf{a} :

- (i) If the prior for \mathbf{a} is the standard reference prior, that is, $p(a_1, a_2) \propto 1$, we have

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \Big| x_t, y_t, \lambda, \sigma^2 \sim N \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma \right],$$

where μ_1, μ_2 and Σ are given in Equations (52)–(59) or equations (49)–(50).

(ii) If the prior for \mathbf{a} is a bivariate normal distribution with mean vector $\boldsymbol{\mu}^0 = (\mu_1^0, \mu_2^0)'$ and covariance matrix

$$\Sigma^0 = \begin{pmatrix} \sigma_{11}^0 & \sigma_{12}^0 \\ \sigma_{12}^0 & \sigma_{22}^0 \end{pmatrix},$$

we then have

$$\begin{aligned} p(a_1, a_2 | \lambda, \sigma^2, \mathbf{x}_t, \mathbf{y}_t) &\propto \exp\left\{-\frac{1}{2}(a_1 - \mu_1, a_2 - \mu_2)' \Sigma^{-1} (a_1 - \mu_1, a_2 - \mu_2)\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}(a_1 - \mu_1^0, a_2 - \mu_2^0)' \Sigma^{0^{-1}} (a_1 - \mu_1^0, a_2 - \mu_2^0)\right\} \\ &\propto \exp\left\{-\frac{1}{2}(a_1 - \mu_1^*, a_2 - \mu_2^*)' \Sigma^{*-1} (a_1 - \mu_1^*, a_2 - \mu_2^*)\right\}, \end{aligned}$$

where

$$\Sigma^* = (\Sigma^{-1} + \Sigma^{0^{-1}})^{-1} \quad (60)$$

$$\boldsymbol{\mu}^* = \Sigma^* (\Sigma^{-1} \boldsymbol{\mu} + \Sigma^{0^{-1}} \boldsymbol{\mu}^0). \quad (61)$$

From (60) and (61), we have

$$\Sigma^* = \Sigma - \Sigma(\Sigma + \Sigma^0)^{-1} \Sigma = \Sigma^0(\Sigma + \Sigma^0)^{-1} \Sigma, \quad (62)$$

and

$$\boldsymbol{\mu}^* = \Sigma^0(\Sigma + \Sigma^0)^{-1} \boldsymbol{\mu} + \Sigma(\Sigma + \Sigma^0)^{-1} \boldsymbol{\mu}^0. \quad (63)$$

Hence, the posterior conditional distribution for the phase parameters is given by

$$(\mathbf{a} | \lambda, \sigma^2, \mathbf{x}_t, \mathbf{y}_t) \sim N[\boldsymbol{\mu}^*, \Sigma^*],$$

where $\boldsymbol{\mu}^*$ and Σ^* can be referred to equations (60)–(61), or (62)–(63).

A.5 Results for Section 6

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

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

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 α_{1t}^s as follows:

$$(\alpha_{1t}^s | \alpha_{1,t-1}^s, \alpha_{1\mathbf{t}}, \alpha_{1,\mathbf{t}-1}, \mathbf{y}_{\mathbf{t}}, \lambda, \sigma^2) \sim N[\alpha_{1,t-1}^s + \Sigma_{12}^*(\lambda)\Sigma_{22}^*(\lambda)^{-1}(\alpha_{1\mathbf{t}} - \alpha_{1,\mathbf{t}-1}), \sigma^2\tau_1^2(\Sigma_{11}^*(\lambda) - \Sigma_{22}^*(\lambda)^{-1}\Sigma_{21}^*(\lambda))]. \quad (64)$$

Similarly, the conditional posterior distribution for α_{2t}^s is

$$(\alpha_{2t}^s | \alpha_{2,t-1}^s, \alpha_{2\mathbf{t}}, \alpha_{2,\mathbf{t}-1}, \mathbf{y}_{\mathbf{t}}, \lambda, \sigma^2) \sim N[\alpha_{2,t-1}^s + \Sigma_{12}^*(\lambda_2)\Sigma_{22}^*(\lambda_2)^{-1}(\alpha_{2\mathbf{t}} - \alpha_{2,\mathbf{t}-1}), \sigma^2\tau_2^2(\Sigma_{11}^*(\lambda_2) - \Sigma_{22}^*(\lambda_2)^{-1}\Sigma_{21}^*(\lambda_2))]. \quad (65)$$

Using the observation equation as in Model (11), we have the conditional predictive distribution for y_t^s as follows:

$$(y_t^s | \mathbf{y}_{\mathbf{t}}, \alpha_{1t}^s, \alpha_{2t}^s, \alpha_{1\mathbf{t}}, \alpha_{2\mathbf{t}}, \beta_t, \lambda, \sigma^2) \sim N[\beta_t + S_{1t}(a_1)\alpha_{1t}^s + S_{2t}(a_2)\alpha_{2t}^s + \Sigma_{12}^*(\lambda)\Sigma_{22}^*(\lambda)^{-1} \times (\mathbf{y}_{\mathbf{t}} - \mathbf{1}_n\beta_t - S_{1t}(a_1)\alpha_{1\mathbf{t}} - S_{2t}(a_2)\alpha_{2\mathbf{t}}), \sigma^2(\Sigma_{11}^*(\lambda) - \Sigma_{12}^*(\lambda)\Sigma_{22}^*(\lambda)^{-1}\Sigma_{12}^*(\lambda))]. \quad (66)$$

B Computational codes

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

References

- [1] P.J. Brown, N.D. Le, and J.V. Zidek. Multivariate spatial interpolation and exposure to air pollutants. *Can. J. Stat.*, 22(4), 1994.
- [2] J.M. Burke, M.J. Zufall, and H. Ozkaynak. A population exposure model for particulate matter: case study results for pm_{2.5} in philadelphia, pa. *J. Exposure Anal. Environ. Epidemiol.*, 11, 2001.
- [3] C.A. Calder, C.H. Holloman, S.M. Bortnick, W.J. Strauss, and M. Morara. *Relating ambient particulate matter concentration levels to mortality using an exposure simulator*. Dept. Statist., Ohio State University, preprint no. 725 edition, 2003.
- [4] C.K. Carter and R. Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3), 1994.
- [5] D. Gamerman and Lopes H.F. *Markov Chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman & Hall, second edition, 2006.
- [6] A.E. Gelfand, A.M. Schmidt, S. Banerjee, and C.F. Sirmans. Nonstationary multivariate process modelling through spatially varying coregionalization (with discussion). *Test*, 13(2), 2004.
- [7] G. Huerta, B. Sanso, and J.R. Stroud. A spatio-temporal model for mexico city ozone levels. *J. R. Statist. Soc. C*, 53(2), 2004.
- [8] N.D. Le, W.M. Sun, and J.V. Zidek. Bayesian multivariate spatial interpolation with data missing by design. *J. R. Statist. Soc. B*, 59(2), 1997.
- [9] J. Moller. *Spatial Statistics and Computational Methods*. Springer-Verlag, first edition, 2003.
- [10] Ozone. *US EPA's Air Quality Criteria for Ozone and Related Photochemical Oxidants (First External Review Draft)*. <http://cfpub2.epa.gov/ncea/cfm/recorddisplay.cfm?deid=137307>, 2005.
- [11] J.R. Stroud, P. Muller, and B. Sanso. Dynamic models for spatio-temporal data. *J. R. Statist. Soc. B*, 63, 2001.
- [12] M. West and J. Harrison. *Bayesian forecasting and dynamic models*. Springer, second edition, 1997.