Submitted to the Annals of Applied Statistics

MODELING HOURLY OZONE CONCENTRATION FIELDS

By Yiping Dou* Nhu D. Le† and James V. Zidek‡,*

*University of British Columbia** *and British Columbia Cancer Research Agency*†

This paper compares two methods built on a hierarchical Bayesian foundation and designed for modeling hourly ozone concentrations over the eastern United States. One, a dynamic linear state space model (DLM) that has been proposed earlier, lies in a very contemporary setting where two historical paths to temporal process models, the Kalman filter and the dynamic system with random perturbations, converge. The other, which we call the Bayesian spatial predictor (BSP), is a Bayesian alternative to the purely spatial method of kriging. The DLM as a dynamic system model, has parameters that are states of the process which generate the ozone and change with time. More specifically the model includes a time–varying site invariant mean field as well as time–varying coefficients for 24 and 12 hour diurnal cyclic components. The resulting model's great flexibility comes at the cost of complexity, forcing the use of an MCMC approach and very time–consuming computations. Thus the size of the DLM's spatial domain of applicability has to be restricted and the number of monitoring sites that can be treated, limited. The paper's assessment of the DLM reveals other difficulties that point to the need to consider a less flexible competitor, a Bayesian spatial predictor (BSP). The two methods are compared in a variety of ways and overall conclusions given. In particular, the conclusions point to the BSP as the more practical alternative for spatial prediction.

1. Introduction. This paper applies and compares two models for mapping hourly ambient ozone concentration fields over subregions of the United States (US), an application whose importance is described below. It focuses primarily on the recently proposed dynamic linear model (DLM) of Huerta, Sansó, and Stroud (2004), because that model seems to have worked well for mapping the hourly ozone field of Mexico City. The second model, whose development began with Le and Zidek (1992), is an alternative to kriging called the Bayesian spatial predictor (BSP; see also Le and Zidek 2006). It was selected to provide a baseline for assessing the first because it has a proven track record in air pollution modeling. Moreover it like the first, has Bayesian foundations.

This paper has two important companions. The first is a technical report that provides a lot more detail about the DLM (Dou, Le, and Zidek, 2007). The second is the statistical software we developed for implementing the DLM that could be used in other applications. The current version, *GDLM.1.0*, is freely available at `http://enviro.stat.ubc.ca` for various platforms, namely Windows, Unix and Linux and comes with a demo. The document for *GDLM.1.0* has been submitted as a supplementary material to this paper (see "Ozone-GDLM.1.0.pdf").

The paper's application poses a challenging methodological problem since the fine scale auto–dependence structure of short term (e.g. hourly) aggregates of space–time process responses can prove difficult to model realistically. "Correlation leakage" exemplifies the difficulties involved (Zidek et al. 2002). However, these short term aggregates are important. In particular, United States regulatory standards for ozone are stated in terms of metrics computed from those hourly averages because of a large body of scientific evidence suggesting a strong link between them and

‡The research reported in this article was supported by the Natural Science and Engineering Research Council of Canada.

AMS 2000 subject classifications: Primary 62P12, 60K35; secondary 62M30

Keywords and phrases: Dynamic linear model, hierarchical Bayes, ozone, space–time fields, spatial prediction

acute health outcomes (Ozone 2005). That evidence concerns both human health and human welfare, the latter referring to such things as crop yield. The importance of these random fields meant that in formulating National Ambient Air Quality Standards (NAAQS) for ozone to protect human welfare, spatial interpolation had to be used in rural areas to characterize them due to the paucity of monitoring sites there (Ozone 2005). A large part of that ozone field over the United States constitutes the application of central interest in this paper although as we see below, practical limitations of the DLM method forces us to restrict attention to clusters of about ten monitoring sites for the comparisons we make of the DLM and BSP methods.

Ideally interpolated fields should also be used in risk analysis of the effects of ozone on human health, as input into the computer models used there to incorporate indoor sources in the estimation of human exposure to an air pollutant. The latter can reduce the underestimation of the health effects of errors resulting from the use of ambient monitoring measurements to represent exposure (Shaddick et al. 2008). The US Environmental Protection Agency (EPA) developed and used such a program (APEX) to explore the health risk of ozone under various regulatory scenarios (Ozone, 2005), albeit without interpolating the hourly ozone concentration fields. In contrast, the BSP was used by the second and third authors of this paper to interpolate a spatial pollution field for another EPA population exposure model called SHEDS (Burke, Zufall, and Özkaynak, 2001). Calder et al. (2008) also interpolated such values in a simplified version of SHEDS. In summary, hourly ozone ambient concentration fields need to be spatially interpolated, predicted or mapped.

The DLM, described in Section 2, is essentially the same as the models of Huerta et al. (2004) and Stroud, Muller, and Sansó (2001), which in turn are state space models (West and Harrison, 1997). As a dynamic system model, it has parameters that represent states of the process that generate the ozone and change with time. More specifically the model includes a time–varying site invariant mean field as well as time–varying coefficients for 24 and 12 hour diurnal cyclic components. The resulting model's great flexibility comes at the cost of complexity, forcing the use of an MCMC approach and very time–consuming computations. Thus the size of the DLM's spatial domain of applicability has to be restricted and the number of monitoring sites that can be treated, limited. The paper's assessment of the DLM reveals other difficulties that led us to consider a less flexible competitor, the BSP and the paper compares these two methods.

A number of criteria were invoked for that comparison and we now summarize these and the results:

Practical applicability. The lack of available software for the DLM hourly ozone model was addressed by the investigators and suitable code had been developed as noted above. Using that software, we found that in contrast to the BSP, the DLM made unduly large computational demands for our application making it unsuitable for mapping large geographical domains.

Spatial prediction accuracy. The objective of mapping hourly spatial ozone fields, led us to compare the accuracy of their respective spatial predictions. We found the BSP having smaller mean square predictive errors with respect to the outcomes in a test set of sites.

Calibration of predictive credibility intervals. Here both methods proved somewhat deficient based on the performance of nominally 95% intervals. The BSP intervals tended to be overly narrow, those of the DLM overly wide.

Accuracy of temporal forecasts. Both methods can produce short term temporal forecasts and that comparison is made in a companion report (Dou, Le, and Zidek, 2009).

The paper is organized as follows. Section 2 introduces the hourly concentration field data modeled in this paper. These data, made at fixed site monitors and reported in the AQS dataset, lead to our DLM. However, choosing the hyperparameters appropriately proves challenging. In fact, for guidance in making those choices we consider a simpler alternative, the FOPM (first–order polynomial model) that is susceptible to theoretical analysis. That analysis also reveals

both natural as well as surprising properties of a simple but representative case. For instance with the type of model considered there and plausibly therefore, the one proposed by Huerta et al. (2004), the predictive posterior variances for successive time points conditional on all the data must be monotonically increasing, a seemingly undesirable property. Theoretical results and algorithms for the DLM are presented in Section 2, which also gives theoretical results for prediction and interpolation at unmonitored (ungauged) sites from their predictive posterior distributions. Section 3 gives a brief overview of the BSP that Le and Zidek (2006) describe in detail. Briefly it is a multivariate spatial predictor with empirical Bayesian elements. Section 4 implements both the DLM and BSP for the ozone data referred to above. Moveover, it compares the results obtained using these two different approaches. Section 5 describes difficulties posed by the DLM approach as revealed by our assessment. We summarize our findings and draw conclusions in Section 6.

2. The dynamic linear model. Although we believe the methods described in this paper apply quite generally to hourly pollution concentration space–time fields with a strong diurnal cycle, the paper focuses on hourly ozone concentrations (ppb) over part of the United States owing to their particular importance. Moreover our assessment is limited to the summer of 1995 for which data had been provided. In all, 375 irregularly located sites (or "stations") monitor that field. To enable a focused assessment of the DLM approach and to make computations feasible, we consider just one cluster of ten stations (Cluster 2), in close proximity to one another. However, in work not reported here for brevity, two other such clusters led to similar findings. Note that by design Cluster 2 has the same number of stations as the one in Mexico City studied by Huerta et al. (2004).

Initially a small amount of randomly missing data were filled in by the conventional approach of spatial regression method. Then an exploratory data analysis, following that of Huerta et al. (2004) showed like theirs, that a square–root transformation of the data is needed to validate the normality assumption for the DLM residuals. The Bayesian periodogram (Dou et al. 2007) for the transformed data reveals a peak between 1 pm and 3 pm each day with a significant 24–hour cycle. We also found evidence of a weak 12–hour cycle. However no obvious weekly cycles or nightly peaks were seen. Moreover the phase seems more or less constant. Thus in the end, the DLM suggested by our analysis turns out to be the one in Huerta et al. (2004) without the covariates they had available in their work; it has states for both local trends as well as periodicity across sites.

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

$$\mathbf{y}_t = \mathbf{1}_n \beta_t + S_{1t}(a_1)\boldsymbol{\alpha}_{1t} + S_{2t}(a_2)\boldsymbol{\alpha}_{2t} + \boldsymbol{\nu}_t \tag{2.1}$$

$$\beta_t = \beta_{t-1} + w_t \tag{2.2}$$

$$\boldsymbol{\alpha}_{jt} = \boldsymbol{\alpha}_{j,t-1} + \boldsymbol{\omega}_t^{\alpha_j}, \tag{2.3}$$

where $\mathbf{1}$ denotes a column vector of 1s, $\boldsymbol{\nu}_t \sim N[\mathbf{0}, \sigma^2 \mathbf{V}_\lambda]$, $w_t \sim N[0, \sigma^2 \tau_y^2]$, $\boldsymbol{\omega}_t^{\alpha_j} \sim N[\mathbf{0}, \sigma^2 \tau_j^2 \mathbf{V}_{\lambda_j}]$, $\mathbf{V}_\lambda = \exp(-\mathbf{V}/\lambda)$, $\mathbf{V}_{\lambda_j} = \exp(-\mathbf{V}/\lambda_j)$, $j = 1, 2$ and $\boldsymbol{\alpha}_{jt} = (\alpha_{j1t}, \ldots, \alpha_{jnt})' : n \times 1, j = 1, 2$. 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, $S_{jt}(a_j) = \cos(\pi t j/12) + a_j \sin(\pi t j/12)$, $j = 1, 2$. Note that $\mathbf{V} = (v_{ij}) : n \times n$ represents the distance matrix for the gauged sites $\mathbf{s}_1, \ldots, \mathbf{s}_n$, that is, $v_{ij} = ||\mathbf{s}_i - \mathbf{s}_j||$ for $i, j = 1, \ldots, n$, where $||\mathbf{s}_i - \mathbf{s}_j||$ denotes the Euclidean distance (km) between sites $\mathbf{s}_i$ and $\mathbf{s}_j$. Note that this model thus assumes a second order spatial stationarity unlike the BSP, the second method considered in this paper. This can be a serious limitation in some

geographical domains although in the application considered in this paper, we found no evidence of non–stationarity.

Models (2.1)–(2.3) can also be written as

$$\mathbf{y}_t = \mathbf{F}_t'\mathbf{x}_t + \boldsymbol{\nu}_t \quad \boldsymbol{\nu}_t \sim N(\mathbf{0}, \mathbf{U}_t) \tag{2.4}$$

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \boldsymbol{\omega}_t \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}_t), \tag{2.5}$$

where with $\mathbf{I}_n$ denoting the $n\times n$ identity matrix, $\mathbf{x}_t' = (\beta_t, \boldsymbol{\alpha}_{1t}', \boldsymbol{\alpha}_{2t}')$, $\mathbf{F}_t' = [\mathbf{1}_n, S_{1t}(a_1)\mathbf{I}_n, S_{2t}(a_2)\mathbf{I}_n]$: $n\times(2n+1)$, $\mathbf{U}_t = \sigma^2\mathbf{V}_\lambda$ and $\mathbf{W}_t = \sigma^2\mathbf{W}$, $\mathbf{W}$ being a block diagonal matrix with diagonal entries τ_y^2, $\tau_1^2\exp(-\mathbf{V}/\lambda_1)$ and $\tau_2^2\exp(-\mathbf{V}/\lambda_2)$.

Let $\mathbf{y}_{1:T} = (\mathbf{y}_{1:T}^m, \mathbf{y}_{1:T}^o)'$, where $\mathbf{y}_{1:T}^m = (\mathbf{y}_1^m, \ldots, \mathbf{y}_T^m)$ represents all the missing values and $\mathbf{y}_{1:T}^o$, all the observed values in Cluster 2 sites for $t = 1, \ldots, T$. The model unknowns are therefore the coordinates of the vector $(\lambda, \sigma^2, \mathbf{x}_{1:T}, \mathbf{y}_{1:T}^m, a_1, a_2)$, in which the vector of state parameters up to time T is $\mathbf{x}_{1:T} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$, the range parameter is λ, the variance parameter is σ^2 and finally the vector of amplitude–phase parameters is $\mathbf{a} = (a_1, a_2)$. Let $\boldsymbol{\gamma} = (\tau_y^2, \tau_1^2, \lambda_1, \tau_2^2, \lambda_2)$ be the vector of parameters fixed in the DLM to achieve computational feasibility.

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

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

Notice that λ and σ^2 have inverse Gamma distributions for computational convenience.[1] Section 4 discusses the choice of the hyperpriors in the context of our application.

2.1. *Parameter specification.* Before turning to the implementation of the DLM in the next section, we explore theoretically, albeit in a more tractable special case, some analytical features of the model. That exploration leads to insight about how the model's parameters should be specified, as well as undesirable consequences of inappropriate choices. Our assessment focuses on the accuracy of the model's predictions.

This simple model we consider is a special case of the so–called "first–order polynomial model (FOPM)", a commonly used model (West and Harrison 1997). It captures many important features and properties of the DLM we have adopted.

Letting 0 label an ungauged site, we assume for $i = 0, \ldots, n$ and $t = 1, \ldots, T$, the FOPM model given by

$$y_{it} = \beta_t + \varepsilon_{it} \tag{2.6}$$

$$\beta_t = \beta_{t-1} + \delta_t, \tag{2.7}$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{0t}, \ldots, \varepsilon_{nt})' \sim N(\mathbf{0}, \sigma_\varepsilon^2\exp(-\mathbf{V}/\lambda))$ 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 known here.

The FOPM is particularly useful for short–term prediction since then the underlying evolution β_t is roughly constant. Observe that the zero–mean evolution error δ_t process is independent over time, so that the underlying process is a random walk. At any fixed time t :

$$\beta_t = \beta_0 + \sum_{k=1}^{t} \delta_k \tag{2.8}$$

$$y_{it} = \beta_0 + \sum_{k=1}^{t} \delta_k + \varepsilon_{it}. \tag{2.9}$$

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

Consequently, the FOPM has the following covariance structure:

$$\text{Var}(y_{it}) = \sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2 \tag{2.10}$$

$$\text{Cov}(y_{it}, y_{jt}) = \sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{ij}/\lambda) \quad (i \neq j) \tag{2.11}$$

$$\text{Cov}(y_{it}, y_{js}) = \sigma_\beta^2 + \min\{t, s\}\sigma_\delta^2 \quad (s \neq t), \tag{2.12}$$

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

This DLM defines a non–stationary spatio–temporal process since for the FOPM to be stationary, the eigenvalues of state transfer matrix, $G = G_t$ in the notation of West and Harrison (1997), must lie inside of the unit circle. But $G_t = 1$, so that conditions fails. Non–stationarity also obtains since $\mathbf{G}_t = \mathbf{I}_{2n+1}$, given all the model parameters in (2.4)–(2.5). The DLM in (2.6)–(2.7) has the important property that its covariance functions in (2.11)–(2.12) depend on the time point $\min\{t, s\}$, not $|t - s|$ and this also renders its non–stationarity.

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

$$\text{Cor}(y_{it}, y_{jt}) = \frac{\sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{ij}/\lambda)}{\sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2} \quad (i \neq j) \tag{2.13}$$

$$\text{Cor}(y_{it}, y_{js}) = \frac{\sigma_\beta^2 + \min\{t, s\}\sigma_\delta^2}{\sqrt{\sigma_\beta^2 + t\sigma_\delta^2 + \sigma_\varepsilon^2}\sqrt{\sigma_\beta^2 + s\sigma_\delta^2 + \sigma_\varepsilon^2}} \quad (s \neq t), \tag{2.14}$$

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

Remarks.

1. The correlations in (2.13) and (2.14) have the following properties when $i \neq j$:

$$\text{Cor}(y_{it}, y_{jt}) > \text{Cor}(y_{it}, y_{js}) \tag{2.15}$$

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

$$\text{Cor}(y_{it}, y_{jt}) - \text{Cor}(y_{it}, y_{js}) \tag{2.16}$$

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

2. By (2.13), $\text{Cor}(y_{it}, y_{jt}) \to 1$ as $t \to \infty$ for $i \neq j$, $i, j \in \{0, \ldots, n\}$. That property seems unreasonable; the degree of association between two fixed monitors should not increase as time goes by. To circumvent this problem we can make some of the model parameters, say σ_δ^2, depend on time. More specifically, (2.13) suggests making $t\sigma_\delta^2 = O(1)$ to stabilize $\text{Cor}(y_{it}, y_{jt})$. Another problem is seen for any two sites in close proximity when $d_{ij} \simeq 0$,

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

a reasonable feature. But for widely separated sites when $d_{ij} \to \infty$,

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

To make this correlation nearly 0, as it should be, we need have $\sigma_\beta^2 + O(1) \ll \sigma_\varepsilon^2$. A sufficient condition for this property to hold is $\sigma_\beta^2 \ll \sigma_\varepsilon^2$ and $t\sigma_\delta^2 = O(1) \ll \sigma_\varepsilon^2$.

In summary, the key result, (2.13), suggests a simple but straightforward way to make the model parameter σ_δ^2 depend on T, namely, to replace it by σ_δ^2/T, an adjustment made necessary

by an artifact of the DLM prior assumptions. Section 6 provides an empirical validation of the benefits of that adjustment, and some of its implications. However from a substantive point of view the adjustment is not sensible — these parameters should not have to be changed just because T is changed.

We now study the behavior of the predictive variances in the FOPM that helps us understand our interpolation results. To that end consider the correlations of responses at an ungauged site $\mathbf{s}_0$ with those at the gauged site $\mathbf{s}_j$, $j \in \{1, \ldots, n\}$, respectively. Note that both (2.15) and (2.16) hold for $i = 0$. The properties of the correlation structure in (2.13)–(2.14), lead us to conjecture that the model's predictive bands will increase monotonically over time as more data become available, in the absence of restrictions on $t\sigma_\delta^2 = O(1)$ as suggested above. Furthermore, even conditioning on all the data, the predictive bands increase over time. In support of these conjectures, we prove that they hold in a simple case where $n = 1$ and $T = 2$ in (2.6)–(2.7) (Dou et al. 2007) but omit the proof here for brevity.

THEOREM 1. *For the FOPM in (2.6)–(2.7) with $n = 1$ and $T = 2$, assume the prior for β_0 is $N(0, \sigma_\beta^2)$. The joint distribution of $\mathbf{y} = (y_{01}, y_{11}, y_{02}, y_{12})'$ is $N(\mathbf{0}, \boldsymbol{\Sigma})$, where*

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

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

$$Var(y_{01}|y_{11}) = \frac{(\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)^2 - (\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{01}/\lambda))^2}{\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2} \tag{2.17}$$

$$Var(y_{02}|y_{12}) = \frac{(\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)^2 - (\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{01}/\lambda))^2}{\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2} \tag{2.18}$$

$$Var(y_{01}|y_{11}, y_{12}) = \frac{M_1}{\Delta} \tag{2.19}$$

$$Var(y_{02}|y_{11}, y_{12}) = \frac{M_2}{\Delta}, \tag{2.20}$$

where

$$\Delta = (\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)(\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2) - (\sigma_\beta^2 + \sigma_\delta^2)^2 \tag{2.21}$$

$$M_1 = (\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)\{(\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)^2 - (\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{01}/\lambda))^2\}$$
$$-2(\sigma_\beta^2 + \sigma_\delta^2)^2(\sigma_\varepsilon^2 - \sigma_\varepsilon^2 \exp(-d_{01}/\lambda)) \tag{2.22}$$

and

$$M_2 = (\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)\{(\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)^2 - (\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2 \exp(-d_{01}/\lambda))^2\}$$
$$-2(\sigma_\beta^2 + \sigma_\delta^2)^2(\sigma_\varepsilon^2 - \sigma_\varepsilon^2 \exp(-d_{01}/\lambda)). \tag{2.23}$$

The above results yield in particular, two inequalities about the predictive variance of y_{01} that can also be obtained directly by elementary reasoning. They show in agreement with intuition, that uncertainty about y_{01} based on more data collected over time is no greater than that based on less:

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

and

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

We would also expect that conditional on the same data, the predictive variances of y_{01} and y_{02}, for example $\text{Var}(y_{01}|y_{11}, y_{12})$ and $\text{Var}(y_{02}|y_{11}, y_{12})$, would be more or less equal. Yet the following theorem shows that is not the case.

THEOREM 2. *For the FOPM in Theorem 1, we have the following properties of the predictive conditional variances:*

$$Var(y_{02}|y_{11}, y_{12}) - Var(y_{01}|y_{11}, y_{12}) = \frac{\sigma_\varepsilon^4 \sigma_\delta^2 (1 - \exp(-d_{01}/\lambda))^2}{\Delta} \geq 0 \tag{2.24}$$

$$Var(y_{02}|y_{12}) - Var(y_{01}|y_{11}) = \frac{\sigma_\varepsilon^4 \sigma_\delta^2 (1 - \exp(-d_{01}/\lambda))^2}{(\sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2)(\sigma_\beta^2 + 2\sigma_\delta^2 + \sigma_\varepsilon^2)} \geq 0 \tag{2.25}$$

$$Var(y_{01}|y_{11}) - Var(y_{01}|y_{11}, y_{12}) \geq Var(y_{02}|y_{12}) - Var(y_{02}|y_{11}, y_{12}). \tag{2.26}$$

Equation (2.24) tells us that the predictive posterior variances conditional on all the data increase monotonically at successive time points. That counterintuitive result leads to monotonically increasing coverage probabilities at ungauged sites. According to (2.25) a similar result holds for the predictive variance at a given time conditional on just the monitoring gauged site data available at that time. As noted above, adding data collected over time reduces predictive uncertainty. However according to (2.26) that benefit decreases over time, another unintuitive result. These surprising properties are discussed in Section 5 in relationship to the DLM model of actual interest in this paper and our empirical findings.

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

COROLLARY 1. *The predictive conditional variances in (2.17)–(2.23) increase as d_{01} increases, or σ_ε^2 increases, or λ decreases.*

Thus, keeping two parameters fixed, these predictive conditional variances are monotone functions of the remaining one. Therefore the DLM can paradoxically lead to larger predictive variances when conditioning on more data. Consider for example, the case of just one gauged site, $n = 1$, and two time points $T = 2$. A second statistician arrives on the scene late at time $T = 2$ believing the process was initiated at time $T = 1$. He has only the data available at that time to use in predicting y_{02}. Being in the same relative position as the first statistician at time $T = 1$, he computes his conditional predictive variance and gets a result identical to the one in Equation (2.17). We denote it by $\text{Var}^*(y_{02}|y_{12})$ to distinguish it from the one in Equation (2.18), which the first statistician would compute at time $T = 2$, if he had not observed y_{01}. Surprisingly under the condition in the next corollary, the late–comer's variance based on just y_{12} is actually less than that of the first statistician using all the data available at time $T = 2$, i.e., $\text{Var}(y_{02}|y_{11}, y_{12})$ in (2.20). This result is stated more precisely in the next corollary.

COROLLARY 2. *For the FOPM in Theorem 1,*

$$Var^*(y_{02}|y_{12}) < Var(y_{02}|y_{11}, y_{12}) \quad \textit{if and only if} \quad \sigma_\varepsilon^2 > \sigma_\beta^2 \left(1 + \frac{\sigma_\beta^2}{\sigma_\delta^2}\right). \tag{2.27}$$

The behavior suggested by Corollary 2 is actually observed in our application (see Section 5).

2.2. *Implementation..* We now briefly describe how to implement our model using the MCMC method, more specifically, the forward–filtering–backward–sampling algorithm of Carter and Kohn (1994). Our approach follows Huerta et al. (2004) and the details can be found in Dou et al. (2007). However unlike them, we use all the samples after the burn–in period, not just the chain containing the accepted samples, thereby avoiding the (small) sampling bias that would

otherwise accrue from, in effect, changing the detailed balance equation of the Metropolis–Hasting algorithm.

The algorithm we use for Cluster 2 based on the AQS dataset is now summarized using '$\sim$' to mean 'from distribution':

1. Initially sample $\lambda^{(1)} \sim IG(\alpha_\lambda, \beta_\lambda)$, $\sigma^{2(1)} \sim IG(\alpha_{\sigma^2}, \beta_{\sigma^2})$ and $\mathbf{x}_{1:T}^{(1)} \sim N(\mathbf{m}_0, \sigma^{2(1)}\mathbf{C}_0)$.
2. Given the $(j-1)^{\text{th}}$ values $\lambda^{(j-1)}$, $\sigma^{2(j-1)}$, $\mathbf{x}_{1:T}^{(j-1)}$, $\mathbf{y}^m_{1:T}{}^{(j-1)}$, $a_1^{(j-1)}$, $a_2^{(j-1)}$ and the observations $\mathbf{y}^o_{1:T}$ sample:
   (1) $(\lambda^{(j)}, \sigma^{2(j)}, \mathbf{x}_{1:T}^{(j)}) \sim p(\lambda, \sigma^2, \mathbf{x}_{1:T} | a_1^{(j-1)}, a_2^{(j-1)}, \mathbf{y}_{1:T}^{(j-1)})$ where $\mathbf{y}_{1:T}^{(j-1)} = (\mathbf{y}^m_{1:T}{}^{(j-1)}, \mathbf{y}^o_{1:T})$.
      (i) Generate a candidate value λ^* from a lognormal 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

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

      Accept the candidate value with probability $\alpha(\lambda^{(j-1)}, \lambda^*)$ and set $\lambda^{(j)} = \lambda^*$; otherwise reject and set $\lambda^{(j)} = \lambda^{(j-1)}$.
      (ii) Sample $\sigma^{2(j)} \sim p(\sigma^2 | \lambda^{(j)}, a_1^{(j-1)}, a_2^{(j-1)}, \mathbf{y}_{1:T}^{(j-1)})$.
      (iii) Sample $\mathbf{x}_{1:T}^{(j)} \sim p(\mathbf{x}_{1:T} | \lambda^{(j)}, \sigma^{2(j)}, a_1^{(j-1)}, a_2^{(j-1)}, \mathbf{y}_{1:T}^{(j-1)})$.
   (2) $\mathbf{y}^m_{1:T}{}^{(j)} \sim p(\mathbf{y}^m_{1:T} | \lambda^{(j)}, \sigma^{2(j)}, \mathbf{x}_{1:T}^{(j)}, a_1^{(j-1)}, a_2^{(j-1)}, \mathbf{y}^o_{1:T})$.
   (3) $(a_1^{(j)}, a_2^{(j)}) \sim p(a_1, a_2 | \lambda^{(j)}, \sigma^{2(j)}, \mathbf{x}_{1:T}^{(j)}, \mathbf{y}_{1:T}^{(j)})$, where $\mathbf{y}_{1:T}^{(j)} = (\mathbf{y}^m_{1:T}{}^{(j)}, \mathbf{y}^o_{1:T})$.
3. Repeat until convergence.

The software, *GDLM.1.0*, developed to implement the DLM approach, enhances the Metropolis–within–Gibbs algorithm by augmenting the R code with C to speed up the computation.

2.3. *Interpolation and prediction.* This section describes how to interpolate hourly ozone concentrations at ungauged sites using the DLM and simulated Markov chains for the model parameters defined above in this section. Suppose $\mathbf{s}_1, \ldots, \mathbf{s}_u$ are u ungauged sites of interest within the geographical region of Cluster 2 sites and we need samples from

$$p(\mathbf{y}^s_{1:T} | \lambda, \sigma^2, \mathbf{x}_{1:T}, a_1, a_2, \mathbf{y}_{1:T}),$$

where $\mathbf{y}^s_{1:T} = (y^s_1, \ldots, y^s_T) : 1 \times T$ while y^s_t denotes the unobserved square–root of ozone concentrations at the ungauged site $\mathbf{s} \in \{\mathbf{s}_1, \ldots, \mathbf{s}_u\}$ and time $t = 1, \ldots, T$. Let $(\alpha^s_{1t}, \alpha^s_{2t})$ denote the unobserved state parameters at time t and site $\mathbf{s}$. The DLM is given by

$$\mathbf{y}_t^{\text{new}} = \mathbf{1}_{n+1}\beta_t + S_{1t}(a_1)\boldsymbol{\alpha}_{1t}^{\text{new}} + S_{2t}(a_2)\boldsymbol{\alpha}_{2t}^{\text{new}} + \boldsymbol{\nu}_t^{\text{new}}, \tag{2.28}$$

where $\mathbf{y}_t^{\text{new}} = (y^s_t, \mathbf{y}'_t)'$, $\boldsymbol{\alpha}_{jt}^{\text{new}} = (\alpha^s_{jt}, \boldsymbol{\alpha}'_{jt})'$ for $j = 1, 2$, and $\boldsymbol{\nu}_t^{\text{new}} \sim N(\mathbf{0}, \sigma^2 \exp(-\mathbf{V}^{\text{new}}/\lambda))$.

In the following two subsections, we illustrate how to sample the unobserved state parameters $\{(\alpha^s_{1t}, \alpha^s_{2t}) : t = 1, \ldots, T\}$ and demonstrate spatial interpolation at the ungauged site $\mathbf{s}$.

2.4. *Sampling the unobserved state parameters.* We first sample α^s_{jt} given $\alpha^s_{j,t-1}$, $\boldsymbol{\alpha}_{jt}$ and $\boldsymbol{\alpha}_{j,t-1}, j = 1, 2$. From the state equation (2.5) for $\boldsymbol{\alpha}_{jt}^{\text{new}}$, we know that the joint density of α^s_{jt} and $\boldsymbol{\alpha}_{jt}$ is Gaussian, with covariance matrix $\sigma^2\tau_j^2 \exp(-\mathbf{V}^{\text{new}}/\lambda_j)$, where $\mathbf{V}^{\text{new}}$ denotes the distance matrix for the ungauged site and gauged sites. The conditional posterior distribution,

$$p(\alpha^s_{jt} | \alpha^s_{j,t-1}, \lambda, \sigma^2, \beta_t, \boldsymbol{\alpha}_{1t}, \boldsymbol{\alpha}_{2t}, a_1, a_2, \mathbf{y}_{1:T}), \tag{2.29}$$

is derived in Appendix A.2.

2.5. *Spatial interpolation at ungauged sites.* We interpolate the square–root of ozone concentration at an ungauged site by conditioning on all the other parameters and observations at the gauged sites. As above, y_t^s and $\mathbf{y}_t$ are jointly normally distributed as a consequence of the observation equation. The predictive conditional distribution for y_t^s, that is,

$$p(y_t^s|\alpha_{1t}^s, \alpha_{2t}^s, \lambda, \sigma^2, \beta_t, \boldsymbol{\alpha}_{1t}, \boldsymbol{\alpha}_{2t}, a_1, a_2, \mathbf{y}_{1:T}), \tag{2.30}$$

is given in Appendix A.2.

3. The Bayesian spatial predictor. For completeness, this section gives a brief description of the BSP, which predicts random space–time response fields (Le and Zidek 2006). The responses are site–specific and may include covariates. Although the application in this paper involves just ozone, the multivariate version of the BSP is needed. We include site–specific covariates here even though we have none in our application, to facilitate possible future extensions of the method to more complete database and these are included in the response vector, unlike the covariates which are constant across the region.

The BSP involves a number of elements beginning with a transformation of the data using a common transformation across all sites. The goal of data transformation is to find an approximately symmetrical form for data histograms at each site for its data over time.

Next comes an exploratory data analysis and the removal of systematic regional components. These components can be such things as trends, periodicity, autocorrelation and models for covariates that are constant over the region. Thus the same model is fitted over all sites in the region using the data from all the sites and times. Usually enough data are available that the estimated parameters are effectively constants. Hence their removal from the stochastic field model has a completely predictable effect on it.

At time t, let $\mathbf{r}_t : 1 \times (g+u)p$ represent the row vector of residuals so obtained for the g gauged, i.e. monitored sites as well as the u ungauged sites in the region where every site has $p \geq 1$ site–specific responses. Assign the j^{th} coordinate of $\mathbf{r}_t$ a column vector of random effects, $\beta_r^j : l \times 1$. With that vector $\mathbf{z}'_{rt}\boldsymbol{\beta}_{\mathbf{r}}^j$ can reflect the effect on the site–response pair represented by j at time t, of the covariates in the transposed column vector $\mathbf{z}'_{rt} : 1 \times l$ which are not site–specific. Conditional on $\mathbf{z}'_{rt}\boldsymbol{\beta}_{\mathbf{r}}^j$, $\mathbf{r}_t$ is assumed to have an arbitrary covariance matrix $\boldsymbol{\Sigma}_{\mathbf{r}}$, representing both the covariances between the sites as well as the covariances between responses within the sites. Note that at this level of modeling, no assumption is made about the separability of the between–site variability and within–site variability. Moreover the site effect vector $\boldsymbol{\beta}_{\mathbf{r}}^j$ allows each site's trend, seasonality and so on to deviate from their regional counterparts.

Conditional on $\beta_{\mathbf{r}}$ and $\boldsymbol{\Sigma}_{\mathbf{r}}$, we assume the $\{\mathbf{r}_t\}$ have a known between–time covariance $\mathbf{A}$ that is separable from the within–time covariance $\boldsymbol{\Sigma}_{\mathbf{r}}$. Furthermore we assume a Gaussian response field across the T time points and all sites and responses so that $\mathbf{r} \doteq (\mathbf{r}'_1, \ldots, \mathbf{r}'_n)' \sim N_{n\times(g+u)p}(\mathbf{z}\boldsymbol{\beta}_{\mathbf{r}}, \mathbf{A} \otimes \boldsymbol{\Sigma}_{\mathbf{r}})$ for the "design matrix" $\mathbf{z} \doteq (\mathbf{z}_1, \ldots, \mathbf{z}_T)' : T \times l$.

Next $\boldsymbol{\beta}_{\mathbf{r}}$ and $\boldsymbol{\Sigma}_{\mathbf{r}}$ are given their conjugate multi–normal and generalized Inverted Wishart prior distributions, respectively. The latter's hypercovariance structure is assumed to be Kronecker product of covariances reflecting the assumed separability of the between–site and within–site covariances. That assumption is implemented in the online R, C and F codes that implement the BSP methodology (go to http://enviro.stat.ubc.ca). Furthermore, they incorporate an empirical Bayesian step which estimates the hyperparameters in the conjugate priors using the EM algorithm. That software is used in the application addressed in this paper. Note however that the separability assumption can be relaxed in other applications of the method at the cost of some recoding of the software.

The resulting posterior predictive distribution with estimated hyperparameters is a matric–t distribution. That means in particular, that the conditional posterior predictive distribution of the responses of direct interest given the site–specific covariates can be derived and the predictive posterior mean will then be a linear combination of regional and site–specific covariates.

As a final step in applying the BSP, the regional models that were removed to get the residuals matrix $\mathbf{r}$ need to be combined with predicted residuals at ungauged sites. These in turn must be squared to get back onto the scale of the raw data.

4. Application. This section applies our models to the hourly ozone concentration field mentioned briefly above.

4.1. *The data.* That ozone field generates the data used in this paper. Those data come from the AQS ozone database created by the EPA. As noted in Section 2, the heavy computational requirements of the DLM force the restriction of our analysis to hourly ground–level ozone concentrations (in ppb) from a cluster of just ten monitoring sites (we call "Cluster 2") and data for that cluster were extracted from that database. That cluster, centered at St Louis, Missouri spans a distance of at most 895 kms. In contrast, that of Huerta et al. (2004) centered in Mexico City, covers no more than 30 kms in any one direction although it has the same number of sites as our cluster.

Data from an additional set of six monitoring sites located at randomly selected and irregularly spaced geographical locations within this area, were extracted to provide a validation sample.

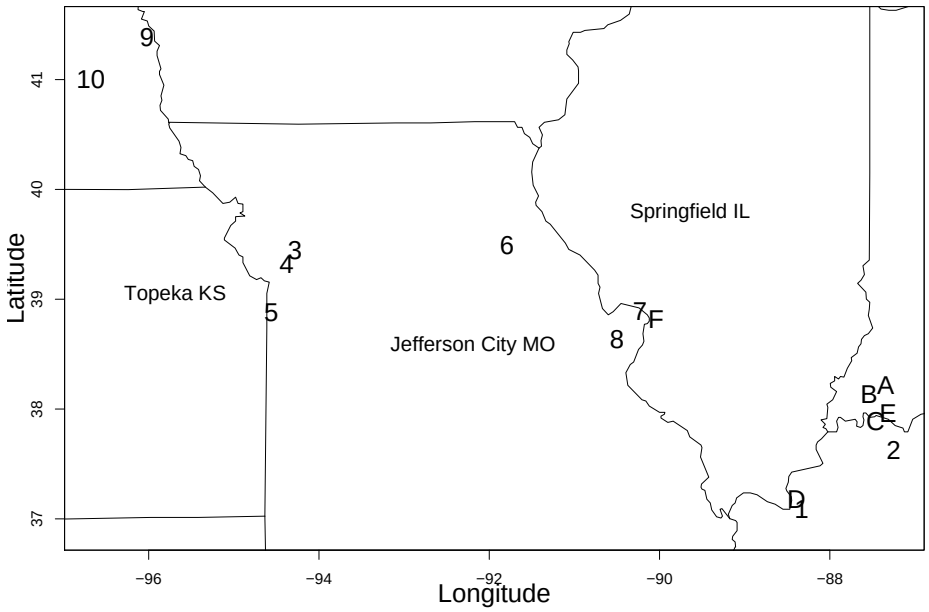

Fig 1: Geographical locations for Cluster 2 sites from the AQS database (1995), where the latitude and longitude are measured in degrees. Integers label gauged sites and letters, ungauged sites.

Figure 1 shows the geographical locations of these ten gauged and six ungauged sites. The percentages of the missing measurements vary between 0% and 24.8% for gauged sites, and between 0% and 11.5% for ungauged sites.

Figure 2 depicts the boxplots of the square–root transformed hourly ozone concentrations at each one of the 16 monitoring stations across all time points. Gauged Site (GS) 9 differs markedly from the others. The authors wondered if its deviation from the rest might be due to some of its geographical features. An examination of the region reveals GS 9's proximity to a body of surface water, the Missouri River. However, GS 7 also lies near water; in fact, it falls between the junction of the Illinois, Mississippi and Missouri Rivers. Similarly GS 1 lies close to Kentucky Lake while GS 6 lies near Mark Twain Lake. Thus in the end we were forced to abandon that potential explanation and have found no other.

We used linear regression to further explore the data, in particular, weekday and hourly effects. We found these to be approximately constant over all gauged sites; the hourly effects from 0 A.M.

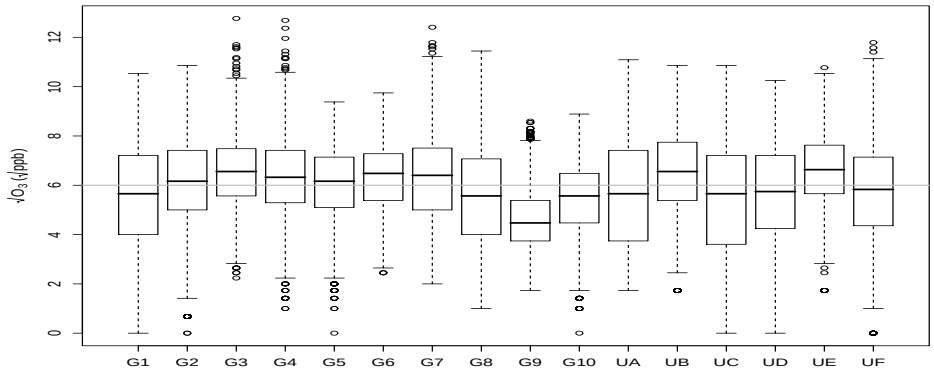

Fig 2: Boxplots for the square–root of hourly ozone concentrations ($\sqrt{ppb}$) at monitoring stations in Cluster 2 from the AQS database during the summer of 1995. Here 'G' stands for 'gauged site' while 'U' stands for 'ungauged site'.

to 10 A.M. vary slightly more than those of the hours following 10 AM, pointing to the relatively strong hourly effects from 10 A.M. to 11 P.M. The weekday effects also seemed constant across GSs. These results suggest that we can model weekday and hourly effects as constant across all gauged sites in the cluster. We used this finding to develop the BSP approach used below. It also has implications for the DLM method, although for brevity we leave details to Dou et al. (2007).

4.2. *The methods.* The issue central addressed in this paper, the need to map the ozone field or in other words, interpolate its values at ungauged sites, leads us to compare the two models we have proposed for that purpose. More precisely we compare the interpolated values they yield of hourly ozone concentrations at the six validation sites. Results for other clusters were generally similar and for one, results are presented in a companion report (Dou et al. 2009).

The DLM. To emulate Huerta et al. (2004) to the maximum feasible extent, we use their initial settings for the starting values, hyperpriors and fixed model parameters but only after confirming that our results would not be unduly sensitive to that choice. In summary we chose:

- 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 $\mathbf{x}_0$, the initial state parameter, is assumed to be normally distributed with mean vector $\boldsymbol{m_0} = (2.85, -0.75\mathbf{1}'_n, -0.08\mathbf{1}'_n)'$ and covariance matrix $\sigma_1^2\mathbf{C}_0$, where $\sigma_1^2 \sim IG(2, 0.01)$ and $\mathbf{C}_0$ 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 $\boldsymbol{\mu}^o = (2.5, 9.8)'$ and a diagonal matrix $\boldsymbol{\Sigma}^o$ with diagonal entries 0.5 and 0.5.
- Some of the model parameters in the DLM are fixed as follows: $\tau_y^2 = 0.02$, $\tau_1^2 = 0.0002$, $\tau_2^2 = 0.0004$, $\lambda_1 = 25$ and $\lambda_2 = 25$.

To test sensitivity of our results to our starting values we experimented with a variety of such values of λ, σ^2, a_1 and a_2. Figure 1 in the supplementary file, "Ozone-MCMC.pdf", compares the results for pairs of two very different values over $4,268$ MCMC iterations. The results demonstrate the adequacy of our burn–in period of $2,269$. We concluded that varying the hyperparameters produced the same results after taking just a few iterations to adjust. In fact, the chains converge in less than five hundreds iterations with an acceptance rate was approximately

62%. The post burn–in samples were the ones used to estimate the posterior distribution that in all cases except for λ resembled closely those obtained by Huerta et al. (2004). In contrast that for λ was centered on a point about ten times larger than theirs, possibly reflecting the much larger spatial domain in our application.

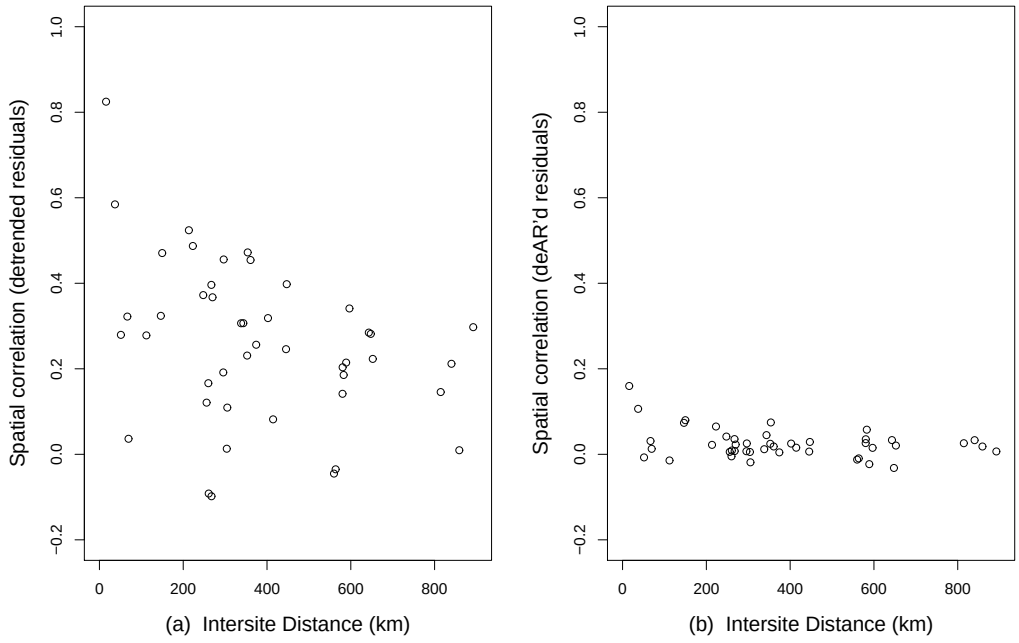

Fig 3: This figure illustrates the correlation leakage problem for intersite spatial correlations for the sites in our analysis. Panel (a) shows those correlations for all site pairs and the residuals obtained by removing regional trends. The second, (b) is the same plot, but this time for the residuals after an additional regional AR(2) model was fitted. Notice the sharp declines in the spatial correlations as a result of "leakage" into spatial cross–correlations at lags 1, 2 and so on.

The BSP. To begin, Figure 3 depicts estimated spatial correlations of the residuals after removing regional (that is common site–model) trends and AR(2) autocorrelation components (dAR'ing). This reveals "correlation leakage", a sharp drop in the lag 0 spatial cross correlation with loss of correlation to lag 1 and longer cross correlations. To prefilter data series as we must to use the BSP when leakage occurs, typically in series based on data collected with short temporal lags, Li et al. (1999) and Zidek et al. (2002) suggest using daily vectors of random hourly response coordinates for selected hours, instead of using the univariate hourly responses themselves. One of their arguments, that the impact of the problem of "correlation leakage" across space and time lags, is reduced using this approach is supported by theoretical results (Zidek et al., 2002). A related argument of theirs is that daily vectors of deAR'd residuals are approximately independent from day–to–day while the need to model fine scale temporal correlation is eliminated.

The benefits promised by these arguments obtain for Cluster 2 and other clusters we have studied. To realize them, we must select the number of hours d between 1 to 24 to include in those daily vectors. The simplest choice $d = 1$ would ensure a 23 hour separation between the successive responses that under our AR(2) structure, would render autocorrelation negligible in the daily sequence. However that choice would also leave us without estimates of the correlation between successive hourly responses, leading to $d = 2$ instead. For that would mean in particular that a spatial interpolator for say hour 12 at an ungauged site, would borrow strength from both hours

11 and 12 measurements at the gauged sites. But that choice like d = 1 has the disadvantage of complexity: twelve models would be needed for the resulting 12 parallel multivariate time series. That points to $d > 2$. But then large values of d such as $d = 24$ would also be undesirable, given our objective of eliminating autocorrelation between successive daily vectors of dimension d. A compromise in the range $2 \leq d \leq 6$ suggests itself as a compromise, although the optimal choice clearly depends on the degree of autocorrelation in the AR(2) process and that varies from one location to another.

Although we do not have a theoretical way of selecting d, we do have an empirical alternative in particular applications. In ours, suppose for example that interest focuses on interpolating hour 12's ozone level at a specified ungauged site using all the observed responses at the gauged sites. The possible response vector sequences corresponding to $d \in \{2, \ldots, 6\}$ would be those from (11:12)–hours to (7:12)–hours, where $(i{:}j)$–hours denotes the sub–data matrix of hourly measurements from hours i to j inclusive, across the gauged sites, $i, j \in \{1, \ldots, p\}$. For each, the spatial correlations between all gauged sites are estimated using the multivariate BSP approach.

As expected, the spatial correlation declines and leakage increases as d the dimension of the response vector increases. The smallest value $d = 2$ in the admissible range $\{2, \ldots, 6\}$ produces the smallest loss of such spatial correlation. We see a further drop when $d = 3$. However, little change is seen for ds beyond that. In other words, nothing is gained by going beyond $d = 2$ and we stop there, making it our compromise choice for hour 12. Similar results obtain for the other hours, strongly supporting the use of the 2–consecutive–hour–block as the response vector. These blocks of data were extracted from the AQS ozone database for the summer of 1995 for Cluster 2 to serve as the observed multivariate responses in a multivariate BSP model framework. While the resulting spatial correlation is not large, the very strong autocorrelation between 12 and its neighbor, hour 11, enables a lot of strength to be borrowed over both space and time in the combined space–time total and this is key to the good performance of the BSP predictor.

Prior to implementing the multivariate BSP approach, a small number of missing measurements were filled in by the conventional method of periodic means. In the BSP model's notation, $p = 2$, $n = 120$, $u = 6$ and $g = 10$. In all, 24 multivariate BSP predictors are derived by successively cycling through the successive two–hour–blocks to predict the hourly ozone levels at the 6 validation sites. We also found the corresponding 95% pointwise predictive intervals along with their empirical coverage probabilities.

4.3. *The results.* Figures 4 and 5 depict plots of the interpolation results for square–root transformed ozone concentrations at Ungauged Site (US) D during the first and fourth weeks in the summer of the study. Overall the BSP proves more accurate than the DLM. Moreover, it avoids the unnatural oscillatory behavior in the 95% predictive interval bounds, an artifact of the harmonic terms in the mean model used in the DLM. To elaborate, on this important point, note that conditional on the mean, the predictive distribution for a response at an ungauged site has a fixed width predictive error band [see Equation (A.3)]. However in the unconditional distribution for that response, the mean's random coefficients contribute to the width of that band. First of all they do so through the harmonic, i.e. cosine and sine, functions associated with them. As these functions are squared in the variance for the unconditional distribution, valleys in the mean become peaks in the variance, giving the impression of a six rather than twelve hour cycle as postulated in the model. That same periodic behaviour can been seen in the paper of Huerta et al. (2004 Figure 6). Secondly they do so through their variances which are re–scaled by their squared harmonic factors. The growth in those variances over time, induced by their random walk evolutionary model, is then inflated by the squared harmonics at peak times. Thus as seen in the figures, by the fourth week we see much larger band widths at those peaks than during the first week. The width of those bands varies further according to the site's distance from a monitoring site as that determines the degree uncertainty in those random coefficients.

The large predictive intervals for the ground–level ozone concentrations reflect the inefficiency of the DLM approach. For example, the 95% predictive bands could be between 0 and 15 for the square–root of ozone levels. These findings lend further support to our conclusion that a more practical alternative to the DLM and needed. In turn that led us to consider the BSP as a possible competitor for predicting ground–level ozone concentrations.

Notice the periodic drop toward zero for the BSP in these figures. That results from BSP's recognition that US D's close "cousin" Gauged Site (GS) 1 with which it is spatially correlated, contains a missing value coded as a '0' in the data at this particular hour. That hour is midnight when we believe, but cannot confirm if quality checking was carried out nightly during the 120 days of study. That also happens at US D and other sites such as US B and GS 2 on 118 out of the 120 days. We imputed the missing observations at midnight by computing its periodic means, leading to the"midnight zeros". For example, the observations at GS 1 and US D are all missing while GS 2 and US C only have one or two observed values at that time. In fact the DLM needs the imputed values for complete data and "updates" the missing values by treating them like the model parameters. However even using our roughly imputed missing values as the periodic means, our results indicate that the BSP works better over those sites.

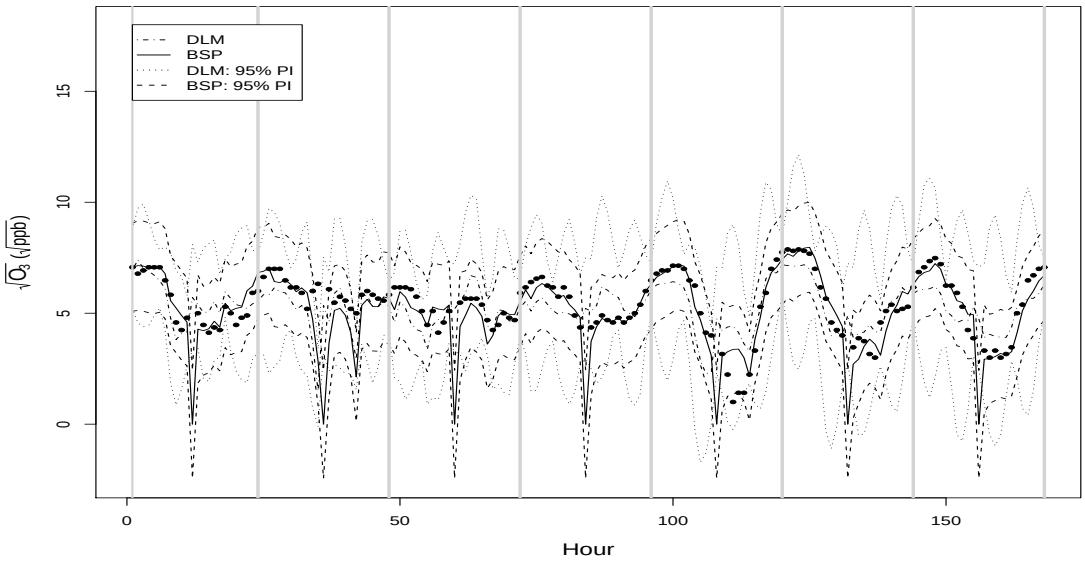

Fig 4: Interpolation at Ungauged Site D for the 1st week. The square–root of hourly ozone concentrations are plotted on the vertical axes and hours on the horizontal axes. Solid and dashed lines represent respectively, BSP interpolation and 95% pointwise predictive intervals; dotdashed and dot lines represent respectively, DLM interpolation and 95% predictive intervals; finally •s represents observations at Ungauged Site D.

Figure 6 plots the ratio of the empirical mean square predictive errors of the DLM over the BSP arising from the prediction of the observations for the six validation site values. Thus values larger than 1.0 mean DLM is less accurate than BSP. Overall the BSP performs uniformly better than the DLM while requiring much less computation time.

Figures 7–8 present the coverage probabilities of the DLM and BSP at 95% nominal level at each one of ungauged sites or 17 weeks involved in the ozone study. Overall the DLM intervals tend to exceed the nominal level of 95% while the BSP intervals cover less than the nominal level, indicating respectively too little and too much confidence in their predicted values.

The coverage probabilities of the model's posterior predictive credibility intervals over successive weeks, conditional on all 17 weeks of data, increase monotonically. That implies the counterintuitive result that uncertainty increases as time evolves. A pragmatic way around this

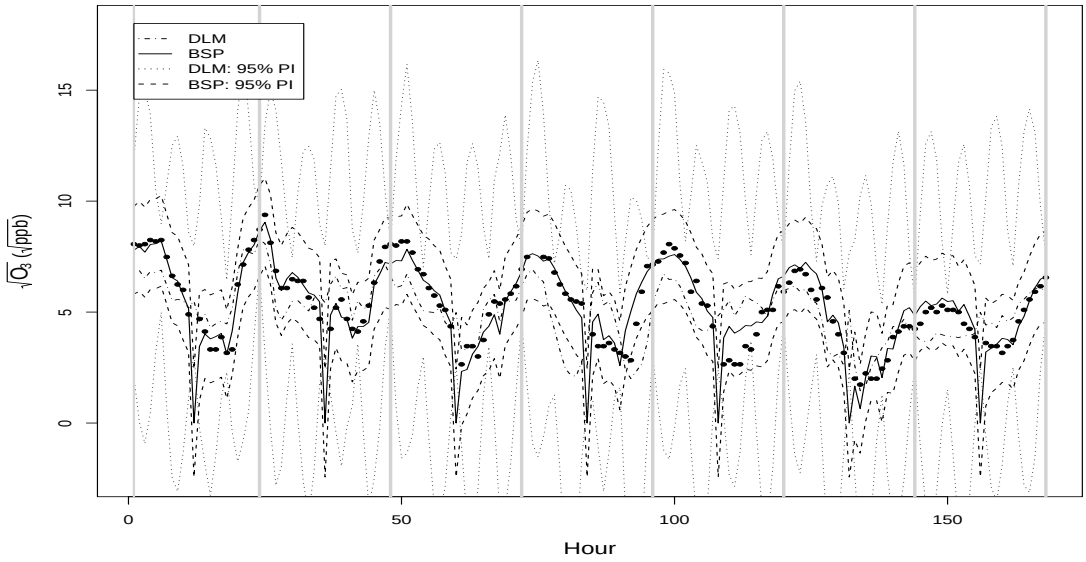

Fig 5: Interpolation at Ungauged Site D for the 4th week. The square–root of hourly ozone concentrations are plotted on the vertical axes and hours on the horizontal axes. Solid and dashed lines represent respectively, BSP interpolation and 95% pointwise predictive intervals; dotdashed and dot lines represent respectively, DLM interpolation and 95% predictive intervals; finally •s represents observations at Ungauged Site D.

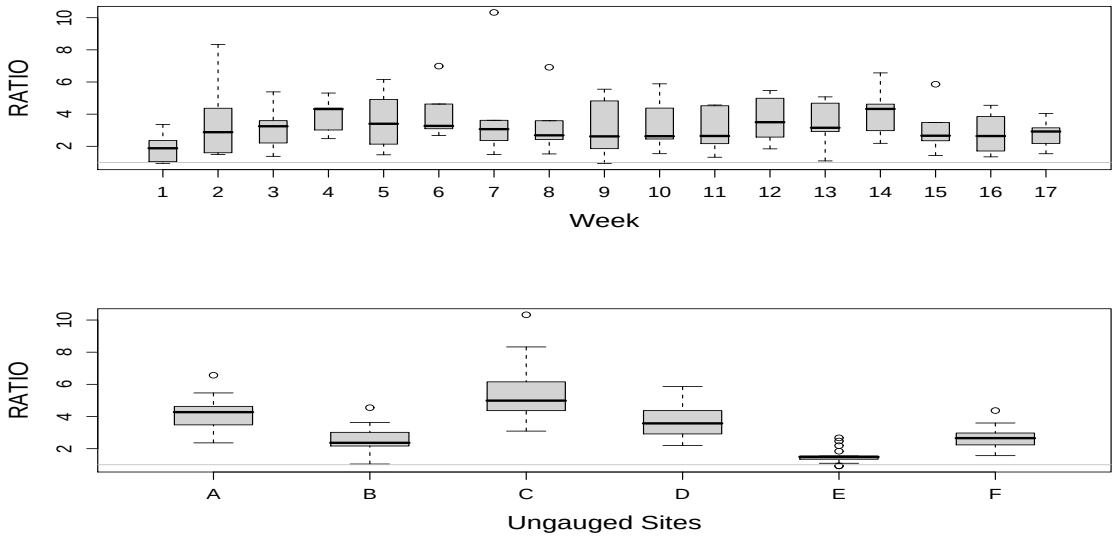

Fig 6: The ratio of the empirical mean squared prediction errors for the DLM over the multivariate BSP interpolators for: (a) the 6 "ungauged" validation sites for each of the 17 weeks; (b) the 17 weeks for each of those 6 sites.

unnatural property comes from Section 2.1 where the correlation structure of an analytically tractable DLM is studied. The section suggests making the model parameters, such as τ_y^2, τ_1^2 and τ_2^2 depend on the time span of the temporal domain T involved. In other words, the hyper-parameters would differ for a study involving one week to those for a study involving the whole summer, quite unreasonable if one views the prior as representing prior knowledge.

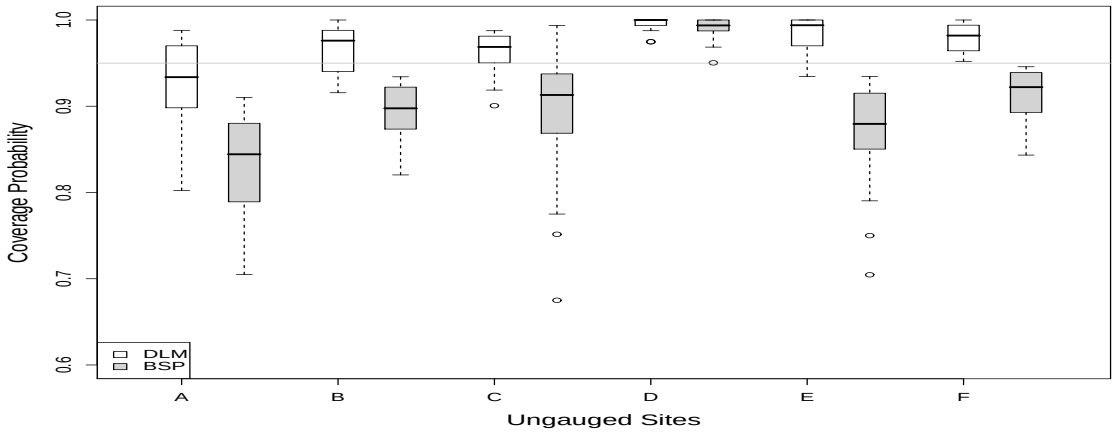

Fig 7: The empirical coverage probabilities of the DLM and multivariate BSP interpolators for each of six ungauged sites at a nominal level of 95%.

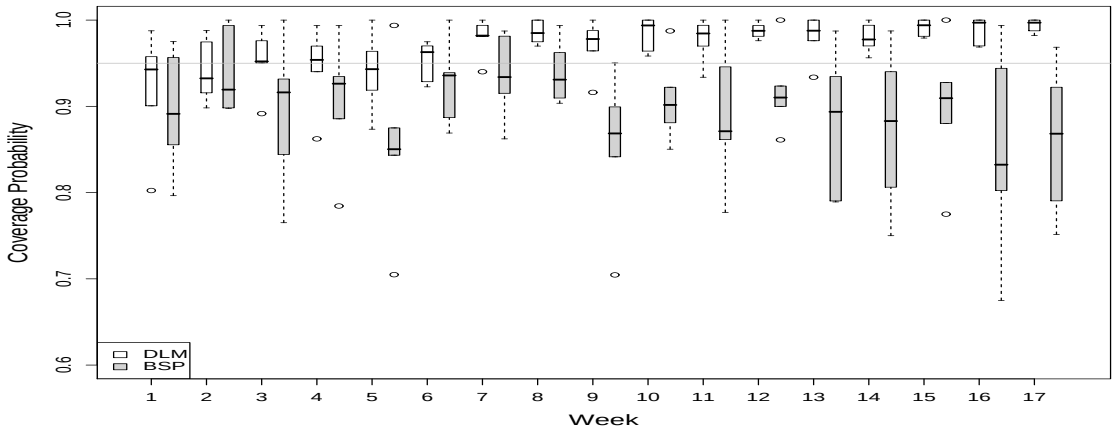

Fig 8: The coverage probabilities of for the DLM and multivariate BSP interpolators for each of 17 weeks at a nominal level of 95%.

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

MCMC convergence. See the supplementary file, "Ozone–MCMC.pdf".

Relationships among parameters. Our prior assumptions make the model parameters λ, σ^2, a_1 and a_2 uncorrelated leading us to investigate that relationship a posteriori. In fact our results, which are omitted for brevity, show all pairs remain uncorrelated except for λ and σ^2, which have a strong linear association. The third author learned about that feature of the DLM from Jonathon Stroud (personal communication). Since σ^2 determines spatial variability while λ determines correlation, this relationship is intriguing. Larger values of σ^2 tend to go with larger λs, i.e. diminished spatial correlation. In kriging type models, it is quite common to have range and sill parameters correlated in the posterior, according to an anonymous reviewer.

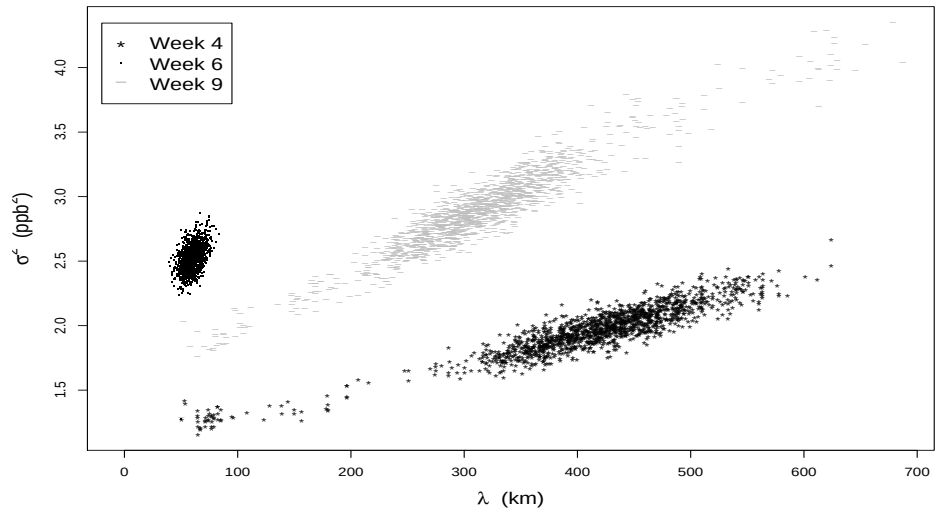

Fig 9: Scatterplots for (λ, σ^2) pairs for various weeks, based on the MCMC samples using a selected week's data, specifically weeks 4, 6 and 9 but starting from the same initial values as those in Section 4.2.

Time varying parameters and coverage 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.

To address these concerns we carried out the following studies:

(i) Study $\tilde{A}$: Implement the DLM at ungauged sites using only weekly data for successive weeks $(W_k : k = 1, \ldots, 17)$. Generate Markov chains for λ, σ^2, a_1 and a_2. Estimate model parameters and interpolate the results at the ungauged sites. Obtain the coverage probabilities at each ungauged site and week for fixed credible interval probabilities using each week's data.

(ii) Study $\tilde{B}$: Implement the DLM at ungauged sites using all the data from weeks 1 to 17 $(\mathbf{W}_{1:17} = \{W_1, \ldots, W_{17}\})$. Estimate model parameters and interpolate the results at ungauged sites. Obtain the coverage probabilities at each ungauged site and week for fixed credible interval probabilities for each week.

(iii) Study $\tilde{C}$: Fix λ_k^* at week k $(k = 1, \ldots, 17)$ using values suggested by the Markov chains generated in Study $\tilde{A}$. Then use these $\boldsymbol{\lambda}^* = \{\lambda_1^*, \ldots, \lambda_{17}^*\}$ as fixed values in the DLM to reduce computation time. In other words, go through all the steps in the algorithm of Section 2.2 but now using only fixed $\boldsymbol{\lambda}^*$s instead of generating them by a Metropolis–Hasting step. Note that we are then only using Gibbs sampling and an MCMC blocking scheme. Compute the corresponding coverage probabilities using $\mathbf{W}_{1:17}$ at each ungauged site and week for fixed credible interval probabilities.

Studies $\tilde{A}$ and $\tilde{B}$ are intended to explore the effect on the interpolation results, of varying the amount of data and its collection time. Study $\tilde{C}$ aims to pick out any significant difference in the interpolation results when using a fixed $\boldsymbol{\lambda}^*$ rather than using the Markov samples of λs. It is also aimed at finding how much time would be saved by avoiding the inefficient Metropolis step. Table 1 on page 18 shows these fixed $\boldsymbol{\lambda}^*$s used in Study $\tilde{C}$. Table 2 on page 18 shows the time saved using a fixed $\boldsymbol{\lambda}^*$s against the one using the Metropolis–Hastings algorithm.

TABLE 1
Fixed values of $\boldsymbol{\lambda}^*$ in Study $\tilde{C}$.

Week	1	2	3	4	5	6	7	8	9
$\boldsymbol{\lambda}^*$	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	
$\boldsymbol{\lambda}^*$	142.2	172.7	187.9	315.8	419.0	99.8	260.3	284.8	

TABLE 2
Summary of computational times (seconds) for complete summer long MCMC runs without spatial prediction in Studies $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$.

				Time (seconds)	
Study	Data	Iteration total	Accept(%)	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

Figure 9 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, implying that the constant λ–σ^2 model is not tenable over a whole summer for this dataset and should not be assumed in general without empirical validation.

Figure 10 typifies figures in Dou et al. (2007) showing the coverage probabilities for various predictive intervals associated with the interpolators in these three studies. The solid line with bullets represents the results for Study $\tilde{A}$, the dotted line with up–triangles for $\tilde{B}$ and the dashed line with squares for $\tilde{C}$. These graphs show that the coverage probabilities of Study $\tilde{B}$ are similar to those of $\tilde{C}$. This suggests that we could use the entries in Table 1 on page 18 as fixed $\boldsymbol{\lambda}^*$s in the DLM to obtain interpolation results similar to those obtained using the Metropolis–within–Gibbs algorithm.

Section 2.1 presents results about the prediction accuracy of the simplest DLM, namely, the FOPM showing in particular, that the predictive variances must increase monotonically at successive time points even though all the variances are conditional on the same 17 weeks of data. Here we see the same phenomenon expressed empirically through the graph of the coverage probabilities in Figure 10. The plots exhibit a monotonic increasing trend in the coverage probabilities of both Studies $\tilde{B}$ and $\tilde{C}$ even though the uncertainty in each case is calculated for distributions that are conditional on all the data. Note that those coverage probabilities for both studies deviate slightly from a strictly monotone trend at some time points because of the time varying relationship between λ and σ^2 seen in Figure 9. This increase can mean that a posteriori, λ and σ^2s vary over the time span of the study, while the prior postulates that they do not. It may also be due to mis–specification of the model parameter values $\gamma = (\tau_y^2, \tau_1^2, \lambda_1, \tau_2^2, \lambda_2)$. (See the initial settings for γ in Section 4.2.) In any case the phenomenon represents a limitation in the applicability of the model since it runs counter to intuition.

As an aside, Study $\tilde{C}$ enjoys significant computational advantages over $\tilde{B}$. Table 2 on page 18 shows computation time of the former to be almost 2.8 times shorter than the latter.

On another point, these studies show that sometimes, paradoxically, the model gives better results using only one week's data rather than all of it. In fact, Corollary 2 in Section 2.1 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, $\sigma_\beta^2 \left(1 + \sigma_\beta^2/\sigma_\delta^2\right) \simeq 0.51$, 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 necessary and sufficient condition in Corollary 2, the predictive variance of Study $\tilde{A}$ is less than that of Study $\tilde{B}$ making the predict and more, not less certain. 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

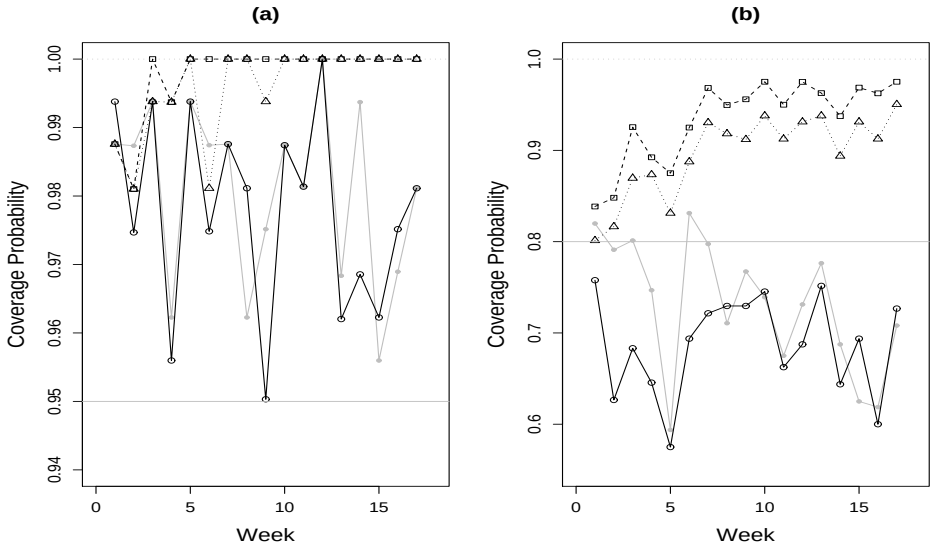

Fig 10: Coverage probabilities over the 17 weeks of summer for: (a) 95% credible intervals at Ungauged Site D; (b) 80% credible intervals at Ungauged Site C. These coverage probabilities are computed for Study $\tilde{A}$: weekly data (solid bullet with solid line); Study $\tilde{B}$: $\mathbf{W}_{1:17}$ (up–triangle with dotted line); Study $\tilde{C}$: $\mathbf{W}_{1:17}$ but with fixed $\boldsymbol{\lambda}^*$ (square with dashed line); and Study $\tilde{D}$: $\mathbf{W}_{1:17}$ but with fixed $\boldsymbol{\lambda}^*$ and modified τ_y^2, τ_1^2 and τ_2^2 (empty circle with solid line).

findings of this section. For example, in panel (b) of Figure 10, the coverage probability for $\tilde{B}$ at the 4^{th} week is larger than that for $\tilde{A}$. From the above discussion, we know that the predictive variance of $\tilde{A}$ should be less than that of $\tilde{B}$. However, σ^2 for $\tilde{A}$ tends to larger than for $\tilde{B}$, leading an inflated predictive variance for $\tilde{A}$. This feature makes it difficult to compare these two predictive variances, but explains the paradox we see in those figures.

6. Concluding remarks. To assess the dynamic linear modeling (DLM) approach to mapping 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 and compared it to an older and computational leaner approach to Bayesian spatial prediction (BSP; Le and Zidek 2006). Practical considerations forced us to focus on small clusters of sites including the one treated in this paper, Cluster 2 during a single ozone summer season. The DLM was the primary focus of the paper since it had already been proposed for modeling hourly ozone fields, albeit over Mexico City (Huerta et al. 2004) implemented through MCMC sampling.

Our assessment reveals some difficulties with this very flexible approach and practical challenges that it presents. We also have made some recommendations for 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 for sampling the model 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.

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). Figure 10 compares Study $\tilde{D}$ with the others. Observe that its coverage probabilities behave like those of Study $\tilde{A}$. This adjustment does seem to eliminate the undesirable property of increasing credibility bands of Studies $\tilde{B}$ and $\tilde{C}$, albeit by an unreasonable modification of the model — and make the prior for these

parameters depend on the time span of the study.

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. We can use six Metropolis–Hasting kernels to sample from the target distribution $\pi(\boldsymbol{\gamma}|\mathbf{y}_{1:T})$, updating each component of $\boldsymbol{\gamma}$ iteratively, where $\boldsymbol{\gamma}$ was defined in Section 2. But, not surprisingly that approach fails because of the extreme computational burden it entails. However, that alternative is the subject of current work along with an approach that admits time varying λs and σ^2s.

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

Finally, given the sophistication of the DLM, the authors were surprised to find that the BSP performed as least as well and in some cases better. This performance was achieved with much lower computational demands, thereby making the BSP suitable for mapping over large geographical domains. So while the BSP was introduced as a "sparring partner" in this paper, it is the only method we can recommend at this time for spatial mapping in the context we have considered. That finding led the authors to enhance the BSP and in particular, to put it into a wholly Bayesian framework. The results are in a manuscript currently in preparation.

An anonymous reviewer supports the comparison of alternative approaches to a problem of practical interest but raises the interesting issue that the results may strongly depend on the details of how the approaches were implemented. We agree. That is why the authors did their best to implement the method/model of Huerta et al. (2004), although neither the software used by those authors, nor their data were available to us. Furthermore we were unable to obtain the kind of weather data they used in their model. Nevertheless, our results are broadly in accord with theirs. As for the BSP, it relies on published software and the approach described in Le and Zidek (2006) so we did not have too much flexibility in our implementation of that method either. Thus we did not have much latitude in our implementation of these methods. By making our software for the DLM available such a comparison could well be replicated by others seeking a spatio–temporal model for hourly ozone fields.

That reviewer also asserted that the comparison does not have a very meaningful purpose since the main objective of the DLM is to model the temporal evolution; the spatial correlation structures come as nuisance parameters so to speak. It is true the genesis of the DLM lay in modeling time–series. However temporal models are now commonly combined over space to get space–time models (Lemos, Sans*ó*, and Los Huertos, 2007). Moreover Huerta et al. (2004) was clearly meant to include spatial mapping (see their Figure 5) in their application. Indeed, it was this feature that made this method so prospectively appealing to the authors. Finally, although the BSP was originally intended for spatial modeling, it can be used for temporal forecasting. In particular, in a companion report (Dou et al. 2009), the DLM and BSP are compared for making the all–important, short term, 24–hour ahead ozone forecasts now common in urban areas. Once again the latter is found to work at least as well or better than the former. So while we are convinced of the value of the DLM approach, in some applications the simplicity of the BSP may make it an advantageous alternative.

APPENDIX A: SUPPLEMENTARY RESULTS

A.1. Results for Theorem 2. Only the results about the predictive variances of $y_{01}|y_{11}$ and $y_{01}|y_{11}, y_{12}$ are shown in this appendix. The other two cases can be obtained by the same

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

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

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

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

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

and

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

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

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

The corresponding first partial derivatives are given as follows:

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

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

and

$$\begin{aligned}\frac{\partial}{\partial \sigma_\varepsilon^2}\mathrm{Var}(y_{01}|y_{11},y_{12}) &= (1-\exp(-\frac{d_{01}}{\lambda}))\left\{2-(1-\exp(-\frac{d_{01}}{\lambda}))\frac{\sigma_\varepsilon^2}{A^2}(c_1A-c_2c_3)\right\}\\ &> \frac{1-\exp(-\frac{d_{01}}{\lambda})}{A^2}c_4,\end{aligned}$$

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

A.2. Results for Equations (2.29)–(2.30). Given the values of phase parameters, range as well as variance parameters and the observations until time t, the joint distribution of $\alpha_{1t}^s, \boldsymbol{\alpha}_{1t}$ is

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

where

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

with $\Sigma_{11}^*(\theta)$ a scalar, $\boldsymbol{\Sigma}_{12}^*(\theta)$ a 1 by n vector and $\boldsymbol{\Sigma}_{22}^*(\theta)$ a n by n matrix. We use $\mathbf{V}^*$ to denote the new distance matrix for the unknown site $\mathbf{s}$ and the monitoring stations $\mathbf{s}_1, \ldots, \mathbf{s}_n$.

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

$$\begin{aligned} (\alpha_{1t}^s | \alpha_{1,t-1}^s, \boldsymbol{\alpha}_{1t}, \boldsymbol{\alpha}_{1,t-1}, \mathbf{y}_t, \lambda, \sigma^2) \quad \sim \quad & N[\alpha_{1,t-1}^s + \boldsymbol{\Sigma}_{12}^*(\lambda_1)\boldsymbol{\Sigma}_{22}^*(\lambda_1)^{-1}(\boldsymbol{\alpha}_{1t} - \boldsymbol{\alpha}_{1,t-1}), \\ & \sigma^2\tau_1^2(\Sigma_{11}^*(\lambda_1) - \boldsymbol{\Sigma}_{12}^*(\lambda_1)\boldsymbol{\Sigma}_{22}^*(\lambda_1)^{-1}\boldsymbol{\Sigma}_{21}^*(\lambda_1))]. \end{aligned} \tag{A.1}$$

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

$$\begin{aligned} (\alpha_{2t}^s | \alpha_{2,t-1}^s, \boldsymbol{\alpha}_{2t}, \boldsymbol{\alpha}_{2,t-1}, \mathbf{y}_t, \lambda, \sigma^2) \quad \sim \quad & N[\alpha_{2,t-1}^s + \boldsymbol{\Sigma}_{12}^*(\lambda_2)\boldsymbol{\Sigma}_{22}^*(\lambda_2)^{-1}(\boldsymbol{\alpha}_{2t} - \boldsymbol{\alpha}_{2,t-1}), \\ & \sigma^2\tau_2^2(\Sigma_{11}^*(\lambda_2) - \boldsymbol{\Sigma}_{12}^*(\lambda_2)\boldsymbol{\Sigma}_{22}^*(\lambda_2)^{-1}\boldsymbol{\Sigma}_{21}^*(\lambda_2))]. \end{aligned} \tag{A.2}$$

Using the observation equation as in (2.1), we have the conditional predictive distribution for y_t^s as follows:

$$\begin{aligned} (y_t^s | \mathbf{y}_t, \alpha_{1t}^s, \alpha_{2t}^s, \boldsymbol{\alpha}_{1t}, \boldsymbol{\alpha}_{2t}, \beta_t, \lambda, \sigma^2) \quad \sim \quad & N[\beta_t + S_{1t}(a_1)\alpha_{1t}^s + S_{2t}(a_2)\alpha_{2t}^s \\ & + \boldsymbol{\Sigma}_{12}^*(\lambda)\boldsymbol{\Sigma}_{22}^*(\lambda)^{-1}(\mathbf{y}_t - \mathbf{1}_n\beta_t \\ & - S_{1t}(a_1)\boldsymbol{\alpha}_{1t} - S_{2t}(a_2)\boldsymbol{\alpha}_{2t}), \\ & \sigma^2(\Sigma_{11}^*(\lambda) - \boldsymbol{\Sigma}_{12}^*(\lambda)\boldsymbol{\Sigma}_{22}^*(\lambda)^{-1}\boldsymbol{\Sigma}_{21}^*(\lambda))]. \end{aligned} \tag{A.3}$$

ACKNOWLEDGEMENTS

The authors are grateful to Prasad Kasibhatla for providing in convenient form, the data use for this paper when the third author was on leave at the Statistical and Applied Mathematics Science Institute. The Editor, an anonymous Associate Editor and an anonymous referee suggested revisions that improved our exposition and we thank them.

REFERENCES

[1] Burke, J. M., Zufall, M. J., and Özkaynak, H. (2001). A population exposure model for particulate matter: case study results for $PM_{2.5}$ in Philadelphia, PA. *J Expo Anal Env Epid.* **11** 470–489.
[2] Calder, C. A., Holloman, C. H., Bortnick, S. M., Strauss, W. J., and Morara, M. (2008). Relating ambient particulate matter concentration levels to mortality using an exposure simulator. *J Am Stat Assoc.* **103** 137–148.
[3] Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika.* **81** 541–553.
[4] Dou, Y. P., Le, N. D., and Zidek, J. V. (2007). A Dynamic Linear Model for Hourly Ozone Concentrations. *TR # 228,* Deptartment of Statistics, University of British Columbia.
[5] Dou, Y. P., Le, N. D., and Zidek, J. V. (2009). Temporal prediction with a Bayesian spatial predictor: an application to ozone fields. *TR #249,* Department of Statistics, University of British Columbia.
[6] Huerta, G., Sansó, B., and Stroud, J. R. (2004). A spatio–temporal model for Mexico city ozone levels. *J Roy Stat Soc C-App.* **53** 231–248.
[7] Le, N. D. and Zidek, J. V. (1992). Interpolation with uncertain spatial covariances: a Bayesian alternative to Kriging. *J Multivariate Anal.* **43** 351–374.
[8] Le, N. D. and Zidek, J. V. (2006). *Statistical analysis of environmental space–time processes.* Springer, New York.
[9] Lemos, R. T., Sansó, B. and Los Huertos, M. (2007). Spatially varying temperature trends in a central California estuary. *J Agric Biol Envir S.* **12** 379–396.
[10] Li, K. H., Le, N. D., Sun, L., and Zidek, J. V. (1999). Spatial–temporal models for ambient hourly PM_{10} in Vancouver. *Environmetrics.* **10** 321–338.

[11] Ozone (2005). Air quality criteria for ozone and related photochemical oxidants. *National Center for Environmental Assessment.* RTP: US Environmental Protection Agency, http://oaspub.epa.gov/eims/eimsapi.dispdetail?deid=149923.
[12] Shaddick, G., Lee, D., Zidek, J. V., and Salway, R. (2008). Estimating exposure response functions using ambient pollution concentrations. *Ann Appl Stat.* **2** 1249–1270.
[13] Stroud, J. R., Muller, P., and Sansó, B. (2001). Dynamic models for spatio–temporal data. *J Roy Stat Soc B.* **63** 673–689.
[14] West, M. and Harrison, J. (1997). *Bayesian forcasting and dynamic models*, 2nd ed. Springer, New York.
[15] Zidek, J. V., Sun, L., Le, N. D., and Özkayn, H. (2002). Contending with space–time interaction in the spatial prediction of pollution: Vancouver's hourly ambient PM_{10} field. *Environmetrics.* **13** 595–613.

Department of Statistics
University of British Columbia
6356 Agriculture Rd
Vancouver, BC
Canada, V6T 1Z2
E-mail: ydou@stat.ubc.ca

BC Cancer Research Center
675 West 10th Avenue
Vancouver, BC
V5Z 1L3, Canada
E-mail: nhu@bccrc.ca

Department of Statistics
University of British Columbia
6356 Agriculture Rd
Vancouver, BC
Canada, V6T 1Z2
E-mail: jim@stat.ubc.ca