

Research Article

Temporal Forecasting with a Bayesian Spatial Predictor: Application to Ozone

Yiping Dou,¹ Nhu D. Le,² and James V. Zidek³

¹Finance, eBay Inc., San Jose, CA 95125, USA

²BC Cancer Agency Research Center, Vancouver, BC, Canada V5Z 4E6

³Department of Statistics, University of British Columbia, Vancouver, BC, Canada V6T 1Z2

Correspondence should be addressed to James V. Zidek, jim@stat.ubc.ca

Received 14 January 2012; Revised 24 April 2012; Accepted 27 April 2012

Academic Editor: Tareq Hussein

Copyright © 2012 Yiping Dou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper develops and empirically compares two Bayesian and empirical Bayes space-time approaches for forecasting next-day hourly ground-level ozone concentrations. The comparison involves the Chicago area in the summer of 2000 and measurements from fourteen monitors as reported in the EPA's AQS database. One of these approaches adapts a multivariate method originally designed for spatial prediction. The second is based on a state-space modeling approach originally developed and used in a case study involving one week in Mexico City with ten monitoring sites. The first method proves superior to the second in the Chicago Case Study, judged by several criteria, notably root mean square predictive accuracy, computing times, and calibration of 95% predictive intervals.

1. Introduction

This paper compares two methods for temporally forecasting next-day hourly ground-level ozone concentrations over spatial regions. Software for implementing both methods along with demo files can be downloaded from <http://enviro.stat.ubc.ca/>. The paper focuses on a case study involving Chicago during the summer of 2000. The methods can be used to forecast the maximum eight-hour average ozone concentration, which is reported for many urban areas. For example, on June 27, 2009 the AIRNow website forecasts a maximum for Chicago of between 0 and 50 ppb, rating that as "Good." In contrast, for that day in one part of Los Angeles, the rating was "Unhealthy for sensitive groups," meaning a forecast maximum of between 101 and 150 ppb.

These forecasts are needed to forewarn susceptible groups of high ozone concentrations that are associated with acute health effects. Such effects are well documented in the air quality criterion document (Ozone [1]). See <http://oaspub.epa.gov/eims/eimsapi.dispdetail?deid=149923>, the basis of the recommendations made in 2007 to the US Environment Protection Agency by its Clean Air Scientific Advisory Committee for Ozone, in which the third author served. In fact,

the accumulated body of evidence was so strong that the committee recommended strengthening the air quality standards for this criterion pollutant to meet the requirements of the US Clean Air Act. In particular, the evidence pointed to a strong association between morbidity as well as reduced lung function and high levels of ground-level ozone concentrations. This points to a need for enhanced near-term forecasting methods.

One general method for making such forecasts relies on the fusion of measured hourly ozone concentration values and simulated values obtained from chemical transport models (CTMs) such as CMAQ. Two papers [2, 3] develop methods for doing this, albeit by different approaches unrelated to those in this paper. In future work, all these methods should be compared in domains where CTM data are available. Even without that hypothetical comparison, those in this paper have the advantage of being available in domains where they are not. These Bayes-empirical Bayes methods offer the flexibility needed to characterize environmental space-time processes, while fully representing the various kinds of uncertainty involved in their construction. Both have been developed for and successfully used to model hourly ozone air pollution concentrations in other contexts.

The first method in this paper denoted by M1 adapts a multivariate method developed for modeling space-time fields [4–7]. A univariate version of that method for hourly ozone concentrations is the subject of a companionable paper [8], which compares it with a state space model but for spatial prediction, not temporal forecasting, and it does so in a different geographical region. The goal there is mapping the ozone field for another requirement of the US Clean Air Act of 1970, namely, the protection of human welfare including such things as crop yields. M1 needs some new theory, which is presented in the sequel along with a demonstration on how it may be applied.

The second method denoted by M2 uses a method originally developed for modeling hourly ozone concentrations in Mexico City [9]. That method and the models on which it is based along with the computational algorithms used to implement it seem to have been quite successful in that application. Moreover, even though it was developed for use in Mexico City for one specific week, a strong *prima facie* case can be made for its applicability in other weeks and jurisdictions and that is why we assess the performance of that method here. Much recent work has been done in modeling random space-time pollution fields [10–12]. As one of the photo-oxidants, ozone is produced in the same way in all temporal and spatial domains by a complex interaction of oxides of nitrogen (NO_x) with volatile organic compounds (VOCs) in the presence of heat and sunlight. In most modern urban environments such as Chicago and Mexico City, vehicle emissions are a prime source of the NO_x and VOCs [1]. Furthermore the *prima facie* case is supported by an exploratory data analysis that shows very similar daily cycles in our domain as those observed in the Mexico City application. This led us to match, to the greatest possible extent [8], the method used there in our adaptation of it, and no originality is claimed for it.

The main finding in this paper is that in the case study M1 outperforms M2 in a number of ways. First is its computational efficiency. To run the M2 approach, it often took about a week or so to get the results, while M1 only took about ten to twelve hours at the same Linux server. Thus, M2 would not be suitable for making 24 ahead forecasts, while M1 running on a faster processor could be used for that purpose. We also found that M1 produced more accurate forecasts than M2, as measured by their root-mean-squared-prediction errors. Moreover, M1’s predictive error bands proved to be better calibrated. In other space-time domains, a similar assessment would have to be made to select a forecasting procedure, and M2 may be superior in some. Overall, we believe that the value of this paper lies in the guidance on how that assessment could be made and the source of software that can be used for it.

The layout of this paper will now be described. Section 2 presents both approaches to forecasting hourly ozone concentrations. Section 2.1 introduces M1. Its forecasting (posterior) distribution is developed, and the corresponding pointwise predictive intervals at each gauged site constructed. Section 2.1 also extends the results to forecasting r -step-ahead responses for any $r \in \mathcal{N}$. Section 2.2 reviews M2. Section 3 implements these two methods in a case study

involving Chicago and data from the US EPA’s AQS air quality database. Section 3.1 presents and compares the results for one-day-ahead predictions. Finally, Section 4 summarizes these results and gives our conclusions.

2. Methodology

This section presents our two temporal forecasting approaches. Although both are general and can be used in other contexts, we develop them as methods for forecasting an hourly response tomorrow given data up to today. The measured value of the response is available and serves as a “test value” at each of g monitoring sites for a comparative assessment of the two forecasted methods.

2.1. Method M1

2.1.1. Basic Theory. The general approach [7], on which M1 is based, assumes a multivariate space-time field of p -dimensional random response vectors indexed by their locations, a finite number of sites in a specified geographical domain. These sites need not lie on a lattice.

The general theory involves a geographical region that includes g monitored (i.e., gauged) sites and u unmonitored (i.e., ungauged) sites. Although this paper requires only the part of that theory for those that are monitored, we state it in generalable to link this paper to its companion publication [8]. The p dimensional response vectors at these sites for times $t = 1, \dots, n$ are combined in the response matrix $\mathbf{Y} : n \times (u + g)p$. This matrix can be partitioned as $(\mathbf{Y}^{[u]}, \mathbf{Y}^{[g]})$, where $\mathbf{Y}^{[u]} : n \times up$ contains the response vectors at the ungauged sites and $\mathbf{Y}^{[g]} : n \times gp$, those at the gauged sites. The theory posits ([7], p. 145–146) that

$$\begin{aligned} \mathbf{Y} \mid \beta, \Sigma &\sim N(\mathbf{Z}\beta, \mathbf{I}_n \otimes \Sigma), \\ \beta \mid \Sigma, \beta_0, \mathbf{F} &\sim N(\beta_0, \mathbf{F}^{-1} \otimes \Sigma), \\ \Sigma &\sim \text{GIW}(\Psi, \delta), \end{aligned} \quad (1)$$

where $\mathbf{Z} : n \times l$ denotes the non-site-specific covariates (their site-specific counterparts could be included in the response vectors); $\beta : l \times (u + g)p$ denotes the matrix of site-specific random covariate coefficients; $\Sigma : (u + g)p \times (u + g)p$ denotes the covariance matrix among the responses at any given time point. The hyperparameters for this hierarchical Bayes model are $\beta_0, \mathbf{F}, \delta$, and Ψ , with \mathbf{F} representing the variance component of β between its l rows. Here, GIW denotes the Generalized Inverted Wishart distribution, a conjugate prior for normal matrix distributions. Separability of the hypercovariant matrices for both the response and random coefficient matrices is assumed for computational simplicity. Invoking Box’s celebrated dictum that all models are wrong, we defend this assumption by the good performance of the resulting method as seen in the empirical assessment provided in the sequel.

Validating the modeling assumptions above will usually require a transformation of the random responses, a square root transformation in this paper’s case study. Then systematic components such as the temporal trend over the whole

regions will need to be removed. These can be accurately inferred from the typically large dataset formed by aggregating the data over all sites and times. Finally, something needs to be done to eliminate autocorrelation in the temporal sequence of responses. For example, the temporal series can often be filtered using a regional time series model without site-specific parameters. However, our relative abundance of data leads us here to a different approach described in detail in the next subsection, that splits the transformed, detrended residuals into separate, disjoint subsequences of responses, which are separated widely enough in time as to be uncorrelated and hence independent under our Gaussian sampling model. In our experience, the residuals obtained after these steps have been taken usually satisfy the model assumptions above, and these comprise the response vectors in M1. So the model above can then be applied to each subsequence and type II maximum likelihood estimators found for the hyperparameters. These can subsequently be averaged across the subsubsequences to get an overall estimate. While this approach would be less efficient than a full-data approach under a correctly specified model, it avoids the risk of model misspecification in complex situations like that of the case study. The forecasting model developed below can then be applied, and the preliminary steps above reversed, to get the forecasts back on the scale of the raw data.

To elaborate on our distributional assumptions, the GIW prior for

$$\Sigma = \begin{pmatrix} \Sigma^{[g,g]} & \Sigma^{[g,u]} \\ \Sigma^{[u,g]} & \Sigma^{[u,u]} \end{pmatrix} : \begin{pmatrix} gp \times gp & gp \times up \\ up \times gp & up \times up \end{pmatrix} \quad (2)$$

can be defined, through the Bartlett decomposition, as follows:

$$\begin{aligned} \Sigma^{[g,g]} &\sim \text{IW}(\Lambda_1, \delta_1), \\ \Gamma^{[u]} &\sim \text{IW}(\Lambda_0, \delta_0), \\ \tau^{[u]} | \Gamma^{[u]} &\sim N(\tau_{00}, \mathbf{H}_0 \otimes \Gamma^{[u]}), \end{aligned} \quad (3)$$

where $\Gamma^{[u]} = \Sigma^{[u,g]} = \Sigma^{[u,u]} - \Sigma^{[u,g]}(\Sigma^{[g,g]})^{-1}\Sigma^{[g,u]}$, and $\tau^{[u]} = (\Sigma^{[g,g]})^{-1}\Sigma^{[g,u]}$. Note that $\Sigma^{[g,g]}$ has an inverted Wishart distribution with hyperparameters (Λ_1, δ_1) ; the matrix τ_{00} is the hypermean of $\tau^{[u]}$, and the matrix \mathbf{H}_0 gives the covariance between the rows of $\tau^{[u]}$. Denote the set of hyperparameters as $\mathcal{H} = \{\mathbf{F}, \beta_0, \Omega, \Lambda_1, \delta_1, \Lambda_0, \delta_0, \tau_{00}, \mathbf{H}_0\}$.

Given the observations at the gauged sites (i.e., $\mathbf{Y}^{[g]}$), the predictive distribution of $\mathbf{Y}^{[u]}$ is completely determined, it is the distribution required in our companionable paper on spatial prediction [8]. Furthermore given partially observed responses at gauged sites, the predictive distribution of the missing responses at gauged sites given these observations can be derived after the hyperparameters have been estimated by an empirical Bayes approach ([7], p. 300–303), and that is how the theory is used in this paper.

Deriving that forecasting model requires a general result that concerns a sequence of n response vectors of which $n-w$ are observed, and w lies in the future. Then with the superscripts “ m ” and “ o ,” respectively, standing for “missing” and “observed,” we may further partition the random response

matrix already partitioned above, as $\mathbf{Y}^{[g]} = (\mathbf{Y}^{[g^m]}, \mathbf{Y}^{[g^o]})' : n \times gp$ with $\mathbf{Y}^{[g^m]} : w \times gp$ and $\mathbf{Y}^{[g^o]} : (n-w) \times gp$. Forecasting requires the predictive posterior distribution of $(\mathbf{Y}^{[g^m]} | \mathbf{Y}^{[g^o]}, \mathcal{H})$ given in the following result ([7], p. 160–161).

Theorem 1. *Let*

$$\begin{aligned} \mathbf{Z}\beta_0^{[g]} &= \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix} : \begin{pmatrix} w \times gp \\ (n-w) \times gp \end{pmatrix}, \\ \mathbf{I}_n + \mathbf{Z}\mathbf{F}^{-1}\mathbf{Z}' &= \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} : \begin{pmatrix} w \times w & w \times (n-w) \\ (n-w) \times w & (n-w) \times (n-w) \end{pmatrix}. \end{aligned} \quad (4)$$

Conditional on the hyperparameters \mathcal{H} , the marginal posterior distribution is a matrix t distribution:

$$\mathbf{Y}^{[g^m]} | \mathbf{Y}^{[g^o]}, \mathcal{H} \sim t_{w \times gp}(\mu_{(u|g)}, \Phi_{(u|g)} \otimes \Psi_{(u|g)}, \delta_{(u|g)}), \quad (5)$$

where

$$\begin{aligned} \mu_{(u|g)} &= \mu_{(1)} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{Y}^{[g^o]} - \mu_{(2)}): w \times gp, \\ \Phi_{(u|g)} &= \frac{\delta_1 - gp + 1}{\delta_1 - gp + n - w + 1} \mathbf{A}_{11 \circ 2} : w \times w, \\ \Psi_{(u|g)} &= \frac{1}{\delta_1 - gp + 1} \left\{ \Psi_{gg} + (\mathbf{Y}^{[g^o]} - \mu_{(2)})' \right. \\ &\quad \left. \mathbf{A}_{22}^{-1}(\mathbf{Y}^{[g^o]} - \mu_{(2)}) \right\} : gp \times gp, \end{aligned} \quad (6)$$

$$\delta_{(u|g)} = \delta_1 - gp + n - w + 1,$$

where $\mathbf{A}_{11 \circ 2} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$, and Ψ_{gg} is assumed to be $\Lambda_1 \otimes \Omega$, denoting spatial and between hour correlations, respectively.

Remark 2. This theorem gives the joint predictive distribution for w future response vectors given $(n-w)$ observed responses. As coordinates of those future vector are observed, this distribution yields in turn the conditional predictive distribution for the unobserved coordinate responses, That is how the theorem is applied in the next section.

2.1.2. One-Day-Ahead Forecasts. For expository simplicity, we describe the general method M1 in terms of the goal of forecasting ozone concentrations at a specific hour on Day 121 and each of $g = 14$ monitoring sites based on data collected at those sites during the preceding days, that being the objective in the case study described in the next section. However, we emphasize that M1 is generally applicable to other days and geographical domains with appropriate modifications. In fact, the approach can be adapted for use with other environmental processes. With that caveat, we now describe M1 in this subsection. Note that in our description, hour 1 refers to the period between midnight and 1 AM and so on.

To begin, we follow the standard practice of transforming the hourly data by taking their square roots to achieve a more nearly Gaussian data distribution [13]. Hereafter these transformed values will be our “responses.” M1 then partitions the sequence of responses into blocs of $p = 24$ hours each. For $k < 24$ (hereafter referred to as Case 2), the last bloc spans the 24 hour period from hour $k + 1$ on day 120 to hour k on Day 121, and the first from hour $k + 1$ of Day 1 to hour k of day 2. However $k = 24$ (Case 1) is different, and there the blocs correspond to the days from 1 to 121. In either case, each bloc yields a 24-dimensional multivariate response vector with an unspecified covariance structure. Reformulating our model in this way gives us the advantage of avoiding the challenging task of specifying the complex short-term autocovariance structure, which varies over the day. But it does require the multivariate theory described in the previous subsection.

The next step in developing the forecast model would generally require the removal of any systematic, regional components in the series. In particular, it is necessary to learn which covariates/predictors to include in the design matrix \mathbf{Z} so that in application of the method, site-specific coefficients can be fitted allowing deviations from the regional baselines established for them at the preliminary stage. Note that the EnviRo.stat software (see [14]) referred to in Section 1 automatically estimates the those baseline coefficients as prior hypermeans, using a maximum likelihood-based approach, the key preliminary step here is actually the identification of \mathbf{Z} . Note also that in our application the covariates need to be adapted in form to conform to the temporal span of the response vectors (the so-called “blocs” introduced below), once hour k has been specified.

Finally, we need to address the autocovariance structure. While the responses are primarily an AR(2) time series after removing their diurnal pattern [13], as we do in the case study, we need to allow for a lag 1 autocorrelation in series of the response vectors to capture small but potentially significant longer-term dependence [15]. To eliminate it, we use the approach described in the previous subsection and split the observed response vectors into two groups: those with odd numbers and those with even numbers. For each of the resulting subseries, the model assumptions will hold approximately, and the hyperparameters can be estimated as described in the previous subsection. More detail is given below where we turn to a more precise description of M1.

In Case 2 where $k < 24$, the 24-hour bloc containing hour k includes measured responses for the hours $k + 1, \dots, 24$ on Day 120. Thus we may first apply Theorem 1 with $w = 1$ designating the last bloc in our construction above, to get a joint predictive distribution for its associated “future” response vector. Then we compute the marginal conditional predictive distribution for hour k implied by that joint distribution, given these data from Day 120 to finish the construction.

However, Case 1 is more difficult and the one we treat first. Among options considered by the authors for this case, was changing the bloc length to say $p = 25$ so that the last bloc could reach back to hour k on Day 120, which contains a measured response. That approach was discarded since the sequence of 7 bloc sequences would not synchronize

with weeks. The latter are important structural features of the process that reflect the changing daily traffic patterns and give us the single covariate we have in the case study. Instead we accomplish the same thing, another way, which does preserve the week, namely, we apply Theorem 1 with $w = 2$ to get the initial joint predictive distribution. The required marginal conditional distribution is then obtained by conditioning on all the data from Day 120.

Note that in both cases, the future response or responses in Theorem 1 depend on the bloc immediately preceding them. To avoid the need to model in that dependence, we simply eliminate that data vector. We did explore the result of keeping it in and ignoring that dependence and found virtually no difference in the Case Study. We now turn to a more precise description of M1.

For that we need some notation. Let $Y_{t,i}^{[g_j^m]}$ denote the i th coordinate of the unobserved response vector for blocs t and gauged site j , while $Y_{t',i}^{[g_j^o]}$ denotes the observed response for bloc t' , $t, t' = 1, \dots, n$, $i = 1, \dots, p$, $j = 1, \dots, g$. For bloc t , Sites g_1, \dots, g_{14} , and hours between i and j inclusive, that is, $i : j$, let $Y_{t,i;j}^{[g_l^o]}$, $l = 1, \dots, 14$ denote the random response. Throughout $p = 24$.

The two cases referred to above are as follows.

Case 1 (predict the response for the last hour $k = 24$ day 121). Hyperparameter estimates for its predictive distribution conditional on observed data are found first for the odd blocs $\mathbf{U}_t = (Y_{2t-1,1}^{[g_1^o]}, \dots, Y_{2t-1,p}^{[g_g^o]}) : gp \times 1$, $t = 1, \dots, 60$ as described in the previous subsection, using $\mathbf{Y}^{[g^o]'} = (\mathbf{U}_1, \dots, \mathbf{U}_{60})$. Repeat this procedure for the even blocs, letting $\mathbf{V}_t = (Y_{2t,1}^{[g_1^o]}, \dots, Y_{2t,p}^{[g_g^o]})$, $t = 1, \dots, 59$. Finally average the resulting pairs of hyperparameter estimates to get overall estimates.

Construction of the predictive distribution of the response at hour $k = 24$, begins with the corresponding matrix-variate observed responses, $\mathbf{Y}^{[g^o]}$ made by extracting the corresponding responses from Day 1 to Day 118 according to the above constructed \mathbf{U}_t or \mathbf{V}_t . Now suppose hypothetically that Day 120's responses have not been observed and suppose the corresponding matrix-variate missing responses, $\mathbf{Y}^{[g^m]}$ are constructed as follows:

$$\mathbf{Y}^{[g^m]} = \begin{bmatrix} Y_{121,1}^{[g_1^m]} & \dots & Y_{121,p}^{[g_g^m]} \\ Y_{120,1}^{[g_1^o]} & \dots & Y_{120,p}^{[g_g^o]} \end{bmatrix} = \begin{pmatrix} \mathbf{Y}_{121,1:p}^{[g_{1:g}^m]} \\ \mathbf{Y}_{120,1:p}^{[g_{1:g}^o]} \end{pmatrix} : 2 \times gp, \quad (7)$$

where $\mathbf{Y}_{121,1:p}^{[g_{1:g}^m]} : 1 \times gp$ is the unobserved future response vectors for days 121 and $\mathbf{Y}_{120,1:p}^{[g_{1:g}^o]} : 1 \times gp$, the observed response vector of Day 120. Hence we have $w = 2$ and $n = 121$ in Theorem 1. Thus the predictive posterior distribution of $\mathbf{Y}^{[g^m]}$ can be obtained by applying (6). To obtain that of $\mathbf{Y}_{121,1:p}^{[g_{1:g}^m]}$ given $\mathbf{Y}_{1:120,1:p}^{[g_{1:g}^o]}$, since in reality the latter are observed, one can decompose $\mu_{(u|g)}$ and $\Phi_{(u|g)}$ as follows:

$$\mu_{(u|g)} = \begin{pmatrix} \mu_{1r} \\ \mu_{2r} \end{pmatrix}, \quad \delta_{(u|g)} \Phi_{(u|g)} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \quad (8)$$

where $\mu_{ir} : 1 \times gp$ and $B_{ij} : 1 \times 1$ for $i, j = 1, 2$. Hence, the predictive posterior distribution of $\mathbf{Y}_{121,1:p}^{[g^m]}$ is given by

$$\begin{aligned} & \mathbf{Y}_{121,1:p}^{[g^m]} \mid \mathbf{Y}_{120,1:p}^{[g^o]}, \mathbf{Y}_{1:119,1:p}^{[g^o]}, \mathcal{H} \\ & \sim t_{1 \times gp} \left(\mu_{1r} + B_{12} B_{22}^{-1} \left(\mathbf{Y}_{120,1:p}^{[g^o]} - \mu_{2r} \right), \frac{B_{11 \circ 2}}{\delta_{(u|g)} + 1} \right. \\ & \quad \otimes \boldsymbol{\Psi}_{(u|g)} \left(\mathbf{I}_{gp} + \boldsymbol{\Psi}_{(u|g)}^{-1} \left(\mathbf{Y}_{120}^{[g^o]} - \mu_{2r} \right)' B_{22}^{-1} \right. \\ & \quad \left. \left. \times \left(\mathbf{Y}_{120}^{[g^o]} - \mu_{2r} \right) \right), \delta_{(u|g)} + 1 \right). \end{aligned} \quad (9)$$

To get the joint predictive distribution of responses at these gauged sites for Day 121's last hour, let $\mathbf{e}_k^l : k \times 1$ be such that $e_{kl} = 1$ and $e_{kj} = 0$ for $j \neq l, j = 1, \dots, k$. Let $\mathbf{E}_1 = \text{blocs-diag-matrix} \{ \mathbf{e}_p^l \} : gp \times g$. The joint predictive distribution of the p th unobserved response $\mathbf{Y}_{121,p}^{[g^m]}$, that is, $\mathbf{Y}_{121,1:p}^{[g^m]} \mathbf{E}_1$, is also a multivariate t -distribution:

$$\begin{aligned} & \mathbf{Y}_{121,p}^{[g^m]} \mid \mathbf{Y}_{120,1:p}^{[g^o]}, \mathbf{Y}_{1:119,1:p}^{[g^o]}, \mathcal{H} \\ & \sim t_{1 \times g} \left(\mu^* \mathbf{E}_1, \phi^* \mathbf{E}_1' \boldsymbol{\Psi}^* \mathbf{E}_1, \delta_{(u|g)} + 1 \right), \end{aligned} \quad (10)$$

where $\mu^* = \mu_{1r} + B_{12} B_{22}^{-1} (\mathbf{Y}_{120,1:p}^{[g^o]} - \mu_{2r})$, $\phi^* = B_{11 \circ 2} / (\delta_{(u|g)} + 1)$ and $\boldsymbol{\Psi}^* = \boldsymbol{\Psi}_{(u|g)} (\mathbf{I}_{gp} + \boldsymbol{\Psi}_{(u|g)}^{-1} (\mathbf{Y}_{120,1:p}^{[g^o]} - \mu_{2r})' B_{22}^{-1} (\mathbf{Y}_{120,1:p}^{[g^o]} - \mu_{2r}))$, that completes the description of M1 in this case.

Case 2 (predict the response for hour $k < 24$ on day 121). For bloc t , Sites g_1, \dots, g_{14} , and hours between i and j inclusive, that is, $i : j$, let $\mathbf{Y}_{t,i;j}^{[g^l]}$, $l = 1, \dots, 14$ denote the random response. Now let $\mathbf{Y}^{[g^m]}$ consist of k unobserved responses and $p - k$ observed ones at each of the gauged sites. What we need to do now at each site is to use the hours leading up to the first hour without data on Day 121, that is, hour k for which the forecast is needed. To do this, we can use any data from hours $1 : k$ on Day 121 that may be available, supplemented by the data from the $p - k$ preceding hours on Day 120, $(k + 1) : p = 21 : 24$. Thus we create a $p = 24$ dimensional response vector, $\mathbf{Y}^{[g^m]} = (Y_{120,(k+1):p}^{[g_1]}, Y_{121,1:k}^{[g_1]}, \dots, Y_{120,(k+1):p}^{[g_{14}]}, Y_{121,1:k}^{[g_{14}]})$. The same routine as in Case 1 using odd and even blocs, is then used with the remaining data to obtain parameter estimates, albeit with these shifted 24 dimensional hourly response vectors.

For hour $k < 24$, $\mathbf{Y}^{[g^o]}$ holds the observed responses from Day 1 to the one ending on day 119. To predict the responses one-day-ahead at gauged sites in this field, we have $w = 1$ and $n = 120$ in Theorem 1 with $\mathbf{Y}^{[g^m]}$ rearranged so that all missing responses are at the beginning of the response vector. Specifically, let $\mathbf{e}_1^{i,j} : gp \times 1$, be 1 at the $(pj - i + 1)$ th element and 0 otherwise, for $i = 1, \dots, k$ and $j = 1, \dots, g$; $\mathbf{e}_2^{i,j} : gp \times 1$, be 1 at the $(p(j - 1) + i)$ th element and 0

otherwise, for $i = 1, \dots, p - k$ and $j = 1, \dots, g$. Let $\mathbf{E}_2 : gp \times gp = (\mathbf{e}_1^{1:k,1}, \dots, \mathbf{e}_1^{1:k,g}, \mathbf{e}_2^{1:(p-k),1}, \dots, \mathbf{e}_2^{1:(p-k),g})$. Applying Theorem 1 yields

$$\begin{aligned} & \mathbf{Y}^{[g^m]} \mathbf{E}_2 \mid \mathbf{Y}^{[g^o]}, \mathcal{H} \\ & \sim t_{1 \times gp} \left(\mu_{(u|g)} \mathbf{E}_2, \Phi_{(u|g)} \otimes \mathbf{E}_2' \boldsymbol{\Psi}_{(u|g)} \mathbf{E}_2, \delta_{(u|g)} \right). \end{aligned} \quad (11)$$

Notice that $\mathbf{Y}^{[g^m]} \mathbf{E}_2$ is $(Y_{n,k;1}^{[g_1^1]}, \dots, Y_{n,k;1}^{[g_{14}^1]}, Y_{n-1,1:(p-k)}^{[g_1^o]}, \dots, Y_{n-1,(p-k)}^{[g_{14}^o]})$. To obtain the predictive distribution for the unobserved responses, we first need to decompose $\mathbf{Y}^{[g^m]} \mathbf{E}_2$, $\mu_{u|g}$ and $\mathbf{E}_2' \boldsymbol{\Psi}_{(u|g)} \mathbf{E}_2$ as follows: $\mathbf{Y}^{[g^m]} \mathbf{E}_2 = (\mathbf{T}_{1c}, \mathbf{T}_{2c}) : (1 \times gk, 1 \times g(p - k))$, where $\mu_{(u|g)} = (\mu_{1c}, \mu_{2c}) : (1 \times gk, 1 \times g(p - k))$, and

$$\begin{aligned} & \mathbf{E}_2' \boldsymbol{\Psi}_{(u|g)} \mathbf{E}_2 \\ & = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix} : \begin{pmatrix} gk \times gk & gk \times g(p - k) \\ g(p - k) \times gk & g(p - k) \times g(p - k) \end{pmatrix}. \end{aligned} \quad (12)$$

Applying standard theory, for the multivariate t distribution yields the predictive distribution of the unobserved response \mathbf{T}_{1c} given data \mathbf{T}_{2c} as a t -distribution given by

$$\begin{aligned} & \mathbf{T}_{1c} \mid \mathbf{T}_{2c}, \mathbf{Y}^{[g^o]}, \mathcal{H} \\ & \sim t_{1 \times gk} \left(\mu_{1c} + (\mathbf{T}_{2c} - \mu_{2c}) \mathbf{C}_{22}^{-1} \mathbf{C}_{21}, \right. \\ & \quad \frac{\delta_{(u|g)}}{\delta_{(u|g)} + g(p - k) + 1} \Phi_{(u|g)} \\ & \quad \times \left\{ 1 + (\delta_{(u|g)} \Phi_{(u|g)})^{-1} (\mathbf{T}_{2c} - \mu_{2c}) \mathbf{C}_{22}^{-1} \right. \\ & \quad \left. \times (\mathbf{T}_{2c} - \mu_{2c})' \right\} \otimes \mathbf{C}_{11 \circ 2}, \\ & \quad \left. \delta_{(u|g)} + g(p - k) + 1 \right). \end{aligned} \quad (13)$$

This completes the description of M1 in this case.

2.1.3. Multi-Day Ahead Forecasts. This subsection generalizes M1 to get a method that provides an r -step-ahead forecast, ($r \in \mathcal{N}$). Let N be the total number of days of observed responses. As before, we consider the multivariate setting, $p = 24$ being the total number of response coordinates and g , the total number of gauged sites. As well, we generalize the forecasting problem in two ways, namely, to forecasting the response on the last hour of the $(N + r)$ th day; hour $k < 24$ on the $(N + r)$ th day.

Case 1 (predict the final response for the last hour on day $N + r$). Here the odd bloc responses are $\mathbf{U}_t^{(1)} = (Y_{2t-1,1}, \dots, Y_{2t-1,p}^{[g^o]})$. Note that $t^O = K$ if $N = 2K$ or $N = 2K - 1$ for some $K \in \mathcal{N}$. The even bloc responses are $\mathbf{V}_t^{(1)} = (Y_{2t,1}^{[g^o]}, \dots, Y_{2t,p}^{[g^o]})$. Note that $t^E = K - 1$ if $N = 2K$ and $t^E = K$ if $N = 2K - 1$ for some $K \in \mathcal{N}$.

Remark 3. Notice that the total number of observations in data submatrices can be different for each N , being an odd or even number. Denote by t^N , the total number of observed responses. Then $t^N = t^O$ for the odd-day-responses and t^E , for the even-day-responses.

As in Section 2.1, we obtain the estimates of hyperparameters by averaging those two sets of estimates given odd or even bloc responses, respectively. Given these final estimates, we now obtain the predictive posterior distributions given all hourly observations up to Day N in Theorem A.1. They are also multivariate t -distributions. The proof and details can be seen in Appendix A.1.

Case 2 (predict the response for hour $k < 24$ on day $N + r$). Here the odds block responses are $\mathbf{U}_t^{(k)} = (Y_{2t-1,k+1}^{[g_i^o]}, \dots, Y_{2t-1,p}^{[g_i^o]}, Y_{2t,1}^{[g_i^o]}, \dots, Y_{2t,k}^{[g_i^o]}, \dots, Y_{2t-1,k+1}^{[g_g^o]}, \dots, Y_{2t-1,p}^{[g_g^o]}, Y_{2t,1}^{[g_g^o]}, \dots, Y_{2t,k}^{[g_g^o]}) : 1 \times gp$. Note that $t^O = K \in \mathcal{N}$ if $N = 2K$ and $t^O = K - 1$ if $N = 2K - 1$ for some $K \in \mathcal{N}$. The even block responses are $\mathbf{V}_t^{(k)} = (Y_{2t,k+1}^{[g_i^e]}, \dots, Y_{2t,p}^{[g_i^e]}, Y_{2t+1,1}^{[g_i^e]}, \dots, Y_{2t+1,k}^{[g_i^e]}, \dots, Y_{2t+1,p}^{[g_i^e]}, \dots, Y_{2t+1,1}^{[g_g^e]}, \dots, Y_{2t+1,k}^{[g_g^e]}) : 1 \times gp$. Note that $t^E = K - 1$ if $N = 2K$ or $N = 2K - 1$ for some $K \in \mathcal{N}$. We do the same thing here to obtain the ‘‘final’’ estimates for hyperparameters. Thus we are able to get the predictive posterior distribution given all observations up to bloc N and estimates for hyperparameters.

Remark 4. We also let t^N be the total number of observed response variables. So t^N is t^O for the odds-day-response blocs and t^E , for the even-day-response blocs.

Given the final estimates in Section 2.1, the predictive posterior distributions given all hourly observations up to N days are also multivariate t -distribution. The proof and details can be seen in Appendix A.2.

Remark 5. From Theorem A.2, the predictive distribution for the unobserved response variables from day $N + 1$ to $N + r$ is the product of a sequence of matrix- t and t distributions. This implies no analytic form can be found for the response variable at the $(k-1)$ th (for $k = 2, \dots, p$) hour of the $(N+r)$ th day at gauged site j ($j = 1, \dots, g$).

2.2. Method M2. An alternative approach to M1, through dynamic linear modeling, can also be used for forecasting and would seem an obvious choice, being an amalgamation of state-space time series models. Let \mathbf{Y}_t be the response vector across all sites at hour t . As in the previous subsections, responses are square root transformed hourly ozone concentrations. Exploratory analysis of these transformed data, found 24- and 12-hour diurnal cycles, pointing to the approach in Huerta et al. [9] for ozone that is based on the same patterns.

Let β_t be the common temporal trend coefficient across the spatial sites at hour t . Furthermore, α_{1t} and α_{2t} denotes the coefficients of the periodic components with respect to 24- and 12-hour diurnal cycles, respectively. These two

components are $S_{jt}(a_j) = \cos(\pi jt/12) + a_1 \sin(\pi jt/12)$ for $j = 1, 2$. Let $\mathbf{x}_t = (\beta_t, \alpha'_{1t}, \alpha'_{2t})' : (2n + 1) \times 1$ be the state vector at hour t and $\mathbf{F}_t : n \times (2n + 1) = [\mathbf{1}'_n, \text{diag}(S_{1t}(a_1)), \text{diag}(S_{2t}(a_2))]$. Furthermore, $\mathbf{V} : n \times n$ denotes the Euclidean intersite distance matrix for all site pairs. In this model, λ is the range parameter, σ^2 , the variance parameter, and a_1, a_2 , the phase parameters. Assume a constant covariance matrix for the state parameter vector \mathbf{x}_t and denote it by \mathbf{W} . That covariance, which accounts for the spatial correlation between sites, is given by $\mathbf{W} = \text{diag}(\tau_y^2, \tau_1^2 \exp(-\mathbf{V}/\lambda_1), \tau_2^2 \exp(-\mathbf{V}/\lambda_2))$. The vector of hyperparameters, $\gamma = (\tau_y^2, \tau_1^2, \lambda_1, \tau_2^2, \lambda_2)$, is identical to that in Huerta et al. [9] based on assessments made for Dou et al. [8].

Thus, the measurement and state equations of the DLM are given by

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{F}'_t \mathbf{x}_t + \nu_t, & \nu_t &\sim N\left(\mathbf{0}, \sigma^2 \exp\left(-\frac{\mathbf{V}}{\lambda}\right)\right), \\ \mathbf{x}_t &= \mathbf{x}_{t-1} + \omega_t, & \omega_t &\sim N(\mathbf{0}, \sigma^2 \mathbf{W}), \end{aligned} \quad (14)$$

with initial information: $\mathbf{x}_0 \mid \mathbf{D}_0 \sim N(\mathbf{m}_0, \sigma_0^2 \mathbf{C}_0)$. The hyperparameters \mathbf{m}_0 , σ_0^2 , and \mathbf{C}_0 are also identical to those in Huerta et al. [9]. One can obtain the posterior distribution of the state parameters at the last known time point, n , that is, $\mathbf{x}_n \mid \mathbf{y}_{1:n}, \theta \sim N(\mathbf{m}_n, \sigma^2 \mathbf{C}_n)$, using the Kalman filter, a smoothing method and the Metropolis-within-Gibbs sampling algorithm [8, 9, 16–18]. We omit details on updating and forecasting the state parameters given the model parameters and observations up to current time point.

Given the distribution of the state parameters at the last time point, n , the observed responses until time n , $\mathbf{y}_{1:n}$, and the model parameters, $\theta = \{\lambda, \sigma^2, a_1, a_2\}$, the r -step-ahead prediction is given by

$$\begin{aligned} &\mathbf{y}_{n+r} \mid \mathbf{y}_{1:n}, \theta \\ &\sim N\left(\mathbf{F}'_{t+r} \mathbf{m}_n, \sigma^2 \left\{ \mathbf{F}'_{t+r} (\mathbf{C}_n + r\mathbf{W}) \mathbf{F}_{t+r} + \exp\left(-\frac{\mathbf{V}}{\lambda}\right) \right\}\right), \end{aligned} \quad (15)$$

for $r \in \mathcal{N}$. Note that \mathbf{F}_{t+r} , \mathbf{m}_n and \mathbf{C}_n can be obtained by application of a standard method [8, 9, 17]. Here $n = 2880$ and $r = 1, \dots, 24$ for the one-day-ahead prediction in the Case Study. For any fixed r , the predictive response, \mathbf{y}_{n+r} , can also be obtained by the MCMC (Markov Chain Monte Carlo) method. More specifically, at iteration j , suppose we have updated the vector of model parameters: $\theta^{(j)} = (\lambda^{(j)}, \sigma^{2(j)}, a_1^{(j)}, a_2^{(j)})$ using the FFBS (forward-filtering-backward-sampling) algorithm [8, 9, 16, 18]. That is, one has

$$\mathbf{x}_n \mid \mathbf{y}_{1:n}, \theta^{(j)} \sim N(\mathbf{m}_n^{(j)}, \sigma^{2(j)} \mathbf{C}_n^{(j)}). \quad (16)$$

Then, the predictive response at iteration j , $\mathbf{y}_{n+r}^{(j)}$ can be drawn from (15), that is,

$$\begin{aligned} \mathbf{y}_{n+r} \mid \mathbf{y}_{1:n}, \theta^{(j)} \\ \sim N \left(\mathbf{F}_{t+r}^{(j)'} \mathbf{m}_n^{(j)}, \sigma^{2(j)} \right. \\ \left. \left\{ \mathbf{F}_{t+r}^{(j)'} \left(\mathbf{C}_n^{(j)} + r \mathbf{W}^{(j)} \right) \mathbf{F}_{t+r}^{(j)} \right. \right. \\ \left. \left. + \exp \left(-\frac{\mathbf{V}}{\lambda^{(j)}} \right) \right\} \right). \end{aligned} \quad (17)$$

Consequently, the predictive responses are obtained by the sample means of $\{\mathbf{y}_{n+r}^{(j)} : j = 1, \dots, J\}$ ($J = 500$, where J denotes the total number of iteration after burn-in period; $r = 1, \dots, 24$). The empirical predictive intervals at the 95% nominal level can be obtained as the corresponding sample quantiles.

3. Case Study

This section implements the forecasting methods in the last section for one Chicago summer (from May 1 to August 31) using data for that urban area taken from the EPA’s AQS database (2000). These extracted data come from fourteen irregularly distributed monitoring stations measuring hourly ozone concentrations in parts per billion (ppb), which, to assure the validity of our Gaussian model assumptions, are square-root-transformed as noted in the last section. Each has few missing values under the EPA 1997 Standard in 1997 (i.e., 80 parts per billion for the eight-hour ground level ozone concentrations) during the overall time span across all available sites in this region.

To assess the model’s performance for temporal forecasting, 14 sites are selected as “gauged” sites (i.e., $g = 14$) and their observed responses on Day 121 are set aside as test values. Figure 1 shows the geographical locations of these fourteen gauged sites.

To explore these data further, weekday and hourly effects were computed for each site by averaging the transformed hour values over each of the seven weekdays over the whole summer. We found these effects to be very similar from one gauged site to the next. Thus, since “bloc” is the unit of time t , our approach puts the appropriate zero-one elements into the \mathbf{Z} to mark off the progression of blocs as the t progressed. Baseline hourly effects are represented in the hypermean function and are automatically fitted by software, EnviroStat.1.0.1. This then represents the overall diurnal pattern, while allowing site-specific deviations within the model.

3.1. Results and Comparisons. The two methods considered in this paper were applied at all fourteen gauged sites (GSs for short) to predict the twenty-four left out (test) and square-root-transformed hourly observations on Day 121. For all sites, plots showing the observations during the six days leading up to test Day 121, as well as the twenty-four forecasts by both methods for that day, may be seen in Dou et al. [19].

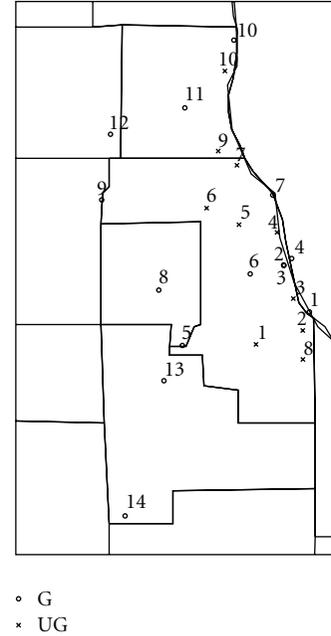


FIGURE 1: Geographical locations for the Chicago AQS database (2000), where the latitude and longitude are measured in degrees. (G: gauged sites; UG: ungauged sites).

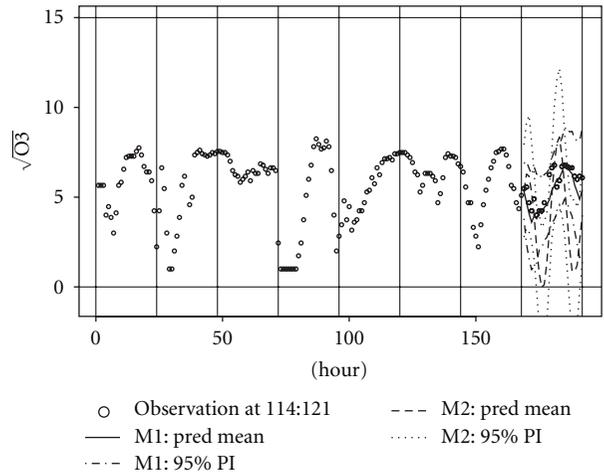


FIGURE 2: The observed square-root of ozone concentrations ($\sqrt{\text{ppb}}$) from Day 114 to Day 121, the predicted values using M1 and M2, and their 95% pointwise predictive intervals at GS 1.

These plots also include the 95% pointwise predictive intervals for that day for M1 and M2. For brevity, we include only figures that present some noteworthy features of these predictors.

To begin, Figure 2 for GS 1 shows a flat ozone field at this location on Day 119 and strongly varying one on Day 120 with two peaks. M1 does better than M2 in following not only the overall trend but tracking the turns quite well. All twenty-four of the test values lie within the 95% predictive credibility band although that for hour 12 overlaps the upper boundary. M2 follows ozone’s peaks fairly well,

but it forecasts valleys that do not turn up—the strong structure provided by the sines and cosines reduces this model’s flexibility and capacity to track the series in this case. Those harmonics point to two twelve-hour cycles despite the general lack of same in the observed series during the days before.

In contrast to the case of spatial prediction [8] where generally four peaks are seen in the sinusoidal curves bounding the DLM’s 95% predictive credibility bands, here there are just two. In fact, four peaks would be expected since the mean model’s components of variation has sines and cosines that are squared when they enter into the posterior variance. Thus their valleys turn into peaks, one every six hours since both twelve- and twenty-four hour cycles are present in the mean. So why only two?

The answer seems to lie in the fact that the random coefficients for the twelve hour components of variance in the forecast at any monitoring site, is not as uncertain in forecasting at that site than they are in spatial prediction at other sites, which may be a substantial distance from the monitoring sites. Consequently these components, although small, would have large posterior variances in the predictor than these components of the forecaster at one of the monitoring sites. In other words, much more information is available in the data leading up to the last day for the forecasts for the test values at that site than is available at a remote and unmonitored sites.

Incidentally, the lower bounds for M2 forecasts can go below zero so in practice would need to be truncated. Moreover, few of the test values lie within the 95% credibility band.

Huerta et al. [9] give similar plots for three stations in their case study for Mexico City using the method that led to our method M2. These plots differ from ours in that their plotted ordinates represent ozone while ours represent square-root-transformed ozone. We elected to keep the latter since that is the scale on which our analysis was done and hence the one that provides maximal diagnostic benefit. Furthermore our square root scale does not risk exaggerating the observed differences between both methods being considered. At the same time, we recognize the practical importance of publishing forecasts on the ozone rather than square-root ozone scale, and hence Table 1 presents comparisons on the former scale.

Huerta et al. [9] focus on just one week, so their forecasts are for their last Day 7, based on the six preceding days (whose observed hourly concentrations are plotted). In contrast, since our forecasts are based on the entire summer, our last day is Day 121 and we plot the observed values for this day (as well as the preceding seven). Although the daily amplitudes of the sinusoids in Huerta et al. [9] vary from day to day, the periodicity is very consistent over those days, with two fairly distinct peaks each day. In contrast, the ozone patterns over the seven days preceding our forecast Day 121, differ markedly from one day to another. Thus while the data series for Day 120 shows two very distinct peaks, that for Day 116 is nearly flat, and that for Day 118 (essentially) shows only one peak, after a monotone increasing trend rising to the end of that day. Finally Day 7 in their plot shows

TABLE 1: The root-mean-square-predictive error (RMSPE) of the one-day-ahead prediction at fourteen gauged sites by using M1 and M2. M1 dominates M2 in all but 1 case.

Gauged site	RMSPE (M1)	RMSPE (M2)
1	0.71	3.06
2	0.63	2.72
3	0.63	2.16
4	0.70	2.06
5	1.73	1.61
6	0.86	1.52
7	0.47	2.05
8	1.00	2.73
9	0.77	2.65
10	0.71	2.35
11	0.70	3.04
12	0.67	1.85
13	0.85	3.50
14	1.04	2.37

good agreement at all three of their stations, between their “predictive median” and the test data values. In contrast, our averages of 500 MCMC generated predicted responses for M2 disagree markedly with many of the hourly test values on Day 121. We would conjecture that these discrepancies occur because the data series on Day 121 tends to be much flatter than on all the preceding days except Day 116. (It may also derive from the additional (temperature) data Huerta et al. [9] had to enhance their forecasts that we did not have).

We do not see in any of the plots in Huerta et al. [9], the four peaks seen in some of our 95% credibility bands, suggesting that the uncertainty in the coefficients of the twelve-hour cyclical components is quite well resolved by the six days of data preceding their test day. The credibility bands for their version of M2 contain all the test values for all three of their sites, as do M2’s bands in Figure 2. (However, that is not the case for all fourteen of our stations as noted below. Moreover, their retrospective 95% credibility bands for all three sites are too narrow and fail to contain a large fraction of the observed values on a number of their days, e.g., on Sep 10 and Sep 12 at the Xalostoc site).

A summary very similar to that above for GS 1 also applies to the omitted figures for GS’s 2–4, 7–13, although for GS 11 M1 forecasts diverge from the test values over the final four hours, and for GS 13 M1 underestimates those test values in the middle of the day.

Figure 3 shows that GS 5 is different. Although M2’s forecast series has two peaks of moderate height, the bounds for the credibility bands have four of them consistent with two twelve-hour cycles. None of the forecast series tracks the series of test values well.

Summaries similar to that for GS 5 applies to GS 6 with the exception that M1 forecasts track the test value series quite well unlike M2 and to GS 14, where both methods underestimate the test values, M1 being closer overall than M2.

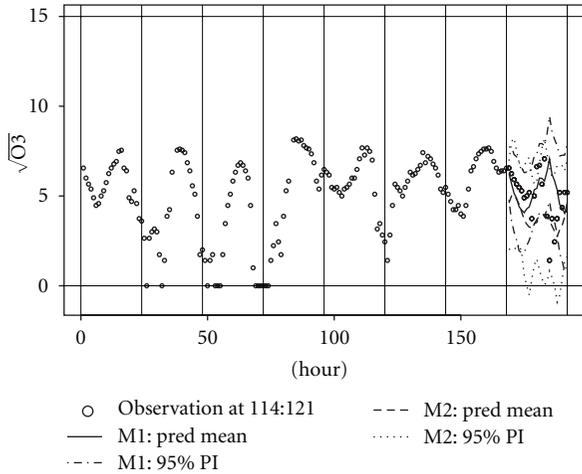


FIGURE 3: The observed square-root of ozone concentrations ($\sqrt{\text{ppb}}$) from Day 114 to Day 121, the predicted values using M1 and M2, and the 95% pointwise predictive intervals using M1 and M2 at GS 5.

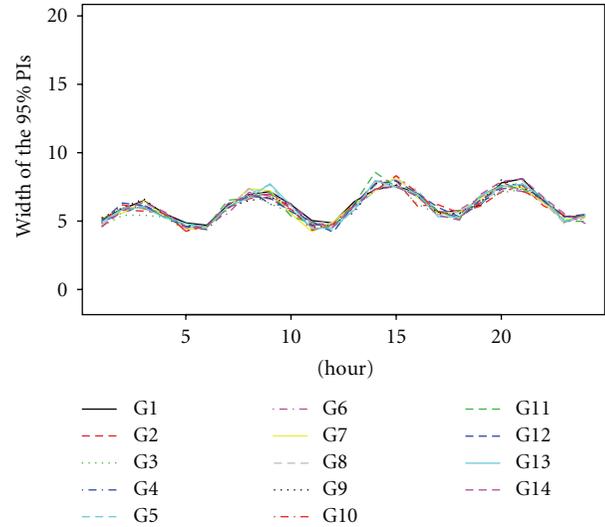


FIGURE 5: The width of the 95% pointwise predictive intervals (PIs) of the one-day-ahead prediction at 14 gauged sites using M2.

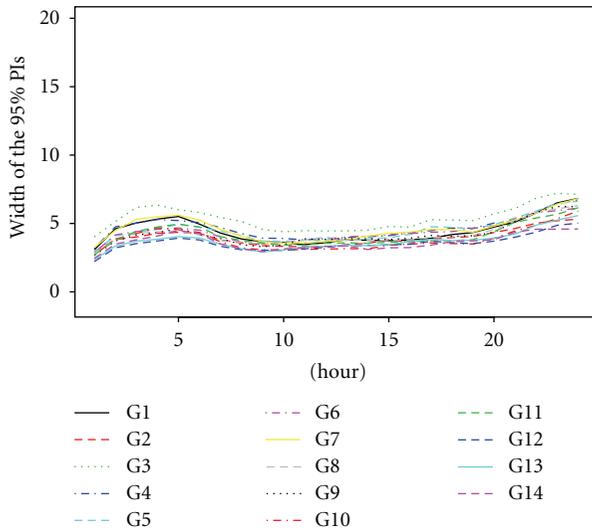


FIGURE 4: The width of the 95% pointwise predictive intervals of the one-day-ahead prediction at 14 gauged sites using M1.

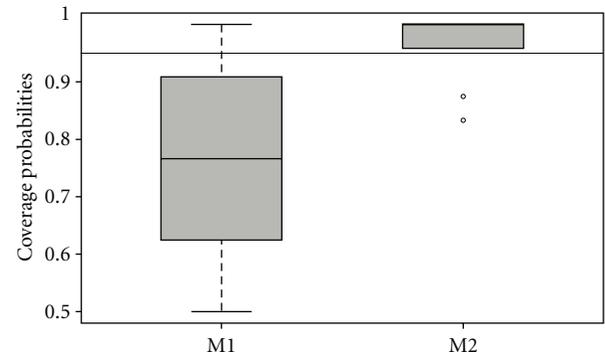


FIGURE 6: Boxplots of the coverage probabilities using M1 and M2 at the 95% nominal level.

Figure 4 plots the width of the 95% pointwise predictive credibility bands generated by the BSP at each of the twenty-four hours of Day 121. Starting from around 9 AM, these bands tend to increase and continue to do so until the last hour at 11 PM, reflecting the increasing uncertainty about the forecasts since increasingly fewer responses are observed as time increasing.

Figure 5 is a similar plot to that above, but this one for M2 instead of M1. These lengths are close to each other for various sites, exhibiting a wiggly periodic behavior across all gauged sites, a characteristic previously observed in Dou et al. [8, 18]. Although these lengths are very close to each other, M2 actually underestimates the predictive variances at gauged sites as seen in Figure 6, which shows the

coverage probabilities of M1 and M2, and also shows a slightly overestimated predictive variance for M1, at the 95% nominal level.

Table 1 presents the root-mean-square-predictive error (RMSPE) of the predictive responses on the 121st day at each one of fourteen gauged sites using these two approaches. At GS j , the RMSPE of the prediction at hour h can be computed by

$$\text{RMSPE}^j = \sqrt{\frac{1}{24} \sum_{h=1}^{24} (\text{PRED}_h^j - \text{OBS}_h^j)^2}, \quad (18)$$

where PRED_h^j is the predictive response at hour h of Day 121 and OBS_h^j , the corresponding observed response at the same hour, same day, and same site. M2 has a larger RMSPE over all the gauged sites compared with M1. M1 has the smallest RMSPE across most gauged sites.

4. Discussion and Conclusions

For forecasting ground-level hourly ozone concentrations in a Chicago summer, M1 seems better than M2. It seems more accurate, and its 95% predictive credibility interval is better calibrated. However, in any practical application M1 and M2 would need to be assessed in the same manner as in this paper before making a final selection. It should be noted that a new model also based on the dynamic linear approach has been proposed by Sahu et al. [11] for ozone modeling. It would be interesting to compare this new approach with the methods in this paper.

Those methods M1 and M2 are two quite different approaches to modeling space-time process and comparing and contrasting them at a more fundamental level seems worthwhile. To begin, both are quasi-Bayesian models in that they rely on some preliminary data analyses. Thus the diurnal cycles are identified for the M2 mean function, while regional non-site-specific weekday effects are found for M1. Both methods can then incorporate predictors or covariates in their parametric mean functions with random coefficients as well as reflect diurnal patterns of variation. M1 proceeds with this in two steps. First, regional time-dependent covariates or predictors are identified for the construction of the design matrix \mathbf{Z} , in his case day-of-the-week bloc effects. Secondly, it estimates hypermeans for this predictor's coefficients as well as for the multivariate bloc vector's responses, in this case the hourly effects. At the same time, it allows site-specific deviations from these baseline estimates through the random mean coefficients. The case study suggests that these random coefficients capture site-specific hourly effects quite well. In contrast, M2 builds regional features and daily variation into its mean response function through the incorporation of mean trends and periodic components before implementation. Thus, its prescribed mean is fairly structured with Fourier components to describe daily 12- and 24-hour cycles. In contrast, M1 incorporates all general trends and diurnal patterns in the hypermean for its random coefficients and then allows site-specific deviations from this hypermean at all sites. The former is more flexible than the latter, in allowing the coefficients to change over time, but the second is more flexible than the first in allowing an arbitrary shape for the daily pattern of variation and allowing site-specific trends.

Both approaches put spatial covariance structures on their mean models as well as on the residuals. In contrast to M2, M1 does not require a nonstationary spatial covariance structure, and the form of the spatial covariance matrix is completely unspecified at level one of the Bayesian hierarchy. This is not important for the Chicago analysis where the spatial ozone field is quite flat, but we believe it would be an important difference between the models in say Los Angeles or Seattle, where M1 would be favored. M2 prescribes its temporal correlation structure through the structure of its mean function, notably a random walk model for its model coefficient vector. In contrast, M1's 24-hour bloc covariance matrix is unspecified at level one of the hierarchical model, leaving the data a big role in determining its form. However, this feature comes at the price of an assumption that

the 24 autocovariance matrix is separable from the spatial covariance. Moreover, the covariance is constant over time. Both of these assumptions are limitations of M1.

Both M1 and M2 rely on both autocorrelation as well as temporal correlation for forecasting next day ozone levels. We believe responses will be somewhat autocorrelated from day to day and that feature can be exploited to enhance the forecasting performance. As formulated, M2 does borrow that additional strength, where M1 loses in the way we have implemented is parent, the BSP, by dropping a day to avoid having to formulate a multivariate time series model for the vectors of daily bloc responses. However, this is not strictly necessary. The more general version of the BSP approach does allow for that correlation, and in principle we would have estimated the hyperparameters that approach, suffered the consequences of possible misspecification and increased the computational burden of implementing M1. Thus M1 was formulated under the assumption of uncorrelated responses between days, unlike M2 which makes no such assumption, with the goal of ensuring timely 24 ahead ozone forecasts.

M2 has a much more general parent, in the dynamic linear model (DLM) and undoubtedly other implementations of the DLM could be made that retained its positive features while overcoming some of the limitations of M2 noted above. For example, a nonstationary spatial covariance could undoubtedly be used. As well the random walk model which has serious limitations could be replaced by say a more reasonable model like an AR(1), albeit with an added parameter burden. That would in turn further restrict the number of monitoring sites it could realistically handle in an urban area. As it stands, M1 computational efficiency enables it to handle a much larger number of sites than M2 in an urban area such as the greater Los Angeles area, which has 30 sites well beyond the reach of M2.

Although any ozone forecast for hourly concentrations 24 hours in advance cannot be much better than the baseline estimate, we have included Case 1 for completeness. Its Equation (9) is the basis of the forecast for that case. That equation actually gives a predictive distribution for all the hourly concentrations in Day 121 and could be used to forecast them all. However, the forecasts would not be very good compared to those given by the method in Case 2. The reason is that the latter exploits the strong AR(2) structure in any consecutive sequence of 24 hourly responses, unlike the former which assumes the daily vectors of responses are conditionally independent as an approximation made mainly for computational expediency. Thus for all other hours the forecaster in Case 2 should be used when the data in Day 120 are available. Note that within the Bayesian framework, the unconditional distribution of 24 dimensional response vectors are not independent, a feature that Case 1 exploits.

We have not considered the realistic case where only a limited number of hours of Day 120 data are available. That is because this case would be just a formalistic extension of methods M1 and M2.

Finally, we would emphasize that the results in Section 2.1 have been generalized in that section, another way in which M1 and M2 go beyond the limited application

in the Case Study. Moreover, our approach for turning BSP into the temporal forecasting tool in M1 could well be used for any univariate time series. The approach would avoid the need to capture autocorrelation at fine temporal scales, something that can be difficult to do as in the ozone case, where the AR structure varies over the day.

Overall, we have found that for forecasting Chicago's next day ozone concentration levels, M1 would be more practical and more accurate than M2. With its well-calibrated forecast intervals, it seems a promising methodology for practical application.

Appendix

A. Supplementary Results

A.1. Theorem A.1 and Its Proof

Theorem A.1. Let $\mathbf{Y}^{[g^m]} = ((\mathbf{Y}_{N+r,1:p}^{[g_{1:g}^m]})', \dots, (\mathbf{Y}_{N+1,1:p}^{[g_{1:g}^m]})', (\mathbf{Y}_{N,1:p}^{[g_{1:g}^m]})')' : (r+1) \times gp$ and $\mathbf{Y}^{[g^o]} = \mathbf{Y}_{1:(N-1),1:p}^{[g_{1:p}^o]} : (N-1) \times gp$. Then, one has the following predictive distributions:

$$(i) (\mathbf{Y}^{[g^m]} \mid \mathbf{Y}^{[g^o]}, \mathcal{H}) \sim t_{(r+1) \times gp}(\check{\mu}_{(u|g)}, \check{\Phi}_{(u|g)} \otimes \check{\Psi}_{(u|g)}, \check{\delta}_{(u|g)}), \text{ where}$$

$$\begin{aligned} \check{\mu}_{(u|g)} &= \mu_{(1)} + \mathbf{A}_{12} \mathbf{A}_{22}^{-1} (\mathbf{Y}^{[g^o]} - \mu_{(2)}), \\ \check{\Phi}_{(u|g)} &= \frac{\delta_1 - gp + 1}{\delta_1 - gp + N + 1} \mathbf{A}_{11 \circ 2}, \\ \check{\Psi}_{(u|g)} &= \frac{1}{\delta_1 - gp + 1} \\ &\quad \times \left\{ \mathbf{\Lambda}_1 \otimes \mathbf{\Omega} + (\mathbf{Y}^{[g^o]} - \mu_{(2)})' \right. \\ &\quad \left. \times \mathbf{A}_{22}^{-1} (\mathbf{Y}^{[g^o]} - \mu_{(2)}) \right\}, \\ \check{\delta}_{(u|g)} &= \delta_1 - gp + N + 1. \end{aligned} \quad (A.1)$$

(ii) The predictive distribution of $Y_{N+r,p}^{[g_j^m]}$, the p th unobserved response on the $(N+r)$ th day at gauged site j , is t -distributed:

$$Y_{N+r,p}^{[g_j^m]} \mid \mathbf{Y}_{1:N,1:p}^{[g_{1:g}^o]}, \mathcal{H} \sim t_{\check{\delta}} \left(\left(\mathbf{e}_r^j \right)' \check{\mu} \mathbf{e}_g^j, \frac{\check{\delta}}{\check{\delta} - 2} \left(\mathbf{e}_r^j \right)' \check{\Phi} \mathbf{e}_r^j \left(\mathbf{e}_g^j \right)' \check{\Psi} \mathbf{e}_g^j \right), \quad (A.2)$$

where

$$\begin{aligned} \check{\mu} &= \check{\mu}_{1r} + \check{\mathbf{B}}_{12} \check{\mathbf{B}}_{22}^{-1} \left(\mathbf{Y}_{N,1:p}^{[g_{1:g}^o]} - \check{\mu}_{2r} \right), \\ \check{\Phi} &= \frac{1}{\check{\delta}_{(u|g)} + 1} \check{\mathbf{B}}_{11 \circ 2}, \end{aligned}$$

$$\begin{aligned} \check{\Psi} &= \check{\Psi}_{(u|g)} \left[\mathbf{I}_{gp} + \check{\Psi}_{(u|g)}^{-1} \left(\mathbf{Y}_{N,1:p}^{[g_{1:g}^o]} - \check{\mu}_{2r} \right)' \right. \\ &\quad \left. \times \check{\mathbf{B}}_{22}^{-1} \left(\mathbf{Y}_{N,1:p}^{[g_{1:g}^o]} - \check{\mu}_{2r} \right) \right], \end{aligned}$$

$$\check{\delta} = \check{\delta}_{(u|g)} + 1. \quad (A.3)$$

Proof. The result is straightforward by Theorem 1, where $w = r + 1$ and $n = N + r$; decompose $\check{\mu}_{(u|g)}$ and $\delta_{(u|g)} \check{\Phi}_{(u|g)}$ as follows

$$\check{\mu}_{(u|g)} = \begin{pmatrix} \check{\mu}_{1r} \\ \check{\mu}_{2r} \end{pmatrix} : \begin{pmatrix} r \times gp \\ 1 \times gp \end{pmatrix}, \quad (A.4)$$

$$\delta_{(u|g)} \check{\Phi}_{(u|g)} = \begin{pmatrix} \check{\mathbf{B}}_{11} & \check{\mathbf{B}}_{12} \\ \check{\mathbf{B}}_{21} & \check{\mathbf{B}}_{22} \end{pmatrix} : \begin{pmatrix} r \times r & r \times 1 \\ 1 \times r & 1 \times 1 \end{pmatrix}.$$

Hence, we have

$$\mathbf{Y}_{(N+1):(N+r),1:p}^{[g_j^m]} \mid \mathbf{Y}_{N,1:p}^{[g_{1:g}^o]}, \mathbf{Y}_{1:(N-1),1:p}^{[g_{1:g}^o]}, \mathcal{H} \sim t_{r \times gp}(\check{\mu}, \check{\Phi} \otimes \check{\Psi}, \check{\delta}), \quad (A.5)$$

where $\check{\mu}$, $\check{\Phi}$, $\check{\Psi}$, and $\check{\delta}$ are given in Theorem A.1.

We have $(\mathbf{e}_r^j)' \mathbf{Y}_{(N+1):(N+r),1:p}^{[g_j^m]} \mathbf{E}_1 \mathbf{e}_g^j = Y_{N+r,p}^{[g_j^m]}$, that is, the unobserved response of the last hour of the $(N+r)$ th day at Gauged Site j ($j = 1, \dots, g$). Hence, we have

$$Y_{N+r,p}^{[g_j^m]} \sim t_{1 \times 1} \left(\left(\mathbf{e}_r^j \right)' \check{\mu} \mathbf{e}_g^j, \left(\mathbf{e}_r^j \right)' \check{\Phi} \mathbf{e}_r^j \otimes \left(\mathbf{e}_g^j \right)' \check{\Psi} \mathbf{e}_g^j, \check{\delta} \right), \quad (A.6)$$

that is, $t_{\check{\delta}} \left(\left(\mathbf{e}_r^j \right)' \check{\mu} \mathbf{e}_g^j, \left(\check{\delta} / (\check{\delta} - 2) \right) \left(\mathbf{e}_r^j \right)' \check{\Phi} \mathbf{e}_r^j \left(\mathbf{e}_g^j \right)' \check{\Psi} \mathbf{e}_g^j \right)$. \square

A.2. Theorem A.2 and Its Proof

Theorem A.2. Let $\mathbf{Y}^{[g^m]} = ((\mathbf{Y}_{N+r-1}^{[g_{1:g}^m]})', \dots, (\mathbf{Y}_{N+1}^{[g_{1:g}^m]})', (\mathbf{Y}_N^{[g_{1:g}^m]})')' : r \times gp$, where $\mathbf{W}_i^{[g_j^m]} = (Y_{i,k+1}^{[g_j^m]}, \dots, Y_{i,p}^{[g_j^m]}, Y_{i+1,1}^{[g_j^m]}, \dots, Y_{i+1,k}^{[g_j^m]}) : 1 \times gp$, and $\mathbf{Y}_N^{[g_j^m]} = (Y_{N,k+1}^{[g_j^m]}, \dots, Y_{N,p}^{[g_j^m]}, Y_{N+1,1}^{[g_j^m]}, \dots, Y_{N+1,k}^{[g_j^m]}) : 1 \times gp$, for $i = N+1, \dots, N+r-1$ and $j = 1, \dots, g$. We also let $\mathbf{Y}^{[g^o]} = \mathbf{Y}_{1:(N-1)}^{[g_{1:g}^o]} : (N-1) \times gp$, where $\mathbf{Y}_i^{[g_j^o]} = (Y_{i,k+1}^{[g_j^o]}, \dots, Y_{i,p}^{[g_j^o]}, Y_{i+1,1}^{[g_j^o]}, \dots, Y_{i+1,k}^{[g_j^o]}) : 1 \times gp$. One then has the following predictive distributions:

(i)

$$\mathbf{Y}^{[g^m]} \mid \mathbf{Y}^{[g^o]}, \mathcal{H} \sim t_{r \times gp}(\check{\mu}_{(u|g)}, \check{\Phi}_{(u|g)} \otimes \check{\Psi}_{(u|g)}, \check{\delta}_{(u|g)}), \quad (A.7)$$

where

$$\begin{aligned}\tilde{\boldsymbol{\mu}}_{(u|g)} &= \boldsymbol{\mu}_{(1)} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\left(\mathbf{Y}^{[g^{\circ 1}]} - \boldsymbol{\mu}_{(2)}\right) : r \times gp, \\ \tilde{\boldsymbol{\Phi}}_{(u|g)} &= \frac{\delta_1 - gp + 1}{\delta_1 - gp + N}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}), \\ \tilde{\boldsymbol{\Psi}}_{(u|g)} &= \frac{1}{\delta_1 - gp + 1} \\ &\quad \times \left\{ \mathbf{\Lambda}_1 \otimes \boldsymbol{\Omega} + \left(\mathbf{Y}^{[g^{\circ 1}]} - \boldsymbol{\mu}_{(2)}\right)' \right. \\ &\quad \left. \times \mathbf{A}_{22}^{-1}\left(\mathbf{Y}^{[g^{\circ 1}]} - \boldsymbol{\mu}_{(2)}\right) \right\}, \\ \tilde{\delta}_{(u|g)} &= \delta_1 - gp + N.\end{aligned}\tag{A.8}$$

(ii)

$$\begin{aligned}&\left(\mathbf{Y}_{(N+1):(N+r-1),1:p}^{[g_1^m]}, \mathbf{Y}_{N+r,1:k}^{[g_1^m]} \mid \mathbf{Y}_{1:N,1:p}^{[g_1^{\circ 1}]} \mid \mathcal{H}\right) \\ &\propto \prod_{j=1}^g p\left(\mathbf{T}_{1j}^r \mid \mathbf{Y}_{N+1,1:k}^{[g_j^m]}, \mathbf{Y}_{1:N,1:p}^{[g_1^{\circ 1}]} \mid \mathcal{H}\right) \\ &\quad \times \left(\mathbf{Y}_{N+1,1:k}^{[g_j^m]} \mid \mathbf{Y}_{1:N,1:p}^{[g_1^{\circ 1}]} \mid \mathcal{H}\right) \\ &\sim \prod_{j=1}^g t_{(r-1) \times p}\left(\tilde{\boldsymbol{\mu}}_{1j}^*, \tilde{\boldsymbol{\Phi}}_j^* \otimes \tilde{\boldsymbol{\Psi}}_j^*, \tilde{\delta}_{(u|g)} + 1\right) \\ &\quad \times t_{1 \times k}\left(\tilde{\boldsymbol{\mu}}_{2j}^*, \tilde{\boldsymbol{\Phi}}_{2j}^* \otimes \tilde{\boldsymbol{\Psi}}_{2j}^*, \tilde{\delta}_{(u|g)} + p - k\right),\end{aligned}\tag{A.9}$$

where $\tilde{\boldsymbol{\mu}}_{1j}^*$, $\tilde{\boldsymbol{\Phi}}_j^*$, and $\tilde{\boldsymbol{\Psi}}_j^*$ are given in (A.16) and $\tilde{\boldsymbol{\mu}}_{2j}^*$, $\tilde{\boldsymbol{\Phi}}_{2j}^*$, and $\tilde{\boldsymbol{\Psi}}_{2j}^*$, in (A.19).

Proof. (i) The result is straightforward by Theorem 1 where $w = r$ and $n = N + r - 1$;

(ii) denote $\mathbf{E}_{2j} = (\mathbf{e}_{gp}^{(j-1)p+1}, \dots, \mathbf{e}_{gp}^{jp}) : gp \times p$ for $j = 1, \dots, g$. And let $\mathbf{E}_3 = (\mathbf{e}_p^1, \dots, \mathbf{e}_p^1) : p \times p$. We will have the following results (details can be referred to in [19]):

$$\begin{aligned}\tilde{\mathbf{Y}}^{[g^m]} &= \mathbf{Y}^{[g^m]}\mathbf{E}_{2j}\mathbf{E}_3 \\ &= \begin{pmatrix} \mathbf{Y}_{N+r,k}^{[g_j^m]} & \dots & \mathbf{Y}_{N+r,1}^{[g_j^m]} & \mathbf{Y}_{N+r-1,1}^{[g_j^m]} & \dots & \mathbf{Y}_{N+r-1,k+1}^{[g_j^m]} \\ \vdots & & \vdots & \vdots & & \vdots \\ \mathbf{Y}_{N+1,k}^{[g_j^m]} & \dots & \mathbf{Y}_{N+1,1}^{[g_j^m]} & \mathbf{Y}_{N,1}^{[g_j^{\circ 1}]} & \dots & \mathbf{Y}_{N,k+1}^{[g_j^{\circ 1}]} \end{pmatrix} : r \times p,\end{aligned}\tag{A.10}$$

for $j = 1, \dots, g$. From (i) in Theorem A.2, we have

$$\tilde{\mathbf{Y}}^{[g^m]} \mid \mathbf{Y}^{[g^{\circ 1}]}, \mathcal{H} \sim t_{r \times p}\left(\tilde{\boldsymbol{\mu}}_j, \tilde{\boldsymbol{\Phi}}_{(u|g)} \otimes \tilde{\boldsymbol{\Psi}}_j, \tilde{\delta}_{(u|g)}\right),\tag{A.11}$$

where

$$\begin{aligned}\tilde{\boldsymbol{\mu}}_j &= \tilde{\boldsymbol{\mu}}_{(u|g)}\mathbf{E}_{2j}\mathbf{E}_3, \\ \tilde{\boldsymbol{\Psi}}_j &= \mathbf{E}_3'\mathbf{E}_{2j}'\tilde{\boldsymbol{\Psi}}_{(u|g)}\mathbf{E}_{2j}\mathbf{E}_3.\end{aligned}\tag{A.12}$$

We first decompose $\tilde{\mathbf{Y}}^{[g^m]}$, $\tilde{\boldsymbol{\mu}}_j$ and $\tilde{\delta}_{(u|g)}\tilde{\boldsymbol{\Phi}}_{(u|g)}$ as follows:

$$\begin{aligned}\tilde{\mathbf{Y}}^{[g^m]} &= \begin{pmatrix} \mathbf{T}_{1j}^r \\ \mathbf{T}_{2j}^r \end{pmatrix} : \begin{pmatrix} (r-1) \times p \\ 1 \times p \end{pmatrix}, \\ \tilde{\boldsymbol{\mu}}_j &= \begin{pmatrix} \tilde{\boldsymbol{\mu}}_{1j} \\ \tilde{\boldsymbol{\mu}}_{2j} \end{pmatrix} : \begin{pmatrix} (r-1) \times p \\ 1 \times p \end{pmatrix}, \\ \tilde{\delta}_{(u|g)}\tilde{\boldsymbol{\Phi}}_{(u|g)} &= \begin{pmatrix} \tilde{\boldsymbol{\Phi}}_{11} & \tilde{\boldsymbol{\Phi}}_{12} \\ \tilde{\boldsymbol{\Phi}}_{21} & \tilde{\boldsymbol{\Phi}}_{22} \end{pmatrix} : \begin{pmatrix} (r-1) \times (r-1) & (r-1) \times 1 \\ 1 \times (r-1) & 1 \times 1 \end{pmatrix}.\end{aligned}\tag{A.13}$$

Consequently, we have

(a)

$$\mathbf{T}_{2j}^r \mid \mathbf{Y}^{[g^{\circ 1}]}, \mathcal{H} \sim t_{1 \times p}\left(\tilde{\boldsymbol{\mu}}_{2j}, \tilde{\boldsymbol{\Phi}}_{22} \otimes \tilde{\boldsymbol{\Psi}}_j, \tilde{\delta}_{(u|g)}\right),\tag{A.14}$$

(b)

$$\mathbf{T}_{1j}^r \mid \mathbf{T}_{2j}^r, \mathbf{Y}^{[g^{\circ 1}]}, \mathcal{H} \sim t_{(r-1) \times p}\left(\tilde{\boldsymbol{\mu}}_{1j}^*, \tilde{\boldsymbol{\Phi}}_j^* \otimes \tilde{\boldsymbol{\Psi}}_j^*, \tilde{\delta}_{(u|g)} + 1\right),\tag{A.15}$$

where

$$\begin{aligned}\tilde{\boldsymbol{\mu}}_{1j}^* &= \tilde{\boldsymbol{\mu}}_{1j} + \tilde{\boldsymbol{\Phi}}_{12}\tilde{\boldsymbol{\Phi}}_{22}^{-1}\left(\mathbf{T}_{2j}^r - \tilde{\boldsymbol{\mu}}_{2j}\right), \\ \tilde{\boldsymbol{\Phi}}_j^* &= \frac{\tilde{\delta}_{(u|g)}}{\tilde{\delta}_{(u|g)} + 1}\left(\tilde{\boldsymbol{\Phi}}_{11} - \tilde{\boldsymbol{\Phi}}_{12}\tilde{\boldsymbol{\Phi}}_{22}^{-1}\tilde{\boldsymbol{\Phi}}_{21}\right),\end{aligned}\tag{A.16}$$

$$\tilde{\boldsymbol{\Psi}}_j^* = \tilde{\boldsymbol{\Psi}}_j\left(\mathbf{I}_p + \tilde{\boldsymbol{\Psi}}_j^{-1}\left(\mathbf{T}_{2j}^r - \tilde{\boldsymbol{\mu}}_{2j}\right)\tilde{\boldsymbol{\Phi}}_{22}^{-1}\left(\mathbf{T}_{2j}^r - \tilde{\boldsymbol{\mu}}_{2j}\right)\right).$$

We then decompose \mathbf{T}_{2j}^r , $\tilde{\boldsymbol{\mu}}_{2j}^r$, and $\tilde{\boldsymbol{\Psi}}_j$ as follows:

$$\begin{aligned}\mathbf{T}_{2j}^r &= \begin{pmatrix} \mathbf{T}_{21}^j & \mathbf{T}_{22}^j \end{pmatrix} : \begin{pmatrix} 1 \times k & 1 \times (p-k) \end{pmatrix}, \\ \tilde{\boldsymbol{\mu}}_{2j}^r &= \begin{pmatrix} \boldsymbol{\mu}_{21}^j & \boldsymbol{\mu}_{22}^j \end{pmatrix} : \begin{pmatrix} 1 \times k & 1 \times (p-k) \end{pmatrix}, \\ \tilde{\boldsymbol{\Psi}}_j &= \begin{pmatrix} \tilde{\boldsymbol{\Psi}}_{11}^j & \tilde{\boldsymbol{\Psi}}_{12}^j \\ \tilde{\boldsymbol{\Psi}}_{21}^j & \tilde{\boldsymbol{\Psi}}_{22}^j \end{pmatrix} : \begin{pmatrix} k \times k & k \times (p-k) \\ (p-k) \times k & (p-k) \times (p-k) \end{pmatrix}.\end{aligned}\tag{A.17}$$

Hence the predictive distribution of \mathbf{T}_{21}^j , that is, $\mathbf{Y}_{N+1,1:k}^{[g_j^m]}$ is given by

$$\mathbf{Y}_{N+1,1:k}^{[g_j^m]} \mid \mathbf{Y}_{1:N,1:p}^{[g^{\circ 1}]}, \mathcal{H} \sim t_{1 \times k}\left(\tilde{\boldsymbol{\mu}}_{2j}^*, \tilde{\boldsymbol{\Phi}}_{2j}^* \otimes \tilde{\boldsymbol{\Psi}}_{2j}^*, \tilde{\delta}_{(u|g)} + p - k\right),\tag{A.18}$$

where

$$\begin{aligned}\tilde{\mu}_{2j}^* &= \mu_{21}^j + (\mathbf{T}_{22}^j - \mu_{22}^j) (\tilde{\Psi}_{22}^j)^{-1} \tilde{\Psi}_{21}^j, \\ \tilde{\Phi}_{2j}^* &= \frac{\tilde{\delta}_{(u|g)}}{\tilde{\delta}_{(u|g)} + p - k} \tilde{\Phi}_{22}^j \\ &\times \left(1 + \left(\tilde{\delta}_{(u|g)} \tilde{\Phi}_{22}^j \right)^{-1} (\mathbf{T}_{22}^j - \mu_{22}^j) \right. \\ &\quad \left. \times \left(\tilde{\Psi}_{22}^j \right)^{-1} (\mathbf{T}_{22}^j - \mu_{22}^j)' \right), \\ \tilde{\Psi}_2^* &= \tilde{\Psi}_{11}^j - \tilde{\Psi}_{12}^j \left(\tilde{\Psi}_{22}^j \right)^{-1} \tilde{\Psi}_{21}^j.\end{aligned}\quad (\text{A.19})$$

Therefore, we have

$$\begin{aligned}p\left(\mathbf{Y}_{(N+1):(N+r-1),1:p}^{[g_{1:g}^m]}, \mathbf{Y}_{N+r,1:k}^{[g_{1:g}^m]} \mid \mathbf{Y}_{1:N,1:p}^{[g_{1:g}^o]}, \mathcal{H}\right) \\ \propto \prod_{j=1}^g p\left(\mathbf{T}_{1j}^r \mid \mathbf{Y}_{N+1,1:k}^{[g_j^m]}, \mathbf{Y}_{1:N,1:p}^{[g_{1:g}^o]}, \mathcal{H}\right) \\ \times p\left(\mathbf{Y}_{N+1,1:k}^{[g_j^m]} \mid \mathbf{Y}_{1:N,1:p}^{[g_{1:g}^o]}, \mathcal{H}\right).\end{aligned}\quad (\text{A.20})$$

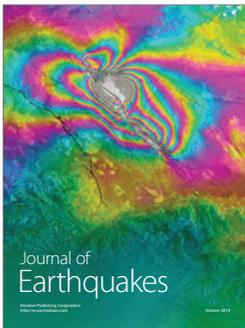
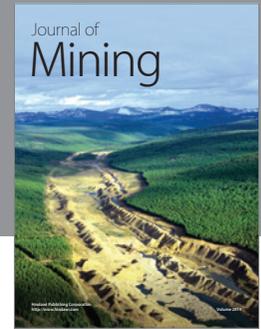
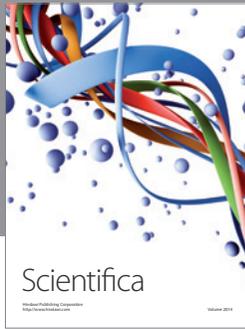
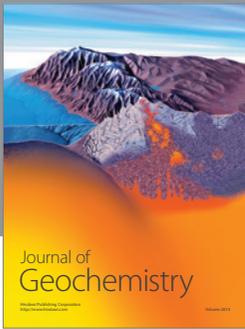
The predictive distribution for $\mathbf{Y}_{N+r,k}^{[g_{1:g}^m]}$ has no analytic form. \square

Acknowledgments

This work was partially supported by funding from the Pacific Institute of the Mathematical Sciences as well as the Natural Science and Engineering Research Council of Canada.

References

- [1] Ozone, *Air Quality Criteria for Ozone and Related Photochemical Oxidants*, US Environmental Protection Agency, 2005.
- [2] S. K. Sahu, S. Yip, and D. M. Holland, "A fast Bayesian method for updating and forecasting hourly ozone levels," *Environmental and Ecological Statistics*, vol. 18, no. 1, pp. 185–207, 2011.
- [3] J. V. Zidek, N. D. Le, and Z. Liu, "Combining data and simulated data for space-time fields: application to ozone," *Environmental and Ecological Statistics*, vol. 19, pp. 37–56, 2012.
- [4] N. D. Le and J. V. Zidek, "Interpolation with uncertain spatial covariances: a Bayesian alternative to Kriging," *Journal of Multivariate Analysis*, vol. 43, no. 2, pp. 351–374, 1992.
- [5] N. D. Le, W. Sun, and J. V. Zidek, "Bayesian multivariate spatial interpolation with data missing by design," *Journal of the Royal Statistical Society B*, vol. 59, no. 2, pp. 501–510, 1997.
- [6] J. Zidek, L. Sun, N. Le, and H. Özkaynak, "Contending with space-time interaction in the spatial prediction of pollution: vancouver's hourly ambient PM10 field," *Environmetrics*, vol. 13, no. 5–6, pp. 595–613, 2002.
- [7] N. D. Le and J. V. Zidek, *Statistical Analysis of Environmental Space-Time Processes*, Springer, New York, NY, USA, 2006.
- [8] Y. P. Dou, N. D. Le, and J. V. Zidek, "Modeling hourly ozone concentration fields," *Annals of Applied Statistics*, vol. 4, Article ID 11831213, 2010.
- [9] G. Huerta, B. Sansó, and J. R. Stroud, "A spatiotemporal model for Mexico City ozone levels," *Journal of the Royal Statistical Society C*, vol. 53, no. 2, pp. 231–248, 2004.
- [10] V. J. Berrocal, A. E. Gelfand, and D. M. Holland, "A bivariate space-time downscaler under space and time misalignment," *Annals of Applied Statistics*, vol. 4, pp. 1942–1975, 2010.
- [11] S. K. Sahu, A. E. Gelfand, and D. M. Holland, "High-resolution space-time ozone modeling for assessing trends," *Journal of the American Statistical Association*, vol. 102, no. 480, pp. 1221–1234, 2007.
- [12] D. Wraith, C. Alston, K. Mengersen, and T. Hussein, "Bayesian mixture model estimation of aerosol particle size distributions," *Environmetrics*, vol. 22, no. 1, pp. 23–34, 2011.
- [13] W. Meiring, P. Guttorp, and P. D. Sampson, "Space-time estimation of grid-cell hourly ozone levels for assessment of a deterministic model," *Environmental and Ecological Statistics*, vol. 5, no. 3, pp. 197–222, 1998.
- [14] N. D. Le and J. V. Zidek, *EnviRo.stat. A software for "Statistical Analysis of Environmental Space-Time Processes"*, 2006, <http://enviro.stat.ubc.ca/statisticalanalysis/>.
- [15] C. S. Hirtzel and J. E. Quon, "Statistical analysis of continuous ozone measurements," *Atmospheric Environment A*, vol. 15, no. 6, pp. 1025–1034, 1981.
- [16] C. K. Carter and R. Kohn, "On gibbs sampling for state space models," *Biometrika*, vol. 81, pp. 541–553, 1994.
- [17] W. West and J. Harrison, *Bayesian Forecasting and Dynamic Models*, Springer, 1997.
- [18] Y. P. Dou, N. D. Le, and J. V. Zidek, "A dynamic linear model for hourly ozone concentrations," Tech. Rep. 228, Department of Statistics, University of British Columbia, 2007.
- [19] Y. P. Dou, N. D. Le, and J. V. Zidek, "Temporal prediction with a Bayesian spatial predictor: an application to ozone fields," Tech. Rep. 249, Department of Statistics University of British Columbia, 2009.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

