The University of British Columbia

Department of Statistics

Technical Report #232

Calibrating Deterministic Modeling Output With Application Of Ozone Fields

Zhong Liu, Nhu D. Le, James V. Zidek

July 12, 2007

Department of Statistics

University of British Columbia

333-6356 Agricultural Road

Vancouver, BC

V6T 1Z2

Canada

Calibrating Deterministic Modeling Output With Application Of Ozone Fields

Zhong Liu[1], Nhu D. Le[2], James V. Zidek[3]

Abstract. In this technical report, we first review a two-step linear regression model which can adjust the outputs from a deterministic model of a random field according to measurements of that field. However this model does not incorporate spatial correlation leading us to propose an ad-hoc spatial-temporal alternative based on Kriging to deal with the spatial correlation. After demonstrating that alternative's promise, we go on to refine it through Bayesian hierarchical spatial-temporal model. The resulting Bayesian model can provide forecasts and spatial predictions along with their credible intervals. As an application, we analyze ozone data from the eastern and central USA. These predictions and forecasts are compared with those from other approaches.

1 Introduction

Scientific researchers have developed numerical models to simulate ozone concentrations using their knowledge of physical and chemical processes. The resulting deterministic models yield a single predetermined output which is always the same if the input is fixed. One such deterministic model is the AQM (air quality model) which can generate regional-scale ozone concentrations. It is the non-hydrostatic version of the MAQSIP (Multiscale Air Quality Simulation Platform) model described in Odman and Ingram (1996). The AQM can simulate hourly ozone concentrations based on a resolution of $6 \times 6\text{km}^2$. One input to the AQM is the output of another deterministic model MM5 described in Grell et al. (1995). Another input is the output of an emission model.

[1]Department of Statistics, U. of British Columbia. Email: zliu@stat.ubc.ca
[2]British Columbia Cancer Research Centre. Email: nle@bccrc.ca
[3]Department of Statistics, U. of British Columbia. Email: jim@stat.ubc.ca

Hourly measurements of the ozone concentration level are available at the Air Quality System (AQS) monitoring network. Both the measurements and modeling output are from May 15 to September 11, 1995, a 120-day period over the eastern and central part of the United States. The modeling output consists of the ozone concentration levels simulated from the deterministic model AQM. In this report, we choose the 78 stations with the fewest missing measurements. The modeling outputs from AQM are for grid cells of resolution 6×6 km^2. Each of these 78 stations is inside one grid cell and each of these 78 grid cell has only one station inside. If one grid cell has more than one stations inside then the measurements from these stations are averaged. Macthing grid cells and monitoring sites in this way enables an empirical assessment of the role of modeling - measurement similarities in spatial prediction. Some of the 78 stations have measurements in the past and we want to forecast the measurements into the future. The measurements at the rest are left as a validation dataset, which are to be predicted for empirical comparison. For these stations, we want to spatially predict their measurements. In the rest of this report, prediction means spatially predicting the measurements at the stations without measurements.

The rest of the report is organized as follows. Section 2 reviews the two-stepp linear regression model on which we base our approach. Section 3 extends this temporal model to an ad-hoc spatial-temporal model in order to assess such a model's potential. Based on the favorable results seen in Section 3, Section 4 develops a more rigorous spatial-temporal model in a Bayesian framework. Section 5 compares the prediction and forecast results between different approaches. The final section includes some discussion.

2 Two-step linear regression

This section reviews the two-step linear regression model proposed by Guillas et al. (2006). The strong linear correlation between the hourly measurements and modeling output at each station makes such a model seem natural. That correlation is seen in Figure 1, a histogram of the Pearson's correlation coefficients at all the 78 stations considered in this

report. Figure 1 shows those correlations to be bigger than 0.5 at most stations, pointing to a linear relationship between the measurements and modeling outputs.

The first model in the two-step linear regression procedure proposed by Guillas et al. (2006) relates the measurements, $\{O(t)\}$, to the model outputs $\{M(t)\}$ by

$$O(t) = c + aM(t) + N_t,\ t = 1, 2, \cdots, T \tag{1}$$

with autocorrelated residuals

$$N_t = \rho N_{t-1} + e_t. \tag{2}$$

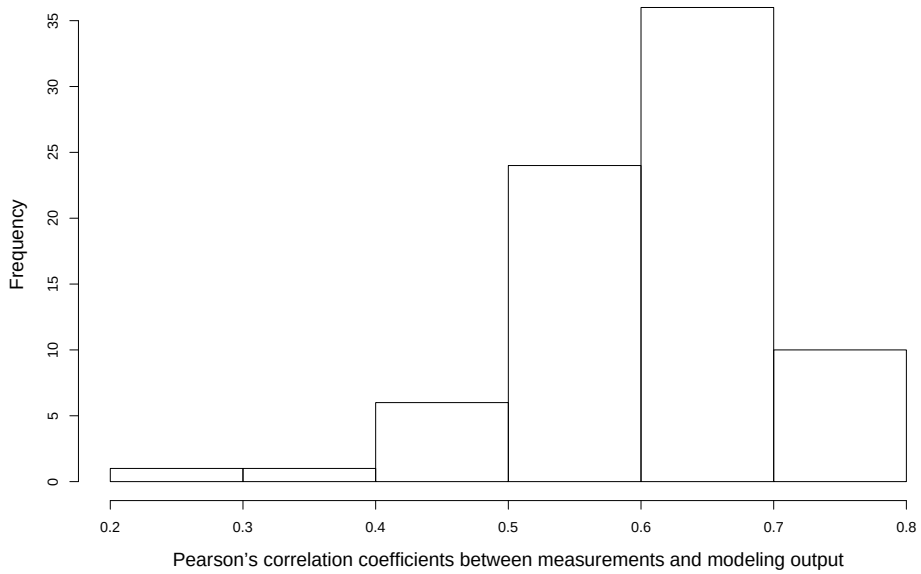

Figure 1: Histogram of Pearson's correlation coefficients between hourly measurements and modeling output for the 78 stations.

The residuals for the autocorrelation model in turn satisfy the following linear regression model:

$$e_t = \sum_{i=1}^{12} \alpha_i m_i(t) + \epsilon_t, \tag{3}$$

$$m_i(t) = \begin{cases} 1 & \text{if t mod 12 = i;} \\ 0 & \text{if t mod 12} \neq \text{i.} \end{cases}$$

$i = 1, \cdots, 12$.

The models (1), (2) and (3) are the same for all stations, while the coefficients are station specific. Both the measurements and modeling outputs are at the hourly level of temporal resolution and thus they have a strong auto-regressive structure. Assuming the $\{N_t\}$ have an AR(1) autoregressive structure provides computational simplicity at the possible expense of realism. We also have tried an AR(2) structure for the $\{N_t\}$ but the forecast results do not improve, leading us to use AR(1) throughout this report. Following Guillas et al. (2006), we first fit Model (1) by using generalized least squares as proposed by Box et al. (1994), $N(t)$ being an AR(1) process. If the modeling output captures the temporal structure of the measurements very well then Model (1) would be enough. However, most of the time, the measurements will still have some temporal structure that the modeling output fails to capture. That is to say, the residuals $\{e_t\}$ will have non-zero means. So, the second step is to fit Model (3). The covariates in this model are indicator functions of the 12 hours. We also tried indicator functions of the 24 hours, but the forecast results do not improve. So, in favor of fewer parameters, we choose indicator functions of the 12 hours. For the residuals of the AR(1) process $\{N_t\}$, $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ independently and identically.

Models (1) and (3) are fitted sequentially and independently. These two models can then be used to forecast the measurements. Suppose at each station, the training data are measurements and modeling output for times $t = 1$ to T. For times $t = T + 1, \cdots, T + T'$, only the modeling output is available and the goal is to forecast the measurement during that future time period. To do so, Models (1) and (3) are fitted to obtain estimators $\hat{c}$, $\hat{a}$, $\hat{\rho}$, $\hat{\alpha}_i$, $i = 1, \cdots, 12$. The forecast of $\hat{O}_t$ is then the following

$$\hat{O}_t = \hat{c} + \hat{a} M_t + \hat{N}_t,$$

where we have

$$\hat{N}_t = \hat{\rho}\hat{N}_{t-1} + \sum_{i=1}^{12} \hat{\alpha}_i m_i(t) \tag{4}$$

$t = T+1, \cdots, T+T^{'}$.

Although the two-step regression above is easy to implement, it has two disadvantages. First, it does not include any spatial correlation between stations. We fit models (1) and (3) at each station independently. Thus we can only forecast the measurement for stations that have measurements in the past, after estimating the requisite coefficients. In other words, it cannot borrow strength across space to forecast future values of stations without past data by exploiting data from neighbors with a past. Second, the two-step regression procedure only gives point forecasts without any indication of their uncertainties.

To address the first disadvantage of the two-step regression, we first consider a simple but ad-hoc approach which Kriges parameter estimates ($\hat{a}$, $\hat{c}$ and $\hat{\alpha}_i$) across space to get values for them at stations without measurements. After affirming the value of our ad-hoc approach, we address the second disadvantage by extending it to a Bayesian spatial-temporal model. The credible intervals of the forecast and prediction can then be obtained from their MCMC posterior samples at considerable computation cost.

3 Two-Step Linear Regression with Kriging

This section presents an ad-hoc approach which is able to not only forecast the measurements but also spatially predict the measurements at those locations without historic measurements. First, we fit models (1) and (3) at the stations with measurements. Then the estimates of parameters a, c and α_i, $i = 1, \cdots, 12$, are interpolated at locations without measurements by Kriging, a celebrated geostatistical approach for spatially interpolating point-referenced data. Given measurements of a random field $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), ..., Y(\boldsymbol{s}_n))^T$, the question is how to predict the random variable Y at a new site $\boldsymbol{s}_0$. The simplest Kriging approach, ordinary Kriging, assumes $E(\boldsymbol{Y}_{\boldsymbol{s}_0}) = E(\boldsymbol{Y}_{\boldsymbol{s}_i}) = \mu$, $i = 1, \cdots, n$. The interpolated

value at $\boldsymbol{s}_0$ is

$$Y^*(\boldsymbol{s}_0) = \sum_{i=1}^{n} w_i Y(\boldsymbol{s}_i),$$

w_i being the weight of $Y(\boldsymbol{s}_i)$ subject to the constraint $\sum_{i=1}^{n} w_i = 1$. Obviously, $Y^*(\boldsymbol{s}_0)$ is an unbiased estimator of $Y(\boldsymbol{s}_0)$. So, to minimize the MSE (mean square error) of $Y^*(\boldsymbol{s}_0)$, we only need to minimize its variance,

$$\begin{aligned} \sigma_E^2 &= E\left[(Y^*(\boldsymbol{s}_0) - Y(\boldsymbol{s}_0))^2\right] \\ &= -\gamma(\boldsymbol{s}_0 - \boldsymbol{s}_0) - \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) + 2\sum_{i=1}^{n} w_i \gamma(\boldsymbol{s}_i - \boldsymbol{s}_0). \end{aligned}$$

$\gamma(\cdot)$, the so-called variogram function, is defined as

$$\mathrm{V}ar\left[Y(\boldsymbol{s}+\boldsymbol{h}) - Y(\boldsymbol{s})\right] = C(\boldsymbol{h}) = 2\gamma(\boldsymbol{h}).$$

For simplicity, this report uses the exponential function to model the spatial covariance

$$C(\boldsymbol{h}) = \begin{cases} \sigma \exp(-|\boldsymbol{h}|/\lambda) & \text{if } |\boldsymbol{h}| > 0; \\ \sigma & \text{if } |\boldsymbol{h}| = 0. \end{cases} \tag{5}$$

where σ^2 is the spatial variance and λ is the range parameter. More complex choices of the spatial covariance function will be considered in future work. The weights $\{w_i\}$ are determined through minimizing σ_E^2. Comprehensive studies of Kriging can be found in Cressie (1993) and Stein (1999).

At the locations without measurements but with modeling output, we have the Kriging interpolated values of the parameters a, c and α_i, $i = 1, \cdots, 12$. Then we can plug these parameters into Formula (4) to obtain spatial predictions of the measurements.

4 A Bayesian hierarchical spatial-temporal model

The ad-hoc approach in the previous section is "self-contradictory" because it fits Models (1) and (3) assuming the parameters a, c and α_i, $i = 1, \cdots, 12$ are independent across stations but then assumes parameters at different stations to be spatially correlated. That ad-hoc approach then uses Kriging to interpolate the parameter values at the locations without measurements assuming parameters at different stations are spatially correlated.

Another disadvantage of the ad-hoc approach is that it fails to provide measures of the uncertainties associated with the forecasts or predictions. Kassteele et al. (2006) use a similar two-step regression model to interpolate PM10 concentrations over western Europe.

In this paper, the authors first fit a linear regression model with PM10 as the response, modeling output and other explanatory variables as covariates. Then the residuals are interpolated using Kriging. The approach in Kassteele et al. (2006) has the same weakness as the ad-hoc approach presented in the previous section. That is, it assumes the residuals are spatially independent to fit the model by using ordinary least square method and then uses Kriging to interpolate the residuals at other locations. Moreover, the model used in Kassteele et al. (2006) does not include temporal correlation because the data are is yearly averages of PM10 concentrations that do not have much temporal correlation. Hence, strictly speaking, both the ad-hoc approach in the previous section and that of Kassteele et al. (2006) do not yield spatial-temporal models. This section presents a Bayesian hierarchical spatial-temporal model that assumes parameters at different stations in Models (1) and (3) are spatially correlated Gaussian processes.

For the Bayesian spatial-temporal model has the following mathematical form:

$$
\begin{aligned}
&O_{\boldsymbol{s},t} = a_{\boldsymbol{s}} + c_{\boldsymbol{s}} M_{\boldsymbol{s},t} + N_{\boldsymbol{s},t} \\
&N_{\boldsymbol{s},t} = \rho N_{\boldsymbol{s},t-1} + \gamma_{1,\boldsymbol{s}} Z_{1,t} + \gamma_{2,\boldsymbol{s}} Z_{2,t} + \gamma_{3,\boldsymbol{s}} Z_{3,t} + \gamma_{4,\boldsymbol{s}} Z_{4,t} + \epsilon_{\boldsymbol{s},t} \\
&\boldsymbol{a} = (a_1, \cdots, a_n)^{\mathrm{T}} \sim MVN(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) \\
&\boldsymbol{c} = (c_1, \cdots, c_n)^{\mathrm{T}} \sim MVN(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \\
&\boldsymbol{\gamma}_1 = (\gamma_{1,1}, \cdots, \gamma_{1,n})^{\mathrm{T}} \sim MVN(\boldsymbol{\mu}_{\gamma_1}, \boldsymbol{\Sigma}_{\gamma_1}) \\
&\cdots \\
&\boldsymbol{\gamma}_4 = (\gamma_{4,1}, \cdots, \gamma_{4,n})^{\mathrm{T}} \sim MVN(\boldsymbol{\mu}_{\gamma_4}, \boldsymbol{\Sigma}_{\gamma_4}) \\
&\text{times } t = 1, \cdots, T, \text{ and} \\
&\text{sites } \boldsymbol{s} = s_1, \cdots, s_n \\
&\boldsymbol{\mu}_a = (\mu_a, \cdots, \mu_a)^{\mathrm{T}} \\
&\boldsymbol{\mu}_c = (\mu_c, \cdots, \mu_c)^{\mathrm{T}} \\
&\boldsymbol{\mu}_{\gamma_i} = (\mu_{\gamma_i}, \cdots, \mu_{\gamma_i})^{\mathrm{T}} \\
&\boldsymbol{\Sigma}_a = \sigma_a^2 \exp(-\boldsymbol{D}/\lambda_a) \\
&\boldsymbol{\Sigma}_c = \sigma_c^2 \exp(-\boldsymbol{D}/\lambda_c) \\
&\boldsymbol{\Sigma}_{\gamma_i} = \sigma_{\gamma_i}^2 \exp(-\boldsymbol{D}/\lambda_{\gamma_i}) \\
&\epsilon_{\boldsymbol{s},t} \text{ independently and identically distributed as } N(0, \sigma_\epsilon^2),
\end{aligned}
\tag{6}
$$

n being the number of monitoring sites, T, the number of hours and D, the Euclidean distance matrix between stations. Vectors $\boldsymbol{\mu}_a$, $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_{\gamma_i}$ have dimension n. In Model (6), $\boldsymbol{s}$ denotes location and t, time. The time-related covariates at time t are $Z_{1,t} = \cos(\pi t/6)$, $Z_{2,t} = \sin(\pi t/6)$, $Z_{3,t} = \cos(\pi t/12)$, and $Z_{4,t} = \sin(\pi t/12)$, instead of the indicator functions in (3) to reduce the number of parameters in the model. The parameter vectors $\boldsymbol{a}$, $\boldsymbol{c}$ $\boldsymbol{\gamma}_1, \cdots \boldsymbol{\gamma}_4$ at different locations are spatially correlated, that correlation being specified by an exponential function as shown in model (6). Given $\boldsymbol{a}$, $\boldsymbol{c}$ and $\boldsymbol{\gamma}_i$, $i = 1, \cdots 4$, the

measurements $O_{\boldsymbol{s},t}$ at each station $\boldsymbol{s}$ are spatially independent AR(1) processes. We make the following assumptions for this Bayesian spatial-temporal model.

- The measurements and modeling output are linearly related.
- The autoregressive parameter ρ is constant across stations.
- The residuals $\{N(t)\}$ follow an AR(1) process.
- The spatial correlation between measurements at different stations can be explained by the modeling output and the spatial correlation between the coefficients.

We use the Gibbs sampling algorithm proposed by Gelfand and Smith (1990) to fit the Bayesian spatial-temporal Model (6). The key to making Gibbs sampling work is writing Model (6) in matrix form, that is

$$\begin{aligned}
&\boldsymbol{O} = \boldsymbol{A}\boldsymbol{a} + \boldsymbol{M}\boldsymbol{c} + \boldsymbol{N} \\
&\boldsymbol{R}\boldsymbol{N} = \boldsymbol{\gamma}_1\boldsymbol{Z}_1 + \boldsymbol{\gamma}_2\boldsymbol{Z}_2 + \boldsymbol{\gamma}_3\boldsymbol{Z}_3 + \boldsymbol{\gamma}_4\boldsymbol{Z}_4 + \boldsymbol{\epsilon} \\
&\boldsymbol{a} \sim MVN(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) \\
&\boldsymbol{c} \sim MVN(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \\
&\boldsymbol{\gamma}_i \sim MVN(\boldsymbol{\mu}_{\gamma_i}, \boldsymbol{\Sigma}_{\gamma_i}) \quad i = 1, \cdots, 4.
\end{aligned} \tag{7}$$

The measurement vector is denoted as $\boldsymbol{O} = (O_{1,1}, \cdots, O_{1,T}, \cdots, O_{n,1}, \cdots, O_{n,T})^{\mathrm{T}}$ and the residual vector, as $\boldsymbol{N} = (N_{1,1}, \cdots, N_{1,T}, \cdots, N_{n,1}, \cdots, N_{n,T})^{\mathrm{T}}$. The matrix form of the

modeling output is

$$\boldsymbol{M} = \begin{pmatrix} M_{1,1} & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ M_{1,T} & 0 & \dots & 0 & \dots & 0 \\ 0 & M_{2,1} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & M_{2,T} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & M_{n,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & M_{n,T} \end{pmatrix}_{nT \times n}.$$

In the matrix $\boldsymbol{M}$, only the elements from row $(i-1) \times T + 1$ to $i \times T$ are non-zero at column i, $i = 1, \cdots, n$. We also define a matrix $\boldsymbol{r}$ as

$$\boldsymbol{r} = \begin{pmatrix} -\rho & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix}_{(T-1) \times T}$$

and a matrix $\boldsymbol{R}$ as a diagonal matrix with matrix $\boldsymbol{r}$ on the diagonal repeated n times.

Define the matrix $\boldsymbol{A}$ as

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 \end{pmatrix}_{nT \times n} .$$

In matrix $\boldsymbol{A}$, only the elements from row $(i-1) \times T+1$ to $i \times T$ are 1 at column i, $i = 1, \cdots, n$ while all other elements in matrix $\boldsymbol{A}$ are zero.

Define the matrix $\boldsymbol{Z}_1$ as

$$\boldsymbol{Z}_1 = \begin{pmatrix} \cos(2\pi/6) & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cos(T\pi/6) & 0 & \dots & 0 & \dots & 0 \\ 0 & \cos(2\pi/6) & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cos(T\pi/6) & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & \cos(2\pi/6) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & \cos(T\pi/6) \end{pmatrix}_{n(T-1) \times n} .$$

In that matrix, only the elements from rows $(i-1) \times (T-1)+1$ to $i \times (T-1)$ are non-zero

at column i, $i = 1, \cdots, n$. The matrices $\boldsymbol{Z}_2$, $\boldsymbol{Z}_3$ and $\boldsymbol{Z}_4$ have the same structure as $\boldsymbol{Z}_1$. In particular, the non-zero elements are replaced by $\sin(\pi t/6)$, $\cos(\pi t/12)$ and $\sin(\pi t/12)$, $t = 2, \cdots, T$.

Let $\boldsymbol{N} = \boldsymbol{O} - \boldsymbol{A}\boldsymbol{a} - \boldsymbol{M}\boldsymbol{c}$ and $\boldsymbol{F} = \boldsymbol{\gamma}_1\boldsymbol{Z}_1 + \boldsymbol{\gamma}_2\boldsymbol{Z}_2 + \boldsymbol{\gamma}_3\boldsymbol{Z}_3 + \boldsymbol{\gamma}_4\boldsymbol{Z}_4$. Then $\boldsymbol{\epsilon} = \boldsymbol{R}\boldsymbol{N}$ is just a vector of independent and identically distributed normal variables with mean 0 and variance σ_ϵ^2. So, the density of $\boldsymbol{O}$ given the other parameters is

$$p(\boldsymbol{O}|\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{R}, \sigma_\epsilon^2) \propto$$
$$\left(\sigma_\epsilon^2\right)^{-\frac{n(T-1)}{2}} \exp\left(-\frac{1}{2\sigma_\epsilon^2}\left[\boldsymbol{R}(\boldsymbol{O} - \boldsymbol{A}\boldsymbol{a} - \boldsymbol{M}\boldsymbol{c}) - \boldsymbol{F}\right]^{\mathrm{T}}\left[\boldsymbol{R}(\boldsymbol{O} - \boldsymbol{A}\boldsymbol{a} - \boldsymbol{M}\boldsymbol{c}) - \boldsymbol{F}\right]\right). \quad (8)$$

The inferences about the parameters, forecasts and predictions are based on their Markov Chain Monte Carlo (MCMC) samples generated by the Gibbs sampling algorithm. That algorithm as implemented is iterative. First, we choose some arbitrary initial values for all the parameters. Then, in each iteration of the algorithm, the parameters are updated by generating a random sample from their conditional distributions given the data and other parameter values from the previous iteration. For computational efficiency, each of the parameter vectors $\boldsymbol{a}$, $\boldsymbol{c}$ and $\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_4$ is updated as a block. The conditional distributions for each of the parameters are listed below from (9) to (15) for completeness.

The prior and conditional distributions of $\boldsymbol{a}$ are

$$\begin{aligned}
&p(\boldsymbol{a}|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) \sim MVN(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) \\
&p(\boldsymbol{a}|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a, \boldsymbol{M}, \boldsymbol{A}, \boldsymbol{R}, \sigma_\epsilon^2) \sim MVN(\boldsymbol{\mu}_a^{'}, \boldsymbol{\Sigma}_a^{'}) \\
&\boldsymbol{\Sigma}_a^{'} = \left[\boldsymbol{\Sigma}_a^{-1} + \boldsymbol{A}^{\mathrm{T}}(\frac{\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}}{\sigma_\epsilon^2})\boldsymbol{A}\right]^{-1} \\
&\boldsymbol{\mu}_a^{'} = \boldsymbol{\Sigma}_a^{'}\left[\boldsymbol{A}^{\mathrm{T}}(\frac{\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}}{\sigma_\epsilon^2})(\boldsymbol{O} - \boldsymbol{M}\boldsymbol{c}) - \frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{F}}{\sigma_\epsilon^2} + \boldsymbol{\Sigma}_a^{-1}\boldsymbol{\mu}_a\right]. \quad (9)
\end{aligned}$$

The prior and conditional distributions of $\boldsymbol{c}$ are

$$
\begin{aligned}
&p(\boldsymbol{c}|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \sim MVN(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \\
&p(\boldsymbol{c}|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c, \boldsymbol{M}, \boldsymbol{A}, \boldsymbol{R}, \sigma_\epsilon^2) \sim MVN(\boldsymbol{\mu}_c^{'}, \boldsymbol{\Sigma}_c^{'}) \\
&\boldsymbol{\Sigma}_c^{'} = \left[\boldsymbol{\Sigma}_c^{-1} + \boldsymbol{M}^{\mathrm{T}}(\frac{\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}}{\sigma_\epsilon^2})\boldsymbol{M}\right]^{-1} \\
&\boldsymbol{\mu}_c^{'} = \boldsymbol{\Sigma}_c^{'}\left[\boldsymbol{M}^{\mathrm{T}}(\frac{\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}}{\sigma_\epsilon^2})(\boldsymbol{O} - \boldsymbol{A}\boldsymbol{a}) - \frac{\boldsymbol{M}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{F}}{\sigma_\epsilon^2} + \boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c\right].
\end{aligned} \tag{10}
$$

The prior and conditional distributions of $\boldsymbol{\gamma}_i$, $i = 1, 2, 3, 4$, are

$$
\begin{aligned}
&p(\boldsymbol{\gamma}_i|\boldsymbol{\mu}_{\gamma_i}, \boldsymbol{\Sigma}_{\gamma_i}) \sim MVN(\boldsymbol{\mu}_{\gamma_i}, \boldsymbol{\Sigma}_{\gamma_i}) \\
&p(\boldsymbol{\gamma}_i|\boldsymbol{\mu}_{\gamma_i}, \boldsymbol{\Sigma}_{\gamma_i}, \boldsymbol{M}, \boldsymbol{A}, \boldsymbol{R}, \sigma_\epsilon^2) \sim MVN(\boldsymbol{\mu}_{\gamma_i}^{'}, \boldsymbol{\Sigma}_{\gamma_i}^{'}) \\
&\boldsymbol{\Sigma}_{\gamma_i}^{'} = \left[\boldsymbol{\Sigma}_{\gamma_i}^{-1} + \frac{\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{Z}_i}{\sigma_\epsilon^2}\right]^{-1} \\
&\boldsymbol{\mu}_{\gamma_i}^{'} = \boldsymbol{\Sigma}_{\gamma_i}^{'}\left[\frac{\boldsymbol{Z}_i^{\mathrm{T}}(\boldsymbol{R}(\boldsymbol{O} - \boldsymbol{A}\boldsymbol{a} - \boldsymbol{M}\boldsymbol{c}) - F + \boldsymbol{\gamma}_i\boldsymbol{Z}_i)}{\sigma_\epsilon^2} + \boldsymbol{\Sigma}_{\gamma_i}^{-1}\boldsymbol{\mu}_{\gamma_i}\right].
\end{aligned} \tag{11}
$$

The prior and conditional distributions of σ_ϵ^2 are both inverse gamma distributions with the following form

$$
\begin{aligned}
&p(\sigma_\epsilon^2|\alpha, \beta) \propto (\sigma_\epsilon^2)^{-\alpha-1}\exp\left(-\frac{1}{\beta\sigma_\epsilon^2}\right) \\
&p(\sigma_\epsilon^2|\boldsymbol{A}, \boldsymbol{O}, \boldsymbol{m}, \boldsymbol{a}, \boldsymbol{c}, \boldsymbol{R}, \alpha, \beta) \propto (\sigma_\epsilon^2)^{-\alpha^{'}-1}\exp\left(-\frac{1}{\beta^{'}\sigma_\epsilon^2}\right) \\
&\alpha^{'} = \alpha + \frac{n(T-1)}{2} \\
&\beta^{'} = \left[1/\beta + \frac{1}{2}\left(\boldsymbol{R}(\boldsymbol{O} - \boldsymbol{A}\boldsymbol{a} - \boldsymbol{M}\boldsymbol{c}) - F\right)^{\mathrm{T}}\left(\boldsymbol{R}(\boldsymbol{O} - \boldsymbol{A}\boldsymbol{a} - \boldsymbol{M}\boldsymbol{c}) - F\right)\right]^{-1}.
\end{aligned} \tag{12}
$$

The prior and conditional distributions of σ_a^2 are both inverse gamma distributions with

the following form

$$
\begin{aligned}
&p(\sigma_a^2|\alpha,\beta) \propto (\sigma_a^2)^{-\alpha-1}\exp\left(-\frac{1}{\beta\sigma_a^2}\right)\\
&p(\sigma_a^2|\alpha,\beta,\boldsymbol{a},\lambda_a,\boldsymbol{\mu}_a) \propto (\sigma_a^2)^{-\alpha^{'}-1}\exp\left(-\frac{1}{\beta^{'}\sigma_a^2}\right)\\
&\alpha^{'} = \alpha + \frac{n}{2}\\
&\beta^{'} = \left[1/\beta + \frac{1}{2}\left(\boldsymbol{a}-\boldsymbol{\mu}_a\right)^{\mathrm{T}}(\exp(-\boldsymbol{D}/\lambda_a))^{-1}\left(\boldsymbol{a}-\boldsymbol{\mu}_a\right)\right]^{-1}. \qquad (13)
\end{aligned}
$$

We cannot find a conjugate prior for the parameter λ_a, so its conditional distribution does not have a closed form. However, we have

$$
p(\lambda_a|\sigma_a^2,\boldsymbol{\mu}_a,\boldsymbol{a}) \propto |\exp(-\boldsymbol{D}/\lambda_a)|^{-1}\exp\{-\frac{1}{2}\left(\boldsymbol{a}-\boldsymbol{\mu}_a\right)^{\mathrm{T}}(\exp(-\boldsymbol{D}/\lambda_a))^{-1}\left(\boldsymbol{a}-\boldsymbol{\mu}_a\right)\}. \qquad (14)
$$

To obtain a random sample from the above distribution of λ_a, we have to use the Metropolis-Hasting algorithm proposed by Metropolis et al. (1953) and Hastings (1970). The conditional distributions of σ_c^2 and λ_c are very similar to σ_a^2 and λ_a. We only need to replace $\boldsymbol{a}$ and $\boldsymbol{\mu}_a$ with $\boldsymbol{c}$ and $\boldsymbol{\mu}_c$ correspondingly in (13) and (14).

To update ρ, its ordinary least square (OLS) estimate given other parameters is used instead of a random sample from its conditional distribution. We do this in part for computational simplicity. Finding a conjugate prior for ρ is very difficult while finding its ordinary least square estimation is very easy. The main reason is that the variance of the conditional distribution of ρ is small enough to be treated as zero because of the large sample size $n \times (T-1)$. The OLS estimate of ρ is

$$
\hat{\rho} = \frac{\sum_{s=1}^{n}\sum_{t=2}^{T}(\tilde{N}_{s,t} - \sum_{j=1}^{4}\tilde{Z}_{j,t}\gamma_{j,s})\tilde{N}_{s,t-1}}{\sum_{s=1}^{n}\sum_{t=2}\tilde{N}_{s,t-1}^2}, \qquad (15)
$$

where $\tilde{N}$ is the matrix form of $\boldsymbol{N}$ in model (7) and $\tilde{Z}_{1,t} = \cos(\pi t/6)$, $\tilde{Z}_{2,t} = \sin(\pi t/6)$, $\tilde{Z}_{3,t} = \cos(\pi t/12)$, $\tilde{Z}_{4,t} = \sin(\pi t/12)$.

At each iteration in Gibbs sampling, the forecast or prediction is obtained in an iterative way. The forecast or prediction is

$$\begin{aligned}
&\hat{O}_{\boldsymbol{s},t} = a_{\boldsymbol{s}} + c_{\boldsymbol{s}} M_{s,t} + N_{s,t} \\
&N_{\boldsymbol{s},t} = \rho N_{\boldsymbol{s},t-1} + \gamma_{1,\boldsymbol{s}} Z_{1,t} + \gamma_{2,\boldsymbol{s}} Z_{2,t} + \gamma_{3,\boldsymbol{s}} Z_{3,t} + \gamma_{4,\boldsymbol{s}} Z_{4,t} + \epsilon_{\boldsymbol{s},t} \\
&\epsilon_{\boldsymbol{s},t} \text{ is a random sample generated from } N(0, \sigma_\epsilon^2).
\end{aligned} \tag{16}$$

At those stations where the historic measurements are available up to time T, $\hat{O}_{\boldsymbol{s},t}$, $t = T+1, \cdots, T+T^{'}$ are the forecasts. At those stations where there is no historic measurement, $\hat{O}_{\boldsymbol{s},t}$, $t = 1, \cdots, T$ are the predictions. The parameters at those stations without measurements can be spatially predicted by the parameter values at stations with measurement. Finally, the credible interval of the forecast or prediction is obtained by taking their quantiles of all their MCMC samples.

5 Data analysis and conclusions

Among all the 78 stations, we use the first 240 hours of measurements and modeling outputs at 15 stations as training data. Then we forecast the ozone concentration levels in the next 240 hours for the 15 stations and also predict the first 240 hours' ozone concentration levels at the remaining 63 stations. To fit the Bayesian spatial-temporal Model (7), we use 500 iterations of the Gibbs sampling algorithm and the first 50 iterations for the "burnin-in" period. Figure 2 shows the time series plots of the posterior MCMC samples for some parameters as example. From that plot, we can see the Markov Chains converge after the "burn-in" period. We use the root mean square forecast error (RMSFE) and root mean square prediction error (RMSPE) to evaluate accuracy of the forecasts or predictions, the smaller RMSFE or RMSPE, the better the forecast or prediction. At each station location $\boldsymbol{s}$, RMSFE or RMSPE is $\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(O_{\boldsymbol{s},t} - \hat{O}_{\boldsymbol{s},t}\right)^2}$, $O_{\boldsymbol{s},t}$ being the real measurement at time

t and $\hat{O}_{\boldsymbol{s},t}$, the forecast or prediction.

5.1 Forecasting and spatial prediction

Table 1 presents parameters estimates for Model (6). Most differ significantly from 0 except σ_a^2, the spatial variance of the process $\boldsymbol{a}$. Table 2 presents the RMSFE of the forecasts by three approaches, Bayesian spatial-temporal model (6), the ad-hoc approach in Section (3) and the unadjusted modeling output. As an example, Figure 3 shows the plot of measurements versus the forecast by the Bayesian spatial-temporal model for one station.

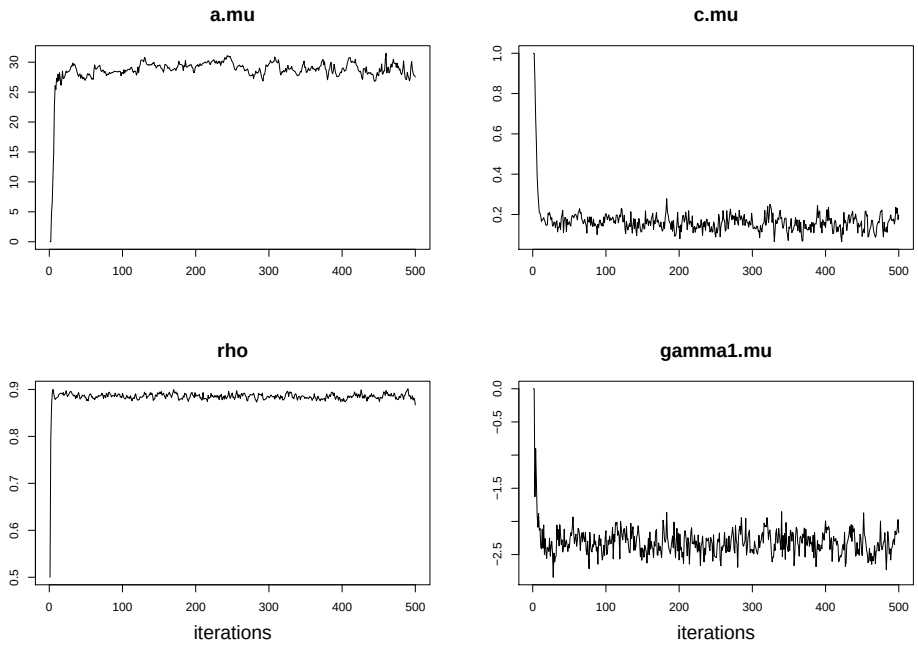

Figure 2: Plots of the parameter values in the Gibbs sampling iterations. The upper left panel is for μ_a; the upper right panel is for μ_c; the lower left panel is for ρ and the lower right panel is for $\mu_{\gamma 1}$.

The spatial predictor predicts the ozone concentration levels over the first 240 hours at the 63 stations that have modeling output only and no historic measurements. The four predictions include the Bayesian spatial-temporal model, the ad-hoc approach, Kriging, and the unmodified modeling output. By kriging, we mean using the Kriging approach to interpolate the measurements at the 63 stations repeatedly for every hour. So, the modeling

Table 1: Posterior mean and standard deviation of μ_a, σ_a^2, μ_c, σ_c^2, ρ and μ_{γ_i}, $\sigma_{\gamma_i}^2$, $i = 1, \cdots, 4$.

	mean	sd		mean	sd
μ_a	28.96	0.94	μ_{γ_4}	1.10	0.13
μ_c	0.16	0.04	$\sigma_{\gamma_1}^2$	1.96	1.06
σ_a^2	2.99	3.81	$\sigma_{\gamma_2}^2$	2.12	1.15
σ_c^2	0.30	0.11	$\sigma_{\gamma_3}^2$	1.06	0.51
μ_{γ_1}	-2.31	0.17	$\sigma_{\gamma_4}^2$	0.84	0.51
μ_{γ_2}	-3.17	0.19	ρ	0.89	0.01
μ_{γ_3}	-0.54	0.15	σ_ϵ^2	39.19	0.98

output is not used in the Kriging approach. Tables 3 and 4 present the RMSPE for these four approaches. As an example, Figure 4 shows the plot of measurements versus the spatial predictions using the Bayesian spatial-temporal model at one station.

The Bayesian spatial-temporal model can give credible predictive intervals for the forecasts and predictions by using the quantiles of the MCMC samples of the forecast and prediction. The coverage probability of the credible interval is defined as the proportion of the true predicted measurements falling into the credible intervals of the forecast or prediction. The coverage probabilities of the 90% credible intervals for the forecasts and predictions are 85% and 84% respectively.

5.2 Remarks

The results above suggest a number of conclusions.

- The two-step linear regression procedure can only forecast the measurements. The Kriging approach can only spatially predict the measurements. But the ad-hoc approach and the Bayesian spatial-temporal model can both forecast and predict the measurements at the same time. Figure 3 and Figure 4 show that the forecasts and predictions from the Bayesian spatial-temporal model track the real measurements reasonably well.

Table 2: The first column gives the indices of 15 stations where the first 240 hours' measurements are available. The second, third and forth columns are the RMSFE (square root of the mean square forecast error) of the next 240 hours' measurements fore-casted by Bayesian hierarchical model, ad-hoc approach and modeling output without calibration. The number followed by $*$ indicates the approach that has the smallest RMSFE in each row.

station	Bayesian	ad-hoc	modeling output
1	16.52*	21.51	23.06
2	14.33*	19.70	21.21
3	17.07	13.73*	14.03
4	17.22*	18.16	18.37
5	15.95	15.61*	17.53
6	17.00	14.54	12.34*
7	16.65*	19.33	21.55
8	14.08*	19.91	20.87
9	12.43*	14.08	14.90
10	13.56	11.28*	14.55
11	12.89	12.04*	13.44
12	13.43*	15.37	15.98
13	13.23	10.95*	12.74
14	13.01	12.57*	13.96
15	13.90*	14.71	17.16
mean	14.75*	15.57	16.78

- Tables 2, 3 and 4 show that in averaging over the stations, the Bayesian spatial-temporal model gives the smallest RMSFE and RMSPE among all the competitive approaches for forecasting and spatially prediction. For forecasting, the mean RMSFE of the Bayesian spatial-temporal model is about 20% smaller than the forecasts from the unadjusted modeling output. The RMSFE of the ad-hoc approach is slightly bigger than the Bayesian spatial-temporal model. In prediction, not surprisingly, Kriging has the biggest RMSPE because we have interpolated 63 stations using only 15 stations. The RMSPE of the Bayesian spatial-temporal model is about 10% smaller than the ad-hoc approach and the modeling output. So, using Model (6) to adjust the modeling output seems worthwhile achieving better forecasts and predictions. The

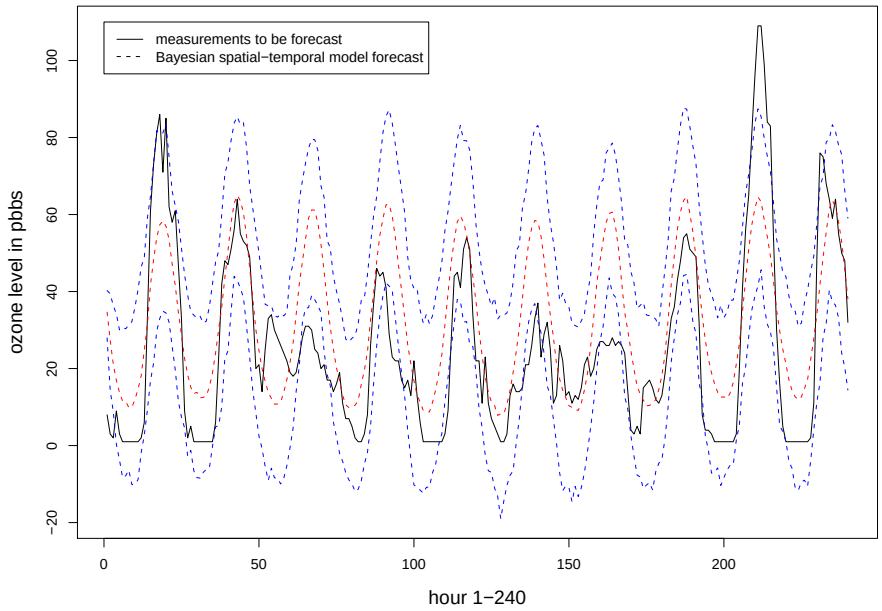

Figure 3: Plot of measurement versus the forecast and the 90% credible interval from the Bayesian spatial-temporal model at one of the 15 stations.

coverage probabilities of the intervals for the Bayesian spatial-temporal model are fairly good, meaning the uncertainties in the forecasts and predictions are accounted for quite well.

- Although in averaging over the stations, the Bayesian spatial-temporal model achieves the smallest RMSFE and RMSPE, it is also marginally better than the ad-hoc approach in forecasting albeit at significant cost in computational time and parsimony judging from the number of parameters. Table 2 shows that the ad-hoc approach has smaller RMSFE than Bayesian spatial-temporal model at 8 out 15 stations. Table 3 and 4 show that ad-hoc approach has smaller RMSPE than Bayesian spatial-temporal model at 24 out 63 stations. So, cost stated above, the major benefit of the Bayesian spatial-temporal model lies in the quality of spatial predictions and in its well calibrated credible predictive intervals.

- As we can see in Table 1, the posterior mean of the auto-regressive parameter ρ is as

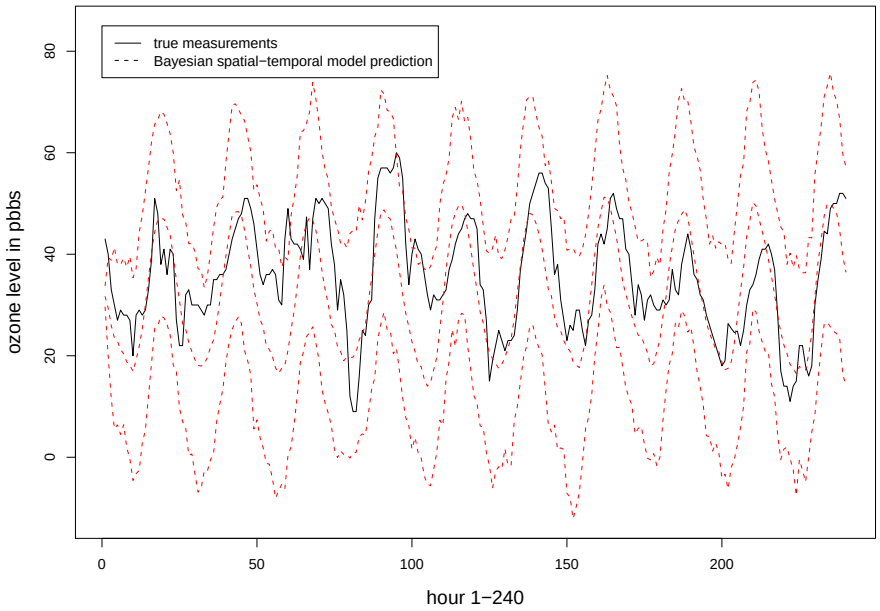

Figure 4: Plot of measurement versus the spatial prediction and the 90% credible interval from the Bayesian spatial-temporal model at one of the 63 stations.

big as 0.89, which suggests strong auto-regression in the residuals $\{N_t\}$.

5.3 Future Work

We assume the residuals $\{\epsilon_{\boldsymbol{s},t}\}$ in Model (6) are independently normal variables. This is assumption implies that the modeling output captures the spatial correlation within the measurements. To verify this assumption, we calculate the Pearson's correlation coefficients between residuals at different stations. We use the residuals for the 15 stations in the training dataset. Figure 5 shows the plot of these correlation coefficients versus the distance between stations and Figure 5 and Figure 6 shows the locations of the 15 stations. We can see in Figure 5 that there are still some spatial correlation left in the residuals. Figure 4 shows that the spatial predictions are very smooth, which also implies that the prediction does not borrow much strength from the neighboring stations. In the future work, we will assume that the residuals at different stations $\{\epsilon_{\boldsymbol{s},t}\}$ are also spatially correlated. Through this

technical report, we assume exponential function for the spatial correlation for simplicity, we will use Matern function in the future for its flexibility.

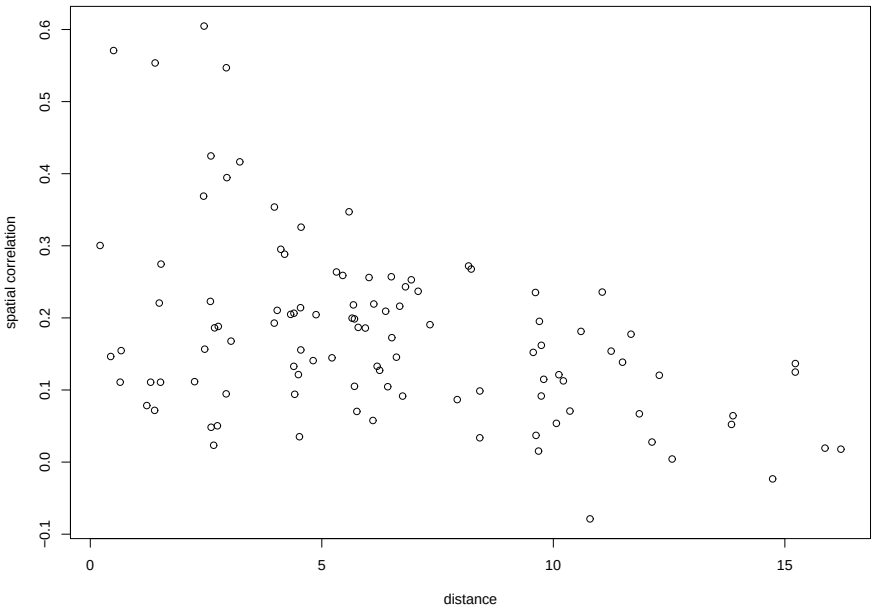

Figure 5: The correlation coefficients of residuals at different stations versus the distance between stations. This plot is for the 15 stations in the training dataset.

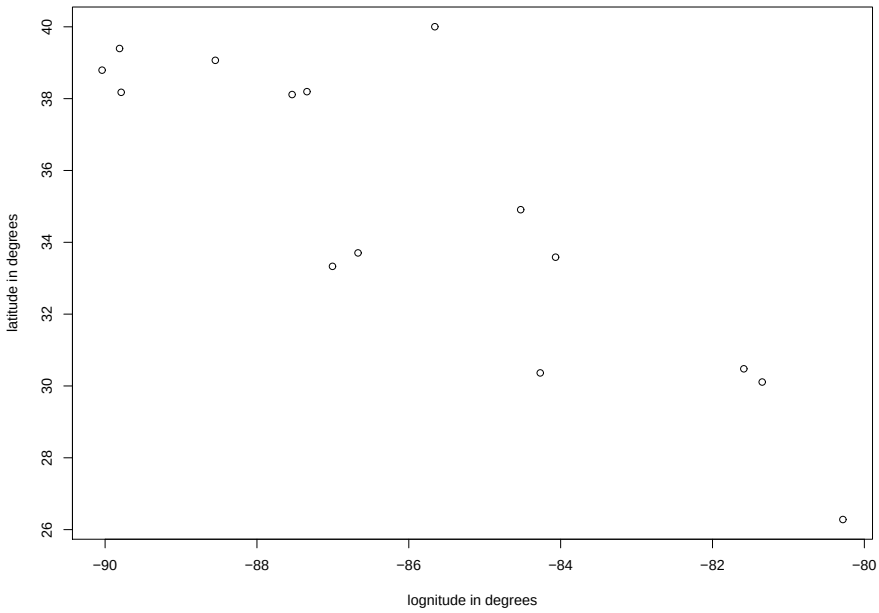

Figure 6: The locations of the 15 stations in the training dataset.

Table 3: RMSPE (square root of the mean square prediction error) at stations 1-35. Column 2: Bayesian spatial-temporal model; column 3: ad-hoc approach; column 4: modeling output; column 4: Kriging approach. The number followed by $*$ indicates the approach that has the smallest RMSPE in each row.

station	Bayesian	ad-hoc	modeling	Kriging
1	10.54*	13.08	12.07	16.51
2	11.10	12.07	10.11*	17.05
3	13.93	13.91	13.25*	24.15
4	14.26*	17.08	17.52	19.17
5	8.88*	29.98	12.32	43.40
6	12.79	18.15	9.94*	40.89
7	17.37	12.98*	19.02	22.80
8	18.88	13.62*	21.64	23.10
9	21.06	17.28*	24.66	27.81
10	19.47	16.26*	18.33	31.90
11	17.68	16.07*	18.54	35.59
12	18.26*	17.94	18.92	29.32
13	16.85*	20.35	19.17	36.71
14	13.36*	15.37	13.50	20.84
15	14.91	11.49*	13.76	16.70
16	14.98	12.91*	13.89	17.11
17	16.11*	18.77	16.28	19.14
18	14.51*	15.76	14.94	23.07
19	14.00*	14.29	14.92	18.33
20	12.79*	14.15	13.61	17.70
21	13.91	13.06*	13.08	18.79
22	13.75	10.26*	11.98	14.83
23	15.48*	16.25	16.46	10.38
24	8.44*	10.75	9.05	19.64
25	20.60	15.93*	23.25	24.78
26	15.51	13.40*	18.58	25.74
27	16.82*	19.10	20.19	27.33
28	18.99	12.9*4	22.49	26.94
29	18.44*	19.17	21.57	32.11
30	19.49	18.70*	23.67	32.67
31	23.61	17.84*	23.32	25.94
32	12.50*	21.61	13.46	34.19
33	14.38*	22.79	17.86	32.95
34	14.57*	15.29	11.87	27.05
35	15.57	14.76*	15.72	24.43

Table 4: RMSPE at stations 36-63. Column 2: Bayesian spatial-temporal model; column 3: ad-hoc approach; column 4: modeling output; column 5: Kriging approach. The bold station number means the Bayesian spatial-temporal does not have the smallest RMSPE in that station.

station	Bayesian	ad-hoc	modeling	Kriging
36	12.73*	22.56	14.94	35.43
37	15.28*	23.24	18.28	33.09
38	13.75	12.52*	12.56	26.54
39	15.90*	21.30	23.26	20.20
40	14.24*	14.86	17.75	16.91
41	13.26*	14.49	18.68	16.57
42	15.69*	14.72	20.83	18.77
43	11.16*	14.66	10.06	26.53
44	8.71*	14.28	9.08	23.84
45	17.30	15.07*	20.56	14.22
46	17.15	14.35*	18.50	31.88
47	18.54	13.80*	19.79	28.34
48	11.86	15.54	11.83*	19.62
49	18.17*	20.88	21.85	25.71
50	14.80*	15.35	14.59	13.38
51	13.94	11.48*	17.41	14.31
52	12.45*	12.60	16.81	13.79
53	16.22*	18.07	23.34	16.47
54	15.92	12.77*	19.68	13.77
55	13.49*	15.47	19.74	15.31
56	15.96*	17.61	19.94	21.02
57	18.54*	19.41	22.73	25.79
58	17.57*	19.02	21.94	25.05
59	17.80*	17.40	22.95	23.48
60	16.62	13.34*	18.60	18.03
61	17.95	14.32*	23.61	19.83
62	15.16*	18.49	15.69	14.62
63	14.31*	15.48	14.14	19.99
mean	15.36	16.13	17.27	23.36

References

Box, G., Jenkins, G. and Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control; 3rd Edition*, Holden-Day.

Cressie, N. A. (1993). *Statistics for Spatial Data*, Wiley-Interscience.

Gelfand, A. and Smith, A. (1990). Sampling-based approaches to calculating marginal densities, *Journal of American Statistical Association* **85**: 398–409.

Grell, G., Dudhia, J. and Stauffer, D. (1995). A description of the fifth-generation penn state/ncar mesoscale model (mm5), *Technical report*, National Center for Atmospheric Research, Boulder, Colorado.

Guillas, S., Tiao, G., Wuebbles, D. and Zubrow, A. (2006). Statistical diagnostic and correction of a chemistry-transport model for the prediction of total column ozone, *Atmospheric Chemistry and Physics* **6**: 527–537.

Hastings, W. (1970). Monte carlo sampling methods using markov chains and their application., *Biometrika* **57**: 97–109.

Kassteele, J. V. d., Koelemeijer, R. B. A., Dekkers, A. L. M., Schaap, M., Homan, C. and Stein, A. (2006). Statistical mapping of pm10 concentrations over western europe using secondary information from dispersion modeling and modis satellite observations, *Stochastic Environmental Research and Risk Assessment (SERRA)* **21**: 183–194.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953). Equation of state calculations by fast computing machines., *Journal of Chemical Physics* **21**: 1087–1092.

Odman, M. and Ingram, C. (1996). Multiscale air quality simulation platform (maqsip) source code documentation and validation, *Technical report*, Environmental Programs, MCNC-North Carolina Supercomputing Center.

Stein, M. (1999). *Interpolation of spatial data : some theory for kriging*, New York : Springer.