Noname manuscript No.
(will be inserted by the editor)

A fast Bayesian method for updating and forecasting hourly ozone levels

Sujit K. Sahu · Stan Yip · David M. Holland

Received: date / Accepted: date

Abstract A Bayesian hierarchical space-time model is proposed by combining information from real-time ambient AIRNow air monitoring data, and output from a computer simulation model known as the Community Multi-scale Air Quality (Eta-CMAQ) forecast model. A model validation analysis shows that the model predicted maps are more accurate than the maps based solely on the Eta-CMAQ forecast data for a two week test period. These out-of sample spatial predictions and temporal forecasts also outperform those from regression models with independent Gaussian errors. The method is fully Bayesian and is able to instantly update the map for the current hour (upon receiving monitor data for the current hour) and forecast the map for several hours ahead. In particular, the eight-hour average map which is the average of the past four hours, current hour and three hours ahead is instantly obtained at the current hour. Based on our validation, the exact Bayesian method is preferable to more complex models in a real-time updating and forecasting environment.
Key Words: Bayesian inference; Eta-CMAQ model; Space-time forecasting; hierarchical model; separable models; spatial interpolation.

1 Introduction

Accurate, instantaneous and high resolution spatial air-quality information can better inform the U.S. public and regulatory agencies about air pollution levels that lead to adverse health effects. The most direct way to obtain accurate air quality information is from measurements made at surface monitoring stations across the United States (U.S.). However, many areas of the U.S. are not monitored and typically, air monitoring sites are sparsely and irregularly spaced over large areas. Thus, it is now important to develop computationally efficient models to combine air monitoring data and numerical model

Sujit K. Sahu
School of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK. E-mail: S.K.Sahu@soton.ac.uk

Stan Yip
Exeter Climate Systems, University of Exeter, Exeter, EX4 4QJ, UK E-mail: C.Y.Yip@exeter.ac.uk

David M. Holland
U.S. Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, NC, 27711, USA. E-mail: Holland.David@epamail.epa.gov

output, in a coherent way for better prediction and forecasting of air pollution over short (e.g. hourly) time periods.

U.S. national air quality forecasts and near real-time predictive spatial maps are currently provided to the general public through the EPA-AIRNow web site: http://airnow.gov/. Current and next day particulate matter and ozone (O_3) air quality forecasts for over 200 U.S. cities are now provided on a daily basis. These forecast maps, however, are based primarily on the output of a computer simulation model known as the Eta-CMAQ model, see e.g. http://www.epa.gov/asmdnerl/CMAQ/. These models use emission inventories, meteorological information, and land use to estimate average pollution levels for gridded cells (12 km^2) over successive time periods. However, it is well known that these computer model may produce biased output and, as a result, this may lead to inaccurate pollution forecasts.

The objective of this paper is to develop a set of Bayesian hierarchical models which are capable of producing instantaneous, but more accurate short term forecast maps of hourly ozone concentration levels. These models combine ground-level observations from the real-time ozone monitoring network (http://airnow.gov) and output from the Eta-CMAQ model. Using data over a two week period in August 2005 we develop a Bayesian model which is shown to provide improved predictions relative to those achieved by Eta-CMAQ alone based on cross-validation. The space-time model lends itself to closed form analytic Bayesian posterior predictive distributions for spatial interpolation of ozone concentration level for the past hours, current hour and forecast for future hours. These predictive distributions provide instantaneous spatial interpolation maps which could be used in a real-time environment such as the U.S. EPA AIRNow system. The predictive distributions are used to obtain the eight-hour average map which is the average of the past four hours, current hour and three hours ahead, see Section 3.3. The forecasts are evaluated by using the model fitted to several weekly data sets. Our approach does not rely on iterative algorithms such as the Markov Chain Monte Carlo (MCMC) methods that are often used in Bayesian model fitting and forecasting. The MCMC methods require considerably more execution time to estimate model parameters, thus eliminating their potential use in real-time forecasting environments.

Although spatial prediction with fused data is a relatively new field, several papers have appeared in the literature on this topic, see the review by Gelfand and Sahu (2009) and the references therein for a recent snapshot of activities. Fuentes and Raftery (2005) developed a hierarchical statistical framework to model the "true" pollutant process as jointly Gaussian random fields. They estimate the parameters for the bias of Eta-CMAQ output, the parameters of the covariance structure for Eta-CMAQ, and the measurement error process. However, this methodology only applies to spatial processes at a fixed time point, without evaluation of the space-time dependence structure. Kang *et al.* (2008) consider Kalman-Filter approaches to improve next day forecasts of ozone concentration at individual U.S. monitoring sites for the summer of 2005.

Zimmerman and Holland (2005) consider the problem of optimal spatial prediction of wet deposition data using data from two monitoring networks with network-specific biases and variances. Cowles and Zimmerman (2003) use a Bayesian modeling approach for spatio-temporal data from two monitoring networks that accounts for possible differences in network measurement error, bias and variances. Jun and Stein (2004) suggest new ways of comparing space-time correlation structure of monitoring observations with Eta-CMAQ numerical model output. McMillan *et al.* (2008) develop their model at the grid cell level simplifying the computation requirements and enabling the use of this model to provide fused predictions for large spatial domains and temporal periods. Unlike the exact, almost instantaneous, computation method proposed in this paper, these methods and those of

Fuentes and Raftery (2005) rely on slower MCMC algorithms. Moreover, none of the above mentioned articles developed forecasting methods in the data fusion setting of this paper.

In an earlier paper (Sahu *et al.*, 2009) we have developed a hierarchical space-time forecast model for the daily 8-hour maximum ozone concentration data for the same two week time period in 2005 from the same study region. Besides the obvious difference in time units, hourly and daily, there are some fundamental differences between the objectives, models, fitting and forecasting methods, and the results of the current paper and the previous paper. The objective here is to instantaneously predict the current 8-hour ozone concentration based on past data upto the current hour while the same there is to obtain accurate forecast for the next day. The model here incorporates cyclic parameters describing the diurnal patterns and the results are obtained within a few seconds of inputting the current hour's data. The previous paper proposes a more complex dynamic model which requires iterative MCMC methods taking about 5-7 hours to produce the next day forecasts. When applied to the same data set these forecasts can be more accurate than those from the simple model proposed in this paper, indeed see Section 5 where these and further differences between the two approaches are presented.

The remainder of this paper is organized as follows. In Section 2 we describe the available data and their use in our modeling development described in Section 3. Model validation results and model based analyses are presented in Section 4. The differences between the methodology of the current paper are contrasted with that of our previous paper which modeled daily data in Section 5. A few summary remarks are placed in Section 6. An Appendix contains derivations of simplifications for various crucial expressions in the posterior distributions.

2 Data descriptions

We use real-time hourly ozone concentrations in parts per billion (ppb) units from $n = 350$ sites covering the eastern U.S. for a two-week test period, August 2–14, 2005. Data from 40 additional sites are set aside for model validation, see Figure 1. There are about 20% missing values in the monitoring data which we impute using a simple regression model.

There are 9119 Eta-CMAQ grid cells (12 km^2) in our study region spanning the eastern U.S. In practice, the hourly output from the Eta-CMAQ model are available up to 24 hours in advance. However, in our post-hoc study for data from 2005, Eta-CMAQ output for all days and hours are available for analysis, though we shall pretend that the output for only the next 24 hours are available to mimic the real situation.

The range of the Eta-CMAQ forecast data is quite similar to the range of the ground-level ozone monitoring data. To compare the Eta-CMAQ forecasts with the observed monitoring data, we plot data from four randomly chosen sites and Eta-CMAQ forecasts from the corresponding grid cells containing the sites in Figure 2. The plots show good agreement between the two at some of the sites but large disagreements at the other sites. This implies that there is spatio-temporal bias in the Eta-CMAQ forecasts and appropriate space-time modeling is needed to remove those biases.

To complete a full weekly cycle, we model data for a running window of seven days starting at any given hour. More distant past data can be included at the expense of increasing the computational burden. For spatial prediction, we use the Eta-CMAQ forecasts for 3000 randomly sampled grid cells out of the available 9119 grid cells. This is for illustration purposes only, as all of the available Eta-CMAQ output could be used to produce spatial maps. We use the square-root scale to stabilize the variance for modeling, but produce the predictions on the original scale for ease of interpretations, as done previously in Sahu, Yip and Holland (2009).

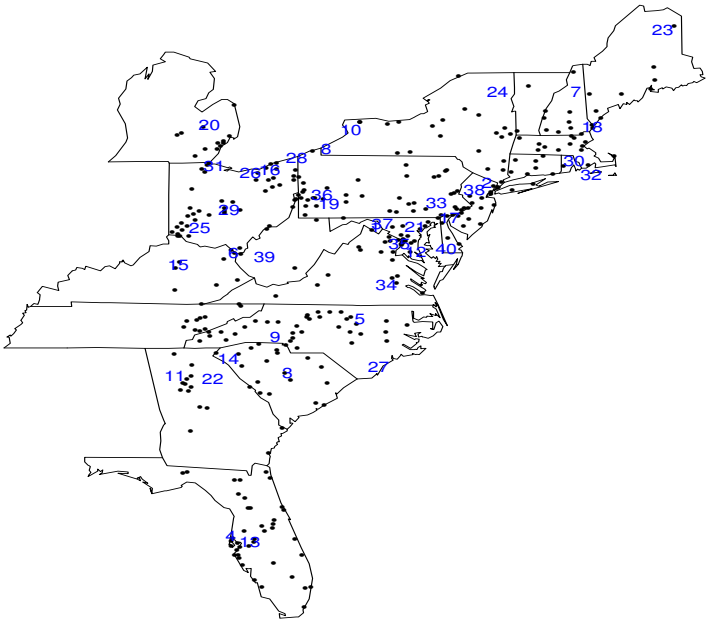

Fig. 1 The 350 data and 40 validation sites (1 to 40).

3 Models

Let $Z(\mathbf{s}, t)$ denote the observed square-root ozone concentration at location $\mathbf{s}$ and at hour t ($t = 1, \ldots, T$). We develop models for data from n stations denoted by $\mathbf{s}_1, \ldots, \mathbf{s}_n$, for a running window of seven days so that $T = 168(24 \times 7)$.

Further, let $X(\mathbf{s}, t)$ denote the square-root of the Eta-CMAQ ozone forecast value at the grid cell covering the site $\mathbf{s}$ and at time t. For simplicity, we adopt this notation scheme to define the location of Eta-CMAQ areal grid cell averages. Figure 2 shows that $X(\mathbf{s}_i, t)$ can be a good predictor of $Z(\mathbf{s}_i, t)$. The figure also shows heavy diurnal cycles in both ozone concentrations and their Eta-CMAQ forecasts.

We model the diurnal patterns by including a different hourly intercept for each hour to have an adequate, but simple model. The hourly intercept at any given hour remains constant for different days. The hourly intercept is defined by $\xi(t) = \beta_j$, where the hour $t(= 1, \ldots, T)$ corresponds to the jth hour of the day, $j = 1, \ldots, 24$. The full model is given by:

$$Z(\mathbf{s}_i, t) = \beta_0\, x(\mathbf{s}_i, t) + \xi(t) + w(\mathbf{s}_i, t), i = 1, \ldots, n,\ t = 1, \ldots, T, \tag{1}$$

where β_0 is an unknown regression co-efficient. Note that the model (1) is in the form: noisy data equal to the true mean level plus a random error where the true mean level is given by $\beta_0\, x(\mathbf{s}_i, t) + \xi(t)$ and the random error term, dependent in space and time, is given by $w(\mathbf{s}_i, t)$. Let $\boldsymbol{\beta}$ denote the unknown parameters $(\beta_0, \beta_1, \ldots, \beta_{24})$ and $p = 25$ denote the dimensionality of $\boldsymbol{\beta}$.

The error term $w(\mathbf{s}_i, t)$ is assumed to be a zero-mean spatio-temporal process with a separable covariance structure, given by:

$$\mathrm{Cov}\big\{w(\mathbf{s}_i, t_k),\ w(\mathbf{s}_j, t_l)\big\} = \sigma_w^2\ \rho_s(|\mathbf{s}_i - \mathbf{s}_j|; \phi_s)\ \rho_t(|t_k - t_l|; \phi_t). \tag{2}$$

We write $\mathbf{w}$ to denote the vector of all the nT $w(s_i, t)$'s. Let $H(\phi) = \Sigma_s \otimes \Sigma_t$ where the $n \times n$ spatial correlation matrix Σ_s has elements $\rho_s(|\mathbf{s}_i - \mathbf{s}_j|; \phi_s)$, for $i, j = 1, \ldots, n$ and $T \times T$ temporal correlation matrix Σ_t has elements $\rho_t(|t_k - t_l|; \phi_t)$, for $k, l = 1, \ldots, T$.

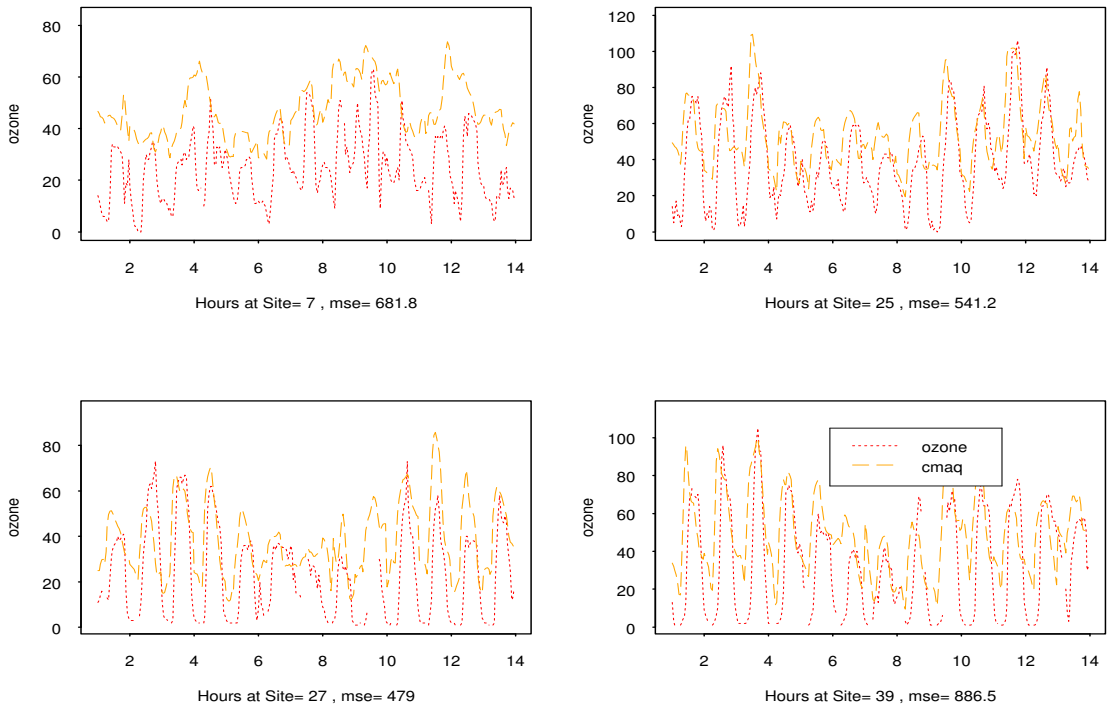

Fig. 2 Observed hourly data are dotted lines and Eta-CMAQ forecasts are dashed lines at four randomly chosen sites. The labels on the X-axis are the days in August, 2005. The mse for each plot is the mean square error between the data and the Eta-CMAQ forecasts.

This model reduces to the usual regression model with independent errors when we take $H(\phi) = I$, the identity matrix. This can be achieved by choosing $\rho_s(d;\phi_s) = \rho_t(d,\phi_t) = 1$ if $d = 0$ and 0 otherwise. This independent error regression model is compared with the spatio-temporal model in Section 4.

We take the two ρ's to be exponential covariance functions, i.e., $\rho_s(d;\phi_s) = \exp(-\phi_s|d|)$ and $\rho_t(d;\phi_t) = \exp(-\phi_t|d|)$. Ideally, $\boldsymbol{\phi} = (\phi_s, \phi_t)'$ should be estimated within the Bayesian model as well. However, in a classical inference setting it is not possible to consistently estimate all the parameters ϕ and σ^2 in a typical model for spatial data with a covariance function belonging to the Matèrn family, see Zhang (2004). Moreover, Stein (1999) shows that spatial interpolation is sensitive to the product $\sigma^2\phi$ but not to either one individually. In Section 4 we choose optimal values of $\boldsymbol{\phi}$ using a validation mean square error criterion and estimate the variances conditional on those values.

For convenience, we work with the precision $\tau_w^2 = 1/\sigma_w^2$. The joint prior distribution of $\boldsymbol{\beta}, \tau_w^2$ is assumed to be:

$$\pi(\boldsymbol{\beta}, \tau_w^2) = N\left(\boldsymbol{\beta}_m, \frac{V}{\tau_w^2}\right) G(a_w, b_w),$$

where $\boldsymbol{\beta}_m$, $p \times 1$, and V, $p \times p$, are suitable hyper-parameters and τ_w^2 follows the gamma distribution $G(a_w, b_w)$ with mean a_w/b_w. In our implementation we take $a_w = 2$ and $b_w = 1$ to have a proper prior specification. We take $\boldsymbol{\beta}_m$ to be the null vector and $V = 10^4 I$ to have a vague prior on the regression parameter $\boldsymbol{\beta}$.

3.1 Posterior distributions

Model (1) can be written as

$$\mathbf{Z} \sim N\left(X\boldsymbol{\beta}, \sigma_w^2 H\left(\phi\right)\right)$$

where X is the associated $nT \times p$ design matrix. The joint posterior distribution of $\boldsymbol{\beta}$ and τ_w^2, $\pi\left(\boldsymbol{\beta}, \tau_w^2|\mathbf{z}\right)$, is:

$$\begin{aligned}
&\propto \left(\tau_w^2\right)^{\frac{nT+p}{2}+a_w-1} \exp\left[-\tfrac{\tau_w^2}{2}\left(\mathbf{z}-X\boldsymbol{\beta}\right)' H^{-1}\left(\phi\right)\left(\mathbf{z}-X\boldsymbol{\beta}\right) - \tfrac{\tau_w^2}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_m)'V^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_m) - b_w\tau_w^2\right] \\
&\propto \left(\tau_w^2\right)^{\frac{nT+p}{2}+a_w-1} \exp\left[-\tfrac{\tau_w^2}{2}\left\{\left(\mathbf{z}-X\boldsymbol{\beta}\right)' H^{-1}\left(\phi\right)\left(\mathbf{z}-X\boldsymbol{\beta}\right) + (\boldsymbol{\beta}-\boldsymbol{\beta}_m)'V^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_m) + 2b_w\right\}\right].
\end{aligned}$$

Now we use the matrix identity:

$$(\mathbf{z}-X\boldsymbol{\beta})'H^{-1}(\phi)(\mathbf{z}-X\boldsymbol{\beta})+(\boldsymbol{\beta}-\boldsymbol{\beta}_m)'V^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_m)+2b_w = (\boldsymbol{\beta}-\boldsymbol{\beta}^*)'(V^*)^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)+2b_w^*$$

where

$$V^* = \left(V^{-1} + X'H^{-1}(\phi)X\right)^{-1}, \quad \boldsymbol{\beta}^* = V^*\left(V^{-1}\boldsymbol{\beta}_m + X'H^{-1}(\phi)\mathbf{z}\right)$$

and

$$b_w^* = b_w + \left\{\boldsymbol{\beta}_m'V^{-1}\boldsymbol{\beta}_m + \mathbf{z}'H^{-1}(\phi)\mathbf{z} - (\boldsymbol{\beta}^*)'(V^*)^{-1}(\boldsymbol{\beta}^*)\right\}/2.$$

Hence the joint posterior distribution is:

$$\pi\left(\boldsymbol{\beta}, \tau_w^2|\mathbf{z}\right) \propto \left(\tau_w^2\right)^{\frac{nT+p}{2}+a_w-1} \exp\left[-\frac{\tau_w^2}{2}\left\{(\boldsymbol{\beta}-\boldsymbol{\beta}^*)'(V^*)^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}^*) + 2b_w^*\right\}\right].$$

Now the full conditional posterior distributions are given by:

$$\begin{aligned}
&\boldsymbol{\beta}|\mathbf{z}, \tau_w^2 \sim N\left(\boldsymbol{\beta}^*, \sigma_w^2 V^*\right) \\
&\tau_w^2|\mathbf{z}, \boldsymbol{\beta} \sim G\left(\tfrac{nT+p}{2} + a_w, \tfrac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)'(V^*)^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}^*) + b_w^*\right).
\end{aligned}$$

By direct integration the marginal posterior distributions are obtained as follows:

$$\boldsymbol{\beta}|\mathbf{z} \sim t\left(\boldsymbol{\beta}^*, 2b_w^*\frac{V^*}{nT+2a_w}, nT+2a_w\right), \quad \tau_w^2|\mathbf{z} \sim G\left(nT/2+a_w, b_w^*\right) \tag{3}$$

where $\mathbf{Y} \sim t(\boldsymbol{\mu}, \Sigma, \nu)$ has the probability density function

$$f(\mathbf{y}|\boldsymbol{\mu}, \Sigma, \nu) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{p/2}}|\Sigma|^{-1/2}\left\{1 + \frac{(\mathbf{y}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})}{\nu}\right\}^{-(\nu+p)/2}.$$

In the univariate case with scalar μ and $\Sigma = \sigma^2$ we note the following two properties of this distribution which we shall require for prediction in the next sub-section:

$$E(Y^2) = \mu^2 + \sigma^2\frac{\nu}{\nu-2}, \mathrm{Var}(Y^2) = \frac{2\sigma^4\nu^2(\nu-1)}{(\nu-4)(\nu-2)^2} + 8\sigma^2\mu^2\frac{\nu}{\nu-2}, \tag{4}$$

when $\nu > 4$.

We use the marginal posterior distributions (3) to make inference. Specifically, $\boldsymbol{\beta}^*$ provides the point estimates for the parameter $\boldsymbol{\beta}$. We obtain a credible interval for the component, β_k, $k = 1, \ldots, p$ by using its marginal posterior distribution which is a t-distribution with $nT + 2a_w$ degrees of freedom having mean β_k^* and scale parameter λ_k^2 where $\lambda_k^2 = \frac{2b_w^*}{nT+2a_w}V_{kk}^*$ where V_{kk}^* is the kth diagonal entry of V^*. Now it is straightforward to see that an equal-tailed $(1-\alpha)100\%$ credible interval for β_k is given by

$$\beta_k^* \pm \lambda_k\, t_{\alpha/2;nT+2a_w}$$

where $P(T > t_{\alpha/2;nT+2a_w}) = \alpha/2$ when T follows the standard t-distribution with $nT+a_w$ degrees of freedom.

Similarly we estimate σ_w^2 by the posterior expectation

$$E(1/\tau_w^2|\mathbf{z}) = \frac{b_w^*}{nT/2 + a_w - 1}$$

which follows from the properties of the Gamma distribution. Here also we can find an equal tailed credible interval for σ^2 by using the probability identity

$$P(g_{\alpha/2;nT/2+a_w,\lambda} \leq \tau_w^2 \leq g_{1-\alpha/2;nT/2+a_w,\lambda}) = 1 - \alpha$$

where $g_{a,\nu,\lambda}$ is such that $P(Y < g_{a,\nu,\lambda}) = a$ for any $0 < a < 1$ when Y follows $G(\nu, \lambda)$.

3.2 Predictive distributions

Using the above models we can interpolate the spatial surface at any time point t' in the future or in the past. Let the p-dimensional vector of values of the regression variables at this new location-time combination be given by $\mathbf{x}_0$. We first construct the joint distribution:

$$\begin{pmatrix} Z(s',t') \\ \mathbf{Z} \end{pmatrix} \sim N\left\{ \begin{pmatrix} \mathbf{x}_0'\boldsymbol{\beta} \\ X\boldsymbol{\beta} \end{pmatrix},\ \sigma_w^2 \begin{pmatrix} 1 & \Sigma_{12} \\ \Sigma_{21} & H(\phi) \end{pmatrix} \right\},$$

where $\Sigma_{21} = \Sigma_{12}'$ and Σ_{12} is the nT dimensional vector with elements given by $\sigma_s(\mathbf{s}_i - \mathbf{s}')\sigma_t(t-t')$ where $\sigma_s(\mathbf{s}_i - \mathbf{s}') = \rho_s(|\mathbf{s}_i - \mathbf{s}'|;\phi_s)$ and $\sigma_t(t-t') = \rho_t(|t-t'|;\phi_t)$. Now we obtain the conditional distribution

$$Z(s',t')\,|\mathbf{z},\boldsymbol{\beta},\sigma_w^2 \sim N\left\{\mathbf{x}_0'\boldsymbol{\beta} + \Sigma_{12}H^{-1}(\phi)(\mathbf{z} - X\boldsymbol{\beta}), \sigma_w^2\left(1 - \Sigma_{12}H^{-1}(\phi)\Sigma_{21}\right)\right\}.$$

We need to integrate out $\boldsymbol{\beta}$ and τ_w^2 from the above distribution to obtain the required predictive distribution. To do this, we note that:

$$\begin{aligned} z(s',t') - \mathbf{x}_0'\boldsymbol{\beta} - \Sigma_{12}H^{-1}(\phi)(\mathbf{z} - X\boldsymbol{\beta}) &= z(s',t') - \Sigma_{12}H^{-1}(\phi)\mathbf{z} - \mathbf{x}_0'\boldsymbol{\beta} + \Sigma_{12}H^{-1}(\phi)X\boldsymbol{\beta} \\ &= z^*(\mathbf{s}',t') - (\mathbf{x}_0' - \Sigma_{12}H^{-1}(\phi)X)\boldsymbol{\beta} \\ &= z^*(\mathbf{s}',t') - \mathbf{g}'\boldsymbol{\beta} \end{aligned}$$

where

$$z^*(\mathbf{s}',t') = z(s',t') - \Sigma_{12}H^{-1}(\phi)\mathbf{z} \text{ and } \mathbf{g}' = \mathbf{x}_0' - \Sigma_{12}H^{-1}(\phi)X.$$

Therefore,

$$\pi(Z(s',t')\,|\mathbf{z},\boldsymbol{\beta},\sigma_w^2) \propto (\tau_w^2)^{1/2}\exp\left[-\tfrac{\tau_w^2}{2C(\mathbf{s}',t')}\left\{z^*(\mathbf{s}',t') - \mathbf{g}'\boldsymbol{\beta}\right\}^2\right]$$

where

$$C(\mathbf{s}',t') = 1 - \Sigma_{12}H^{-1}(\phi)\Sigma_{21}.$$

This shows that

$$Z^*(\mathbf{s}',t')|\mathbf{z},\boldsymbol{\beta},\tau_w^2 \sim N\left(\mathbf{g}'\boldsymbol{\beta}, \sigma_w^2 C(\mathbf{s}',t')\right).$$

Hence by integrating out $\boldsymbol{\beta}$ we have

$$Z^*(\mathbf{s}',t')|\mathbf{z},\tau_w^2 \sim N\left(\mathbf{g}'\boldsymbol{\beta}^*, \sigma_w^2(C(\mathbf{s}',t') + \mathbf{g}'V^*\mathbf{g})\right).$$

By integrating this with respect to the marginal posterior distribution of τ_w^2 in Equation (3), we obtain the posterior predictive distribution of Z^* given $\mathbf{z}$ as:

$$Z^*(\mathbf{s}',t')|\mathbf{z} \sim t(\mathbf{g}'\boldsymbol{\beta}^*, 2b_w^*\frac{C(\mathbf{s}',t') + \mathbf{g}'V^*\mathbf{g}}{nT + 2a_w}, nT + 2a_w).$$

Now the posterior predictive distribution of $Z(\mathbf{s}', t')|\mathbf{z}$ is obtained as:

$$Z(\mathbf{s}', t')|\mathbf{z} \sim t\left(\mathbf{x}_0'\boldsymbol{\beta}^* + \Sigma_{12}H^{-1}(\phi)(\mathbf{z} - X\boldsymbol{\beta}^*), 2b_w^* \frac{C(\mathbf{s}', t') + \mathbf{g}'V^*\mathbf{g}}{nT + 2a_w}, nT + 2a_w\right). \quad (5)$$

Observe that we model ozone on the square root scale. Hence the predictions using the posterior predictive distribution (5) will be on the square-root scale as well. We can predict on the original scale by evaluating:

$$\begin{aligned} E(Z^2(\mathbf{s}', t')|\mathbf{z}) &= \{E(Z(\mathbf{s}', t')|\mathbf{z})\}^2 + \mathrm{Var}\{Z(\mathbf{s}', t')|\mathbf{z})\} \\ &= \left\{\mathbf{x}_0'\boldsymbol{\beta}^* + \Sigma_{12}H^{-1}(\phi)(\mathbf{z} - X\boldsymbol{\beta}^*)\right\}^2 + 2b_w^* \frac{C(\mathbf{s}', t') + \mathbf{g}'V^*\mathbf{g}}{nT + 2a_w - 2}, \end{aligned}$$

See the Appendix for simplified expressions for $\Sigma_{12}H^{-1}(\phi)$ and $\Sigma_{12}H^{-1}(\phi)\Sigma_{21}$ and (4) for properties of the t-distribution. The variance of the prediction, $\mathrm{Var}(Z^2(\mathbf{s}', t')|\mathbf{z})$, is calculated using the expression for variance also noted down in (4).

To obtain the prediction intervals we can adopt one of the two approaches. The first method is to find a Monte Carlo estimate of the interval by sampling from the t-distribution (5) and using appropriate averages. This approach, however, will be slower than the method based on the normal approximation for the square of the t-distribution (5) we adopt here. The approximation is justified by the fact that the degrees of freedom $nT + 2a_w$ is very large (more than 2500 in our application). The approximate 95% prediction interval is given by

$$E(Z^2(\mathbf{s}', t')|\mathbf{z}) \pm 1.96 \times \sqrt{\mathrm{Var}(Z^2(\mathbf{s}', t')|\mathbf{z})}.$$

3.3 Predicting the 8-hour map at the current hour

A useful application of the proposed methods is the ability to predict the 8-hour average ozone concentration at the current hour. In the EPA AIRNow environment, the 8-hour average ozone concentration at the current hour t is the simple average of the 8-hourly concentrations at the current hour t, four past hours $(t-1, t-2, t-3, t-4)$, and three future hours $(t+1, t+2$, and $t+3)$. Accordingly, the 8-hour ozone level at time t, location $\mathbf{s}'$ is given by:

$$O_8(\mathbf{s}', t) = \frac{1}{8}\sum_{k=-4}^{3} Z^2(\mathbf{s}', t+k).$$

(Here we use Z^2 since ozone is modeled in the square-root scale.) Note that at any unobserved site $\mathbf{s}'$, $Z^2(\mathbf{s}', t)$ for any t is the square of the non-central t-distribution with parameters as given in (5). The posterior predictive distribution of $O_8(\mathbf{s}', t)$, defined as the sum of the non-central F-distributed random variables, is not available in closed form. As a result, we use Monte Carlo simulation to find the mean and standard deviation of the posterior predictive distribution of $O_8(\mathbf{s}', t)$ given the observed data $\mathbf{z}$ as follows. We generate a large number B of independent random variables, $Z^{(j)}(\mathbf{s}', t+k), j = 1, \ldots, B$ for each $k = -4, -3, \ldots, 3$ at each hour t at the given location $\mathbf{s}'$. Now we obtain $O_8^{(j)}(\mathbf{s}', t) = \frac{1}{8}\sum_{k=-4}^{3} {Z^{(j)}}^2(\mathbf{s}', t+k)$ for each $j = 1, \ldots, B$. The 8-hour average is estimated by the sample mean, $\bar{O}_8(\mathbf{s}', t) = B^{-1}\sum_{j=1}^{B} O_8^{(j)}(\mathbf{s}', t)$ and the sample standard deviation of $O_8^{(j)}(\mathbf{s}', t)$ is used as an uncertainty estimate of the posterior predictive distribution. In the Monte Carlo simulation, we use independent samples ${Z^{(j)}}^2(\mathbf{s}', t+k)$, $k = -4, \ldots, 3$ for each j. In effect, we perform marginal predictions of $Z^2(\mathbf{s}', t+k)$ for each k, just as we do marginal predictions at all the different locations $\mathbf{s}'$ in the predictive grid of 3000 sites. Joint predictions and forecasting is computationally prohibitive in the instantaneous prediction problem of this paper and are not pursued here.

4 Analysis

We use the set-aside validation data from 40 stations to select the decay parameters ϕ_s and ϕ_t. Let $\hat{Z}^2(\mathbf{s}_i^*, t)$ denote the model based validation estimate for $Z^2(\mathbf{s}_i^*, t)$ where $\mathbf{s}_i^*$ denotes the ith validation site, $i = 1, \ldots, 40$. Again, recall that we model ozone in the square root scale. The validation mean-square error is given by

$$\text{VMSE} = \frac{1}{n_v} \sum_{i=1}^{40} \sum_{t=1}^{T} \left(Z^2(\mathbf{s}_i^*, t) - \hat{Z}^2(\mathbf{s}_i^*, t) \right)^2 I(Z(\mathbf{s}_i^*, t)) \tag{6}$$

where $I(Z(\mathbf{s}_i^*, t)) = 1$ if $Z(\mathbf{s}_i^*, t)$ is available, and 0 otherwise, and $n_v = \sum_{i=1}^{40} \sum_{t=1}^{T} I(Z(\mathbf{s}_i^*, t))$ is the total number of available observations at the 40 validation sites. For ϕ_s, we searched for the optimal value in a grid formed of the values of ϕ_s corresponding to the spatial ranges of 50, 250, 500 and 1000 kilometers. For the temporal decay parameter ϕ_t, we searched for the optimum value in a grid formed of the values of ϕ_t corresponding to the temporal ranges of 3, 6, 9, 12 and 24 hours.

As described in Section 2, we model data for a running window of seven days, so $T = 168(7 \times 24)$ starting at one of the five hours between 2PM and 6PM. We do this for each of the six starting days, August 2–7. Thus, we have performed model fitting for 30 start day and start hour combinations. For each of the 30 model fitting combinations we choose the optimal values of ϕ_s and ϕ_t from a fresh grid search using the VMSE criterion in (6). The average spatial and temporal ranges were approximately 600 kilometers and 20 hours, respectively, providing evidence of strong spatial and temporal dependencies.

Corresponding to each of the 30 model fitting combinations we have calculated the validation interpolations at the 40 sites for the last hour of data and for each of the next three future hours. For example, starting at 3PM of August 4th, we model $T = 168$ hourly observations from 3PM on August 4th to 2PM on August 10 using data from all the 350 monitoring sites. Predictions are obtained at the 40 validation sites for the last hour, 2PM on August 10, and forecasts at these sites are calculated for 3PM, 4PM and 5PM on August 10.

Figure 3 provides a comparison of spatial predictions obtained from our model, Eta-CMAQ, and the regression model with independent error distribution assumption by taking $H(\phi) = I$, see the discussion below (2). In terms of VMSE, our proposed Bayesian space-time model clearly outperforms the other two approaches. We note that VMSE tends to increase as the length of the forecast period increases. Table 1 details uniform reductions in mean-square error that result in using the proposed model relative to the regression model.

Figures 4 and 5 provide detailed validation plots for times when our model performs the best, and the worst. In both cases, we see that the Eta-CMAQ forecasts are upwardly biased. The Bayesian model predictions are closer to the observations relative to the other two approaches (the $y = x, 45^o$, line is superimposed). The validation plot for the 8-hour averages is provided in Figure 6 where again similar conclusions are drawn. In these three plots, it may seem that the Bayesian model is producing slightly biased predictions. Predicted values are a bit high when the observed values are low, and low when the observed values are high. This can be attributed to the 'regression to the mean' effect of the simple spatio-temporal regression model we have adopted here for instantaneous prediction and forecasting. To address this issue we have included higher order polynomial terms for the Eta-CMAQ forecasts in the model. However, this change did not improve the validation predictions and made the linear term for Eta-CMAQ forecasts non-significant. We have also examined the nominal coverage of the 95% prediction intervals and that has turned out to be adequate, see particular results in the next section.

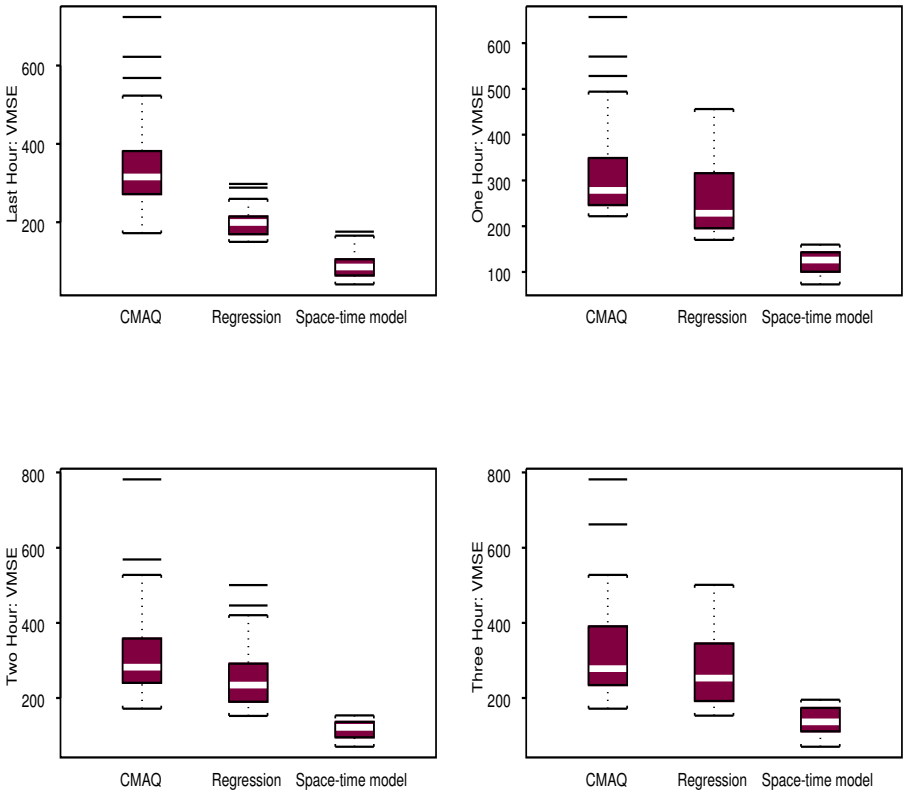

Fig. 3 Boxplots of the VMSE's: top left for the last modelling hour, top right for the one-hour, bottom left for the two-hours and bottom right for the three hours ahead forecasts.

Parameter estimates are illustrated in Table 2 and Figure 7 when the data to be modeled is started at 3PM on August 3. The regression parameter β_0 for the Eta-CMAQ output co-variate is significant showing positive association between Eta-CMAQ output and hourly ozone concentration values. The β for the midnight hour is taken as the baseline. The estimates of β_j for all other hours are relative to the β_{24} for the midnight hour. The estimates of the 23 βs corresponding to 1AM to 11PM (Figure 7) show that the mean hourly ozone concentration diminishes until 9AM in the morning and then it starts to rise with the peak reaching around 4PM and 5PM. We observed similar parameter estimates for other starting date and hour, especially for the variance parameter σ_w^2.

We illustrate the 3-hours ahead forecasts at 2PM on August 13 in Figures 8 and 9. (The graphics parameters and the color-key bars are same for the three forecasts maps in these two figures.) Here we use the seven days data ending at 2PM on August 11th to do the model fitting and then obtain the 3-hours ahead forecasts using the fitted model. The Eta-CMAQ forecasts predict generally higher level of ozone concentrations than those from the independent error regression model (Figure 8) and the proposed spatio-temporal model (left panel of Figure 9). The superimposed actual ozone observations make clear that our model greatly improves the forecasts using both the Eta-CMAQ and independent error regression model. In fact, the VMSE's are 240.58, 187.84 and 70.45, respectively for

Table 1 Mean-Square Error Reductions using Bayesian Space-Time Model relative to the Independent Error Regression Model

Day	Start Hour				
	2PM	3PM	4PM	5Pm	6PM
Last hour interpolation					
Aug 2	124.7	124.8	141.0	103.0	84.3
Aug 4	94.1	99.6	118.6	141.4	192.4
Aug 6	117.4	82.1	65.1	74.2	148.0
One hour ahead forecasts					
Aug 2	80.0	76.9	63.7	86.4	137.6
Aug 4	117.5	144.4	206.1	282.3	327.8
Aug 6	118.5	125.0	150.6	160.4	150.7
Two hours ahead forecasts					
Aug 2	88.1	71.3	40.4	82.7	148.3
Aug 4	119.9	137.1	199.1	298.7	375.1
Aug 6	104.4	117.0	171.3	190.3	141.6
Three hours ahead forecasts					
Aug 2	44.5	40.3	72.0	125.1	168.1
Aug 4	109.2	164.6	267.0	339.3	277.6
Aug 6	117.4	172.5	182.2	127.9	114.1

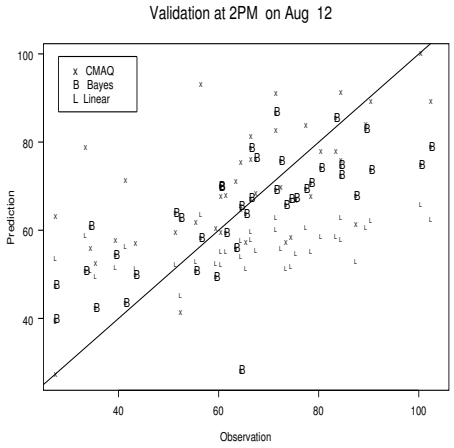

Fig. 4 The validation plot for 2PM on August 12 when the proposed model performs the worst.

Table 2 Parameter estimates when the modeled data starts at 3PM on August 3.

	Estimate	95% Credible interval
β_0	0.163	(0.150, 0.176)
β_{24}	3.451	(3.252, 3.650)
σ_w^2	2.532	(2.503, 2.561)

the three models. The standard deviation of the model based predictions are plotted on the right panel of Figure 9. Since a constant-variance model was fit on the square-root scale, it is not surprising that, after transformation back to the original scale, larger standard deviations of prediction are associated with larger predicted values.

We illustrate the 8-hour average map predictions at 3PM on August 11th in Figures 10 and 11. (The graphics parameters and the color-key bars are same for the three forecast maps in these two figures.) We use the seven days data ending at 3PM on August

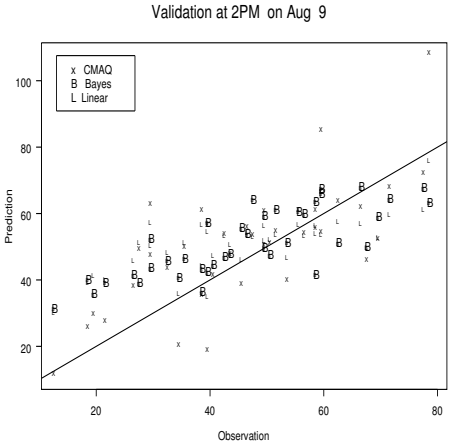

Fig. 5 The validation plot for 2PM on August 9 when the proposed model performs the best.

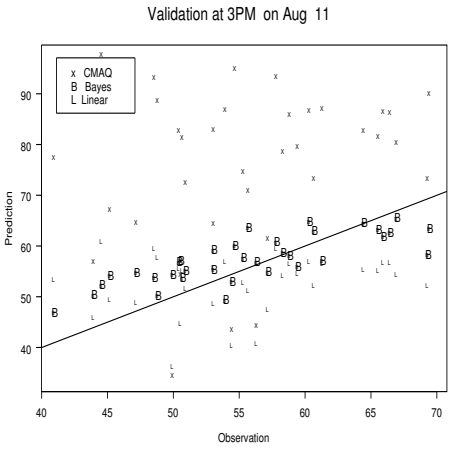

Fig. 6 The validation plot for the 8-hour average at 3PM on August 11th.

11th to do the model fitting and forecasting. In particular, to calculate the 8-hour average prediction at 3PM we spatially interpolate the value at each of the five hours 11AM-3PM and temporally forecast for the three hours 4-6PM, see Section 3.3 for the details. Our model clearly outperforms both the Eta-CMAQ and the independent error regression model; the VMSE's are 605.87, 80.89 and 22.40, respectively for the Eta-CMAQ, the independent error regression model and the proposed model. See also the validation plot in Figure 6 for the 8-hour averages. The standard deviation of the model based predictions are provided in the right panel of Figure 11. The 8-hour average predictions have smaller uncertainties than the 3-hour ahead forecasts, as expected.

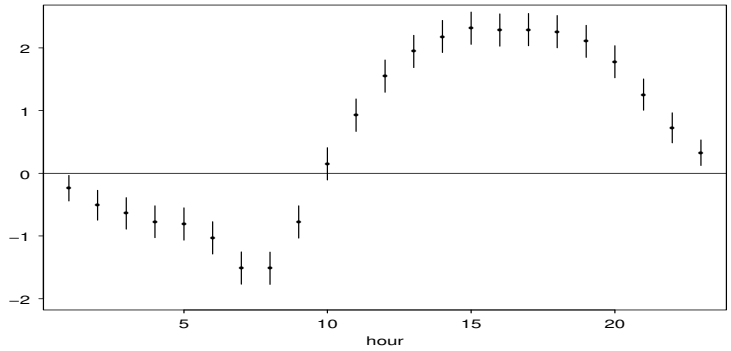

Fig. 7 The posterior means and 95% credible intervals for the parameters $\beta_1, \ldots, \beta_{23}$ corresponding to the hours 1AM to 11PM.

5 Comparison with the dynamic modeling approach

Sahu *et al.* (2009) have proposed a dynamic model approach for forecasting the next day's 8-hour ozone concentration based on modeling daily data. Their model requires 5-7 hours computing time and, as a result, is not suitable for the instantaneous forecasting problem of this paper. However, the current model with suitable modifications for differences in time unit can be fitted to the daily data and the forecasts can be compared. This is taken up below.

To adapt the current model for daily data we must replace the hourly intercept term $\xi(t)$ by an overall intercept ξ from (1) and treat the time unit t as daily. Hence, the hourly model adapted for daily data is given by:

$$Z(\mathbf{s}_i, t) = \beta_0\, x(\mathbf{s}_i, t) + \xi + w(\mathbf{s}_i, t), \tag{7}$$

for $i = 1, \ldots, n,\ t = 1, \ldots, T$. The specification for the space-time error term $w(\mathbf{s}_i, t)$ can remain to be the same separable process, although the temporal correlation now will have interpretations in daily time units. This model is to be compared with the Sahu *et al.* (2009) model given by:

$$Z(\mathbf{s}_i, t) = O(\mathbf{s}_i, t) + \epsilon(\mathbf{s}_i, t), \tag{8}$$

for $i = 1, \ldots, n,\ t = 1, \ldots, T$, where $\epsilon(\mathbf{s}_i, t)$ is a white noise process, assumed to follow $N(0, \sigma^2_\epsilon)$ independently. The model for $O(\mathbf{s}_i, t)$ has been assumed as:

$$O(\mathbf{s}_i, t) = \xi + \rho\, O(\mathbf{s}_i, t-1) + \beta_0\, x(\mathbf{s}_i, t) + \eta(\mathbf{s}_i, t),$$

for $i = 1, \ldots, n,\ t = 1, \ldots, T$ where ξ, ρ and β_0 are unknown parameters and $\eta(\mathbf{s}_i, t)$ is a spatially correlated, but temporally independent error term. The grand mean of the data has been chosen as the initial condition for $O(\mathbf{s}, 0)$.

A few remarks regarding the differences between the two sets of models are now appropriate. The current model (7) does not incorporate the top level white noise process (usually called the nugget term) and only has one variance parameter for $w(\mathbf{s}_i, t)$ while the dynamic model has two variance components one for each of $\epsilon(\mathbf{s}_i, t)$ and $\eta(\mathbf{s}_i, t)$. There is a concept of true process given by $O(\mathbf{s}_i, t)$ in the dynamic models while there is only a concept of mean process described by the mean function $\beta_0\, x(\mathbf{s}_i, t) + \xi$ in the current model. The two sets of models require different sets of hyper-parameters; the current model has

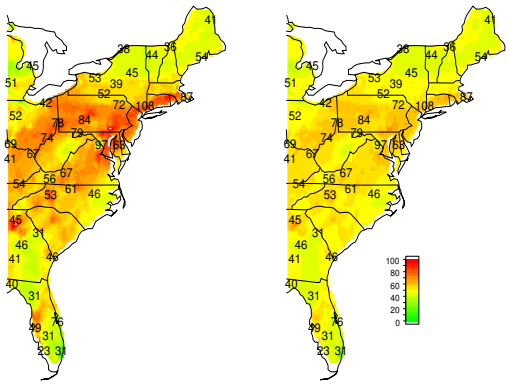

Fig. 8 The 3-hours ahead forecasts at 2PM on 13th August: left panel is using the Eta-CMAQ and right panel is using an independent error regression model. Observed values from some selected sites are superimposed. (For visual clarity we present only a subset of the monitoring data.)

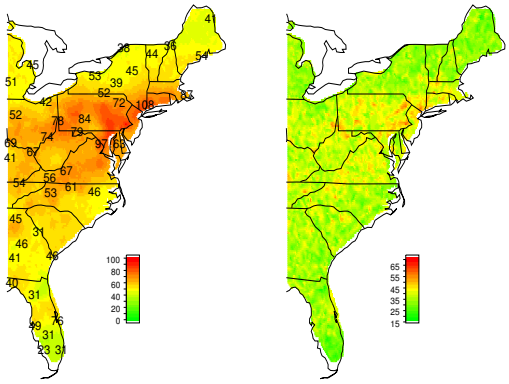

Fig. 9 Left panel is the model based 3-hours ahead forecasts at 2PM on August 13th and the right panel provides the standard deviation map. Observed values from some selected sites are superimposed. (For visual clarity we present only a subset of the monitoring data.)

two spatial and temporal correlation decay parameters ϕ_s and ϕ_t, see (2) and the hyper-parameters in the prior for β_0, ξ and σ_w^2. On the other hand, the dynamic models require a spatial decay parameter for the correlation of the $\eta(\mathbf{s}, t)$ process and the hyper-parameters for the prior of β_0, ξ, ρ, σ_ϵ^2 and σ_w^2. The introduction of the hierarchy in the dynamic models is advantageous in handling missing data since those are simply filled up using the top-level model (8) in each iteration of the MCMC algorithm. The missing data in the current model must be imputed beforehand so that the exact computing methods of Section 3 can be implemented. The expected consequence of these modeling assumptions is that the forecasts using the dynamic models can be more accurate than those from the current set of models.

To illustrate the differences between the two sets of models we compare the mean-square errors and the median of the standard deviations of the predictions and the nominal coverage of the 95% prediction intervals for 3 different 7-day data sets and one day ahead forecasts published in Sahu *et al.* (2009) in their Table 1. We compare the three sets of model fitting and forecasting for August 3-12; we do not compare the data for August 2 and 13 since those were unavailable. The spatial decay parameter was chosen the same for both the hourly and daily model. The temporal decay parameter for the hourly model was

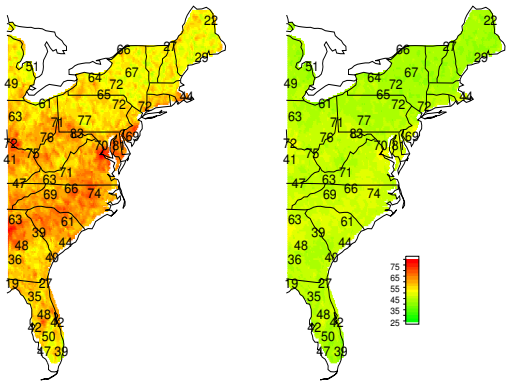

Fig. 10 Left panel is for the 8-hour average Eta-CMAQ map at 3PM on August 11th and right panel is for the same map using an independent error regression model. Observed values from some selected sites are superimposed. (For visual clarity we present only a subset of the monitoring data.)

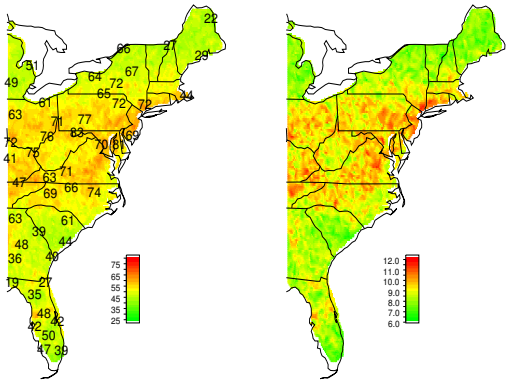

Fig. 11 Left panel is for the 8-hour average model based map at 3PM on August 11th and right panel is the standard deviation map. Observed ozone values from some selected sites are superimposed. (For visual clarity we present only a subset of the monitoring data.)

Table 3 Properties of the Validation predictions. Med.sd are the median of the standard deviation of the predictions and Cover is the nominal coverage of the 95% prediction intervals.

Days	Daily Model			Hourly Model		
	VMSE	Med.sd	Cover	VMSE	Med.sd	Cover
Aug 3–10	50.0	34.6	0.98	51.3	27.0	1.00
Aug 4–11	64.5	34.0	0.95	68.1	26.7	1.00
Aug 5–12	62.1	33.4	0.98	49.0	26.1	0.98

chosen as 0.13 corresponding to roughly 24 hours decay in temporal correlation. Table 3 provides the results for the daily and the hourly model applied to daily data. The dynamic daily model has better VMSE results except for the last set of days. The standard deviation of the predictions are smaller for the simpler hourly model of this paper. This is expected since the more complex daily model will imply higher prediction variability. The nominal coverages from both models are seen to be adequate. We may see better prediction performance by the daily model than the hourly model, however, the daily model is unsuitable for the instantaneous hourly prediction problem since it requires longer computing time.

6 Summary

As we enter a new age where air pollution data can be accessed in real-time, new space-time models are needed to provide continuous, updated maps of current and future air pollution levels. This is a departure from previous research on retrospective space-time analyses of air pollution data where computational time for fitting models is not a constraint. To meet this modeling challenge, we develop a flexible Bayesian spatial-temporal model which can be fit by exact methods, eliminating the need to use computationally intensive MCMC. Given the limited time available to provide credible predictions in a real-time environment, we use a minimum number of parameters in the model, but do account for ozone diurnal variation, the influence of Eta-CMAQ numerical output, and space-time random variation. Validation analyses show that this model provides improved predictions of hourly ozone spatial patterns, and can be used to predict 8-hour average ozone concentrations surrounding any hour of the day, including 3 forecasted hours in the 8-hour average. Future efforts will focus on improving the model further by incorporating spatially varying coefficients for Eta-CMAQ to evaluate regional effects and thereby adjust for possible regional biases. It will be also interesting to see how the model will perform for other pollutants such as the fine particulate matter.

Appendix: Simplifying the expressions: $\boldsymbol{\Sigma_{12}H^{-1}(\phi)}$ and $\boldsymbol{\Sigma_{12}H^{-1}(\phi)\Sigma_{21}}$

Note that

$$\begin{pmatrix} 1 & \Sigma_{12} \\ \Sigma_{21} & H(\phi) \end{pmatrix} = \begin{pmatrix} 1 & \Sigma_s'(\mathbf{s}-\mathbf{s}') \otimes \Sigma_t'(\mathbf{t}-t') \\ \Sigma_s(\mathbf{s}-\mathbf{s}') \otimes \Sigma_t(\mathbf{t}-t') & \Sigma_s \otimes \Sigma_t \end{pmatrix}$$

where $\Sigma_s(\mathbf{s}-\mathbf{s}')$ is an $n\times 1$ column vector with the ith entry given by $\sigma_s(\mathbf{s}_i-\mathbf{s}')$ and $\Sigma_t(\mathbf{t}-t')$ is a $T\times 1$ column vector with the kth entry given by $\sigma_t(t-t')$. Here $H^{-1}(\phi) = \Sigma_s^{-1}\otimes\Sigma_t^{-1}$. Hence the $1\times nT$ vector $\Sigma_{12}H^{-1}(\phi)$ will have elements (for $j=1,\ldots,n$ and $k=1,\ldots,T$)

$$\begin{aligned} b_{jk}(\mathbf{s}',t') &= \sum_{i=1}^{n}\sum_{m=1}^{T}\sigma_s(\mathbf{s}_i-\mathbf{s}')\sigma_t(m-t')(\Sigma_s)_{ij}^{-1}(\Sigma_t)_{mk}^{-1} \\ &= \sum_{i=1}^{n}\sigma_s(\mathbf{s}_i-\mathbf{s}')(\Sigma_s)_{ij}^{-1}\sum_{m=1}^{T}\sigma_t(m-t')(\Sigma_t)_{mk}^{-1} \\ &= b_s(j,\mathbf{s}')\ b_t(k,t'), \end{aligned}$$

where

$$b_s(j,\mathbf{s}') = \sum_{i=1}^{n}\sigma_s(\mathbf{s}_i-\mathbf{s}')(\Sigma_s)_{ij}^{-1}, \text{ and } b_t(k,t') = \sum_{m=1}^{T}\sigma_t(m-t')(\Sigma_t)_{mk}^{-1}.$$

The quantity $b_t(k,t')$ simplifies considerably by noting that it resembles the inner product of a multiple of a particular column of Σ_t and a particular row of Σ_t^{-1}. First, consider the case $t'\le T$. In this case $b_t(k,t')$ is the inner product of the t'th column of Σ_t and kth row of Σ_t^{-1}. Hence $b_t(k,t')$ will be 1 if $t'=k$ and 0 otherwise. Now consider the case $t'>T$. Suppose that we can write

$$\sigma_t(m-t') = \sigma_t(t'-T)\sigma_t(T-m) \tag{9}$$

for $m = 1, \ldots, T$, thus $b_t(k, t')$ will be $\sigma_t(t'-T)$ times the inner product of the Tth column of Σ_t and kth row of Σ_t^{-1}. Observe that (9) holds for the exponential covariance function adopted here. Thus we have proved the following result:

$$b_t(k, t') = \begin{cases} \delta_{k,t'}, & \text{if } t' \leq T \\ \delta_{k,T}\, \sigma_t(t' - T), & \text{if } t' > T \end{cases}$$

where $\delta_{i,j} = 1$ if $i = j$ and 0 otherwise.

Now we obtain simplified expressions for a quantity like $\Sigma_{12} H^{-1}(\phi)\, \mathbf{a}$ where $\mathbf{a}$ is nT by 1 with elements a_{jk}, $j = 1, \ldots, n$ and $k = 1, \ldots, T$. We have:

$$\begin{aligned} \Sigma_{12} H^{-1}(\phi)\, \mathbf{a} &= \textstyle\sum_{j=1}^{n} \sum_{k=1}^{T} b_{jk}(\mathbf{s}', t') a_{jk} \\ &= \textstyle\sum_{j=1}^{n} \sum_{k=1}^{T} a_{jk} b_s(j, \mathbf{s}')\; b_t(k, t') \\ &= \textstyle\sum_{j=1}^{n} b_s(j, \mathbf{s}') \sum_{k=1}^{T} a_{jk} b_t(k, t'). \end{aligned}$$

Now

$$\sum_{k=1}^{T} a_{jk} b_t(k, t') = \begin{cases} \sum_{k=1}^{T} a_{jk} \delta_{k,t'}, & \text{if } t' \leq T \\ \sum_{k=1}^{T} a_{jk} \delta_{k,T}\, \sigma_t(t' - T), & \text{if } t' > T. \end{cases}$$

Thus we have,

$$\sum_{k=1}^{T} a_{jk} b_t(k, t') = \begin{cases} a_{jt'}, & \text{if } t' \leq T \\ a_{jT}\, \sigma_t(t' - T), & \text{if } t' > T. \end{cases}$$

Finally,

$$\Sigma_{12} H^{-1}(\phi)\, \mathbf{a} = \begin{cases} \sum_{j=1}^{n} b_s(j, \mathbf{s}') a_{jt'}, & \text{if } t' \leq T \\ \sigma_t(t' - T) \sum_{j=1}^{n} b_s(j, \mathbf{s}') a_{jT} & \text{if } t' > T. \end{cases}$$

Now we simplify the expression for the conditional variance. Note that $\Sigma_{12} H^{-1}(\phi)\, \Sigma_{21}$ is exactly equal to $\Sigma_{12} H^{-1}(\phi)\, \mathbf{a}$ where $\mathbf{a} = \Sigma_{21}$. For this choice we have, $a_{jt} = \sigma_s(\mathbf{s}_j - \mathbf{s}')\sigma_t(t - t')$. Hence,

$$\Sigma_{12} H^{-1}(\phi)\, \Sigma_{21} = \begin{cases} \sum_{j=1}^{n} b_s(j, \mathbf{s}') \sigma_s(\mathbf{s}_j - \mathbf{s}') \sigma_t(t' - t'), & \text{if } t' \leq T \\ \sigma(t' - T) \sum_{j=1}^{n} b_s(j, \mathbf{s}') \sigma_s(\mathbf{s}_j - \mathbf{s}') \sigma_t(T - t'), & \text{if } t' > T. \end{cases}$$

Let

$$a_s(\mathbf{s}') = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_s(\mathbf{s}_i - \mathbf{s}') \left(\Sigma_s^{-1}\right)_{ij} \sigma_s(\mathbf{s}_j - \mathbf{s}').$$

Thus,

$$\Sigma_{12} H^{-1}(\phi)\, \Sigma_{21} = \begin{cases} a_s(\mathbf{s}'), & \text{if } t' \leq T \\ a_s(\mathbf{s}') \sigma_t^2(t' - T) & \text{if } t' > T. \end{cases}$$

Now

$$C(\mathbf{s}', t') = 1 - a_s(\mathbf{s}')\; a_t(t'),$$

where

$$a_t(t') = \begin{cases} 1, & \text{if } t' \leq T \\ \sigma_t^2(t' - T) & \text{if } t' > T. \end{cases}$$

References

Cowles, M. K. and Zimmerman, D. L. (2003). A Bayesian space-time analysis of acid deposition data combined from two monitoring networks. *Journal of Geophysical Research-Atomospheres*. **108**, doi: 10.1029/2003JD004001.

Fuentes, M. and Raftery, A. (2005). Model Evaluation and Spatial Interpolation by Bayesian Combination of Observations with outputs from Numerical Models. *Biometrics*, **61**, 36–45.

Gelfand, A. E. and Sahu, S. K. (2009). Combining Monitoring Data and Computer model Output in Assessing Environmental Exposure. To appear in, *Handbook of Applied Bayesian Analysis*, edited by O'Hagan, A. and West, M.

Jun, M. and Stein, M. L. (2004). Statistical comparison of observed and CMAQ modeled daily sulfate levels. *Atmospheric Environment*, **38**, 4427–4436.

Kang, D., Mathur, R., Rao, S. T., and Yu, S. (2008). Bias adjustment techniques for improving ozone air quality forecasts. Journal of Geophysical Research, 113, D23308, doi:10.1029/2008JD010151.

McMillan, N., Holland, D. M., and Morara, M., and Feng, J. (2008). Combining numerical model output and particulate data using Bayesian space-time modeling. *Environmetrics*, to appear.

Sahu, S. K., Gelfand, A. E. and Holland, D. M. (2009). Fusing point and areal space-time data with application to wet deposition. *J. Roy. Stat. Soc.*, Series C, to appear.

Sahu, S. K., Yip, S. and Holland, D. M. (2009). Improved space-time forecasting of next day ozone concentrations in the eastern U.S. *Atmospheric Environment*, **43**, 494-501, DOI information: 10.1016/j.atmosenv.2008.10.028.

Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging:* Springer Verlag.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, **99**, 250–261.

Zimmerman, D. L. and Holland, D. M. (2005). Complementary co-kriging: spatial prediction using data combined from several environmental monitoring networks. *Environmetrics*, **16**, 219–234.

Disclaimer:

The U.S. Environmental Protection Agency's Office of research and Development partially collaborated in the research described here. Although it has been reviewed by EPA and approved for publication, it does not necessarily reflect the Agency's policies or views.