Available online at www.sciencedirect.com
SCIENCE DIRECT
Atmospheric Environment 40 (2006) 4883–4893
ATMOSPHERIC ENVIRONMENT
www.elsevier.com/locate/atmosenv

A Bayesian statistical approach for the evaluation of CMAQ

Jenise L. Swall[a,*,1], Jerry M. Davis[b,2]

[a]Atmospheric Sciences Modeling Division, Air Resources Laboratory, National Oceanic and Atmospheric Administration, MD E243-01, RTP, NC 27711, USA
[b]Atmospheric Modeling Division, U.S. Environmental Protection Agency, MD E243-01, RTP, NC 27711, USA

Received 3 March 2005; received in revised form 17 October 2005; accepted 22 December 2005

Abstract

Bayesian statistical methods are used to evaluate Community Multiscale Air Quality (CMAQ) model simulations of sulfate aerosol over a section of the eastern US for 4-week periods in summer and winter 2001. The observed data come from two U.S. Environmental Protection Agency data collection networks. The statistical methods used here address two problems that arise in model evaluation: the sparseness of the observational data which is to be compared to the model output fields and the comparison of model-generated grid cell averages with point-referenced monitoring data. A Bayesian hierarchical model is used to estimate the true values of the sulfate concentration field. Emphasis is placed on modeling the spatial dependence of sulfate over the study region, and then using this dependence structure to estimate average grid cell values for comparison with CMAQ. For the winter period, CMAQ tends to underpredict the sulfate concentrations over a large portion of the region. The CMAQ simulations for the summer period do not show this systematic underprediction of sulfate concentrations.

Keywords: Spatial analysis; Bayesian methods; Air quality model; Aerosol sulfate; Model validation

1. Introduction

The U.S. Environmental Protection Agency (U.S. EPA) has a long standing interest in monitoring and controlling particulate matter levels in the atmosphere. Numerous studies in recent years have examined the connection between high particulate pollution levels and human morbidity and mortality (see, for example, Dominici et al., 2000; Smith et al., 2000; Styer et al., 1995). Particulate pollution is also known to contribute to regional climate changes, visibility impairment, and acidic deposition (U.S. Environmental Protection Agency, 1995). U.S. EPA has been mandated by the Clean Air Act (and its amendments) to develop National Ambient Air Quality Standards (NAAQS) for particulate matter.

The purpose of the current research is to show how an advanced statistical technique can be used to evaluate the ability of an air quality model to simulate pollution levels in the atmosphere. The

*Corresponding author. Tel.: +1 919 541 7655; fax: +1 919 541 1379.
E-mail addresses: Jenise.Swall@noaa.gov (J.L. Swall), Davis.Jerry@epa.gov (J.M. Davis).
[1]On assignment to National Exposure Research Laboratory, U.S. Environmental Protection Agency, MD E243-01, RTP, NC 27711, USA.
[2]Visiting from Department of Marine, Earth & Atmospheric Sciences, Box 8208, North Carolina State University, Raleigh, NC 27695, USA.

1352-2310/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2005.12.058

model considered here is the Community Multiscale Air Quality (CMAQ) model developed by the U.S. EPA. A complete description of the science contained in CMAQ can be found in Byun and Ching (1999). For the purposes of this study, we focus on sulfate aerosol, since typically the distribution of this pollutant is more homogeneous spatially than many other pollutants of interest. The gas-phase and aqueous-phase chemistry of sulfate production is discussed in detail in Seinfeld and Pandis (1998) and Finlayson-Pitts and Pitts (2000).

We examine the performance of CMAQ by comparing the values of the air concentration of sulfate simulated by CMAQ with the estimates of sulfate provided by our statistical model based on the observational data. The statistical model allows us to account for fine-scale variability and provides estimates for sulfate levels in areas without monitors. The approach uses Bayesian methods to explore the spatial correlation structure inherent in the observational data, so that the estimates can be made accordingly. Another advantage of this particular statistical technique is its ability to quantify the uncertainty in both grid cell and point estimates. However, our method does not address aspects of temporal variability or assess correlation patterns over time.

There have been very few studies addressing CMAQs ability to simulate spatial trends in sulfate aerosol. One recent study by Mebust et al. (2003) evaluated CMAQs skill in simulating observed visibility indices and aerosol species concentrations. The speciated aerosol evaluation used observations of sulfate, nitrate, $PM_{2.5}$, PM_{10}, and organic carbon. These observations were obtained from 18 sites in the Interagency Monitoring of Protected Visual Environments (IMPROVE) network for June 1995. Results indicated that of all the constituents, sulfate was simulated best by CMAQ. The mean bias was $0.15 \mu g\, m^{-3}$, while the normalized mean bias (NMB) was 3.1%. The authors found that across time the NMB was within ±25% on all but 1 day. Across space the CMAQ simulation was within ±25% at 15 of the 18 monitors. The mean error was found to be considerably larger than the mean bias. The normalized mean error (NME) was found to be within 50% on all but 2 days and at all but two locations.

Jun and Stein (2004) propose techniques to assess the ability of CMAQ to capture spatiotemporal patterns in the variability of sulfate aerosol. In slightly older work, Haas (1998) provides a Monte Carlo hypothesis testing methodology which allows the consideration of a variety of hypotheses concerning spatiotemporal correlation structure, spatial or temporal drift, and the adequacy of particular statistical models. Spatiotemporal random field models for sulfate deposition are treated in detail in Vyas and Christakos (1997).

2. Data

The 2001 calendar year CMAQ simulation run was done on a 36 km horizontal grid using the Lambert conformal projection with 14 vertical layers based on a sigma coordinate system. The run used the CB-IV gas-phase chemical mechanism. The meteorological data came from MM5 (Grell et al., 1994), which was developed cooperatively by Penn State University and the National Center for Atmospheric Research. The MM5 output is processed by MCIP v2.3 (Meteorology-Chemistry Interface Processor) to generate inputs to the chemical transport model processor. Emissions data came from two sources: the 2001 National Emissions Inventory (NEI) for anthropogenic emissions and BEIS 3.12 for biogenic emissions (Houyoux, 2004). For more information about CMAQ and associated modeling systems mentioned here, see U.S. Environmental Protection Agency (1998, Section 2.3).

The observed data come from the monitoring sites in two U.S. EPA data collection networks: the Clean Air Status and Trends Network (CASTNet) and the Speciated Trends Network (STN). The CASTNet data are weekly aggregated samples, running from Tuesday morning to Tuesday morning. The STN data are collected every third or sixth day. Fig. 1 shows the locations of the monitoring sites and the bounds of the region used for the evaluation of CMAQ. The spatial region outlined in Fig. 1 was selected to encompass a portion of the major source region and the areas downwind from that region. This area also has a large number of monitors.

The observed and CMAQ simulated fields were averaged over 4-week periods, each of which constitutes a lunar month. These periods were chosen to compliment the CASTNet sampling schedule. This averaging procedure made it possible to compare the various data sources on the same temporal scale. Fig. 2 shows the average observed sulfate for selected sites by lunar month. In this research, lunar month one (2–30 January 2001) and

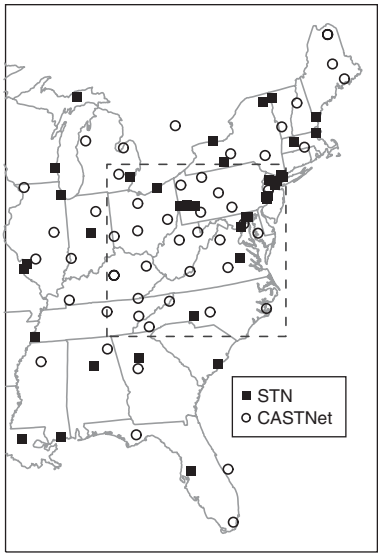

Fig. 1. Monitoring sites and the bounds of the study region.

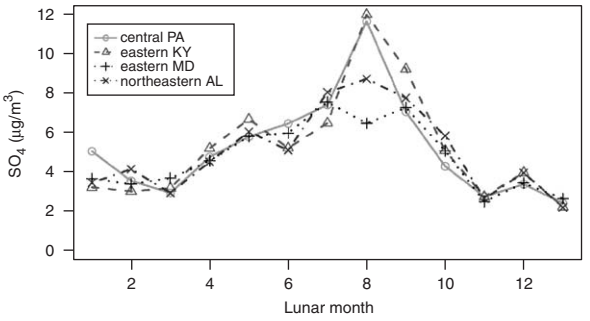

Fig. 2. Average observed sulfate by lunar month for selected sites.

lunar month eight (17 July 2001 to 14 August 2001) were used to assess CMAQ. These two periods provide examples from different seasons and from lower and higher sulfate levels.

3. Methodology

In comparing the CMAQ output with the monitoring data, we are faced with two major problems. One is the availability of only sparse observational data from limited monitoring networks to compare with an extensive structure of CMAQ grid cells. Since many common model evaluation methods (for example, Mebust et al., 2003) compare each monitoring value with the output given by CMAQ for the grid cell in which the monitor lies, the sparseness of the data means that the model's performance can only be ascertained for a small subset of grid cells. Our objective is to use statistical methods with the monitoring data to estimate the "true" concentration of sulfate aerosol at sites or grid cells for which we have no monitoring data.

A related concern is the practice of comparing an observation taken at a particular monitoring site with the average given by CMAQ for the grid cell in which the monitor lies (Mebust et al., 2003; Jun and Stein, 2004). Even if CMAQ is performing perfectly, we should not expect that an observation at a single point within a 36 km^2 region will be exactly equal to the grid cell average given by CMAQ. This issue is sometimes referred to as a "change of support" problem (Gelfand et al., 2001). Since we are focusing on sulfate aerosol, which has a smoothly varying distribution (Jun and Stein, 2004) across our study region, the support issue may not have a large impact on the statistical estimates obtained. However, the errors associated with estimates for grid cell averages vs. point-referenced locations differ somewhat even in the case of the relatively smooth distribution of sulfate aerosol; accounting for this allows us to obtain a more accurate estimate of variability.

The statistical methods we propose for model evaluation address both of these concerns. We employ a Bayesian hierarchical model, based on that used by Fuentes and Raftery (2005), to estimate the sulfate concentration field from the observational data. Emphasis is placed on understanding the spatial dependence of sulfate values across the region, and on using this dependence structure to estimate the values for specific locations or grid cells. Using the ideas of Fuentes and Raftery (2005) and Gelfand et al. (2001), we adjust for the change of support by adding estimates of the dependence between point observations and grid cells based on a suitable sample of points within each grid cell.

3.1. Statistical model description

Our analysis makes use of a Bayesian hierarchical model which takes into account parameters for trend estimation, spatial dependence, and small-scale error; it can be viewed as a type of Bayesian kriging technique. Similar models have been used by numerous Bayesian practioners; for example, Handcock and Stein (1993) focus on the subject of Bayesian kriging, and Banerjee et al. (2004, Section 5.1) provide a good explanation of Bayesian hierarchical spatial models. A key difference between Bayesian

and classical statistical inference is the Bayesian treatment of all unknown parameters as random variables. This facilitates the incorporation of previous scientific knowledge into the model through prior distributions on these parameters and allows us to better ascertain uncertainty associated with certain model components. After data are obtained, we update these prior distributions using this data and Bayes' Theorem. The resulting posterior distribution can then be used to give estimates about the values of the parameters and the uncertainty associated with them. References describing the general principles of Bayesian inference are numerous; these might include, but are not limited to, Lee (1997) and Gelman et al. (1995).

We begin by representing the observed data, denoted by the vector y, as a sum of the vector of "true" values z and fine-scale "error" or variability, ε. Assuming that these errors are independent and normally distributed with variance σ^2, we can write

$$y|z,\sigma^2 \sim \mathrm{N}(z,\sigma^2 I), \tag{1}$$

where I is the identity matrix. Our likelihood is then given by Eq. (1); this forms the top level in the model's hierarchical structure. It is worthwhile to note that this likelihood could be extended to account for additive or multiplicative biases in the observed data by adjusting the mean in Eq. (1); a similar approach was used by Fuentes and Raftery (2005).

In the next level, we define the prior distribution for the parameter z, which is the chief parameter of interest

$$z|\beta,\theta_1,\theta_2 \sim \mathrm{N}(X\beta,\Sigma). \tag{2}$$

The mean $X\beta$ serves to capture a trend component (linear in β) that exists between covariate information and the sulfate aerosol field. Such covariates might include the location coordinates of the monitors, elevation, or other information. Each of these covariates is stored in a column of the X matrix, so that the β vector contains a coefficient for each covariate.

The covariance matrix Σ is calculated based on a model which captures the spatial dependence among the sulfate observations. Based on our preliminary analyses, we assume that the correlation structure is stationary and isotropic. Many possible covariance models could be used for Σ, but probably the most common way to choose such a model is to select the one that most closely fits the empirical variogram. In our case, a good fit was provided by the exponential covariance structure, which is easy to implement and adequately captures the dependence in this and many similar situations. This covariance model depends only on two governing parameters (θ_1 and θ_2) and the distance between each pair of monitoring locations. For more details concerning variogram fitting, see Banerjee et al. (2004, Chapter 2.1.4) or Cressie (1993, Chapter 2.6). For more information about stationarity and isotropy, and a list of other common choices for covariance models, see Banerjee et al. (2004, Chapter 2) or Cressie (1993).

The ith row and jth column of the matrix Σ gives the covariance between the true sulfate concentration at point i and point j. Note that Σ must be a symmetric matrix, since $\mathrm{Cov}(z_i,z_j)=\mathrm{Cov}(z_j,z_i)$. Using an exponential covariance structure, we have

$$\Sigma_{ij}=\mathrm{Cov}(z_i,z_j)=\frac{1}{\theta_1}\exp\left\{-\frac{d_{ij}}{\theta_2}\right\}, \tag{3}$$

where d_{ij} is the distance between any two locations i and j. This covariance structure for z can also be expressed in semivariogram form

$$\gamma(d)=\frac{1}{\theta_1}\left(1-\exp\left\{-\frac{d}{\theta_2}\right\}\right), \tag{4}$$

where θ_1 may be referred to as the reciprocal of the sill and θ_2 as the range. The nugget parameter is not needed in this covariance model, because the variance associated with the fine-scale error is accounted for by the σ^2 in the likelihood. For further explanation of these parameters, see Banerjee et al. (2004, Section 2.1.3) or Cressie (1993, Section 3.2.1).

Our prior for z introduced three new parameters, β, θ_1, and θ_2, which introduce an additional level in the hierarchy. We also have not yet addressed the prior for σ^2, introduced in Eq. (1). The prior distributions for these parameters complete the hierarchy of the model. Our implementation makes use of the following prior distributions; however, choice of the prior can be changed to better reflect the state of knowledge available in a given situation

$$\sigma \sim \mathrm{U}(0.05\,\mu\mathrm{g\,m^{-3}},0.2\,\mu\mathrm{g\,m^{-3}}),$$
$$\theta_1 \sim \Gamma(1\,\mathrm{m^6\,\mu g^{-2}},0.005\,\mathrm{m^6\,\mu g^{-2}}),$$
$$\theta_2 \sim \mathrm{U}(50\,\mathrm{km},500\,\mathrm{km}), \quad \beta \sim \mathrm{N}(0,2500I).$$

Our prior for σ reflects the relatively small level of measurement error and fine-scale variability expected in the observations taken by the monitors. We have very little information about θ_1, so we use

a conjugate prior with large variance. Similarly, for θ_2 (the range parameter) we also allow a wide range of possible values. The elements of the vector β are assumed a priori to be independent, all with zero mean and large variance to reflect our uncertainty about these values. As is the case with regression models, the units for each element of β depend on the units of the covariate for which it is the coefficient.

3.2. Estimation using the model

Estimation is performed by applying Bayes' Theorem, a basic law of probability that dictates how probabilities should be updated on the basis of new information. Using this theorem, we can obtain the distribution of each parameter based on the data and the other parameters in the model. We summarize this procedure for the main parameter of interest, z.

Using Bayes' Theorem, we have

$$p(z|y,\sigma^2,\beta,\Sigma) \propto p(y|z,\sigma^2)p(z|\beta,\Sigma). \quad (5)$$

Note that we can make use of the "$\propto$" symbol here, because the denominator specified by Bayes' Theorem is not a function of z. Leaving out the multiplicative terms that do not depend on z, we can continue in this vein, noting that the expressions on the right side of the $\propto$ symbol require continuation onto additional lines:

$$\begin{aligned} p(z|y,\sigma^2,\beta,\Sigma) &\propto \exp\left\{-\frac{1}{2}(y-z)^{\mathrm{T}}\left(\frac{1}{\sigma^2}I\right)(y-z)\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}(z-X\beta)^{\mathrm{T}}\Sigma^{-1}(z-X\beta)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[-y^{\mathrm{T}}\left(\frac{1}{\sigma^2}I\right)z - z^{\mathrm{T}}\left(\frac{1}{\sigma^2}I\right)y \right.\right. \\ &\quad \left.\left. + z^{\mathrm{T}}\left(\frac{1}{\sigma^2}I\right)z\right]\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}[z^{\mathrm{T}}\Sigma^{-1}z - z^{\mathrm{T}}\Sigma^{-1}X\beta - (X\beta)^{\mathrm{T}}\Sigma^{-1}z]\right\}. \end{aligned}$$

Recall that, as in Eq. (1), I represents the identity matrix. After rearranging the terms, we recognize the kernel of this distribution for z as multivariate normal, so that

$$p(z|y,\sigma^2,\beta,\Sigma) = (2\pi)^{-n/2}|V|^{-1/2} \times \exp\left\{-\frac{1}{2}[z-Vm]^{\mathrm{T}}V^{-1}[z-Vm]\right\},$$

where

$$V = \left(\frac{1}{\sigma^2}I + \Sigma^{-1}\right)^{-1}, \quad m = \left(\frac{1}{\sigma^2}I\right)y + \Sigma^{-1}X\beta$$

and n is the length of the vector z. So, we can say that, given the other parameters and the data, z follows the distribution:

$$z|y,\sigma^2,\beta,\Sigma \sim \mathrm{N}(Vm, V). \quad (6)$$

Using the same method, we can obtain the conditional distributions (at least up to a constant factor) for each of the other parameters.

Once we have stepped through the above process for each of the parameters, we make use of Markov Chain Monte Carlo (MCMC) methods to sample from the joint posterior distribution of all the parameters. It is worth noting that these samples, since they come from the joint distribution, will reflect any correlations that exist between the parameters. Given our parameters and the priors we have chosen, our procedure makes use of both the Gibbs sampler, when feasible, and the Metropolis algorithm otherwise. For more information about MCMC methods and their implementation, references include Gelman et al. (1995), Gilks et al. (1998), and Gamerman (1997).

It is instructive to consider a bit more closely the parameters σ^2, β, θ_1, and θ_2. Though these parameters are not the focus of interest, their distributions will also be updated either directly based on the data through the likelihood (σ^2) or through their relationship with z. Unlike the classical kriging strategy, the parameters which govern the covariance structure are not estimated using a variogram and then assumed known for the remainder of the analysis. The "true" sulfate field, the trend, and the covariance structure are estimated at the same time, and the range of possible values is explored. This allows for a more thorough and accurate accounting of variability and error at the end of the estimation process.

3.3. Estimation of sulfate levels for unobserved locations and for grid cells

In addition to estimating the values of the parameters z, σ^2, β, θ_1, and θ_2, we would like to make statistical predictions. In other words, based on the values of the parameters as estimated using the observed monitoring data, we would like to predict the "true" concentration of sulfate over each grid cell or at unobserved point locations.

In the case of predictions of "true" sulfate concentrations to be made at point locations, the procedure is similar to that used by many other authors (e.g. Handcock and Stein, 1993; Gelfand et al., 2001), and it also shares some of the features of classical kriging. For the purposes of this section, we denote the concentrations at these prediction locations as z_2, and we denote the concentrations at the observed locations (for which we earlier calculated the full conditional in Eq. (6)) as z_1.

Samples from the predictive distribution, z_2, can be estimated conditionally on the estimated values z_1 at each step of the MCMC routine. Recalling Eq. (2), we remember that the concentration field is assumed multivariate normal, so that

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} X_1\beta \\ X_2\beta \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right). \tag{7}$$

This partitioning is based on standard multivariate normal theory (Anderson, 1984, Chapter 2.5). Again using properties of the multivariate normal, we can write the predictive distribution as

$$z_2|z_1, \beta, \Sigma \sim \mathrm{N}(X_2\beta + \Sigma_{21}\Sigma_{11}^{-1}(z_1 - X_1\beta), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}). \tag{8}$$

Here, X_2 represents the values of the covariates for the prediction locations, and Σ_{22} represents the covariance among the prediction locations, based on the values θ_1, θ_2, and the distances among the locations, as given in Eq. (3). Lastly, $\Sigma_{21} = \Sigma_{12}^{\mathrm{T}}$ gives the covariance between the prediction locations and the observation locations, also based on Eq. (3). Samples are taken from this predictive distribution, based on the posterior joint distribution of the other parameters. Further inspection shows that the mean of the distribution given in Eq. (8) is similar to the predictor obtained using simple kriging. However, this Bayesian method takes into account the variability in the predictive distribution and the distributions of the other parameters.

To predict the average sulfate concentration for a grid cell volume, we need to adjust our method, with particular focus on the covariance structure. Now, Σ_{12} (and its transpose, Σ_{21}) represent covariances between cell averages and pointwise values, while Σ_{22} represents covariances among the cells. As explained by Gelfand et al. (2001) and Fuentes and Raftery (2005), for each grid cell, we take a sample of points from within it. In the case of the covariance between a monitoring site and a cell, we then calculate the covariances between each sample point in the cell and the monitoring site. We then average these covariances to obtain an estimate for the covariance between the grid cell and the monitoring location. For two grid cells, we calculate and then average the covariances for all pairings of the sample points between the two cells. The predictive distribution still has the form given in Eq. (8); it is only our method for calculating the partitions of Σ that must change.

4. Results

As discussed in Section 2, we apply this method to sulfate concentrations averaged over a 4-week period in winter and a 4-week period in summer. Fig. 3 shows the averaged monitoring data in our region for the winter period. This figure, as well as all those that follow, remain in the original units of $\mu g\,m^{-3}$. Further examination of the figure shows that for the most part, the sulfate concentrations seem to be higher as we move in a more northerly direction. It also appears that observations taken in the Appalachians tend to be a bit lower. As a result, we include elevation and the y-coordinate of the locations as covariates in the model.

When applied to this data, our statistical model yielded the estimates shown in Fig. 4. It should be noted that the values shown are the estimated posterior mean values for the predictive distribution of the sulfate concentrations for the grid cells. These grid cells have the same boundaries as those used by CMAQ, so we can directly compare our model's estimate of the "true" concentrations with those simulated by CMAQ, which can be found in Fig. 5. Fig. 6 displays the differences between the CMAQ output and estimates made by our Bayesian model.

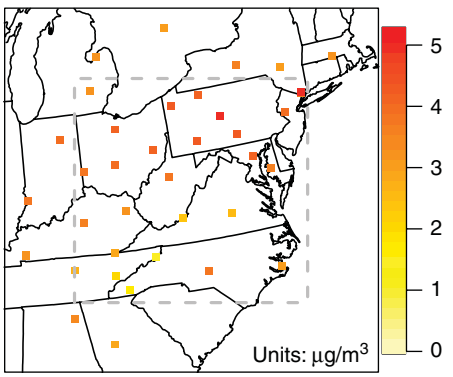

Fig. 3. Sulfate monitoring data (winter).

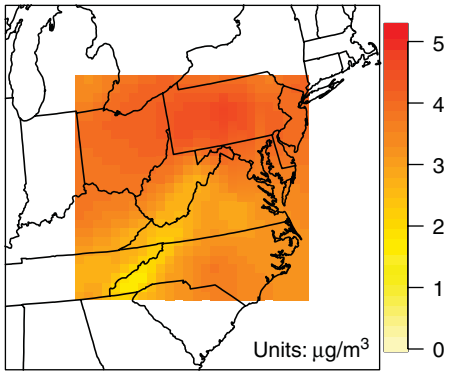

Fig. 4. Statistical estimates (winter).

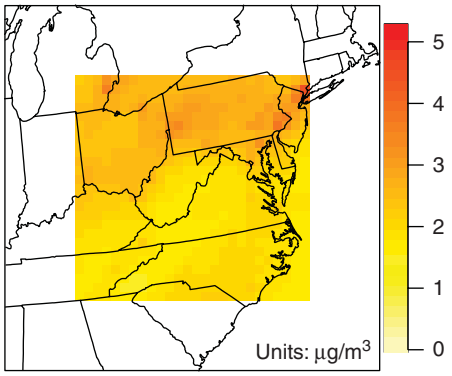

Fig. 5. CMAQ output (winter).

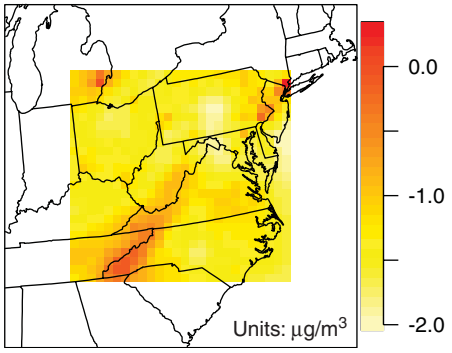

Fig. 6. CMAQ output—statistical estimates (winter).

This last figure indicates that CMAQ is tending to underpredict the sulfate concentration over the majority of the region in winter. Exceptions can be found in much of the Appalachian mountains, where the discrepancies between CMAQ output and the statistical estimates based on the observations are smaller.

For the summer period of 17 July 2001 through 13 August 2001, the CMAQ model performs quite well. Sulfate concentrations observed at the monitors and averaged over this period are displayed in Fig. 7. The presence of substantially more monitors is noticeable when comparing this figure with the observed concentrations during the winter period in Fig. 3. This is due largely to the fact that a larger number of STN monitors were operational by the second half of 2001. As with the winter time period, we see a general tendency for sulfate concentrations to be higher in the more northerly areas of the region, so the y-coordinate of each location is again included as a covariate. There also appears to be some reason to include the x-coordinate as a covariate, since values along the coast are a bit lower than values further inland. Interestingly, we do not observe notable differences in the Appalachian region in this summer time span. Runs of the model were performed with and without elevation as a covariate, with little discernible difference in the estimates. The results that follow correspond to the estimated model without elevation as a covariate.

The mean predicted sulfate concentrations as estimated by our statistical model are shown in Fig. 8. These statistical estimates can be compared with the CMAQ output averaged over the same summer period in Fig. 9. The CMAQ output and the estimates are similar enough that it is hard to see noticeable differences at a first glance. It should be noted that the differences in summer are larger than those in winter, in absolute terms, but still small relative to the range of sulfate values observed in the summertime. A glance at Fig. 10 also reveals that these differences do not indicate a more systematic underestimate of sulfate concentrations in the region, as we saw in the winter (compare with Fig. 6).

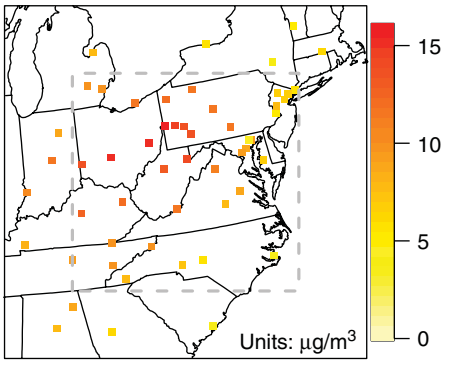

Fig. 7. Sulfate monitoring data (summer).

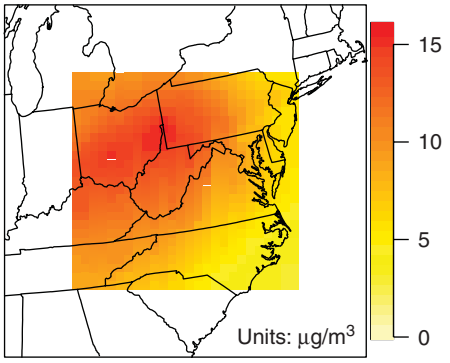

Fig. 8. Statistical estimates (summer).

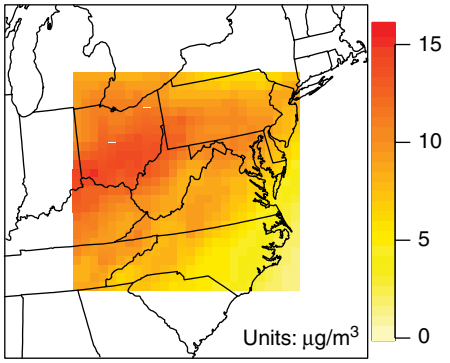

Fig. 9. CMAQ output (summer).

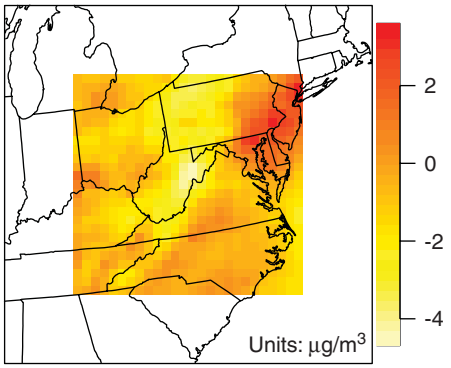

Fig. 10. CMAQ output—statistical estimates (summer).

Since there is error inherent in all statistical estimates, we may ask whether the differences between CMAQ output and the estimates are significant, or whether they may be attributed solely to error. One advantage of the Bayesian model we have implemented is the ease with which this error can be estimated by using the predictive distribution in Eq. (8). Figs. 11 and 12 display the standard

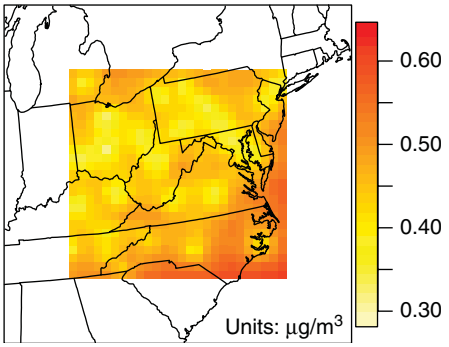

Fig. 11. Standard deviation of predictive distribution (winter).

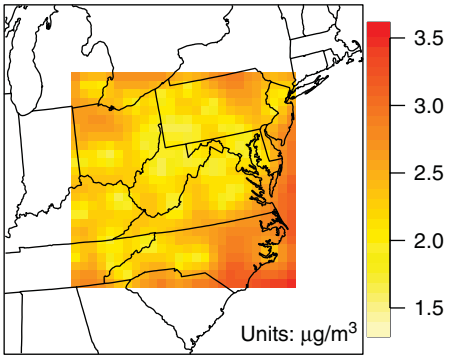

Fig. 12. Standard deviation of predictive distribution (summer).

deviation of the predictive distribution for each grid cell for the winter and summer time periods, respectively. As we would expect, the variability of the predictive distribution is lowest for grid cells which are near monitoring sites, and highest for grid cells which are far from monitors. Also, as we observed earlier, the standard deviations tend to be higher in the summer time period, since sulfate values take a larger range during this time.

We also use the predictive distribution to find the 95% Bayesian credible interval, which gives the interval within which 95% of the density of the predictive distribution falls. In particular, we are interested in grid cells in which the CMAQ simulated value falls outside of the 95% credible interval, because the statistical model can be said to strongly indicate an error in CMAQ for these locations. Figs. 13 and 14 show, for the winter and summer periods, the differences between CMAQ output and our mean statistical estimate only for grid cells at which the CMAQ output falls outside the 95% credible interval.

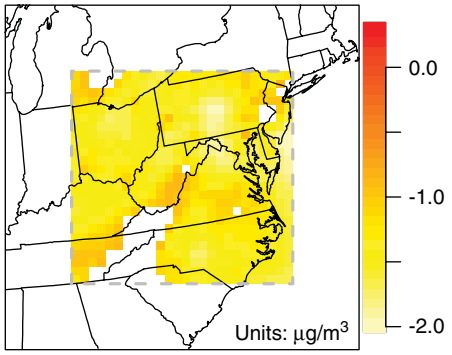

Fig. 13. Significant differences (winter).

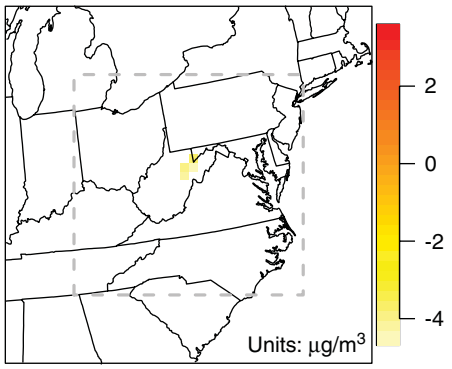

Fig. 14. Significant differences (summer).

Fig. 13 shows that in the winter period, the CMAQ simulated sulfate values differ substantially from our statistical estimates based on monitoring data. This is true for most of the region under consideration, with the exception of the mountainous areas of North Carolina and southwestern Virginia. This is consistent with our initial comparisons of Figs. 3 and 5, in which CMAQ appeared to be underestimating sulfate levels during this January period throughout most of the region. In the summer period, we have only a very small area which the statistical model indicates as a possible area for further investigation. This consists of a small area in northeastern West Virginia, where the CMAQ model is slightly underestimating sulfate levels.

5. Discussion

There are notable discrepancies between the CMAQ simulated sulfate concentrations and the estimates made based on the observational data in the January time period, but very few such discrepancies in the summer time frame. The good performance of CMAQ during the latter period is consistent with the findings of others, such as Mebust et al. (2003), who also confirmed CMAQs ability to simulate sulfate in a summer time period (June 1995). However, the systematic underestimation of sulfate during the winter time period in this region is more perplexing. It is unlikely that emissions are the cause, since CMAQ incorporates the continuous emissions monitoring (CEM) data into the model runs. These data account for the majority of sulfur dioxide emitted into the atmosphere. It is more likely that the cause lies within the aqueous chemical mechanisms used in the model, and with the model simulation of precipitation. An examination of these causes is beyond the scope of this work. The papers by Dennis (1991) and Dennis et al. (1993) should be consulted. To investigate the winter underestimation problem further, one strategy is to use a different year for another retrospective study. In 2002, for example, more STN monitors were available; additional observational data would tend to improve the statistical estimates as well.

The Bayesian method utilized in this work has two main advantages: (1) its use of the sparse observational data to estimate sulfate levels over the entire region and (2) its ability to account for the differences between point-referenced and grid cell averages. By using the available monitoring data to estimate the parameters governing the covariance structure, the microscale error, and the true sulfate levels, we are able to make statistically justifiable estimates of the sulfate level for both unobserved sites and for CMAQ grid cell averages. In addition to making sulfate estimates, Bayesian analysis enables us to obtain information about the posterior distributions for each of the parameters, allowing us to better quantify the uncertainty associated with those estimates and to identify areas in which differences between estimates based on monitoring data and CMAQ output are significantly different.

Another advantage of our statistical model is its flexibility. For instance, our situation allowed use of a simple isotropic and stationary covariance model (the exponential). In other situations, in which this assumption may not be appropriate, we could substitute other, less simple covariance models, which allow for anisotropy or nonstationarity.

Compared to classical kriging, our method is much more computationally intense. In the

traditional approach, the parameters of the covariance structure are estimated at the beginning of the analysis (typically through a variogram), and treated as if known for the remainder of the analysis. Using MCMC, the Bayesian approach samples these parameters at the same time as the sulfate values are estimated. This requires multiple calculations of the covariance matrix Σ, as θ_1 and θ_2 are repeatedly sampled. When there is a large amount of monitoring data or a large number of grid cells at which to make predictions, the calculations involved can progress slowly, owing largely to the need to invert the covariance matrix Σ, or its partitions. (See Eqs. (6) and (8).) Together with the stationarity assumption (as mentioned above), the computational demands limit the size of the region over which CMAQ can be evaluated using this method.

Acknowledgments

The authors express their sincere appreciation to Robin Dennis, Brian Eder, Peter Finkelstein, Alice Gilliland, and Shawn Roselle for many fruitful discussions. They also gratefully acknowledge constructive comments from several reviewers.

Disclaimer

The research presented here was performed under the Memorandum of Understanding between the U.S. Environmental Protection Agency (EPA) and the U.S. Department of Commerce's National Oceanic and Atmospheric Administration (NOAA) and under agreement number DW13921548. This work constitutes a contribution to the NOAA Air Quality Program. Although it has been reviewed by EPA and NOAA and approved for publication, it does not necessarily reflect their policies or views.

References

Anderson, T.W., 1984. An Introduction to Multivariate Statistical Analysis, second ed. Wiley, New York.

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2004. Hierarchical Modelling and Analysis for Spatial Data. Chapman & Hall, Boca Raton.

Byun, D.W., Ching, J.K.S., 1999. Science algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) modeling system. Report EPA-600/R-99/030, U.S. Environmental Protection Agency, U.S. Government Printing Office, Washington, DC.

Cressie, N.A.C., 1993. Statistics for Spatial Data, revised ed. Wiley, New York.

Dennis, R.L., 1991. Bounding: estimating upper and lower limits of deposition reduction. In: Irving, P.M. (Ed.), Acidic Deposition: State of Science and Technology, vol. I: Emissions, Atmospheric Processes, and Deposition. U.S. National Acid Precipitation Assessment Program, Washington, DC.

Dennis, R.L., McHenry, J.N., Barchet, W.R., Binkowski, F.S., Byun, D.W., 1993. Correcting RADM's sulfate underprediction: discovery and correction of model errors and testing the corrections through comparisons against field data. Atmospheric Environment 27A, 975–997.

Dominici, F., Samet, J.M., Zeger, S.L., 2000. Combining evidence on air pollution and daily mortality from the twenty largest U.S. cities: a hierarchical modeling strategy (with discussion). Journal of the Royal Statistical Society, Series A 163, 263–302.

Finlayson-Pitts, B.J., Pitts, J.N., 2000. Chemistry of the Upper and Lower Atmosphere: Theory, Experiments, and Applications. Academic Press, San Diego.

Fuentes, M., Raftery, A.E., 2005. Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. Biometrics 61, 37–46.

Gamerman, D., 1997. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. Chapman & Hall, London.

Gelfand, A.E., Zhu, L., Carlin, B.P., 2001. On the change of support problem for spatio-temporal data. Biostatistics 2, 31–45.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. Bayesian Data Analysis. Chapman & Hall, London.

Gilks, W.R., Spiegelhalter, D.J., Richardson, S., 1998. Markov Chain Monte Carlo in Practice. Chapman & Hall, Boca Raton.

Grell, G.A., Dudhia, J., Stauffer, D.R., 1994. A description of the fifth generation Penn State/NCAR mesoscale model (MM5), NCAR Technical Note NCAR/TN-398+STR, National Center for Atmospheric Research, Boulder, CO.

Haas, T.C., 1998. Statistical assessment of spatio-temporal pollutant trends and meteorological transport models. Atmospheric Environment 32, 1865–1879.

Handcock, M.S., Stein, M.L., 1993. A Bayesian analysis of kriging. Technometrics 35, 403–410.

Houyoux, M., 2004. CAIR Emissions Inventory Review. U.S. Environmental Protection Agency. Office of Air Quality Planning and Standards, Emissions Monitoring and Analysis Division, Research Triangle Park, NC 27711.

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

Lee, P.M., 1997. Bayesian Statistics: An Introduction, second ed. Wiley, New York.

Mebust, M., Eder, B.K., Binkowski, F.S., Roselle, S.J., 2003. Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component, 2, model evaluation. Journal of Geophysical Research 108 (D6), 4184.

Seinfeld, J.H., Pandis, S.N., 1998. Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. Wiley, New York.

Smith, R.L., Davis, J.M., Sacks, J., Speckman, P., Styer, P., 2000. Regression models for air pollution and daily mortality: analysis of data from Birmingham, Alabama. Environmetrics 11, 719–743.

Styer, P., McMillan, N., Gao, F., Davis, J., Sacks, J., 1995. Effect of outdoor airborne particulate matter on daily death counts. Environmental Health Perspectives 103, 490–497.

U.S. Environmental Protection Agency, 1995. Air Quality Criteria for Particulate Matter, Report EPA/600/P-95/001aF, Research Triangle Park, NC.

U.S. Environmental Protection Agency, 1998. EPA Third-Generation Air Quality Modeling System, Models-3 Volume 9b User Manual, Report EPA-600/R-98/069(b), Research Triangle Park, NC.

Vyas, V.M., Christakos, G., 1997. Spatiotemporal analysis and mapping of sulfate deposition data over eastern U.S.A. Atmospheric Environment 31, 3623–3633.