ELSEVIER

Atmospheric Environment 38 (2004) 4427–4436

ATMOSPHERIC ENVIRONMENT

www.elsevier.com/locate/atmosenv

Statistical comparison of observed and CMAQ modeled daily sulfate levels

Mikyoung Jun*, Michael L. Stein

Department of Statistics, University of Chicago, Chicago, IL 60637, USA

Received 29 October 2003; received in revised form 20 May 2004; accepted 26 May 2004

Abstract

New statistical procedures to evaluate the Models-3/Community Multiscale Air Quality (CMAQ) using observed data are introduced. Certain space–time correlations are used to assess dynamic aspects of CMAQ and to compare the space–time structure of CMAQ to that of observations. Our analyses show that, overall, CMAQ matches the space–time correlation structure of observed sulfate concentrations well. Analyses on the space–time correlation of the difference of observations and CMAQ output show that CMAQ partially captures time lagged spatial variation of sulfate concentrations. Separable covariance functions are shown to provide a poor description of the observations.
© 2004 Elsevier Ltd. All rights reserved.

Keywords: Air quality model; Model evaluation; Space–time process; Separable covariance function

1. Introduction

The Models-3/Community Multiscale Air Quality (CMAQ) modeling system has been developed by the United States Environmental Protection Agency (US EPA) to understand and evaluate air quality by producing maps of multiple pollutants' distributions over the United States (and some parts of Canada). CMAQ takes terrain, land use, emissions, meteorology and other information as its inputs and produces various pollutants' ambient concentrations or wet/dry depositions. Its output is in the form of block averages in multiple layers; the resolution can be set by the user to a certain extent. There have been many improvements made to the science and mechanisms of CMAQ since its first release to the public in June 1998. For example, the mechanisms for aerosol and cloud chemistry have been updated constantly and the Plume-In-Grid model has been added to the system to generate more accurate flow of emissions from point sources. A version of CMAQ that can be run in parallel across multiple processors is now available. More detailed information about CMAQ and its development can be found in Dennis et al. (1996), Models-3 website (http://www.epa.gov/asmdnerl/models3/) or Community Modeling and Analysis System (CMAS) website (http://www.cmascenter.org).

To evaluate the accuracy of CMAQ, it is essential to compare the output with observations from monitoring stations. Eder et al. (2002) and Mebust et al. (2003) compute bias and error statistics of CMAQ output relative to observations, plot observations against CMAQ output and fit linear regression lines. These simple comparisons are valuable but are not effective for understanding many aspects of the spatial and temporal patterns in the model errors. Fuentes and Raftery (2001) model the "true" process of air pollutants, the observations and CMAQ output process as jointly Gaussian random fields. They estimate the parameters for the bias of the CMAQ output, the parameters for the covariance structure of CMAQ and observation error processes and simulate the conditional distribution of the "true"

*Corresponding author. Tel.: +1-773-702-0959; fax: +1-773-702-9810.

E-mail addresses: jun@galton.uchicago.edu (M. Jun), stein@galton.uchicago.edu (M.L. Stein).

1352-2310/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2004.05.019

process given CMAQ output and the observations. This approach is attractive in principle but requires a number of strong assumptions about the statistical characteristics of the processes. Meiring et al. (1998) compares an air quality model with observations by temporal prewhitening and spatial deformation. Haas (1998) evaluates meteorological pollutant transport and deposition models through a Monte Carlo hypothesis test.

In this paper, we suggest new ways of comparing CMAQ output to observations that assess CMAQ's ability to capture dynamic aspects of the process accurately. Sampson and Guttorp (1999) point out the necessity of comparing the space–time correlation structure of observations and numerical model output in the evaluation of numerical models, but there has not been much work done in this regard. The methodology used in Fuentes and Raftery (2001) only applies to spatial processes at a fixed time point, so evaluation of the dynamic aspect of CMAQ is not possible. Meiring et al. (1998) and Haas (1998) do consider air quality models in a space–time context, but neither is intended to evaluate the dynamic behavior nor the space–time correlation structure of the physical model output.

One important aspect of space–time correlations of atmospheric processes on the time-scale of days is that they are likely to be asymmetric in space–time in the sense that the correlation between site A today and site B tomorrow will often be different than the correlation between site B today and site A tomorrow. For example, Gneiting (2002) finds such an asymmetry in daily wind speeds in Ireland. We suggest ways of assessing the performance of CMAQ with regard to its dynamic behavior by looking at correlations over different spatial locations at different time points. We also suggest examining correlations of various linear combinations of observations or CMAQ output to understand the structure of space–time correlation of each process. Meiring et al. (1998) and Haas (1998) require developing a complete spatial or spatial–temporal stochastic model for the process under study, which we do not require. Our methods give researchers new tools to assess the fits of models such as those described in Meiring et al. (1998) and Haas (1998), and thus provide an important complement to these more model-based approaches.

2. Data

We have daily averages from the Eulerian Model Evaluation Field Study (EMEFS) and CMAQ output of the ambient concentration of aerosol sulfate in $\mu g\ m^{-3}$ over the eastern United States and parts of Canada from March 1990 through May 1990. These data were provided by the US EPA. For a detailed description of EMEFS, see McNaughton and Vet (1996). The observations and CMAQ output used for the analysis in this paper are available at http://galton.uchicago.edu/~stein/corr/. There is a total of 61 monitoring stations over the whole region. Their locations are given in Fig. 1. This period is especially useful for this study because of intensive monitoring undertaken during EMEFS. For example, Fig. 1 shows two clusters of monitoring stations that only operated during this period.

Since CMAQ output is averaged over large square blocks and observations are effectively averages over much smaller regions, CMAQ output and observations have a mismatch of support. We ignore this problem here and act as if the CMAQ output is point-valued. We believe this simplification is reasonable because it is known that the concentration of sulfate generally varies fairly smoothly in space compared to that of SO_2 or NO_x The reason for this is that the conversion rates of SO_2 (gas and aqueous phase combined) to sulfate gives SO_2 a lifetime of 2–4 days (Warneck, 1999), so that SO_2 has sufficient time to mix well regionally and thus produces relatively smooth sulfate aerosol distributions in space and time. We can give a quantitative assessment of the greater smoothness of sulfate aerosol through the spatial variogram (definition and explicit expression are given in Appendix A) using CMAQ output. The spatial variogram values of sulfate aerosol and SO_2 concentration at spatial lag 36 km divided by the square of the respective mean for each pollutant (to make the quantities unit-free) are 0.027 and 5.249, respectively. The values at spatial lag 51 km are 0.051 and 6.698 and the values at spatial lag 72 km are 0.090 and 7.094, respectively. This implies that SO_2 has much higher (by about two orders of magnitude) relative spatial variation at spatial lags under 100 km

About 7.4 of the sulfate observations during March–May 1990 are missing. We decided to drop four stations (station numbers 29, 37, 47 and 49) where more than 20 days are missing and the missing values happen over long continuous time periods. This drops the percentage of missing data to 5.0 There are of course no missing values in the CMAQ output.

All of our analyses are on the logarithmic scale, which has the advantage of making the variances of the observations and CMAQ output more nearly constant across space and their distributions more nearly Gaussian. In addition to the log transformed observations and CMAQ output processes, we examine the difference between the log transformed observations and the log transformed CMAQ output. Hereafter, we will just say "observations" and "CMAQ output" for the base 10 logarithms of these quantities and "CMAQ error" for the difference of these logarithms. Since we compare the CMAQ output to the observations, we only use CMAQ output for the locations in Fig. 1 even though CMAQ output covers the whole domain. Thus, stations located in the same cell of CMAQ output are matched to the same CMAQ output values.

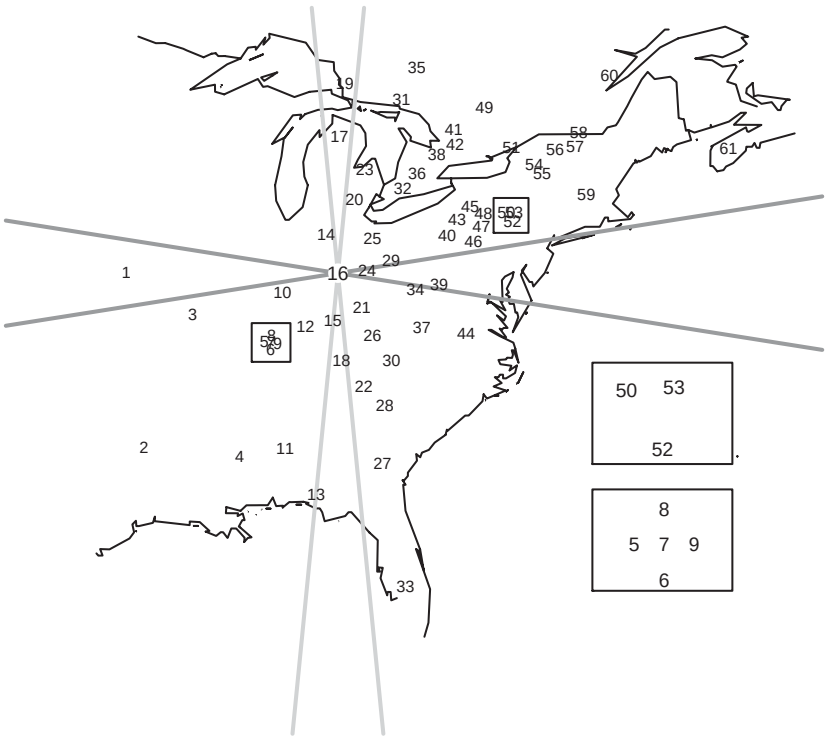

Fig. 1. Location of 61 monitoring stations over the eastern US and Canada. There are two clusters of stations (inside of squares) and these two clusters are magnified in the two subplots on the right. Stations are numbered 1-61 from west to east. Lines from site 16 show the bands used in Fig. 4 for that site.

The focus of this work is on the space–time dynamics of air pollution. Thus, we are not interested in the fact that some locations will have consistently higher sulfate levels than others or in seasonal variations. Fig. 2 shows that CMAQ does a good job at capturing consistently higher or lower pollution levels at the various monitoring sites. Sites such as 19, 35, 60 and 61 have lower values compared to other sites and CMAQ captures this pattern quite well. There is not much seasonality in our 3 month period, although observations and CMAQ output show a somewhat similar temporal pattern (not shown). To remove these purely spatial and purely temporal effects from each process under study, for an observation at location $\mathbf{s}$ and time t we subtract the spatial average (the average across time at each site) and temporal average (the average across space at each time), see Appendix B for details. These spatial and temporal averages would also be of interest, for example, to find evidence for biases in CMAQ, but are less useful and are, we contend, a distraction when studying dynamics. One way that we can think about spatial–temporal processes is to write them as

$$Z(\mathbf{s},t) = m(\mathbf{s}) + \mathrm{m}(t) + e(\mathbf{s},t) \qquad (1)$$

that is, as an additive effect in space, $m(\mathbf{s})$ an additive effect in time, $\mathrm{m}(t)$ and a space–time interaction, $e(\mathbf{s},t)$ We believe that comparing the interactions, $e(\mathbf{s},t)$ for CMAQ output and observations is a better measure of CMAQ's ability to capture dynamics than looking directly at Z For example, systematic errors in emissions, which are known to be a serious problem with CMAQ, are more likely to affect $m(\mathbf{s})$ than $e(\mathbf{s},t)$ In computing the spatial and temporal averages, we imputed values for the missing observations to reduce biases due to spatial and temporal patterns in the missing observations. The details of the imputation procedure are given in Appendix B. After the removal of spatial and temporal averages, the imputed values are no longer used and the analyses are done using only sites and times at which observations are available. We tried various other approaches to handling the missing observations in the computation of spatial and temporal averages, including just using the available observations, and the results were almost unchanged. However, the approach to dealing with missing data may be more critical when the fraction of missing observations is higher.

3. Spatial variations

Fig. 3(a) shows the spatial variogram (Appendix A) for the observations, CMAQ output and the CMAQ

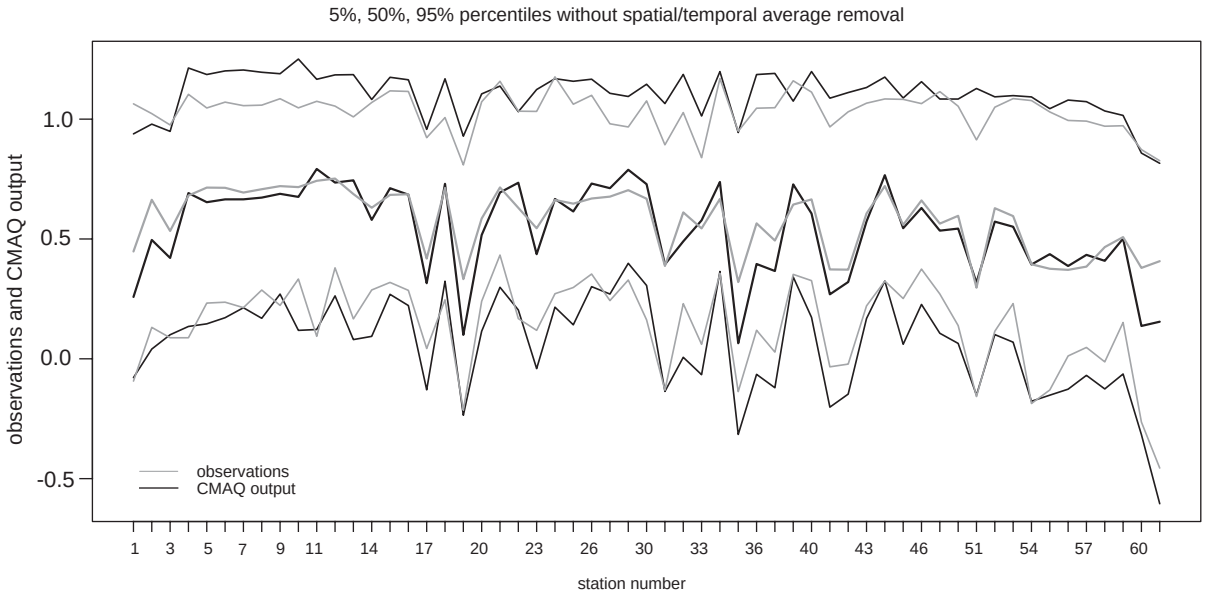

Fig. 2. Comparison of 5%, 50% and 95% percentiles of the temporal distribution of observations and CMAQ output without removal of spatial and temporal averages.

error averaged over time. While the variogram for the observations and CMAQ output are roughly linearly increasing in distance, the variogram for the CMAQ error is relatively constant beyond around 600 km This implies that there is almost no dependence between values of CMAQ error farther than 600 km apart, which suggests that the process of CMAQ error is statistically simpler than the observations or CMAQ output.

Fig. 3(a) also shows that the variogram of the CMAQ error is larger than the other two processes at spatial lags shorter than 700 km or so and smaller at spatial lags longer than 700 km This indicates that CMAQ does better at explaining large-scale variations in observed levels of sulfate than it does local variations. If CMAQ output and observations were independent, the variogram of the CMAQ error would be the sum of the variograms of CMAQ output and observation. Thus, the ratio, (variogram of the CMAQ error)/(variogram of CMAQ output variogram of observations), is a sensible measure of how much of the spatial variation of the observations CMAQ captures. Values of the ratio o 1 indicate a positive correlation between observations and CMAQ output, with smaller values indicating stronger dependence. This ratio is plotted against the distance lags in Fig. 3(b), which shows that the ratio (after smoothing) is o 1 at all distances and is decreasing as distance increases, so that CMAQ captures an increasing fraction of the spatial variation as distance increases. Therefore, there is some dependence even at short lags, but much stronger dependence at larger lags. Moreover, since we have already subtracted off the spatial and temporal averages, we can say that CMAQ captures not only consistently higher or lower pollution levels in some regions but also more complicated spatial–temporal patterns of the observations.

For distance lags up to about 200 km Fig. 3(a) shows that observations have higher variogram values than CMAQ output, which may be attributable to block-averaging of CMAQ output. On the other hand, for lags longer than 200 km or so, CMAQ output has higher variogram values than the observations. This implies CMAQ has more large-scale spatial variation than the observations. Similarly, the temporal variances of CMAQ output are higher than the temporal variances of the observations at most monitoring sites (results not shown). This larger variation is perhaps a sign of a problem with CMAQ, since if the model were correct, one would generally expect the lack of measurement error and the spatial and temporal averaging of the model output to produce smaller variations than in the data.

One approach to simplifying the modeling of the joint relationship between spatially varying quantities is to assume separability (Sun et al., 1998); that is, the covariance function for two spatial processes factors into a function depending only on space and another depending only on the indices of the processes. Fig. 3(b) provides a diagnostic for assessing this assumption. Specifically, Appendix C shows that separability implies that the true value of the ratio plotted in Fig. 3(b) is independent of distance. The clear decrease of the estimate of this ratio with distance strongly suggests

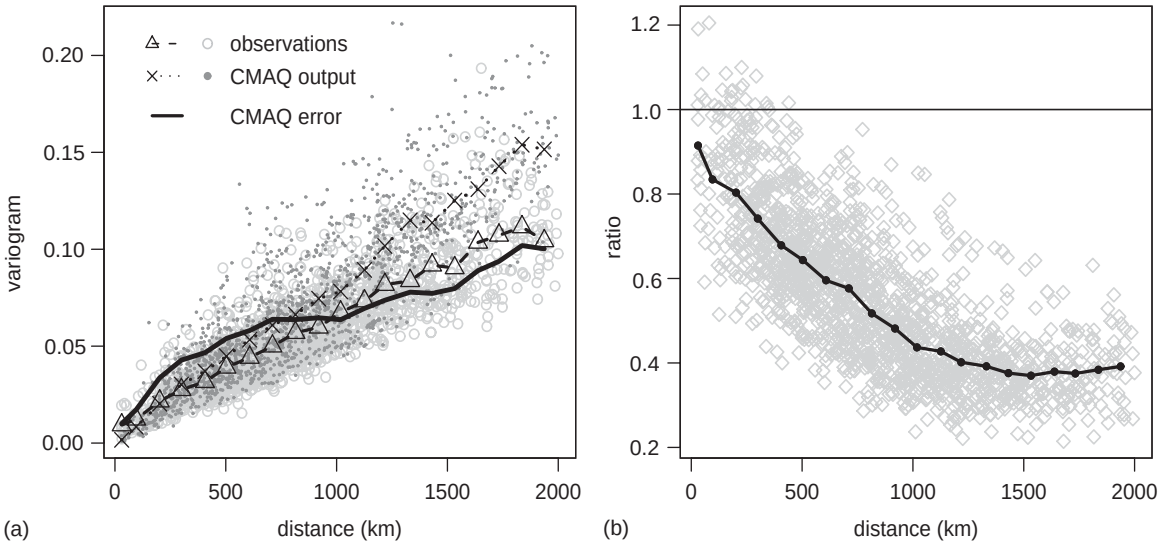

Fig. 3. (a) Spatial variograms for the observations, CMAQ output and the CMAQ error. After grouping the variogram values into bins based on distance, the average distance vs. the average correlation inside of each bin is represented by the dark symbols and connecting lines are drawn between the symbols. (b) The ratio of variogram of the CMAQ error and the sum of variograms of the observations and the CMAQ output. The line represents the ratio calculated using the binned averages of each variogram.

that separability is not a tenable assumption for the joint distribution of observations and CMAQ output.

4. Space–time structure

CMAQ is intended to give an accurate representation of the dynamics of air pollution. Therefore, it needs to track the flow of air pollutants with the prevailing winds. We suggest that simple space–time correlations can be used to assess how well CMAQ captures the dynamics of air pollutant processes. There are two different types of questions we can ask. First, we can ask whether CMAQ explains the spatial–temporal variation in the observations. This question can be answered by comparing the spatial–temporal variations in the observations to those in the CMAQ error. Second, we can ask if CMAQ produces sulfate values whose spatial–temporal variations mimic the kinds of spatial–temporal variations shown in the observations. This question can be examined by comparing measures of spatial–temporal variations in the observations to similar measures in the CMAQ output. If CMAQ does well with the first question, it will do well with the second, but the converse need not hold.

We will assume throughout this section that the processes we are considering are statistically homogeneous in time over the period of study. Let $Z_a(\mathbf{s}, t)$ be the process of interest with spatial and temporal averages removed (see Appendix B). Define $D(\mathbf{s}_1, \mathbf{s}_2, j) = \mathrm{Cor}(Z_a(\mathbf{s}_1, t), Z_a(\mathbf{s}_2, t+j))$ where Cor indicates correlation and j an integer is the time lag in days. When Z_a is the observations or the CMAQ output and there are predominant winds blowing from $\mathbf{s}_1$ to $\mathbf{s}_2$ for j positive, we expect $D(\mathbf{s}_1, \mathbf{s}_2, j)$ positive and bigger than $D(\mathbf{s}_1, \mathbf{s}_2, -j)$. By examining these correlations for different spatial locations and values of j for each process, we can compare the dynamics of the observation process to that of CMAQ output. The empirical value, $\hat{D}$ for D is calculated by

$$\hat{D}(\mathbf{s}_1, \mathbf{s}_2, j) = \frac{N_1^{-1} \sum_{i=1}^{T-j} Z_a(\mathbf{s}_1, i) Z_a(\mathbf{s}_2, i+j)}{N_2^{-1} \sum_{i=1}^{T-j} Z_a(\mathbf{s}_1, i)^2 \, N_3^{-1} \sum_{i=1}^{T-j} Z_a(\mathbf{s}_2, i)^2} \qquad (2)$$

where any summand that includes a missing observation is set to 0. The constants N_1, N_2, and N_3 are the actual number of available terms in the corresponding sums. For example, N_1 is the number of values of i in the range $1 \le i \le T-j$ for which $Z_a(\mathbf{s}_1, i)$ and $Z_a(\mathbf{s}_2, i+j)$ are both available.

Fig. 4 compares the empirical space–time correlation functions $\hat{D}$ for the observations and CMAQ error at different time lags with respect to the two directions, E-W and N-S. The average, $[\hat{D}(\mathbf{s}_1, \mathbf{s}_2, j) + \hat{D}(\mathbf{s}_1, \mathbf{s}_2, -j)]/2$ along with the difference, $\hat{D}(\mathbf{s}_1, \mathbf{s}_2, j) - \hat{D}(\mathbf{s}_1, \mathbf{s}_2, -j)$ is plotted against the difference of latitude and longitude (in deg) of $\mathbf{s}_1$ and $\mathbf{s}_2$. The pairs of spatial locations are selected so that the difference of longitude or latitude of the two locations has an angular restriction. For example, for the plot with respect to the latitude difference, we restrict $|\mathrm{lon}(1) - \mathrm{lon}(2)| \le \frac{1}{8}|\mathrm{lat}(1) - \mathrm{lat}(2)|$ to

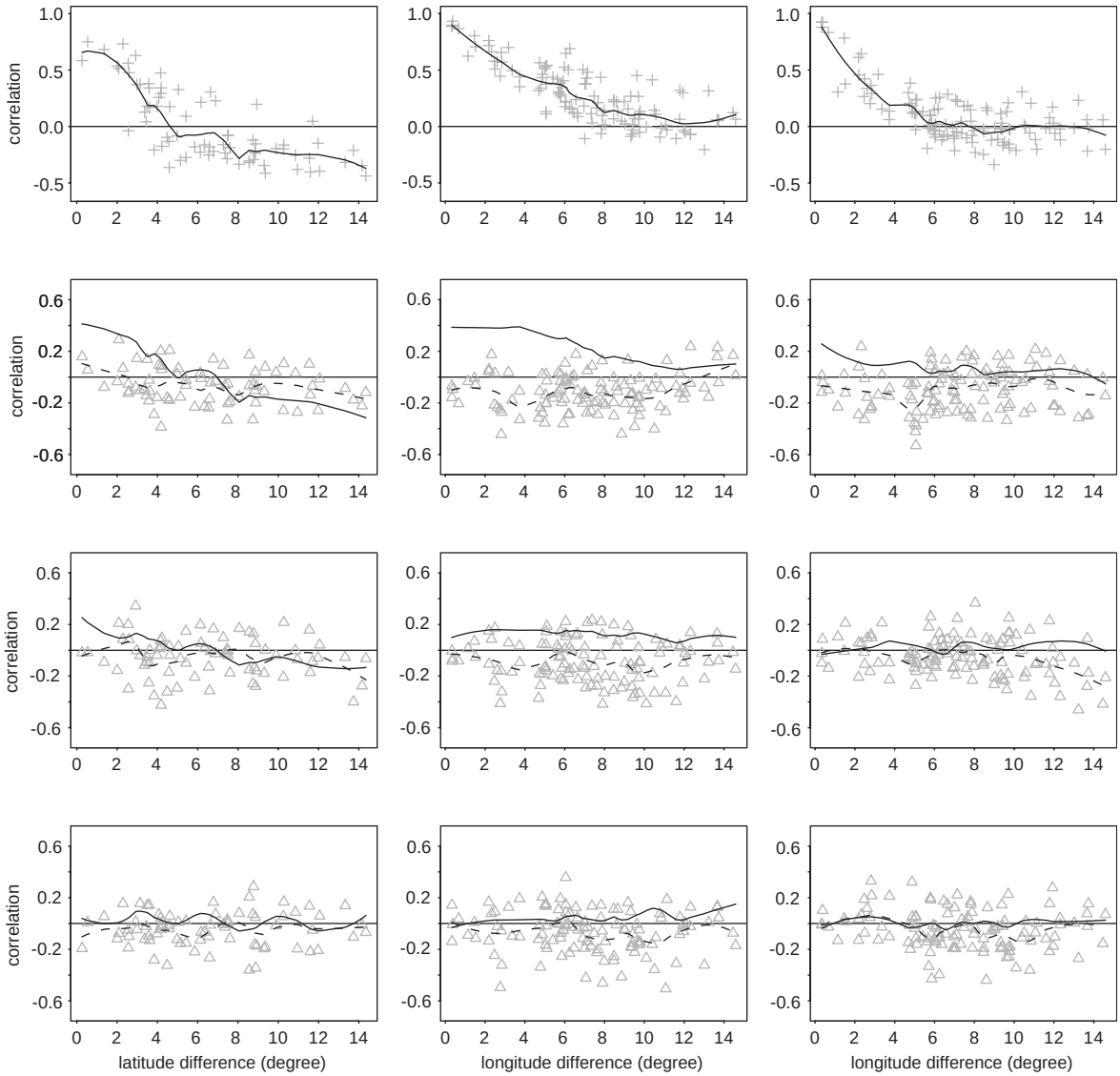

Fig. 4. Comparison of $D\,\mathbf{s}_1\ \mathbf{s}_2\ j$ and $D\,\mathbf{s}_1\ \mathbf{s}_2\ \ j$ for observations and CMAQ error for $j\ \ 0\ 1\ 2\ 3$ (top to bottom). Left two columns are for observations and the right column is for CMAQ error. Crosses (shown only for $j\ \ 0$) represent the average and triangles represent the difference of the two correlations. Solid lines are the smoothing lines of the crosses and dashed lines are the smoothing lines of the triangles. The smoothing is done by the procedure *loess* (Cleveland et al., 1992) with the smoother span 0.75.

get meridional "bands", where lon i and lat i denote the longitude and latitude of $\mathbf{s}_i\ \ i\ \ 1\ 2$ For the plot with respect to the longitude difference, we give a similar restriction with lon i and lat i switched. Fig. 1 illustrates the meridional and zonal "bands" for site 16. There are 82 pairs of sites that fall in the meridional bands and 127 pairs in the zonal bands. Since $D\,\mathbf{s}_1\ \mathbf{s}_2\ j\ \ \ D\,\mathbf{s}_2\ \mathbf{s}_1\ \ j$ we set $\mathbf{s}_1$ to be always north of $\mathbf{s}_2$ for the plot with respect to latitude difference and $\mathbf{s}_1$ to be east of $\mathbf{s}_2$ for the plot with respect to longitude difference.

The middle column in Fig. 4 shows clear evidence of asymmetry in the correlation for sites in a zonal band. Most of the triangles, which plot values of $D\,\mathbf{s}_1\ \mathbf{s}_2\ j$ $D\,\mathbf{s}_1\ \mathbf{s}_2\ \ j$ are below 0 for time lags 1, 2 and 3. This implies that there is a flow from west to east, since $\mathbf{s}_1$ is located east of $\mathbf{s}_2$ For $j\ \ 2$ the largest negative differences are at larger longitude differences than when $j\ \ 1$ This again shows the movement of the air flow from west to east. The left column in Fig. 4 shows at best a weak asymmetry indicating a flow from south to north.

If CMAQ is capturing the dynamics of air pollutant processes correctly, the space–time asymmetry should not show up in the correlation pictures of the CMAQ

error. The right column of Fig. 4 shows that there remains some degree of asymmetry for the CMAQ error for a time lag of one day. The migration of the largest negative differences to a larger longitude difference as time lag increases from $j \quad 1$ to $j \quad 2$ is even clearer than in the correlation plot of observations. However, in contrast to the results for the observations, we see hardly any asymmetry in the CMAQ error for a time lag of two days up to the longitude difference around 9 This indicates that CMAQ does a reasonable job of capturing dynamic effects on time scales longer than one day. Furthermore, the spatial correlations for the CMAQ error decrease to zero much faster than the spatial correlations of the observations and the level of space–time correlations with respect to each time lag for the CMAQ error is fairly low compared to the observations. These findings again show that it should be easier to develop statistical models for the CMAQ error than for the process itself.

For further evaluation of CMAQ, we suggest comparing correlations of linear combinations of the processes. We specifically consider the spatial correlation of temporal differences,

$$\mathrm{ST}\ \mathbf{s}_1\ \mathbf{s}_2\ j \quad \mathrm{Cor}\ Z_a\ \mathbf{s}_1\ t \quad Z_a\ \mathbf{s}_1\ t \quad j \quad Z_a\ \mathbf{s}_2\ t \quad Z_a\ \mathbf{s}_2\ t \quad j \qquad 3$$

and the temporal correlation of spatial differences,

$$\mathrm{TS}\ \mathbf{s}_1\ \mathbf{s}_2\ j \quad \mathrm{Cor}\ Z_a\ \mathbf{s}_1\ t \quad Z_a\ \mathbf{s}_2\ t \quad Z_a\ \mathbf{s}_1\ t \quad j \quad Z_a\ \mathbf{s}_2\ t \quad j$$

The idea behind looking at ST and TS correlations is to focus on space–time interactions (the $e\ \mathbf{s}\ t$ term in Eq. (1)). Statistics such as the space–time correlation function or space–time variogram of the unfiltered processes, because they are affected by temporal or spatial averages, describe some mixture of the various terms in Eq. (1). The ST and TS correlations of the filtered processes are two simple summary statistics that focus on the joint variation of the $e\ \mathbf{s}\ t$ process over space and time. Other statistics would also have this property and could be worth considering. The ST and TS correlations are estimated using estimates analogous to Eq. (2) for D

We can compare the ST and TS correlations for the observations and CMAQ output processes to check how the covariance structure of the processes match. We are interested in how each correlation varies as the distance between $\mathbf{s}_1$ and $\mathbf{s}_2$ or the time lag j changes. Fig. 5 compares the values of the empirical space–time correlations **ST** and **TS** for the observations, CMAQ output and the CMAQ error. We see that these correlations for the observations and CMAQ output are not so different. In the top row of Fig. 5, note that the ST correlations for CMAQ error are nearly zero after 500 km or so while the correlations for both observations and CMAQ output are noticeably negative from about 600 km to 2000 km The bottom row of Fig. 5 shows that the TS correlations of the CMAQ error do not increase nearly as strongly with distance as those of the observations. In particular, for $j \quad 2$ the TS correlations for the CMAQ error are nearly 0 but are clearly positive for the observations. Both ST and TS correlations show the greater simplicity of the space–time correlations of CMAQ error compared to those of the observations, which indicates that CMAQ is capturing much of the pattern in the original process properly.

We showed earlier that separability was not a tenable assumption for describing the joint distribution of observations and CMAQ output. Separability has also been used as a simplifying assumption about space–time covariance functions (Posa, 1993; Haas, 1995; Mitchell