1.15 COMPARISON OF SULFATE CONCENTRATIONS SIMULATED BY TWO REGIONAL-SCALE MODELS WITH MEASUREMENTS FROM THE IMPROVE NETWORK

John S. Irwin[1], Edith Gego[2], Christian Hogrefe[3], Jennifer M. Jones[3] and S. Trivikrama Rao[1*]*
[1]*NOAA Atmospheric Sciences Modeling Division, RTP, NC, U.S.A.*
[2]University Corporation for Atmospheric Research, Idaho Falls, ID, U.S.A.
[3]Atmospheric Sciences Research Center, University at Albany, Albany, NY. U.S.A.

INTRODUCTION

The complexity of turbulent flow is so formidable that a description of the flow characteristics at all points in space and time is not feasible. Consequently, any study of turbulent flows (either in the form of observations or solutions of the conservation equations) is directed towards describing the flow properties and the evolution of mean and turbulent quantities (*Haugen*, 1973; *Nieuwstadt and van Dop*, 1982; *Garratt*, 1992). This means that we can expect air quality models to hopefully agree with observations "in general" (i.e., compare average patterns in time and space), but we cannot expect models to match with observations "in particular" (i.e., at a particular point in time and space). Thus, the basis for development of meaningful model evaluation metrics is to devise analyses that avoid direct comparison of observations with predictions paired in time and space, as this would be requiring models to simulate "in particular." In this discussion, we test a method for assessing model performance that compares modeling results with observations in a general sense.

Previous investigations of observations of sulfate suggested there are locally contiguous subregions where monitoring results have similar temporal behavior (*Gego et al*., 2003), which had been detected through the use of a principal component analysis. Previous investigations of performance suggest that grid-based models are most skillful in simulating longer-term variations in time and space, as they lack the resolution and physics to simulate finer-scale variations (*Hogrefe et al.*, 2001ab). This suggests that if we sort the observations and modeling results into geographical subregions and compare running averages (computed over a long time window), we might have a robust first-order method for assessing regional-scale model performance that is faithful to the concept of testing a model on its ability to predict "in general."

DISCUSSION

Rotated principal component analysis

Principal component analysis (PCA) is a multivariate technique designed to facilitate interpretation of large data sets involving numerous mutually dependent variables. By summarizing the correlations (i.e., identifying the redundancies) between all variables, PCA allows determination of the 'true' dimensionality of a data set. It also allows building of a new data set (the principal components data set) whose dimensions reflect the true dimensionality of the original data set and whose variables are mutually orthogonal. *Eder* (1989) provided insights on how to use PCA to analyze and summarize the temporal correlation of time series of a given air contaminant measured at numerous monitored sites. In *Eder's* (1989) approach, a sample individual corresponds to a sampling event (date) and a variable is a monitoring site. PCA used in this framework allows classification of all monitoring sites into a limited number of categories (or subregions), each of which

[*] On assignment to the U.S. Environmental Protection Agency

corresponds to a specific contaminant's temporal evolution (specific succession of rises, falls and plateaus), i.e., a specific mode of variation.

One may consider that the information included in the original data set can be reasonably described by a limited number of PCs. The number of PCs retained is representative of the true dimensionality of the original data set. In our case, it also represents the number of 'distinct modes of variations' or the number of clusters we wish to differentiate in the data set. There are several methods for deciding the number of PCs to retain, among them the "Rule N" method (*Overland and Preisendorfer*, 1982), and the Scree test (*Cattell*, 1966; *Wilks*, 1995). No one approach is thought superior to the others. In this study, the number of clusters retained for each air pollutant and network is the number of eigenvalues greater than 1 as in *Eder* (1989). Orthogonally rotating the PCs retained so as to increase their correlation with the original data, a procedure often referred to as varimax (*Kaiser*, 1958), has been shown to facilitate interpretation of the principal components (*Horel*, 1981). We, therefore, chose to use it as well. Details on the application of PCA to monitoring data can be found in *Gego et al*., (2004).

Interagency Monitoring of PROtected VIsual Environment (IMPROVE) network

Initiated in 1985, the IMPROVE network essentially aims at monitoring air quality conditions in Class I areas, i.e., in national parks and wilderness areas that receive special protection from adverse air quality impacts through the U.S. Environmental Protection Agency's Prevention of Significant Deterioration (PSD) program (*U.S. Environmental Protection Agency*, 1980). The air sampler at IMPROVE sites consists of 4 modules located 3 m above ground level and equipped with a device that excludes particles larger than 2.6 micrometers from the sampler. Sulfate concentration is calculated by stoichiometry from the mass of sulfur extracted from a teflon filter and analyzed for by X-Ray fluorescence. A 24-hour integrated air sample is collected every three days. Measured concentrations are reported at ambient temperature and pressure conditions.

This study utilizes the sulfate concentrations reported by IMPROVE at sites located within the contiguous US for 1996. Only those sites having 85 or more out of 104 observations (82%) were retained for analysis. Because PCA cannot handle missing data, missing data at a given site were substituted for using a temporal linear interpolation scheme. A total of 57 out of 72 IMPROVE sites were used in the RPCA. From our PCA analysis of the IMPROVE observed sulfate concentrations values, we identified six subregions within the US having similar temporal behavior: Pacific Coast states, Four-Corners states, Idaho-Wyoming-Montana, New England states, Kentucky-Virginia, and Central Florida, which are shown in Figure 1.

CMAQ and REMSAD modeling results

One-year simulation results of sulfate for the contiguous US for 1996 for two regional-scale models were recently made available: Community Multi-Scale Air Quality Model (CMAQ, 2002 release, http://www.epa.gov/asmdnerl/models3/cmaq.html) and Regional Modeling System for Aerosols and Deposition (REMSAD, 7.06 version, http://remsad.saintl.com/). Both models used the CB IV chemical mechanism, both employed 36 km horizontal grids with 12 vertical layers where the lowest layer was approximately 38 meters in thickness. The meteorology for both models was provided by the Mesoscale Meteorological Model (MM5, version 2.12, http://www.mmm.ucar.edu/mm5/mm5-home.html). The MM5 output was processed by the Meteorology Chemistry Interface Processor (MCIP, v4.2.0, http://www.cmascenter.org/modelclear.shtml#cmaq). It was later detected that there were

errors in MCIP in defining the winds for use by CMAQ, so the results by both modeling systems should be viewed with caution (see http://www.cmascenter.org/workshop/). For our purposes of testing a method for assessing model performance, these 1996 simulation results are considered satisfactory.

For the analyses to be discussed, we paired the simulated 24-hour average sulfate value whose grid center was nearest the IMPROVE observation site, which are only available every three days. A five-week running average was applied to the predicted and observed sulfate values, and then averaged over all sites for each subregion to produce the illustrations displayed in Figure 2. The western subregions are the top three panels, and the eastern subregions are in the bottom three panels. Both models underpredict the summer increase in sulfate values in the Pacific Coast and Southwest states, whereas all models reproduce the temporal patterns reasonably well elsewhere. Discussion of these results with the modeling team who performed the model simulations revealed that a "clean" annual boundary condition of 0.5 micrograms per cubic meter was applied for sulfate. Investigations are underway to explore whether the sulfate boundary condition should have a seasonal peak during the summer.

CONCLUSIONS

To explore a possible method for assessment of model performance that is faithful to the concept of testing a model on its ability to predict "in general", we used the 1-year simulation results of sulfate for the contiguous US for 1996 for two regional-scale models: CMAQ (2002 release) and REMSAD (7.06 version). Previous investigations of observations of sulfate suggested there are local contiguous subregions where monitoring results have similar temporal behavior, and previous investigations of regional-scale modeling results suggests models are most skillful in simulating longer-term variations in time and space, as they lack the resolution and physics to simulate finer-scale variations. Using a principle component analysis of the IMPROVE observed sulfate concentrations values, we identified six subregions within the US having similar temporal behavior. Following the application of a 5-week running average to the observed and modeled values, we summarized the performance of the models in each subregion. Both models are seen to underpredict the summer increase in sulfate values in the Pacific Coast and South West states, whereas all models are reproducing the temporal pattern reasonably well elsewhere. It is hypothesized that the "clean" boundary condition used in these simulations for sulfate contributed to the bias seen in the results for the western subregions. The differences in magnitude of the sulfate values and differences in seasonal variations were clearly detected by the analyses performed. Assessing regional-scale model performance on their ability to characterize the longer-term variations in the time series within subregions is seen to provide valuable clues towards mitigating gross sources of bias and allows for a quantitative comparison of the relative performance of several models.

DISCLAIMER

The United States Environmental Protection Agency through its Office of Research and Development partially funded and collaborated in the research described here under Interagency Agreements (DW 13938634 and DW 13938483) with the Department of Commerce. The Department of Commerce partially funded and collaborated in the research described here under contracts with Dr. E. Gego (EA133R-03-SE-0710), and with the State University of New York to Dr. C. Hogrefe (EA133R-03-SE-0650). This paper has been reviewed in accordance with US EPA's peer and administrative review policies and has been approved for presentation and publication.

REFERENCES

Cattell R. B. (1966), The Scree test for the number of factors, Multivariate Behavioral Res., 1. 245-276.

Eder, B. K. (1989), A principal component analysis of SO42- precipitation concentrations over the eastern United States, Atmos. Environ. 23 (12), 2739-1750.

Garratt, J.R., (1992): The Atmospheric Boundary Layer. Cambridge Atmospheric and Space Science Series. Cambridge University Press. 316 pages.

Gego E., C. Hogrefe, P.S. Porter, J.S. Irwin, and S. Trivikrama Rao. (2003): Comparison of the space-time signatures of air quality data from different monitoring networks, Proceedings of the 26th NATO/CCMS International Technical Meeting on Air Pollution Modeling and Its Application, May 26 – 30, 2003, Istanbul, Turkey

Gego, E.L., P.S. Porter, J.S. Irwin, C. Hogrefe and S. Trivikrama Rao, (2004): Assessing the comparability of ammonium, nitrate and sulfate concentrations measured by three air quality monitoring networks. Accepted for publication in Pure and Applied Geophysics.

Haugen, D.A. (Editor), (1973): Workshop on Micrometeorology. American Meteorological Society, Boston, MA. 392 pages.

Hogrefe, C., S. T. Rao, P. Kasibhatla, G. Kallos, C. Tremback, W. Hao, D. Olerud, A. Xiu, J. McHenry, and K. Alapaty (2001): Evaluating the Performance of Regional-Scale Photochemical Modeling Systems: Part I - Meteorological Model Outputs. Atmos. Env., 35, 4159-4174

Hogrefe, C., S. T. Rao, P. Kasibhatla, W. Hao, G. Sistla, R. Mathur, and J. McHenry (2001): Evaluating the Performance of Regional-Scale Photochemical Modeling Systems: Part II - Ozone predictions. Atmos. Env., 35, 4175-4188

Horel, J.D. (1981), A rotated principal component analysis of the interannual variability of the Northern Hemisphere 500 mb height field, Mon. Wea. Rev., 109, 2080-2092.

Kaiser H. F. (1958), The varimax criterion for analytical rotation in factor analysis, *Psychometrika*, 23, 187-201.

Nieuwstadt, F.T.M. and H. van Dop. (1982): Atmospheric Turbulence and Air Pollution Modelling. D. Reidel Publishing Company. Boston, MA.358 pages.

Overland, J.E., and R.W. Preisendorfer, (1982), A significance test for principal components applied to a cyclone climatology, Mon. Wea. Rev., 110, 1-4.

Wilks, D.S., (1995): Statistical Methods in the Atmospheric Sciences: an Introduction, International Geophysic Series, Vol. 59 (Academic Press San Diego California).

U.S. Environmental Protection Agency. (1980): Prevention of significant deterioration Workshop Manual. EPA-450/2-80-081, NTIS PB-81-201410, U.S. Environmental Protection Agency, Research Triangle Park, NC.

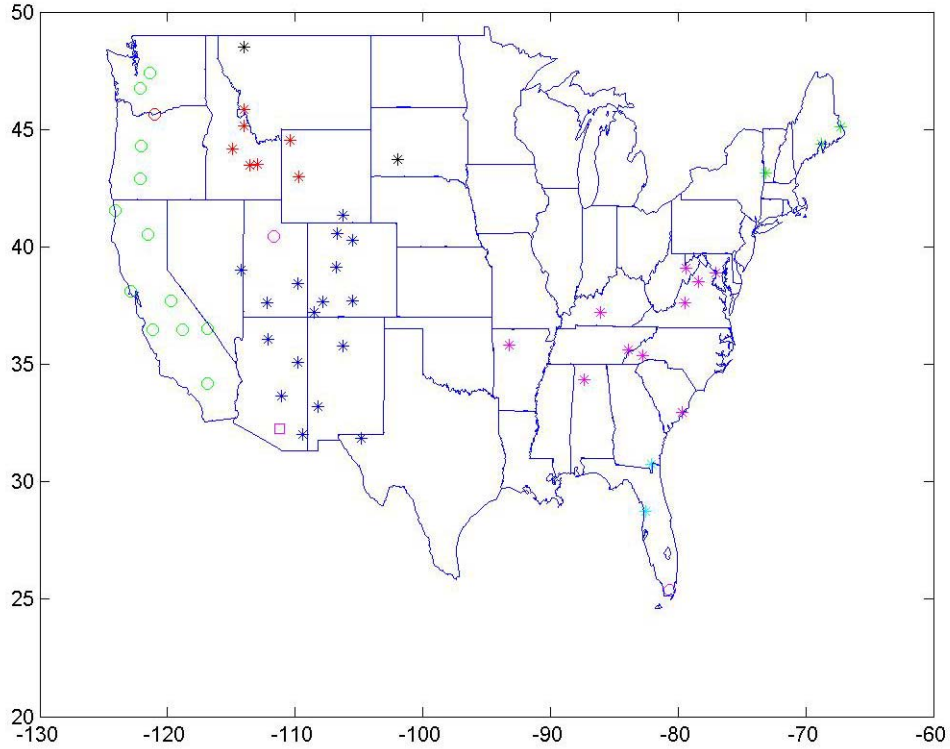

Figure 1. PCA results for the IMPROVE sulfate observations for 1996. There are six subregions differentiated using different symbols, three in western US and three in eastern US.

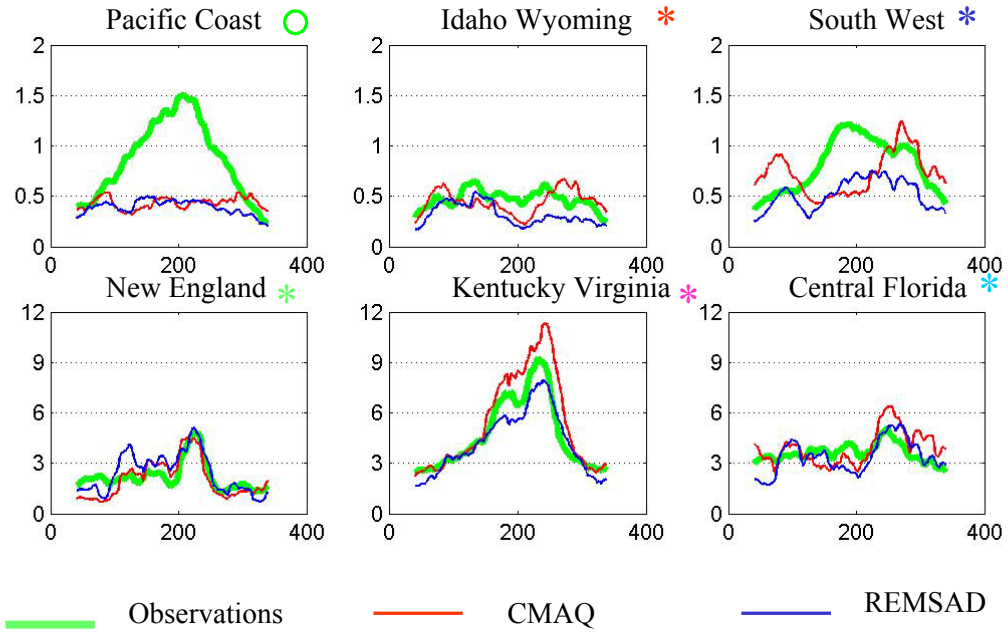

Figure 2. Five-week average sulfate concentration (micrograms per cubic meter) as a function of Julian day number for each of the subregions identified by the PCA analysis.