EM Feature

Spatial Prediction of Air Quality Data

by David M. Holland, William M. Cox, Rich Scheffe, Alan J. Cimorelli, Douglas Nychka, and Philip K. Hopke

This article is motivated by discussions at a recent workshop on spatial data analysis sponsored by EPA's Office of Air Quality Planning and Standards and the Office of Research and Development. A general consensus emerged at the workshop that it is now possible to model the spatial dependence of air pollution data to predict pollutant concentrations at unobserved locations, and that using these models would enhance regulators' ability to develop scientifically supportable public policy.

The demand for spatial models has grown rapidly over the past few years, as the need for spatial prediction in the regulatory environment has become reality. For example, how will the U.S. Environmental Protection Agency (EPA) (1) evaluate regional progress in air quality under the proposed Clear Skies Initiative and other national emissions control strategies; (2) redesign monitoring networks to ensure that it has sufficient data to describe changes in environmental quality that result from regulatory decisions; and (3) develop spatial models that combine data from urban and rural monitoring networks? It is imperative that recent developments in the theory and application of spatial models be used to provide meaningful answers to these problems of national scope and impact.

Currently, most air quality regulatory practices use monitoring data as independent point measurements with an assumed area of representativeness (e.g., a county). However, there is an increasing need, in part due to resource constraints, to develop regional air quality management policies and regulations using site measurements to evaluate air quality at locations that are not regularly monitored. Based on deliberate statistical research over the past two decades and the advent of inexpensive but powerful computing resources, there now exist accepted methods for predicting pollutant values at unobserved locations across the entire spatial field of interest based on available data. This spatial information, coupled with prediction uncertainties, will enable air quality managers to construct better emissions control programs and will be invaluable in estimating population exposure to support future health effects studies.

A recent workshop on spatial data analysis[1] that was jointly sponsored by EPA's Office of Air Quality Planning and Standards and the Office of Research and Development provided the motivation for this article. Attendees at the workshop included scientists and policy staff from EPA and experts in spatial statistics from academia and national laboratories. A general consensus emerged from the workshop that it is now possible to model the spatial dependence of air pollution data to predict reliable surfaces with uncertainties, and that using these models would enhance our ability to develop scientifically supportable public policy. These models need to be flexible, adapting to the underlying spatial variation in the data that might vary with location.

Spatial prediction is essentially a problem of spatial inference away from irregularly spaced points of measurement. Often there is a spatial continuity among monitoring measurements taken in close proximity, particularly in comparison to measurements taken at sites far apart. This phenomenon is typically due to physical attributes of the formation and transport of a pollutant that produces spatially coherent air quality patterns. Given this type of correlated spatial data structure, it is reasonable to expect that the pollutant values at unobserved locations spanning a large spatial field can be predicted based on the observed data. Thus, it is possible to obtain reliable predictions at many nonmonitored areas, estimate the interpolation error, and choose good locations for adding and/or deleting monitoring sites. This article supports the continued development and application of spatial models for prediction of air pollution at nonmonitored locations from the observed data. For readers who are unfamiliar with the technical terminology of spatial prediction, we also include a glossary (see sidebar on page 32) with nontechnical definitions of terms used in this article.

Glossary of Spatial Prediction Terms

Bayesian models Statistical models that include prior knowledge of unknown quantities in the form of a probability distribution.

Empirical orthogonal functions Empirical basis functions are used to express the spatial covariance function.

Geostatistics A methodology for the analysis of spatially correlated data. The characteristic feature is the use of statistical models to quantify the spatial correlation structure.

Isotropy The spatial relationship (e.g., correlation) between the field at two locations only depends on the distance of separation. Isotropic spatial fields have no preferred spatial orientation.

Kernel smoothing A method of spatial prediction using weighted averages of the observations. The weights are determined explicitly by a kernel function, a smooth bump-shaped function that weights observations based on their distance to the location of prediction.

Kriging A widely-used averaging method for prediction at nonmonitored locations. Kriging is based on spatial dependence, where observations that are more correlated with the spatial field at nonmonitored locations are given higher weight. The exact form of the weights depends on the spatial covariance function. Under the assumption that the covariance function is correct, the Kriging method provides unbiased predictions with minimum prediction variances.

Markov-random field models A statistical model for a spatial field defined on a regular or irregular grid of points. The model assumes that the distribution of any point on the grid depends on the nearest neighboring locations and a quantity that is statistically independent from the rest of the field. Much less statistical research has focused on this approach for spatial prediction.

Mean function A function for the large-scale variation of the data, usually the arguments of the function are the coordinates of the measurement location.

Parametric covariance A covariance function expresses the covariance between two values usually as a function of the coordinates of the two corresponding sites. Empirical covariances are estimated from the data.

Process convolution Modeling spatial data with Gaussian assumptions is common to many geostatistical analyses. A common approach in spatial statistics is to model the spatial dependence through a variogram. An alternative method for creating a Gaussian process is to take independent Gaussian variables and convolve them with an arbitrary kernel.

Residual process A random process characterizing small-scale variation in the data.

Stationarity A property of a statistical model for a spatial field that implies a degree of homogeneity. One consequence is that the distribution for two locations of a spatial field only depends on the difference of the location coordinates. Isotropy is a special case of stationarity.

Thin-plate spline models A method of spatial prediction equivalent to Kriging, but is based on a simplified covariance model. The functional form of the model tracks the data, but is also constrained to be smooth.

IMPORTANCE OF SPATIAL PREDICTION

Data from large-scale monitoring networks are used to inform the public of air quality levels and levels of exposure, establish compliance with standards, and evaluate the progress of air quality control programs. New or revised National Ambient Air Quality Standards (NAAQS) challenge data analysts to predict the effects of reduced emissions within nonmonitored areas. To better utilize spatial information for environmental decision-making, the air pollution scientific community can use spatial prediction methods to show important gradients of air pollution, design monitoring networks, and refine the definition of nonattainment boundaries. Further, we can embrace not only the use of spatial predictions, but output from air quality deterministic models to support conclusions on regulatory actions. For example, over the past decade, several observational-based models have been developed to address regulatory questions, such as the relative effectiveness of nitrogen oxides versus volatile organic compounds controls to reduce ozone (O_3). By integrating all available information, we can implement new approaches for holistic decision-making that integrate multiple perspectives and allow for the testing of different options.[2]

The spatial configuration of monitoring sites where air quality measurements are taken is an important component of inference from spatial data. For example, many air pollution monitoring sites are positioned to detect high concentrations and support decision-making relative to an area's attainment of pollution standards. However, when viewed as independent point measurements, these monitoring sites can give an incomplete picture of potential pollution problems even in areas of high monitor density. By using spatial statistical methods, it is possible to combine information from an array of monitoring sites and infer pollution levels across a much broader region than the isolated point or site measurements. Additionally, these methods provide a companion estimate of uncertainty attached to the spatial prediction. This allows us to distinguish areas where we can place confidence in the predictions based on low uncertainty levels.

In the United States and other countries, there are extensive large-scale networks of monitoring stations that collect data on atmospheric concentrations of pollutants such as O_3, sulfur dioxide (SO_2), and fine particulate matter ($PM_{2.5}$). The large number of stations, and thus the high costs of monitoring, is in part due to the use of these data as independent point measurements. A major criterion for modifying an existing network or deploying a new network is the quality of the resulting spatial predictions—and minimizing the monitoring costs of obtaining such predictions. The result will be networks that can better quantify people's exposure across the entire spatial field of interest. To accomplish this, it is important to ask what monitoring coverage (i.e., where, when, and what to monitor) is required to allow, in some quantitative sense, optimal predictions of the entire air quality field. This requires modeling the spatial dependence among monitoring sites to calculate interpolations and prediction errors, and then finding a network that optimizes some suitably defined measure of spatial information at nonmonitored locations.[3-7]

SPATIAL MODELS FOR AIR POLLUTION DATA

Spatial models are not entirely new tools for air quality planning. In fact, their potential use for air quality management was discussed in 1985 by Greenland and Yorty,[8] who suggested that attainment/nonattainment designations incorporate the use of spatial fields. Also, in a December 1994 Report to Congress,[9] EPA discussed the potential improvements in decision-making using predicted spatial fields. In the following section, a brief summary is provided of spatial and spatial-temporal models applied to air pollution and deposition data.

Geostatistics is the area of spatial statistics that addresses prediction of unknown values at specified locations or aggregations of locations. The prefix "geo" originally implied statistics pertaining to the earth,[10] but these methods have now been applied extensively to solve problems in many fields associated with the science of spatial prediction. These include meteorological and natural resource mapping problems, air pollution, and remote sensing. Geostatistical data are usually defined as point observations of a continuously varying quantity over a region in space. The strength of geostatistics over many other approaches is that it recognizes two types of spatial variability: large-scale variability due to known factors and small-scale variability. The first type of variability (mean structure) is usually modeled with a function of the site coordinates and the second type (residual process) has associated with it a spatial covariance function.[11,12]

The best interpolation methods appear to be those derived from treating the observed and unobserved values as a collection of correlated random variables[13] to produce spatial predictions at desired locations with minimum mean squared errors. These approaches, together with numerous refinements, are generally referred to as Kriging[14] in the literature. Kriging is a form of spatial averaging where neighboring sites are assigned weights such that the prediction error is minimized. The weights depend on the location of the monitoring sites and the spatial covariance estimated from the data. Many environmental processes, including processes underlying pollution, can be better understood by considering both their spatial and temporal dependence. Recent research has focused on the development of space-time models to solve problems that are inherently multivariate (i.e., more than one pollutant), and incorporate both the spatial and temporal correlation structure of the data.

MODELS FOR SPATIAL DEPENDENCE

The spatial covariance of atmospherically driven pollutant processes is affected by spatially varying features of landscape, topography, and the complex interaction of meteorology and emissions. Because of this complexity, one would not expect the spatial covariance function to have a simple form. However, early models for spatial data focused on simple correlation functions that depended only on the distance between monitoring sites. As noted by Sampson et. al.,[15] the lack of general models for spatial covariance and limited computing resources led to almost complete reliance on stationary models. However, recent research has developed more flexible models that accommodate nonstationary, or heterogeneous, covariance structure when considered over a large spatial range. Many approaches to modeling nonstationary covariance begin by smoothing locally stationary models over space or kernel smoothing of empirical covariances estimated from a finite number of monitoring sites. Haas[16-18] used a moving window approach based on assuming first and second order stationarity within each window to predict spatial patterns of wet sulfate deposition in the United States. This approach recognizes that spatial stationarity does not exist everywhere, but does exist within a given window of sites. Loader and Switzer[19] used modified covariance estimates obtained by shrinking the empirical covariances toward parametric covariances in a Bayesian context to predict patterns of wet sulfate deposition data in the eastern United States. Oehlert[20] applied a kernel smoothing approach for smoothly interpolating empirical covariances to estimate regional trends in wet sulfate deposition.

Over the past decade, more sophisticated models for global nonstationarity have been presented in the literature. A recent manuscript by Fuentes[21] described an approach for representing nonstationary covariance as a spatially weighted combination of local orthogonal stationary covariances in different regions. Covariance parameters are estimated in a Bayesian context to allow for predictive distributions accounting for the uncertainty of estimating these parameters. Decomposition of spatial processes in terms of empirical orthogonal functions has been applied by Nychka and Saltzman[22] and Holland et al.[23] to model O_3 and SO_2 spatial fields. In both of these papers, the spatial covariance is represented as a sum of a stationary isotropic process and a linear combination of basis functions with random coefficients designed to represent the deviation from stationarity.

Higdon[24] used a process-convolution approach for modeling nonstationary covariance structure based on the concept that any Gaussian stationary process can be expressed as the convolution of a Gaussian white noise process with a kernel function. To account for nonstationarity, the kernel function can be allowed to vary smoothly over space. Higdon[25] constructed a space-time process by spatially smoothing the white noise process defined over space and time with a purely spatial kernel function. After assigning prior distributions for all model parameters, this approach is used to predict daily spatial patterns of maximum 8-hr average O_3 concentrations.

A spatial deformation model was considered by Sampson and Guttorp[26] for modeling nonstationary covariance. These authors propose a nonlinear transformation to deform the geographic coordinates so that the process is approximately stationary under the new coordinates. Meiring et al.[27] discuss the use of thin-plate spline models to relate the two spaces, which appears to improve upon the transformation criterion used by Sampson and Guttorp. These techniques have been applied to analyses of acid rain and O_3 concentrations.[28-31] Royle and Berliner[32] developed a hierarchical model for the joint prediction of daily 8-hr average O_3 concentrations and maximum temperature in the Midwest. The method incorporates the dependence of O_3 on temperature and emphasizes the parameterization of relationships between variables in the mean as opposed to the covariance.

In recent years, hierarchical Bayesian approaches for both temporal and spatial interpolation have been developed, starting with Le and Zidek.[33] Brown et al.[34] and Le et al.[35] extended this method to deal with the multivariate setting where not all monitored sites measure the same set of pollutants. A major advantage of Bayesian methods is that the uncertainty in the estimation of both the mean and spatial covariance of the spatial field can be incorporated in the predictive distribution of pollution for any point of interpolation. Sun et al.[36] developed predictive distributions for nonmonitored PM_{10} concentrations in Vancouver using a Bayesian approach. They noted the underprediction of extreme values in the pollution field, but the Bayesian methodology provides useful estimates of uncertainties for large values. These uncertainties are needed in health impact analyses to evaluate the association between air pollution exposures and health outcomes. Cressie et al.[37] compared Kriging and Markov-random field models in the prediction of PM_{10} concentrations around Pittsburgh. Meanwhile, Kibria et al.[38] developed a multivariate spatial prediction methodology in a Bayesian context for the prediction of $PM_{2.5}$ data in Philadelphia. This approach was applied to $PM_{2.5}$ and PM_{10} data measured at monitoring sites with different start-up times.

FUTURE RESEARCH

The areas of spatial statistics in need of further research include

1. improving existing collection techniques for spatial-temporal data sets with high percentages of missing data;
2. developing better diagnostic methods for evaluating the fit of alternative spatial models and characterizing the nature of underlying nonstationarity;
3. combining atmospheric model output with monitoring data in a coherent way for improved spatial prediction and validation of model output;
4. incorporating the time-dependence (if it exists) in space-time models for better comparison to atmospheric model output and to allow forecasting of pollution over the short-term; and
5. developing specialized software for fast and optimal analyses of large data sets.

The use of output from atmospheric models should be emphasized. Given that many monitoring sites are located in areas of high pollution levels, model output that characterizes the low and high values of the underlying atmospheric processes might give a more reasonable representation of the mean surface. Also, a numerical model that explains most of the spatial variation would allow modeling differences in the data and model output that might be more stationary in the mean and covariance structure. Hierarchical Bayesian models offer powerful approaches for modeling complex pollution processes. However, these models are computationally demanding and innovative Monte Carlo approaches must be developed to allow wider implementation of these models for spatial prediction.

CONCLUSIONS

Over the past two decades, the application of spatial models has increased, particularly in the environmental sciences. The primary motivation for this development is the growing awareness that environmental data collected in space over an array of monitoring sites, like data collected over time, are likely to be correlated. However, much of this research has yet to percolate through to the routine analyses of air quality data. The perspective that a clear demarcation exists between the scientific and regulatory monitoring communities is counterproductive to optimizing the collective value of research and monitoring. Both spatial and spatial-temporal statistical models provide a bridge for broader interaction and synthesis between these two groups. An anticipated outcome of this collaboration is an improved ability of the air regulatory community to apply spatial models in a statistically rigorous fashion to uncover new insights for environmental managers. Spatial prediction has the potential to improve our response to increasingly complex air quality issues, to provide a sound basis for resource allocation decisions (particularly with respect to network design), to refine the boundaries of nonattainment areas, and to suggest new perspectives in the development of emissions control strategies.

DISCLAIMER

The U.S. Environmental Protection Agency (EPA) collaborated in the research described here. This article has been peer reviewed by EPA and approved for publication.

REFERENCES

1. Spatial Data Analysis Technical Exchange Workshop, Research Triangle Park, NC, December 2001; http://www.epa.gov/ttn/amtic/spatlwrks.html.
2. Stahl, C.H.; Cimorelli, A.J.; Chow, A.H. A New Approach to Environmental Decision Analysis: Multicriteria Integrated Resource Assessment (MIRA); *Bulletin of Science, Technology, and Society* **2002**, *22*, 443-459.
3. Cressie, N.; Gotway, C.A.; Grondona, M.O. Spatial Prediction from Networks; *Chemometrics and Intelligent Laboratory Systems* **1990**, *7*, 251-271.
4. Caselton, W.F.; Kan, L.; Zidek, J.V. Quality Data Networks that Minimize Entropy. In *Statistics in the Environmental and Earth Sciences*; A. Walden, P. Guttorp, Eds.; Halsted Press: New York, 1992; pp 10-38.
5. Holland, D.M.; Baumgardner, R.; Haas, T.; Oehlert, G. Design of the Clean Air Act Deposition Monitoring Network. In *Environmental Statistics, Assessment, and Forecasting*; R. Cothern, N.P. Ross, Eds.; Lewis Publishers: Ann Arbor, 1994; pp 147-162.
6. Müller, W.V. *Collecting Spatial Data: Optimum Design of Experiments for Random Fields*; Second Edition; Physica-Verlag: Heidelberg, 2000.
7. Zidek, J.V.; Sun, W.; Le, N.D. Designing and Integrating Composite Networks for Monitoring Multivariate Gaussian Pollution Fields; *Appl. Stats*. **2000**, *49*, 63-79.
8. Greenland, D.; Yorty, R.A. The Spatial Distribution of Particulate Concentrations in the Denver Metropolitan Area; *Annals Assoc. Am. Geog*. **1985**, *75*, 69-82.
9. *Clean Air Act Ozone Design Value Study: A Final Report*; EPA 454/R-94-035; U. S. Environmental Protection Agency: Research Triangle Park, NC, 1994.
10. Matheron, G. Principles of Geostatistics; *Econ. Geol*. **1963**, *58*, 1246-1266.
11. Cressie, N. The Many Faces of Spatial Prediction. In *Geostatistics, Volume 1*; M. Armstrong, Ed.; Kluwer: Dordrecht, 1989; pp 163-176.
12. Zimmerman, D.L. Spatial Design, Optimal. In *Encylopedia of Statistics, Volume 4*; A.H. El-Shaarawi, W. Piegorsch, Eds.; John Wiley & Sons: New York, 2001; pp 2067-2071.
13. *Encyclopedia of Statistical Sciences; Volume 3*; S. Kotz, N.L. Johnson, Eds.; John Wiley & Sons; New York, 1983.
14. Matheron, G. *The Theory of Regionalized Variables*; Centre de Morphologie Mathématique de Fontainebleau: Paris, 1971.
15. Sampson, P.D., Damian, D., Guttorp, P. *Advances in Modeling and Inference for Environmental Processes with Nonstationary Spatial Covariance*; NRCSE-TRS No. 061, 2001; National Research Center for Statistics and the Environment Technical Report Series, 2001.
16. Haas, T.C. Lognormal and Moving-Window Methods for Estimating Acid Deposition; *J. Am. Stats. Assoc*. **1990**, *85*, 950-963.
17. Haas, T.C. Kriging and Automated Variogram Modeling within a Moving Window; *Atmos. Environ*. **1990**, *24A*, 1759-1769.
18. Haas, T.C. Local Prediction of a Spatio-Temporal Process with an Application to Wet Sulfate Deposition; *J. Am. Stats. Assoc*. **1995**, *90*, 1189-1199.
19. Loader, C.; Switzer, P. *Spatial Covariance Estimation for Monitoring Data*; SIAM Institute for Mathematics and Society (SIMS) Technical Report No. 133, 1989; Stanford University: Stanford, CA, 1989.
20. Oehlert, G.W. Regional Trends in Sulfate Wet Deposition; *J. Am. Stats. Assoc*. **1993**; *88*, 390-399.
21. Fuentes, M. A New High Frequency Kriging Approach for Nonstationary Environmental Processes; *Environmetrics* **2001**; *12*, 469-483.
22. Nychka, D.; Saltzman, N. Design of Air Quality Networks. In *Case Studies in Environmental Statistics*; D. Nychka, W. Piegorsch, L. Cox, Eds.; Springer-Verlag: New York, 1998; pp 51-76.
23. Holland, D.M.; Saltzman, N.; Cox, L.; Nychka, D. Spatial Prediction of Sulfur Dioxide in the Eastern United States. In *GeoEnv II: Geostatistics for Environmental Applications*; J. Gómez-Hernández, A. Soares, R. Froidevaux, Eds.; Kluwer: Dordrecht, 1999; pp 65-76.
24. Higdon, D. A Process-Convolution Approach to Modeling Temperatures in the North Atlantic Ocean; *J. Environ. & Ecolog. Stats*. **1998**, *5*, 173-190.
25. Higdon, D. Space and Space-Time Modeling Using Process Convolutions. In *Quantitative Methods for Current Environmental Issues*; C. Anderson, V. Barnett, P.C. Chatwin, A.H. El-Shaarawi, Eds.; Springer Verlag: London, 2002; pp 37-56.
26. Sampson, P.D.; Guttorp, P. Nonparametric Estimation of Nonstationary Spatial Covariance Structure; *J. Am. Stats. Assoc*. **1992**, *87*, 108-119.
27. Meiring, W.; Monestiez, P.; Sampson, P.D.; Guttorp, P. Developments in the Modelling of Nonstationary Spatial Covariance Structure from Space-Time Monitoring Data. In *Geostatistics Wollongong '96, Volume 1*; E.Y. Baafi, N. Schofield, Eds.; Kluwer; 1997; pp 162-173.
28. Guttorp, P.; Le, N.D.; Sampson, P.D.; Zidek, J.V. Using Entropy in the Redesign of an Environmental Monitoring Network. In *Multivariate Environmental Statistics*; G.P. Patil, C.R. Rao, Eds.; Elsevier Science: New York, 1993; pp 175-202.
29. Guttorp, P.; Meiring, W.; Sampson, P.D. A Space-Time Analysis of Ground-Level Ozone Data; *Environmetrics* **1994**, *5*, 241-254.
30. Guttorp, P.; Sampson, P.D. Methods for Estimating Heterogeneous Spatial Covariance Functions with Environmental Applications. In *Handbook of Statistics, Volume 12*; G.P. Patil, C.R. Rao, Eds.; Elsevier Science: New York, 1994; pp 663-690.
31. Mardia, K.V.; Goodall, C.R. Spatial-Temporal Analysis of Multivariate Environmental Monitoring Data. In *Multivariate Environmental Statistics*; G.P. Patil, C.R. Rao, Eds.; Elsevier Science: New York, 1993; pp 347-386.
32. Royle, J.A.; Berliner, M.L. A Hierarchical Approach to Multivariate Spatial Modeling and Prediction; *J. Agri. Biolog. & Environ. Stats*. **1999**, *4*, 29-56.
33. Le, N.D.; Zidek, J.V. Interpolation with Uncertain Spatial Covariances: A Bayesian Alternative to Kriging; *J. Multivar. Anal*. **1992**, *43*, 351-374.
34. Brown, P.J., Le, N.D., Zidek, J.V. Multivariate Spatial Interpolation and Exposure to Air Pollutants; *The Canadian Journal of Statistics* **1994**, *22*, 489-510.
35. Le, N.D.; Sun, W.; Zidek, J.V. Bayesian Multivariate Spatial Iinterpolation with Data Missing by Design; *Journal of the Royal Statistical Society* **1997**, *59* (Series B), 501-510.
36. Sun, L.; Zidek, J.V.; Le, N.D.; Özkaynak, H. *Interpolating Vancouver's Daily Ambient PM_{10} Field*; NRCSE-TRS No. 033, 1999; National Research Center for Statistics and the Environment Technical Report Series, 1999.
37. Cressie, N.; Kaiser, M.S.; Daniels, M.J.; Aldworth, J.; Lee, J.; Lahiri, S.N.; Cox, L. Spatial Analysis of Particulate Matter in an Urban Environment. In *GeoEnv II: Geostatistics for Environmental Applications;* J. Gómez-Hernández, A. Soares, R. Froidevaux, Eds.; Kluwer: Dordrecht, 1999; pp 41-52.
38. Kibria, Golam B.M.; Sun, L.; Zidek, J.V.; Le, N.D. *A Bayesian Approach to Backcasting and Spatially Predicting Unmeasured Multivariate Random Space-Time Fields with Application to $PM_{2.5}$*; The University of British Columbia, Department of Statistics, Technical Report #193, 2000.

About the Authors

David M. Holland (*corresponding author; e-mail: holland.david@epamail.epa.gov*) is a senior statistician at EPA's Office of Research and Development in Research Triangle Park, NC. **William M. Cox** is a statistician and **Rich Scheffe** is leader for the monitoring and quality assurance group at EPA's Office of Air Quality Planning and Standards in Research Triangle Park, NC. **Alan J. Cimorelli** is a lead meteorologist at EPA Region III in Philadelphia, PA. **Douglas Nychka** is section head of the Geophysical Statistics Project at the National Center for Atmospheric Research in Boulder, CO. **Philip K. Hopke** is the Bayard D. Clarkson Distinguished Professor in the Department of Chemical Engineering and Director of the Center for Air Resources Engineering and Science at Clarkson University in Potsdam, NY.