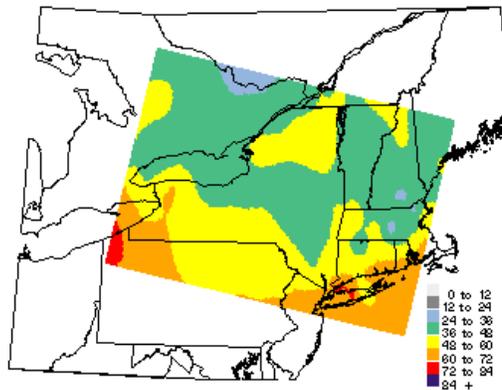
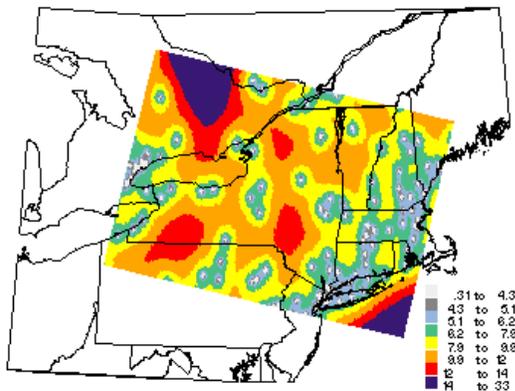


Developing Spatially Interpolated Surfaces and Estimating Uncertainty

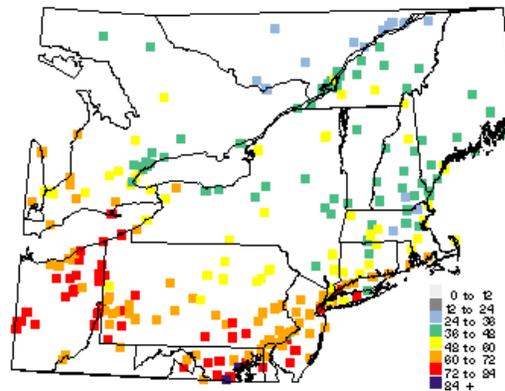
Interpolated O₃ 18JUN01



Standard Error O₃ 18JUN01



Observed O₃ 18JUN01



EPA-454/R-04-004
November 2004

Developing Spatially Interpolated
Surfaces and Estimating Uncertainty

U.S. ENVIRONMENTAL PROTECTION AGENCY
Office of Air and Radiation
Office of Air Quality Planning and Standards
Research Triangle Park, NC 27711

DISCLAIMER

The information in this document has been reviewed in accordance with the U.S. EPA administrative policies and approved for publication. Mention of trade names or commercial products in this document does not constitute endorsement or recommendation for use.

ACKNOWLEDGMENTS

The majority of the work performed to prepare this document was accomplished under contract no. 68-D-02-061 with Battelle. We would like to express our appreciation to Dave Holland, EPA Office of Research and Development for insights and helpful comments.

TABLE OF CONTENTS

DISCLAIMER	i
ACKNOWLEDGMENTS	ii
1.0 INTRODUCTION	1
2.0 OVERVIEW OF SPATIAL INTERPOLATION MODELS	3
2.1 General Characteristics of the Data	6
2.2 General Characteristics of Spatial Interpolation Methods	6
2.3 Specific Methods of Spatial Interpolation	9
2.4 References	14
3.0 THE ORDINARY KRIGING MODEL	16
3.1 Step 1: Build the Analysis Data Set	17
3.2 Step 2: Summarize and Understand the Data	21
3.3 Step 3: Conduct a Variogram Analysis	25
3.4 Step 4: Apply Spatial Prediction and Uncertainty Formulas	35
3.5 Step 5: Evaluate Model Performance Through Diagnostics	38
3.6 References	41
4.0 COMMON EXTENSIONS TO THE ORDINARY KRIGING MODEL	42
4.1 Covariance Extensions	42
4.2 Trend Extensions	56
4.3 Dimensional Extensions	64
4.4 References	66
5.0 MODEL EVALUATION AND ACCEPTANCE	68
5.1. Specific Performance Measures for Model Evaluation	68
5.2 References	76
6.0 SOFTWARE TO DEVELOP SPATIAL INTERPOLATION MODELS	78
6.1 SAS	78
6.2 S-Plus	82
6.3 Venables and Ripley	85
6.4 Fields	86
6.5 EPA Design Interface (EPA-DI)	88
6.6 ArcGIS Software by ESRI	90
6.7 GSLIB	92
6.8 Summary	94
6.9 References	95

7.0	KRIGING MODEL EXAMPLES	96
7.1	PM _{2.5} Annual Average Data	96
7.2	Ozone 8-Hour Design Value Data	110
7.3	References	123
8.0	SPATIAL INTERPOLATION MODELING LIMITATIONS	124
8.1	Air Pollutant of Concern	124
8.2	Spatial Scale	125
8.3	Spatial Prediction Uncertainty	126
8.4	Monitoring Network Design	128
8.5	References	130
9.0	SPATIAL INTERPOLATION MODELING ALTERNATIVES	131
9.1	Alternative Non-Statistical Approaches	131
9.2	Other Statistical Methodologies	134
9.3	Flexible Models Handling Nonstationarity	137
9.4	References	138
10.0	SPATIAL INTERPOLATION MODELING REFERENCES	139
10.1	General Spatial Interpolation	139
10.2	Ordinary Kriging	141
10.3	Extensions to Ordinary Kriging	142
10.4	Software References	144
10.5	Alternative Approaches to Spatial Interpolation	147

LIST OF TABLES

Table 3.1	Example analysis data set for two-dimensional spatial interpolation	18
Table 3.2	Distribution (summary statistics) of annual average PM _{2.5} values by spatial sub-region within the domain of interest for the case study example	23
Table 3.3	Comparison of variogram models	32
Table 3.4	Comparison of model parameters (using a robust empirical variogram) across temporal scales (PM _{2.5} data)	32
Table 4.1	Comparison of percentiles of standard errors between isotropic and anisotropic model based on annual PM _{2.5} data	52
Table 4.2	Numeric comparison of model parameters for relative variograms	55
Table 4.3	Comparison of percentiles of standard errors between ordinary and universal kriging models for same data set	62
Table 6.1	Summary of Software Packages Reviewed in Section 6	95
Table 7.1	Summary of Variogram Model Parameters	100
Table 7.2	Effect estimates for covariates in the universal kriging.	118
Table 7.3	Design values, interpolated values, and residuals for locations within region.	122

LIST OF FIGURES

Figure 2.1.a	Spatial interpolation of annual average $PM_{2.5}$ concentrations for 2000.	4
Figure 2.1.b	Standard errors of estimated annual average $PM_{2.5}$ concentrations for 2000.	5
Figure 2.2	Geographic areas estimated to exceed (dark gray) or not exceed (white) an annual average $PM_{2.5}$ level of $16 \mu\text{g}/\text{m}^3$ in 2000. The conclusion is less certain for those areas shaded light gray.	5
Figure 2.3	Illustration of Gradual vs. Abrupt Interpolator.	8
Figure 2.4	Examples of polygonal interpolations in one spatial dimension (left) and two spatial dimensions (right)	10
Figure 2.5	Comparison of Spline and Polynomial Interpolation Methods	11
Figure 3.1	Spatial distribution of $PM_{2.5}$ monitoring sites within domain of interest for case study example.	23
Figure 3.2	Partitioning of domain of interest for case study example.	24
Figure 3.3	Distribution (side-by-side box plots) of annual average $PM_{2.5}$ values by spatial sub-region within domain of interest for case study example.	24
Figure 3.4	Direct comparison of three model families	27
Figure 3.5	Empirical variogram (circles) and parametric variogram model (solid curve) based on annual $PM_{2.5}$ data.	33
Figure 3.6	Comparison of variogram model parameters across temporal scales ($PM_{2.5}$ data)	34
Figure 3.7	Ordinary kriging predictions based on annual $PM_{2.5}$ data from case study example.	37
Figure 3.8	Ordinary kriging standard errors based on annual $PM_{2.5}$ data from case study example.	37
Figure 4.1	Comparison of Isotropy to Geometric Anisotropy	44
Figure 4.2a	Geometric anisotropy plot — in each subplot the upper label indicates the ratio and the lower label indicates the angle (angles 0, 45, 90, and 135; ratios 2, 4, 8, and 16) — annual average $PM_{2.5}$	46
Figure 4.2b	Directional variograms using S-Plus code — annual average $PM_{2.5}$	47
Figure 4.2c	Geometric anisotropy plot — in each subplot the upper label indicates the ratio and the lower label indicates the angle (angles 17, 62, 107, and 152; ratios 2, 4, 8, and 16) — coal seam data	48
Figure 4.2d	Directional variograms using S-Plus code — coal seam data	48
Figure 4.3	Comparison of empirical variograms and variogram models under the isotropic and anisotropic covariance assumptions	50
Figure 4.4	(a) Exponential (solid line, $a=1.333$) versus Gaussian (dashed line, $a=2.3094$) variogram model (range=4, sill=7, nugget=2). (b) Geometric anisotropy versus isotropic variogram model (range=4, dashed line vs. range=2, solid line; sill=7, nugget=2).	51
Figure 4.5	Kriging predictors for (a) isotropic and (b) anisotropic model based on annual $PM_{2.5}$ data	52
Figure 4.6	Graphical representations of relative variogram models	55

Figure 4.7	Graphical comparison of full region variogram model versus sub-region empirical variograms.	56
Figure 4.8	Examples of increasingly complex polynomial trend surfaces	59
Figure 4.9	Empirical variogram and variogram model for universal kriging model.	61
Figure 4.10	Universal kriging predictions based on annual PM _{2.5} data from case study example.	61
Figure 4.11	Contour plot of feature differences between ordinary and universal kriging interpolations of annual PM _{2.5} data.	63
Figure 5.1	Example Histogram of Kriging Residuals	70
Figure 5.2	Example Plus-Minus Plot	71
Figure 6.1	S+SpatialStats dialog box for ordinary kriging.	84
Figure 6.2	ArcGIS Spatial analyst kriging dialog box.	91
Figure 7.1	S-Plus empirical variogram window.	97
Figure 7.2	S-Plus model variogram window.	98
Figure 7.3	Comparison of empirical variograms.	99
Figure 7.4	Baseline ordinary kriging predictions for annual PM _{2.5} data.	101
Figure 7.5	S-Plus command window, illustrating the directional variogram code.	101
Figure 7.6	Directional variogram plot.	102
Figure 7.7	Sub-region variograms.	104
Figure 7.8	Sub-region empirical variograms overlaid with an overall variogram model . .	106
Figure 7.9	S-Plus universal kriging window.	107
Figure 7.10	Universal Kriging predictions based on annual PM _{2.5} data.	108
Figure 7.11	Side-by-side box plots of PM _{2.5} concentrations at urban versus rural monitoring sites.	109
Figure 7.12	Locations of monitoring stations over southeastern United States.	110
Figure 7.13	Eight-hour ozone design values for 1999-2001 observed at stations across southeastern United States.	111
Figure 7.14	Empirical semi-variogram for full dataset.	113
Figure 7.15	Theoretical and empirical semi-variograms for the full dataset.	114
Figure 7.16	North/South empirical semi-variogram for the full dataset.	115
Figure 7.17	East/West empirical semi-variogram for the full dataset.	115
Figure 7.18	Spatially Interpolated Surface of eight-hour ozone design values.	117
Figure 7.19	Standard errors for the spatially interpolated surface of eight-hour ozone design values.	117
Figure 7.20	Leave-one-out cross validation (single variogram) residuals.	118
Figure 7.21	Plus/minus plot of leave-one-out cross validation residuals.	118
Figure 7.22	Leave-one-out cross validation residuals in the Birmingham area.	120
Figure 7.23	Scatter plot of observed and interpolated design values.	120
Figure 7.24	Residuals (observed minus interpolated) in the Birmingham area.	122

1.0 INTRODUCTION

The need for spatial interpolation models in the regulatory environment has grown in the past few years. The EPA is using these models to review decisions on monitoring network design and to predict the efficacy of emission control programs. Due to the limited number of monitoring sites across the country for ambient concentrations of ozone and fine particles, there is a need to use spatial interpolation to predict ambient concentrations in unmonitored locations. Support for these methods has emerged from scientists and state/local/EPA agencies in recent workshops¹. The general consensus is that it is now possible to model the spatial dependence of air pollution data to reliably predict concentrations in unmonitored locations along with associated uncertainties for use in developing regulatory policy². EPA recognizes the merits of these methods, more specifically kriging, for use in the modeled attainment tests for the 8-hour ozone and PM 2.5 National Ambient Air Quality Standards attainment demonstrations. These methods provide environmental decision makers the opportunity to show important gradients of air pollution, review the location of monitoring networks and refine the definition of nonattainment boundaries.

The purpose of this document is to provide an overview and better understanding of spatial interpolation methods. Key to selecting an interpolation method and understanding the results is understanding the data. Characteristics of the data that are important to consider are spatial representativeness, temporal sampling frequency, measurement accuracy, and existence of spatial relationships or behaviors at varying scales. This document discusses whether there is a need to force the interpolated surface to pass through the measured values, whether the data contain a global trend across the entire area of interest, or whether short-range variation is significant. These are important features to consider when interpolating spatial data. General interpolation methods and data considerations are discussed in Section 2. Ordinary kriging, a geostatistical spatial interpolation method, is discussed in Section 3, along with an example of developing an interpolated surface of PM_{2.5} concentrations in the eastern U.S. This section also briefly touches on limitations of this approach and methods for evaluating model performance through diagnostics. Common extensions to ordinary kriging such as including spatial trends, temporal dynamics, non-stationary covariance structures, use of covariates and multivariate modeling are presented in Section 4. Section 5 contains more details on model evaluation. Section 6 discusses software for performing spatial interpolation analysis. Section 7 provides an example use of S-Plus to identify an optimal kriging estimate for annual average PM_{2.5}

¹“*Spatial Data Analysis Technical Exchange Workshop*”, Research Triangle Park, NC, December 2001; <http://www.epa.gov/ttn/amtic/spatwrks.html> and “*2nd Particulate Matter/Regional Haze/Ozone Modeling Workshop*”, Santa Fe, NM, June 2003; <http://www.epa.gov/scram001/pmwork.htm> .

²Holland, D.M., W.M. Cox, R. Scheffe, A.J. Cimorelli, D. Nychka, and P.K. Hopke, (2003), “*Spatial Prediction of Air Quality Data*”, EM, August 2003, p. 31-35.

concentrations and an example use of SAS to krig 8-hour ozone concentrations. Section 8 discusses limitations to spatial interpolation. Section 9 discusses alternative methods that may be applied when certain limitations to kriging are relevant. Section 10 provides a summary of references on various aspects of spatial interpolation. This document should be a helpful reference and synopsis of methods and issues to be addressed when interpolating ambient air quality concentrations. For more details, a list of references has been provided within each section. Finally, this document is not intended to be static. As new information becomes available, EPA will update the document to reflect significant advances in spatial interpolation technology.

2.0 OVERVIEW OF SPATIAL INTERPOLATION MODELS

Spatial interpolation as applied to air pollution data is loosely defined as the procedure for estimating ambient air concentrations at unmonitored locations throughout an area based on available observations within some proximity of the area. The justification underlying spatial interpolation is the assumption that points closer together in space are more likely to have similar values than points more distant. This observation is known as Tobler's First Law of Geography [1]. Spatial interpolation is a very important component of many geographical information systems (GIS), frequently used as a tool to aid spatial decision making both in (1) physical and human geography and (2) related disciplines, such as air quality research and mineral prospecting. Many of the techniques of spatial interpolation are two-dimensional developments of the one-dimensional methods originally developed for time series analysis [2].

Figures 2.1.a and 2.1.b provide an example of the type of information that can be provided by a spatial interpolation modeling exercise. Figure 2.1.a displays a contour map of spatially-interpolated (statistically modeled) annual average $PM_{2.5}$ concentrations over much of the eastern United States. The contours indicate the spatial gradients of the annual $PM_{2.5}$ process across the domain, where state boundaries have been overlaid for perspective. Figure 2.1.b displays a similar map of the spatial prediction errors (uncertainty) associated with the interpolation.³ Uncertainty estimation can be a critical component of spatial interpolation because the outputs from such an exercise are model estimates, not true values. These results are based on a universal kriging model applied to annual average $PM_{2.5}$ ambient air monitoring data generated by EPA's Federal Reference Method (FRM) network during calendar year 2000. These data will be used for a running case study example presented for illustration purposes throughout this and other sections of the document.

The contours of Figure 2.1.a provide a general indication of spatial $PM_{2.5}$ air quality at the annual scale. A spatial interpolation exercise can also be used to answer more specific questions. For example, which areas of the depicted region exceed an annual arithmetic mean $PM_{2.5}$ level of $16 \text{ : } g/m^3$? Figure 2.2 provides an answer to that question, along with an indication of uncertainty. Based on the spatial interpolation model's results (same results as those used to generate Figures 2.1.a and 2.1.b), Figure 2.2 identifies those geographic areas that are estimated to exceed $16 \text{ : } g/m^3$ for an annual $PM_{2.5}$ level (dark gray) by at least one standard deviation, those areas estimated to fall below $16 \text{ : } g/m^3$ (white) by one standard deviation, and those areas within +/- one standard deviation of $16 \text{ : } g/m^3$ for which the conclusion is less certain (light gray).

³ Note in Figure 2.1.b that some contiguous contour lines display the same standard error value (e.g., a 1.5 contour line next to a 1.5 line). Such results are caused by rounding, as the default setting in S-Plus for displaying values of the contour lines is one decimal place. Some manipulation of the S code running the default program would be required to expand the number of decimal places displayed.

The remainder of this section provides a more detailed overview of spatial interpolation and various approaches. However, the primary focus of this document is the spatial interpolation method known as kriging, in particular, ordinary kriging. Section 3 discusses ordinary kriging in detail and Section 4 provides some kriging-based and other extensions to the ordinary kriging approach. One of the main reasons for focusing on kriging in this document is that it is a member of the stochastic class of spatial interpolation schemes (discussed in Section 2.1). In particular, kriging is a statistical model that produces both a spatial surface of predictions for the process of interest as well as the uncertainty associated with those estimates. It is an advantage of stochastic methods that they can provide measures of uncertainty in the spatial interpolation model's output, which serves to guide spatial decision making.

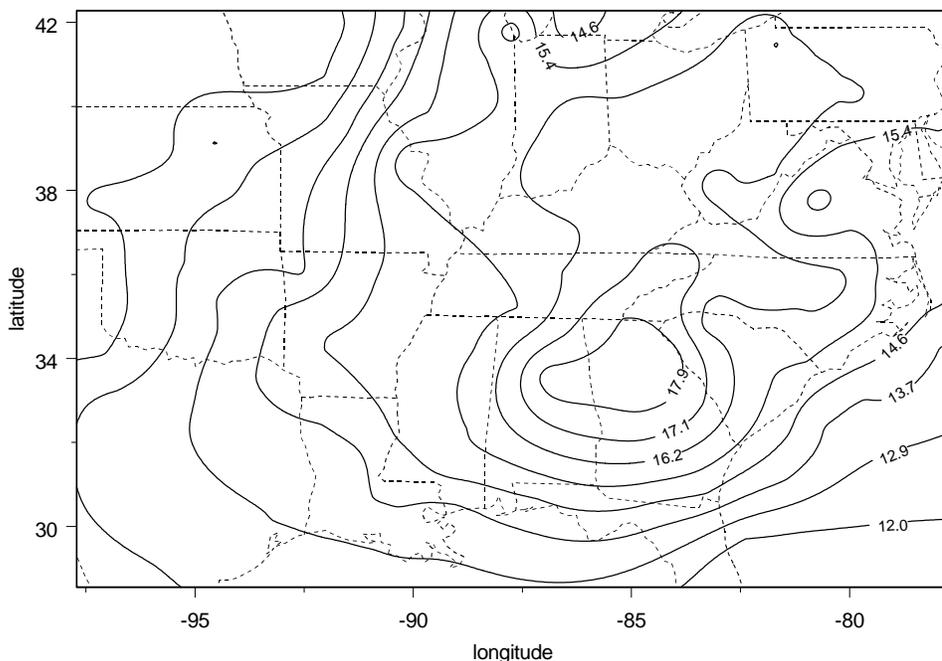


Figure 2.1.a Spatial interpolation of annual average PM_{2.5} concentrations for 2000.

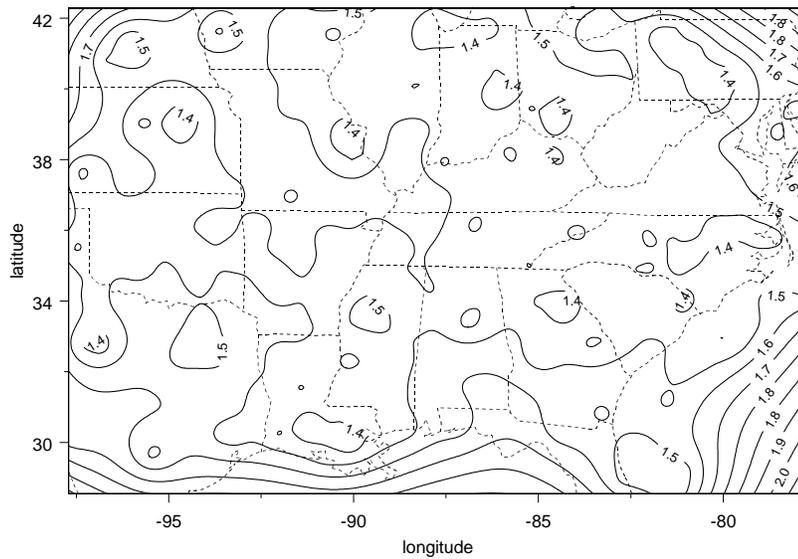


Figure 2.1.b Standard errors of estimated annual average PM_{2.5} concentrations for 2000.

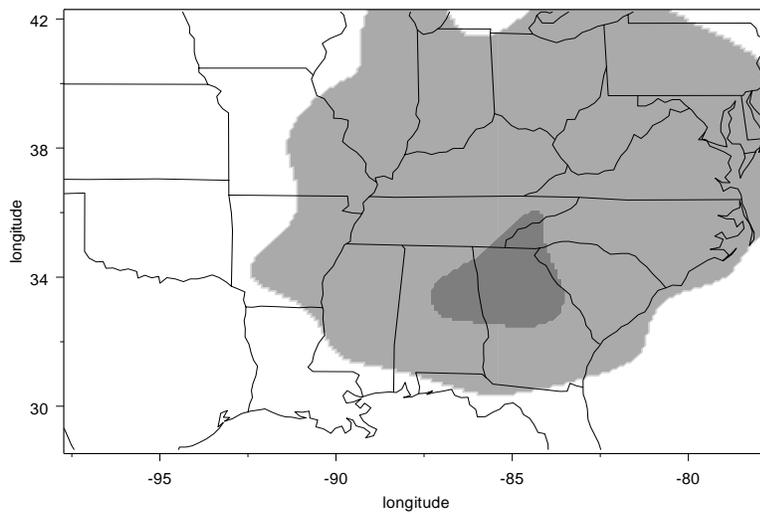


Figure 2.2 Geographic areas estimated to exceed (dark gray) or not exceed (white) an annual average PM_{2.5} level of 16 : g/m³ in 2000. The conclusion is less certain for those areas shaded light gray.

2.1 General Characteristics of the Data

Knowledge about the data that are to be interpolated is critical to selecting an appropriate interpolation method and to understanding the results produced by the interpolation. Characteristics of the data that are important to consider include spatial representativeness, temporal sampling frequency, measurement accuracy, and existence of spatial relationships or behaviors at varying scales. If data are gathered for a specific purpose, such as measuring broad pollutant levels in densely populated areas, an interpolation based on those data may not perform well at predicting pollutant levels in rural areas. If data are not gathered for purposes of measuring pollutants issued from point sources, an interpolation may not perform well at predicting peaks in the surface caused by point sources. Also, there may not be a need to force the interpolated surface to pass through the measured values if there is a high degree of imprecision in those measurements. Another trait to consider is the true nature of the spatial process being measured; for example, whether the data are expected to contain a global trend across the entire area of interest or whether short-range variation is significant. Such conditions might affect decisions regarding the use of global or local interpolators.

2.2 General Characteristics of Spatial Interpolation Methods

There are several characteristics of spatial interpolation methods that are useful to review prior to discussing specific methods. These characteristics include point-based versus areal-based, global versus local, exact versus approximate, stochastic versus deterministic, gradual versus abrupt, and the ability to consider covariates [3]. Each of these classifications is described below.

Point-based versus Areal-based: Point-based interpolation methods predict values at specific points in space based on the values and locations of other individual points in space. Predicting an ozone concentration at a specific latitude and longitude based on the measurements from a monitoring network is an example of this. On the other hand, areal interpolation methods estimate values for entire zones or areas based on data available for a different set of zones or areas. For example, given the population in surrounding Counties A, B, C, and D, estimate the population in County E. As another example, many commonly-used atmospheric dispersion models predict areal averages (i.e., average pollution levels over grid cells).

Global versus Local: Whether an interpolation method utilizes a single function that is mapped across the entire area of concern or breaks up the area into smaller blocks that are evaluated individually is another important characteristic. Global interpolators develop and utilize a single function that estimates values for the entire sample area. Thus, changing one input data point affects the predictions for the entire area. Local interpolators break the full sample area into smaller pieces that are each evaluated individually by a particular function. With a local interpolator, a change in a single data point affects only those areas that consider that point in the prediction algorithm. Local interpolation methods that consider a large percentage of the measured data points in each localized calculation become similar to global

methods. One example of a global interpolator is global polynomial interpolation, a method that fits a smooth surface based on a single mathematical function over the measured points. This type of interpolation may be useful for interpolating surfaces with gradual variation over the area of interest.

Exact versus Approximate: Interpolation methods vary based on whether the predicted surface must include the exact values of the measured data points or not. Exact interpolators do match the measured values on which the interpolation is based. Thus, the predicted surface must pass through each measured data point. Approximate interpolators utilize the measured values in calculating the predicted surface, but the surface is not restricted to passing through the measured values at those locations. This feature of approximate interpolators may make them more attractive for some users as they can produce a surface that avoids sharp peaks and troughs in the estimated surface.

Approximate interpolators may be more appropriate when there is uncertainty about the accuracy of the measured values (i.e., measurement error). On the other hand, if a user has confidence in the values of the measured data points, an exact interpolator may be preferred. For example, if an analyst is interpolating rainfall based on exact measurements obtained from weather stations with perfect accuracy, the analyst may feel comfortable using an exact interpolator that reflects those measurements. Keep in mind, however, that the choice of an interpolation method is driven by a user's needs and even with precise, accurate data a user may prefer to not restrict the predicted surface to reproducing the measured values exactly. It is possible that by allowing the predicted surface to deviate from the measured values, the predictions for non-sampled locations may be more accurate [2].

Stochastic versus Deterministic: Whether methods utilize the concept of randomness is another important characteristic to consider. Stochastic methods incorporate the idea of randomness into the interpolation process. These methods, which include kriging, allow the uncertainty of the predicted values to be calculated. Deterministic methods do not incorporate statistical probability theory into development of the predictions. Instead, these methods use mathematical formulas or other relationships to interpolate values. An example of a deterministic method would be one that derives a predicted value by a simple averaging of nearby measured points. Inverse Distance Weighted (IDW) is a deterministic method that uses a weighted average of nearby points with distance being the only factor influencing calculation of the weight. The advantage of stochastic methods is the ability to provide estimates of uncertainty for the spatial interpolation model's output. Kriging is a stochastic method because it assigns weights based not only on the distance between surrounding points but also on the spatial autocorrelation among the measured points, which is determined by modeling the variability between points as a function of separation distance.

Gradual versus Abrupt: Another distinguishing characteristic of spatial interpolators is the smoothness of the predicted surface that is produced. A gradual interpolator produces a surface with gradual, relatively smooth, changes. An abrupt interpolator produces a discontinuous surface with sharp changes. Figure 2.3 is an illustration of the difference between

a gradual and abrupt interpolator. The proximal “nearest-neighbor” method, which sets unknown points equal to the nearest measured point, is an example of an abrupt interpolator (see Section 2.2). Note also that local interpolators can produce discontinuous surfaces with abrupt changes.

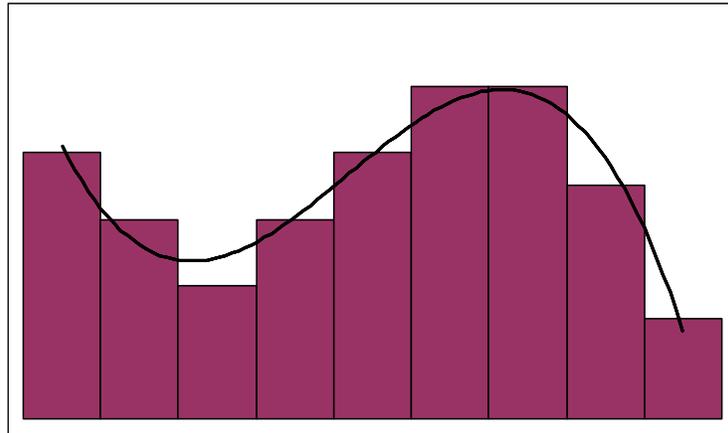


Figure 2.3 Illustration of Gradual vs. Abrupt Interpolator.

Inclusion of Covariates: A final characteristic is the ability of an interpolator to include additional variables other than distance between points in the interpolation process. There are cases where inclusion of other variables in the spatial interpolation model can lead to a better predicted surface because the additional information helps identify highs and lows associated with the spatial process that otherwise would be missed. Another potential benefit is the ability to reduce the uncertainty of the interpolation model’s output if, in fact, the covariates explain some of the spatial variability in the process of interest. Geographic variables of use in explaining spatial variation might include elevation, topography, and land use. When modeling air pollution, other covariates that might improve predictions include other pollutants (that are correlated with the pollutant of interest), meteorology, or emissions information. For example, if ozone concentrations are correlated with temperature, adding temperature to an ozone spatial interpolation model should improve the accuracy of the predictions.

2.3 Specific Methods of Spatial Interpolation

There are various methods available to perform spatial interpolation — some more scientific than others. The general theory behind spatial interpolation is the previously-mentioned Tobler’s First Law of Geography — the closer together two points are in space the more likely those points are to be similar. Techniques range from a “nearest-neighbor” approach that involves identifying the closest measured point to an unmeasured point and assigning the value of the measured point to the unmeasured point, to an IDW approach that involves estimating values by averaging nearby data points with closer points receiving more weight, to a statistical or geostatistical approach such as kriging that involves producing prediction surfaces and accuracy measures for those predictions by evaluating the autocorrelation of measured points. The different methods offer various combinations of the characteristics listed in the previous section and should be applied as appropriate dependent upon the data, the purpose of the analysis, and the planned use of the predicted surface. Since the primary focus of this document is on kriging, this section provides a review of several primary types of interpolation methods followed by a more detailed overview of kriging and some examples of kriging applications. All of the methods discussed are point-based as opposed to areal.

2.3.1 Interpolation Methods Other Than Kriging

Although, as stated above, kriging is the primary focus of this document, it is useful to be aware of some of the other common point-based spatial interpolation methods that have been developed. The following provides an overview and references for further exploration.

Polygonal (Nearest Neighbor): Polygonal or proximal techniques are deterministic methods that utilize no information about the system being analyzed other than the measured data points. They are relatively simple to implement in that all points in an area are set equal to one value, whether it be the value of the nearest measured point, an average of the cell and its surrounding points, or the mode of the cell and its surrounding cells. Thus, polygonal interpolators are local in that predictions are based on values from a subset of the area of interest. When using the value of the nearest point or the mode, all predicted values are equal to a measured value. Thus, based on the technique used to assign values to cells, polygonal methods can be either exact or approximate. The interpolated values generated by these methods will also be restricted to the range of the measured values, i.e., the maximum predicted value will not exceed the maximum measured value and vice versa.

These methods are more formally called by a few names including Thiessen Polygons, Voronoi diagrams or maps, and Delauney triangulation. The output of these methods is a set of contiguous polygons whose values change abruptly at the boundaries between them, which defines these methods as abrupt interpolators as opposed to gradual. For a two-dimensional spatial situation, the polygons are drawn by connecting neighboring points with a line and intersecting that line with a perpendicular line. If the sampled data points are in a rectangular grid, then the resulting polygons will be of equal size and regularly spaced. If the measured data points are irregularly spaced, then the resulting predictive surface will be an irregular lattice of

polygons. This type of method may be appropriate for interpolating data that are more discrete than continuous in nature. Figure 2.4 presents different views of spatial surfaces predicted by a polygonal method.

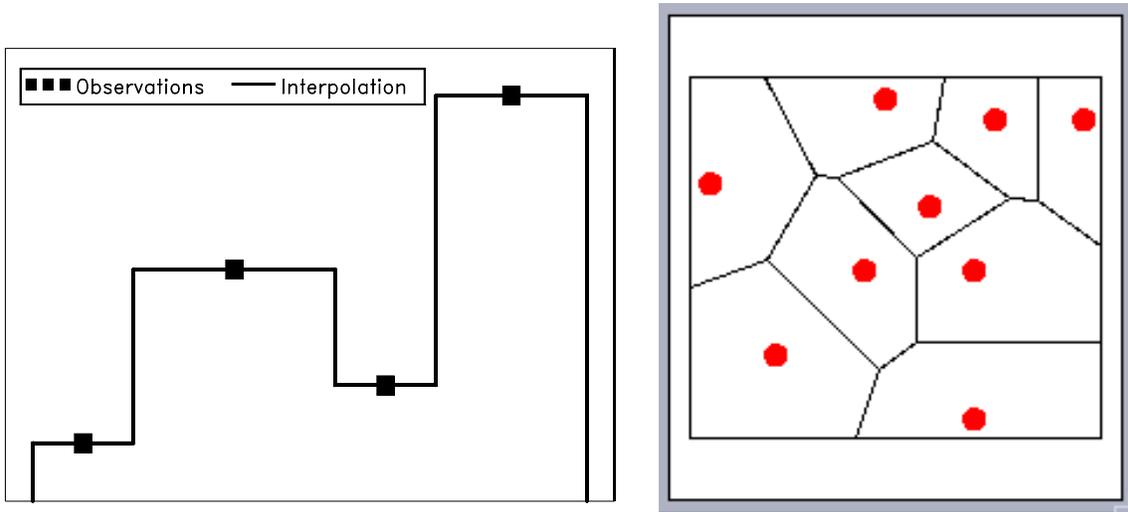


Figure 2.4 Examples of polygonal interpolations in one spatial dimension (left) and two spatial dimensions (right).

Inverse Distance Weighted (IDW): Another set of deterministic interpolation methods are based on mathematical formulas. Estimates are based on averages of the measured values at n known points. IDW is an example of a gradual, exact, mathematical interpolator in which points closer to the measured data points receive more weight in the averaging formula. The formula can be adjusted to change the relative importance of the nearest points as opposed to those that are further away, i.e., the power. Specifying a higher power places more weight on the nearer points while a lower power increases the influence of points that are further away. Using a lower power will result in a smoother interpolated surface being generated. Other variables within the IDW formula that can be altered include the number of measured points that can be considered in the averaging, the zone of influence or search area within which measured data points will be considered, and the direction from which measured points are selected. IDW is a local interpolator except in the case where the zone of influence for all parts of the area of interest includes all measured points.

There are a few areas of concern with the IDW and other non-statistical averaging methods. First, the range of the interpolated values is constrained by the range of the measured values, i.e., no interpolated values will fall outside the observed data range. This means that high or low points of the area under consideration will be missed if they are not sampled. Also, because of the nature of the averaging formula, areas outside of the sampled area will flatten to the mean value.

EPA's AIRNow program uses IDW techniques to produce real-time ozone maps and air quality forecasts. The surfaces are generated using a power of 2 in the denominator of the weighting variable, i.e., $1/r^2$, with r representing distance from a measured point. For additional information about the IDW method and application of the method to air monitoring data, see references [4,5]. The first paper, [4], compares interpolations of PM_{10} concentrations using four different weighting schemes — $1/r$, $1/r^2$, $1/r^3$, and $1/r^4$. As the exponent in the denominator increases, increasing weight is placed on measured values closest to the point being estimated.

Splines: "Spline" interpolation is another type of deterministic interpolation method. Splines are part of a family of exact interpolation models called radial basis functions (RBF). RBF methods include thin-plate spline, regularized and tension spline, and inverse multiquadratic spline. RBF methods seek to minimize the overall curvature of the estimated surface while passing through the measured data points. A mathematical function is utilized that produces a surface with continuous elevation and slope and minimum curvature; thus, these are gradual interpolators. This method performs best when the surface is relatively smooth and a large number of measured data points are available. RBFs will not perform as well when there are large changes in the surface within short distances. RBF interpolation methods are local in that a subset of measured values can be used to generate each prediction, with the actual search area being flexible. Allowing more measured values in the calculation will result in a smoother predicted surface.

Unlike IDW methods, the values predicted by RBFs are not constrained to the range of measured values, i.e., predicted values can be above the maximum or below the minimum measured value. For additional details on spline interpolation and a specific application of the Thin Plate Spline method, see reference [6].

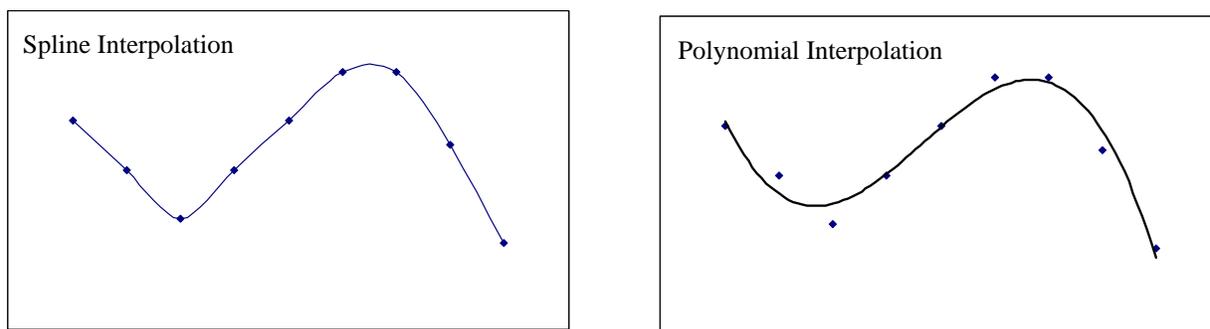


Figure 2.5 Comparison of Spline and Polynomial Interpolation Methods.

Polynomial Interpolation: Polynomial interpolation is an approximate, deterministic interpolation method that fits a mathematical function to the measured points. Options range from a first-order polynomial (linear) to a second-order polynomial (quadratic) to higher-order polynomials. The predictive surface is typically generated by using a least-squares regression fit

that minimizes squared differences between the surface and measured points. Because it is an approximate interpolator, the surface is not constrained to going through the measured points as with RBF interpolation (see Figure 2.5). In addition, because the method generates the best fit (least squares criterion) between the measured points, it is unlikely that the fitted line will run outside the minimum or maximum measured value, except once it goes beyond the measured area (i.e., extrapolation).

There are two types of polynomial interpolation — global and local. Global polynomial interpolation fits a polynomial model to the entire surface based on all measured points. Local polynomial interpolation fits multiple polynomials using subsets of the measured points. Global polynomial interpolation is more appropriate for a surface that varies slowly over the area of interest, while local polynomial interpolation captures more of the short-range variation in addition to the long-range trend. Global polynomial interpolation accounts for bends in the data — one bend with quadratic, two bends with cubic, and so forth. Surfaces that do not display a series of bends, however, such as one that increases, flattens out, and increases again, can be better represented using local polynomial interpolation. Both the global and local methods produce a gradual predicted surface.

2.3.2 Kriging

Geostatistical interpolation methods are stochastic methods, with kriging being the most well-known representative of this category. Kriging methods are gradual, local, and may or may not be exact (perfectly reproduce the measured data). Also, they are not by definition set to constrain the predicted values to the range of the measured values. Similar to the IDW method, kriging calculates weights for measured points in deriving predicted values for unmeasured locations. With kriging, however, those weights are based not only on distance between points, but also the variation between measured points as a function of distance. The kriging process is composed of two parts — analysis of this spatial variation and calculation of predicted values.

Spatial variation is analyzed using variograms, which plot the variance of paired sample measurements as a function of distance between samples. An appropriate parametric model is then typically fitted to the empirical variogram and utilized to calculate distance weights for interpolation. Kriging selects weights so that the estimates are unbiased and the estimation variance is minimized. This process is similar to regression analysis in that a continuous curve is being fitted to the data points in the variogram. Identifying the best model may involve running and evaluating a large number of models, a process made simpler by the geostatistical software packages or features discussed later in this document.

After a suitable variogram model has been selected, kriging creates a continuous surface for the entire study area using weights calculated based on the variogram model and the values and location of the measured points. The analyst has the ability to adjust the distance or number of measured points that are considered in making predictions for each point. A fixed search radius method will consider all measured points within a specified distance of each point being

predicted, while a variable search radius method will utilize a specified number of measured points within varying distances for each prediction.

Because kriging employs a statistical model, there are certain assumptions that must be met. First, it is assumed that the spatial variation is homogenous across the study area and depends only on the distance between measured sites. There are different kriging methods and each has other assumptions that must be met. Simple kriging assumes that there is a known constant mean, that there is no underlying trend, and that all variation is statistical. Ordinary kriging is similar except it assumes that there is an unknown constant mean that must be estimated based on the data. Universal kriging differs from the other two methods in that it assumes that there is a trend in the surface that partly explains the data's variations. This should only be utilized when it is known that there is a trend in the data.

A major benefit of the various forms of kriging (and other stochastic interpolation schemes) is that estimates of the model's prediction uncertainty can be calculated, considered in the analysis, and plotted along with the predicted surface. Such uncertainty information is an important tool in the spatial decision making process.

This section has provided a brief introduction to kriging interpolation. Section 3 of this document presents the details of the ordinary kriging process. Section 4 will review other types of kriging, including universal kriging, kriging with external drift, and co-kriging, in discussing extensions to the ordinary kriging model.

2.3.3 Kriging Applications

The purpose of this section of the document is to demonstrate that the spatial interpolation technique of kriging (in its various forms) is a well-established tool for spatial data analysis and decision making. For many years kriging has been used in various applications, including many in the air quality area and, more generally, in other environmental applications as well.

In a paper published in 1998, James Mulholland and other researchers from the Georgia Institute of Technology and Emory University utilized a universal kriging approach to interpolate ozone levels in the Atlanta region for purposes of investigating the relationship between ozone concentrations and increases in pediatric asthma rates [7]. They used a linear drift function to represent the trend in their data. Kriging has also been utilized to assist in the design of pollution monitoring networks. A 1991 paper by two Johns Hopkins University researchers highlighted the use of kriging to assist in optimizing the location of air pollutant (in this case, sulfur dioxide, nitrogen oxides, particulate matter, and unsaturated hydrocarbons) monitoring stations for a network in Tarragona, Spain [8]. Similarly, a study by Holland, et al., published in 1998 utilized kriging and other methods to evaluate the CASTNet network in the eastern U.S. [9].

Kriging analysis is also applied for other environmental applications such as soil and groundwater modeling. EPA's Preparation of Soil Sampling Protocols document recommends the use of block kriging for "soil mapping, isopleth development, and spatial distribution of soil

and waste properties” [10]. A 1997 report by the Kansas Geological Survey utilized universal kriging to estimate water table elevation levels in Kansas [11]. In this case, universal kriging was applied because of a systematic decrease in the water table toward the east of the state. These researchers used kriging not just for predicting water levels, but for identifying outlier measured values that might contain measurement error. They performed this quality control analysis through cross-validation, i.e., estimating a value for a measured point based on all other data except that point, and then comparing the predicted value to the measured value.

For a 1990 inter-agency National Acid Precipitation Assessment Program report, researchers utilized kriging techniques extensively to evaluate proposed monitoring networks and acid deposition models [12]. One application of kriging in this report was to produce maps of observed and predicted levels of different pollutants for comparison purposes. Another application was to evaluate kriging uncertainty estimates to determine the suitability of different network configurations.

2.4 References

The sources for kriging examples are numerous. In addition to the specific cases listed here, we encourage the reader to consider plentiful sources such as *Mathematical Geology* or the NATO ASI book series for additional references.

- [1] Tobler, Waldo (1970). “A computer movie simulating urban growth in the Detroit region;” *Economic Geography*, 46(2), pp 234-240.
- [2] Ripley, B. D. (1981). Spatial Statistics. John Wiley & Sons, New York, New York.
- [3] Goodchild, Michael, University of California at Santa Barbara; Lecture on Spatial Interpolation; <http://www.geog.ucsb.edu/~good/176b/a15.html>
- [4] Husar, Rudolf, and Falke, Stefan, (1996a). “Uncertainty in the Spatial Interpolation of PM10 Monitoring Data in Southern California.” CAPITA Reports. Center for Air Pollution Impact and Trend Analysis (CAPITA), Washington University. <http://capita.wustl.edu/CAPITA/CapitaReports/CaInterp/CaINTERP.HTML>
- [5] Husar, Rudolf, and Falke, Stefan, (1996b). “Uncertainty in the Spatial Interpolation of Ozone Monitoring Data.” CAPITA Reports. Center for Air Pollution Impact and Trend Analysis (CAPITA), Washington University. <http://capita.wustl.edu/CAPITA/CapitaReports/O3Interp/O3INTERP.HTML>
- [6] Ionescu, Anda, Candau, Yves, Mayer, Eric, and Colda, Iolanda, (1998). “Analytical Determination and Classification of Pollutant Concentration Fields using Air Pollution Monitoring Network Data: Methodology and Application in the Paris area During Episodes with Peak Nitrogen Dioxide Levels.” International Conference on Air Pollution Modelling and Simulation; Champs sur Marne, France.

- [7] Mulholland, James, et al. (1998). "Temporal and Spatial Distributions of Ozone in Atlanta: Regulatory and Epidemiologic Implications." James Mulholland, Andre Butler, James Wilkinson, and Armistead Russell (Georgia Institute of Technology) and Paige Tolbert (Emory University). *Journal of the Air and Waste Management Association*, Vol. 48, May.
- [8] Trujillo-Ventura, Arturo, and Ellis, J. Hugh, (Johns Hopkins University) (1991). "Multiobjective Air-Pollution Monitoring Network Design." *Atmospheric Environment*, Vol. 25A.
- [9] Holland, David M., et al. (1998). "Spatial Prediction of Sulfur Dioxide in the Eastern United States." David M. Holland and Lawrence Cox (EPA/ORD), Nancy Saltzman (National Institute of Statistical Sciences) and Douglas Nychka (North Carolina State University). *Geostatistics for Environmental Applications*, Vol. VII, pp 65-76.
- [10] U.S. EPA (1992). "Preparation of Soil Sampling Protocols: Sampling Techniques and Strategies." Report No. EPA/600/R-92/128, Environmental Monitoring Systems Laboratory, Office of Research and Development.
- [11] Miller, Richard, et al. (1997). "Acquisition Activity, Statistical Quality Control, and Spatial Quality Control for 1997 Annual Water Level Data Acquired by the Kansas Geological Survey." Richard Miller, John Davis, and Ricardo Olea, Kansas Geological Survey, <http://magellan.kgs.ukans.edu/WaterLevels/CD/Reports/OFR9733/rep00.htm>.
- [12] Dennis, Robin L., et al., (1990). "Acidic Deposition: State of Science and Technology Report 5, Evaluation of Regional Acidic Deposition Models and Selected Applications of RADM," National Acid Precipitation Assessment Program.

3.0 THE ORDINARY KRIGING MODEL

As discussed in Section 2, kriging is a geostatistical spatial (and temporal, potentially) interpolation method that derives predicted values based on the distance between points in space and the variation between measurements as a function of distance. Ordinary kriging is a version of kriging that assumes the mean is constant but unknown across the spatial domain of interest. This section will describe some of the theory behind kriging and walk the reader through some of the mechanics of an ordinary kriging exercise.

Before beginning, it is important to recognize that the real-world is made up of four dimensions; three spatial dimensions and a time dimension. Furthermore, real-world spatial correlations often exhibit directional behavior (anisotropy) and/or lack of stationarity. For the purposes of illustration, this section and much of this document focus on a simpler representation of the real-world; namely, two-dimensional space, a fixed point in time, and spatial correlations that do not depend on direction (isotropy). More complex spatial interpolation approaches that address issues such as temporal dynamics and anisotropy, as well as other issues such as trends, covariates, and multivariate modeling, are discussed in Section 4.

Let Z represent the spatial process of interest (e.g., ozone 8-hour average concentrations or $PM_{2.5}$ annual average concentrations). Let x represent some measure of west-to-east relative location such as longitude, and let y represent some measure of north-to-south relative location such as latitude. (Strictly speaking, kriging is defined on coordinate systems for which distances are linear. Except as an approximation, this is not true for latitude and longitude. We address this concept in more detail in Section 3.1) Thus, $Z(x,y)$ represents the realization of the process of interest at the point in space (x,y) . Ordinary kriging provides a statistical model for the process of interest, $Z(x,y)$, at all points in space, (x,y) , within some well-defined spatial domain or region (e.g., an EPA Region, a State, some arbitrary rectangle or polygon, etc.). Ordinary kriging defines the process as follows:

$$Z(x,y) = u + e(x,y) , \quad (\text{Eq. 1})$$

where

- u = the overall, large-scale mean of the process across the spatial domain; and
- $e(x,y)$ = the small-scale random fluctuation of the process within the spatial domain.

Conceptually, a model like Equation (1) divides the spatial process into two components: a large-scale mean trend component, u , and a small-scale random fluctuation component, $e(x,y)$. Unlike many statistical modeling exercises, ordinary kriging places very little emphasis on the large-scale mean trend component, choosing to treat this component of the model as a simple constant, u , that does not depend on location, (x,y) . An ordinary kriging modeling exercise is instead focused on carefully modeling the structure and behavior of the small-scale random fluctuation component, $e(x,y)$. Variations and extensions of the ordinary kriging model are

presented and discussed in Section 4, including approaches that place more emphasis on modeling the structure of the large-scale mean component.

In practice, ordinary kriging, for the purpose of spatial interpolation, might be implemented via the following five steps:

- Step 1: Build the analysis data set
- Step 2: Summarize and understand the data
- Step 3: Conduct a variogram analysis
- Step 4: Apply spatial prediction and uncertainty formulas
- Step 5: Evaluate model performance through diagnostics

The first two steps are necessary and important before embarking on an ordinary kriging modeling or other spatial interpolation exercise. The purpose of the variogram analysis in the third step is to model the structure and behavior of the small-scale random process, $e(x,y)$, of model Equation (1). These results are then used as input for the fourth step, which applies ordinary kriging equations for spatial prediction and uncertainty estimation to generate a spatially-interpolated surface of the process $Z(x,y)$ and, likewise, a surface of the model's prediction uncertainty. In the fifth and final step, formal and informal diagnostics may be checked as part of a quality assurance/quality control (QA/QC) assessment of the model's performance. (See also Section 5 for a broader perspective on model evaluation.) Depending on the level of stakeholder satisfaction with the initial results of this step-by-step process, it may be necessary to re-visit and possibly modify some or all of these steps. The following sections discuss them in turn.

3.1 Step 1: Build the Analysis Data Set

The most basic data set required for any two-dimensional spatial interpolation exercise consists of three variables: the process of interest (Z), a measure of location in the first spatial dimension (x), and a measure of location in the second spatial dimension (y). The variables x and y are often longitude and latitude, respectively. Table 3.1 provides an abbreviated example for a data set of annual average $PM_{2.5}$ concentrations that will be used in a case study example presented throughout this section and other parts of this document. There often may be one or more data management or QA/QC issues to address even in those cases that are as seemingly straightforward as that of Table 3.1. Several common issues are discussed below.

Table 3.1 Example analysis data set for two-dimensional spatial interpolation

Z = PM_{2.5} annual average (: g/m³)	x = longitude	y = latitude
12.0759	-81.3456	28.5508
11.8871	-81.3631	28.5994
10.9991	-81.3100	28.7456
10.4941	-82.7000	28.9806
11.0013	-82.1008	29.1703
.	.	.
.	.	.
.	.	.
.	.	.
15.8305	-83.2092	42.2283
14.9955	-89.0894	42.2672
14.8023	-85.5419	42.2781

Temporal Averaging and Completeness Criteria

The basic assumption of this document is that the purpose of the modeling exercise, be it ordinary kriging or otherwise, is spatial interpolation. No attempt is made to address any temporal dynamics associated with the problem. With that perspective in mind, it remains important to define and understand the temporal scale of the problem. In general, this scale should be chosen to match the management decision needs. For example, the data set to be used for analysis will be different if the goal is to spatially interpolate a PM_{2.5} 24-hour average concentration versus a PM_{2.5} annual arithmetic mean concentration. Furthermore, the results of the spatial interpolation exercise and their associated interpretation may vary dramatically depending on the temporal scale of the data.

Often it is the case that an initial data set will require pre-processing to generate analysis data that better match the spatial process of interest with respect to temporal scale. This can be accomplished via some sort of temporal averaging of the initial data. For example, a data set of 1-hour ozone concentrations might be averaged in some manner to yield 8-hour concentrations for analysis. Note that some loss of information occurs when temporally averaging data, as opposed to modeling the original data and somehow taking advantage of the information provided by its inherent temporal variation. For simplicity, this section and most of the document assume that the simpler approach of temporal averaging will typically be employed to model the spatial process at the direct temporal scale of interest. This document does not necessarily advocate such an approach, but recognizes its practicality. The discussion that

follows highlights some of the issues that arise due to temporal averaging. See Section 4 for more discussion on incorporating temporal dynamics into a spatial interpolation modeling exercise.

In many instances, the approach to temporal averaging may be dictated by some form of guidance or regulation, or by the management decision at hand. In other instances, the approach will be determined in a more subjective manner, preferably through the informed consensus of engaged, knowledgeable stakeholders. For the $PM_{2.5}$ case study example presented throughout this section, the annual average analysis data set summarized in Table 3.1 was generated from an initial data set of 24-hour averages. (As stated above, this document does not necessarily advocate data pre-processing via temporal averaging, but this approach was taken for the $PM_{2.5}$ case study example for illustrative purposes.) Site-specific annual averages were calculated as the arithmetic mean of the 24-hour average data collected at the given site throughout calendar year 2000. The majority of the sites included in the case study example data set provided data consistent with a sampling schedule of once every three days.

When pre-processing an initial data set via temporal averaging, some consideration must be given to the temporal completeness of the resulting average. If the initial data used to generate an endpoint such as an annual average are somehow temporally incomplete, the calculated endpoint may be biased in some manner. For example, many air pollutants exhibit significant seasonal fluctuations, so an annual average estimated from only a single quarter's worth of data (e.g., January through March) may not be representative of the true year-long average. It may, therefore, be important to define certain temporal completeness criteria that determine whether the temporally-averaged data at specific locations should be included in the analysis data set. Ultimately, a balance must be struck between a more (less) strict temporal completeness requirement and a more (less) sparse spatial analysis data set. For the $PM_{2.5}$ case study example presented throughout this section, it was decided that all four quarters throughout the year had to provide at least 75 percent of the expected number of 24-hour average observations according to a given site's known or estimated sample frequency. For example, a sampling schedule of every three days leads to approximately 120 samples per year or 30 per quarter, which results in a 75 percent completeness requirement of at least 23 observations in each of the year's four quarters.

Another issue to consider when pre-processing the data via temporal averaging is the impact of the averaging on the assumed variation (variability + uncertainty) structure of the resulting data points. Ordinary kriging and many other statistical spatial interpolation procedures assume a common variogram model (and, hence, variance) across all the analysis data points. An initial data set may indeed satisfy this basic assumption if the data have been generated via consistent monitoring technologies with common QA/QC oversight. However, any temporally-averaged data derived from the initial data set might not strictly satisfy the common variance assumption if the sample frequencies of different sites vary. For example, all else being equal, it is expected that a site-specific annual average calculated from 24-hour average samples collected every three days will be more precisely estimated than one calculated from samples

collected every twelve days. While this issue merits some consideration, it is beyond the scope of this document and will not be addressed further.

Method Detection Limits

For the purposes of this document, a method detection limit (MDL) is defined as that threshold level below which a laboratory analysis does not report an air pollution concentration data point, but instead censors the output and reports only the MDL and the fact that the concentration in question falls below the MDL. This censoring occurs because a laboratory cannot state with confidence that a given observation is significantly different from zero. In practice, a laboratory will typically use data generated as part of its routine quality assurance/quality control (QA/QC) procedures to calculate a MDL as some multiple of a standard deviation representing analytical uncertainty. As of the writing of this document, there remains some variation in exactly how laboratories calculate MDLs and how they report below-MDL data. Nonetheless, the issue of censored data or below-MDL data may need to be addressed in some cases before beginning an ordinary kriging exercise, because such data may induce bias in the overall analysis. (Further details on this topic are beyond the scope of this document.)

The most common data analysis approach to handling below-MDL data is to impute a value for the censored data then treat the imputed values as real data. A typical imputation technique is to replace a censored data point by one-half of the reported MDL value for that observation. More sophisticated techniques include estimating the parameters of an assumed statistical distribution using the available uncensored data, then simulating values for the censored data according to the estimated parameters. Regardless of the imputation (or other) approach to addressing below-MDL data, two general recommendations can be made. First, consideration should be given to the level of MDL beyond which it may not be appropriate to include a given site's data or, more importantly, beyond which it may not be appropriate to proceed with the intended ordinary kriging exercise. Second, some sort of sensitivity analysis should be conducted as part of the overall modeling exercise to assess the degree to which the method, or alternative methods, of addressing below-MDL data impact the results of the analysis.

Spatial Coordinate System

Air pollution data with spatial context, monitoring data or otherwise, are commonly reported in a longitude by latitude coordinate system in units of degrees, minutes, and seconds. Such a coordinate system accounts for the curvature of the earth's surface. Meanwhile, statistical spatial interpolation techniques such as ordinary kriging typically assume some sort of spatial correlation structure defined with respect to the linear distance between two points in space. Therefore, it is not strictly accurate to calculate the distance between two points in a longitude by latitude coordinate system using a simple linear distance function. In many cases, if the spatial domain under study is small enough in geographic extent, the error in such a calculation will be minimal. Perhaps counter to intuition, this is not due to an equality of degrees latitude to degrees longitude at small scale, but rather due to the fact that for small scale the distance represented by one degree longitude is approximately equal at all points in the space. This is not necessarily true

at large scales. In other cases, it may be more appropriate to take the time to convert from a latitude by longitude coordinate system to a relative coordinate system of west-to-east and north-to-south linear distance (e.g., kilometers, miles, etc.) from some arbitrary reference point (e.g., the first data point in the analysis data set).

When working in (or starting from) a longitude by latitude coordinate system, the following set of formulas may prove useful for calculating approximate linear distances between observations:

$$a = \left\{ \sin \left[\frac{\pi}{360} (y_2 - y_1) \right] \right\}^2 + \cos \left(\frac{\pi}{180} y_1 \right) * \cos \left(\frac{\pi}{180} y_2 \right) * \left\{ \sin \left[\frac{\pi}{360} (x_2 - x_1) \right] \right\}^2$$

$$c = 2 * \arcsin[\min(1, \text{sqrt}(a))]$$

$$\text{distance} = R * c$$

where $\mathbf{B} = \pi . 3.14159$, $x_i =$ longitude of location i ($i=1,2$), $y_i =$ latitude of location i ($i=1,2$), \cos represents the cosine function, and \sin represents the sine function. R represents the radius of the earth and is approximately equal to 6,367 kilometers or 3,956 miles. The distance results from the formula are more accurate than relying on longitude by latitude coordinates alone because the formula accounts for the curvature of the earth. It can be used as a tool, for example, to ascertain constant longitude by latitude degree distances throughout the spatial domain under study. To the extent that those distances vary across the domain, some modeling error will be introduced due to ordinary kriging using the original longitude by latitude coordinate system. The larger the spatial domain, the more this becomes an issue. (Note that the above formula is still only an approximation because it treats the earth as if it were shaped like a perfectly round ball. Additional formulas for estimating such distances can be found at: <http://www.census.gov/cgi-bin/geo/gisfaq?Q5.1>)

3.2 Step 2: Summarize and Understand the Data

Once the spatial analysis data set has been built, it is important to generate some initial summaries of the available data prior to analysis in order to obtain a better understanding of its empirical structure and behavior. Reasonable summaries include, but are not limited to, graphical information systems (GIS) maps of the spatial domain and available data locations, a histogram of the overall data distribution, and summary statistics such as the data's mean, standard deviation, and various percentiles (e.g., minimum, median, maximum, etc.). Moving-window statistics (i.e., summarizing different sub-regions of the spatial domain) are particularly helpful for assessing the ordinary kriging stationarity assumptions (i.e., constant mean and constant variogram model across the spatial domain).

Consider the $PM_{2.5}$ annual average case study example discussed throughout this section. Defining a rectangle covering much of the eastern United States as the spatial domain of interest, Figure 3.1 displays monitoring locations within this domain that provide data for use in an ordinary kriging exercise. For perspective, Figure 3.1 overlays state boundary lines. A total of 363 locations provide data. While the total number of monitoring locations is a reasonably high number and most of the domain exhibits some monitoring, the spatial distribution of available data for analysis is far from evenly distributed across the domain. In fact, as expected, much of the available data tend to come from clusters of monitors within urban centers or areas of generally higher population density.

If information about the entire spatial domain is equally important, and if no prior knowledge about the spatial surface is available, then for spatial interpolation via ordinary kriging or other methods, it can be desirable to have a uniform, dense spatial grid of data for the purposes of prediction. In reality, the data analyst is constrained to whatever useful data might be available. In the instance of the $PM_{2.5}$ case study example, the FRM monitoring network was not necessarily designed for the express purpose of conducting spatial interpolation modeling exercises. As such, to the degree that $PM_{2.5}$ annual average concentrations correlate with population density, ordinary kriging results from these data will be influenced by the population-oriented siting tendencies of the network. To address this issue, other modeling approaches that account for explanatory factors such as population density (or terrain, emission sources, meteorology, etc.) might be considered in order to reduce the uncertainty in the results that is due to the data collection design. For the most part, this issue is beyond the scope of this document. It is, however, addressed in further detail within Section 4.

Continuing with the $PM_{2.5}$ annual average case study example, Figure 3.2 divides the spatial domain along longitudinal and latitudinal lines into nine sub-regions of roughly equal area. The figure indicates the number of monitoring sites falling within each sub-region. The purpose of this type of data partition (i.e., moving windows) is to provide further summaries of the data overall and within each spatial sub-region in order to support the ordinary kriging model's two stationarity (homogeneity) assumptions; namely, that the mean is constant across the spatial domain and that the variogram is the same model for all data across the domain. In general, such analyses provide the data analyst with a better understanding of the data.

Corresponding to the partition indicated by Figure 3.2, Table 3.2 and Figure 3.3 provide similar information in tabular and graphical format, respectively, on the distribution of the case study data both overall and specific to each spatial sub-region. While no strong trends appear, sites in the western third of the region appear to exhibit somewhat lower concentration distributions, while sites in the central, northern, northeastern, and eastern sub-regions appear to exhibit somewhat higher concentration distributions. To the degree that significant large-scale spatial trends and/or non-stationary covariances exist, a more sophisticated model than ordinary kriging may be warranted (see Section 3.5 and Section 4 for further discussion).

Table 3.2 Distribution (summary statistics) of annual average $PM_{2.5}$ values by spatial sub-region within the domain of interest for the case study example

Block	Number of Sites	Spatial Variation in Annual $PM_{2.5}$ Concentrations (: g/m^3)						
		Mean	SD	Min	Q1	Median	Q3	Max
ALL	363	14.85	2.43	9.71	13.15	15.10	16.43	22.28
NW	38	11.38	1.03	9.82	10.76	11.07	11.92	13.45
N	88	15.82	1.63	12.01	14.95	15.75	16.75	20.64
NE	50	16.22	1.84	10.39	15.33	15.97	17.61	20.87
W	25	12.40	1.82	9.71	10.86	12.29	13.53	15.83
C	32	16.63	2.20	13.29	15.12	16.26	17.86	22.28
E	59	16.10	1.68	12.44	14.87	16.11	17.22	21.01
SW	29	12.95	0.87	11.24	12.47	12.91	13.38	15.05
S	21	14.40	1.68	12.22	13.47	14.03	14.93	19.26
SE	21	13.53	2.37	10.49	11.86	13.37	15.34	18.31

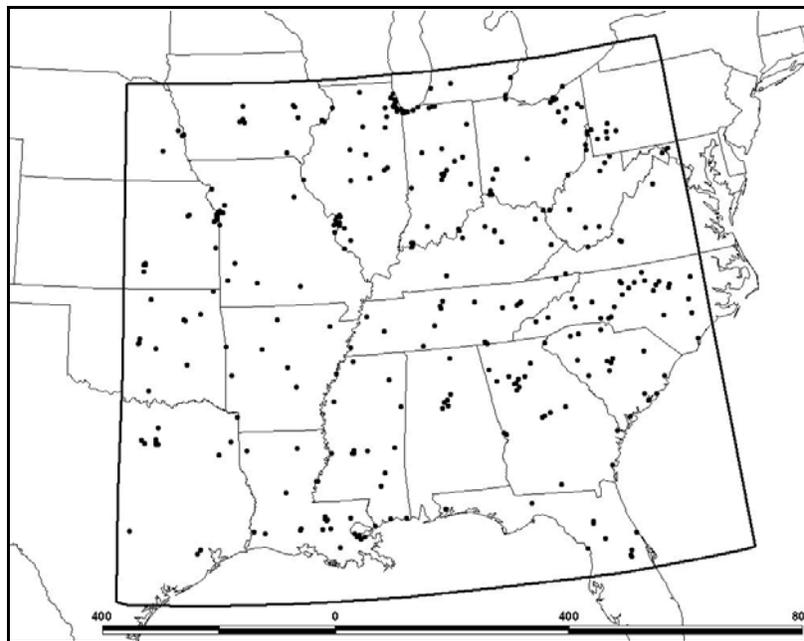


Figure 3.1 Spatial distribution of $PM_{2.5}$ monitoring sites within domain of interest for case study example.

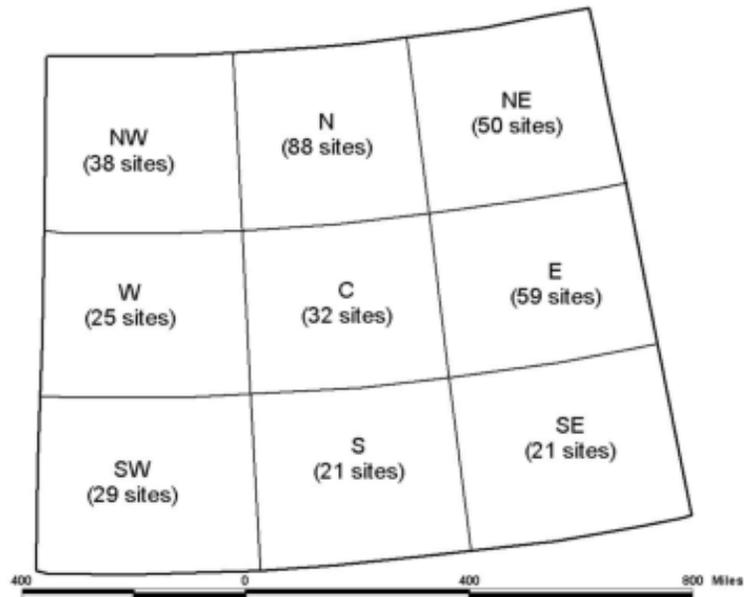


Figure 3.2 Partitioning of domain of interest for case study example.

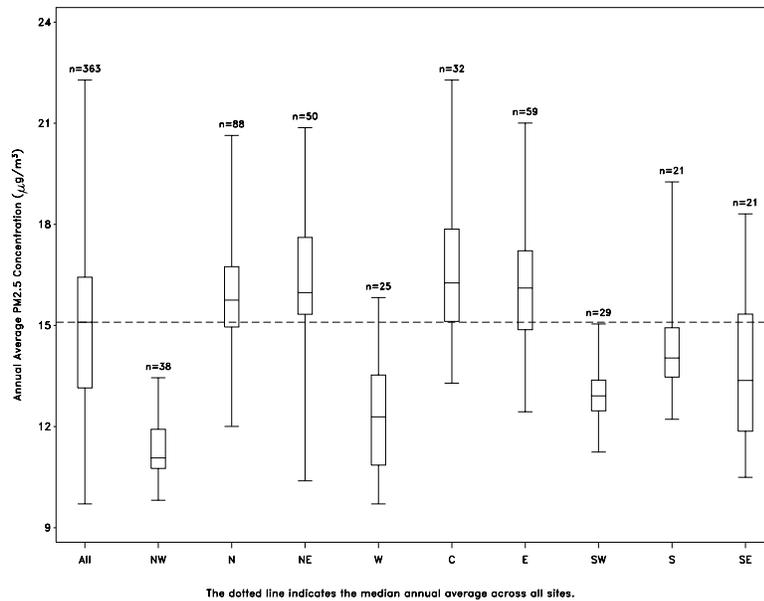


Figure 3.3 Distribution (side-by-side box plots) of annual average $PM_{2.5}$ values by spatial sub-region within domain of interest for case study example.

3.3 Step 3: Conduct a Variogram Analysis

Once the analysis data set has been built and explored to gain some basic understanding, a formal ordinary kriging modeling exercise can begin. The exercise begins by conducting a variogram analysis. This analysis typically consists of first generating an empirical variogram and then fitting a parametric model that adequately captures the structure of the empirical variogram. [Note that non-parametric models might be entertained as well, but they are beyond the scope of this document. See Cressie (1993) for further discussion of non-parametric estimates.] Ultimately, the estimated parameters of the variogram will be input into for the ordinary kriging spatial prediction and uncertainty formulas from which a spatially interpolated surface is generated. In other words, once you have chosen a satisfactory variogram model, you then use that model as an input to the actual kriging process as described in Section 3.4. To illustrate the concept of variogram analysis and to aid the reader in understanding the process, the following discussion presents an ordinary kriging exercise conducted using the annual PM_{2.5} case study example data discussed previously.

The first step in generating an empirical variogram is to partition the data according to the distance between distinct pairs of observation locations. There appears to be little guidance on the optimal size of the partitions (or bins). Cressie (1993) recommends that the bins be as small as possible to retain spatial resolution, and yet large enough that the empirical variogram estimate is stable. Journel and Huijbregts (1978) recommend that the number of pairs in each bin be at least 30. An alternative recommendation is based on the similarity of the process of generating an empirical variogram to the process of generating a histogram. This recommendation is to use a number of bins approximately equal to the square root of the number of available data points. For an example of 400 total pairs, this approach would yield 20 bins with 20 pairs per bin. It represents a compromise between the number of bins and the number of pairs in each bin. Finally, certain software packages provide a default number of bins. For example, the empirical variogram window in S-Plus's spatial module defaults to 20 bins, though this can be changed easily.

A completely different choice would be to use no bins at all. This approach is referred to as a variogram cloud and plots the squared difference of two observations on the y-axis against the distance between the same two observations on the x-axis. This is repeated for all possible pairs of observations in the data set. The resulting scatter plot (or cloud) of observations will tend to have the general shape of the appropriate variogram model (see below). As a result, the variogram cloud may often be useful to an analyst who wishes to predetermine what model family is most appropriate for the ordinary kriging process.

Once the bin structure for distances has been determined, apply the following formula to the contents of each bin to generate a function that changes across bins:

$$\gamma(h) = \text{average} [(Z(x_1,y_1) - Z(x_2,y_2))^2], \quad (\text{Eq. 2})$$

where γ (i.e., gamma) is the resulting function dependent on bin, h is the bin number, Z is the response of interest (e.g., a $PM_{2.5}$ annual average), and (x_1, y_1) and (x_2, y_2) describe locations in space in terms of the spatial coordinate system of choice (e.g., longitude and latitude). The average is taken over all the pairs of points whose distance falls into bin number “ h ”. The bins can be defined in terms of the distances they include. For example, a set of bins for a hypothetical analysis might represent 0-10 kilometers (km) with midpoint 5 km, 10-20 km with midpoint 15 km, 20-30 km with midpoint 25 km, and so forth. Plotting this set of averages against the midpoint distance of the associated bin provides a graph of the empirical variogram (i.e., the value of γ as a function of distance). See Figure 3.4 for an example of such a graph.

Next, consider a suite of parametric models to find the best fitting estimate for the empirical variogram. Though other parameter fitting methods are possible, in general, the most appealing and common approaches are least-squares-minimization via the objective function and trial-and-error via visual inspection. For the former, one numerically computes the set of parameters that minimizes the objective function. The objective function is defined as the sum of the squared residuals (obtained when the proposed parametric model is subtracted from the empirical variogram), and provides a measure of the fit of the proposed model to the empirical variogram [see Cressie (1993) for further details]. For the latter, one chooses the set of parameters that gives the best visual fit. Occasionally, it may be necessary to use a combination of these two methods. For instance, one might choose the type of model from the options discussed below and then compute the variogram parameters that provide the best least squares fit for that model family. This will be discussed in more detail later in this document.

A number of parametric models suit the variogram modeling process. [Note that the formal definition of the variogram requires that certain statistical properties hold. All of the models discussed here fulfill these properties. These required properties also make non-parametric variograms more difficult, but not impossible, to fit. See Cressie (1993) for further details on this topic.] The most common models used in the variogram modeling process are: linear, spherical, exponential, rational quadratic, wave (or hole-effect), power, and Gaussian. Cressie (1993) discusses all of these models except the Gaussian in some detail, providing mathematical formulas and further details. Of these seven variogram models, three are used most commonly: spherical, exponential, and Gaussian (see Figure 3.4b). Exponential models fit best when the spatial autocorrelation decreases exponentially with increasing distance. Spherical models provide a better fit when spatial autocorrelation decreases to a point after which it becomes zero. Gaussian variograms tend to have an “S” shape, with a gradual upward slope at short distances, followed by a sharper upward slope at middle distances and, finally, by another gradual upward slope at long distances. A more complicated approach uses nested models and, in essence, sums two or more of the above types of models together to determine the final variogram model. [See Cressie (1993) for more details.] The final model selected will represent the spatial correlation of the measured data and guide the assignment of the kriging weights in the second part of the process (discussed below in Section 3.4).

The three most common models also share defining parameters: nugget, sill, and range. See Figure 3.4a for a visual depiction of these parameters. Additional technical details about

these terms can be found in Cressie (1993) and Ripley (1981). Briefly, the nugget parameter represents the variation due to measurement error and micro scale variation (i.e., variation at very short distances). It is the point at which the variogram model appears to intercept the y-axis. Theoretically, in a measurement-free environment, the variogram value could be zero at zero separation distance; however, at very small separation distances variograms often appear to have positive variances because of measurement error or spatial sources of variation that are smaller than the distance between sampling locations. Put another way, the variogram will always be defined to be equal to zero at zero separation distance, unless the model formally includes measurement error and/or micro scale variation. (Note that including a non-zero nugget parameter in a kriging model results in a spatially interpolated surface that does not necessarily pass directly through the observations.) The range parameter represents the spatial distance after which data are effectively no longer spatially correlated, or autocorrelated. The sill parameter describes the maximum value of the variogram minus any nugget effect. In some sense, the sill can be interpreted as the maximum variance associated with the surface itself, after the effects of measurement error have been removed. Some authors refer to the maximum variance, when the measurement error is included, as the total sill. These parameters will define how spatial predictions are calculated in the next step of the kriging process (discussed below in Section 3.4).

Several important issues to consider when conducting a variogram analysis are discussed in the following sub-sections. In particular, we consider manually estimating variogram parameters, software, interpreting results, and the effects of temporal averaging.

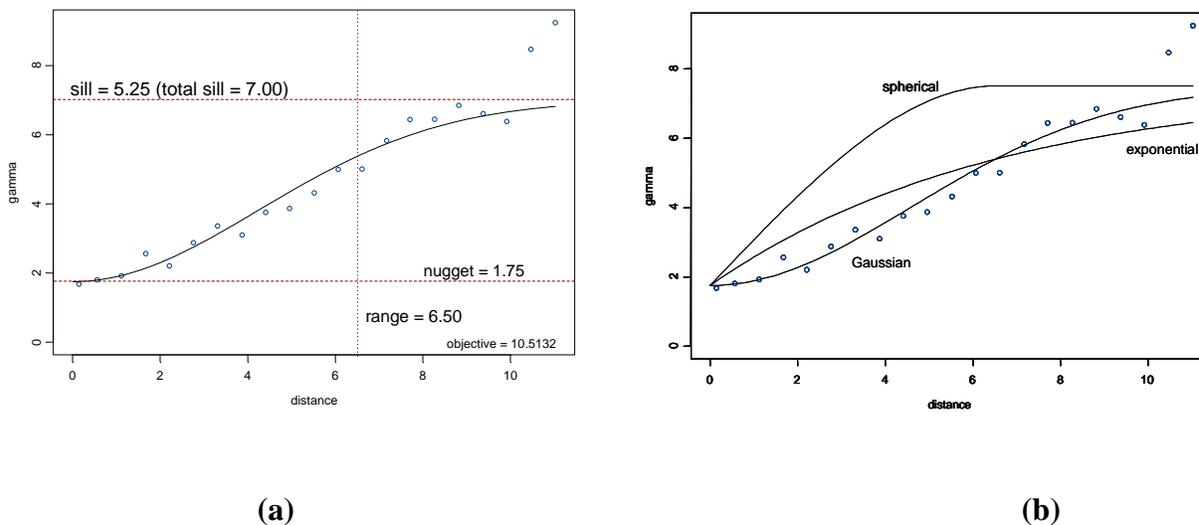


Figure 3.4 Direct comparison of three model families.

Manually Estimating Variogram Parameters

Consider the three most commonly used parametric variogram models: spherical, exponential, and Gaussian. When attempting to model one of these types of variograms via trial-and-error visual inspection, there are some guidelines that can be applied when starting the process. In general, when visually estimating variogram parameters, it is important to note that not all empirical variogram points are equally important when it comes to developing model variograms. Short distances are most important since they have the greatest impact on prediction and prediction errors. Long distances may be generated with fewer observation pairs due to the geometry of the spatial sampling locations, and, therefore, the variogram model fit at such distances may be more uncertain as a result. In addition, a large number of values at or near the sill will tend to dominate any automatic fitting algorithms (and, thus, the objective function). Thus, one might consider trimming some of the empirical values that are beyond the range out of the variogram modeling process.

Model Family. When attempting to determine what model to apply to an empirical variogram, one can get information by a visual inspection of the shape of the variogram (see Figure 3.4b). For example, as stated previously, Gaussian variograms tend to have an “S” shape. That is, they exhibit a gradual upward slope from distance zero, followed by a sharper upward slope toward the middle of the variogram, and finally another gradual upward slope at the end of the variogram. On the other hand, both the spherical and exponential variograms start sloping upward more sharply at distance zero. Of the two, the exponential variogram tends to have more gradual behavior. The exponential curve tends to be more sharp than the Gaussian and spherical models at the beginning. The exponential curve also tends to become shallow more gradually than the spherical variogram, which tends to have the same slope until it nears the sill at which point it tends to become nearly flat.

Nugget. If co-located data are available in a data set, then an estimate of the measurement error can be obtained. In this case, the average of observations at a single location are used to compute the value applied to the computation of the empirical variogram, but the average of the sum of the squared deviations from those co-located observations is an estimate of the variance of the measurement error. The estimate can be used as a lower bound for the nugget, or even as an estimate of the nugget itself, if one assumes that there is no micro scale variation. As an example, consider the PM_{2.5} annual average data set used throughout this document. In this case, the estimated variance of only the co-located observations in the annual data is 0.18772 [g/m^3]². However, upon using S-Plus software to estimate the nugget that minimizes the objective function, the value is an order of magnitude larger. This suggests that, for this example, there is either an important micro scale variation term in this spatial analysis or an inappropriate numerical optimization. In situations like this, where one value is a great deal smaller than the other, it is recommended that one take the approach of using the average of the two values to estimate the nugget. Choosing one extreme or the other will have implications for both the associated kriging surface and on the estimates of the standard deviations. As a result, the reader is cautioned to take care to examine the data and the results before choosing one extreme or the other for the nugget.

If the values are approximately the same, this suggests that both are estimating the true nugget value with error and, thus, in this case it may be appropriate to use the average of the two values as a revised estimate of the nugget. As a final note on estimating the nugget, another, perhaps more crude, estimate is the straight line extrapolation back from the minimum bin distance of the empirical variogram to the y-axis.

Sill. For any data set, a reasonable estimate of the sill can be obtained by inspecting the empirical variogram at large values of the distance. In practice, the empirical variogram tends to flatten out at large distances, indicating an empirical sill. Thus, it may be appropriate to use either the largest computed empirical variogram value, or the value of the empirical variogram associated with the longest distance, to compute an estimate of the sill. To gain a true estimate of the sill, it is then necessary to adjust for the nugget by subtracting an estimate of the nugget. Both of these suggestions are based on a single value. Another suggestion would be to use the average of the empirical variogram values that make up the plateau at long distances. For example, one might use the average of the last seven empirical variogram values shown in Figure 3.4a.

Range. An empirical estimate of the range is less obvious. One might compute the optimal (least-squares) estimate of the range given any estimates of the sill and nugget computed as discussed above. In other words, fix the values of the sill and the nugget, then let software such as S-Plus provide a solution for the range parameter given the assumed values of the sill and nugget. If one is confident in the sill and nugget estimates used in this manner, choosing a numerically optimal range might be a way to improve one's overall estimate of the variogram model. If one chooses to use empirical, or inspection-based, parameters in modeling the variogram, several iterations may be necessary to refine the parameters so that the visual match to the empirical variogram is improved, or so that the objective function is minimized. Another way to empirically estimate the range is to consider the plateau of empirical values that occur at long distances, as discussed above. By definition, the range is the distance at which the data are no longer autocorrelated. Another way to think about this is the distance in the variogram after which no more variance accumulates. On the variogram, this is the distance at which the plateau starts. Thus, identifying such a plateau in an empirical variogram might provide empirical information about the range as well as the sill.

Software Considerations

Various software packages, including Surfer, GMS, and others, will compute the empirical variogram and assist in the modeling process. Among these, two statistical packages are noteworthy and were used in the development of this document: SAS and S-Plus. Both packages have different strengths and weaknesses. SAS, in general, appears to be more flexible than S-Plus, but it requires understanding of SAS programming. SAS does not provide much in the way of automatic defaults or computation of parameters, though standard parameter estimation functions can be used, in conjunction with the variogram functionality, to model the process.

S-Plus, on the other hand, with the addition of its spatial module, has been designed to minimize the level of technical knowledge required on the part of the user. S-Plus provides a series of menu-accessed functions to process a data set. In many cases, these functions will cause parameters to default to functional values if they are not specified by the user. For instance, when computing the empirical variogram, S-Plus defaults to 20 bins. In addition, when investigating variogram models, S-Plus's spatial module provides functionality to estimate the necessary parameters for a given model of interest. For example, when considering a spherical, exponential, or Gaussian model, S-Plus will provide estimates for the nugget, sill, and range upon request. S-Plus also provides the value of the objective function, the numeric measure of model fit, so that the user can consider model fits using both visual inspection and numerical comparison. Using S-Plus, Figure 3.5 plots a variogram model (solid curve) fit to the associated empirical variogram (circles). In determining if the model fit is appropriate, the data analyst should consider both how well the proposed model line matches the points of the empirical variogram as well as the magnitude of the objective value (shown in the lower right corner of Figure 3.5).

The reader is cautioned that though defined properly in its documentation, S-Plus uses "range" interchangeably for the "a" parameter used to define the spherical, exponential, and Gaussian variogram functions. It is important to understand that "a" is identically equal to "range" only for the spherical variogram definition. In general, S-Plus's definition will be used in the examples in this text. For further understanding of the "a" parameter and its relationship to the "range", compare the on-line S-Plus documentation for the various variogram models to the variogram models presented in Cressie (1993).

It should also be noted, looking forward to the computation of a kriging estimate, that S-Plus only has built in functionality to fit kriging models to variograms of the spherical, exponential, and Gaussian families. It is possible to program additional covariance forms, but this will require knowledge of the S programming language, or the involvement of additional personnel with such experience. Finally, regardless of the software package used, the user needs to be aware of how the software handles latitude and longitude. As discussed in Section 3.1, caution must be exercised when working with this coordinate set. Refer to Section 6 for further discussion of software packages for ordinary kriging and spatial interpolation.

Interpreting Results

It is important to recognize that relying on automated software packages to generate estimators using numerical techniques can potentially lead to undesirable results. Thus, care should be taken to evaluate any such results before proceeding. For example, consider the apparently best fitting model to the empirical variogram as computed by S-Plus for the case study example. A direct application of the tool suggests an exponential model with a range of 67,886 degrees, a sill of 36,101 [g/m^3]², and a nugget of 1.68 [g/m^3]². This model has the best objective value of the three model families applied (spherical, exponential, and Gaussian) with an objective value of 2.48. Unfortunately, the range and sill, while numerically optimal, are nonsensical in terms of the real world (e.g., the range exceeds the circumference of the Earth).

When dubious parameters such as these appear in an analysis, steps may need to be taken to stabilize the results. (Note that such undesirable variogram parameters may not practically impact the appearance of the kriging surface, but they are likely to impact the standard errors and the evaluation of uncertainty.)

One way to stabilize the empirical variogram and, thus, the parameters of the variogram model, is suggested in Cressie (1993). This alternative methodology, called robust estimation, is supported by S-Plus's spatial module. In general, robust estimation takes the fourth root of the squared difference within the average and raises the average to the fourth power. [Compare to Equation (2).] This results in an estimate that is less susceptible to contamination by outliers. See Cressie (1993) for the specific formulas and technical details. Table 3.3 lists the model and model parameters that provide the best fits to the empirical variograms generated with the robust estimation method. (See the caution regarding S-Plus's definition of range earlier in this section.) In general, the robust analysis provides a more realistic set of parameter values than the baseline analysis.

Effect of Temporal Averaging and Scale

It is important to consider the impact temporal scale can have on the estimation of the variogram. Most importantly, averaging over different temporal scales significantly changes the problem being analyzed. To illustrate this issue, similar variogram analyses were applied to PM_{2.5} data averaged over a year (annual data), averaged over four different quarters (quarterly data), and for individual days within each of the four quarters (daily data) for a total of nine variogram analyses. Table 3.4 and Figure 3.6 present the model parameters when a Gaussian model family is applied to an empirical variogram generated using the robust estimation method suggested by Cressie (1993). (See the caution regarding S-Plus's definition of range earlier in this section.) Note that the fit of the variogram model to the robust empirical variogram tends to improve, in the sense that the objective function tends to decrease with the amount of time over which the average concentrations are taken. This makes a certain amount of sense from a statistical point of view. That is, averaging over more data means less overall variability in the resulting data set, which can have the effect of reducing the nugget and the sill (both measures of variation). The reduced variability may also lead to stronger spatial correlation in the data set. If this is the case, a stronger correlation will have the effect of increasing the range (a measure of how far two locations need to be separated so that they are effectively no longer correlated). Journel and Huijbregts (1978) refer to this averaging as regularization and discuss the impact in some detail. (Note that the empirical variograms and objective functions for data taken at various temporal averages are not necessarily compatible and will not necessarily improve a kriging estimate.)

Regardless of the specific equations used to compute the objective function, they all have the same goal. Namely, they seek to provide a measure of how well a particular variogram model fits a particular set of empirical variogram points. This means that the objective functions for two variogram models compared to two empirical variograms with the same number of points might be roughly comparable.

Overall, it is important to keep in mind that the temporal scale of interest should be chosen to fit management decisions; that is, it should not be chosen for the sole purpose of improving the variogram model fit. In addition, by averaging over time one loses information available in the data set; which, from a statistical point of view, is potentially an undesirable occurrence. Careful analysis is always recommended in order to clarify the total impact of temporal averaging and changing temporal scales.

Table 3.3 Comparison of variogram models

	Baseline	Robust Estimator
Model Type:	Exponential	Gaussian
range:	67,886 degrees	9.538 degrees
sill:	36,101 (: g/m ³) ²	8.5 (: g/m ³) ²
nugget:	1.68 (: g/m ³) ²	2.02 (: g/m ³) ²
Objective:	2.4785	4.063

Table 3.4 Comparison of model parameters (using a robust empirical variogram) across temporal scales (PM_{2.5} data)

Gaussian Robust	Year	Quarter 1	Day 1	Quarter 2	Day 2	Quarter 3	Day 3	Quarter 4	Day 4
range:	9.54	6.04	8.94	12.47	4.14	5.11	2.59	7.59	3.38
sill:	8.50	3.26	42.58	18.79	101.81	6.90	32.99	13.53	68.62
nugget:	2.02	3.24	4.37	2.04	6.37	1.64	3.14	3.79	13.12
Objective:	4.06	3.36	40.51	7.29	1,176.52	5.86	331.77	13.55	494.95

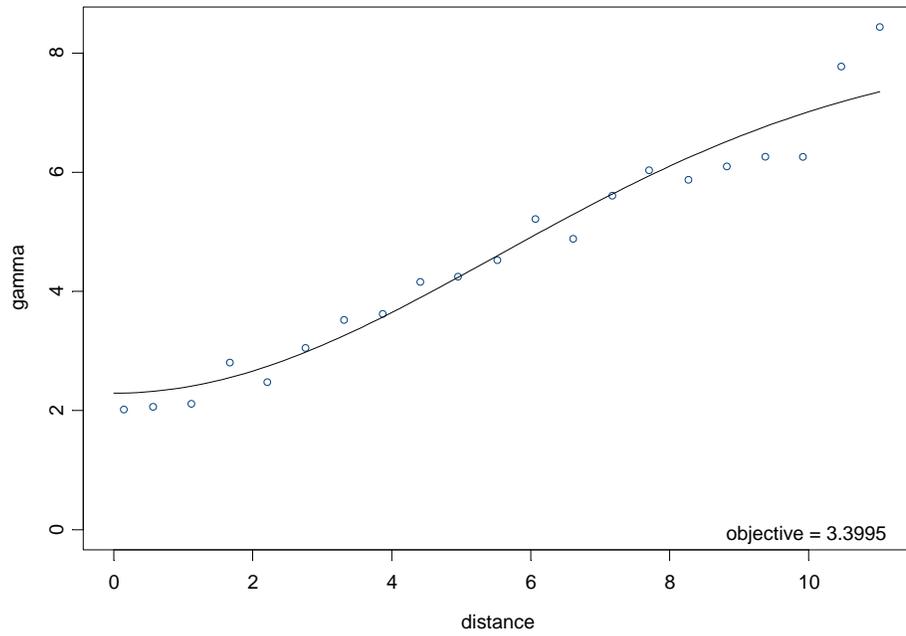


Figure 3.5 Empirical variogram (circles) and parametric variogram model (solid curve) based on annual $PM_{2.5}$ data.

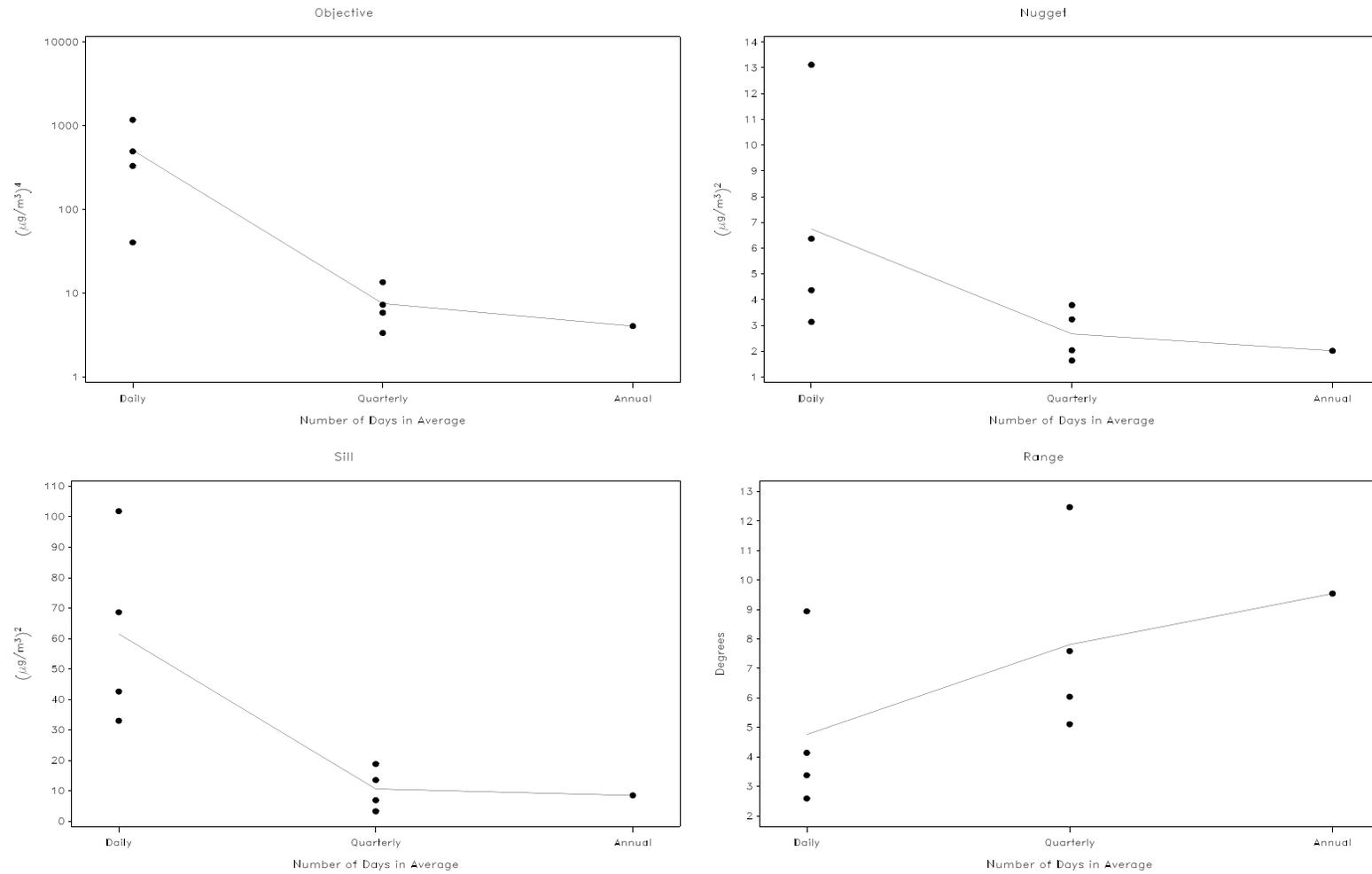


Figure 3.6 Comparison of variogram model parameters across temporal scales (PM_{2.5} data).

3.4 Step 4: Apply Spatial Prediction and Uncertainty Formulas

Kriging in general is a linear, global, gradual, stochastic spatial interpolator (see Section 2.1 for definitions). Recall that ordinary kriging defines a statistical model for the process of interest at a point in space (x, y) in terms of an overall, large-scale mean of the process across the spatial domain, u , and the small-scale random fluctuations of the process within the spatial domain, $e(x,y)$:

$$Z(x,y) = u + e(x,y) . \tag{Eq. 3}$$

This model places very little emphasis on the large-scale mean trend component, instead focusing on carefully modeling the structure and behavior of the small-scale random fluctuation. In the previous step (Section 3.3), the spatial structure associated with $e(x,y)$ was modeled using a variogram analysis. That variogram model is now applied to the task of spatial prediction.

Conceptually, the goal of kriging is to find linear combinations of the data that are, in some sense, optimal and consistent with the observed data points. Kriging determines optimal weighting by considering correlation of the process between data locations and the location to be estimated, and by considering the correlation (i.e., redundancy) among data locations. These weights incorporate the information contained in the assumed variogram model. Cressie (1993) and Ripley (1981) present technical developments of the kriging estimators, but from different perspectives. The reader is encouraged to read these texts for further understanding. Both developments represent the kriging estimate at a location (x_0, y_0) as follows:

$$Z.\text{est}(x_0, y_0) = \mathbf{8Z} , \tag{Eq. 4}$$

where $\mathbf{8}$ is a vector of kriging data weights and \mathbf{Z} represents the vector of all observations. The exact equations for this estimate and the associated variances (i.e., uncertainty equations) can be found in either text referenced above, and are reproduced in Appendix A.

Note that the bulk of the technical effort devoted to an ordinary kriging exercise occurs within the previous step of variogram modeling. The current step of applying the variogram analysis results via the kriging formulas in order to obtain spatial predictions and uncertainty estimates is relatively straightforward, given that the ordinary kriging equations already exist and are well known. This assumes, however, that software is available for applying the equations.

Software Considerations

Surfer, GMS, S-Plus, SAS, and several other software packages have the capability to compute kriging estimates given data and a variogram function. Focusing on S-Plus and SAS as in the previous section, note that S-Plus allows both command line and menu-driven kriging, while SAS only supports command line (or program-based) kriging. The menu-driven ordinary kriging in S-Plus accepts as inputs the data and a variogram model (including the model family

and the range, sill, and nugget). Kriging predictions and standard errors can then be generated everywhere in either a default or user-defined grid. The menu-driven option additionally provides several printing options, including contour (black and white or color) and surface plots for the predictions and standard errors of the kriging. These plots can include the prediction and/or observation locations at the user's option. Figures 3.7 and 3.8 depict the kriging predictions and kriging standard errors for the model variogram generated in the previous section. As is typical with any statistical estimates, the prediction standard errors tend to be smallest where there are plentiful data, in this case, plentiful monitoring observations. A more detailed example of using S-Plus's spatial module for kriging is provided in Section 7.

Limitations

It is important to keep in mind that kriging is a "smoother." That is, the variability in the kriging estimates is less than the variability of the unobserved, true spatial process. This does not mean that kriging is constrained to fall within the range of observed values, as the only constraint on the weights used to generate the kriging estimate is that they sum to one (see Appendix A). However, the reduced variability of the kriging estimates (as compared to the true spatial process) means that the sampling scheme (e.g., a monitoring network's design) is critically influential on the results of the modeling analysis. The sampling scheme should, in some way, provide representative data for the spatial region that is to be interpolated. (We comment on this further in Section 8.)

For example, in the case of a monitoring network designed to provide population-based neighborhood-scale air pollution information, it may be inappropriate to place monitors near point sources (yielding higher concentrations and less spatially representative information) or in rural/background areas (yielding lower concentrations). However, if the goal of ordinary kriging is to accurately model a detailed air pollution concentration surface throughout an entire spatial domain, then such a network may limit the accuracy of the resulting interpolation at many specific points in space because it fails to sample the extreme sub-regions of concentrations within the overall spatial domain. In general, the issue of the mechanism by which the data were generated, of which sampling scheme is a component, should be considered carefully when embarking on a spatial interpolation exercise, kriging or otherwise. From a statistical point of view, the best monitoring design for spatial interpolation (in the absence of other information, such as information about spatial gradients) is a dense uniform grid with coverage over the entire spatial domain of interest.

Finally, ordinary kriging also does not account for any spatial trend in the data. If it is believed that a spatial trend is critical to adequately describing the spatial process of interest, then universal kriging or other spatial interpolation approaches may be more appropriate. Universal kriging and, more generally, the issue of accounting for large-scale spatial trends are discussed in more detail in Section 4. In addition, a discussion and example of how sensitive results are to assumptions, fitting methods, and variogram models is included in Section 7.

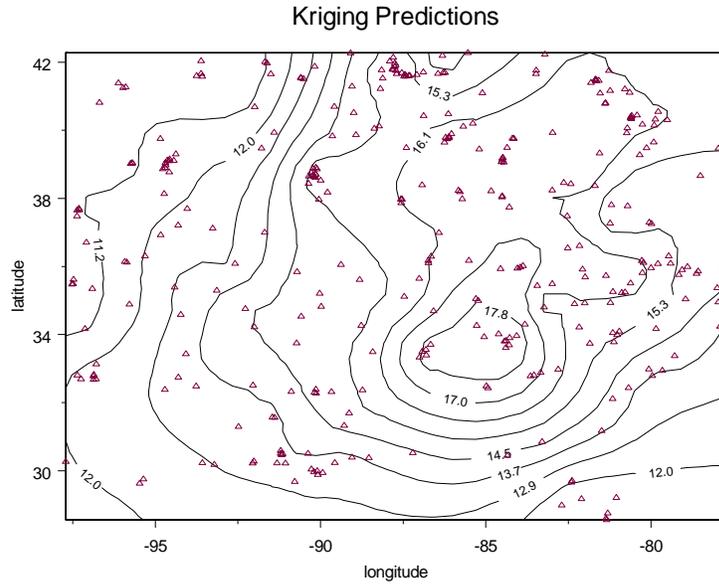


Figure 3.7 Ordinary kriging predictions based on annual PM_{2.5} data from case study example.

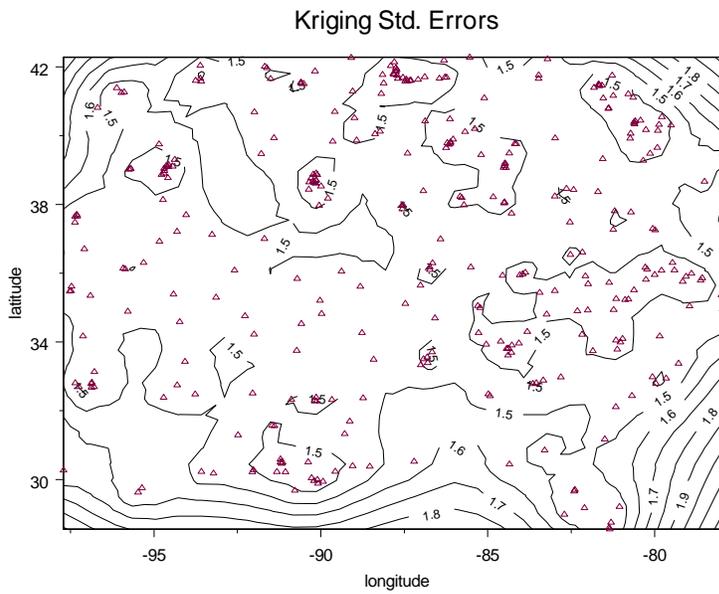


Figure 3.8 Ordinary kriging standard errors based on annual PM_{2.5} data from case study example.

3.5 Step 5: Evaluate Model Performance Through Diagnostics

Before continuing, review the ordinary kriging process thus far. Ordinary kriging is implemented via five steps. First, a data set must be identified and prepared for analysis. Then, as with all statistical processes, one should summarize and understand the data, looking for outliers or data abnormalities that may adversely impact the remaining steps. Once the data have been prepared and summarized, active analysis can begin. Ordinary kriging depends heavily on the identification of the spatial covariance structure of a data set. This is accomplished using a variogram analysis. Given the spatial covariance structure, the spatial prediction and uncertainty formulas associated with ordinary kriging can be applied to provide a predictive spatial surface. Finally, the model performance should be evaluated through diagnostics. It is on this final step that this section focuses. Observe that there are no formal approaches or answers to the model evaluation issue. However, some common sense ideas can be used, in conjunction with output from the ordinary kriging, to do some basic comparisons of performance.

Recognize that a variogram model can have a large amount of uncertainty associated with it. This uncertainty can result from your sample size, the target variable, the domain, and the underlying physics of the problem. As in most statistical analyses, outliers can be problematic. While one can spend a lot of time refining a variogram, in some cases this may be an inefficient use of resources, as small changes in a variogram may not have an appreciable affect on the kriging predictions. (We note that small changes can greatly impact the error surface, but caution the reader to balance the effort made to refine a model with the potential gain of changes in the final model.) With that in mind, the following section presents several diagnostics that may be used to evaluate model performance.

Compare Objective Functions

Certain evaluations and diagnostics are built into the ordinary kriging process as described in this document. Specifically, consider the objective function. This is a least squares measure of the fit of a proposed variogram model to an empirical variogram. (Recall that S-Plus provides the value of the objective function automatically.) By considering several variogram models, one can choose the one model of the proposed models that has the lowest objective function. Note that the absolute level of an objective value may be difficult to judge. (There is some evidence that the objective value of the best fit may decrease, or improve, with the number of days included in the temporal averaging time of the data. See Figure 3.6.) However, the relative levels of the objective functions when comparing across models is more meaningful. When considering a best fit, keep in mind that in practice there is little difference in the quality of a fit between a model with an objective function value of, for example, 1.2 versus a model with an objective function value of 0.888. Only when the differences between objective functions are reasonably large should one consider that one model might be truly superior to another. It is also important to keep in mind that the best objective function may not always produce the best model. That is, a good fit to the empirical variogram can be, but is not necessarily, a good indicator of how well the model describes the true spatial process. And, the final model selection

should strive to honor both the fit to the empirical variogram and consistency with the conceptual model for the underlying process (e.g., honor anisotropy, etc.).

Compare Visual Fits

A similar comparison process can be conducted by plotting several variogram models against the empirical variogram and visually looking for a best fit. As with many statistical analyses, visual inspection can provide insight that simple numerical summaries cannot. For instance, one might find that the model with the best numerical fit passes through or near all of the central empirical variogram values, but completely misses the behavior at the extreme distances of a variogram, while another, less numerically optimal model, follows the overall empirical variogram behavior better without precisely matching any of the empirical values. Additionally, visual inspection can be useful when the empirical variogram (or the variogram cloud, discussed earlier) has a distinct shape. For example, if an s-shaped curve is apparent in the empirical variogram as it is in Figure 3.5, this may be an indication that a Gaussian model is most appropriate, as the Gaussian variogram family is the only variogram family that exhibits this behavior. In such a case, it may be appropriate to focus one's analytical attention to Gaussian models, regardless of the numeric fit suggested by the objective function. Thus, one will often find it useful to compare both the values of the objective function and the visual fit of the empirical variogram to a proposed model.

Assess Parameter Estimates

Another important factor to consider when evaluating an ordinary kriging model's output is the set of parameters used to define the variogram model. As noted earlier, it is important to inspect the values of the sill, range, nugget, and any other parameters used to define the variogram used in the kriging process. Even if they are numerically optimal (i.e., they minimize the objective function), they may not make sense in the context of the analysis. For example, if a numerically optimal parameter set suggests a range in excess of the analysis space (e.g., a measurement in degrees latitude or longitude that exceeds the circumference of the Earth), one must strongly question those results. In such cases, it may be more appropriate to manually choose the variogram parameters. Then, an iterative refinement of those parameters will likely be part of this diagnostics step.

From a purely statistical or mathematical perspective, the real-world interpretation of the variogram parameters may be of little importance. That is, in performing an ordinary kriging exercise, the most important feature is how well a variogram model fits, not necessarily what parameters are used to obtain that fit. However, in most applications, the variogram parameters can be directly related to real-world physical and other characteristics of the underlying spatial process, which in turn speak directly to the non-statistical scientists and stakeholders involved in the analysis. Furthermore, one can often obtain several variogram models that have approximately the same fit, but that have very different sets of parameters. With these issues in mind, when presenting ordinary kriging models to a non-statistical audience, it may be of value to choose a model that has parameters the audience will find most easy to interpret.

Cross-Validate

Given that kriging is a spatial interpolation, one possible diagnostic approach would be to test its robustness to errors in the data. One might accomplish this by repeatedly applying the kriging equations using the variogram model defined for the full data set to a reduced data set (i.e., remove one or more observations from the analysis), and then considering the differences between the reduced data model results and the observations left out of the analysis. This method of diagnostics is sometimes referred to as cross-validation (Cressie, 1993).

For illustration, consider a small spatial interpolation problem with ten data points. Compute and model the empirical variogram. Then remove the first data point and compute the ordinary kriging estimate and standard error at that data location, using the already chosen variogram model. Compute the squared difference between the first data value and the kriging prediction at that location. Repeat the process for data points two through ten, each time computing a kriging prediction and squared error for the remaining nine data points. These differences (or residuals) form an empirical distribution of the kriging variance, (i.e., the standard errors squared). Compare this empirical distribution to the theoretical error distribution generated from the kriging equations. If they are dissimilar, this suggests that the variogram may have been incorrectly modeled. More information on cross-validation (and other performance measures) are presented in Section 5.1

Consider More Complex Models

Another type of diagnostic is the investigation of more sophisticated spatial interpolation models. There are a number of variations that may be applied to the ordinary kriging process, which may lead to a better spatial prediction. For example, universal kriging refers to a kriging prediction that incorporates some sort of spatial trend. This may be appropriate when the investigator has prior knowledge that suggests that a simple trend (perhaps a linear trend or polynomial of low degree) is indeed evident for the process across the space of interest. Another example of additional model sophistication is the use of an anisotropic (directional) covariance structure. The basic variogram model discussed in this document focuses on variogram structures that are identical in every direction (i.e., isotropic models). Anisotropic variogram structures incorporate directional impact (as may arise from prevailing winds, for instance).

Incorporating such complexities into a kriging predictive surface will almost certainly provide what appears to be a better overall model fit, as they represent an increase in the overall number of parameters in the model. If, in the opinion of the data analyst and other stakeholders and decision makers, the model improvement is good enough, this may indicate that the baseline model is not sufficient. As discussed above, determination of whether a model is good enough may be done through the comparison of objective functions (i.e., decide whether the additional model complexity significantly improves the objective function). An additional method of comparing the fits of two competing models is to consider the prediction standard errors resulting from each model. In general, a model with consistently lower standard errors (given an

appropriate variogram fit) might be considered an improvement. The use of more sophisticated models is discussed in more depth in Section 4.

A Diagnostic Algorithm

As a final guideline on diagnostics, one might consider the following general algorithm incorporating the suggestions above for evaluating an ordinary kriging prediction.

- Step A: Using both standard and robust estimation, fit three variogram models to the empirical variogram: exponential, Gaussian, and spherical. Consider the value of the associated objective functions and the visual fit. Choose the model with the best fit and the most applicability to your situation.
- Step B: Consider a more sophisticated model, e.g., incorporating a trend and/or anisotropic variogram structure. Fit three variogram models as in Step A. If the objective value of the best fit is dramatically lower than the previous best fit, this suggests that the more sophisticated structure is a better fit.
- Step C: Optional: Test for the robustness of the chosen model to data errors by performing a cross validation.

This diagnostic algorithm is only a suggestion. Any diagnostic process must be tempered by local knowledge about the spatial region and variable of interest. As a result, other diagnostic processes may be appropriate for specific applications or organizations.

3.6 References

- [1] Chamberlain, Robert G., Caltech (JPL); “What is the best way to calculate the distance between 2 points?”; <http://www.geog.ucsb.edu/~good/176b/a15.html>.
- [2] Cressie, Noel A. C. (1993). Statistics for Spatial Data. John Wiley & Sons, New York, New York.
- [3] Journel, A. G., And Huijbregts, Ch. J. (1978). Mining Geostatistics. Academic Press, London, England.
- [4] Ripley, Brian D. (1981). Spatial Statistics. John Wiley & Sons, New York, New York.

4.0 COMMON EXTENSIONS TO THE ORDINARY KRIGING MODEL

The preceding section provided a detailed description for the spatial interpolation method of ordinary kriging. This section offers an overview of extensions to that method. The reason for considering such complexities is that in some cases the ordinary kriging model may not be able to adequately describe the nature of the variability of a given spatial process. For example, ordinary kriging assumes that a constant population mean applies to the entire spatial domain under study, which may not be an appropriate assumption if the process exhibits a noticeable large-scale trend across the domain. The main advantage of employing a more complex modeling approach is the ability to address a wider array of spatial scenarios. The main disadvantage is the additional analysis burden that might come with such complexities.

A number of common extensions are considered in this section, including spatial trends, temporal dynamics, non-stationary covariance structures, use of covariates, and multivariate modeling. Given this document's emphasis on the importance of addressing spatial interpolation uncertainty as part of the modeling exercise, this section's presentation is centered on statistical methods for incorporating such extensions. Statistical methods naturally (typically) provide uncertainty estimates. In particular, several longstanding geostatistical kriging techniques are available to handle the types of extensions considered here. The method of *universal kriging* accounts for spatial trends, *kriging with external drift* incorporates covariates into the model, and *co-kriging* is a form of multivariate modeling. Space-time kriging approaches have also been developed. These geostatistical spatial interpolation methods and other statistical modeling extensions to the ordinary kriging model are discussed below.

4.1 Covariance Extensions

Our initial example (in Section 3) assumed that an isotropic and stationary covariance function (model) applied throughout the study domain. These assumptions simplify the analysis, but may not always accurately reflect the underlying behavior (e.g., physics, biology, atmospheric science, etc.) of the spatial process of interest. For example, in a region with prevailing winds, it may be unreasonable to assume that the spatial covariance of fine particulate matter is the same moving from west to east as it is moving from north to south. Likewise, it may not be appropriate to assume that the underlying spatial relationships are the same in predominately urban areas as they are in rural areas.

It is also important to note that the estimation of the covariance (or variogram) is likely the single most important and difficult step of any kriging procedure. We have demonstrated that this analysis process may be reasonably simple for a two-dimensional spatial surface that has isotropic and stationary covariance. However, each deviation from this simple case introduces additional complexity to the modeling process. The following sub-sections consider the issues of anisotropy and non-stationary covariance.

4.1.1 Anisotropy

An isotropic covariance structure is one in which the magnitude of the covariance between measured data at two locations depends only on the distance between the two locations. In contrast, anisotropic covariance is a structure in which the magnitude of the covariance between the observations at two locations depends both on the distance and the direction between the locations. This directional covariance structure can be caused by underlying physical processes that evolve differentially in space, such as prevailing winds. When modeling to spatially interpolate, the implementation of an anisotropic covariance model might provide a better overall description of the data by putting additional structure in the covariance component of the model (rather than in the trend component of the model, a possibility that is discussed in Section 4.2).

Generally speaking, there are two types of anisotropy: geometric anisotropy and zonal anisotropy. Geometric anisotropy occurs when the range, but not the sill, of the variogram changes in different directions (see Section 3.3 for range and sill definitions). Zonal anisotropy exists when both the range and the sill of the variogram change with direction. Of the two, we shall focus on geometric anisotropy in this section since, according to Kaluzny, et al. (1998), zonal anisotropy can often be corrected by detrending the data, a process that is closely related to universal kriging. [Universal kriging is discussed in more detail later in this section. Also, note that zonal anisotropy can be corrected for by using a nested variogram model. Nested variogram models are beyond the current scope of this document, but more details can be found in Cressie (1993).]

Geometric anisotropy means that the correlation is stronger in one direction than it is in other directions. Mathematically, if one were to plot the directional ranges (that is, the correlation ranges associated with specific directions), they would fall on the surface of an ellipsoid. (In two dimensions, they fall on the edge of an ellipse.) Thinking of things in this way, one can visualize the three-dimensional ranges of a geometric anisotropic model forming a watermelon while an isotropic model forms a basketball (see Figure 4.1 for a two-dimensional depiction).

One way in which geometric anisotropy can be identified and modeled is by calculating and plotting experimental directional variograms. If using a graphical technique, a directional variogram plots the empirical variogram for a specific band of angles. Plotting multiple disjoint directional variograms can indicate if one direction is significantly different from another and, hence, should be modeled by a different covariance structure. Specifically speaking, if one identifies a shape that seems to be consistent across the majority of the directional variograms, but there are one or more directions that deviate from the basic shape, anisotropy is a distinct possibility. If the difference in shape appears to occur primarily in the range, then this is an indication of geometric anisotropy.

When geometric anisotropy is indicated, the variogram must be modeled with two additional parameters: the directions and ratio of minimum and maximum correlation ranges.

Observe that an isotropic model can be thought of as a specific case of an anisotropic model; namely, one with angle = 0 and ratio = 1. In practice, the angle of adjustment is typically identified as the direction corresponding to the longest correlation range. The ratio is then the value calculated by dividing the shorter correlation range into the longer correlation range. For example, if the range in the east-west direction is 20 km and the range in the north-south direction is 5 km, then the ratio would be equal to 4.

Practically speaking, geometric anisotropy can arise in air quality data as a result of prevailing winds and/or other meteorological factors. Thus, it may be of interest to consider how one might conduct an analysis of a special case. For example, suppose prior analysis or expert local knowledge suggests that winds travel in a single direction across a region of interest. Further suppose that winds in this region tend to blow from south to north. Measured in degrees, this direction is 90 degrees away from a default direction of east to west. By computing an empirical variogram limited in direction to, say, 10 degrees to either direction of 90 degrees, one can identify the approximate variogram structure associated with that direction. Likewise, by computing the direction-limited variogram associated with 0 (zero) degrees, one can identify the approximate variogram structure associated with the perpendicular direction. (These directional variograms can be created using the command line “variogram” function in S-Plus. See Kaluzny, et al., 1998 for more information.) Assuming geometric anisotropy holds, the angle of the anisotropy in this example is 90 degrees and the ratio of the anisotropy is the ratio of the range of the 90-degree variogram to the range of the perpendicular variogram (zero degrees).

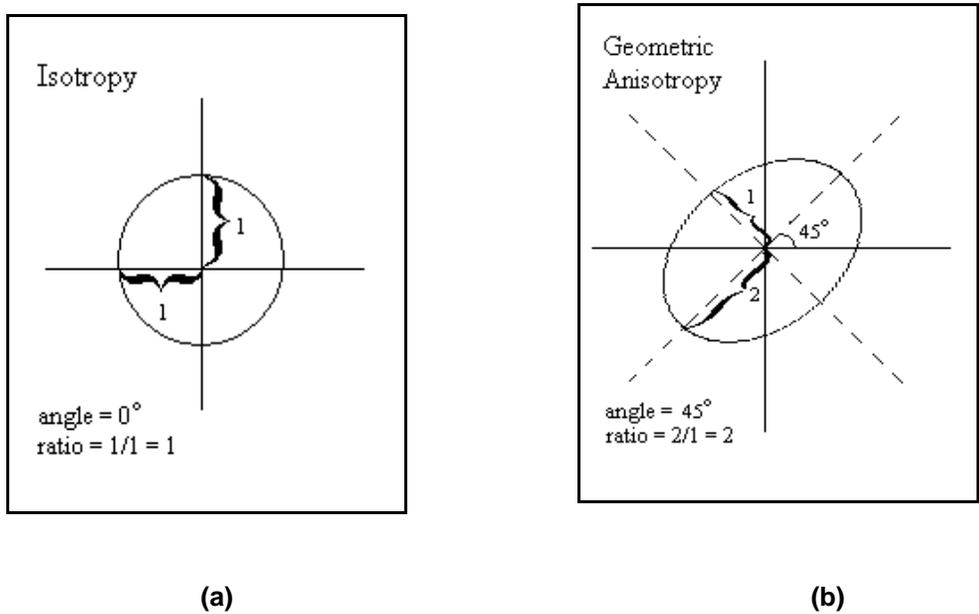


Figure 4.1 Comparison of Isotropy to Geometric Anisotropy.

Once a proper ratio and angle have been identified, most software packages will correct for geometric anisotropy internally. Thus, in general, it will not be necessary for one to make

alteration to one's data manually. Note that if manual manipulation is necessary in a specific circumstance, Cressie (1993) and Kaluzny, et al., (1998) discuss the technical and theoretical details of adjusting data for geometric anisotropy in more detail.

In practice, identifying the nature of an anisotropic covariance can require an extended study, often involving trial and error. Statistical packages with spatial functionality (such as S-Plus) can assist in this investigation, but are unlikely to give direct answers. Investigation must be done to identify which anisotropy (if any) provides the best variogram fit to the data. This process can be aided by the use of graphics, but often requires subjective judgement. Fortunately, once the proper parameters are identified, such packages will often compute the necessary linear transformations based on those parameters. Specifically, to create an anisotropy plot in the S-Plus spatial package, choose Spatial $\hat{\circ}$ Geometric Anisotropy from the main menu. (Similar effects can be generated using command line code. See the S-Plus spatial documentation for more details on how to do this.) Choose the data set of interest. Select the variable, Location 1, and Location 2 fields. Enter the angles and ratios to explore and press OK. A multipanel plot appears in a graph sheet, with several directional variograms for all combinations of ratio values and directions entered. This tool for investigating geometric anisotropies primarily conducts a linear transformation of the spatial region and then computes an omnidirectional variogram based on the transformed space. [See Kaluzny, et al., (1998) for details on this transformation.] This pull-down menu-based function is easy to use, but can lead to results that are not necessarily easy to interpret.

Using S-Plus, we produce an example multipanel plot in Figure 4.2a. Each individual plot in this figure displays the empirical or modeled variogram value on the vertical axis (i.e., gamma) versus distance on the horizontal axis. Recall that the empirical gamma ($\hat{\circ}$) function was described in Section 3.3 [see Equation (2)]. Figure 4.2a is based on the annual average PM_{2.5} data introduced in Sections 2 and 3. The angles chosen are 0, 45, 90, and 135 degrees. The ratios chosen are 2, 4, 8, and 16. Note that an isotropic model has ratio=1 and angle=0, so our chosen set of ratios and angles allows comparison of various potential anisotropies to the isotropic model. Observe that the anisotropic empirical variograms are all visually similar. The x-scale tends to increase with the ratio, but this is to be expected. (Keep in mind that the anisotropic adjustment transformation tends to stretch the distances, thus increasing the overall distance scale.) These observations suggest that this data set does not demonstrate any strong anisotropies.

Another, possibly better, tool for investigating geometric anisotropy involves generating directional variograms by invoking the `variogram()` command at the S-Plus command prompt. Figure 4.2b was generated using the following S-Plus code:

```
pm25.dvar1 <- variogram(conc ~ loc(lat, lon), data=ann.avg, azimuth=c(0,45,90,135),
  tol.azimuth=22.5)
plot(ann.avg)
```

Note that *ann.avg* is the annual average PM_{2.5} data set, *conc* is the concentration, and *lat* and *lon* are the latitude and longitude, respectively. The *azimuth* and *tol.azimuth* subcommands control the direction and width of the arcs for which the directional variograms are computed. (See the on-line help within S-Plus for more information on this command.) The resulting Figure 4.2b shows four directional variograms that are quite similar in shape, which is indicative of an isotropic correlation structure. In situations such as this, one would typically assume an isotropic model, but for the purpose of illustration in the PM_{2.5} discussions that follow, we shall assume that an anisotropy exists with a ratio of 1.5 and an angle of 90 degrees.

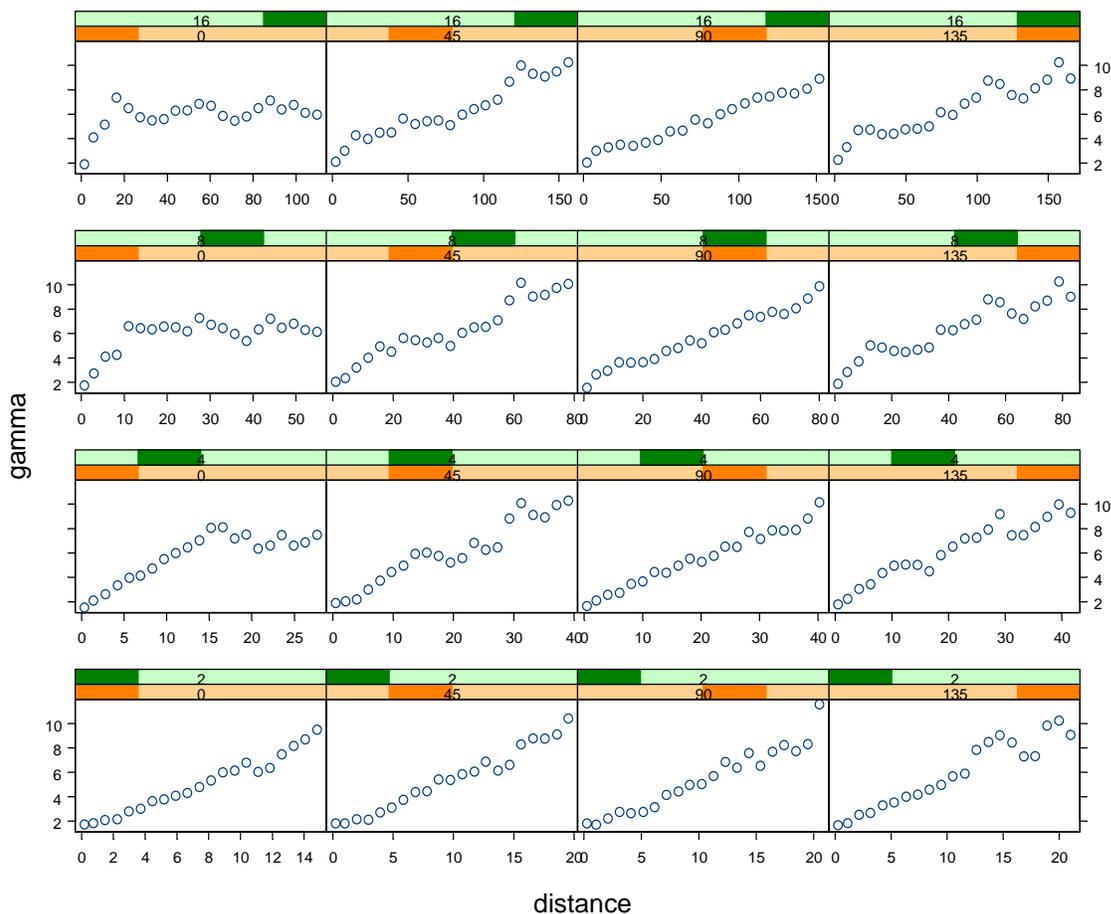


Figure 4.2a Geometric anisotropy plot — in each subplot the upper label indicates the ratio and the lower label indicates the angle (angles 0, 45, 90, and 135; ratios 2, 4, 8, and 16) — annual average PM_{2.5}

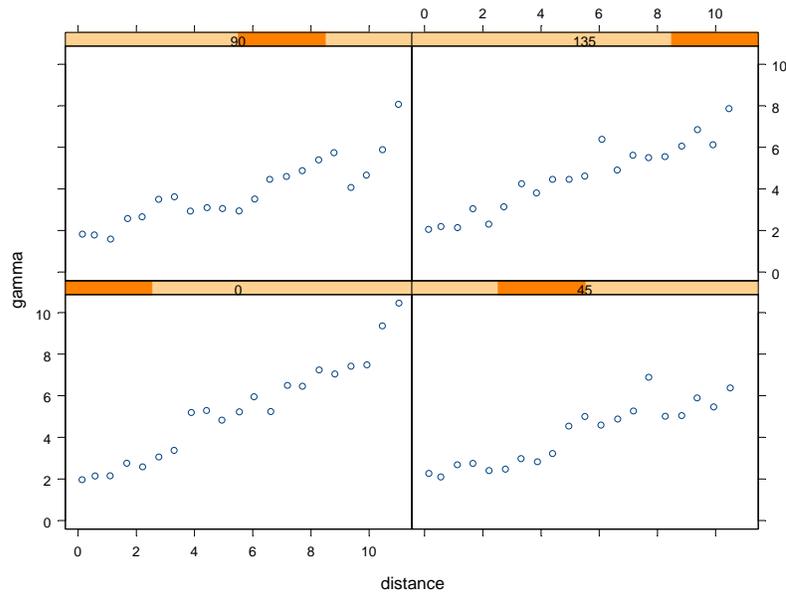


Figure 4.2b Directional variograms using S-Plus code — annual average $PM_{2.5}$

To better understand the relative diagnostic utility of the two graphical methods highlighted in Figures 4.2a and 4.2b, we consider a set of distinctly anisotropic coal seam data originally published by the U.S. Bureau of Mines in 1970 (see Gomez and Hazen, 1970). A number of authors have analyzed this data set over the years and it is well-documented to have an anisotropy. (Specifically, the direction of this anisotropy along the long axis is approximately 17 degrees east of north. For the purposes of simplicity, we will focus our analyses on the angles 17, 62, 107, and 152 degrees.) We analyze the potential anisotropies in these data using the Geometric Anisotropy and directional variogram plots discussed above, and present the results in Figures 4.2c and 4.2d, respectively.

An experienced eye may be required to detect the anisotropy using only Figure 4.2c. At the time of this writing, the authors do not see any clearly interpretable information in these plots. Investigation of S-Plus documentation does not illuminate the issue of how to identify the (known) anisotropy in this data set using this tool.

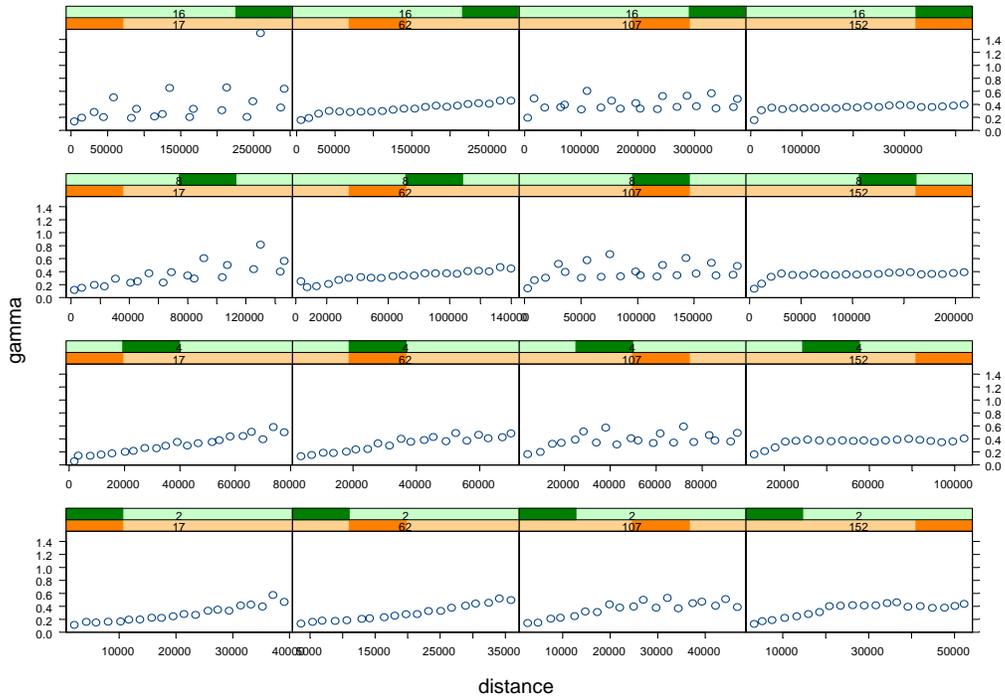


Figure 4.2c Geometric anisotropy plot — in each subplot the upper label indicates the ratio and the lower label indicates the angle (angles 17, 62, 107, and 152; ratios 2, 4 8, and 16) — coal seam data

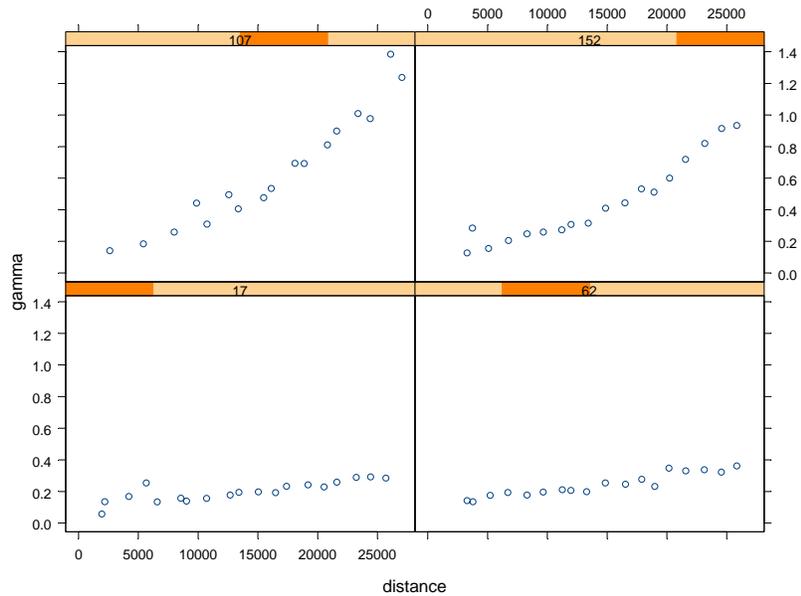


Figure 4.2d Directional variograms using S-Plus code — coal seam data

Figure 4.2d is easier to interpret. This plot was created with a modification of the S-Plus code presented earlier. When interpreting this, recall Figure 4.1. It is important to compare directional variograms that are at right angles to one another. If all four directions have visually similar variograms, this corresponds to the circle in Figure 4.1 and, thus, isotropy. On the other hand, if the variograms in perpendicular directions increase at different rates, a geometric anisotropy may be indicated. This case corresponds to the oval in Figure 4.1. Observe that in the left column of Figure 4.2d, the directional variograms are reasonably similar in shape and slope. Also, observe that the directional variograms in the right column of the figure are visually quite different from each other (as well as from the left column). In the figure, images that are above/below one another are directional variograms at right angles to one another. Thus, it appears that a geometric anisotropy, oriented along the 17- to 107-degree axes, is associated with this data set. Since the 17-degree variogram is increasing much more slowly (that is, it is associated with the longer axis of the oval in Figure 4.1), we say that the anisotropy is in the 17-degree direction.

It now remains to consider the ratio of the anisotropy. While it cannot be computed directly from these plots, a heuristic estimate can be gained by comparing the 17- and 107-degree images. Observe that, in the 17-degree plot, a gamma value of approximately 0.25 ft^2 is reached at the maximum observed range of 25,000 ft. Also, note that a similar gamma is reached in the 107-degree plot at a range of 6,250 ft. The ratio of these approximate ranges suggests a ratio of $25,000/6,250=4$ might be appropriate for describing this geometric anisotropy.

In summary, the multipanel plots produced by the S-Plus Geometric Anisotropy function are easy to produce using convenient drop down menus. However, analysis of the distinctly anisotropic coal ash data suggests this graphical diagnostic tool may be difficult to interpret, and the S-Plus documentation does not provide guidance on this matter. A better diagnostic tool for evaluating geometric anisotropy is the calculation and graphical comparison of directional variograms (compare Figures 4.2c and 4.2d). S-Plus can conduct such analyses and produce such plots, but it must be done from the command line (see the example code presented earlier).

We now plot robust empirical variograms and a Gaussian model for the isotropic and anisotropic annual $\text{PM}_{2.5}$ case mentioned above and shown in Figures 4.2a and 4.2b. Observe two important points. First, the empirical variogram points appear to be clustered more closely around the model line for the proposed anisotropic model than for the isotropic proposal. (In this example, the impact is minimal, with only two points that show practical difference.) Secondly, the objective function of the baseline, isotropic variogram is approximately 4.1, while the objective function value for the anisotropic model is closer to 1.6. Both of these observations suggest that the anisotropic model may be an improvement over the isotropic case.

However, as with all statistical and other modeling procedures, it is important to balance the improvement in fit with the simplicity of the model. More sophisticated modeling of the variogram to better match the empirical variogram is analogous to increasing the degree of a polynomial regression to more closely match the observed data. Devore (1995) observes that while adding parameters to a regression can decrease the overall sum of squared differences

between the regression line and the observed values, and thus improve the overall model fit, the effective improvement can be minimal. Likewise, an improved fit of an empirical variogram may not necessarily provide a meaningful improvement to the kriging and associated uncertainty estimates that ultimately result.

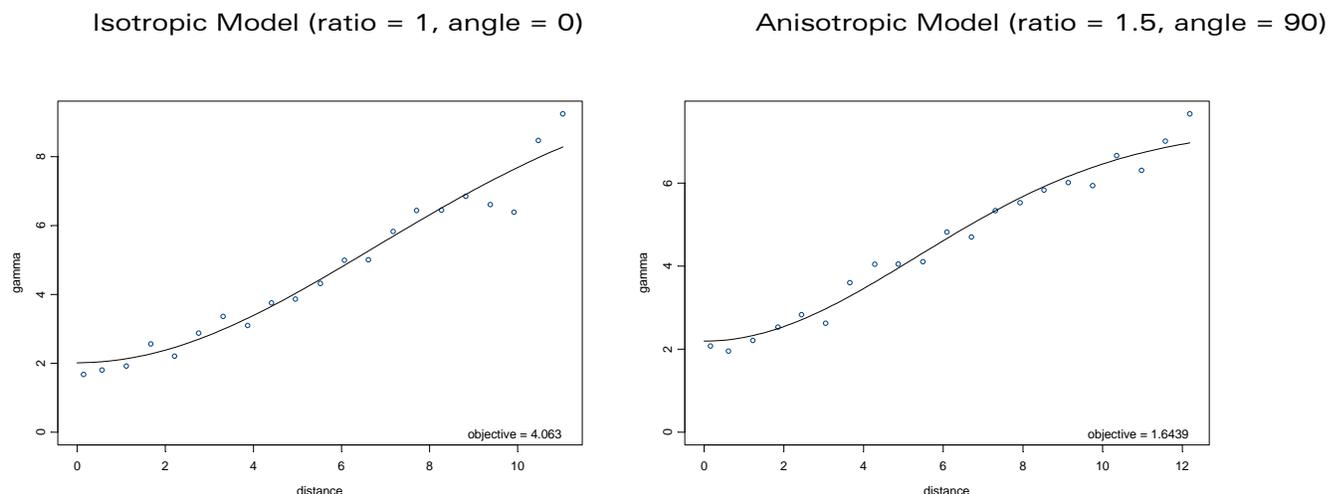


Figure 4.3 Comparison of empirical variograms and variogram models under the isotropic and anisotropic covariance assumptions.

Practically speaking, small differences in variogram models are unlikely to have significant impacts on the results of a kriging analysis. One must consider how much of a change in the variogram model is required to cause a meaningful change in the kriging matrix (see Appendix A). And, given a change in the kriging matrix, one must consider how that might impact the kriging weights. In addition, any given change on the kriging standard errors is ultimately evidenced as the square root of the kriging variance, thereby lessening any practical ramifications.

Keeping in mind the above discussion, let us now consider the changes in a variogram model that are in fact likely to have a substantive impact. An important aspect to keep in mind is that changes to the variogram at small distances will likely have greater overall effect than changes at long distances. Thus, we observe that large changes in the nugget will often have a noticeable impact on the kriging surface. Next, observe that using a Gaussian variogram model instead of a spherical or exponential variogram model may have an impact due to the difference in curvature (see Figure 4.4a). Finally, with regard to geometric anisotropy, observe that a fairly large ratio is necessary to have an impact (see Figure 4.4b). This is because the variograms at angles between the angle of the anisotropy and the perpendicular angle will tend to be close to the associated isotropic variogram. The differences at the extremes are more pronounced for high ratio anisotropic models and, thus, such models are more likely to be effectively different from the associated isotropic variograms.

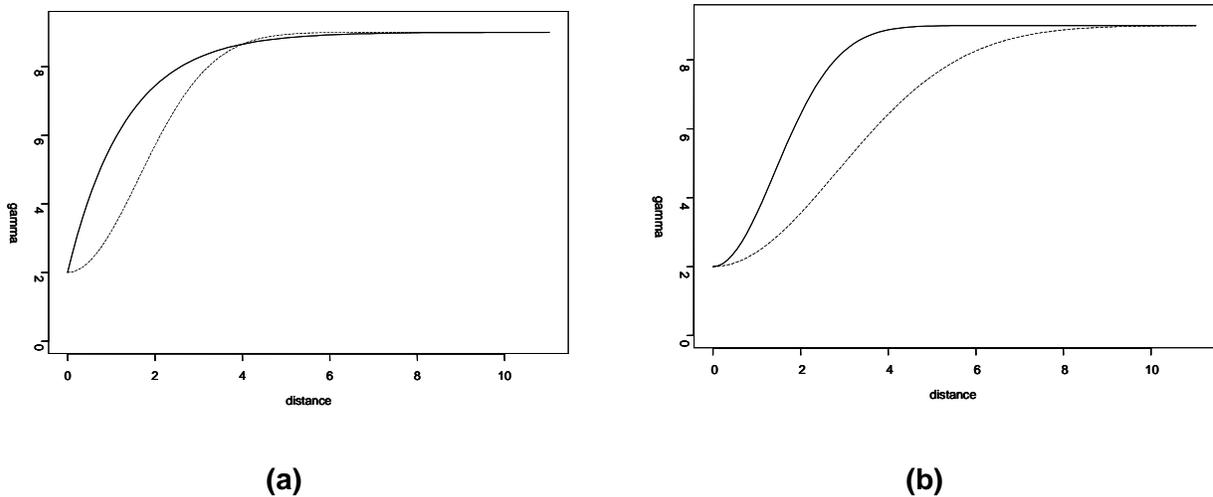


Figure 4.4 (a) Exponential (solid line, $a=1.333$) versus Gaussian (dashed line, $a=2.3094$) variogram model (range=4, sill=7, nugget=2). (b) Geometric anisotropy versus isotropic variogram model (range=4, dashed line vs. range=2, solid line; sill=7, nugget=2).

Returning to our $PM_{2.5}$ case study example, Figure 4.5 depicts the kriging predictions under the isotropic model of Section 3 (left panel) versus those of the chosen anisotropic model (right panel). Observe that the contour lines from the anisotropic model are smoother. Visually, this result could be either more or less appealing depending on one's subjective perspective. The anisotropic kriging standard errors are also different, being somewhat larger and less variable across the domain in relation to those of Figure 3.8. In particular, note the results in Table 4.1, which displays various percentiles of the distribution of standard errors associated with each model's spatial interpolation surface. In summary, while the distributions of the standard errors are roughly the same, the contour maps of the kriging surfaces exhibit certain differences. Without additional information, neither model demonstrates clear superiority over the other.

This may suggest that it is unnecessary to choose the more complex anisotropic model over the simpler isotropic model, despite the apparently better variogram fit of the anisotropic model as suggested by Figure 4.3. With parsimony and interpretability in mind, the simpler model (isotropic) may be preferred.

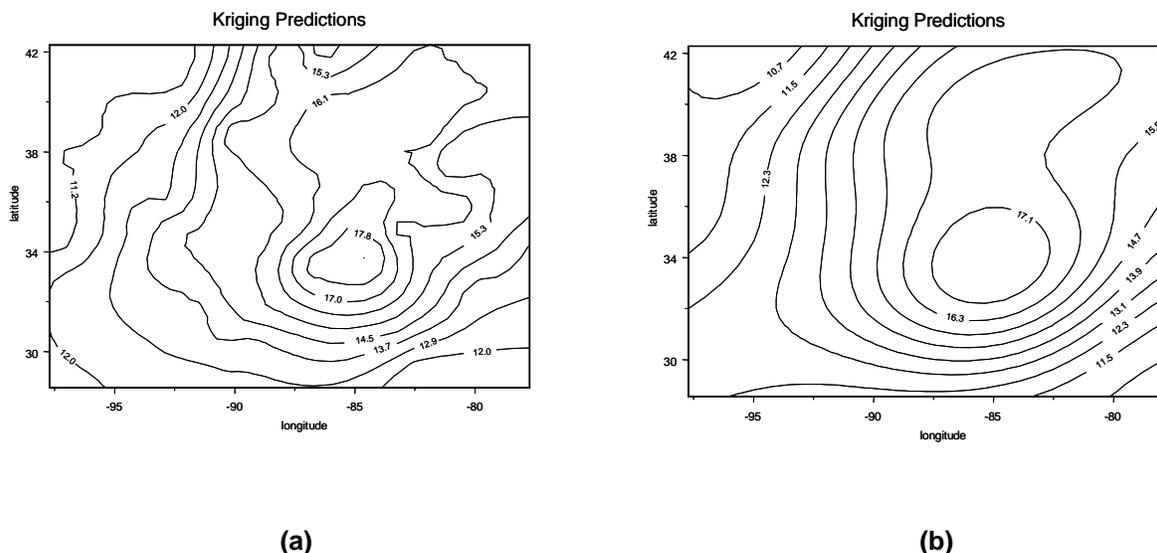


Figure 4.5 Kriging predictors for (a) isotropic and (b) anisotropic model based on annual $PM_{2.5}$ data.

Table 4.1 Comparison of percentiles of standard errors between isotropic and anisotropic model based on annual $PM_{2.5}$ data

Model	Prediction Standard Error Percentiles (: g/m^3)				
	Minimum	25th	Median	75th	Maximum
Isotropic Kriging	1.433	1.440	1.448	1.473	1.824
Anisotropic Kriging	1.500	1.513	1.527	1.560	2.020

4.1.2 Non-Stationary Covariance

Both the isotropic and the anisotropic (covariance) models assume that the spatial covariance structure is essentially identical throughout the spatial domain of interest. This property is called second-order stationarity. The official definition requires that the mean of all locations of interest be equal and that the covariance between any two locations depends only on

the distance and direction between them and not specifically on the actual location. When it is suspected that this is not the case, either for physical reasons (e.g., a spatial region that contains both highly urban and highly rural areas) or for data reasons (e.g., trend removal and anisotropy do not improve the kriging estimates), more complicated models may be required. One way of addressing these more complicated scenarios is identifying regions that are locally stationary, both in terms of trend and spatial covariance structure, and then dealing with the sub-regions piecewise.

Section 3.2 of this document introduced the idea of using “moving windows” to analyze a spatial region. Heuristically, one can extend this concept to identify regions with similar (and dissimilar) spatial covariance structures. Consider partitioning a spatial domain of interest into disjoint regions in some convenient manner. In Section 3.2, a three-by-three regular partition along latitude and longitude lines was selected. More generally, one might partition a region along whatever coordinate system is being used in the analysis. Another suggestion is to partition the spatial domain in one direction (e.g., by latitude or the y-coordinate) yielding a number of horizontal rows or vertical columns of spatial sub-regions. Whatever partition structure is chosen, it is important to find a balance between the number of partitions (when exploring for non-stationary covariance, more is better) and the number of observations in a sub-region (if too few observations are available, a sub-region’s variogram model may be misleading or impossible to compute).

Once an overall spatial region of interest has been partitioned in some manner, the next step is to estimate the variogram models for each of the spatial sub-regions. Once the estimation is complete, compare the resulting models to one another. This comparison can be as formal as one wishes, but for the purposes of this discussion, we suggest keeping the exercise fairly simple. That is, compare the model family, sill, range, and nugget. Are the individual sub-region models similar? If so, regions with similar models can be combined for the final analysis. If not, they should be kept separate. (A quick check to see if a proper combination has been made is to check the fit of a combined region variogram to the individual sub-regional variograms. If the two models are similar, the combination was appropriate.) After checking and combining all sub-regions as appropriate, one can consider using the relative variogram techniques described in Cressie (1993, pp 64-66) to complete the analysis.

These types of investigations and associated model extensions can quickly become highly complex. While statistical software packages can often be programmed to assist with such analyses, few, if any, have user-ready options to conduct them automatically. Specifically, when using S-Plus, one must subdivide a region manually before fitting and plotting the distinct sub-regional variograms.

Figure 4.6 and Table 4.2 illustrate the proposed “moving-window” technique as applied to the annual average $PM_{2.5}$ data discussed in previous sections. (See the caution regarding S-Plus’s definition of range in Section 3.3.) For this example, the spatial domain was divided into nine sub-regions (see Figure 3.2). A robust empirical variogram was computed for each sub-region, and a Gaussian variogram model was applied to each empirical variogram. An initial

inspection indicates that each of the sub-regions has a different variogram. However, one must be careful in drawing any firm conclusions. For instance, Figure 4.7 compares the model variogram estimated for the full region of interest to the same sub-region empirical variograms of Figure 4.6. The solid lines in this figure's plots correspond to the full region variogram model and the circles correspond to the sub-region empirical variograms. From this set of plots in Figure 4.7, it is apparent that the full region model variogram provides a reasonable fit in a number of cases. Keep in mind these are real-world data, so a perfect fit should not be expected. Furthermore, the lack of fit between the full region model and sub-region empirical variograms may be due, in part, to a high degree of uncertainty in the sub-region calculations, i.e., there are less data available when dividing the full region into nine sub-regions. (Refer to Section 7.1 for an alternative regional decomposition for these data, namely four sub-regions.) In this case, it appears that (at least at this scale) regional differences in the covariance structure may be somewhat limited. In addition, the full region model variogram appears to provide a rather good fit to the full region empirical variogram (see Figure 3.5). As a result, we chose not to explore for non-stationary covariance further. Specifically, for the case study example data considered here, the stationary model of Section 3 appears to provide a reasonable fit to the annual $PM_{2.5}$ data. The issues of non-stationarity and associated modeling complexities are discussed in more detail in Journel and Huijbregts (1978).

Again, we stress here that the x- and y-scales associated with each of the sub-region variograms are different. This is a result of the default plotting functionality of S-Plus, and is maintained both to illustrate what a user can expect from default plots and to provide a warning. For variograms to be optimally comparable, they should be compared on consistent scales. Figure 4.7 illustrates this issue explicitly. While the same full region variogram model is plotted against each sub-region empirical variogram, the appearance of the full model curve in each graph varies significantly due to the variation in the x- and y-scales between each plot.

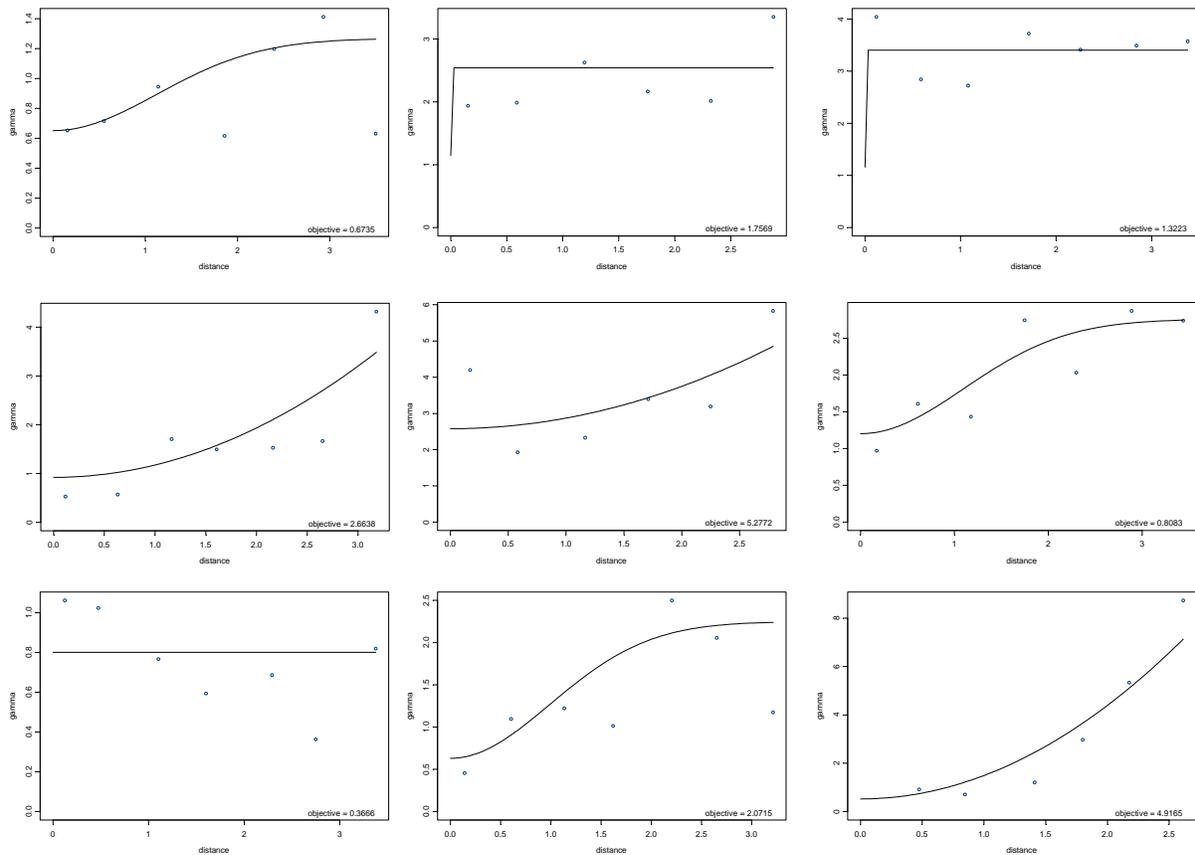


Figure 4.6 Graphical representations of relative variogram models.

Table 4.2 Numeric comparison of model parameters for relative variograms

Location	Objective	Range	Sill	Nugget
NW	0.6735	1.5885819	0.6061077	0.6513840
N	1.7569	0.000004030064	1.394803	1.145689
NE	1.3223	0.000003410967	2.245727	1.158600
W	2.6638	200.2026388	10,149.7677185	0.9215201
C	5.2772	207.155232	12,549.164065	2.578686
E	0.8083	1.555260	1.555038	1.202254
SW	0.3666	0.004201473	0.000008270449	0.8009557
S	2.0715	1.39807314	1.6155037579	0.6309632
SE	4.9165	180.6909998	31,585.8357172	0.5258113

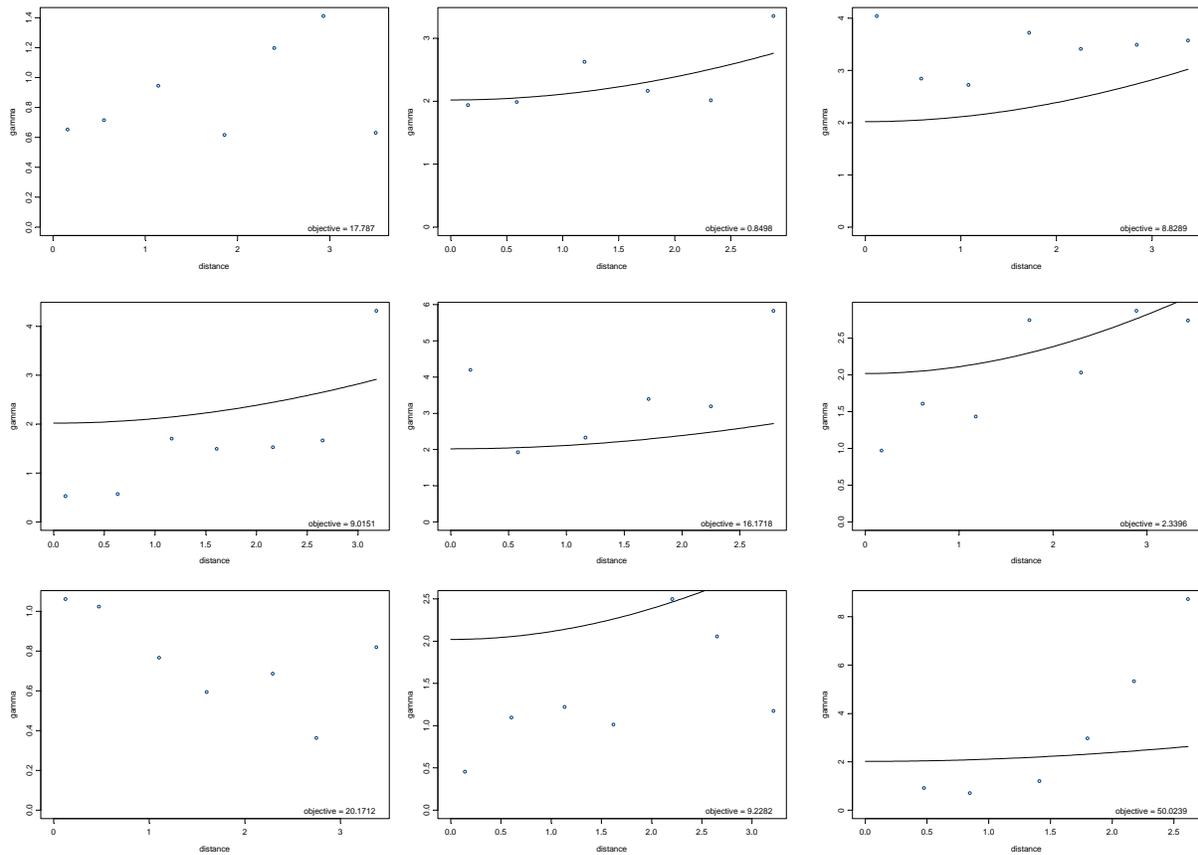


Figure 4.7 Graphical comparison of full region variogram model versus sub-region empirical variograms.

4.2 Trend Extensions

Our initial example (in Section 3) assumed a constant mean, uninfluenced by any external variables. Specifically, recall the u term in Equation (1) from Section 3. We now consider extensions to more complex models that account for large-scale spatial trend in the process of interest. In particular, we focus on two kriging models to address trends; universal kriging and kriging with external drift.

4.2.1 Universal Kriging

Recall that kriging allows the prediction of unknown values of a random function as a linear interpolation of observed values at known locations, using a model of the covariance of the random function when calculating predictions of those unknown values. While ordinary kriging assumes a constant trend, universal kriging is an adaptation that accommodates a spatially

varying trend. Universal kriging can be used to both produce local estimates in the presence of trend, and to estimate the underlying trend itself. (Note that universal kriging with a constant mean is equivalent to ordinary kriging.) Strictly speaking, universal kriging refers to kriging estimation using (a linear combination of) functions of the location, though we address briefly in this section the use of covariates in a limited way. For more details on the use of covariates see the next sub-section, which introduces the interpolation method of kriging with external drift.

Universal kriging may be used when one has prior knowledge of the physical properties of a process that indicates a large-scale spatial relationship exists. For instance, when investigating fine particulate matter in a region that is primarily urban in the north and primarily rural in the south, it may be reasonable to expect that there is a large-scale spatial trend with higher $PM_{2.5}$ concentrations at higher latitudes. Even if prior knowledge does not indicate such a relationship, inspection of the data may suggest that a universal kriging approach is still worthy of consideration. For instance, if a preliminary three-dimensional plot of a data set depicts distinctly higher data in the center of the region of interest, surrounded by gently decreasing data away from the center, one might consider a universal kriging approach using a linear (or some other polynomial) trend in both the east and north directions.

In certain circumstances, universal kriging can provide a better fit to the spatial process of interest. We must be careful in defining better here, as all kriging methods are interpolators and, thus, tend to provide estimates consistent with observations. Thus, by a better fit, we suggest that under appropriate conditions a universal kriging estimator can more accurately depict the true underlying process of interest. This is accomplished by increasing the complexity of the trend model. Specifically, Equation (1) from Section 3 is modified as:

$$Z(x,y) = u(x,y) + e(x,y). \quad (\text{Eq. 5})$$

Notice that the mean term in this model, $u(x,y)$, now depends on location unlike in Section 3.

Statistically speaking, care must be taken when considering more complex models of any sort, whether they are attempting to describe trends or any other behaviors. In this case, we consider specifically universal kriging models, but these cautions should be applied to other methodologies as well. As with many statistical analyses, the introduction of additional parameters will tend to reduce the overall uncertainty of the process of interest left unexplained by the model. However, as a result, one must be careful to avoid over specification of the model that can result in fitting not to the true process, but, for example, to the random pattern of imprecision associated with the measurement of the data.

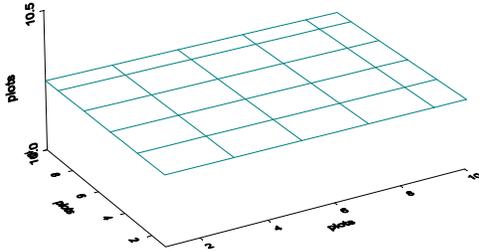
To understand this issue more clearly, consider the analogy of linear regression. Several authors discuss the dangers of overfitting regression models in detail. We summarize the discussion in Seber (1977) here. By considering the least squares regression estimate, the author proves that adding even a single extra regressor to a regression model results in model prediction intervals that are at least as wide, if not wider, than the original, reduced model. Seber concludes that while bias can be reduced and the fit can be improved by a more complex model, the

variance of the predictor is not reduced. In fact, Walls and Weeks (1969) provide an example in which the variance of the prediction at a particular point is increased tenfold when the model is enlarged by a single parameter. Thus, while it is true that the mean square error (MSE) of a regression model may increase or decrease when extra regressors are added, it is important to avoid overfitting. Similar issues must be considered when expanding kriging, or indeed, any statistical model.

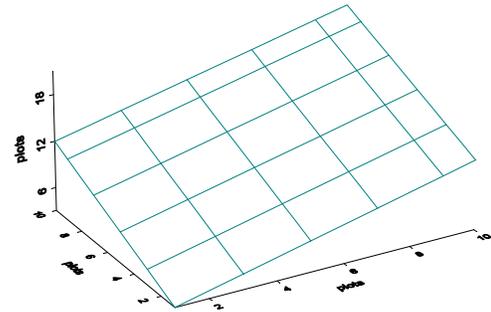
When considering specific types of universal kriging models, several possibilities present themselves. Most commonly, a low-order polynomial on spatial location is chosen to model the large-scale spatial trend. This choice has certain advantages. Foremost of these advantages is interpretability. Often, one can easily picture what a linear or quadratic trend might look like, where more complicated functions are less clear (see Figure 4.8). Likewise, it can be easier to imagine what physical or environmental factors might cause a simple trend in a spatial process. For example, if a universal kriging analysis of $PM_{2.5}$ results in a positive trend coefficient associated with the west-to-east component of variation, this indicates that particulate matter concentrations are generally higher in the east than in the west. If inspection of the spatial region then indicates a predominance of point sources, such as industrial plants, in the east and lack of such in the west, then one might presume that the point sources are influencing this trend. As an aside, note that the universal kriging function available as a menu option in S-Plus can automatically include up to quadratic trends in location.

Other more complicated large-scale spatial trend models include locally weighted regression (lowess) and splines. Fitting more complicated models can provide better overall fit of the trend to the observations, but may reduce interpretability and increase the danger of model over-specification as discussed above. Lowess and spline models, in particular, potentially involve many more parameters than a low-order polynomial regression and, thus, may pick up on spatial trends that appear to exist, but in reality are due largely to the data's measurement imprecision. On the other hand, such complex models may stabilize and simplify an otherwise highly complex covariance function (see discussion in Section 4.1).

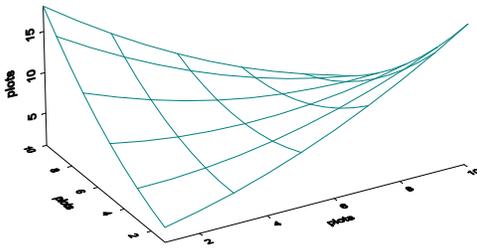
No Trend (ordinary kriging)



Linear Trend



Quadratic Trend



Cubic Trend

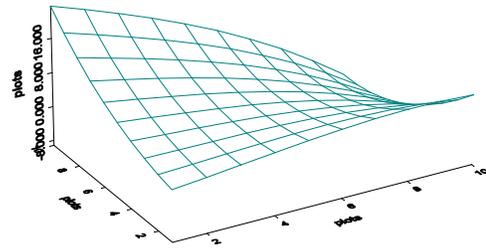


Figure 4.8 Examples of increasingly complex polynomial trend surfaces.

As always, care must be taken when attempting to extrapolate estimates outside the range of data used to generate a universal kriging interpolation. The addition of a trend can, like in ordinary least squares regression, indicate nonsensical values outside the range of the data. For example, a negative linear trend in the north-south direction will eventually suggest negative values outside of the data. Consider Figure 4.8. These plots depict the impact of adding terms to a polynomial trend. Note that the constant and linear trends generally look like a sheet of paper, possibly rotated through space, but that as additional quadratic and cubic terms are added additional bumps and warps appear in the surface. Note that these plots depict only the $u(x,y)$ trend in Equation (5). The addition of the $e(x,y)$ terms to the plot may conceal that trend somewhat, but generally inspection of the resulting universal kriging predictions should still reveal the general shape of the trend surface.

As an example, we conduct a simple universal kriging exercise. Consider again the annual average $PM_{2.5}$ data example. Suppose we believe that instead of a constant mean, a quadratic function of the location is the best description of the large-scale spatial trend. The first step in the analysis, as with ordinary kriging, is to identify an appropriate covariance function to describe the data. This is more difficult than in the ordinary kriging case, as the inherent trend in the process can influence the apparent covariance structure. In practice, this can be problematic. The key is to identify a region in space where one believes the trend is reasonably stable. If it appears that the slope in one direction (say north-to-south) is close to zero, but the slope in the perpendicular direction is non-zero, one could develop the empirical variogram and, thus, model the variogram using only data aligned in the north-south direction.

If alternative subsets of the data do not provide such stability, another approach should be considered. For instance, one might estimate the covariance by conducting a preliminary ordinary least squares regression to remove the trend and then estimate the covariance structure from the residuals (i.e., the data with trend removed). By definition, the residuals from such a regression will have a constant trend and, thus, are appropriate, in general, for modeling a variogram. [Unfortunately, one risks the sort of overfitting discussed earlier using this method. This problem is discussed in further detail in Journel and Huijbregts (1978).] Universal kriging estimators are then computed using the original data. That is, the non-detrended data are combined with the variogram model and the form of the trend surface to compute the universal kriging estimate. Cressie (1993) and Ripley (1981) provide formulas and technical details about computing universal kriging estimators given the data and a variogram model. Appendix A of this document reproduces these formulas.

For the purpose of this example, we first fit a quadratic regression (ordinary least squares) to the annual average $PM_{2.5}$ data. Taking the residuals of this regression, we computed a robust empirical variogram. Figure 4.9 illustrates one variogram model plotted against the robust empirical variogram. Observe that the model, while having a smaller objective value than that for the ordinary kriging model in Section 3 (see Figure 3.5), is not as visually appealing. Also, observe that the gamma function values are smaller overall. [Recall that gamma is the empirical variogram value calculated for a given bin of distances, as defined in Section 3.3 and Equation (2).] This is a direct result of the more complicated trend surface included in the universal kriging model. That is, a more complicated trend surface shifts a certain amount of overall data variation away from the variogram model, and explains it through the mean component of the model, i.e., $u(x,y)$.

The variogram model depicted by the line in Figure 4.9 was then incorporated into a universal kriging model where, for the purpose of this example, a quadratic trend was specified. More specifically, the original non-detrended data are used as the data for the final universal kriging model fitting, but the covariance structure specified in this model is based on an analysis of residuals from an initial quadratic regression. A contour plot of the resulting universal kriging surface is displayed in Figure 4.10. Observe that the universal kriging estimate of the surface is visually quite similar to the ordinary kriging estimate discussed in Section 3 (see Figure 3.7).

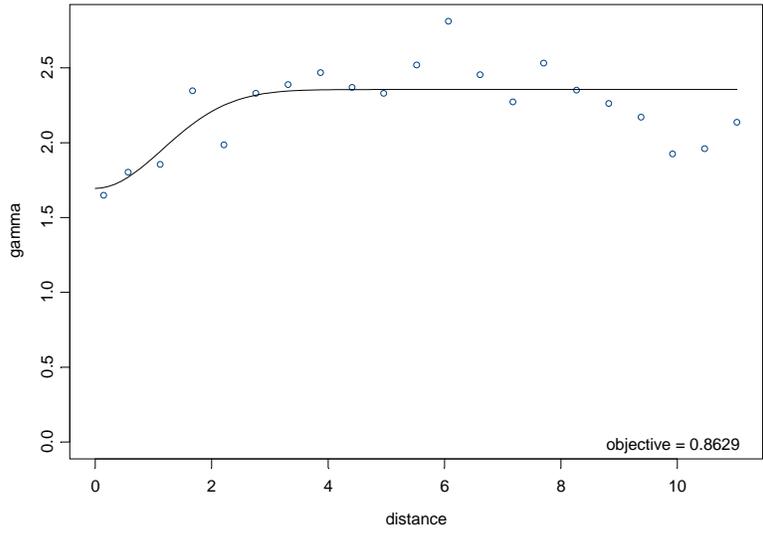


Figure 4.9 Empirical variogram and variogram model for universal kriging model.

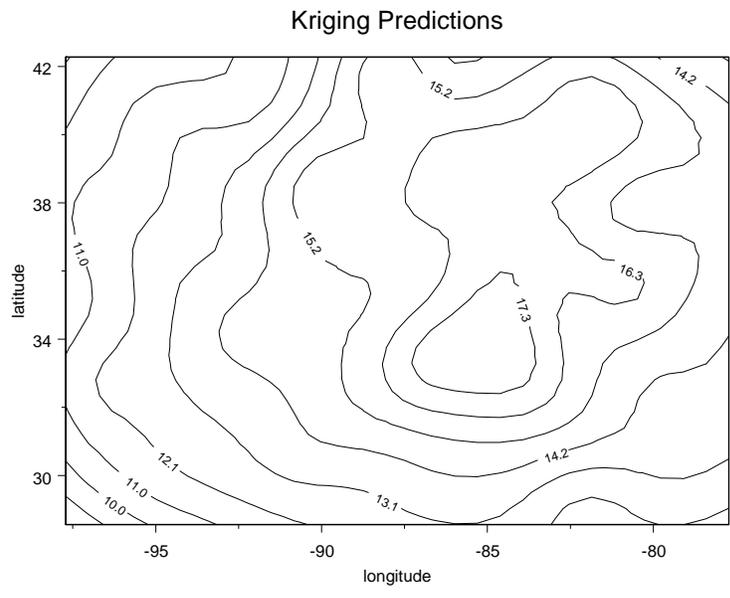


Figure 4.10 Universal kriging predictions based on annual PM_{2.5} data from case study example.

Table 4.3 and Figure 4.11 compare the standard errors and model fits of the ordinary kriging (Section 3) and universal kriging models for the annual average PM_{2.5} data. Table 4.3 presents the five-number summaries (minimum, 25th percentile, median, 75th percentile, maximum) of the spatial prediction standard errors across the predicted surface for the ordinary kriging (Section 3) and universal kriging models. Perhaps surprisingly, the distributions of the standard errors from the two models are very similar. Ordinary kriging appears to have a slightly larger range of standard error values. However, in general, the standard errors for universal kriging appear to be slightly larger than those for ordinary kriging. This suggests that the quality of the fits provided by the two models is similar. As a result, pending additional information, it might be recommended that the simpler model, ordinary kriging, be used to interpolate this spatial region.

Table 4.3 Comparison of percentiles of standard errors between ordinary and universal kriging models for same data set

Model	Prediction Standard Error Percentiles (: g/m ³)				
	Minimum	25th	Median	75th	Maximum
Ordinary Kriging	1.433	1.440	1.448	1.473	1.824
Universal Kriging	1.450	1.510	1.537	1.570	1.815

Figure 4.11 further confirms the above recommendation. This graph represents the contrast between the ordinary kriging (Section 3) and universal kriging spatial predictions. At each predicted location, the universal kriging estimate was subtracted from the ordinary kriging estimate. If the value was less than -0.28 : g/m³, the region was colored dark grey, indicating the ordinary kriging estimate was generally lower than the universal kriging estimate. If the difference was greater than 0.46 : g/m³, the region was colored white, indicating the ordinary kriging estimate was generally higher than the universal kriging estimate. The remaining areas were shaded with a light grey, suggesting that the two models were generally in agreement. These breakpoints were chosen based on the quantiles of the differences. That is, -0.28 : g/m³ represents a value greater than 25 percent of the differences, while 0.46 : g/m³ represents a value greater than 75 percent of the differences. In other words, in terms of absolute difference, the two models (ordinary and universal kriging) agree to within a factor of 0.46 : g/m³ for more than half of the prediction surface. While the high and low regions tend to be clustered, the midrange of the differences is fairly evenly distributed, suggesting that the universal kriging estimate did not detect any important trend features missed by the ordinary kriging model. Thus, relative to the ordinary kriging model of Section 3, it may be reasonable to believe that the universal kriging does not necessarily represent an important revision of the predicted spatial surface.

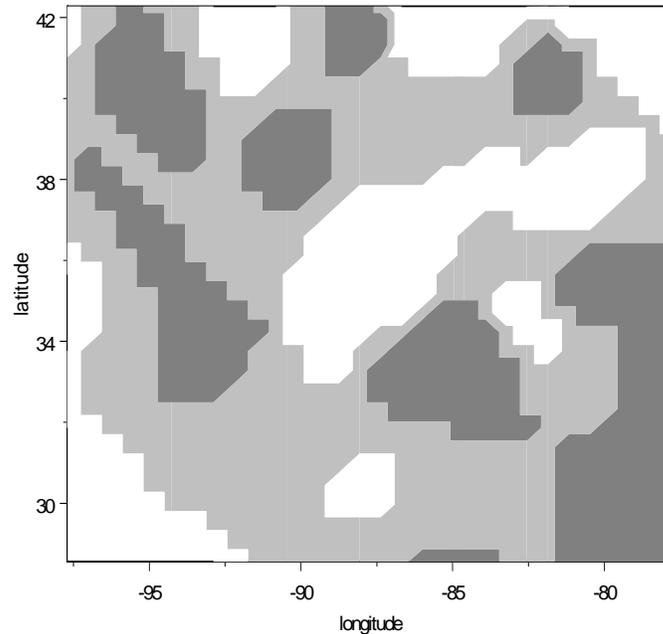


Figure 4.11 Contour plot of feature differences between ordinary and universal kriging interpolations of annual $PM_{2.5}$ data.

4.2.2 Kriging with External Drift

In kriging with external drift (KED), a primary variable is singled out to be predicted while another available variable is treated as a covariate. This notion has been considered by several authors, including Ahmed and De Marsily (1987); Brown, Le, and Zidek (1994); Bourennane, King, Chery, and Bruand (1996); and Gotway and Hartford (1996).

Since the covariate is not necessarily modeled as a spatial process, predictions of the primary variable can be constructed only at sites at which the covariate (covariable) is observed. For predicting at locations for which the covariate is not observed, use of interpolated, predicted, or some manner of imputed missing values has been suggested. These approaches are somewhat ad hoc, however, and in many cases result in underestimation of prediction variances, since the uncertainty of the predicted covariate is not taken into account.

The KED model focuses on conditioning the primary variable on an available covariate (as opposed to co-kriging, discussed in the next section, which focuses on simultaneously kriging

multiple variables). Specifically, the most commonly applied KED model extends the constant mean term of Equation (1), u , as the sum of that constant with another constant times the value of the covariate. Goovaerts (1997) provides an in-depth discussion of this topic including several guidelines. The author observes that the relationship between the primary variable and the external variable should be linear, the value of the external variable must be known at all primary locations and at all locations being predicted, and the external variable should vary smoothly in space. Additional discussions, including the specific formulas associated with KED can be found in Royle and Berliner (1999) and Wackernagel (1995). Note that KED modeling equations can be extended to multiple primary variables and external covariates.

As an aside, the use of an analysis of variance (ANOVA) classification as a so-called external variable is an interesting possibility when developing kriging or other spatial interpolation models. This technique looks at covariates, such as an indicator of different sub-regions or land use, and assigns mean values to certain classes of these covariates. For instance, a rural site might be assigned one mean value and an urban site assigned another mean value. These values might then be used to model the variogram in the process of interest. To the degree that the analysis data set specifies such factors in a non-continuous or polynomial-like fashion (e.g., labeling census tracts as urban or rural), the resulting spatial prediction surface will behave like an abrupt interpolation (see Section 2). This technique is similar to KED (or universal kriging, for that matter) in that it amounts to including external variables or covariates in the model to help explain the primary variable's overall spatial variation. However, unlike with KED, an external variable with an ANOVA-like classification will not necessarily be linearly related to the primary variable and will not necessarily vary smoothly in space.

In general, KED modeling issues are in some ways similar to those presented for universal kriging. Specifically, introducing the (x,y) component to the trend term, $z(x,y)$, in Equation (5) for universal kriging amounts to including an external variable or covariate to help explain the variation in the primary variable. Conceptually, from the KED perspective, the spatial coordinate system, (x,y) , represents an external variable or covariate that is exhaustively sampled and free of measurement error. KED is therefore not discussed in further detail in this document. For further discussion, refer to the numerous references provided within this section. Also, refer to Section 4.2.1 and Appendix A for more details on universal kriging.

4.3 Dimensional Extensions

Our initial example (in Section 3) assumed a single response of interest observed over only two spatial dimensions, with no temporal component. Indeed, the primary scope of this document has been to reduce the spatial interpolation problem to fit within this context. However, it is easy to imagine more complicated scenarios in which one is interested in multiple responses in all three spatial dimensions, or interested in space-time estimates. For example, the Multi-Stage Community Air Quality Modeling System (CMAQ) is an air pollution dispersion model that produces three-dimensional estimates of $PM_{2.5}$ and other pollutants over time. One might be interested in a statistical estimate that provides similar results based on data from $PM_{2.5}$ monitoring stations. Kriging and other spatial interpolators can be extended to interpolate in

three dimensions and time, or to address multiple responses simultaneously. These issues are generally beyond the scope of this document, but are touched on briefly in the following discussions. Specifically, Section 4.3.1 discusses three-dimensional spatial and space-time kriging, while Section 4.3.2 introduces the concept of co-kriging.

4.3.1 3-D Spatial or Space-Time Kriging

In principle, three-dimensional spatial kriging or space-time kriging is conducted in a manner identical to ordinary kriging in two spatial dimensions. First an empirical variogram is computed. Then a variogram model is estimated that well approximates the empirical variogram. Finally, a linear combination of the observations is made, using weights computed based on the variogram model. The equations used to calculate these weights are understandably more complex than those used in the two-dimensional spatial case (see Appendix A), but are well established. Cressie (1993) develops these equations and discusses variogram models for these more complex kriging scenarios.

As with other kriging extensions, the complexity in higher-dimensional kriging comes in the estimation of the variogram model. While the types of models are, in general, natural extensions of the variogram models used in the two-dimensional spatial case, the representations can become more complicated. In space-time cases, one must remain particularly aware of the directional nature of time. Unlike in space, which flows in multiple directions, time flows in a single direction, i.e., from past to present. For example, intuitively one might think of today's daily $PM_{2.5}$ concentration at a given location as being predictive of tomorrow's concentration, but not yesterday's. Nonetheless, some sort of correlation (covariance) structure might be expected among concentrations from different days. Chatfield (1995) discusses the nature of temporal covariance structures in more detail. Spatial interpolation modeling extensions to three-dimensional space or space-time applications are not currently considered in further detail as part of this document.

4.3.2 Co-Kriging

Our primary example used throughout this document considers only a single variable of interest (annual $PM_{2.5}$). However, there are scenarios in which it may be desirable to describe the spatial relationship between two or more variables (and their interactions) based upon the available measurements on all the variables of interest. For example, one may be interested in the spatial distribution of not only fine particulate matter, but also other pollutants such as those classified as hazardous air pollutants (HAPs). In an attempt to study multi-pollutant interactions, monitoring stations may measure not only $PM_{2.5}$, but other pollutants as well. The multiple surfaces described by such multivariate observations can then be interpolated in a manner similar to the ordinary kriging techniques described earlier in this document. In the context of kriging, this process of modeling multiple variables simultaneously is called co-kriging.

In principle, the co-kriging estimates are very similar to those of ordinary kriging (see Appendix A). The predictor of a given covariate at a given location is a linear combination of all

the available data values of all of the covariates of interest. The resulting co-kriging equations are slightly more complicated than those of ordinary kriging, but are still relatively straightforward. For example, a variogram model must be developed for each variable of interest. However, a cross-variogram model that describes the co-variation (or correlation relationship) between each pair of variables is now required as well. Cressie (1993), Journel and Huijbregts (1978), Wackernagel (1995), Meyers (1983), and Goovaerts (1997) all discuss co-kriging, providing technical details and formulas.

Perhaps obviously, the problem of estimating the appropriate covariance functions to use in co-kriging is more complicated. Cressie (1993) discusses the potential problems associated with using the various formulations of cross-variograms. He suggests specific formulations that will work directly with the co-kriging equations. In general, however, building valid, flexible models for cross-variograms and fitting them to the available data is a problem that requires further research. As a result, there does not appear to be much in the way of readily available software packages to assist with this spatial modeling procedure. A notable exception is GSLIB and software based on GSLIB, such as GMS. (See Section 6 for further discussion of GSLIB.)

In conclusion, we have attempted in this section to review the landscape of kriging extensions. These complexities were reviewed to provide the reader with options for those cases when an ordinary kriging model is not adequate to describe the true nature of a given spatial process. We have addressed, to varying degrees, spatial trends, temporal dynamics, non-stationary covariance structures, use of covariates, and multivariate modeling. Given the scope of this document, we have provided a number of references for the interested reader to pursue further as their kriging and spatial interpolation needs dictate.

4.4 References

- [1] Kaluzny, Stephen P., Vega, Silivia C., Cardoso, Tamre P., and Shelly, Alice A. (1998). S+ Spatial Statistics. Springer-Verlag, New York, New York.
- [2] Cressie, Noel A. C. (1993). Statistics for Spatial Data. John Wiley & Sons, New York, New York.
- [3] Devore, Jay L. (1995). Probability and Statistics for Engineering and the Sciences. Duxbury Press, Pacific Grove, California.
- [4] Seber, G. A. F. (1977). Linear Regression Analysis. John Wiley & Sons, New York, New York.
- [5] Walls, R. C., and Weeks, D. L. (1969). "A Note on the Variance of a Predicted Response in a Regression;" *American Statistician* 23 (3), pp 24-25.
- [6] Journel, A. G., and Huijbregts, Ch. J. (1978). Mining Geostatistics. Academic Press, London, England.

- [7] Ripley, Brian D. (1981). Spatial Statistics. John Wiley & Sons, New York, New York.
- [8] Ahmed, S., and De Marsily, G. (1987). "Comparison of Geostatistical Methods for Estimating Transmissivity Using Data on Transmissivity and Specific Capacity;" *Water Resources Research* 23, pp 1717-1737.
- [9] Brown, P. J., Le, N. D., and Zidek, J. V. (1994). "Multivariate Spatial Interpolation and Exposure to Air Pollutants;" *The Canadian Journal of Statistics* 22, pp 489-509.
- [10] Bourennane, H., King, D., Chery, P., and Bruand, A. (1996). "Improving the Kriging of a Soil Variable Using Slope Gradient as External Drift;" *European Journal of Soil Science*, 47, pp 473-483.
- [11] Gotway, C. A., and Hartford, A. H. (1996). "Geostatistical Methods for Incorporating Auxiliary Information in the Prediction of Spatial Variables;" *Journal of Agricultural, Biological, and Environmental Statistics*, 1, pp 17-39.
- [12] Royle and Berliner (1999). "A Hierarchical Approach to Multivariate Spatial Modeling and Prediction;" *Journal of Agricultural, Biological, and Environmental Statistics*, 4(1), pp 29-56.
- [13] Chatfield, C. (1995). The Analysis of Time Series. Chapman & Hall, London, England.
- [14] Wackernagel, H. (1995). Multivariate Geostatistics, Springer-Verlag, Berlin, Germany.
- [15] Goovaerts, P. (1997). Geostatistics for Natural Resources Evaluation, Oxford University Press, New York, New York.
- [16] Meyers, D.E. (1983). "Estimation of Linear Combinations and Co-Kriging," *Mathematical Geology*, 15(5), pp 633-637.
- [17] Gomez, M., and Hazen, K. (1970). "Evaluation of Sulfur and Ash Distribution in Coal Seams by Statistical Response Surface Regression Analysis." U.S. Bureau of Mines, Report of Investigation 7377, Washington, D.C., 120 pages.

5.0 MODEL EVALUATION AND ACCEPTANCE

This document has provided a general overview of spatial interpolation (Section 2), with more focused attention on the spatial statistics approach of ordinary kriging (Section 3) and common extensions to that method (Section 4). Under any spatial interpolation methodology, it is important to evaluate one's candidate model to determine whether it is acceptable for its intended use. For example, Section 3.5 considered various model diagnostics and offered a general diagnostic algorithm for evaluating the ordinary kriging model. In particular, the proposed algorithm recommended considering more complex models, such as those discussed in Section 4, as part of the model diagnostic/evaluation process. Although presented in the context of ordinary kriging and other kriging models, many of the model diagnostics discussed in Section 3.5 are applicable to other types of spatial interpolation models as well.

The current section of this document enhances and goes beyond the concepts presented in Section 3.5. Section 5.1 will present specific performance measures that can be applied following a kriging exercise to evaluate the accuracy of the model, extending the discussion of cross-validation introduced in Section 3.5. Following Section 5.1, we broaden the discussion of model diagnostics viewing them as just one component of a larger process of model evaluation and acceptance. Section 5.2 offers some further perspective on this viewpoint and proposes EPA's data quality objectives (DQO) process as an appropriate framework for model evaluation and acceptance. Section 5.3 provides an overview of the DQO process, with discussion and examples tailored to the current context of developing a spatial interpolation model. Section 5.4 lists references.

5.1. Specific Performance Measures for Model Evaluation

In this section, we present five methodologies for evaluating kriging models. Each of these methodologies is similar to the more familiar methodology for evaluating regression models; residuals (the difference between fitted values and original observations) are obtained and evaluated using a battery of tests. However, in spatial kriging models, the method of obtaining residuals is not obvious. Due to the two-stage model-fitting process for kriging models (first estimating the variogram using the data, then predicting new values using the same data), residuals calculated as the difference between the estimated surface and the observed values can give misleading information. As an example of the most extreme case, in a model with no nugget effect all of the observed values will be interpolated perfectly (i.e., the estimated spatial surface takes the value of the true observation at each location). In this situation, all of the residuals will be zero, but the model may do a poor job of predicting values at unobserved locations. Since these ordinary residuals are of no use, the main focus of the methodologies that follow is on calculating meaningful residuals.

Once these residuals have been obtained, a variety of different tests and evaluation methods can be applied to them. The available tests depend on the method used for obtaining the residuals. We discuss tests and evaluation methodologies in each section after describing each method of obtaining residual values. It is worth noting that these methods will not always help

identify problems with the assumptions of stationarity and isotropy. These assumptions are important, and sometimes their validity is difficult to assess.

Throughout the descriptions of the methods, we use consistent notation. In all cases, $z(s_i)$ denotes the true observation at location s_i , and $\hat{z}(s_i)$ denotes the predicted (kriged) value at that location. The kriging standard error at location s_i , is denoted $\mathbf{F}(s_i)$. We denote the residual at location s_i by $r(s_i)$, and we denote the total number of available observations for analysis by N .

Leave-one-out Cross Validation: Single Variogram

By far, the most popular technique for generating useful residuals in kriging analyses is leave-one-out cross validation. A good review of leave-one-out cross validation may be found in Isaaks and Srivastava (1989). Leave-one-out cross validation residuals are generated using the following procedure:

1. Create an empirical variogram using all of the N available observations.
2. Estimate the theoretical variogram from the empirical variogram.
3. For each observation $z(s_i)$, $i = 1, \dots, N$ in the dataset,
 - a. Remove the observation from the dataset.
 - b. Predict the kriged value $\hat{z}(s_i)$ at the location of the removed observation using the remaining $N - 1$ observations.
 - c. Calculate the difference between the predicted value and the true value, and divide this difference by the kriging standard error:
 $r(s_i) = [z(s_i) - \hat{z}(s_i)] / \mathbf{F}(s_i)$. This form for the residuals is described in Bradley and Haslett (1992).
 - d. Record the value $r(s_i)$ as the standardized residual at the location of the removed observation.

Calculating residuals in this way gives information about how well the model will perform when making predictions at new spatial locations since predictions are made without the benefit of knowing the true value at the predicted location. In addition, the leave-one-out cross validation residuals provide information about whether kriging modeling assumptions are valid and whether standard errors estimated by the kriging model are accurate.

Once the residual at each location has been calculated as described, several different methods can be used to assess the model. The simplest method is to examine a histogram of the standardized residuals. This histogram should give an indication of the types of prediction errors obtained by the model. If several outliers exist, it is an indication that the model is not fitting the data well. In this case, a spatial model may be inappropriate, or a more robust error distribution, such as a t-distribution, might produce a better fit to the data. If the standardized residuals do not appear to follow a symmetric form resembling a normal distribution, some of the kriging model assumptions may be incorrect, and other modeling techniques should be considered. Figure 5-1 is an example of a histogram of residuals with a normal curve overlaid. Q-Q plots of the

standardized residuals are also useful for assessing whether the kriging model assumptions are satisfied.

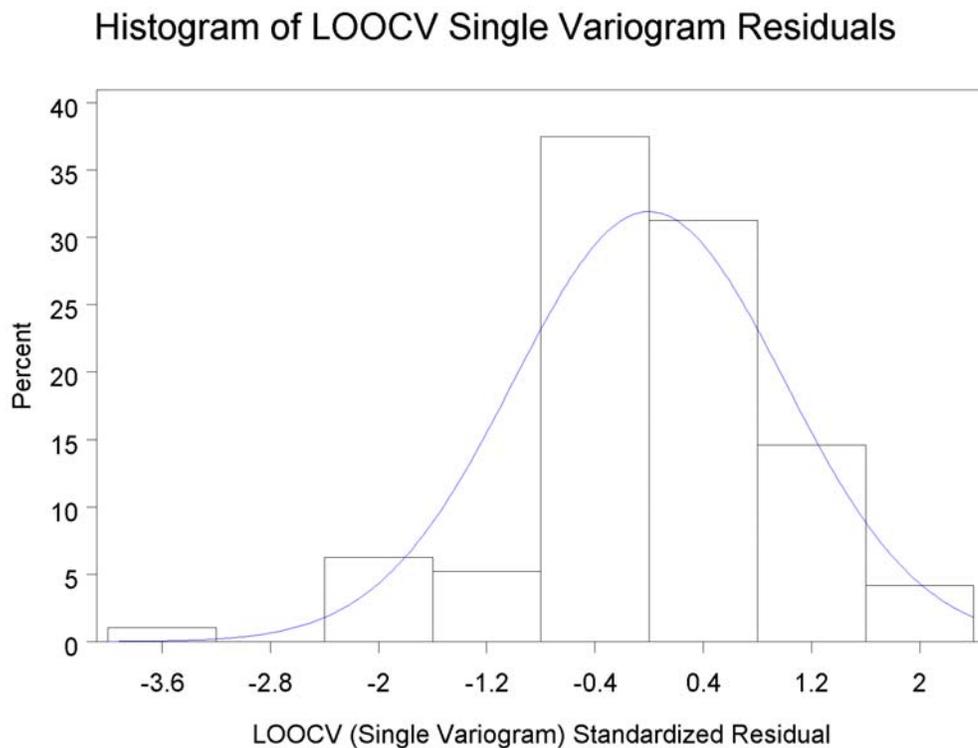


Figure 5.1 Example Histogram of Kriging Residuals

In addition to univariate assessments, the modeler should examine the spatial properties of the cross-validation residuals. One popular method is to construct a plus/minus plot as illustrated in Isaaks and Srivastava (1989). At each location with a positive residual, a plus sign is plotted on a diagram of the field. At each location with a negative residual, a negative sign is plotted. The resulting plot indicates areas of consistent overestimation (locations where clusters of negative signs appear) and underestimation (locations where clusters of positive signs appear). If too many clusters appear, it is an indication that the model is invalid. Another method useful for evaluating the residuals spatially is to plot the residuals against their longitude or latitude separately. These plots can indicate whether a pattern exists along a one-dimensional line. As before, any pattern other than random noise might indicate an invalid model. Figure 5-2 is an example of a plus/minus plot of residuals from an ordinary kriging exercise. In this example, the analyst might be justifiably concerned about the tight clustering of positive residuals (e.g. Atlanta and Birmingham regions) that suggest assumptions of spatial independence are not strictly met.

Using leave-one-out cross validation with a single variogram has many advantages. Leave-one-out cross validation is the most recognized method for obtaining residuals in kriging

analyses; its utility for assessing the validity of spatial models is widely accepted. Also, calculating the residuals is not computationally burdensome [some algorithms exist for calculating these residuals without refitting the kriging model each time; see Green (1985)]. Finally, it is simple to examine graphical summaries since only one residual is produced at each observation location.

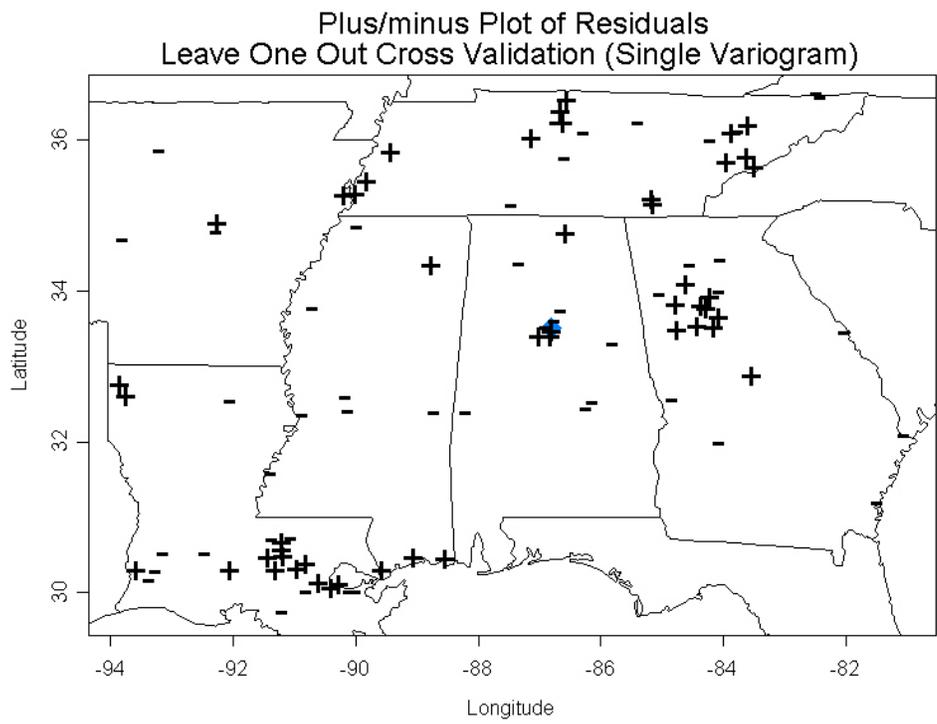


Figure 5.2 Example Plus-Minus Plot

There are, however, a few disadvantages to the technique. The first disadvantage to leave-one-out cross validation is that N observations are used to determine the variogram, but only $N - 1$ observations are used to predict the value at the location of an observation that has been left out. If the goal is to determine how well a new observation will be predicted, it makes no sense to use the observation to be predicted in developing the variogram. The second disadvantage to this technique is that it is susceptible to clustering. If observations appear in clusters, leaving out one observation from a cluster will not have a large effect on the predicted value at the observation's location. As a result, the model will appear overly optimistic about the accuracy of predictions. The final disadvantage is that the residuals obtained using leave-one-out cross validation are inherently correlated with each other (Kitanidis, 1991). As a result of this correlation, residual histograms and Q-Q plots may appear to indicate a violation of kriging assumptions even if nothing is wrong. The appearance of asymmetric residuals can be especially problematic with long correlation distances or small amounts of data. Another result of the correlation in the residuals is that derivation of distributional properties for the mean and

variance of the standardized residuals is complicated. This complication makes objective testing of the model assumptions difficult [though objective testing is still possible, see Kitanidis (1991) for details].

In order to account for these three disadvantages, other methods for calculating residuals have been proposed. To remedy the inconsistency of predicting an observation using a variogram built using that observation, leave-one-out cross validation with multiple variograms can be applied. To remedy the clustering problem, an alternative method known as leave-n-out cross validation is useful. Finally, to create uncorrelated residuals with superior statistical properties (and thus allow more objective assessment of the model), the method of orthonormal residuals has been proposed.

Leave-one-out Cross Validation: Multiple Variograms

Leave-one-out cross validation with multiple variograms was introduced as a way to avoid using more observations for fitting the variogram than for predicting new values. The method is very similar to the single variogram method described above:

For each observation $z(s_i)$, $i = 1, \dots, N$ in the dataset,

1. Remove the observation from the dataset.
2. Create an empirical variogram using the remaining $N - 1$ observations.
3. Estimate a temporary theoretical variogram from the empirical variogram.
4. Predict the kriged value $\hat{z}(s_i)$ at the location of the removed observation using the remaining $N - 1$ observations and the temporary theoretical variogram.
5. Calculate the difference between the predicted value and the true value, and divide this difference by the kriging standard error: $r_i(s_i) = [z(s_i) - \hat{z}(s_i)] / \mathbf{F}(s_i)$.
6. Record the value $r_i(s_i)$ as the standardized residual at the location of the removed observation.

Once the residuals are obtained, the same univariate and spatial methods described for assessing residuals in the single variogram version can be used to assess these residuals. Standardized residuals should again appear somewhat symmetric, outliers should be investigated, and clusters of positive or negative residuals should be evaluated.

The main advantage to the multiple variogram method is that it does not use more data for fitting the variogram than it uses for predicting the value at the new location. As a result, the multiple variogram method should give a more accurate picture of how well new observations will be predicted. Another advantage of the multiple variogram method is that, like the single variogram method, a single residual is produced at each location making spatial assessment of residuals simple.

There are some disadvantages to leave-one-out cross validation with multiple variograms as well. The most obvious problem is that the variogram must be fit separately each time an observation is taken out, so calculating the residuals is computationally intensive. In most

situations, fitting separate variograms also means that the variogram-fitting process must be automated. Also, like the single variogram method, this method is susceptible to clustering effects. The next method, leave-n-out cross validation, was developed as a way to combat this weakness in leave-one-out cross validation strategies.

Leave-n-out Cross Validation

Leave-one-out cross validation, whether performed with a single variogram or multiple variograms, can be misleading when the data are clustered. When data are clustered, removing a single observation has little effect on prediction performance at the location of the removed observation since nearby observations provide most of the information for prediction. As a result, leave-one-out cross validation may produce residuals that are overly optimistic and fail to diagnose model problems (Isaaks and Srivastava, 1989).

Leave-n-out cross validation was developed as a remedy to the clustering problem when calculating residuals. The steps in the method are as follows:

1. Create an empirical variogram using all of the N available observations.
2. Estimate the theoretical variogram from the empirical variogram.
3. For several sets $\{z(s_{(1)}), \dots, z(s_{(n)})\}$ of n observations (here, $s_{(1)}, \dots, s_{(n)}$ are a set of locations in a cluster),
 - a. Remove the n observations from the dataset.
 - b. Predict the kriged values $\{\hat{z}(s_{(1)}), \dots, \hat{z}(s_{(n)})\}$ at the n locations of the removed observations using the remaining $N - n$ observations.
 - c. Calculate the difference between the predicted values and the true values at these locations and divide each by its kriging standard error:

$$r(s_{(i)}) = [z(s_{(i)}) - \hat{z}(s_{(i)})] / \mathbf{F}(s_{(i)})$$
 - d. Record these values as residuals at the locations of the removed observations.

When data are clearly clustered, this method is superior to the previous two. The main advantage of leave-n-out cross validation is that clustering has less influence on the predicted values.

There are several disadvantages to performing leave-n-out cross validation, however. If the data are not clearly clustered, the modeler must choose which observations to remove in each group of n . Although data collection methods may sometimes provide an obvious clustering (e.g., several core samples taken from a single borehole in a mining application), clusters are not always apparent. When clusters are not apparent, a moderate amount of data can provide a large number of possible combinations of n elements to remove. For instance, in an application with only 10 observations, there are 45 unique ways to pick 2 observations to remove. The method of choosing clusters is important because the way in which groups of n observations are chosen can have a strong impact on the residuals obtained.

A second disadvantage to leave-n-out cross validation is that several different residuals $r_i(s_i)$ can be obtained at each location s_i . Since the sets of n observations chosen in Step 3 of the method need not be mutually exclusive, one observation may be left out in several different groups of n observations removed. As a result, conflicting information may be obtained at a single location. For example, a single location may have both positive and negative residuals. In addition, it is difficult to perform graphical summaries and assess model assumptions since multiple residuals calculated at a single location are correlated.

Finally, leave-n-out cross validation shares a disadvantage with the single variogram leave-one-out cross validation in that the variogram is estimated using more information than is used for prediction.

Orthonormal Residuals

The method of orthonormal residuals was proposed by Kitanidis (1991) as an alternative to leave-one-out cross validation for assessing the adequacy of a kriging model. Kitanidis (1991) points out that while distributional properties can be obtained for residuals created using leave-one-out cross validation, they are difficult to calculate. Orthonormal residuals were proposed as an alternative for calculating residuals with better statistical properties. Orthonormal residuals are created as follows:

1. Create an empirical variogram using all of the N available observations.
2. Estimate the theoretical variogram from the empirical variogram.
3. Choose an ordering for the observations. This ordering may be chosen randomly, although in some situations a natural ordering exists. We denote the ordered sites $s_{[1]}, \dots, s_{[n]}$.
4. For each ordered observation $z(s_{[i]})$ starting with the second ordered observation $z(s_{[2]})$ in the dataset and continuing to the last ($i = 2, \dots, N$),
 - a. Predict the kriged value at the location of the observation using the previous observations in the ordering. For example, the value at the location of the second observation $z(s_{[2]})$ is predicted using only the value of the first observation $z(s_{[1]})$, and the value at the location of the third observation $z(s_{[3]})$ is predicted using the values of the first and second observations $\{z(s_{[1]}), z(s_{[2]})\}$.
 - b. Calculate the difference between the predicted value and the true value at the location being predicted and divide this value by the kriging standard error to produce a standardized residual: $r_i(s_{[i]}) = [z(s_{[i]}) - \hat{z}(s_{[i]})] / \mathbf{F}(s_{[i]})$.
 - c. Record the value $r_i(s_{[i]})$ as the orthonormal residual at location $s_{[i]}$.

Note that this method produces $N - 1$ residuals since the value at the first ordered observation location is never predicted. After obtaining the orthonormal residuals, two summary values for the residuals are calculated. The summary values are

$$Q_1 = \frac{1}{N-1} \sum_{k=2}^N \epsilon(s_{[k]})$$

and

$$Q_2 = \frac{1}{N-1} \sum_{k=2}^N \epsilon(s_{[k]})^2$$

If the kriging model is adequate, Q_1 should be approximately normally distributed with mean 0 and variance $1/(N-1)$, and $(N-1)Q_2$ should approximately follow a chi-squared distribution with $N-1$ degrees of freedom. As a result, two hypothesis tests can be performed on the model. If

$$|Q_1| > \frac{2}{\sqrt{N-1}}$$

the kriging model can be rejected. Similarly, if

$$|Q_2 - 1| > \frac{2.8}{\sqrt{N-1}}$$

the model can be rejected. In addition, p-values may be calculated using distributional theory to give a measure of the weight of evidence against the kriging model. In the two preceding equations, the cut-offs for rejecting the model approximate a Type I error level of 5 percent. These approximations should only be applied if $N > 50$. Alternatively, exact cut-off values for rejecting the model can be calculated using the appropriate normal and χ^2 distributions for any value of N .

The above discussion assumes an ordinary kriging model where only a mean parameter is fit. In universal kriging situations, other parameters relating covariates to observations are estimated in addition to the overall mean level. In this case, if the total number of parameters (including the mean) is $p > 1$, p observations are required to fit the model. As a result, the algorithm begins by predicting the $(p+1)^{\text{th}}$ observation given the first p ordered observations and proceeds as described above. Thus, when universal kriging is performed $N-1$ should be replaced with $N-p$ and the summations should begin at $k=p+1$ in the preceding equations. Otherwise, the calculations are the same.

Orthonormal residuals provide several advantages to the modeler. The first advantage is the availability of the Q_1 and Q_2 statistics. The theoretical properties of these two statistics make testing the model simple and objective. Another advantage to orthonormal residuals is that it is simple to examine the residuals graphically. As with leave-one-out cross validation, histograms of residuals and spatial plots of residuals can be obtained and evaluated. Finally, orthonormal residuals are not as susceptible to clustering effects since values are not predicted using all neighboring points except at the end of the ordering.

There are several disadvantages to orthonormal residuals as well. It is impossible to calculate multiple variograms as can be done for leave-one-out residuals since the first few points in the ordering are typically insufficient for estimating a variogram. Secondly, the method of orthonormal residuals is more complex to program than leave-one-out cross validation. Yet another disadvantage is that an ordering must be chosen for the points. Since this ordering is often chosen randomly, different results can be obtained with the same data. The result is that the decision to keep or reject the model may depend on the random ordering of the points which is completely unrelated to the adequacy of the model. Kitanidis (1991) recommends calculating the orthonormal residuals using several different random orderings to evaluate the impact of reordering.

Methods with a Known Covariance

In rare situations, the modeler may actually know the covariance structure of the data. This situation is possible if pilot studies have been conducted to determine the covariance structure or if an experiment has been designed in such a way that the covariance is known. In these situations, kriging is equivalent to a generalized least squares problem with a known covariance matrix (Christensen, et al., 1992).

There are many well known methods for assessing the validity of models that fall into the generalized least squares framework. These methods include the calculation of an R^2 statistic, lack-of-fit tests, hypothesis tests on individual predictor variables (where appropriate), tests for non-constant variance, and visual methods of assessment such as examination of histograms and Quantile - Quantile (Q-Q) plots of residuals. Several types of residuals including standardized residuals and Cook's distance can be calculated and assessed as well. We recommend Weisberg (1985) for an overview of generalized least squares and standard diagnostic techniques.

In addition, residuals can be analyzed spatially using plus/minus plots and other spatial displays of residuals. Spatial displays should be examined because although the covariance matrix may be known, a spatial model still may be inappropriate. Areas of consistent overprediction and underprediction may indicate a problem with the model. If the covariance structure is known to be correct, the modeler may choose to include different covariates in a universal kriging analysis to remedy the problem.

5.2 References

- [1] Isaaks, E. H., and Srivastava, R. M. (1989). An Introduction to Applied Geostatistics. New York: Oxford University Press.
- [2] Bradley, R. , and Haslett, J. (1992). "High-interaction diagnostics for geostatistical models of spatially referenced data," *The Statistician*, 41, 371-380.

- [3] Green, P. J. (1985). "Linear models for field trials, smoothing and cross-validation," *Biometrika*, 72 , 527-537.
- [4] Kitanidis, P .K., (1991). "Orthonormal residuals in geostatistics: model criticism and parameter estimation," *Mathematical Geology*, 23 (5),741-758.
- [5] Christensen, R., Johnson, W., and Pearson, L. M. (1992). "Prediction diagnostics for spatial linear models," *Biometrika*, 79, 583-591.
- [6] Weisberg, S. (1985). Applied Linear Regression. New York: Wiley.

6.0 SOFTWARE TO DEVELOP SPATIAL INTERPOLATION MODELS

Analysts interested in performing spatial interpolation have many choices of products to use to conduct their analyses. A person's choice may depend on their programming ability, as some products require skill with a programming language such as S, SAS, or R. Other products, such as the ArcGIS products developed by ESRI, are more menu-driven and, thus, generate the necessary code behind the scenes. Another driver of choice may be the purpose of the analysis. If the purpose of a spatial interpolation exercise is to produce maps of a predicted surface to aid decision makers, a product that easily generates graphical output may be preferable. On the other hand, if the purpose of an interpolation exercise is theoretical research on new techniques, then a product that provides a user with the flexibility to manipulate all aspects of the modeling process, including considering novel or non-standard techniques, may be preferable.

In the section below, we review the characteristics of multiple geostatistical analysis products that each can perform kriging and other types of interpolation. We also provide references from which more detailed information on the individual products can be obtained. Users can utilize this information to assist with making an informed decision regarding which product best fits their spatial interpolation needs. The products listed in this document, however, are not meant to represent the full universe of products available, but rather a few of the more well-known products to which the majority of users might already have access. There are many other products available, both free and for purchase, including Surfer from Golden Software and GS+ from Gamma Design Software. Detailed information on, and links to, these and many other software packages can be found at the website <http://www.i.ai-geostats.org/>.

Inclusion of a product or technique in this document does not constitute EPA endorsement of any commercial product. Users are free to explore and/or employ other techniques. However, EPA encourages those who chose to do so to evaluate, document, and characterize the uncertainties associated with those techniques at a level of rigor and detail comparable to the treatment given the techniques described in this document.

6.1 SAS

SAS Institute develops and markets various software products that enable users to analyze data and make informed decisions. Below we discuss three SAS products that can be used in performing spatial interpolation and developing graphical output: SAS/STAT, SAS/GRAPH, and SAS/INSIGHT. These products are directly relevant, but they are not meant to represent all SAS products that might be of assistance in an interpolation exercise. Other SAS products might also be useful, such as SAS/GIS, which could be used to develop detailed maps of your data.

Functionality

SAS/STAT provides extensive capabilities for performing a wide-range of statistical analyses. SAS has its own language in which users develop programs that utilize

pre-programmed procedures for performing specific analyses. The most directly relevant procedures within SAS/STAT for performing spatial interpolation include VARIOGRAM and KRIGE2D. The VARIOGRAM procedure has the capability to generate isotropic and anisotropic regular semivariogram, robust semivariogram, and covariance measures. The generated values are written to an output data set. The KRIGE2D procedure then can use that output to perform ordinary kriging. It supports four semivariogram models — Gaussian, exponential, spherical, and power. Note that the KRIGE2D procedure only performs ordinary kriging. Currently, SAS does not contain any standard procedures for performing other types of kriging such as universal or KED. Also note that other procedures within SAS/STAT, such as MIXED for example, may be used for kriging or other spatial interpolation analysis, but may require additional programming complexities to generate the desired quantitative or graphical output.

The KRIGE2D procedure writes kriging estimates and associated standard errors to an output data set. At that point, SAS/GRAPH can be used to graphically display the results with procedures such as GPLOT, GCONTOUR, GCHART, GMAP, and G3D. The variogram model fitting results will also be displayed using the SAS/GRAPH procedures. SAS/INSIGHT is another tool that can be used to graphically explore and analyze data. Some limited interpolation can be performed with SAS/INSIGHT, including a linear interpolation method and a thin-plate spline method; however, a powerful feature is that graphs and analyses are dynamic. A user can spin a three-dimensional plot to explore data from different angles. Individual points can be colored and tracked across analyses. This dynamic aspect of SAS/INSIGHT might make it a useful tool for analyzing spatial interpolation results.

The SAS system can be run on many different platforms including PC, Unix, mainframe, and OpenVMS.

Example Code

The SAS code below is an example of a program that performs a variogram analysis and plots the variogram model's fit to the data, and creates a contour plot of the kriging predicted surface based on the variogram model. As suggested above, those with little to no SAS programming experience may have difficulty interpreting and applying such code.

```
* Estimate variogram parameters
proc variogram data=in.pm25_2000_4qtrcmplt_aa_unique outv=outv;
  compute lagd=0.04409 maxlag=275 robust;
  coordinates xc=latdd yc=londd;
  var aaconc;
run;

title 'OUTVAR= Data Set Showing Variogram Results';
proc print data=outv label;
  var lag count distance variog rvario;
run;
```

```

data outv2; set outv;
  vari=variog; type = 'regular'; output;
  vari=rvario; type = 'robust'; output;
run;

```

*Use the Gaussian Semivariogram Model with nugget;

```

data outv3; set outv;
  n0=1.8; c0=18; a0=16;
  vari =n0+ c0*(1-exp(-distance*distance/(a0*a0)));
  type = 'Gaussian'; output;
  vari = variog; type = 'regular'; output;
  vari = rvario; type = 'robust'; output;
run;

```

title 'Theoretical and Sample Semivariogram for Annual Data(Gaussian with nugget)';

```

proc gplot data=outv3;
  plot vari*distance=type / vaxis=axis2
      haxis=axis1;
  symbol1 i=join l=1 c=cyan v=triangle ;
  symbol2 i=join l=1 c= blue v=dot ;
  symbol3 i=join l=1 c= yellow v=square ;
  axis1 minor=none
    label=(c=black 'Lag Distance') /* offset=(3,3) */;
  axis2 minor=none
    label=(angle=90 rotate=0 c=black 'Variogram')
    /* offset=(3,3) */;
run;
quit;

```

*After an appropriate variogram model is chosen, local kriging is performed with a search radius of 6.

```

*(a) use 16*22=352 grid cells;
proc krige2d data=in.pm25_2000_4qtrcmlt_aa_unique outest=est;
  pred var=aaconc r=6;
  model nugget=1.8 scale=18 range=16 form=gauss;
  coord xc=latdd yc=lond;
  grid x=28 to 43 by 1
      y=-77 to -98 by -1;
run;

```

```

proc g3d data=est;
  title 'Surface Plot of Kriged Annual data(Gaussian with nugget, 352 grid cells)';
  scatter gxc*gyc=estimate / grid;
  label gyc = 'longitude'
      gxc = 'latitude'
      estimate = 'Concentration';
run;

```

*Plot the standard errors (they are smaller where more data are available);

```

proc g3d data=est;
  title 'Surface Plot of Standard Errors of Kriging Estimates(Gaussian with nugget, 352 grid cells)';
  scatter gxc*gyc=stderr / grid;
  label gyc = 'longitude'
      gxc = 'latitude'

```

```

        stderr = 'Std Error';
run;

*create contour plots of the kriged estimates and standard errors

*(a) use 352 grid cells;
goptions reset=goptions noprompt device=cgmof97l ftext=swissl htext=1.3
        rotate=landscape
        gsfname=grafout gsfname=replace;

filename grafout "C:\Est_352.cgm";

axis1 order=(28 to 43 by 3) minor=none;
axis2 order=(-98 to -77 by 3) minor=none;
proc gcontour data=est;
    title 'Contour Plot of Kriged Annual data(Gaussian with nugget, 352 grid cells)';
plot gyc*gxc=estimate/ grid haxis=axis1 vaxis=axis2 levels=10 to 17 by 0.5;
    label gyc    = 'longitude'
          gxc    = 'latitude'
          estimate = 'Estimate'
    ;
run;
quit;

goptions reset=goptions noprompt device=cgmof97l ftext=swissl htext=1.3
        rotate=landscape
        gsfname=grafout gsfname=replace;

filename grafout "C:\STE_352.cgm";

axis1 order=(28 to 43 by 3) minor=none;
axis2 order=(-98 to -77 by 3) minor=none;
proc gcontour data=est;
    title 'Contour Plot of Standard Errors of Kriging Estimates(Gaussian with nugget, 352 grid cells)';
plot gyc*gxc=stderr/grid haxis=axis1 vaxis=axis2
    levels=1.354,1.356,1.358,1.36,1.362,1.364,1.366,1.368, 1.37, 1.38, 1.39, 1.40,1.45, 1.50, 1.60, 1.70,
    1.80, 1.90;;
    label gyc    = 'longitude'
          gxc    = 'latitude'
          stderr = 'Std Error'
    ;
run;
quit;

```

Graphics Capability

Graphics can be produced using the stored procedures in SAS/GRAPH. Presentation options regarding colors, titles, legends, etc., can be set using appropriate SAS commands (as suggested by the GPLOT and GCONTOUR code above). In order to utilize this functionality, non-experienced users will likely need to review the appropriate documentation in SAS manuals,

help pages, or other sources. For users who might prefer a more menu-driven tool, SAS/INSIGHT can also be used to create and analyze graphical output.

Strengths

SAS is a very broad, powerful data analysis package with many statistical and graphics capabilities. Performing interpolation exercises in SAS places the user in an environment within which other functionality outside spatial analysis can easily be exploited. SAS is widely used and well supported. Thus, it provides users with the security of working in a stable environment that will continue to be maintained and improved for the foreseeable future.

Weaknesses

Performing analyses with some SAS products requires knowledge of the SAS programming language. Although this is not a problem for experienced users, new users or non-regular users may be intimidated by the challenge of learning the language. The reader may assess this issue for him/herself based on the example code provided above. Another potential weakness is that the KRIGE2D procedure currently only performs ordinary kriging. For users who need to employ other kriging methods, this may be a limitation; otherwise, other procedures in SAS, such as MIXED, would need to be programmed manually to fit the desired model.

Availability

SAS software products are available for purchase from SAS Institute. Detailed information on SAS products and services is available at www.sas.com or by calling 919/677-8000.

Support

SAS provides documentation with the purchase of its products. In addition, SAS customer support can be contacted for assistance with specific questions. SAS offers training on specific types of analyses at various locations around the U.S.

6.2 S-Plus

The standard S-Plus software and the S-Plus spatial module (referred to as S+SpatialStats) developed by Insightful Corporation provide a full set of tools for performing various spatial interpolation exercises. While the standard S-Plus software allows spatial interpolation to be performed through the use of the S programming language, S+SpatialStats provides menu-driven access to a wide set of interpolation and graphical analysis tools. Because of its relative ease of use, this section will focus on the S+SpatialStats module.

Functionality

The S+SpatialStats module, commercially available for both the Windows and Unix computing environments, contains a comprehensive suite of tools for performing spatial interpolation. S+SpatialStats provides a graphical user interface (GUI) that allows users to perform spatial data analysis and interpolation without actually writing the S code. The S+SpatialStats Version 1.5 Supplement manual from Insightful [2] does, however, provide a function reference in the appendix that details the format and example code for the functions utilized by the interface.

The manual identifies three classes of spatial data that the software is designed to analyze: geostatistical data, lattice data, and point pattern data. The text S+ Spatial Stats by Kaluzny, et al. [3], describes the following functionality in the area of geostatistics:

1. “Estimate and visualize standard or robust, omnidirectional or directional variograms;
2. Model empirical variograms; fit theoretical variogram models to empirical data;
3. Perform ordinary kriging to obtain point estimates for unmonitored locations and kriging prediction variances;
4. Perform universal kriging to model large-scale trend while calculating prediction.”

Five common theoretical variogram models are supported: exponential, spherical, Gaussian, linear, and power. As noted in the quote from the Kaluzny text, S+SpatialStats provides dialog boxes to assist users with performing ordinary and universal kriging. It does not include pre-programmed options for performing other types of kriging. There also does not appear to be a built-in function for performing cross-validation.

Example Code

As mentioned, users perform analyses in S+SpatialStats with the assistance of a graphical interface that utilizes menus and dialog boxes to generate the S code necessary to perform the calculations. Figure 6.1 is an example of the S+SpatialStats dialog box that will generate ordinary kriging predictions. Users also have the option to use the command line within the standard S-Plus package to perform analyses. Listed below is sample S code based on examples from the S+SpatialStats manual that plots an empirical variogram and performs a universal kriging exercise.

Fit a theoretical variogram model to an empirical variogram:

```
vg.iron <- variogram(residuals ~ loc(easting, northing), data=iron ore)
vfit.iron <- variogram.fit(vg.iron, param=c(range=8.7, sill=3.5, nugget=4.8), fun=spher.vgram)
plot(vg.iron)
plot(vfit.iron, add=T)
```

Perform universal kriging:

```
# krige the Iron Ore data with a quadratic trend in the x direction using a spherical covariance function
kiron <- krige(iron ~ loc(x,y) + x + X^2, data = iron ore, covfun = spher.cov, range=8.7, sill=3.5,
  nugget=4.8)
# predictions over default 30 x 30 grid
piron <- predict(kiron)
# plot prediction surface
wireframe(fit ~ x * y, data=piron, screen=list(z=300,x=-60, y=0), drape=T)
```

(See the caution regarding S-Plus’s definition of range in Section 3.3.)

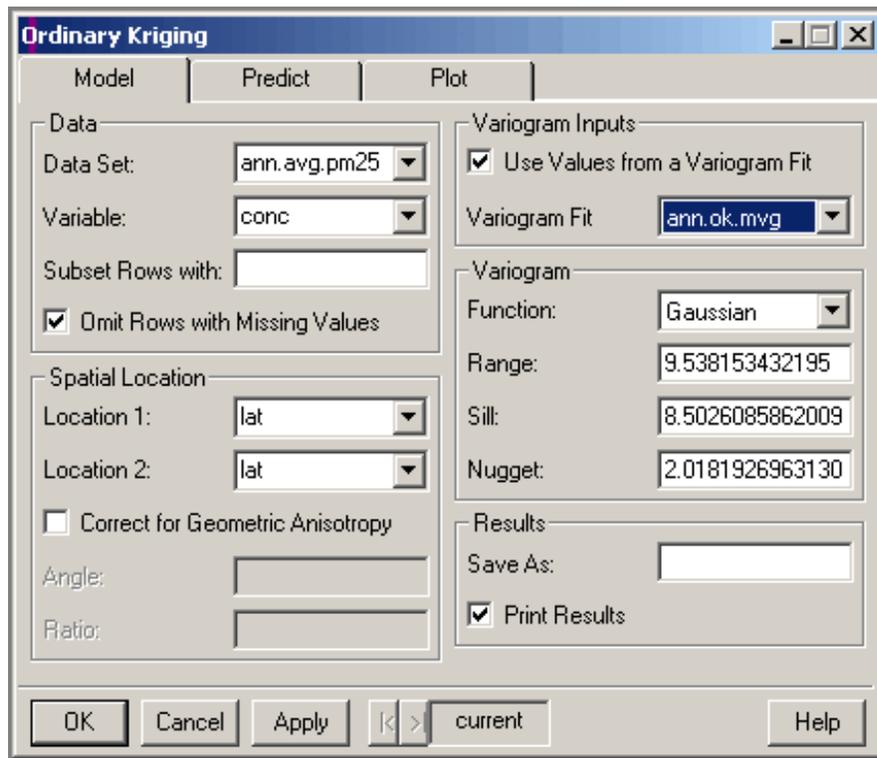


Figure 6.1 S+SpatialStats dialog box for ordinary kriging.

Strengths

Experienced S users have the flexibility to write code and run it from the S-Plus command line or to utilize the menu-driven system to generate the code automatically. The Insightful website states that the S language (developed by Bell Labs) is the only programming language “designed specifically for data visualization and modeling.” S has over 4,200 built-in functions that provide users with a wide array of tools for analyzing and displaying data.

Weaknesses

Performing spatial data analysis and interpolation outside the S+SpatialStats module requires knowledge of the S programming language. Regarding kriging, S-Plus currently only has standard functions for ordinary and universal kriging. Functions are not available for other types of kriging.

Availability

S-Plus and S+Spatial Stats can be purchased from Insightful Corporation. Their products can be viewed at <http://www.insightful.com/products/default.asp>.

Support

Customers can request technical support from Insightful Corporation via e-mail (support@insightful.com), telephone (1-800-569-0123), or fax (206-283-8691).

6.3 Venables and Ripley

As part of their text, Modern Applied Statistics with S-Plus, Third Edition (1999), W.N. Venables and B.D. Ripley provide software libraries that enhance S-Plus. The software tools are also available from Brian Ripley's website at <http://www.stats.ox.ac.uk/~ripley/>. Note that a 4th Edition of Venables and Ripley's text was published in July 2002, Modern Applied Statistics with S, but had not been reviewed at the time this document was prepared.

Functionality

The current suite of programs included with Venables and Ripley's text (3rd Edition) [1] covers a wide range of statistical topics from survival analysis to smoothing methods to regression techniques. A few libraries are relevant to spatial data analysis such as the spatial point process library and penalized spline smoothing library. With improvements that have been made to S-Plus in the area of spatial interpolation, Venables and Ripley's programs in this area are not as extensive as they were with previous editions of the text. With the first Edition, for example, Venables and Ripley included programs for plotting empirical variograms and performing a kriging interpolation. As seen in the discussion of S-Plus in Section 6.2, those programs are no longer necessary.

Strengths

For experienced S users, Venables and Ripley provide additional functions that can be utilized within S-Plus.

Weaknesses

Venables and Ripley do not specialize in or specifically focus on spatial interpolation. They provide software to perform many different types of analyses. Finding the programs that focus on spatial interpolation from Ripley's website is not straightforward. Of the programs that are currently included on Ripley's site, while some appear to perform spline or polygonal interpolation, it does not appear that any focus specifically on kriging.

Availability

As mentioned, the programs are available for free from Ripley's website provided above.

Support

Ripley provides his e-mail address (ripley@stats.ox.ac.uk) and other contact information on his website; however, as he is a professor at the University of Oxford in England and an accomplished author, it is possible that he may not have time to answer all questions about his software.

6.4 Fields

Fields is a collection of programs developed by various researchers at the Geophysical Statistics Project (GSP) within the NSF-funded National Center for Atmospheric Research. Doug Nychka is the GSP project leader. Fields evolved from an earlier suite of S programs known as FUNFITS, used for fitting curves and surfaces to data. Fields is written in the R programming language. The GSP website states that R was selected as the programming language of choice because it is "widely supported, existing S code migrates trivially, and R packages are easily built, distributed, and installed." Additional information on the R statistical programming language is available from the R Project website (<http://www.r-project.org/>).

Functionality

Potential covariance functions provided by Fields include exponential, Gaussian, spherical, cylindrical, Poisson, and others. Users also have the ability to define their own covariance model as an S function. The vgram function performs a least-squares fitting of the variogram to aid in identifying the best covariance function. The major interpolation methods included within Fields are a Thin Plate Spline regression and kriging. The Krig function allows a user to perform either ordinary or universal kriging. It takes the data and the specified covariance function and returns the kriging predictions. Fields also includes generic functions that enable display and analysis of the data. Two examples are the Surface function, which displays the fitted surface, and the Predict.se function, which displays the prediction standard errors.

Fields allows specification of a few different methods for performing "generalized cross-validation." Methods differ based on the points they exclude during the cross-validation

process. Fields is available for the Unix, Linux, and Windows operating systems. Fields is no longer supported for S-Plus; however, an older version may still be downloaded.

Example Code

Below is an example of some Fields code that produces a kriging surface.

```
#2-d example
# fitting a surface to ozone
# measurements. Range parameter is 10 (in miles)
fit <- Krig(ozone$x, ozone$y, exp.cov, theta=10)
summary( fit) # summary of fit
plot(fit) # diagnostic plots of fit
surface( fit, type="C") # look at the surface
out.p<- predict.surface.se( fit)
image(out.p)

# predict at data predict( fit)
# predict on a default surface to make a surface object
out<- predict.surface( fit)
persp( out)

# predict at arbitrary points
xnew<- cbind( c( 10, 20), c( -10, 15))
predict( fit, xnew)

# standard errors of prediction based on covariance being OK.
predict.se( fit, xnew)
```

Graphics Capability

The Fields demo/tutorial on the GSP website (<http://www.cgd.ucar.edu/stats/Software/Fields/fields.demo.shtml>) includes sample code and graphics produced from that code. The graphics are sophisticated and the presentation is flexible for an S or R programmer who can manipulate the various graphics options.

Strengths

A strength of Fields is that was developed by researchers who specialize in the field of geostatistical analysis. Thus, it contains some sophisticated tools to specifically aid spatial data analysis exercises. As the researchers at the GSP continue to perform research into new methods and specific applications, it is likely that the Fields software will continue to be updated with new and improved functionality.

Weaknesses

One weakness of the Fields programs is the limited set of analyses that currently can be easily performed. Users may consider another weakness to be the fact that the software was

developed and is maintained by researchers whose primary job is to perform research, not to provide assistance to Fields users with questions.

Availability

The Fields software is free and can be downloaded from the GSP website (<http://www.cgd.ucar.edu/stats/Software/Fields/>).

Support

The GSP website provides various types of documentation to assist users. An on-line manual is available that discusses each of the major methods provided, including the Krig function. Examples of spatial data analysis exercises using Fields software are also provided in an on-line demo/tutorial. The R Project website contains documentation and manuals on the R language.

6.5 EPA Design Interface (EPA-DI)

EPA Design Interface (EPA-DI) software is a library of S-Plus functions and a graphical user interface that run with the S+SpatialStats module discussed in Section 6.2. EPA-DI allows a user to predict a spatial process, quantify the accuracy of the predictions, and optimize network design by allowing users to assess the effects of adding or deleting monitoring locations.

Functionality

The covariance functions available for use in EPA-DI are the same as those available in the S+SpatialStats module: Gaussian, exponential, spherical, linear, and power. For kriging, EPA-DI also utilizes the S+SpatialStats kriging functionality, which is the S-Plus function krige. Krige performs ordinary or universal kriging.

The unique functionality of EPA DI is the ability to edit network locations by adding and/or removing monitoring locations with a mouse click, and subsequently analyze the effects on the prediction standard errors. A user does this by first plotting the network and then clicking on locations to add or remove locations.

Regarding cross-validation, the Network Analysis module returns two types of residuals: trend residuals and cross-validated residuals. These can be compared to evaluate model or network quality. Because EPA-DI runs with S+SpatialStats, and S+SpatialStats runs on both Unix and Windows systems, EPA-DI can also be used with Unix or Windows operating systems.

Example Code

Most of the functions within EPA-DI are accessed via dialog boxes and menus. There are some features, however, that must be run by typing S code in the command line. These features are clearly labeled in the EPA-DI Tutorial and Help Documentation. The S code is similar to the example presented in Section 6.2.

Graphics Capability

Within EPA-DI, graphics can be generated using the menus and dialog boxes. Following fitting of a kriging model, users can select the types of plots they would like to use to view the kriging predictions. Available plots include contour plots, filled contour plots, and 3-D surface plots. Users can utilize the context menu to change the look of a plot by changing colors, changing contour spacings, and adding titles or text. Users are also able to easily add geographic (state and county) boundaries to plots. Other menu options within EPA-DI such as Interactive Network Plot and Redesign Network allow users to interact with individual measured points on the predicted surface to gain additional information on those points or to select them for removal. Within the Redesign Network dialog box, a user can specify a radius that acts as a tolerance for determining the areas in which monitoring stations should be added to a network and areas from which stations should be removed.

Strengths

For users that have a specific need to evaluate monitoring networks, EPA-DI provides some easy-to-use tools. They allow users to analyze the spatial predictions generated by a particular monitoring network and manipulate the network to assess the effects of adding or removing monitoring locations. EPA-DI also allows users to easily conduct a variogram analysis and generate kriging predictions using ordinary or universal kriging.

Weaknesses

Performing some functions within EPA-DI does require interaction with S-Plus, and typing and running S commands. There is guidance on this in the EPA-DI documentation, but users may need to have some skill with programming in S to implement everything they require. EPA-DI does require a user to have S-Plus and S+SpatialStats installed, which might be a limitation for some potential users.

Availability

EPA-DI is available from EPA on the SCRAM bulletin board. The original DI software created by researchers at the GSP is free and can be downloaded from the GSP website (<http://www.cgd.ucar.edu/stats/Software/DI/>) .

Support

EPA has EPA-DI Tutorial and Help Documentation available to assist with using the EPA-DI software.

6.6 ArcGIS Software by ESRI

ESRI offers multiple geographical information systems (GIS) products that can be utilized in performing spatial interpolation. ArcGIS Spatial Analyst software provides various tools to perform spatial data analysis including some limited interpolation functionality. [2] ArcGIS Geostatistical Analyst provides a full set of tools for performing both spatial data analysis and interpolation. [3] The output from these two products can be utilized in other ArcGIS products such as ArcGIS 3D Analyst.

Functionality

The Spatial Analyst software offers three types of interpolation methods: IDW, spline, and kriging. For each of the methods, the process of generating the predicted surface is a matter of drop-down menus and dialog boxes. For example, to perform an IDW interpolation, a user selects (1) the input data set, (2) the power, (3) the search radius (if fixed, number of points and maximum distance must be inserted), and (4) optionally, a barrier that limits the search area for measured data points. Although Spatial Analyst provides five semivariogram models to choose from in its Kriging procedure (circular, exponential, Gaussian, linear, and spherical), it does not perform the variogram analysis. That analysis must be performed outside the Spatial Analyst software. Users can select to perform either an ordinary or universal kriging procedure. Spatial Analyst offers two types of spline procedures — a Regularized method that creates a smooth surface with gradual changes and a Tension method that is less smooth and more tightly constrained by the measured values.

The Geostatistical Analyst software contains much more extensive functionality for performing spatial interpolations. Methods include polygonal (Voronoi mapping), IDW, global and local polynomial interpolation, and five types of radial basis functions or spline methods, as well as various types of kriging (ordinary, universal, simple, indicator, probability, disjunctive, and co-kriging). Geostatistical Analyst provides a Geostatistical Wizard tool that assists users with fitting a model to a semivariogram. The wizard contains eleven models from which to select including the standard circular, exponential, Gaussian, and spherical models, as well as the less frequently used tetraspherical, pentaspherical, rational quadratic, hole effect, K-bessel, J-bessel, and stable models.

The software allows for the identification and removal of global and local trends from the data through the use of Trend Analysis and Detrending tools, respectively. Geostatistical Analyst provides a cross-validation comparison function that allows users to compare how a single model performs with different parameters or how one method performs versus another method.

The ArcGIS products are only available for the Windows operating system.

Example Code

There is no example code for the ESRI products, as they are menu and dialog box driven. Figure 6.2 contains a sample dialog box from the Spatial Analyst software that allows users to select various options in preparation for creating a kriging predicted surface.

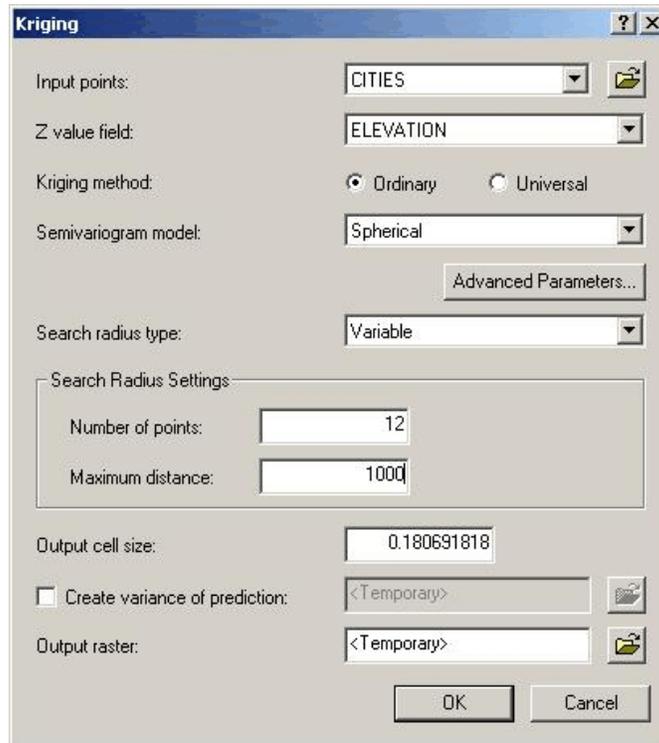


Figure 6.2 ArcGIS Spatial analyst kriging dialog box.

Graphics Capability

The graphics generated by the ArcGIS packages are relatively straightforward to produce and are of high quality.

Strengths

The ArcGIS products are relatively easy to use because of their menu/dialog-box driven nature. Users are essentially walked through the process of performing a spatial interpolation.

Computer programming knowledge is not necessary to use these packages as the graphics are generated through the menu system.

Weaknesses

Users are limited to the functionality provided within the dialog boxes that drive the various features. If a user needs to manipulate the finer details of a model or perform non-standard analyses, they may not find the necessary flexibility within the ArcGIS products.

Availability

ESRI's products can be purchased from their website at <http://www.esri.com/index.html>.

Support

ESRI also sells users manuals for their products from their website. For more extensive instruction, ESRI offers instructor-led and web-based training courses. Customer support is available for purchase.

6.7 GSLIB

GSLIB is a collection of geostatistical programs developed by researchers at Stanford University's Center for Reservoir Forecasting (SCRF) within the School of Earth Sciences. GSLIB is not a commercial product, which means it is not only free, but also not supported. Although it is not supported, two of its developers at Stanford have written a user's guide, which provides guidance on how to use the GSLIB programs. [4] The source code for GSLIB, written in the FORTRAN programming language, and also some help pages are available from the Stanford University Center for Reservoir Forecasting website.

Functionality

GSLIB version 2.0 allows for five different variogram models to be considered in the kriging programs: spherical, exponential, Gaussian, power, and hole effect. Kb2d is a program that can utilize those five variogram models to perform simple and ordinary kriging. Kt3d is a program that can perform stationary or nonstationary simple kriging, ordinary kriging, or kriging with external drift. It allows for up to nine drift indicators. There are additional programs to perform co-kriging (cokb3d) and indicator kriging (ik3d). The Option parameter within kt3d allows for cross-validation to be performed on a data set.

There are no restrictions on the type of operating system on which the GSLIB programs can be run. One restriction, however, is that the computer running GSLIB must have a compiler for ANSI standard Fortran 77 (or any later release).

Example Code

The way the GSLIB works is that the default programs read in a parameter file specific to each program. The user must set the values of all of these parameters before running the program. An example parameter file for the kt3d program (copied from Figure IV.6 of the user's guide) is listed below:

Parameters for kt2d

```
START OF PARAMETERS:
../data/cluster.dat          \file with data
1      2      3             \columns for x, y, and variable
-1.0e21      1.0e21         \trimming limits
3                                     \debugging level: 0,1,2,3
kt2d.dbg                \file for debugging output
kt2d.out                \file for kriged output
5      5.0      10.0       \nx,xmn,xsiz
5      5.0      10.0       \ny,ymn,ysiz
1      1                                     \x and y block discretization
4      8                                     \min, max data for kriging
20.0                                     \maximum search radii
1      2.302                \0=SK, 1=OK (mean if SK)
1      2.0                  \nst, nugget effect
1      8.0      0.0      10.0      10.0    \it, c, azm, a_max, a_min
```

WinGslib is a software tool that provides a graphical interface for setting parameters. Battelle Memorial Institute has also developed a graphical interface for running GSLIB called BATGAM. Battelle will provide the software upon request.

Graphics Capability

GSLIB contains a set of programs that generate various types of statistical and geostatistical plots. Programs are available for creating histograms, scatter plots, semivariograms, and gray-scale maps. The GSLIB programs generate graphics in PostScript format that can be directly printed or previewed. It also contains a function that allows between one and eight plots to be automatically displayed on a single page.

Strengths

GSLIB contains a full set of geostatistical functions. It provides a user with the ability to perform many different types of kriging, which might be appealing to more experienced researchers.

Weaknesses

Working with GSLIB in the standard environment with no graphical interface may be difficult for some users. The process of editing parameters, then compiling and running programs might be unfamiliar to users who have not programmed in Fortran or similar languages. The use of a graphical interface, such as with WinGslib, might make the environment and process less daunting.

Availability

The suite of GSLIB programs is available for free from the SCRF website at <http://pangea.stanford.edu/PetEng/SCRFweb/supporting/index.html>. The GSLIB 2.0 software, written in Fortran 77, is included on a CD-ROM with the purchase of the user's guide. GSLIB 2.907, the most recent version of GSLIB, is written in Fortran 90 and is available for download at <http://www.gslib.com/>.

WinGslib software provides a front-end to the GSLIB programs. It is available for purchase for approximately \$250 per license from Staios, LLC, a private company founded in part by Clayton Deutsch. See the Staios website at <http://www.staios.com/Staios/index.html> to purchase WinGslib, download the latest version of the GSLIB software, or register for on-line or in-house Geostatistics training.

Support

Information and guidance regarding compiling programs, file formats, and specific programs is available at the SCRF website at <http://pangea.stanford.edu/PetEng/SCRFweb/GSLIB/gslibhlp.html>. Also, as mentioned above, a user's guide written by Stanford researchers Clayton Deutsch and Andre Journel is available for purchase from Oxford University Press.

6.8 Summary

To repeat what was mentioned in the introduction to Section 6, a user's spatial interpolation software selection likely will be determined by factors such as the purposes of the interpolation exercise, the skills of the analyst, and the user's access to certain software packages. Software choices certainly go beyond those discussed in this section, but this should provide an introduction to the features contained in some of the more well-known packages. These features can then be compared to those of other packages that a user might be considering. Table 6-1 below summarizes the seven packages discussed in this section.

Table 6.1 Summary of Software Packages Reviewed in Section 6

Software	Contact Info	Main Strength	Main Weakness
SAS	www.sas.com	Multi-purpose statistics package	Knowledge of SAS language required
S+SpatialStats	www.insightful.com	Many tools in menu-driven format	Some knowledge of S required for advanced applications
Venables & Ripley	www.stats.ox.ac.uk/~ripley/		Knowledge of S required
Fields	www.cgd.ucar.edu/stats/Software/Fields/	Developers specialize in spatial analysis & interpolation	Knowledge of R/S required; limited support
DI (Original) EPA DI	www.cgd.ucar.edu/stats/Software/DI/ www.epa.gov/ttn/scram/	Useful for monitoring evaluation	Limited range of uses
ArcGIS	www.esri.com	Menu-driven format, easily generated graphics	Inflexible for non-standard applications
GSLIB	http://pangea.stanford.edu/PetEng/SCRFweb/supporting/index.html	Full set of geostatistical tools	Fortran environment

6.9 References

- [1] W.N. Venables and B.D. Ripley (1994). Modern Applied Statistics with S-Plus, Springer-Verlag, New York, New York.
- [2] McCoy, Jill, and Johnston, Kevin (2001). Using ArcGIS Spatial Analyst. ESRI Press, Redlands, California.
- [3] Johnston, Kevin, Ver Hoef, Jay M., Krivoruchko, Konstantin, and Lucas, Neil (2001). Using ArcGIS Geostatistical Analyst. ESRI Press, Redlands, California.
- [4] Deutsch, Clayton V., and Journel, Andre G. (1998). GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, New York, New York.

7.0 KRIGING MODEL EXAMPLES

We consider two primary examples in this section. In the first, we use S-Plus to identify an optimal kriging estimate of the annual average $PM_{2.5}$ data set introduced earlier in this document. In the second example, we use SAS to perform a similar analysis on an ozone data set.

7.1 $PM_{2.5}$ Annual Average Data

For the $PM_{2.5}$ case study example presented throughout this section, the annual average analysis data set summarized in Table 3.1 was generated from an initial data set of 24-hour averages. (As stated previously, this document does not necessarily advocate data pre-processing via temporal averaging, but this approach was taken for the $PM_{2.5}$ case study example for illustrative purposes. Recall that temporal averaging can have multiple effects, including averaging out any potential geometric anisotropies associated with seasonal prevailing winds.) Site-specific annual averages were calculated as the arithmetic mean of the 24-hour average data collected at the given site throughout calendar year 2000. The majority of the sites included in the case study example data set provided data consistent with a sampling schedule of once every three days, yielding approximately 120 daily observations for calculating a site's annual average concentration.

The goal of this example analysis is to generate a spatial surface that approximately represents the true annual average $PM_{2.5}$ levels throughout the region of interest. This shall be accomplished by identifying the best ordinary kriging model available for the data. Once such a model is identified, we shall explore other, more complex, models. Such an exploration represents a form of model evaluation (i.e., "Is a more complex model a significant improvement?"). (See Section 3.5 and Section 5 for further discussion of model evaluation.) Model refinements shall be considered based on variogram and spatial complexities and shall be evaluated in terms of the amount of change in the final surface. As a general approach to model evaluation, if the amount of model improvement is small compared to the effort required to implement a given change, said refinement will be discarded.

7.1.1 Variogram Modeling Choices

Recall that the estimation of the covariance (or variogram) is likely the single most important and difficult step of any kriging procedure. In previous sections, it has been noted that there are a number of issues to consider when developing a variogram estimate. One must choose a model family and whether to use robust or normal estimation. One must also consider issues such as directional covariance (anisotropy) or non-stationary covariance. We revisit our analyses from previous sections here, comparing various variogram models under various levels of complexities, to identify a model that seems optimally useful. Note that while we will consider visual fits and objective function values in our evaluation in this section, final decisions should not be made until the kriging estimates are generated and compared.

The first step to any variogram estimation process is to generate an empirical variogram. Here, S-Plus gives us two choices on the options tab for the estimation method, normal or robust. In order to evaluate the sensitivity of the overall analysis to the choice of estimation method, we shall generate one empirical variogram for each method. Figure 7.1 illustrates the Empirical Variogram window from S-Plus. Next we use S-Plus Model Variograms functionality to automatically estimate the numerically optimal coefficients for each of the spherical, exponential, and Gaussian model families for both the normal and robust empirical variograms. Figure 7.2 provides an example of the S-Plus Model Variograms window. (See the caution regarding S-Plus's definition of range in Section 3.3.) Recall that by default S-Plus generates an empirical variogram based on 20 bins of distances. We shall use this default for now.

Observe in Figure 7.3 that visually the model fits to the empirical variograms are approximately the same for all six cases. Indeed, the objective values support this, providing scores that are all roughly equivalent. Recall that, in order to be truly meaningful, we look for objective values that are roughly an order of magnitude smaller (or larger) than the other choices. Since we do not see any such large differences, we shall instead turn our attention to the parameter estimates associated with these variogram models. Table 7.1 presents a summary of the numeric results. (See the caution regarding S-Plus's definition of range in Section 3.3.) (Note that these values were drawn from the report generated by S-Plus as part of the variogram modeling process.)

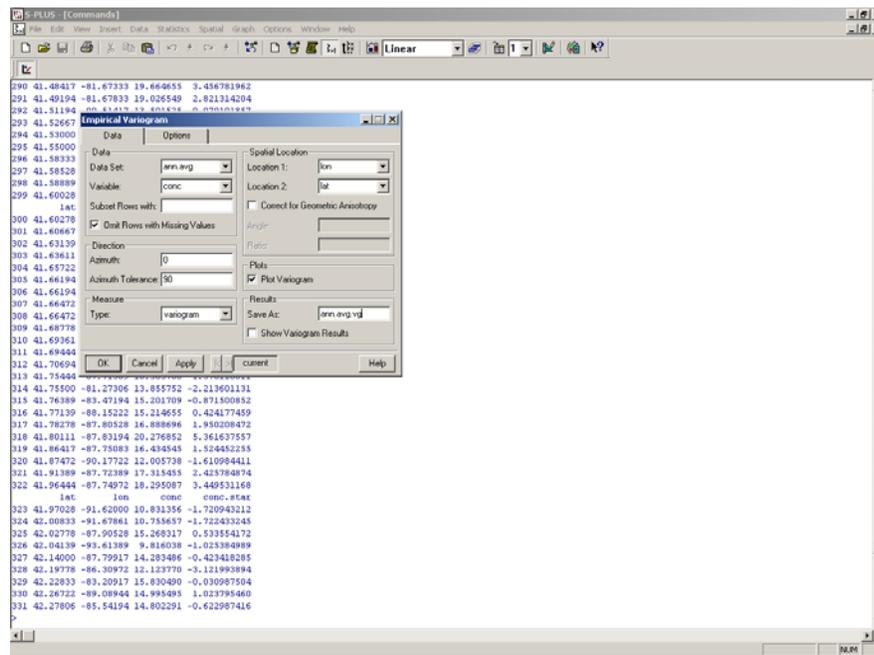


Figure 7.1 S-Plus empirical variogram window.

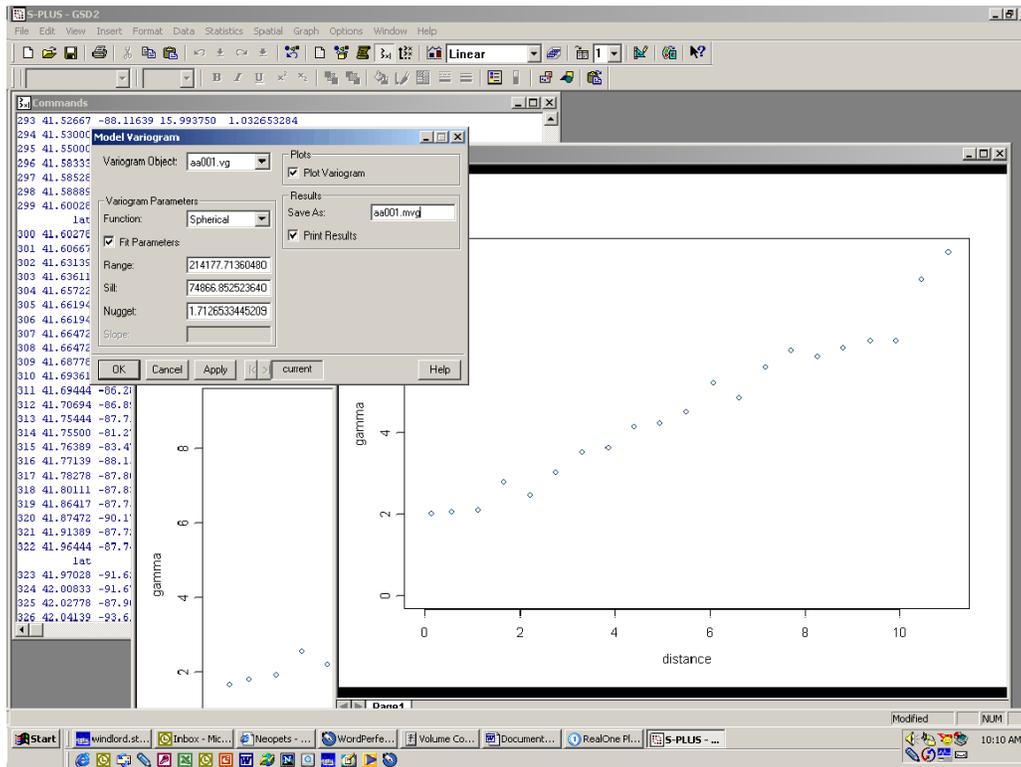


Figure 7.2 S-Plus model variogram window.

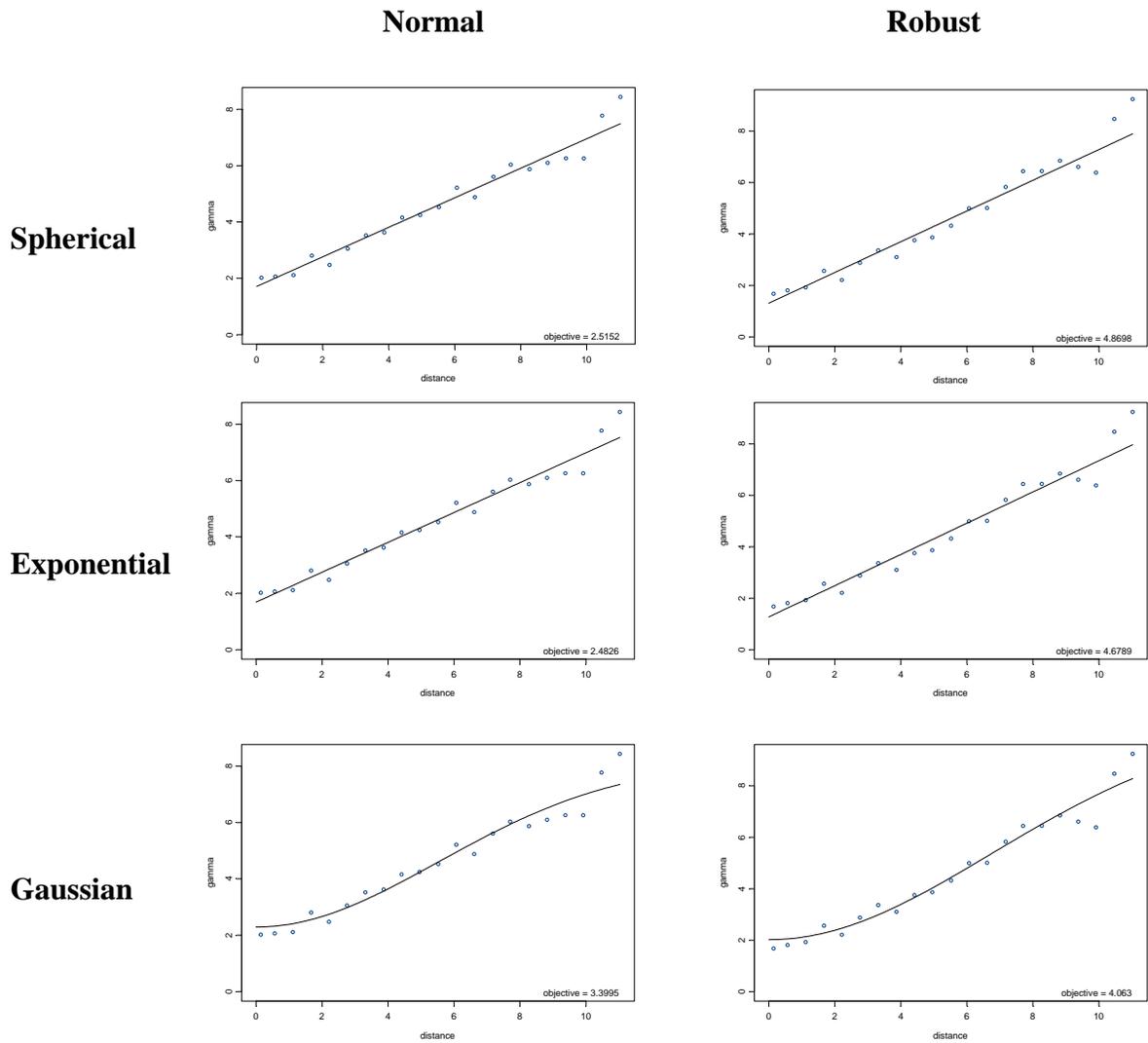


Figure 7.3 Comparison of empirical variograms.

Table 7.1 Summary of Variogram Model Parameters

NORMAL ESTIMATION

*** Variogram Model ***
 Function: Spherical
 range: 214177.713605
 sill: 74866.852524
 nugget: 1.712653
 Objective: 2.5152

*** Variogram Model ***
 Function: Exponential
 range: 96112.236677
 sill: 50990.137111
 nugget: 1.686132
 Objective: 2.4826

*** Variogram Model ***
 Function: Gaussian
 range: 7.796528
 sill: 5.854140
 nugget: 2.291343
 Objective: 3.3995

ROBUST ESTIMATION

*** Variogram Model ***
 Function: Spherical
 range: 281182.397096
 sill: 112035.966880
 nugget: 1.308906
 Objective: 4.8698

*** Variogram Model ***
 Function: Exponential
 range: 26024.464221
 sill: 15817.318979
 nugget: 1.272222
 Objective: 4.6789

*** Variogram Model ***
 Function: Gaussian
 range: 9.538153
 sill: 8.502609
 nugget: 2.018193
 Objective: 4.063

An important consideration when choosing a variogram model is the intuitiveness of the parameters used to estimate that model. In this example, both the spherical and exponential models have extremely large ranges and sills for both the normal and robust empirical variogram estimates (Note: units of $\mu\text{g}/\text{m}^3$ ²). As discussed previously, ranges of this magnitude exceed the circumference of the Earth. Thus, common sense tells us to be skeptical of these models and, therefore, we focus instead on the Gaussian model. Furthermore, the slight s-shape curve to the empirical variogram used for Figure 7.3 is consistent with the type of curve modeled by a Gaussian variogram model. Note that if the objective values had varied more significantly, if one model had been distinctly more visually appropriate, or if expert knowledge had indicated that a specific model was more appropriate, then we might have made an alternate decision. As it stands, we chose the Gaussian variogram with parameters estimated based on the robust empirical variogram. Robust estimation has been shown to be more stable than normal estimation (Cressie, 1993) and, thus, since the objective values are close, this choice is somewhat more desirable. This variogram model shall be considered our baseline model as we proceed

through this section. The kriging estimate resulting from this model is displayed in Figure 7.4 below.

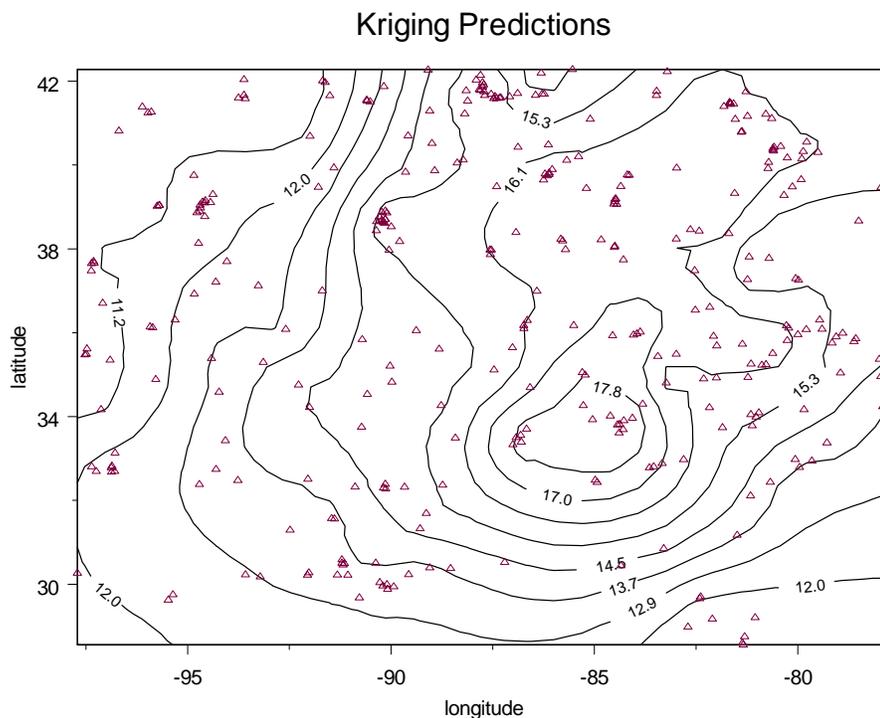


Figure 7.4 Baseline ordinary kriging predictions for annual $PM_{2.5}$ data.

We now consider refinements to the covariance model used to analyze the data. We start by considering anisotropy. Often, data sets will have geometric or other anisotropies associated with them. Thus, one should always consider testing for anisotropies in an analysis. In this example, we used the directional variogram code first discussed in Section 4 in the S-Plus command window (see Figure 7.5). Figure 7.6 depicts the directional variograms associated with the angles 0, 45, 90, and 135 degrees. (Recall the definition of the angle parameter from Section 4.1.) These focused variograms can help the investigator determine if there is a distinct directional impact of covariance. Kaluzny, et al., (1998) provide an extended discussion on this technique and other related diagnostic topics. In this example, we observe that the directional variograms are all approximately the same and, thus, assume that an isotropic model is reasonable for these data. We make this decision because a simple model is preferable to a complex model when the results are qualitatively similar.

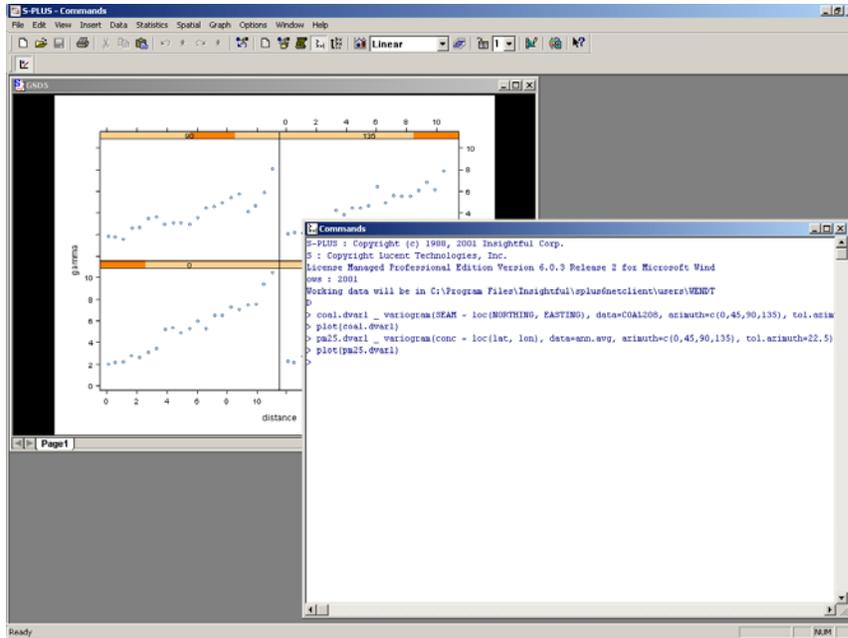


Figure 7.5 S-Plus command window, illustrating the directional variogram code.

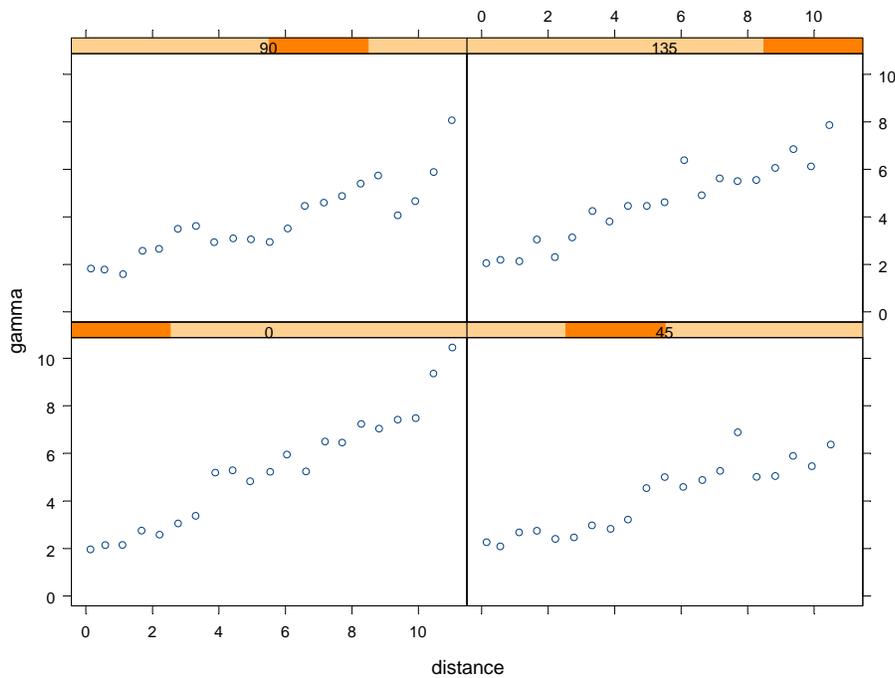
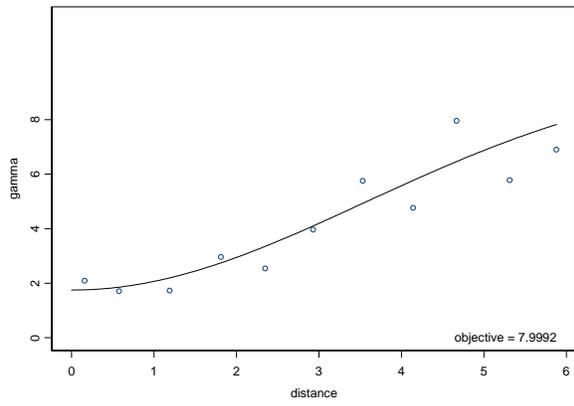


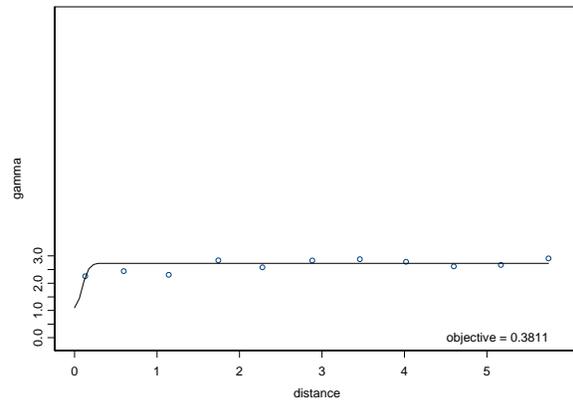
Figure 7.6 Directional variogram plot.

Having explored anisotropy, we consider alternative extensions to the variogram model. In particular, we revise the idea of using sub-region variograms (first discussed in Section 4.1) as a way to identify non-stationary covariance. The standard theoretical developments associated with kriging assume that the covariance throughout a space is stationary (that is, the covariance across the spatial domain is independent of the specific location within the spatial domain). For some applications, it may be reasonable to assume that the covariance through a space is stationary and only explore the issue of regional stationarity when expert knowledge or other analyses indicated that this might be a possibility. One example of such expert knowledge might include a recognition of one sub-region of the space of interest with dramatically different wind patterns than in the rest of the region.

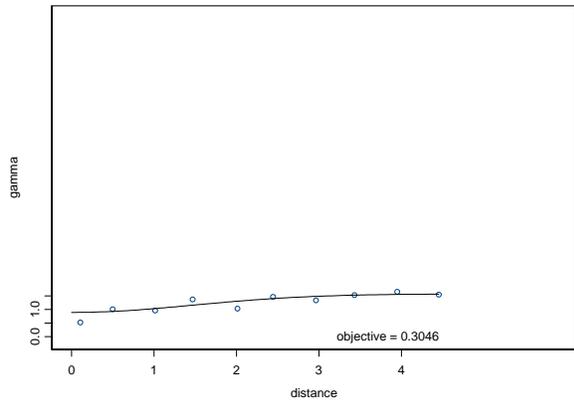
We first discussed sub-regional variogram analysis in Section 4, where we considered dividing the full spatial domain of the annual average $PM_{2.5}$ data set into nine sub-regions. In that example we found little evidence of region-specific covariance. In this section, we revisit the analysis. This time, we consider dividing the overall domain into four quarters along latitude and longitude (i.e., sub-regions). This was accomplished by creating separate data sets at the command line prompt in S-Plus. An empirical variogram and model variogram were then computed for each sub-regional data set separately. Note that, for this example, we computed the empirical variograms on 10 lags rather than 20 (the S-Plus default) because less data are available within each sub-region. Figure 7.7 illustrates the resulting Gaussian variogram models associated with those sub-region empirical variograms. Observe that these variogram models appear to vary widely. This suggests that there is a region-specific covariance structure that one might consider adjusting for in this model. (See the caution regarding S-Plus's definition of range in Section 3.3.)



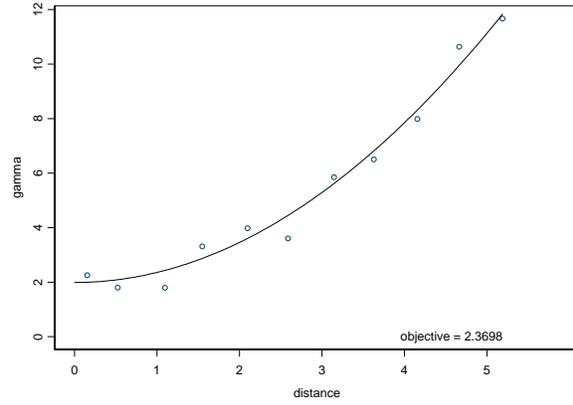
Gaussian(range: 5.01, sill: 8.12, nugget: 1.75)



Gaussian(range: 0.12, sill: 1.63, nugget: 1.10)



Gaussian(range: 2.08, sill: 0.68, nugget: 0.89)



Gaussian(range: 53.98, sill: 1069.93, nugget: 1.99)

Figure 7.7 Sub-region variograms.

Consider, however, the plots of the normal Gaussian variogram model estimated on the overall spatial domain versus the sub-region empirical variograms as depicted in Figure 7.8. In this example, there appears to be clear differences among the regions especially for the larger separation distances. These differences should be investigated; however, for simplicity, this document will use the overall simplified model.

7.1.2 Trend Modeling Choices

Now that candidates for the variogram have been chosen, we consider the impact of a more complicated spatial trend model. Specifically, we shall consider the impact of modeling large scale variability with a two-degree polynomial and the possible information to be gained by considering an urban versus rural categorical covariate. We shall compare the overall appearance of the surfaces to identify the best fit, but we also recognize that expert knowledge may factor into this decision. For instance, an individual with expansive understanding of the behavior of $PM_{2.5}$ in the region in question may recognize a model discarded as numerically sub-optimal as best describing the true spatial behavior of interest.

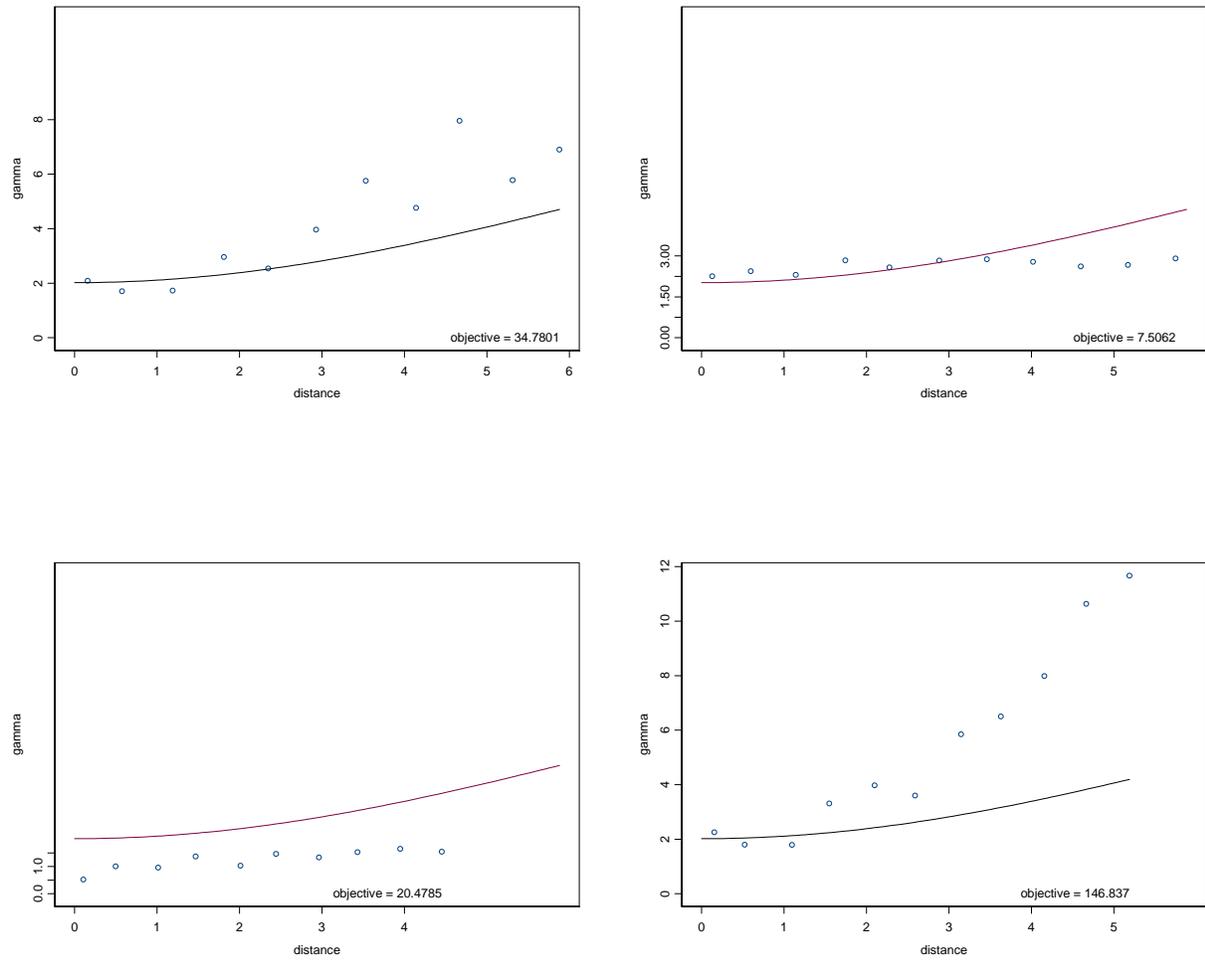


Figure 7.8 Sub-region empirical variograms overlaid with an overall variogram model.

Recall that Figure 7.4 plotted the kriging predictions generated using ordinary kriging and the Gaussian variogram model discussed earlier. Thus far, this is the model of choice. We have considered various variogram refinements and have rejected them in favor of a simpler covariance structure. Now we consider the possible refinement of using a large-scale spatial trend to better describe the overall behavior. To accomplish this we will compute an estimate using universal kriging. We use the same data as in our original estimates, measured over the same spatial region. We also apply the previously described covariance structure, noting that, in general, it is likely that the covariance structure associated with a trend model (i.e., universal kriging) will be different than that associated with an ordinary kriging. Using the Spatial -> Universal Kriging functionality in S-Plus, we can indicate the data set to use and polynomial terms to consider (see Figure 7.9). In this example, we choose all of the terms proposed automatically by the S-Plus functionality: longitude, longitude squared, latitude, latitude squared, and latitude times longitude. S-Plus can then compute the optimal large-scale trend surface. In this case, the long-term trend can be described by the following formula:

$$PM_{2.5} = 15.39 + 0.93 * \text{longitude} + 2.26 * \text{latitude} - 1.37 * \text{longitude}^2 - 2.80 * \text{latitude}^2$$

(Observe that a latitude times longitude term is not included in the equation above. This is because such a term was not significant in the large-scale trend surface estimated by the universal kriging process.)

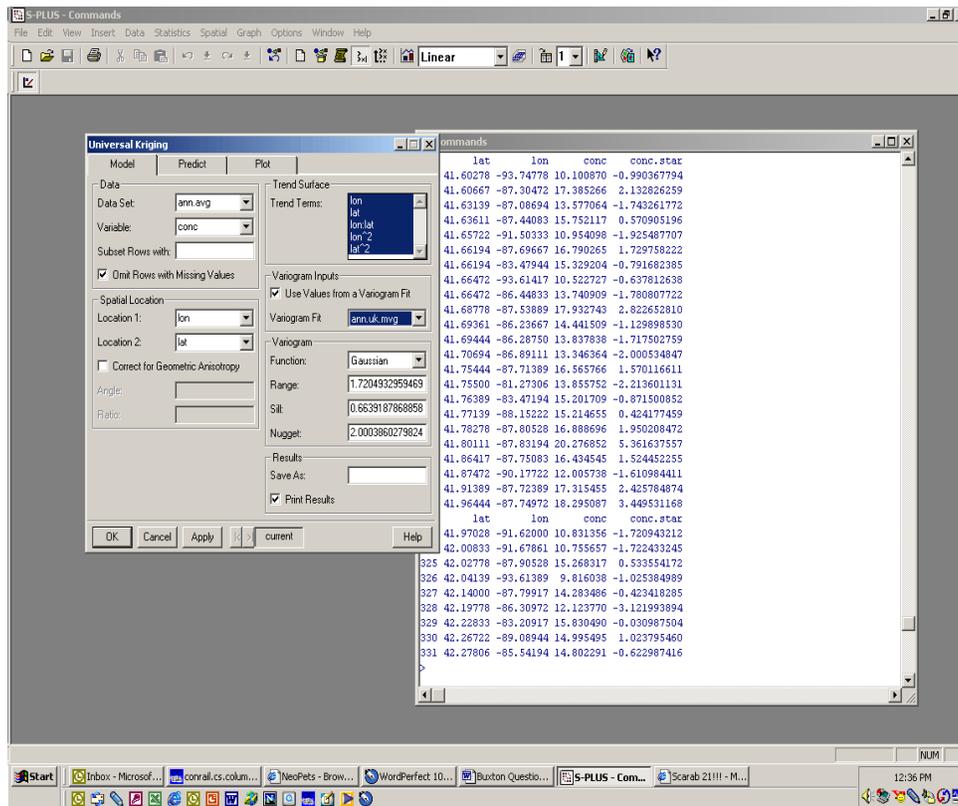


Figure 7.9 S-Plus universal kriging window.

Figure 7.10 illustrates the kriging surface that results from this model. While the model is different from the ordinary kriging model, the resulting kriging surfaces are very similar (compare Figures 7.4 and 7.10). While a monitoring or other $PM_{2.5}$ air quality expert might have more insight on the estimates depicted in these graphs, visual inspection does not reveal any obvious differences of note. Therefore, for reasons of simplicity, it may be reasonable to remain with the choice of an ordinary kriging model in this case.

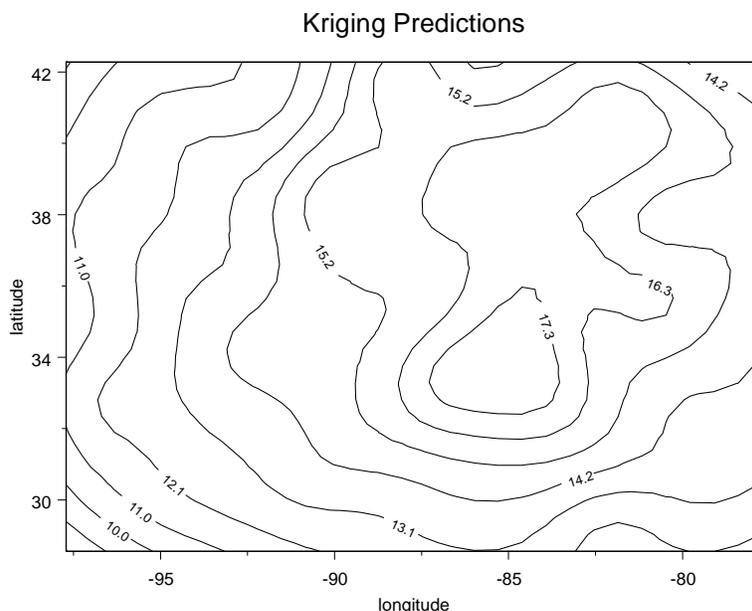


Figure 7.10 Universal Kriging predictions based on annual $PM_{2.5}$ data.

Now consider the possibility of using a covariate to refine the ordinary kriging model. By examining the available data, we notice that we can identify each monitor as being located in an urban or rural area. Since this seems like a covariate that might reasonably impact $PM_{2.5}$ concentrations, we explore whether refining the kriging with this covariate makes sense. To make this determination, we consider the distribution of concentrations in rural areas (code 0) versus urban areas (code 1) using side-by-side box plots (see Figure 7.11). This was accomplished by subsetting the data at the command line prompt and then using the built-in S-Plus functionality to develop the box plots. This plot makes it clear that there is some difference between the two subpopulations, but it is not dramatic, so we do not pursue this possible refinement any further. That is, since the two box plots overlap for most of their ranges, there appears to be no large practical difference in $PM_{2.5}$ values for rural as opposed to urban areas in the case of this particular data set. We note, however, that the desire to fit a more refined model to these data might result in the alternative decision to explicitly incorporate urban versus rural differences into the modeling approach. (A more formal statistical hypothesis test could be

considered at this point to further quantify the distributional differences, or lack thereof, between the two sub-populations.)

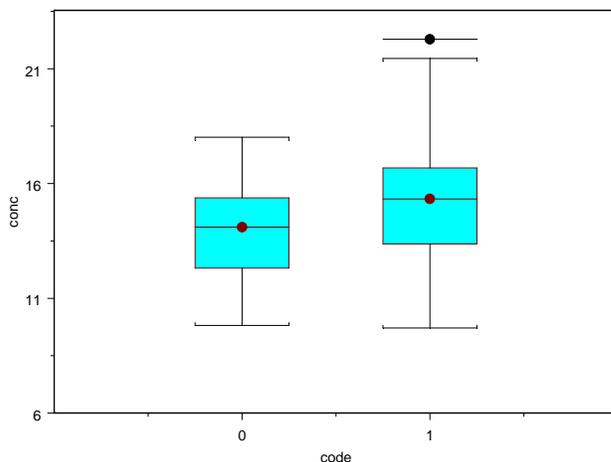


Figure 7.11 Side-by-side box plots of PM_{2.5} concentrations at urban versus rural monitoring sites.

7.1.3 Summary of Decisions and Results

Though we have examined a number of possibilities for more complex kriging estimation models, for this example involving annual average PM_{2.5} concentrations, it appears that an isotropic ordinary kriging model may be sufficient to reasonably describe the regional PM_{2.5} concentration surface (subject to the limitations of the data set). The biggest advantage of the resulting model choice is its simplicity. Note that this example may be unusual in its simplicity. Though it is generated using real data, it does not have the level of complexity that one can sometimes encounter in a spatial analysis of this sort. (This may be a result of the temporal averaging conducted prior to the analysis. For instance, any geometric anisotropies due to seasonal prevailing winds would be averaged out over the course of a year's worth of data.)

Observe that we did not explore any spatial or temporal extensions in this example. This is partly due to the limitations of the available data, but was also done to keep the example relatively simple. Specifically, the PM_{2.5} data available for this analysis did not have any elevation information available. Therefore, extensions into a third spatial dimension were not possible. Also, for simplicity, the data were averaged to the annual scale to remove any temporal dynamics. This reduced the example to a simpler spatial exercise (as opposed to a spatial-temporal exercise).

Still, a number of important observations can be made on the nature of such analyses. First, in this example, the kriging surface tends not to be very sensitive to changes in the model. We have examined a number of variations, only to find that the end result is qualitatively unchanged. Second, and resulting from the first observation, we observe that simplicity can have great value in spatial interpolations. Adding complexity to the model always increases the amount of effort required to complete the analysis and can often lead to very small changes (and possibly little improvement) over a simpler model. Finally, it is important to remember that kriging will always attempt to fit the surface through the observations. As the nugget of the variogram increases, there will be more difference between the kriging estimate and the observation at a given monitor location, but as a general rule the kriging surface will always tend toward the observations.

In conclusion, observe that while S-Plus's Spatial package has many strengths, it also has a number of limitations. Variograms and kriging estimates can be generated with relative ease for reasonably simple model structures. However, changes to the resulting graphs and more complicated analyses can take potentially complex command line programming.

7.2 Ozone 8-Hour Design Value Data

In this example we interpolate a spatial surface of eight-hour ozone design values over the southeastern United States. An eight-hour ozone design value is the arithmetic mean of the 4th highest daily eight-hour maximum concentrations for each of 3 consecutive years (40CFR Part 50.10). The dataset used to perform this interpolation includes eight-hour ozone design values for 2001 (based on ozone observations in 1999, 2000, and 2001) for ozone monitoring stations in Georgia, Alabama, Mississippi, Louisiana, Arkansas, and Tennessee — a total of 96 monitoring stations in all. Figure 7.12 shows the locations of the monitoring stations. A gray diamond has been added to the plot to show the location of Birmingham, Alabama. The interpolation of design values is performed using universal kriging in which covariates and spatial covariance information are used together to predict values between monitor locations. Universal kriging differs from ordinary kriging in that spatial covariates can be included in the model to improve predictions. The covariates are included in a way similar to the way covariates are included in an ordinary regression; the response variable (design values in our case) is predicted by a linear combination of covariates (although the values of other observations play a role in prediction as well). We choose to use universal kriging based on previous experience modeling the same dataset with ordinary kriging. The ordinary kriging models do not account for the differences between ozone levels found near cities and those found in rural areas, and universal kriging provides a simple method of including the proximity of observations to major cities as a covariate.

Locations of Monitoring Stations

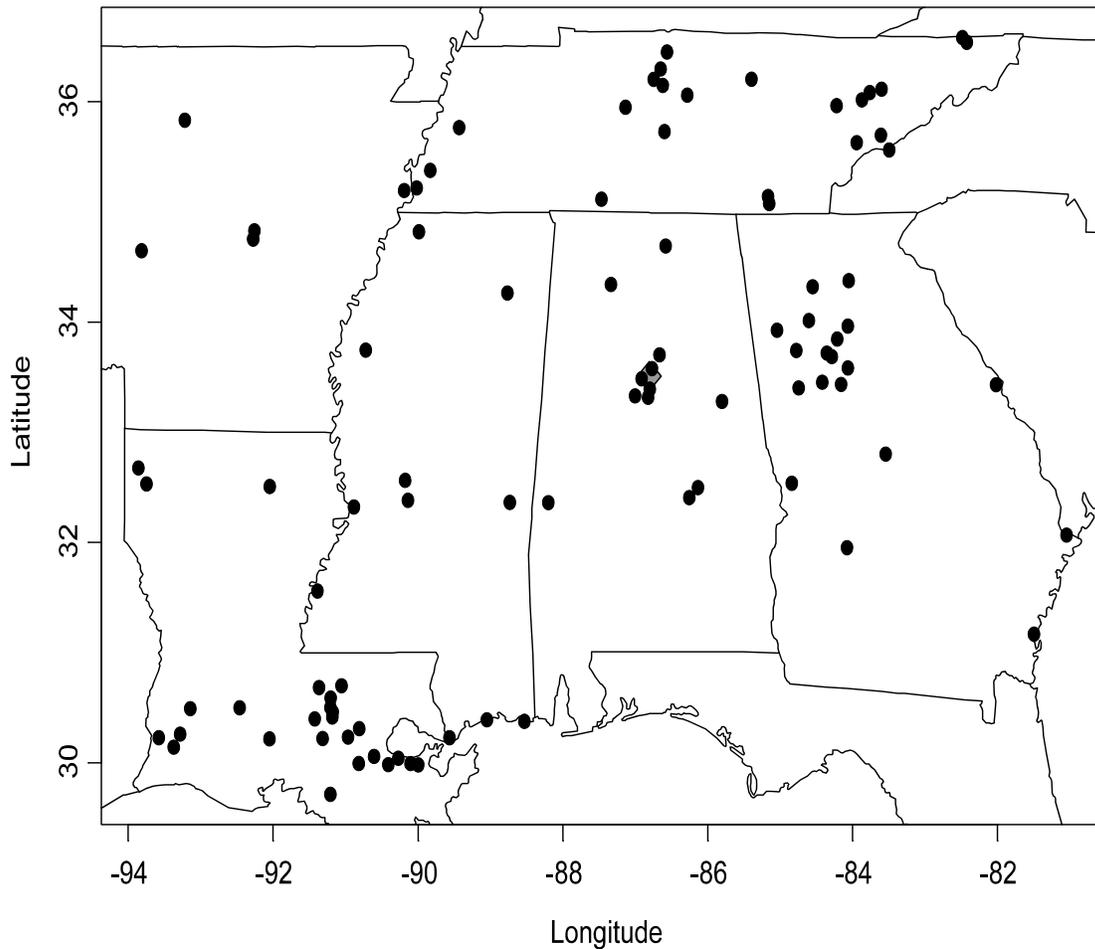


Figure 7.12 Locations of monitoring stations over southeastern United States.

Before developing a spatial model, we examine some plots of the data to search for patterns that might give an indication of an appropriate model. Figure 7.13 shows the data graphically. It is apparent from the figure that some spatial correlation exists since values close to each other tend to be the same color (i.e. magnitude of the concentrations). This figure also seems to indicate that ozone design values calculated closer to major cities are larger. For instance, note the darker squares near the locations of Atlanta, Georgia, and Knoxville, Tennessee, (indicated with black diamonds in the figure). A blue diamond has been added to show the location of Birmingham, Alabama, on this figure. The design values shown are recorded in parts per billion.

Design Values, 1999-2001

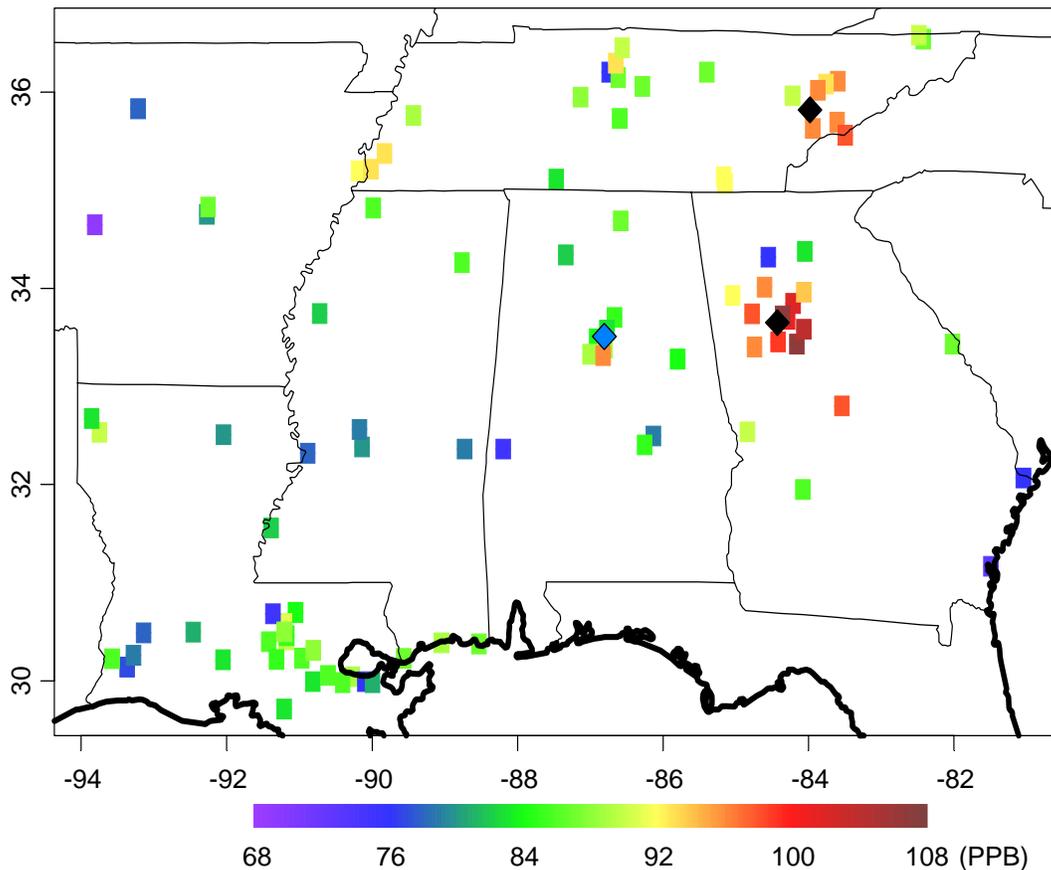


Figure 7.13 Eight-hour ozone design values for 1999-2001 observed at stations across southeastern United States.

In this example, we do not give all of the details associated with constructing the spatial model, but we do describe the steps taken in the construction of our spatial interpolation model for these data.

The first step is to explore the correlation structure of the data using a semi-variogram analysis. Using the full dataset, we construct an empirical semi-variogram. Since the field explored is approximately thirteen longitudinal degrees by seven latitudinal degrees, a maximum lag distance of six degrees was used in construction of the semi-variogram. In constructing the semi-variogram, latitudinal and longitudinal degrees were considered equivalent, which leads to some deformation of the surface. The empirical semi-variogram value was calculated at distances of 0.25, 0.5, 1, 1.5, ..., 6 degrees. Figure 7.14 illustrates the empirical semi-variogram calculated using all of the available observations in the dataset. Points in the figure are proportional in size to the number of observations used to calculate their value.

Empirical Semi-variogram

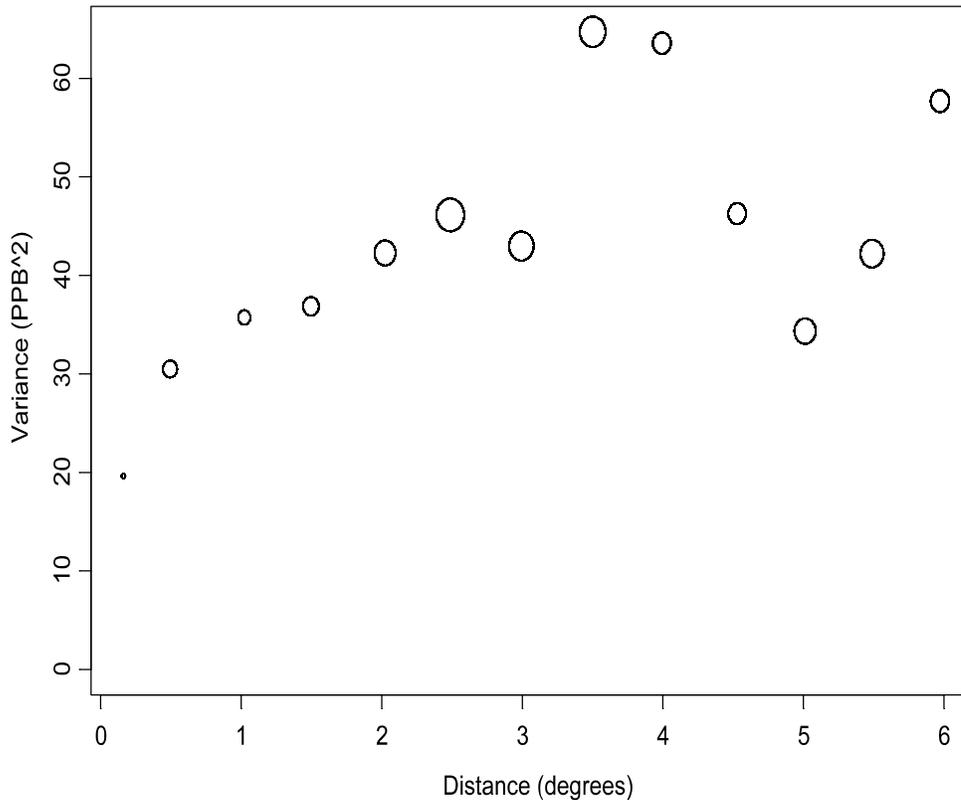


Figure 7.14 Empirical semi-variogram for full dataset.

Based on visual inspection of the empirical semi-variogram in Figure 7.14, it was decided that a Gaussian semi-variogram would be appropriate for describing the correlation structure in the data. Based on visual inspection, a sill of 50 PPB² was chosen for the variogram. The nugget and range of the variogram were selected using an iterative procedure with the ultimate goal of obtaining an average of 2 percent spatial interpolation error at the monitoring sites. Each step of this iterative procedure involved selecting a value for the nugget, estimating a corresponding range value by eye, fitting the universal kriging model (described in more detail below) using those values, and calculating the average percent error over the monitoring sites. These steps were iterated adjusting the nugget value each time until nugget and range values corresponding to an average error of 2 percent were found. The final values were 4.5 PPB² for the nugget and 1 degree for the range. Note, that the 2 percent criterion is used for illustration purposes. It was chosen to reflect the opinion of some monitoring analysts that ozone design values are probably accurate to within +/- 2 ppb which is approximately 2 percent of the larger recorded values.

We chose the Gaussian semi-variogram based on its popularity for this type of modeling and the shape of the empirical variogram as determined by visual inspection. However, due to

the small amount of data available, it is difficult to have confidence in any variogram choice. The Gaussian theoretical semi-variogram is described by the following equation:

$$V(d) = (nugget) + (sill - nugget) \left(1 - \exp \left\{ -\frac{d^2}{range^2} \right\} \right) tt$$

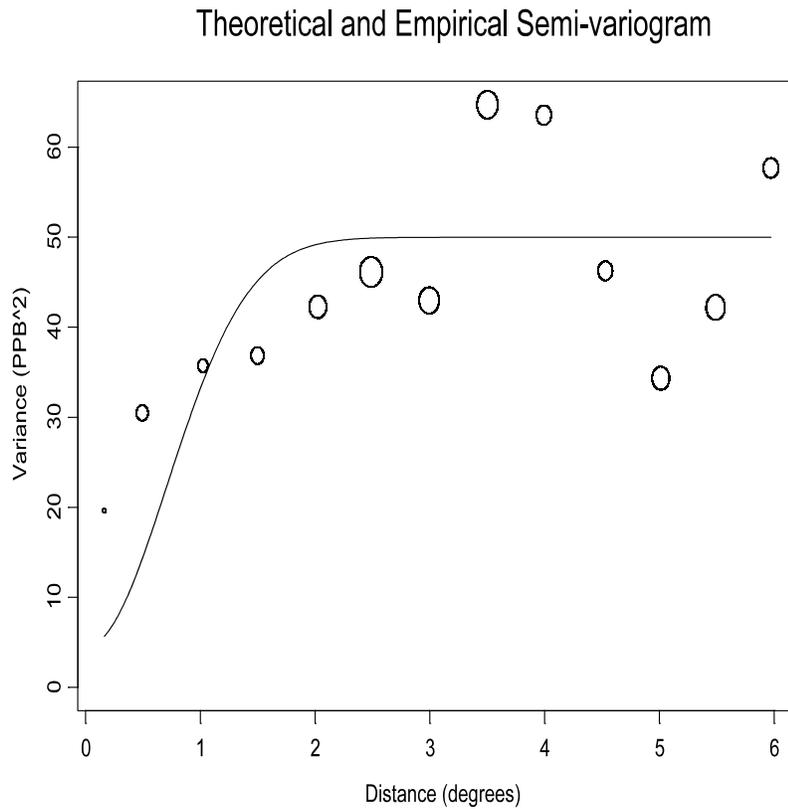


Figure 7.15 Theoretical and empirical semi-variograms for the full dataset.

where V is the value of the semi-variogram and d is the distance between observations. Figure 7.15 shows the empirical semi-variogram again with the theoretical semi-variogram superimposed.

While the Gaussian theoretical semi-variogram seems to fit the empirical semi-variogram relatively well, some investigators of this dataset have raised concerns about anisotropy in the field (i.e., the variogram may be different in different directions). To explore this possibility, we constructed one north/south and one east/west semi-variogram and compared the two. Directional semi-variograms are constructed in the same way as a full semi-variogram except that we make the additional restriction that the angle between any two observations used in the construction of a directional semi-variogram must be within certain limits. For example, if two points were arranged such that one was directly to the north of the other, that pair would be included in the construction of the north/south semi-variogram but not the east/west variogram. Since points seldom lie exactly along north/south or east/west lines, we specify a tolerance of 45 degrees for directional semi-variogram construction. In other words, all sets of two points with an angle between compass heading 315 and 45 would be used for the north/south semi-variogram. All sets of two points with an angle between compass heading 45 and 135 would be used for the east/west semi-variogram. Construction of more than two variograms was not considered due to the small size of the dataset. The resulting variograms appear in Figures 7.16 and 7.17. The theoretical variogram from the full dataset is superimposed as a dotted line on each of the directional empirical semi-variograms in the figures. While the

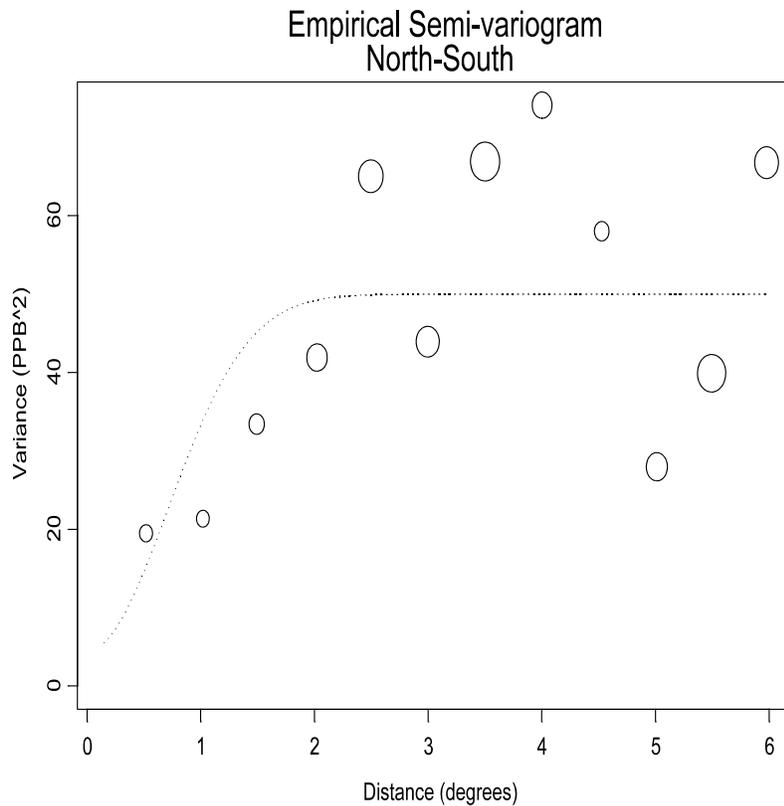


Figure 7.16 North/South empirical semi-variogram for the full dataset.

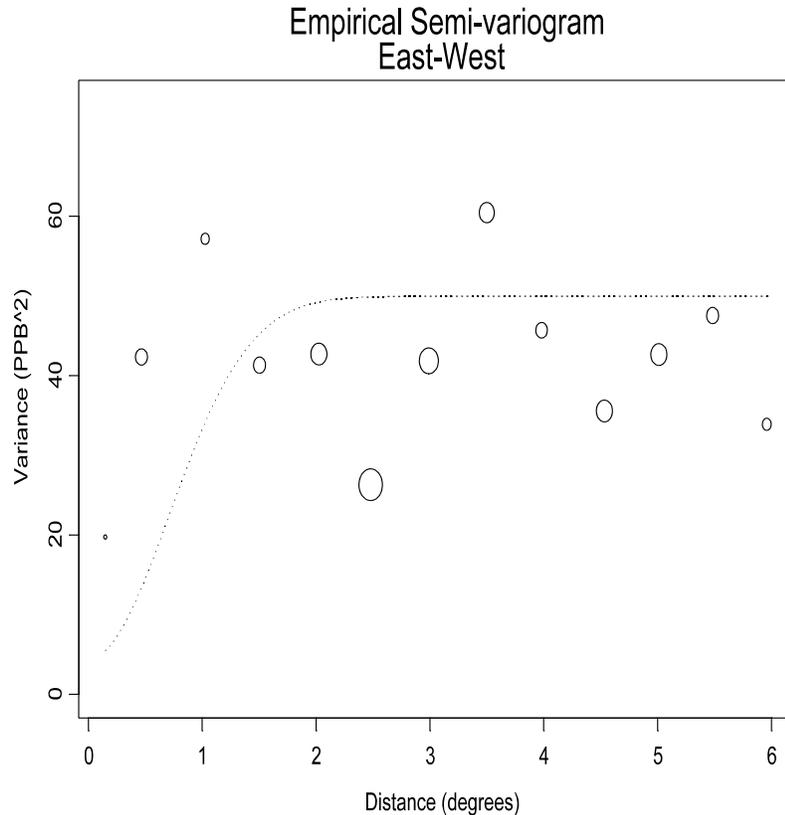


Figure 7.17 East/West empirical semi-variogram for the full dataset.

patterns that appear using the full dataset are not as apparent in the directional variograms, it still seems reasonable to use a Gaussian semi-variogram with a nugget of 4.5 PPB², a sill of 50 PPB², and a range of 1 degree in each case (dotted lines showing these variograms are included in the figures). For this reason, we decided not to incorporate anisotropy into our study, and instead use only the theoretical semi-variogram illustrated in Figure 7.15.

With the theoretical semi-variogram constructed, we turn our attention to the selection of covariates that may be used as predictors in the universal kriging model that we use for interpolation. Recall that in a universal kriging model, we hypothesize that the value at any location is a linear combination of spatial covariates and a spatially correlated error field. Previous analyses of the current dataset (see EPA, 2003) have indicated that the spatial field behaves differently near large cities than it does in rural areas. As a consequence, we include the distance from the nearest major city (defined as a city with a population over 200,000 people in the 2000 census) as a covariate. These cities include Birmingham, Alabama; Montgomery, Alabama; Atlanta, Georgia; New Orleans, Louisiana; Baton Rouge, Louisiana; Shreveport, Louisiana; Nashville, Tennessee; and Memphis, Tennessee. To allow for long range trends in the data, we also include linear terms for the latitude of a point, the longitude of a point, and the interaction between latitude and longitude. In other words, we allow for a long-range flat

trend surface in the data.

We fit the described model producing kriging estimates on a 47×47 grid with 4 km resolution as illustrated in Figure 7.18. The accompanying standard errors for the spatial surface are presented in Figure 7.19. In each figure, we include pluses to show the locations of monitoring stations and a light blue diamond showing the location of the city center of Birmingham, Alabama.

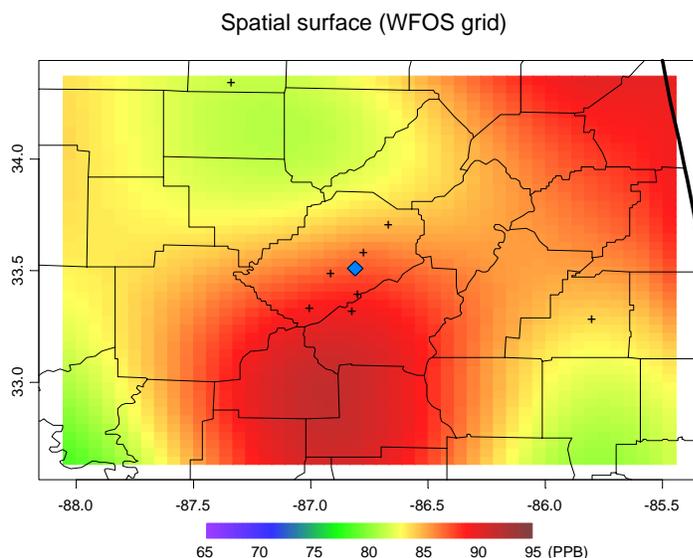


Figure 7.18 Spatially Interpolated Surface of eight-hour ozone design values.

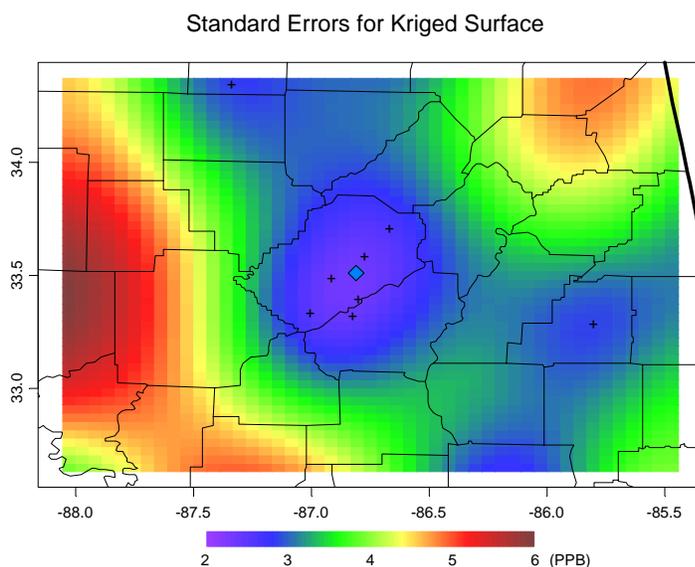


Figure 7.19 Standard errors for the spatially interpolated surface of eight-hour ozone design values.

Estimates of the coefficients of the covariates in the model are presented in Table 7.2. The effect of the distance to the nearest major city is negative indicating that design values are smaller away from cities after adjusting for the spatial correlation in the data. We retain the other effects in the model even though their p-values are slightly non-significant (at the 0.05 level).

Table 7.2 Effect estimates for covariates in the universal kriging.

Effect	Estimate	Standard Error	p-value
Distance to nearest city (degrees)	-3.1752	1.0656	0.0037
Latitude	22.8623	15.1471	0.1347
Longitude	-7.5726	5.7127	0.1883
Latitude*Longitude	0.2494	0.1714	0.1491

While several techniques exist to evaluate the fit of a kriging model, we only apply the most common technique, leave-one-out cross validation (LOOCV) with a single variogram, to determine whether the model fits the data well. Information on LOOCV, as well as several other model validation techniques, can be found in Chapter 5. In short, each observation in the dataset is removed, in turn, and the kriging prediction at that location is calculated using the remaining 95 observations. For each observation, the error in the prediction is divided by the standard error of prediction resulting in 96 standardized residuals. We use several different visual methods to evaluate whether any of these residuals appear suspect. The goal in examining these residuals is to identify possible problems with the model such as asymmetry of the error distribution or poor interpolation in specific areas. We note that this residual analysis does not give information about several kriging model assumptions that could be questioned (such as the selection of covariates, the isotropy of the field, and the uniformity of the field over the entire study region).

The first method used to validate the model is to examine a histogram of standardized leave-one-out cross validation residuals. Figure 7.20 shows a histogram of the standardized residuals for our analysis with a standard normal density curve superimposed. While the residuals appear to be slightly left-skewed, this skewness is not dramatic. Overall, the residuals appear relatively symmetric, so we do not consider our model invalid. One concern that might arise is that the residual distribution appears to be relatively heavy-tailed (i.e., there are several residuals far from zero). This might result from an incorrect theoretical variogram specification since we chose the theoretical variogram to produce an average error of 2 percent rather than selecting the theoretical variogram based purely on observation of the empirical variogram. However, since the distribution is not dramatically heavy-tailed, we do not consider the model invalidated.

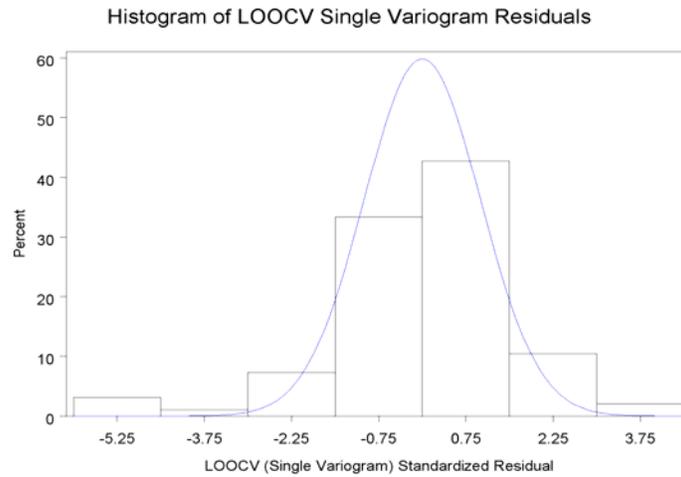


Figure 7.20 Leave-one-out cross validation (single variogram) residuals.

The second method used to validate our model is to examine a plus/minus plot of the residuals. A plus/minus plot shows graphically the locations where the model is underestimating (pluses) and overestimating (minuses) the true observations. In a plus/minus plot, we hope to see a good mix of plus and minus symbols in all locations. Figure 7.21 illustrates a plus/minus plot for our data. In this plot, the size of the plus and minus symbols is proportional to the magnitude

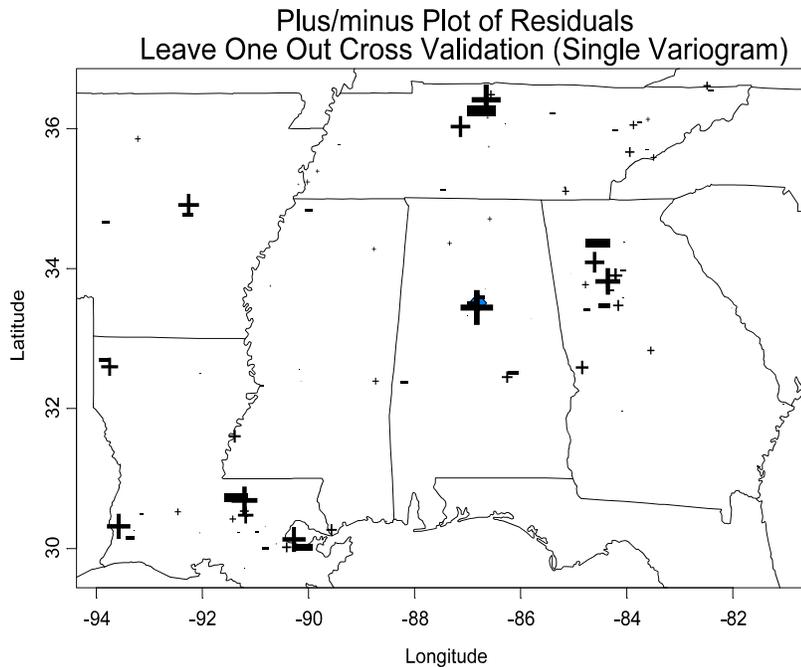


Figure 7.21 Plus/minus plot of leave-one-out cross validation residuals.

of the residual. Most of the residuals appear only as small dots since the residuals are so small; however, we can easily see locations where large prediction errors are found. Since we do not see clusters of large residuals in any place, the plot causes no concern in this respect. Another sign of an incorrect model that can appear in a plus/minus plot is clustering of either pluses or minuses. While there appear to be some clusters of pluses in urban areas, the plus/minus plot does not show any severe clustering. We do not consider our model invalidated by this plot. Since we are focusing on the Birmingham area, we present the values for the leave-one-out cross validation residuals in that area in Figure 7.22.

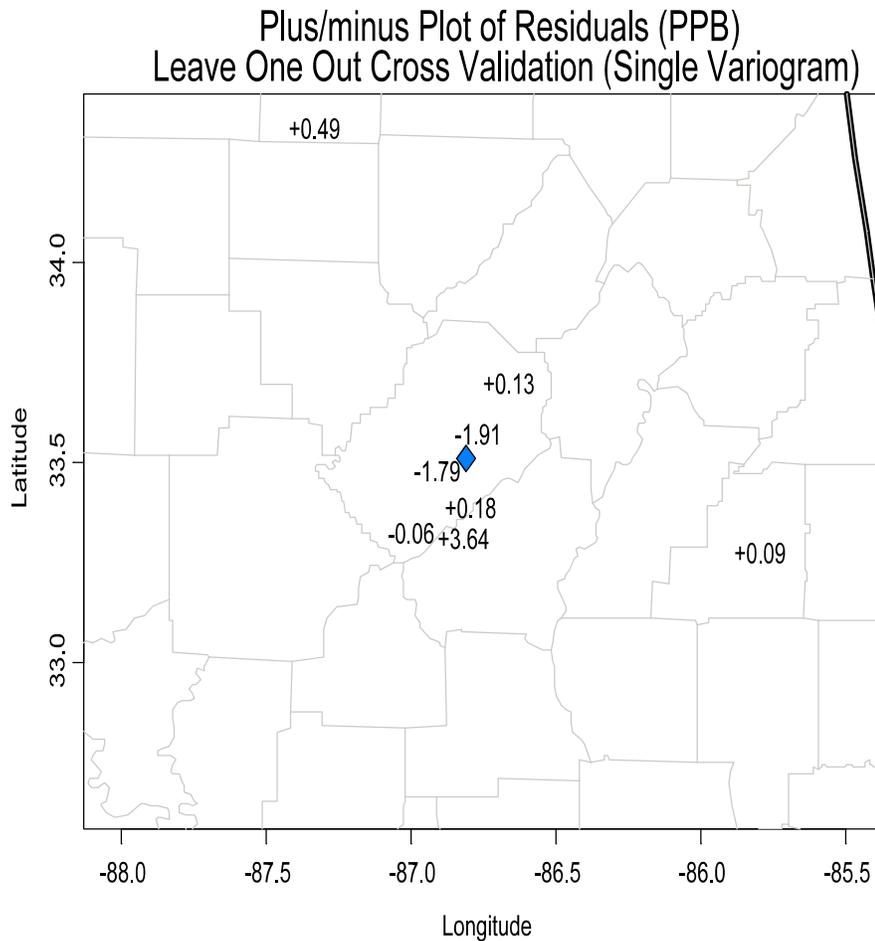


Figure 7.22 Leave-one-out cross validation residuals in the Birmingham area.

As an additional review, we compared the observed design values to the interpolated design values at the monitor site locations. Recall that the model was chosen to produce an average error of 2 percent at the monitor locations. The errors range from -10.3 PPB to +6.5 PPB with a standard deviation of 2.6 PPB. Figure 7.23 presents a scatterplot of the observed values versus the interpolated values at the corresponding locations. A 45-degree line has been added to the plot to better illustrate deviations from perfect prediction. The plot shows that predictions are

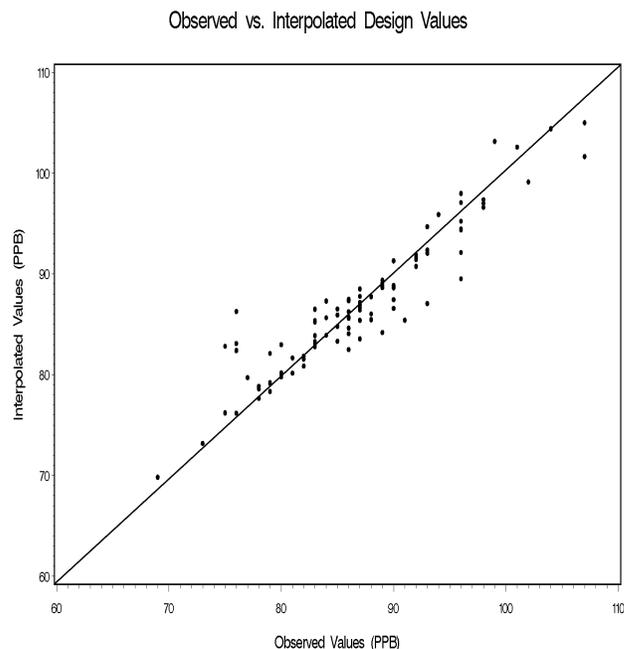


Figure 7.23 Scatter plot of observed and interpolated design values.

relatively close to the line, and predictions are not particularly worse at extremes of the observed value range. Appendix B contains a full table of observed design values and corresponding interpolated design values. Table 7.3 presents a subset of these observed and interpolated values that fall within the region. In addition, Figure 7.24 shows a plot of the difference between the observed and interpolated values. The values in the figure are the observed value minus the interpolated value. Positive numbers indicate that the interpolated surface underestimates the observed value, and negative numbers indicate that the interpolated surface overestimates the observed value.

Some concern may arise from the fact that the interpolated design values do not exactly match the observed design values at the monitor locations. There are several justifications for including error in the interpolation at the monitor locations. First, there is inherent uncertainty in the observed values since the monitors record the ozone level at any given moment with some measurement error. Second, additional information is provided by monitors in the area surrounding any given monitor, so the spatial interpolation may actually be a more representative value for the level at the monitor.

Table 7.3 Design values, interpolated values, and residuals for locations within region.

State	County	Latitude	Longitude	Interpolated value (PPB)	Design value (PPB)	Residual (PPB)
Alabama	CLAY	33.28	-85.80	83.92	84.00	0.08
	JEFFERSON	33.49	-86.91	87.32	84.00	-3.32
		33.33	-87.01	89.10	89.00	-0.10
		33.39	-86.80	88.66	89.00	0.34
		33.70	-86.67	84.80	85.00	0.20
		33.58	-86.77	86.50	83.00	-3.50
	LAWRENCE	34.34	-87.34	81.56	82.00	0.44
	SHELBY	33.32	-86.83	89.53	96.00	6.47

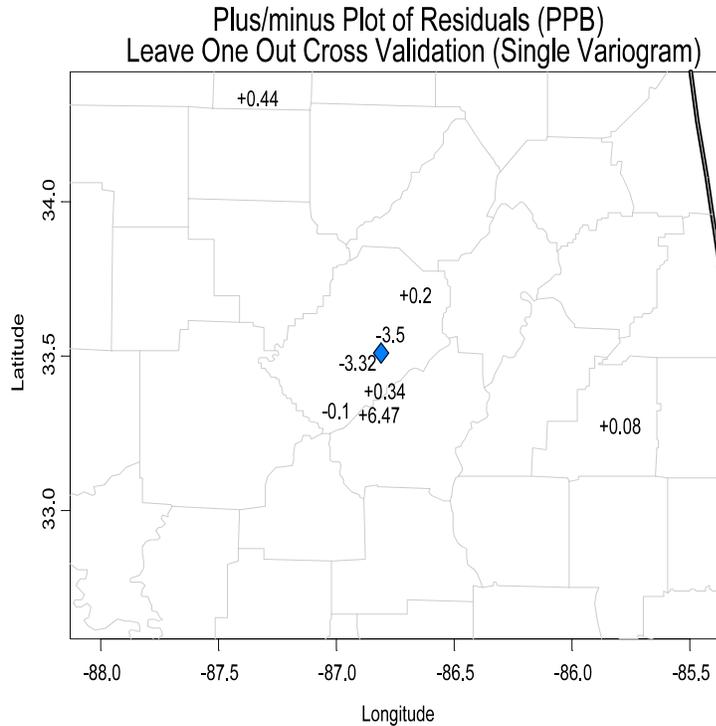


Figure 7.24 Residuals (observed minus interpolated) in the Birmingham area.

Since differences are small and are within the range of uncertainty in the observed data, the results support a conclusion that estimates in the unmonitored cells are valid or adequate for use. In fact, using spatial interpolation might provide an improvement over the use of the area-wide maximum design value for determining nonattainment status.

7.3 References

- [1] Cressie, Noel A. C. (1993). Statistics for Spatial Data. John Wiley & Sons, New York, New York.
- [2] Kaluzny, Stephen P., Vega, Silivia C., Cardoso, Tamre P., and Shelly, Alice A. (1998). S+ Spatial Statistics. Springer-Verlag, New York, New York.

8.0 SPATIAL INTERPOLATION MODELING LIMITATIONS

Section 2 of this document provided a general overview of spatial interpolation methods. The remainder of the document has focused on the spatial statistics approach of kriging and variants of this modeling approach. The assumption throughout has been that such spatial interpolation models will be developed using ambient air monitoring measurements as their primary source of data. The focus on kriging models and monitoring data is appropriate given the prevalence and strengths of such applications. However, there are certain limitations associated with this approach to spatial interpolation. The following sections offer discussion on some of the most pertinent issues. Specifically, Section 8.1 considers different air pollutants. Section 8.2 discusses issues associated with spatial scale. Next, Section 8.3 provides further discussion on spatial prediction uncertainty. Section 8.4 talks about monitoring network design issues, which impact and depend on the issues discussed in the preceding sections. Finally, Section 8.5 lists references. (Refer to Section 9 for a discussion of some alternative modeling approaches and associated issues.)

8.1 Air Pollutant of Concern

In the most general sense, spatial interpolation provides a valuable tool for better understanding air quality and the extent of our nation's air pollution problems. Among other applications, spatial interpolation models can be used to define the spatial extent of episodes of unhealthy air quality, to illuminate relationships between different air pollutants, and to aid in the design or re-design of ambient air monitoring networks. However, the degree to which such models can be developed in a successful, applicable manner may strongly depend on the air pollutant of concern. Put another way, the chemical properties and atmospheric fate and transport of certain pollutants may present unique problems with important ramifications for spatial interpolation modeling.

Among other factors, such as emissions and meteorology, concentrations of air pollutants are affected by atmospheric transport. More specifically, according to Main and Roberts (1999), a pollutant's residence time in the atmosphere, or atmospheric lifetime, determines its range of transport. A residence time on the order of several days generally yields long-range transport, which in turn leads to more uniform spatial patterns. (More uniform spatial patterns may also be consistent with longer averaging times; see Section 3.3 for an example and further discussion.) For example, $PM_{2.5}$ particles, sulfates in particular, exhibit typical residence times of three to five days within the atmospheric boundary layer (i.e., the lowest one to two kilometers). As a result, on average, $PM_{2.5}$ particles can be transported 1,000 or more kilometers from their sources (e.g., a residence time of four days and a mean transport speed of ten kilometers per hour yields a transport distance of about 1,000 kilometers). Under these conditions, a large-scale spatial interpolation exercise such as the annual $PM_{2.5}$ case study example presented throughout this document would seem reasonable.

Next, consider ozone. There exist certain similarities in the formation of ozone and $PM_{2.5}$, in that a substantial fraction of $PM_{2.5}$ depends on photochemical gas phase reactions that

also produce ozone. More generally, existing scientific evidence suggests that regional haze, fine particles, and ozone have common precursor pollutants, emission sources, atmospheric processes, spatial scales of transport, and geographic areas of concern (Main and Roberts, 1999). For these reasons and others, informative spatial interpolations of ozone at regional-scale domains have been developed in a number of applications (see Section 10 for summaries of ozone-related applications). For example, EPA's AIRNow website (<http://www.epa.gov/airnow/>) routinely provides regional maps of spatially interpolated ozone concentrations generated using a non-stochastic inverse-distance weighted (IDW) approach (see Section 2 for a description of the IDW method). Likewise, certain air toxics such as carbon tetrachloride, which is naturally emitted and considered to be relatively spatially ubiquitous (it has a residence time greater than five days), may also exhibit atmospheric behavior conducive to large-scale spatial interpolation applications (Spicer, et al., 2002).

The pollutant-specific spatial characteristics discussed above may not hold true for other pollutants. In particular, other air toxics such as 1,3-butadiene, acrolein, and formaldehyde may tend to exhibit more locally-influenced behavior due in part to residence times of less than one day (Spicer, et al., 2002). Spatial interpolation models can still provide utility in these cases. However, their value may be limited to smaller spatial scales (e.g., urban or neighborhood). Or, more sophisticated modeling approaches may be required in order to obtain useful results. Furthermore, attempts to interpolate may suffer from a lack of spatially dense monitoring data, at least to the degree that spatial patterns are less uniform and a higher level of prediction accuracy is required. In fact, unlike the current relatively extensive monitoring networks in place for ozone, particulate matter, and other criteria pollutants, a national air toxics monitoring network does not currently exist as of the writing of this document. [Refer to Table 5.3 of Spicer, et al., (2002) for a summary of residence times (atmospheric lifetimes) for the 33 air toxics targeted by EPA as presenting the greatest threat to public health in urban areas.]

In summary, different air pollutants may present different technical problems with respect to developing spatial interpolation models. Likewise, spatial interpolation modeling needs and the ultimate utility of such models may vary from pollutant to pollutant. These issues must be considered before fully embarking on any spatial interpolation exercise. Careful planning, such as utilizing the DQO process summarized in Section 5, can help in these matters.

8.2 Spatial Scale

Depending on the size of the spatial domain to be interpolated, it can be important to consider different spatial scales in the modeling process. Related to the discussion above in Section 8.1, it is possible that the air pollutant or other process being modeled exhibits differential behaviors at varying scales. For example, fine particulates and ozone may present relatively smoothly varying large-scale trends at the regional level due in part to long-range transport, more textured variation at the urban scale due in part to the distribution of emissions sources, and still further variation at the neighborhood or micro scale due to terrain, surrounding infrastructure, and other factors. It can be important to model these different spatial scales of variation for the purpose of better understanding the impact of the various causal factors. Or,

explicit modeling of such spatial effects may lead to increased accuracy in the resulting spatial predictions (see discussions of model overfitting throughout Section 4).

In general, the kriging approach to spatial interpolation modeling is equipped to potentially address different spatial scales. Consider the specific method of universal kriging, as described previously in Section 4.2 and Equation (5). First, the $u(x,y)$ term in Equation (5) is typically included in this model to capture large-scale smooth fluctuations in the spatial process of interest. Next, the variogram model assumed to hold for the covariation of the $e(x,y)$ term in Equation (5) is designed to capture smaller-scale spatial variation. Depending on the application and domain of interest, this level of spatial variation may correspond to urban or neighborhood scale variability. Finally, by including a nugget term in the variogram model assumed for $e(x,y)$, universal kriging attempts to address some mixture of micro scale variation and measurement error. Refer to Cressie (1993) for a similar partition and further discussion of spatial variation at different scales, as addressed in the kriging setting.

Although, theoretically at least, kriging methods can address different spatial scales, some real-world scenarios may present more difficult modeling challenges not easily addressed at certain scales. For example, as discussed above in Section 8.1, some pollutants such as certain air toxics may exhibit little to no atmospheric transport beyond their urban and surrounding domain of origin. Cases like this may not require regional-scale spatial interpolation, or model development across such larger domains may be difficult and possibly inappropriate. As stressed in Section 3, kriging methods use mathematical equations to best match the empirical behavior of a given data set. Outside of some quantitative model fitting criterion, there is no requirement for these equations to be consistent with any underlying atmospheric or physical processes. It is, therefore, important for researchers to keep in mind both the utility and appropriateness of a potential spatial interpolation application at different spatial scales.

8.3 Spatial Prediction Uncertainty

One of the advantages of stochastic interpolation methods (see Section 2), as promoted throughout this document, is their ability to provide uncertainty estimates along with their spatial interpolation predictions. A limitation of the uncertainty estimates provided by kriging methods (i.e., spatial prediction standard errors) is that they are likely to underestimate the true uncertainty of the associated spatial predictions. This is true because these methods are essentially two-stage processes, where a variogram model is estimated in the first stage and prediction and uncertainty equations are applied in the second stage. The second stage equations treat the first stage variogram model as true and known. However, it is only an approximation of the data's empirical co-variation. The variogram model itself is estimated with uncertainty. Thus, the spatial prediction uncertainty estimates output from kriging methods do not account for the inherent uncertainty in estimating the required variogram model. These methods, therefore, are likely to underestimate the true uncertainty of their resulting spatial interpolation surface. Also, the estimated uncertainty of interpolated values can be quite sensitive to the variogram used to describe the spatial structure. This can be true even though the predictions themselves may be quite similar.

Any attempt to formally rectify the above limitation is beyond the current scope of this document. However, a couple of simple suggestions are worth noting. At a minimum, the documentation of any spatial interpolation exercise based on kriging methods can explicitly recognize their bias toward underestimating uncertainty. Decision makers and other users of the modeling results can judge for themselves how best to temper their interpretation based on this information. Another simple approach would be to inflate the resulting spatial prediction standard errors by some set amount (e.g., a fixed percentage). A reasonable choice for the amount of inflation may not be obvious. Some guidance might be provided by further consideration of the adequacy of the variogram model fit. In some sense, a better variogram fit suggests less of a need to inflate the final spatial prediction standard errors. More formally, some quantitative measure of the variogram model's fit to the empirical variogram (e.g., objective function, R^2 , etc.) could be used to define a framework for inflation adjustment. Any such choice of ad hoc adjustment, formal or informal, is itself a subjective decision and, therefore, should be supported by documented rationale.

There is another aspect to spatial prediction uncertainty that should be mentioned, and it is somewhat related to both the air pollutant of concern and spatial scale. This aspect is less obvious and more difficult to describe. Basically, kriging methods by definition are smoothing operators. Although the formulas are true to the data in some sense, the relatively smooth spatial interpolation surface they provide may not capture all the underlying variation of the process they are attempting to describe. Nonetheless, given that kriging methods are continuous interpolators as well, they are capable of providing spatial predictions at any point in space. It is, therefore, tempting to use the resulting kriging model to predict air pollution levels at specific points in space that may be of interest. While there is nothing inherently wrong with doing so, caution must be exercised when interpreting results specific to any one location, especially a location without nearby monitoring data.

A kriging spatial interpolation surface is expected to provide a reasonable spatial description of the pollution process in general, but to require it to be highly accurate at each specific location may be an unreasonable expectation. To some extent, the kriging method's spatial prediction standard errors are intended to capture such uncertainty. However, outside of an attempt to quantify uncertainty, the method generally cannot provide highly accurate predictions at every point in space across the entire domain of interest. This limitation will vary by pollutant and is probably more acute when interpolating across large spatial domains. Furthermore, this limitation is likely to be highly dependent on the spatial density and general network design associated with the available monitoring data.

8.4 Monitoring Network Design

In general, monitoring network design is highly interrelated with spatial interpolation modeling, in that the ability to interpolate well should depend strongly on the design of an existing network, while an explicit objective of spatial interpolation should impact how one designs new networks or re-designs existing ones. Thus, the issue of network design cuts across the limitations discussed in the previous sections. Specifically, factors such as the air pollutant of concern and the desired spatial scale of interpolation will affect spatial interpolation quality with respect to existing networks, and will affect design choices with respect to new networks or networks under consideration for re-design. In either case, all three factors (pollutant, scale, and network design) affect the spatial prediction uncertainty resulting from the application of an interpolation model.

In a very real sense, and stated in an overly-simplified fashion, spatial interpolation amounts to “connecting the dots” to fill in or complete a picture. All else being equal, the more space between “dots” the more uncertainty in our understanding of the picture for those areas filled in via spatial interpolation. Interpolation at larger spatial scales may require broad national or regional monitoring network data, while smaller scales may require dense urban or local networks. Furthermore, this statement must be qualified by the degree of spatial uniformity associated with different air pollutants. Monitoring network design encompasses the relative spatial density or sparsity of such “dots” along with other design issues. The discussion below first considers the issue of monitoring network density in more detail. Finally, some discussion is offered regarding quantitative methods for evaluating, and possibly improving, the overall design of existing monitoring networks for, among other uses, spatial interpolation modeling.

A primary issue of concern with respect to a network or collection of monitoring sites providing data for spatial interpolation modeling is the relative spatial density of the network. Generally speaking, the more monitoring locations providing data within the domain of interest, the better. Consistent with Tobler’s 1st Law of Geography (see Section 2), kriging methods often assume greater spatial correlation with shorter distances between points in space. Based on the kriging equations (see Appendix A), this translates to larger spatial prediction uncertainty at interpolation locations further from available monitoring data. Hence, more spatially dense monitoring networks tend to yield more certain spatial interpolation predictions throughout the domain of interest.

At some point, the relative spatial sparsity of a monitoring network can hinder a researcher’s ability to construct a desired spatial interpolation surface based solely on kriging methods applied to monitoring data. This, of course, will depend on the application at hand and the desired quality of the resulting interpolated surface. No simple rule is available to determine whether a monitoring network’s data are too sparse spatially to reliably apply kriging methods for a given application. A formal evaluation process, such as the DQO process described in Section 5, can provide useful insight for making this determination. Such an evaluation may lead researchers to conclude that kriging methods and monitoring data alone are insufficient for their given spatial interpolation needs. In this case, supplemental or new data might be pursued or a

different interpolation approach might be considered. Consistent with the discussion of Section 4, additional monitoring or other data that are more dense spatially can be acquired, and an approach of co-kriging or kriging with external drift can be entertained. Another option would be to consider more complex or sophisticated modeling approaches, such as those discussed in Section 9. The ability to pursue these alternatives will obviously depend on available time and resources.

Aside from spatial density, which is obviously an important design component, one must take into consideration the many other facets of monitoring network design. In particular, the problem of selecting optimal sampling designs is complicated by considerations such as current pollution patterns, meteorology, cost, accessibility, security of sampling sights, and politics. As a result, researchers such as Montserrat Fuentes and others have considered various approaches to design, and the resulting implications of these approaches on the estimation of spatial structures (e.g., spatial interpolation). In particular, Fuentes considers entropy approaches to network design as a way of generating optimal monitoring designs for SLAMS/NAMS and other air monitoring networks. Given its potential utility, this particular approach to monitoring network design is discussed in more detail below. Obviously, it is not the only approach to network design.

Fuentes defines the entropy of a random variable X as the negative of the expected value of the log of the density of X (Fuentes, et al., 2001). Practically speaking, this definition of entropy is meant to explain the uncertainty about X . Using this definition of entropy, it is then proposed that one choose the network design that maximizes the entropy (i.e., minimizes uncertainty). Fuentes assumes for her example that X is the vector of observations at all monitoring network sites, and demonstrates this entropy-based approach both for adding and for removing network sites. In each case, the entropy of the network is compared under each network design of interest and the design with the largest entropy is chosen. Observe that, in most cases, it is necessary to use data from an existing network to estimate the entropy (or any other optimization measure). Therefore, the particular focus in this application is the inter-relationship between network design and spatial interpolation (Fuentes, et al., 2001).

On the other hand, the design of a network can impact the results obtained from a spatial interpolation. As noted earlier in this document, spatial interpolations are limited by the scale and spatial domain of the data obtained from a monitoring network. In addition, if a monitoring network does not measure key features of a spatial region (for instance, if no monitors are placed near point sources or in highly rural areas), then the spatial interpolation techniques discussed in this document cannot accurately estimate the levels of interest in these key areas. As defined, the goal of Fuentes' entropy approach is to find the most informative set of monitoring sites. Thus, this and similar approaches can be useful for designing networks that will provide spatial interpolations with desirable properties.

8.5 References

- [1] Main, H. H., and Roberts, P. T. (2001). "PM2.5 Data Analysis Workbook." Draft workbook prepared for the U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards. Contract No. STI-900242-1988-DWB, Research Triangle Park, North Carolina.

- [2] Spicer, Chester W., Gordon, Sydney M., Holdren, Michael W., Kelly, Thomas J., and Mukind, R. (2002). Hazardous Air Pollutant Handbook: Measurements, Properties, and Fate in Ambient Air. Lewis Publishers, Boca Raton, Florida.

- [3] Cressie, Noel A. C. (1993). Statistics for Spatial Data. John Wiley & Sons, New York, New York.

- [4] Fuentes, M., Chaudhuri, A., Boos, D., and Holland, D. (2001). "Entropy Approaches for Air Pollution Monitoring Network Design", EPA Workshop, Research Triangle Park, North Carolina.

9.0 SPATIAL INTERPOLATION MODELING ALTERNATIVES

Section 8 of this document provided a discussion of the limitations associated with kriging using monitoring data, one common approach to spatial interpolation. This section addresses alternatives that might be applied when certain of these limitations are relevant. Throughout this document an attempt has been made to stress that any spatial interpolation requires assuming and fitting some sort of model. Thus, when additional complexities are needed, or when limitations of existing models present themselves, it is always possible to develop more complex models to address these issues. The following sections offer discussion on some of the issues that may be of interest. Section 9.1 discusses alternate non-statistical modeling approaches, such as air pollution dispersion modeling and ad hoc, or less formal, GIS-based analyses. Then, Section 9.2 talks about alternative statistical methodologies, including general linear models and (hierarchical) Bayesian techniques. Section 9.3 discusses a number of current and future research areas presented in a paper by Holland, et al. Finally, Section 9.4 lists relevant references cited in the text.

9.1 Alternative Non-Statistical Approaches

In the bulk of this document, the discussed spatial interpolation models have been based on data from monitoring networks. These sorts of models can be thought of as receptor-based. Specifically, they rely on monitoring receptors, which measure the impact of various source emissions upon dispersion through the atmosphere. However, there are other approaches to interpolation modeling. For instance, one can consider air pollution dispersion models. Rather than spatially modeling the outputs of a system (i.e., measurements of ozone or $PM_{2.5}$), dispersion models address the inputs of the system (e.g., information about point sources of pollutants or environmental parameters). Another alternative might be to focus on less formal GIS-based approaches. Although not rigorously quantitative, such ad hoc approaches can easily address much more complex data structures, such as combining a large number of potentially useful interpolation inputs. We discuss both air pollution dispersion models and GIS-based approaches in the following sections.

9.1.1 Air Pollution Dispersion Models

Recall that while (due to government regulations and QA/QC practices) the measurements from monitoring networks tend to be accurate (in the sense that they tend to be precise with little to no bias), they also tend to be expensive to obtain, which leads to them being spatially and temporally sparse. On the other hand, estimates generated by air pollution dispersion models are relatively inexpensive to obtain and, thus, can provide measurements everywhere of interest at any time of interest. Unfortunately, while dispersion models can describe specific sorts of behaviors well (for instance, concentration spikes of a given pollutant near a source) they are not guaranteed to be as accurate as the measurements from a monitoring network. Specifically, dispersion models tend to be less precise and have the potential to be biased, due to the limitations of the inputs required for such models (e.g., emissions inventories).

EPA and others make use of a number of dispersion models, each with its own strengths and weaknesses. Included among these models are CMAQ, the community multi-pollutant air quality model; ASPEN, the assessment system for population exposure nation wide; REMSAD, a regulatory modeling system for aerosols and deposition; and others. CMAQ is a multi-pollutant, multi-stage model. ASPEN is a large-scale Gaussian dispersion model. REMSAD is a three-dimensional grid-based Eulerian air quality model that simulates concentrations and deposition of atmospheric pollutants over large spatial scales.

Air pollution dispersion models are part of a much larger class of mathematical and physical models that can be used for spatial interpolations. In contrast to kriging (and other statistical models) that can become less reliable away from observed monitoring data, the proper physical models can describe a process anywhere for which the model is well-defined. This does not imply that such models are accurate everywhere; merely that they can represent specific sorts of behavior well. For example, air pollution dispersion modeling well-describes the impact of point sources of air pollution on a spatial environment, whereas certain monitoring networks are designed so that monitors are located in populated areas and away from immediate pollutant impacts. This is so that they are not unduly influenced by point sources, but it implies that they cannot describe the point source peaks nearly as well when used for interpolation. Theoretically, mathematical and physical models can be generated to describe nearly any desired property of a spatial (or other) process, though often they may lack accuracy, precision, or other useful properties.

In addition to being spatial interpolations of a sort by themselves, air pollution dispersion models might be used in conjunction with monitoring data in various ways. For instance, consider the approaches discussed in Section 4 of this document. The results from a dispersion model might be used as a covariate when performing kriging with external drift. The results from air pollution dispersion models might be used in other ways as well, as we discuss below.

9.1.2 GIS-Based Approaches

As noted in Section 6, Geographic Information Systems (GIS) can be quite useful when conducting spatial investigations. They can provide graphical depiction of data, and some such systems can actually do kriging directly (for example, the suite of products from ESRI). However, the utility of GISs does not necessarily end there. If the analyst is willing to consider less formal techniques, much information contributing to spatial interpolations can easily be capitalized on through the use of GISs.

GIS technology represents, at its simplest, computerized mapping and geographic analysis techniques that provide the ability to integrate data from various sources and spatial/temporal scales. GIS analysis lets you see patterns and relationships in spatial data. The results of such analyses can provide insight into a place or help in the process of choosing between options. Mapping the extremes and the density of features may facilitate the understanding of the relationships between places and the patterns of concentrations of interest. GISs typically have tools to identify what is inside or near a particular location or region. This

ability allows the identification, summarization, and comparison of information within and around key regions. It can also assist in identifying which regions are affected by a particular activity (for instance, point source air pollution emissions). Finally, GISs can help map where things move, or how conditions change in a region over time. Knowing what has changed can lead to an understanding of how things are behaving over time, assist in anticipating future conditions, and facilitate the evaluation of the results of an action or policy.

One of the key strengths of GISs is the ability to combine information via data layers. A data layer associates a specific set of information to spatial locations (for instance, the location of monitoring sites, annual average air temperature by county, etc.). Thus, by combining multiple data layers in a GIS map of a region of interest, various combinations of information can be displayed on a single map quickly and easily. This ability to incorporate a broad class of information, linked by location, in a spatial analysis is similar to kriging with external drift as discussed in Section 4, though it is a much more informal approach. The flexibility of geographic information systems allow them to store information in a continuous (or point-based) fashion and in an abrupt fashion (such as on a lattice, in grid cells, or by county). This flexibility facilitates the combination of information at various spatial scales, a problem which can be complex in standard statistics-based analysis. All of these features add up to the ability of GISs to allow a broader integration of inputs than the statistical methodologies discussed in this document, which in turn can facilitate a wide range of spatial interpolation utilities.

Consider Hogsett, et al., (1997) as an example of GIS utility in a spatial analysis and risk characterization application. In this paper, the authors were interested in the spatial characterization of ozone impact to forests. In their preliminary examination of the data, they observed that, in general, monitoring sites tended not to be located in rural forested areas. This meant that standard spatial interpolation techniques might not work in an optimal manner. (Recall that interpolation tends to become unreliable as you move away from data locations. See Section 8 for a more in-depth discussion of monitoring network design.) They also observed that a GIS could be used to integrate empirical data of growth effects, regional ozone exposure, and environmental factors across forested areas. Thus, Hogsett and his fellow authors developed and implemented an informal method of combining a number of spatial inputs to estimate monthly ozone concentrations in non-monitored areas.

In the above example, the authors took full advantage of the capability of a GIS. They were able to map monitoring locations and forested areas to identify the lack of representativeness of monitoring observations in regions of interest. They spatially interpolated observed and estimated values to consider spatial behavior. They combined layers of information visually and mathematically to explore properties of interest. All told, they were able to use informal methodology and GIS technology to provide a dense grid of exposure estimates that does not necessarily suffer with distance from monitoring sites. In summary, although GIS-based interpolation methods are limited in their quantitative rigor, which can result in a lack of uncertainty characterization, they make up for this limitation through their facilitation of spatial data integration.

9.2 Other Statistical Methodologies

Almost by definition, environmental science is a statistics problem. For example, many scientists observe that climate-scale predictions are not really point predictions, but rather the construction of probability distributions of future weather. Recognizing the relationships between such modeling and the spatial interpolation problems discussed in this document, we consider other statistical methodologies used in environmental science for application to spatial interpolation. In particular, we focus on generalized linear modeling and hierarchical Bayesian models in the following two sections.

9.2.1 General Linear Models

In many fields of science, relationships are so exact that they can be expressed by a function such as $Z = f(\mathbf{x})$. Consider as examples Ohm's laws, Boyle's gas law, Kirchoff's law in electricity, Newton's laws of force and acceleration, Newton's law of cooling, and so forth. When Z cannot be observed directly, but is instead observed with measurement error, we can replace Z with Y , and obtain $Y = f(\mathbf{x}) + \epsilon$. Observe that another justification for using this form arises when quantities are not related functionally, but in a much more obscure manner. For example, suppose $Z = g(\mathbf{x}, \mathbf{x}_1) = f(\mathbf{x}) + h(\mathbf{x}, \mathbf{x}_1)$. When $h(\mathbf{x}, \mathbf{x}_1)$ is unknown and small relative to $f(\mathbf{x})$, the formulation $Y = f(\mathbf{x}) + \epsilon$ provides a reasonable approximation. Many approximation methods, including Taylor's polynomial approximation, treat functions in just this manner; namely, as a simpler, well-understood function plus an error term. Specifically, a common form of $f(\mathbf{x})$ is $\mathbf{X}'\boldsymbol{\beta}$, the linear combination of the variables, \mathbf{X} , and some combination of constants, $\boldsymbol{\beta}$. Thus, in many physical problems, including spatial interpolation, a general linear model (GLM) may apply.

To understand how a general linear model is defined, consider the following matrix equation:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where \mathbf{Y} is an $n \times 1$ observable random vector, \mathbf{X} is an $n \times p$ matrix of observable numbers, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unobservable parameters, and $\boldsymbol{\epsilon}$ is an $n \times 1$ unobservable random vector with a mean of 0 (i.e., $E[\boldsymbol{\epsilon}] = \mathbf{0}$) and some covariation structure denoted by \mathbf{E} (i.e., $\text{cov}[\boldsymbol{\epsilon}] = \mathbf{E}$). These specifications define a general linear model. Consider how these specifications might apply to a spatial analysis of air quality data. Specifically, if one were to apply a general linear models approach to the spatial interpolation of ozone given monitoring observations, one could interpret the parameters of the model as follows. \mathbf{Y} would be the vector of ozone monitoring observations throughout the region of interest. We recognize that such observations are observed with error, so this is an observable random vector. \mathbf{X} is a matrix of numbers associated with those observations. Specifically, if one considers a spatial interpolation on location alone, for example, each row of \mathbf{X} would include the spatial coordinates associated with the locations of the ozone monitoring data (possibly expanded to include the spatial coordinates multiplied by one another and/or raised to various powers). $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$ have the same interpretation as before, but now one

would estimate \mathbf{Y} in order to have the parameters needed to produce the spatial interpolation surface. That is, using the observed ozone data, one estimates the \mathbf{Y} values. Then, by plugging in values of \mathbf{X} in locations where estimates of the surface are desired, $\mathbf{X}\mathbf{Y}$ provides a spatial interpolation surface. Note, however, that this implies that a complete set of values of \mathbf{X} are required at every location where a spatial prediction is desired. This is trivial when only location is included in the analysis; however, general linear models are not limited to location.

In many ways, the form of \mathbf{X} is limited only by the investigator's imagination. \mathbf{X} can include, in addition to location information, any covariates that may be of interest (e.g., land use, climate data, and even other observations of the property of interest). This bears a certain similarity to kriging techniques discussed in Section 4, such as co-kriging and kriging with external drift. However, it is important to note that general linear modeling treats large- and small-scale trends in a different manner than kriging techniques. [For more on general linear models, see Graybill (1976).]

It is interesting to note that \mathbf{E} , the data's co-variation structure in the general linear model, can be made analogous to the variogram models discussed in Section 3. One of the strengths of general linear models over more common linear models techniques is that GLMs do not assume independence between observations. In fact, \mathbf{E} can take nearly any form, subject to certain theoretical constraints on the form of a variance matrix. Thus, if an investigator has a strong sense of the correlation between monitoring sites within the region of interest, they can use the variogram estimation techniques discussed in Section 3.3 to design the \mathbf{E} for a given GLM.

It is also important to note that many statistical software packages, including both SAS and S-Plus, can compute GLMs. That is, S-Plus, SAS, and other such packages can compute estimates of the values of \mathbf{Y} as well as values of $\mathbf{X}\mathbf{Y}$ for any fully specified \mathbf{X} . In particular, the SAS procedure PROC MIXED is quite powerful, allowing the user to specify many options, including the variogram parameters (nugget, sill, and range) described in Section 3.3.

9.2.2 Hierarchical Bayesian Models

Bayesian models are a powerful class of statistical models for all types of environmental data, especially for the modeling of the spatial and temporal distribution of air pollutants. Bayesian analysis is well-suited for estimation and prediction based on the combination of relevant scientific understanding and data. One can draw important parallels between the scientific method and the Bayesian paradigm. Think about the scientific method as an iterative process. First one formulates a hypothesis, rooted in current knowledge. Then data are collected, against which the hypothesis can be tested. Finally, one updates their knowledge about the truth in light of the data, repeating the process as appropriate. The Bayesian paradigm is similar, replacing current knowledge with prior probability distributions, data with a likelihood, and updated knowledge with posterior probability distributions, which may now serve as prior probability distributions for future studies.

Hierarchical Bayesian analysis represents a larger modeling paradigm for the environmental sciences into which all of the techniques discussed in this document might be incorporated. The hierarchical Bayesian viewpoint offers a natural structure for dealing with the complexities arising in the environmental sciences. For instance, one can envision writing down a sequence of conditional probability models, for example, the marginal distribution of a large-scale trend; the conditional distribution of a regional-scale trend given the large-scale trend; the conditional distribution of a neighborhood-scale trend given the region behavior; and the distribution of measurement error conditional on the neighborhood-scale levels. Probability theory can then provide formulas for the combination of these distributions yielding an overall distribution that can be used to estimate spatial surfaces of interest, across different spatial scales, given the observed pollutant measurements.

Hierarchical Bayesian methods can also be used to combine multiple sources of information about a given spatial process using probability distribution theory. This concept is similar to the data layering procedure of GIS methods, but done in a more quantitative and rigorous fashion. For instance, consider the relative strengths and weaknesses of network monitoring data and air pollution dispersion modeling as discussed in Section 9.1.1. The measurements from monitoring networks tend to be accurate but they also tend to be expensive to obtain, which leads to them being spatially and temporally sparse. On the other hand, air pollution dispersion model predictions are relatively inexpensive to obtain and, thus, can provide measurements everywhere of interest at any time of interest. Unfortunately, while dispersion models can describe specific sorts of behaviors well, they are not guaranteed to be as accurate as the measurements from a monitoring network. Similar to, but more rigorous than, the GIS approach discussed above in Section 9.1.2, hierarchical Bayesian models can be used to combine data about the same spatial region from these two information sources to potentially produce a dense spatial interpolation with improved overall accuracy.

As powerful as these methods can be, unfortunately they can also be quite complicated, requiring advanced technical and theoretical statistical training. Specifically, hierarchical Bayesian solutions typically require highly-intensive computing techniques, such as MCMC (Markov Chain Monte Carlo) simulations. Fortunately, WinBUGS, a software package developed at Cambridge for performing MCMC estimation of Bayesian models is freely available. WinBUGS uses an S-Plus-like syntax to define probability distributions. A strength of WinBUGS is that it was developed by researchers who specialize in Bayesian simulations. Thus, it contains some sophisticated tools to automatically optimize MCMC estimation. In addition, WinBUGS also incorporates a graphical interface that allows the user to use a point-and-click interface (instead of a programming interface) to build hierarchical models. The package is capable of handling most Bayesian models, but occasionally it has difficulties with complex models being applied to large data sets. This weakness is confounded by the fact that WinBUGS was not specifically designed for spatial applications, thus limiting the sorts of spatial models that might be applied. In addition, data entry into WinBUGS is awkward and must be conducted each time an analysis is run, making large data sets (such as are associated with many

spatial interpolations) difficult to work with. WinBUGS is available electronically at <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml#intro> . As freeware, support for WinBUGS is limited. Documentation, examples, and information about the users group are available from the website. In conclusion, WinBUGS is a useful tool for MCMC estimation of Bayesian models, but it does have certain limitations.

Other than WinBUGS, software for developing hierarchical Bayesian models is not generally available. Thus, theoretical development of such models can require potentially complex custom computer programming. Therefore, we stress that hierarchical Bayesian solutions, despite their advantages, are not always easily obtained and not necessarily practical to implement.

9.3 Flexible Models Handling Nonstationarity

In a recent paper by Holland, et al. (2002), the authors review the current state of the science of spatial interpolation and discuss a number of innovative techniques for moving beyond spatial interpolation techniques that focus on simple correlation functions. They present various models for accommodating nonstationarity and heterogeneous covariance structures. Sampson, et al., in their EPA-funded report, reference some of these same techniques and others, many being “essentially nonparametric approaches to the modeling of nonstationary covariance structure” in the authors’ words.

Holland, et al., begin by noting that “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.” These approaches include a moving window approach that separately estimates first and second order stationarity within each window of space and a kernel smoothing approach for estimating covariances. They then present a series of more sophisticated models for estimating global nonstationarity using empirical orthogonal functions, process-convolution, spatial deformation, and thin-plate splines. The paper also discusses hierarchical Bayesian approaches, introduced in Section 9.2, for temporal and spatial interpolation. They cite a major advantage of these methods to be the ability to incorporate the uncertainty in the mean and spatial covariance of the spatial field in the predictive distribution.

Holland, et al., also list a few areas that still require further research. These include:

- imputation techniques for spatial-temporal data sets with large amounts of missing data;
- diagnostic methods for evaluating alternative spatial models and characterizing nonstationarity;
- methods to combine monitoring data and dispersion model output in order to improve spatial prediction and validate model output;
- methods to incorporate time-dependence in space-time models; and
- development of software to improve analyses of large data sets.

In conclusion, Holland, et al., argue that spatial and spatial-temporal models can provide an important link between the scientific community and the regulatory monitoring community. Regulators can apply these models to gain a better understanding of complex air-quality issues and to assist with apportioning scarce resources.

9.4 References

- [1] Berliner, L. M., Wikle, C. K., and Cressie, N. (2000). "Long-lead prediction of Pacific SST's via Bayesian dynamic modeling." *Journal of Climatology*.
- [2] Fuentes, M., and Smith, R. (2002). "A new class of nonstationary spatial models." Under review for *JASA*.
- [3] Graybill, F. A. (1976). Theory and Application of the Linear Model. Wadsworth & Brooks, Pacific Grove, California.
- [4] Hogsett, W. E., Weber, J. E., and Tingey, D. (1997). "Environmental Auditing: An Approach for Characterizing Tropospheric Ozone Risk for Forests." *Environmental Management* (21) No. 1, pp 105-120. Springer-Verlag, New York.
- [5] Holland, D.M., Cox, W.M., Scheffe, R., Cimorelli, A.J., Nychka, D., and Hopke, P.K. (2003). "Spatial Prediction of Air Quality Data." *EM*, pp 31-35, August.
- [6] Mitchell, A. (1999). The ESRI Guide to GIS Analysis. ESRI Press, Redlands, California.
- [7] Sampson, P.D., Damian, D., and Guttorp, P. (2001). "Advances in Modeling and Inference for Environmental processes with Nonstationary Spatial Covariance." National Research Center for Statistics and the Environment Technical Report Series. NRCSE-TRS No. 061.
- [8] Stern, H., and Cressie, N. (2000). "Posterior predictive model checks for disease mapping models." *Statistics in Medicine*.
- [9] Wikle, C.K., Milliff, R.F., Nychka, D., and Berliner, L.M. (2001). "Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds." *JASA*, Vol. 96, pp 382-397.

10.0 SPATIAL INTERPOLATION MODELING REFERENCES

The following section contains summaries of references on various aspects of spatial interpolation. The references are categorized by topic to loosely correlate with the sections of this document. References that discuss multiple topics are located in the section that seems most appropriate. This set of references is certainly not exhaustive, as the topic of spatial interpolation has been researched and written about extensively, especially in the last 10 years. These references can, however, serve as a starting point for learning more about specific topics. Some of the web sites included in this section and also in other locations within the document contain links to other spatial interpolation resources that users can investigate to learn more.

10.1 General Spatial Interpolation

The references listed below contain more general background information on spatial interpolation as opposed to being specific to a particular interpolation method or issue.

"Environmental Auditing, An Approach for Characterizing Tropospheric Ozone Risk to Forests;" William E. Hogsett, James E. Weber, David Tingey, Andrew Herstrom, E. Henry Lee, and John A. Laurence; in Environmental Management Vol. 21, No. 1, pp 105-120, 1997. <http://link.springer.de/link/service/journals/00267/papers97/21n1p105.pdf>.

The authors of this paper perform a spatial interpolation of ozone exposure in order to estimate potential biomass loss in forested areas as a result of that ozone exposure. Because only two percent of all ozone monitoring sites are located in forested areas, the authors felt that estimates based on traditional interpolation techniques (IDW, kriging) that utilize only distance data would be highly questionable. Thus, they used a GIS (the paper does not say what specific system was used) to predict an ozone surface across the eastern U.S. for 1988 and 1989 based on emissions of NO_x, temperature, cloud cover, wind direction, elevation, and distance from sources. Subsequently, they fed the ozone predictions into a risk characterization model that calculated predicted biomass loss in eight tree species. The risk characterization interpolation model utilized the ozone predictions, empirical data on biomass change as a function of ozone exposure, and information on species' locations. Findings indicated that "annual biomass losses ranged from 0 to 33 percent depending on species and concentration of ozone across the range of species."

The authors point out that there is uncertainty present in their predictions, which is difficult to quantify. One cause is the ozone predictions for which "there is no means to estimate the certainty of the predictions in nonmonitored forests." Nevertheless, they conclude that "even with the uncertainties, this preliminary assessment has indicated the usefulness of and potential for a spatial approach to assessing the extent and magnitude of risk that tropospheric ozone presents to the forested areas of the eastern United States based on current estimated ozone concentrations."

“Clean Air Act Ozone Design Value Study: A Final Report to Congress,” EPA 454/R-94-035, December 1994.

This document titled “Clean Air Act Ozone Design Value Study: A Final Report to Congress” is a report prepared by EPA/OAQPS on the methodology currently in use for calculating design values to determine if the calculated design value provides a reasonable indicator of the ozone air quality of ozone attainment areas. The document contains much discussion regarding different ways to calculate design values, but does not contain much direct discussion of spatial interpolation or kriging. Chapter 8 on “Assessing the Impact of Transported Ozone and Precursors” includes discussion of illustrating ozone transport by performing spatial interpolation that translates monitoring data from a discrete number of monitoring sites (usually located in or around major cities). Single hour ozone predictions were generated using the Regional Oxidant Model (ROM). Discussion also covers the Transported Ozone Design Value computer model, which assists in determining likely source regions associated with high ozone concentration events.

Chapter 10 on “Alternative Air Quality Indicators” discusses temporal and spatial variability in ozone indicators and how well the temporal and spatial distribution is represented by the design value. Ozone data from the South Coast Air Basin is analyzed to illustrate spatial variability. “Given the spatial variability in ozone observed across urban areas, one cannot expect a single numerical value, such as the design value, to adequately describe concentration gradients.”

Center for Spatially Integrated Social Science Meeting on Spatial Data Analysis Tools
[\(\[www.csiss.org/events/meetings/spatial-tools\]\(http://www.csiss.org/events/meetings/spatial-tools\)\)](http://www.csiss.org/events/meetings/spatial-tools).

The Center for Spatially Integrated Social Science (CSISS) is a NSF-funded project based at the University of California at Santa Barbara. The website states that the CSISS Mission “recognizes the growing significance of space, spatiality, location, and place in social science research.” CSISS’s mission statement pronounces that the goal of CSISS “is to integrate spatial concepts into the theories and practices of the social sciences by providing infrastructure to facilitate: 1) the integration of existing spatial knowledge, making it more explicit, and 2) the generation of new spatial knowledge and understanding.”

As part of the “Spatial Analytic Tools” program, CSISS is hosting a two-day “Specialist Meeting on Spatial Data Analysis Software Tools” on May 10-11, 2002. The website states that the “meeting will bring together software developers from both the public/academic sector as well as the private sector who deal with tools to visualize spatial data (geovisualization), carry out exploratory spatial data analysis (ESDA) and facilitate spatial modeling (spatial regression modeling, spatial econometrics, geostatistics), with a special focus on the potential for social science applications. These tools include a range of different approaches, such as macros and scripts for commercial statistical packages or GISes, modules developed in open source statistical and mathematical toolkits, and free standing software programs. The focus of the meeting is on software ‘tools’ rather than on the methods per se.”

Although CSISS's mission is not specific to spatial interpolation of air quality or environmental data, the tools discussed at this meeting are applicable to this document. The Proceedings Volume from this meeting, which will contain all the papers submitted for discussion, and the work of CSISS in general may be a useful reference for those interested in spatial interpolation.

21st University of Leeds Annual Statistical Research Workshop
(www.maths.leeds.ac.uk/statistics/workshop/).

In recent years, the themes of the Leeds Annual Statistical Research Workshop sponsored by the University of Leeds (England) Statistics Department have reflected the growing interest in image and shape analysis. This year the workshop being held on July 3-5, 2002, is focused on statistical analysis of large data sets. Specific sub-topics include bioinformatics, data mining, functional and image analysis, and speech recognition. Session topics on applications includes environmental modeling, image analysis, and industrial applications.

10.2 Ordinary Kriging

This set of references on ordinary kriging provides some historical perspective on the development of optimal interpolation and kriging methods, and also discusses specific applications of kriging methods.

“Optimum Interpolation of Spatial Air Pollution Data;” N.D. Van Egmond and O. Tissing, National Institute of Public Health, Bilthoven, The Netherlands; Proceedings of the 4th International Clean Air Congress; Tokyo, Japan; May 16-20, 1977.

This is an older paper that discusses finding an “optimal interpolation scheme for daily average spatial data.” The authors use SO₂ data from the National Monitoring System in the Netherlands to test an optimal interpolation scheme “in which the coefficients are chosen so as to minimize the mean square interpolation error.” The interpolation scheme presented obtained better results than another method presented for comparison purposes which used a “linear combination with negative exponential distance weighting coefficients.” Because it is 25 years old, we do not believe the work is necessarily relevant to recent methods and techniques being developed for spatial interpolation.

“Multiobjective Air-Pollution Monitoring Network Design;” Arturo Trujillo-Ventura and J. Hugh Ellis; Johns Hopkins University; Atmospheric Environment, Vol. 25A, 1991.

The authors discuss the utilization of objective functions that optimize the locations of air monitoring stations. In optimizing the design of a monitoring network, “pollutant concentrations at any point in the domain must be interpolated from concentration data at locations where stations exist.” The interpolation method the authors describe is “simple punctual kriging.” Thus, kriging is used as an integral step in determining the optimal design of an air pollution monitoring network. The kriging process provides estimates of the interpolation error of

pollution i at point x given that measurement stations are located at points u . The model was applied to design an air quality monitoring network measuring four pollutants (SO_2 , nitrogen oxides, PM, and unsaturated hydrocarbons) for a region surrounding the city of Tarragona, Spain.

Acidic Deposition: State of Science and Technology (Report 5) - "Evaluation of Regional Acidic Deposition Models (Part I) and Selected Applications of RADM (Part II);"
Robin Dennis, U.S. EPA, et al; National Acid Precipitation Assessment Program;
September 1990.

This report summarized the existing knowledge regarding the evaluation of regional acidic deposition models. Authors investigated how well models predict "1) the change in regional acidic deposition that would result from changes in precursor emission, 2) the influence of sources in one region on acidic deposition in other sensitive receptor regions, 3) the levels of acidic deposition at certain sensitive receptor regions, and 4) the acidic deposition due to emissions transported across geographical/political boundaries, including the United States-Canadian border." The report describes the use of kriging to generate interpolations of wet sulfur deposition in the International Sulfur Deposition Model Evaluation (ISMDE) that evaluated the ability of models to replicate the spatial patterns of wet sulfur deposition within the uncertainties of the data. Contour maps of kriging predictions for each of the models evaluated are presented. Kriging was also used to evaluate a proposed surface monitoring network. Based upon the kriging uncertainty analysis, it was determined that additional sites needed to be added in Canada.

10.3 Extensions to Ordinary Kriging

The references below provide information on extending the ordinary kriging model, as discussed in Section 4.

"Temporal and Spatial Distributions of Ozone in Atlanta: Regulatory and Epidemiologic Implications;" ***James A. Mullholland, Andre J. Butler, James G. Wilkinson, and Armistead G. Russell, Georgia Institute of Technology, and Paige E. Tolbert, Emory University; Journal of Air and Waste Management Association, Vol. 48, May 1998.***

The study discussed in this paper utilized a universal kriging approach to interpolate near-surface ozone levels in the Atlanta area to a regular grid. The authors' purpose for doing this was to investigate relationships between ozone concentrations and increases in the pediatric asthma rate because "quantitative assessments of relationships between ambient levels of air pollutants, such as ozone, and health outcomes, such as asthma exacerbation, are needed to help establish protective levels for the NAAQS."

The authors state that, "spatial resolution of the ozone data was carried out by a universal kriging method using the GIS ARC/Info. In kriging, a smooth surface is estimated from irregularly spaced data points on the assumption that the spatial variation in the feature is homogenous over the domain and depends only on the distance between sites. The universal

kriging procedure, unlike ordinary kriging, incorporates a drift function to account for a structural component in the spatial variation.” The study found that, “on average, the temporal variation in daily maximum ozone concentration was greater than the spatial variation estimated using a universal kriging method, particularly when assessed on a population basis.”

“Preparation of Soil Sampling Protocols: Sampling Techniques and Strategies;” EPA/600/R-92/128; 1992.

The Environmental Monitoring Systems Laboratory within the Office of Research and Development (ORD) at EPA sponsored the development of this document titled “Preparation of Soil Sampling Protocols: Sampling Techniques and Strategies.” Within Section 3, “Statistical Concepts that Pertain to Soil Sampling,” there is a discussion of using geostatistical techniques such as kriging, for “soil mapping, isopleth development, and spatial distribution of soil and waste properties.” The use of punctual and block kriging to interpolate within the system of soil samples is discussed, with block kriging being suggested as the most useful for pollutant studies.

“A Hierarchical Approach to Multivariate Spatial Modeling and Prediction;” J. Andrew Royle, U.S. Fish and Wildlife Service, and Mark Berliner, Ohio State University; Journal of Agricultural, Biological, and Environmental Statistics, Vol. 4, No. 1., pp. 29-56; 1999.

The authors propose a multivariate interpolation method that is similar to co-kriging and KED methods. They apply their proposed methodology to jointly predict daily ozone levels and maximum daily temperatures. The ozone data are from the AIRS database and contain data from 147 sites over 89 days in 1987. The temperature data are from the Solar and Meteorological Surface Observation Network (SAMSON) database. To illustrate their approach, they predict temperature and ozone levels for a single day.

The authors discuss co-kriging and KED and then introduce their hierarchical approach that attempts to address some of the limitations and difficulties faced when using co-kriging or KED. Much of the paper explains the statistical foundation of their methodology. Subsequently, the model is tested through the joint prediction of ozone concentrations and daily maximum temperature. The results observed using this hierarchical model reduced the prediction standard error by about 30 percent over that found using ordinary kriging. The authors indicate that a useful application of the model would be in facilitating the design of more efficient ozone monitoring networks.

10.4 Software References

Some of the references below are also discussed in Section 6 of the document. The GEO-EAS software package is not referenced in Section 6, but is discussed briefly below.

Geophysical Statistics Project website (<http://www/cgd.ucar.edu/stats/index.shtml>).

The Geophysical Statistics Project (GSP) is part of the Climate and Global Dynamics Division of the NSF-funded National Center for Atmospheric Research (NCAR). The website states that the mission of the GSP is to pursue “the innovative application and development of statistical methodology to address problems faced in the Earth sciences. A complementary activity is to generalize specific problems in the geophysical sciences to broad based statistical research.” The major research areas include:

1. Extension of statistical methodology for spatial processes and space/time processes
2. Application of modern regression and model selection to the analysis of geophysical data
3. Deriving a statistical basis for forecasting including the assimilation of observational data with numerical models
4. Modeling complicated physical processes through Bayesian hierarchical models
5. Understanding physical processes through the use of dynamical systems and nonlinear time series.

The principal investigators at GSP include Rick Katz, Joe Tribbia, and Doug Nychka. Researchers at GSP have developed software for analyzing spatial data that are available via the website. GSP researchers developed FUNFITS, which evolved into FIELDS, “a suite of [R,S] functions focused on the analysis of spatial data including large data sets and nonstationary covariance models and simulations. The major methods implemented include cubic and thin plate splines, universal Kriging, and Kriging for large data sets. The main feature is that any covariance function implemented in R,S can be used for spatial prediction.”

GSP researchers also developed the Design Interface (DI) software that contains “graphical and interactive tools for evaluating and modifying spatial designs.” The website states that current version “supports EPA in determining the sensitivity and coverage of air quality monitoring networks for different pollutants. DI is written in the S language with S-PLUS GUI extensions and currently runs on S-PLUS for Windows.” Also, “part of DI’s usefulness is that an experienced S-PLUS user can do some spatial analysis using command-line functions and then present the results using the DI GUI functions. At the lowest conceptual level, DI manipulates two kinds of objects: a covariance object and a network object The functions in DI are used to create or modify a network object or are used to examine the accuracy when spatial predictions are made from observations on this network.”

GSP funds post-doctoral opportunities to provide young statisticians the chance to “branch out into new areas and develop applications in the geophysical and environmental sciences.” Former GSP post-docs hold positions at universities and agencies across North America.

GEO-EAS Software.

The Geostatistical Environmental Assessment Software (Geo-EAS) system is a spatial interpolation software package developed by EPA. Evan Englund is one of the developers of the Geostatistical Environmental Assessment Software (Geo-EAS) system and co-author of the Geo-EAS 1.2.1 User’s Guide. The March 1991 version of the User’s Guide (EPA 600/4-88/033) was authored by Mr. Englund, of the U.S. EPA Environmental Monitoring Systems Laboratory located within the Office of Research and Development, and Allen Sparks, of Computer Sciences Corporation.

The abstract of the user’s guide states “the principal functions of the package are the production of 2-dimensional grids and contour maps of interpolated (kriged) estimates from sample data. Other functions include data preparation, data maps, univariate statistics, scatter plots/linear regression, and variogram computation and model fitting. Extensive use of screen graphics such as maps, histograms, scatter plots, and variograms help the user search for patterns, correlations, and problems in a data set. Data maps, contour maps, and scatter plots can be plotted on an HP compatible pen plotter. Individual programs can be run independently; the statistics and graphics routines may prove useful even when a full geostatistical study is not appropriate.”

In Section 1, Introduction, the guide states that “estimation of the variogram from sample data is a critical part of a geostatistical study. The procedure involves interpretation and judgement, and often requires a large number of ‘trial and error’ computer runs. The lack of inexpensive, easy-to-use software has prevented many people from acquiring the experience necessary to use geostatistical methods effectively. This software is designed to make it easy for the novice to begin using geostatistical methods and to learn by doing, as well as to provide sufficient power and flexibility for the experienced user to solve real-world problems.”

A very useful section of the User’s Guide for readers of this spatial interpolation document is Section 4, Using Geo-EAS in a Geostatistical Study: An Example in which the guide presents an example of how to conduct a geostatistical study. The example data in the guide are composed of 60 soil samples being analyzed for cadmium. The guide walks through the process of exploring the data set, computing and modeling variograms, producing grids of interpolated point or block estimates through kriging, and producing contour maps of kriging estimates.

Using ArcGIS Spatial Analyst, Jill McCoy and Kevin Johnston, ESRI, Redlands, CA; 2001.

Chapter 7 of the ArcGIS Spatial Analyst Manual is titled “Performing Spatial Analysis” and contains a concise discussion of three interpolation methods available in the Spatial Analyst package — Inverse Distance Weighted, Spline, and Kriging. The chapter includes a description of the theory behind interpolation and then presents an easy-to-understand review of what each interpolation method is. The kriging section describes the two tasks necessary to make a prediction: (1) to uncover the dependency rules (autocorrelation) and (2) to make the predictions. The manual quickly covers the two steps of creating variograms and covariance functions to estimate autocorrelation and predicting the unknown values. This manual does not discuss the assumptions behind kriging except that with Ordinary Kriging it is assumed that a constant mean is unknown. It does include a good discussion of how to evaluate semivariograms including definitions of the sill, nugget, and range.

The ESRI web site contains the following comparison of the Spatial Analyst and Geostatistical Analyst software packages: “Where ArcGIS Spatial Analyst includes rudimentary interpolation methods, ArcGIS Geostatistical Analyst expands the number of deterministic and geostatistical interpolation methods and provides many additional tools. In addition, Geostatistical Analyst provides a variety of different output surfaces such as prediction, probability, quantile, and error of predictions. All surfaces can be displayed as grids, filled contours, contours, and hillshades or any combination of these renderings. These surfaces can be exported in raster and shapefile formats for working together with other extensions such as ArcGIS Spatial Analyst. Geostatistical Analyst also includes a powerful set of exploratory spatial data analysis (ESDA) tools for exploring the distribution of the data, identifying local and global outliers, looking for global trends in the data, and understanding the spatial dependence in the data.”

Using ArcGIS Geostatistical Analyst, Kevin Johnston, Jay M. Ver Hoef, Konstantin Krivoruchko, and Neil Lucas; ESRI, Redlands, CA (2001).

ESRI’s ArcGIS Geostatistical Analyst software provides an extensive set of tools for performing spatial interpolation. In addition to discussions of deterministic and kriging methods, the Using ArcGIS Geostatistical Analyst manual also includes discussion and instruction for performing advanced techniques such as universal kriging, co-kriging, and disjunctive kriging. Similar to Section 3 of this document, Chapter 2 of the manual provides a step-by-step tutorial for performing a spatial interpolation exercise. It walks the user through a five step process — represent the data, explore the data, fit a model, perform diagnostics, and compare the models — for fitting a surface. The data used in the tutorial are 1996 maximum eight-hour ozone concentrations from the California Air Resources Board. The Geostatistical Analyst software also provides functionality to perform cross-validation — comparing two different types of models, the same model with different parameters, or the modeled data vs. measured values.

10.5 Alternative Approaches to Spatial Interpolation

The references below provide some further discussion of topics raised in Sections 8 and 9, issues to consider when using spatial interpolation techniques, and alternative approaches to interpolation beyond kriging. These references are separated into three groups: those dealing with interpolation methods beyond kriging, those discussing different types of dispersion or deposition modeling, and those discussing methods to utilize data from different spatial scales.

Spatial Interpolation Methods Beyond Kriging

“Objective Analysis of Air Pollution Monitoring Network Data: Spatial Interpolation and Network Density;” N.D. Van Egmond and D. Onderdelinden; National Institute of Public Health, Bilthoven, The Netherlands; Atmospheric Environment, Vol. 15, No. 6; 1981.

The authors compare the results of three interpolations of data from the Dutch Air Pollution Monitoring Network. The objectives of the study were to evaluate the errors involved with different interpolation techniques and to select a spatial analysis method that would work best for verifying transport models and testing public health criteria. The first method was an Optimal Interpolation method that minimized the discrepancy between the estimated and true concentrations, the interpolation error. The second method was an Eigenvector Interpolation in which “the correlations are derived from the local statistical characteristics of the field, avoiding unjustified assumptions about homogeneity and isotropy.” The third method is what they call “distance and density weighting function interpolation” in which the coefficients are determined by both distance and local network density. The performance of the three interpolation techniques was evaluated and differences were found to be small.

“On the Spatial Interpolation of Air Pollution Fields;” Manfred Zier, Meteorological Service of the GDR, Meteorological Observatory; Symposium on the Development of Multi-Media Monitoring of Environmental Pollution, Riga, Latvia, USSR; December 1978.

This is an older paper that explored a new method of spatial interpolation, called the “quantile interpolation” method by the author. It is meant to allow spatial analysis over inhomogeneous, anisotropic random fields. “This method permits the long-term mean emission value to be obtained for a point of which just a few measuring values are available, which is, however, surrounded by measuring points for which there are available long time series. Long-term mean values (i.e., in the present case: mean values taken over a number of years) are, owing to their close stochastic relations to other probability distribution parameters, of primary significance as to the stochastic description of air-pollution fields. The bases of the quantile interpolation method are frequency statistical correlation parameters, which render it independent of the type of probability distribution of the values.” The author compares the quantile method to a regression method using suspended dust data from the air pollution measuring network of the GDR and notes that the accuracies of the extrapolation are approximately equal. He notes that an “advantage from the quantile interpolation method is, therefore, that by inclusion of an arbitrary

number of additional long-term measuring points the accuracy of the estimates can in a simple manner be substantially improved.”

“Data Assimilation in Air Pollution Modeling;” Xue-Fen Zhang, Delft University of Technology, The Netherlands; Doctoral Thesis; November 1996.

This thesis focuses on developing various spatial interpolation methods to accurately estimate and predict methane gas distribution in Europe. The purpose of the thesis is to develop a “suitable solution to the longstanding problem of how to appropriately assimilate the temporal observations on atmospheric pollutant concentration into large-scale dynamic system models such that: 1) the current status of air pollution can be continuously monitored with feasible accuracy, 2) future potential trends of air pollution distribution can be predicted, and 3) the emissions can be identified.” An optimal interpolation scheme using kriging is employed first. Then methods based on kriging and Kalman filtering are used. Additionally, two new methods based on Kalman filtering are developed and utilized. Chapter 3 presents a good summary of simple and universal kriging and semivariograms. Chapter 4 contains a summary of Kalman filter techniques and reduced rank square root filter algorithm.

“Analytical Determination and Classification of Pollutant Concentration Fields using Air Pollution Monitoring Network Data: Methodology and Application in the Paris area During Episodes with Peak Nitrogen Dioxide Levels;” Anda Ionescu, Yves Candau, Eric Mayer, and Iolanda Colda; International Conference on Air Pollution Modelling and Simulation; Champs sur Marne, France; October 1998.

The authors analyze NO₂ data from the AIRPARIF network using the thin plate splines interpolation method. Their objective is to develop a methodology to estimate the non-homogeneous pollutant concentration fields over the Paris region using values from a monitoring network. They determined the thin plate splines method to be the best for this set of data in an earlier paper in which they compared it against simple kriging and other schemes. They state that “a thin plate spline is equivalent to kriging an intrinsic spatial random field ISRF-1 (Christakos, 1993).” They estimate the quality of the estimation when applying splines through the use of “Leave-One-Out (LOO) Cross Validation Tests.” Using n-1 observations from their data set they determine the interpolation function, calculate its value at the nth point, compare the estimated value to the measured value, and repeat the process for all n observations. They found that the interpolation quality is good when pollution at the estimated location is close to the average level, but that results are less reliable at the borders or when input information is insufficient.

"A Composite Space/Time Approach to Studying Ozone Distribution over Eastern United States", George Christakos and Vikram Vyas, School of Public Health, University of North Carolina at Chapel Hill; Atmospheric Environment, Vol. 32, No. 16; 1998.

The authors of this paper raise the question of “how to obtain adequate representations of ozone distributions with varying space/time trends.” In answering this question, the paper

describes the application of a combined spatial/temporal approach to modeling ozone distribution. They utilize a spatial-temporal random field (S/TRF) model that allows the analysis of a space/time continuum of ozone concentration. This method improves upon the prediction of ozone concentration beyond that achievable by a purely temporal or purely spatial analysis. They state that “prediction of ozone values on the basis of a purely spatial or purely temporal data set may lead to large prediction errors.”

“Spatial and temporal trends are interrelated, because dispersion is also governed by temporal processes—such as temporal dependence of pollution sources, amount of sunlight available for photolytic reaction, and wind patterns.” The authors also express the possibility of using the S/TRF model to assist in future network design. “Assuming that the S/TRF parameters can be estimated (from previous experience with similar situations, etc.), an optimal ozone monitoring network can be selected before any monitoring stations are constructed.”

Using AIRS ozone data from the eastern U.S., the authors also compare the results from the S/TRF model with those of earlier studies that used only spatial data from one time period. They find that the spatiotemporal estimates are significantly more accurate than the estimates produced by a purely spatial model using kriging especially in cases where fewer monitoring locations are available.

"BME Representation of Particulate Matter Distribution in the State of California on the Basis of Uncertain Measurements;" George Christakos, Marc Serre, and Jordan Kovitz, University of North Carolina at Chapel Hill; Journal of Geophysical Research, Vol. 106, No. D9; May 2001.

Christakos, et al., discuss their Bayesian Maximum Entropy (BME) method of spatial interpolation, which they believe yields more accurate predictions than classical methods. The BME approach “has certain advantages over other methods of mapping space/time pollutant distributions. It rigorously takes into consideration many forms of physical knowledge, which improve the accuracy and scientific content of space/time mapping and provide the means to avoid the circular problem of empirical geostatistics and spatial statistics.” On kriging, they note that “kriging techniques are based on the minimum mean squared error criterion that may fail in the case of heavy-tailed random fields with large variances. In contrast, BME permits more flexible estimation criteria that are well-defined even for heavy-tailed fields. In general, BME is a nonlinear estimator which does not impose any constraints on the estimator sought, non-Gaussian laws are automatically incorporated, and by taking into account physical models, BME possesses global estimation features.” They think this is an improved approach because “linear estimators commonly used in spatial statistics can be highly inefficient compared to nonlinear estimators associated with non-Gaussian random fields.”

Also on kriging, “unlike the classical kriging variance that is independent of the data values (and, as a consequence, has been the subject of some criticism among geostatisticians), the BME variance depends on the specific data set considered.” In this paper, the authors apply the

BME approach to daily PM₁₀ concentrations in California obtained from the California Air Resources Board.

"A Study of the Spatiotemporal Health Impacts of Ozone Exposure", George Christakos and Alexander Kolovos, School of Public Health, University of North Carolina at Chapel Hill; Journal of Exposure Analysis and Environmental Epidemiology; 1999.

This paper continues the discussion of the S/TRF model summarized in the first Christakos reference listed above. This paper extends the discussion of spatial-temporal analysis to the study of the health effects of pollutants over time, with the example pollutant in the article being ozone. "The exposure distribution and the biological variables are represented in terms of spatiotemporal random fields." The authors categorize questions about ozone health hazards into three areas: exposure conditions, relating exposure to burden, and detecting adverse health effects and population damage. The S/TRF model can be expanded to estimate the burden map and health effects, in addition to exposure maps. The authors note that kriging is a special case of the BME approach, but that BME leads to novel and more general results that could not be obtained with kriging or other traditional analyses.

"Bayesian Spatial Prediction of Random Space-Time fields With Application to Mapping PM_{2.5} Exposures," B.M. Golam Kibria, Florida International University; Li Sun, James Zidek, and Nhu Le, University of British Columbia; Journal of the American Statistical Association, Vol. 97, No. 457; 2002.

This paper presents an alternative spatial interpolation methodology to kriging. The authors note that "a Bayesian methodology for both temporal and spatial interpolation has been developed as an alternative to the well-known method of kriging. Like kriging, the Bayesian interpolation method assumes an underlying spatial process with responses that are functions of their locations. However, the Bayesian method uses both the location of the sites and the distances between them for interpolation. That is, unlike kriging, it realistically avoids the assumptions that the spatial field is either isotropic or stationary." They note that the method "produces the joint predictive distribution for several locations and different time points using all available data, thus allowing for simultaneous spatial and temporal interpolation. Furthermore, it allows incorporation of uncertainty associated with the mean and the spatial covariance of the field into the predictive distribution." These and other authors have explored the Bayesian spatial interpolation method in prior papers. This paper extends the univariate theory to the multivariate case. The method is validated using PM_{2.5} and PM₁₀ data from eight monitoring locations around Philadelphia.

“Spatial Prediction of Sulfur Dioxide in the Eastern United States;” David M. Holland and Lawrence Cox, EPA/ORD; Nancy Saltzman, National Institute of Statistical Sciences; Douglas Nychka, North Carolina State University; Geostatistics for Environmental Applications, Vol. VII, pp 65-76; 1998.

This EPA-funded paper discusses modeling options for cases where there is a nonstationary covariance structure. It does this through investigating new ways of reducing the standard error of predicted concentrations of a pollutant in non-monitored areas. “This paper evaluates the predictive performance of several spatial covariance functions, including one with nonstationary components, for different configurations of sampling sites that are constructed based on space-filling designs.” Based on analysis of prediction variances, it may be necessary to change network configurations or add sites to or delete sites from the network. For their evaluation, the authors use SO₂ data from 34 rural CASTNet monitoring sites in the eastern U.S.

The authors evaluate different covariance functions: (1) an isotropic exponential covariance function, (2) an isotropic exponential function that includes a surface of marginal variances, and (3) a marginal covariance model that includes nonparametric terms that adjust for additional nonstationary covariance. They found that “the nonstationary model provided large reduction in the mean and median prediction error relative to the exponential and marginal model.” The reason for this is the existence of “an underlying nonstationary covariance structure” in the CASTNet SO₂ data analyzed for the paper. The nonstationary covariance suggests that “the exponential covariance model is not valid for these data and all results derived from this model should be viewed with caution.”

“Designing and Integrating Composite Networks for Monitoring Multivariate Gaussian Pollution Fields,” James V. Zidek and Nhu D. Le, University of British Columbia, and Weimin Sun, Statistics Canada; Journal of the Royal Statistical Society Series C -Applied Statistics; Vol. 49, Part I, 2000.

This paper also focuses on the optimal design of pollution monitoring networks. Their model maximizes the amount of uncertainty that can be eliminated from a network with the constraint of considering cost. They use a multiattribute utility analysis to measure the two objective criteria on a common scale and seek to maximize a linear combination of them. From the abstract, “we describe a Bayesian framework for integrating the measurements of these stations to yield a spatial predictive distribution for unmonitored sites and unmeasured concentrations at existing stations. Furthermore, we show how this network can be extended by using an entropy maximization criterion.” The model is demonstrated by evaluating measurement of ozone and sulfate at a network of 31 monitoring locations in Ontario, Canada.

***“Model Validation and Spatial Interpolation by Combining Observations with Output from Numerical Models via Bayesian Modeling;”* Montserrat Fuentes, North Carolina State University, and Adrian Raftery, University of Washington; Technical Report No. 43, Department of Statistics, University of Washington, November 2001.**

In this report, the authors present a Bayesian method for combining pollution monitoring data and emissions data in an effort to yield better predictions of pollution levels. Their objective is to perform model validation and bias removal for their dispersion model, which is based on emissions data, and to construct reliable air pollution maps combining the emissions and monitoring data. The dispersion model used is the regional scale air quality models (Models-3), while the monitoring data are from the CASTNet. In the article, their models are tested using SO₂ concentrations in the eastern U.S. The air quality models are validated by obtaining the posterior predictive distribution of the measurements at the monitoring sites given the numerical models’ output. The bias in the air quality models is removed by obtaining the posterior distribution of the bias parameters given the measurements at the monitoring sites and the numerical models’ output. Reliable air pollutant maps are generated by simulating values from the posterior predictive distribution of the true values given the monitoring data and dispersion model output.

***“Spatial Prediction of Air Quality Data.”* David M. Holland, EPA/ORD; William M. Cox and Rich Scheffe, EPA/OAQPS; Alan J. Cimorelli, EPA/Region III; Douglas Nychka, NCAR; Philip K. Hopke, Clarkson University. EM, August 2003.**

The authors review the current state of the science of spatial interpolation and discuss a number of innovative techniques for moving beyond spatial interpolation techniques that focus on simple correlation functions. They present various models for accommodating nonstationarity and heterogeneous covariance structure. The authors begin by noting that “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.” These approaches include a moving window approach that separately estimates first and second order stationarity within each window of space and a kernel smoothing approach for estimating covariances. They then present a series of more sophisticated models for estimating global nonstationarity using empirical orthogonal functions, process-convolution, spatial deformation, and thin-plate splines. The paper also discusses hierarchical Bayesian approaches for temporal and spatial interpolation. They cite a major advantage of these methods to be the ability to incorporate the uncertainty in the mean and spatial covariance of the spatial field in the predictive distribution. In conclusion, Holland, et al., argue that spatial and spatial-temporal models can provide an important link between the scientific community and the regulatory monitoring community. Regulators can apply these models to gain a better understanding of complex air-quality issues and to assist with apportioning scarce resources.

Dispersion/Transport/Deposition Modeling

“Estimating the Direction of an Unknown Air Pollution Source Using a Digital Elevation Model and a Sample of Deposition;” Oleg Antonic and Tarzan Legovic, Rudjer Boskovic Institute, Zagreb, Croatia; Ecological Modelling; Vol. 124, 1999.

This paper explores estimating the direction of an air pollution source through the use of a model that estimates topographic exposure to wind. A model of this type is necessary in cases where there is not enough field data to support using interpolation methods. Their hypotheses are that “when the research area is uneven, its spatial units are differently exposed to the air pollution that is coming from some distant pollution source. Direction of the source can be estimated by maximizing correlation between the sampled pollution variables and topographic exposure to a particular direction. Suitable estimators of the topographic exposure to the given direction can be derived from the digital elevation model (DEM). This could enable the explanation of the pollution spatial variability and also spatial prediction of contamination values, even in cases when pollution source, atmospheric conditions, and/or real wind flux are not known.” They used soil pollution data from Risnjak National Park in Croatia to test their model. Their results did confirm their hypotheses. They note that an increase in the number of pollution sources would require an increase in field sampling intensity.

“A Numerical Analysis of Los Angeles Basin Pollution Transport to the Grand Canyon Under Stably Stratified, Southwest Flow Conditions;” Gregory Poulos and Roger Pielke, Colorado State University; Atmospheric Environment; Vol. 28, No. 20, November 1994.

This article discusses approaches for modeling the dispersion/transport of pollution from the Los Angeles region east to the Grand Canyon. The authors compare modeling simulations using both flat and realistic terrains “to show the importance of terrain effects on transport.” The simulations were run using the Colorado State University Regional Atmospheric Modeling System (CSU-RAMS). The output of the meteorological models were used as input into a Lagrangian particle dispersion model that simulated the release and subsequent transport of pollutants in the atmosphere.

“An Integrated Air Pollution Modeling System: Application to the Los Angeles Basin;” Rong Lu and Richard Turco, UCLA; Numerical Simulations in the Environmental and Earth Sciences: Proceedings of the Second UNAM-CRAY Supercomputing Conference; Cambridge, England, Cambridge University Press, 1997.

The authors of this article study an air pollution modeling system that integrates a meteorological model, a gas dispersion model, a photochemistry model, and a radiative transfer model. They use the Surface Meteorology and Ozone Generation (SMOG) model to simulate meteorological conditions, dispersion of passive tracers, and pollutant distributions over Southern California. They evaluate performance by comparing simulation results with observational data. They find that “the overall agreements between predictions and observations show that the SMOG modeling system is able to reproduce the main features of mesoscale meteorology,

pollutant dispersion, and transformations in the atmosphere. The modeling system is capable of handling complicated situations such as the photochemical smog over complex topography in the Los Angeles Basin. The integrated modeling system is shown to be a powerful tool for studying coupled dynamical, chemical and microphysical processes on urban and regional scales.”

“The Danish Eulerian Hemispheric Model - A Three-Dimensional Air Pollution Model Used for the Arctic,” Jesper H. Christensen, National Environmental Research Institute, Department of Atmospheric Environment, Roskilde, Denmark; Atmospheric Environment, Vol. 31, No. 24, 1997.

This paper describes the development of a Eulerian long-range transport model (as opposed to a trajectory model) to describe the transport of air pollution to the Arctic. Trajectory models present problems because trajectories longer than four to five days (the case for connecting emission areas with arctic areas) in the lower troposphere are highly unreliable. The Danish Eulerian Hemispheric Model (DEHM) developed considers factors such as precipitation, vertical and horizontal diffusion, and oxidation. The model reproduced very well the measured concentrations of sulphur species in Europe, Canada, and the Arctic.

“Lagrangian Particle Modeling of Air Pollution Transport in Southwestern United States;” Marek Uliasz, Warsaw University of Technology, Poland; Roger Stocker and Roger Pielke, Cooperative Institute of Research in the Atmosphere, Colorado State University; American Meteorological Society Annual Meeting, 1994.

This article focuses on two methods of dispersion modeling using the Regional Atmospheric Modeling System (RAMS). A source-oriented approach allows calculation of air pollution characteristics for receptors located within the modeling domain based on a set of given emission sources. The receptor-oriented modeling approach depends on meteorology, deposition, and atmospheric transformations of pollutants but is independent of emission sources. The authors note that this may be a more effective approach when air pollution at the receptor is of primary interest. The specific goal of the study was to assess the impact of the Mohave Power Project and other potential sources of air pollution on the Southwestern U.S. including the Grand Canyon.

“GATOR-GCMM: A Global- Through Urban-scale Air Pollution and Weather Forecast Model;” Mark Jacobson, Stanford University; Journal of Geophysical Research, Vol. 106, No. D6, March 2001

The GATOR-GCMM (gas, aerosol, transport, radiation, general circulation, and mesoscale meteorological) model is a one-way nested meteorological model “designed to treat gases, size-/composition-resolved aerosols, radiation, and meteorology from the global to the urban (<5 km) scale.” Further, “the model accounts for radiative feedback from photochemically active gases, size-resolved aerosols, and size-resolved liquid water and ice particles to meteorology on all scales. The model treats subgrid soil and surface classes, including rooftops and road surfaces

for ground-temperature calculations.” This paper presents the model but does not evaluate it using actual data.

“Demonstrating Attainment of the Air Quality Standards: Integration of Observations and Model Predictions into the Probabilistic Framework;” Christian Hogrefe, SUNY-Albany, and S. Trivikrama Rao, New York State Department of Environmental Conservation; Journal of Air and Waste Management Association; Vol. 51, July 2001.

This paper focuses on comparing the results of different meteorological/photochemical modeling systems “to assess whether the model-to-model differences in the predicted relative changes (in ozone) are indeed relatively small, as envisioned in EPA’s draft modeling guidance for the 8-hr. ozone NAAQS.” They present spatial maps of percent ozone reduction given certain emissions reductions predicted by two sets of modeling systems — RAMS/UAM-V and MM5/UAM-V.

“Interpreting the Information in Ozone Observations and Model Predictions Relevant to Regulatory Policies in the Eastern United States,” by Christian Hogrefe, S. Trivikrama Rao, and Igor G. Zurbenko, SUNY-Albany, and P. Steven Porter, University of Idaho; Bulletin of the American Meteorological Society, Vol. 81, No. 9, September 2000.

This paper is focused on comparing the results of different time scales of ozone monitoring — intraday, diurnal, synoptic, and long-term. The authors simulated three months of ozone measurements and compared observations and modeling results. They found that “timescales greater than diurnal need to be considered in order to understand the nature of the ozone problem, to evaluate model performance on different timescales, and to quantify the efficacy of emission control strategies.” For the study, ozone data were extracted from the AIRS database, meteorological data was prepared using the RAMS system, and modeled ozone concentrations were obtained from the Urban Airshed Model, Variable Grid Version (UAM-V).

“IRS-1C LISS III Land Cover Maps at Different Spatial Resolutions Used in Real-time Accidental Air Pollution Deposition Modelling;” C.B. Hasager and S. Thykier-Nielson, Riso National Laboratory, Roskilde, Denmark; Remote Sensing of Environment, Vol. 76, 2001.

This paper explores the use of satellite imagery to assign roughness values to land cover types in order to improve deposition modeling. “Land cover classes with variations in aerodynamic roughness lengths are mapped at a high resolution. The high resolution roughness map is degraded into lower spatial scales and used in a real-time air pollution model with different grid resolution and a model step time of 10 min. The objective is to assert which spatial scale is necessary as input for real-time air pollution.” The study is focused on deposition in Lithuania from nearby nuclear reactor accidents. The authors found that “high-resolution mapping of land cover does increase the spatial details in deposition mapping.” They note that the land cover input should be at a comparable resolution to the deposition model.

Spatial Scale

“Combining Incompatible Spatial Data;” Carol A. Gotway, National Center of Environmental Health, CDC, and Linda J. Young, University of Nebraska; Journal of the American Statistical Association, Vol. 97, No. 458, June 2002.

This article reviews various statistical methods for combining spatial data from different resolutions (point, line, area, surface). The issue of utilizing data sources of different scales is referred to as the “change of support problem” (COSP) in geostatistics. Support refers to the size or content of each data value. The support of a variable changes by averaging or aggregating the available data. The authors note that block kriging can be a solution for the point-to-area COSP as it can be used “to predict the average value of a process at a larger scale, accounting not only for the size, but also for the shape and orientation of the blocks.” They also note that cokriging can be used for point-to-point and point-to-area COSPs, although there is less agreement on how to carry it out. In addition to geostatistical solutions to the COSP problem, nonlinear methods are discussed as well. Methods described include a multi-Gaussian approach, indicator kriging, and covariance-matching constrained kriging. In regards to multiscale modeling, or generating predictions across scales, the paper reviews spatial tree models and Bayesian hierarchical models that can unite observational data and weather forecasting models.

Another concept explored in the paper is the polygon overlay problem, i.e., the problem of superimposing source and target units. Techniques for handling these map or polygon overlay problems include probabilistic potential mapping, pixel aggregation and areal weighting, spatial smoothing methods, and Bayesian areal regression models.

APPENDIX A:
KRIGING MODEL FORMULAS

Appendix A: Kriging Model Formulas

Ordinary Kriging

Ordinary kriging assumes that the variogram is represented by:

$$2\gamma(h) = \text{var} [Z(s+h) - Z(s)], h \in \mathbb{R}^2.$$

Recall that s is a location and h is a shift from that location. Also recall that $Z(s)$ is a measurement taken at s . Notice that the variogram is only dependent on h . Given the variogram, the kriging predictors are:

$$p(Z; s_0) = \sum_{i=1}^n \lambda_i Z(s_i).$$

The optimal λ can be then obtained from:

$$(\lambda_1, \dots, \lambda_n) = \left(\gamma + \mathbf{1} \frac{(\mathbf{1} - \mathbf{1}\Gamma^{-1}\gamma)}{\mathbf{1}\Gamma^{-1}\mathbf{1}} \right)' \Gamma^{-1},$$

where $\mathbf{C} = [C(s_0-s_1), \dots, C(s_0-s_n)]'$ and Γ is the $n \times n$ matrix whose (i, j) th element is $C(s_i-s_j)$.

The kriging (or prediction) variance can then be written as:

$$\sigma_k^2(s_0) = \gamma' \Gamma^{-1} \gamma - (\mathbf{1}\Gamma^{-1}\gamma - 1)^2 / (\mathbf{1}\Gamma^{-1}\mathbf{1}).$$

Universal Kriging

Write the universal kriging model as follows:

$$\mathbf{Z} = \mathbf{X}\beta + \epsilon,$$

and assume

$$2\gamma(h) = \text{var} [Z(s+h) - Z(s)], \text{ where}$$

$Z(s)$, s , and h are defined as above. Then the universal kriging predictors are of the form:

$$p(Z; s_0) = \sum_{i=1}^n \lambda_i Z(s_i).$$

Then the optimal λ can be obtained from:

$$(\lambda_1, \dots, \lambda_n) = \left[\gamma + X(X' \Gamma^{-1} X)^{-1} (x - X' \Gamma^{-1} \lambda) \right]' \Gamma^{-1},$$

where $\mathbf{C} = [C(s_0 - s_1), \dots, C(s_0 - s_n)]'$ and Γ is the $n \times n$ matrix whose (i, j) th element is $C(s_i - s_j)$.

The kriging (or prediction) variance can then be written as:

$$\sigma_k^2(s_0) = \gamma' \Gamma^{-1} \gamma - (x - X' \Gamma^{-1} \gamma)' (X \Gamma^{-1} X)^{-1} (x - X' \Gamma^{-1} \gamma).$$

The optimal estimation of the mean parameters β can be accomplished easily as follows:

$$\hat{\beta} = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} Z,$$

where \mathbf{E} is the variance of Z , recalling that Z is the set of all observations.

TECHNICAL REPORT DATA

(Please read Instructions on reverse before completing)

1. REPORT NO. EPA-454/R-04-004		2.		3. RECIPIENT'S ACCESSION NO.	
4. TITLE AND SUBTITLE Developing Spatially Interpolated Surfaces and Estimating Uncertainty				5. REPORT DATE November 2004	
				6. PERFORMING ORGANIZATION CODE	
7. AUTHOR(S) Shelly Eberly, Jenise Swall, David Holland, Bill Cox, Ellen Baldrige				8. PERFORMING ORGANIZATION REPORT NO.	
9. PERFORMING ORGANIZATION NAME AND ADDRESS U.S. Environmental Protection Agency Office of Air Quality Planning and Standards Emissions, Monitoring and Analysis Division Research Triangle Park, NC 27711				10. PROGRAM ELEMENT NO.	
				11. CONTRACT/GRANT NO.	
12. SPONSORING AGENCY NAME AND ADDRESS Director Office of Air Quality Planning and Standards Office of Air and Radiation U.S. Environmental Protection Agency Research Triangle Park, NC 27711				13. TYPE OF REPORT AND PERIOD COVERED	
				14. SPONSORING AGENCY CODE EPA/200/04	
15. SUPPLEMENTARY NOTES					
16. ABSTRACT This document provides an overview of spatial interpolation methods and their use with air quality data. Characteristics of the data that are important to consider are spatial representativeness, temporal sampling frequency, measurement accuracy, and existence of spatial relationships or behaviors at varying scales. This document discusses whether there is a need to force the interpolated surface to pass through the measured values, whether the data contain a global trend across the entire area of interest, or whether short-range variation is significant. General interpolation methods and data considerations are discussed. Ordinary kriging, a geostatistical spatial interpolation method, is discussed, along with example of developing an interpolated surface of PM2.5 concentrations in the eastern U.S. and methods for evaluating model performance through diagnostics. Common extensions to ordinary kriging such as including spatial trends, temporal dynamics, non-stationary covariance structures, use of covariates and multivariate modeling are presented. Software for performing spatial interpolation analysis are summarized. Two example applications are presented, S-Plus to krig estimates for annual average PM2.5 concentrations and SAS to krig 8-hour ozone concentrations. Limitations to spatial interpolation and alternative methods are presented. A summary of references on various aspects of spatial interpolation is provided. This document should be a helpful reference and synopsis of methods and issues to be addressed when interpolating ambient air quality concentrations.					
17. KEY WORDS AND DOCUMENT ANALYSIS					
a. DESCRIPTORS		b. IDENTIFIERS/OPEN ENDED TERMS		c. COSATI Field/Group	
spatial interpolation, multivariate modeling, kriging, covariates, cross validation, variogram		ozone, PM2.5			
18. DISTRIBUTION STATEMENT Release Unlimited		19. SECURITY CLASS (<i>Report</i>) Unclassified		21. NO. OF PAGES 159	
		20. SECURITY CLASS (<i>Page</i>) Unclassified		22. PRICE	

United States
Environmental Protection
Agency

Office of Air Quality Planning and Standards
Air Quality Strategies and Standards Division
Research Triangle Park, NC

Publication No. EPA-454/R-04-004
October 2004
