



ArcGIS™ Geostatistical Analyst: Statistical Tools for Data Exploration, Modeling, and Advanced Surface Generation

An ESRI® White Paper • August 2001

Copyright © 2001 ESRI
All rights reserved.
Printed in the United States of America.

The information contained in this document is the exclusive property of ESRI. This work is protected under United States copyright law and other international copyright treaties and conventions. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage or retrieval system, except as expressly permitted in writing by ESRI. All requests should be sent to Attention: Contracts Manager, ESRI, 380 New York Street, Redlands, CA 92373-8100, USA.

The information contained in this document is subject to change without notice.

U.S. GOVERNMENT RESTRICTED/LIMITED RIGHTS

Any software, documentation, and/or data delivered hereunder is subject to the terms of the License Agreement. In no event shall the U.S. Government acquire greater than RESTRICTED/LIMITED RIGHTS. At a minimum, use, duplication, or disclosure by the U.S. Government is subject to restrictions as set forth in FAR §52.227-14 Alternates I, II, and III (JUN 1987); FAR §52.227-19 (JUN 1987) and/or FAR §12.211/12.212 (Commercial Technical Data/Computer Software); and DFARS §252.227-7015 (NOV 1995) (Technical Data) and/or DFARS §227.7202 (Computer Software), as applicable. Contractor/Manufacturer is ESRI, 380 New York Street, Redlands, CA 92373-8100, USA.

ESRI, ArcView, and the ESRI globe logo are trademarks of ESRI, registered in the United States and certain other countries; registration is pending in the European Community. ArcGIS, ArcInfo, ArcEditor, and ArcMap are trademarks and @esri.com and www.esri.com are service marks of ESRI.

The names of other companies and products herein are trademarks or registered trademarks of their respective trademark owners.

ArcGIS Geostatistical Analyst: Statistical Tools for Data Exploration, Modeling, and Advanced Surface Generation

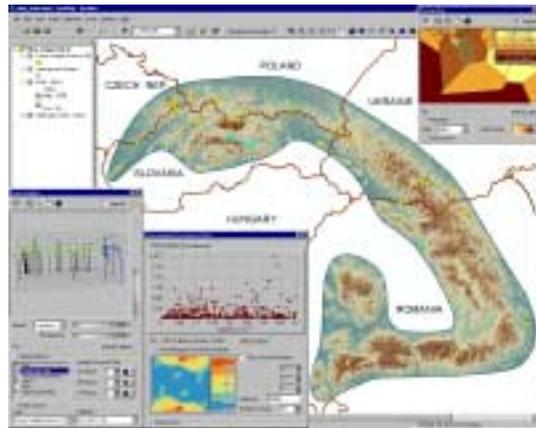
An ESRI White Paper

Contents	Page
1. Overview	1
What Does Geostatistical Analyst Do?	2
Who Uses Geostatistical Analyst?	2
Why Use Geostatistical Analyst?	3
Saves Lives.....	3
Increases Efficiency	4
Multiple Tools for Data Representation.....	4
Exploratory Environment for Data Investigations	4
Wizard-Driven Interface for Fast Efficient Interpolation Processes	5
Perform Diagnostics.....	5
2. Using Geostatistical Analyst.....	5
Working with the Data	5
Represent the Data	6
Explore the Data: Exploratory Spatial Data Analysis.....	6
Fit a Model (create a surface).....	7
Deterministic Methods	7
Geostatistical Methods	8

Contents	Page
Kriging	8
Cokriging.....	9
Output Surfaces	9
Analytical Tools for Creating Surfaces	10
Investigation of Spatial Data Structure: Variography	10
Detrending Tools.....	12
Search Neighborhood Specification.....	13
Data Transformations.....	13
Declustering Method.....	14
Bivariate Normal Distribution.....	14
Perform Diagnostics.....	15
Cross-Validation and Validation.....	15
Model Comparison.....	16
Display Options.....	17
Display Formats	17
3. Conclusion.....	18
Data Used in Screen Shots	18

ArcGIS Geostatistical Analyst: Statistical Tools for Data Exploration, Modeling, and Advanced Surface Generation

- 1. Overview** ESRI® ArcGIS™ Geostatistical Analyst is a new extension to ArcGIS 8.1 (ArcInfo™, ArcEditor™, and ArcView®) that provides a dynamic environment to help solve such spatial problems as improving estimation temperature values, assessing environmental risks, or predicting the existence of any geophysical element. Geostatistical Analyst provides a wide variety of tools for spatial data exploration, identification of data anomalies, evaluation of error in prediction surface models, statistical estimation, and optimal surface creation. Geostatistical Analyst gives anyone with spatial data the freedom to investigate, visualize, and create surfaces using sophisticated statistical methods.



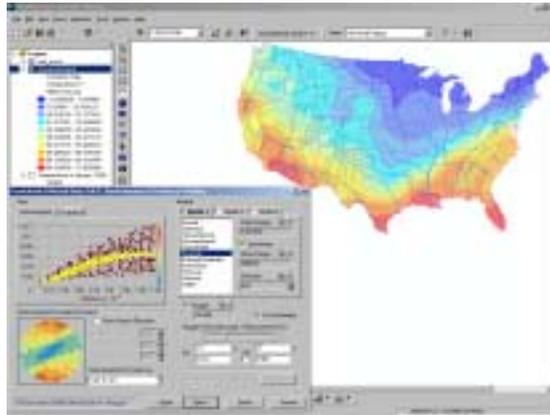
Using a variety of exploratory spatial data analysis tools, the user can investigate the properties of ozone measurements taken at monitoring stations in the Carpathian Mountains.

Investigate ozone levels.

From analyzing severity of ozone air pollution to identifying soil contamination, Geostatistical Analyst can be used to create predictive surfaces for many different types of data. By exploring the data, determining the best interpolation method, and by utilizing diagnostic techniques, one can successfully and easily create optimal interpolated surfaces that statistically represent the data. Geostatistical Analyst can also provide additional statistical information that helps users assess the quality of their analysis using prediction standard error, probability, quantile, and standard error of indicator surfaces.

"...The inclusion of more spatial interpolative methods in ArcGIS is a huge feature. Now, to do powerful interpolations of data trends, ArcGIS users no longer have to deal with the preprocessing of data for export to another application..."

Daniel M. Petrecca
Senior Staff Hydrogeologist/GIS Specialist
Langan Engineering and Environmental Services, Inc.



The semivariogram/covariance dialog was used to fit a geostatistical prediction model to winter temperature data for the USA. This model was then used to create the temperature distribution map.

Analyze temperature data.

What Does Geostatistical Analyst Do?

Geostatistical Analyst determines the probability of certain variables occurring over an area where identifying every possible location would be impossible. For example, in California, air quality monitoring stations are set up around the State. These measure the amount of particulates in the air and how harmful they are to surrounding areas. Geostatistical Analyst can determine the approximate amount of particulates in the area of interest and can also determine where these particulates may be moving by creating an optimal interpolated surface. Geostatistical Analyst provides a multitude of powerful interpolation methods with advanced analytical tools for generating optimal interpolated surfaces from discrete spatial data measurements.

Who Uses Geostatistical Analyst?

Geostatistical Analyst is a flexible software package that allows any user with spatially continuous data to explore and analyze their data using statistical tools and interpolate optimal surfaces. Some of the various fields that utilize Geostatistical Analyst include environmental, agriculture, exploration, geology, meteorology, hydrology, archaeology, forestry, health care, mining, and real estate.

"I work in the environmental side where our services include hydrogeologic, geologic, and geochemical characterizations and remedial investigations of contaminated or potentially contaminated properties. Our clients include utilities, chemical companies, petroleum companies, and basically anyone that has been fortunate enough to have toxic substances on their site. So, we use the Geostatistical Analyst to interpolate many things:

- Site stratigraphy and geologic surfaces from borehole logs
- Groundwater surfaces and flow from water elevation data

- Extent of contamination from chemical analysis of soil and water samples
- Topography from spot elevations
- Trends in other field data measurements (e.g., ground penetrating radar and electromagnetic metal detector data)"

Daniel M. Petrecca
Senior Staff Hydrogeologist/GIS Specialist
Langan Engineering and Environmental Services, Inc.

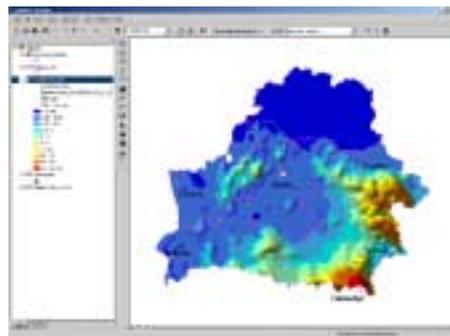
Why Use Geostatistical Analyst?

Geostatistical Analyst creates statistical interpolated continuous surfaces from measured samples. These surfaces represent a statistical estimation or prediction of where a certain phenomenon may occur. From determining the impact of biohazard releases to identifying where a plant species may be thriving, Geostatistical Analyst provides unique interpolation tools for the user's predictive needs. Not only are interpolated surfaces created, but also a wide range of analytical and exploratory tools are incorporated to extract useful information from the data. Geostatistical Analyst provides a cost-effective, logical solution for analyzing a variety of data sets that would otherwise cost an enormous amount of time and money to accomplish.

Saves Lives

Geostatistical Analyst can assess potential environmental hazards by interpolating the possible flow and direction of radiation, air pollution, biohazard releases, aquifer contamination, and any potentially harmful waste that may be introduced into areas of human habitation.

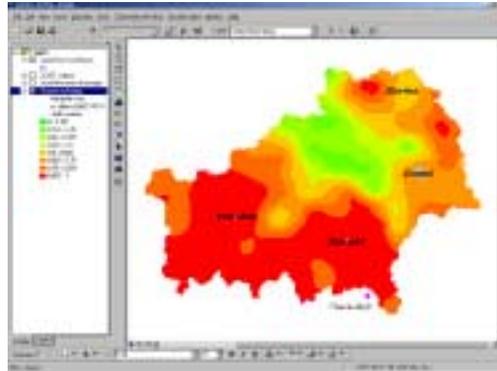
For example, Geostatistical Analyst was used to interpolate the intensity and direction of the hazardous radiation that resulted from the Chernobyl accident. Cases of thyroid cancer, soil radionuclides, and food contamination were a few of the variables used to estimate the severity of the Chernobyl accident on the surrounding areas.



This map illustrates predicted radiocesium soil contamination levels in the country of Belarus after the Chernobyl nuclear accident in 1994.

Assess environmental impact predictions.

Using Geostatistical Analyst, areas that have received environmental contamination can be identified to ultimately prevent exposure to health risks.



In this example, locations shown with darker colors indicate a probability greater than 62 percent that radioceasium contamination exceeds the upper permissible threshold for human consumption in forest berries (a popular food source) in the southern part of Belarus in 1993.

Calculate probabilities of exceeding environmental thresholds.

Increases Efficiency

Geostatistical Analyst provides users with the capability to predict optimal conditions for efficient and more reliable production. For example, a corn farmer can utilize Geostatistical Analyst to determine why crop yields in a certain area of his farm are below potential. Crop rotations, extra fertilizer, enhanced irrigation techniques, and so forth, can then be applied to the declining area. Geostatistical Analyst provides cost-effective interpolation methods for predicting and calculating areas of potential growth and/or areas of decline.



By interpolating the amount of phosphorous in the soil throughout this farm in Illinois, the user can more effectively allocate fertilization for future crops.

Assess agricultural productivity.

Multiple Tools for Data Representation

Geostatistical Analyst is an ArcMap™ extension that enables the user to take full advantage of the multitude of visualization tools in ArcMap. ArcMap enables users to create high-quality cartographic maps and explore and analyze data in order to make more effective decisions. Visualization is particularly important when using Geostatistical Analyst because inferences about the data can be acquired for building better models and creating more accurate surfaces.

Exploratory Environment for Data Investigations

The more the user understands about the phenomenon being investigated the better the surface produced from the sample points will be. Geostatistical Analyst provides a full suite of Exploratory Spatial Data Analysis (ESDA) tools to explore the data. Each tool provides a view of the data in a separate window, and each tool is linked to each other

and with the map. The ESDA tools allow the user to explore the distribution of the data, look for global trends in the data, identify local and global outliers, and understand the spatial structure in the data.

Wizard-Driven Interface for Fast Efficient Interpolation Processes

Geostatistical Analyst provides a wizard-driven environment that guides the user through the interpolation process. By selecting the input data and the interpolation method and following the wizard instructions, a representative surface is created for the data. It is recommended that the user first use the ESDA tools provided, which assist in assessing which interpolation model is best for the data.

"With Geostatistical Analyst, you get to adjust the interpolation methods and parameters and see a preview of the surface in real time as the changes are made in the wizard. So, once the final methods and parameters are determined, the surface is created once as a layer."



Daniel M. Petrecca
Senior Staff Hydrogeologist/GIS Specialist
Langan Engineering and Environmental Services, Inc.

Perform Diagnostics

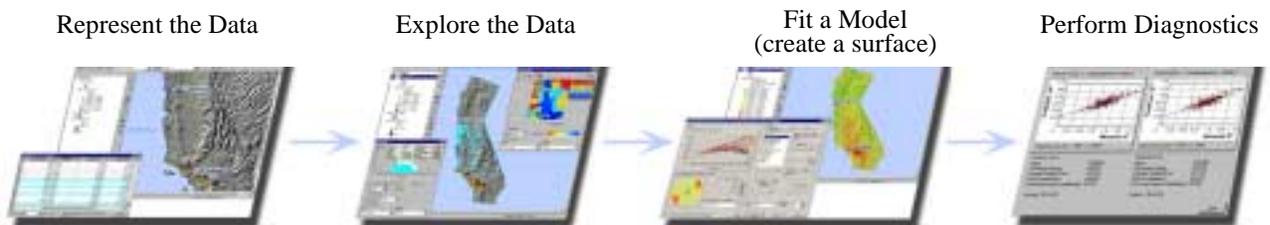
After creating a prediction surface, it is important to identify how statistically sound the surface is. Geostatistical Analyst provides cross-validation and validation tools that allow the user to evaluate the model and predictions that were used to create the surface. The tools quantify the statistical significance of the model and the model can be changed by refining the parameters.

Geostatistical Analyst also provides the user with comparative tools for choosing the best interpolated surface for the data. These tools are provided so that the user can quantify the predictions based on one model relative to another. By visually analyzing the prediction errors of the different models, the optimal model can be utilized.

2. Using Geostatistical Analyst

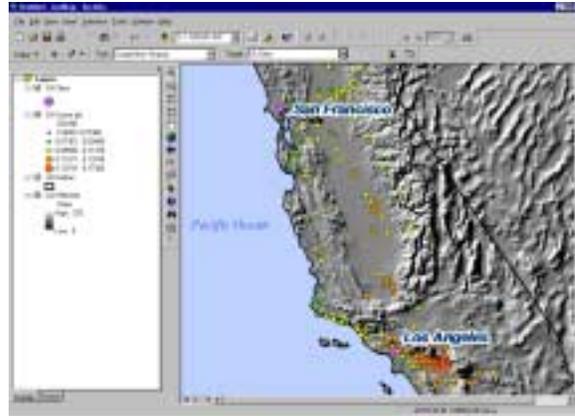
Working with the Data

There are four basic steps to creating an interpolated surface.



Represent the Data

There are many different tools for visualizing the data, which provide a great deal of information before creating a surface. By visualizing the data, useful information such as inferences from oceans, elevation, roads, and polygon edges can be acquired. Representing the data is a vital first step in assessing the validity of the data and identifying external factors that may ultimately play a role in the distribution of data.



In this display we can notice that topographic and coastal influences could play a role in the severity of air pollution in California.

Explore the Data: Exploratory Spatial Data Analysis

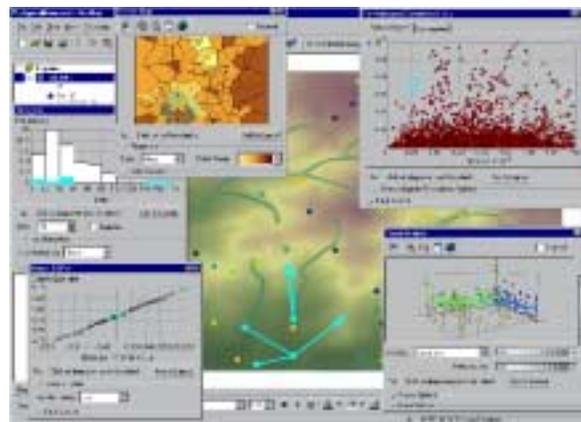
The ESDA tools are designed to explore spatial data. Visualizing the distribution of the data, looking for data trends, looking for global and local outliers, examining spatial autocorrelation, and understanding the covariation among multiple data sets are all useful tasks to perform on the data. ESDA is a powerful set of exploratory tools for determining which interpolation method is appropriate for the data. For example, an environmentalist may notice a global trend in the data caused by a drastic elevation difference in the trend analysis dialog box. This user may lean toward utilizing the universal kriging method to account for this variation.

The views in ESDA are selectively interactive within ArcMap. Data selected or highlighted with these tools will also be selected or highlighted in the ArcMap display window. This provides the user with a unique capability to visually interact with the data to better understand the relationships represented in the various displays.

Examine Local Data
Variability

Examine Distribution
Statistics

Compare Data
Distributions



Analyze Spatial
Data Distributions

Analyze Data
Trends

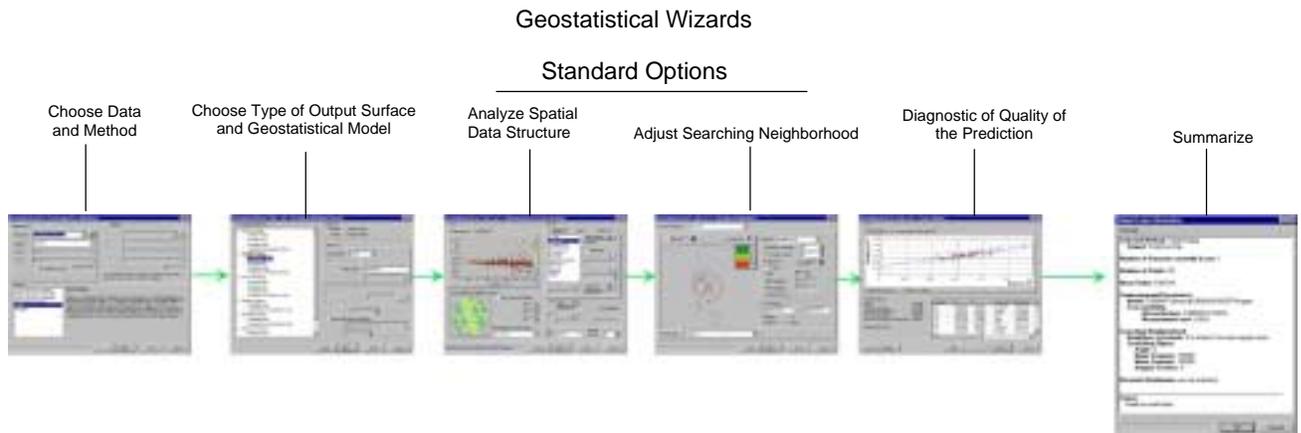
All of the ESDA graphs are dynamically linked with each other and the map.

Fit a Model (create a surface)

Once the user has thoroughly explored the data for anomalies such as global and local outliers and data trends, Geostatistical Analyst provides a wizard that makes the interpolation process easy to complete. A wide variety of interpolation methods for creating surfaces are available.

There are two main groups of interpolation techniques, deterministic and geostatistical. The deterministic interpolation technique is used for creating surfaces from measured points based either on the extent of similarity (e.g., inverse distance weighted [IDW]) or the degree of smoothing (e.g., radial basis functions and global and local polynomials). The geostatistical interpolation technique is based on statistics and is used for more advanced prediction surface modeling that also includes errors or uncertainty of predictions. For more information on the various interpolation methods provided, please see Chapters 5 and 6 in *Using ArcGIS Geostatistical Analyst*.

Depending on the method, the following output surfaces can be produced: prediction, prediction standard error (uncertainty), quantile, probability, or standard error of indicators. Geostatistical Analyst gives the user full control over the parameters of the associated models and reliable defaults are provided.



The geostatistical wizard guides the user through the interpolation process. Standard options are shown and a wide variety of analytical tools, such as detrending, declustering, and transformations, are also included.

Deterministic Methods

The deterministic interpolation technique is used for creating surfaces from measured points based either on the extent of similarity (e.g., IDW) or the degree of smoothing (e.g., radial basis functions and global and local polynomials).

- Inverse Distance Weighted
- Global Polynomial
- Local Polynomial
- Radial Basis Functions

Deterministic interpolation techniques can be divided into two groups: global and local. Global techniques calculate predictions using the entire data set. Local techniques calculate predictions from the measured points within specified neighborhoods, which are

smaller spatial areas within the larger study area. Geostatistical Analyst provides the global polynomial as a global interpolator and the IDW, local polynomials, and radial basis functions as local interpolators.

An interpolation can either force the resulting surface to pass through the data values or not. An interpolation technique that predicts a value identical to the measured value at a sampled location is known as an exact interpolator. An inexact interpolator predicts a value at a sampled location that is different from the measured value. The latter can be used to avoid sharp peaks or troughs in the output surface. IDW and radial basis functions are exact interpolators, while global and local polynomial are inexact. A typical usage of a deterministic method might be a retail analyst that needs to determine the purchasing power of distant retail locations based on the proximity from the consumer's home. By applying weights with the IDW method, this analyst may be able to effectively identify the optimal locations for retail stores.

Geostatistical Methods

The geostatistical interpolation technique is based on statistics and is used for more advanced prediction surface modeling that also includes errors or uncertainty of predictions.

- Kriging and Cokriging
 - Algorithm
 - ◆ –Ordinary
 - ◆ –Simple
 - ◆ –Universal
 - ◆ –Indicator
 - ◆ –Probability
 - ◆ –Disjunctive
- Output Surfaces
 - Prediction and prediction standard error
 - Quantile
 - Probability and standard error of indicators

Kriging

Geostatistical methods create surfaces incorporating the statistical properties of the measured data. Because geostatistics is based on statistics, these methods produce not only prediction surfaces but also error or uncertainty surfaces, giving the user an indication of how good the predictions are.

Kriging is divided into two distinct tasks: quantifying the spatial structure of the data and producing a prediction. Quantifying the spatial data structure, known as variography, is fitting a spatial-dependence model to the data. To make a prediction for an unknown value for a specific location, kriging will use the fitted model from variography, the spatial data configuration, and the values of the measured sample points around the prediction location. Geostatistical Analyst provides many tools to help determine which parameters to use, and defaults are also provided so that a surface can be created quickly. A typical usage of the kriging method might be an environmentalist who is sampling aquifers and identifies that a spatial correlation between sample points is apparent. This

analyst may then use the probability kriging method to determine the likelihood of exceeding the critical arsenic level established by the Environmental Protection Agency.

Many methods are associated with geostatistics but they are all in the kriging family. Ordinary, simple, universal, probability, indicator, and disjunctive kriging along with the multivariate versions in cokriging are available in Geostatistical Analyst.

Kriging is a moderately quick interpolater that can be exact or smoothed depending on the measurement error model. It is very flexible and allows the user to investigate graphs of spatial autocorrelation. Kriging uses statistical models that allow a variety of map outputs including predictions, prediction standard errors, standard error of indicators, and probability. The flexibility of kriging can require a lot of decision making. Kriging assumes the data comes from a stationary stochastic process. A stochastic process is a collection of random variables that are ordered in space and/or time such as elevation measurements. Some methods, such as ordinary, simple, and universal kriging, assume normally distributed data.

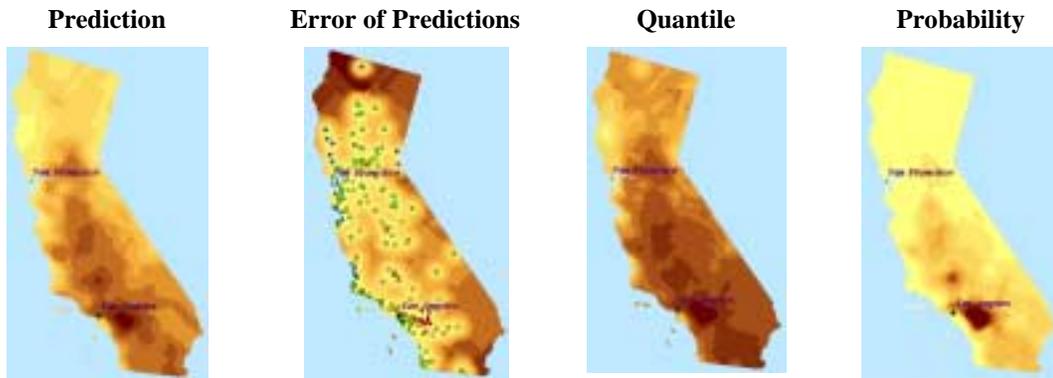
Cokriging

Cokriging is the multivariate equivalent to kriging. By using multiple data sets it is a very flexible interpolation method, allowing the user to investigate graphs of cross-correlation and autocorrelation. The flexibility of cokriging requires the most decision making. Cokriging can use either semivariograms or covariances. It can use transformations and remove trends, and it can allow for measurement error in the same situations as for the various kriging methods.

Output Surfaces

There are four interpolation surfaces available with Geostatistical Analyst. These are prediction, quantile, probability of exceeding thresholds, and errors of predictions. These allow the user to analyze the data in a variety of ways.

- **Prediction map:** Produced from the interpolated values to display random variables at locations where data has not been collected.
- **Error of predictions map:** Produced from the standard errors of interpolated values or the standard error of interpolated indicator values to display the uncertainty of the predictions.
- **Quantile map:** Produced when the user specifies a probability and wants a map of the values where predictions exceed (or do not exceed) the values at the specified probability.
- **Probability map:** Produced when the user specifies a threshold and wants a map of probabilities that the values exceed (or do not exceed) the specified threshold.



Various Surfaces Produced Using Ordinary Kriging

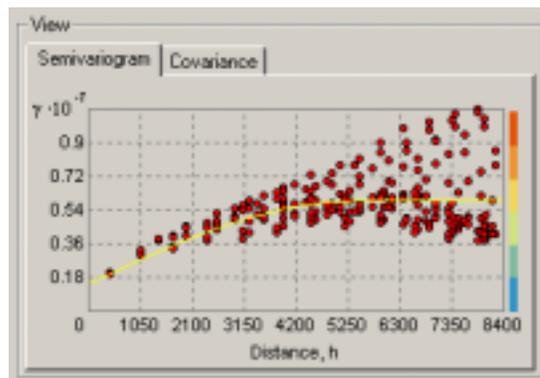
Analytical Tools for Creating Surfaces

Geostatistical Analyst provides a wizard containing analytical tools to assist the analyst in determining the values for the parameters used in creating each surface type. Some of these dialog boxes and tools are applicable to almost all interpolation methods such as specifying the search neighborhood, cross-validation, and validation. Others are specific to the geostatistical methods (kriging and cokriging) such as modeling semivariograms, transformations, detrending, declustering, and checking for bivariate normal distributions.

Investigation of Spatial Data Structure: Variography

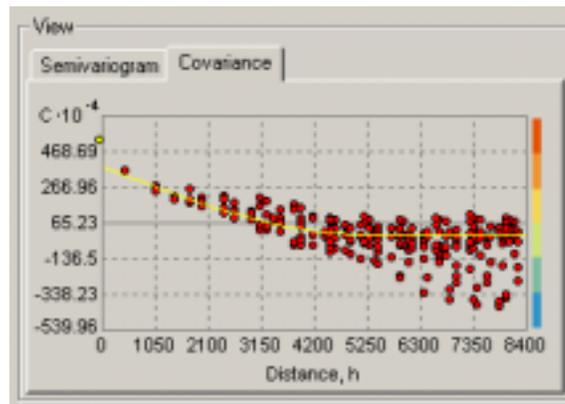
The semivariogram and covariance functions measure the strength of statistical correlation as a function of distance. Geostatistical Analyst provides the user with a semivariogram/covariance preview. This makes adjusting the parameters of the model, including anisotropy (see page 13) and modeling measurement errors, easier and more efficient.

- The semivariogram displays the statistical correlation of nearby data points. As the distance increases, the likelihood of these data points being related becomes smaller.



The variance of the difference increases with distance, so the semivariogram can be thought of as a dissimilarity function.

- When two locations are close to each other, then we expect them to be similar and so their covariance will be large. As the locations get farther apart, they become less similar and so their covariance becomes zero.



The autocorrelation decreases with distance, so the covariance can be thought of as a similarity function.

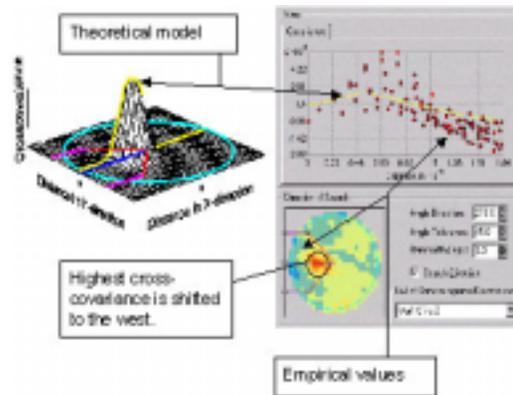
- Error Modeling.

Regardless of the interpolation method there will always be some error in the analysis. This error can be influenced by many factors such as inaccurate base data, variation in physical properties of data samples, human error, and temporal changes between measurements. All of these factors play a role in the severity of the measurement errors. Error modeling is used to minimize the effect of these measurement errors. When measurement error is specified, kriging is an inexact interpolation technique. Therefore, predictions in the measurement locations are different from the actual measurements. Geostatistical Analyst allows the analyst to adjust the measurement error to optimize the error model. For coincidental data (multiple measurements at one location) Geostatistical Analyst can estimate measured variation for the user. Three of the kriging methods—ordinary, simple, and universal—allow the use of the measurement error model.

- Cross-Covariance.

Cross-covariance is the statistical tendency of variables of different types to vary in ways that are related to each other. Positive cross-covariance occurs when both variables tend to be above their respective means together, and negative cross-covariance occurs if one variable tends to be above its mean when the other variable is below its mean.

When the user has multiple data sets and wants to use cokriging, then the user needs to develop models for cross-covariance.



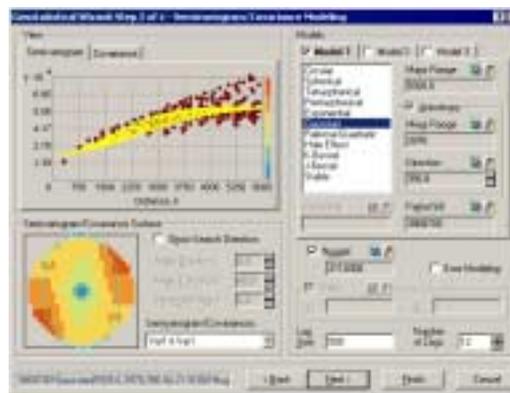
Cross-covariance modeling is used to define the local characteristics of spatial correlation between two data sets and used to look for spatial shifts in cross-correlation between two data sets. In this example, there is a strong spatial shift in the values of two data sets toward the west.

Identify spatial correlation between two data sets.

■ Anisotropy: Directional semivariogram and covariance functions.

Anisotropy is similar to global directional trends except that the cause of the directional influence is not usually known and it is modeled as a random error. Anisotropy is a characteristic of a random process that shows higher autocorrelation in one direction than another. Even without knowing the cause, anisotropic influences can be quantified and accounted for.

An isotropic model reaches the level at which the semivariogram levels off (sill) at the same distance all directions. However, with an anisotropic model, some directions reach the sill more rapidly than others.



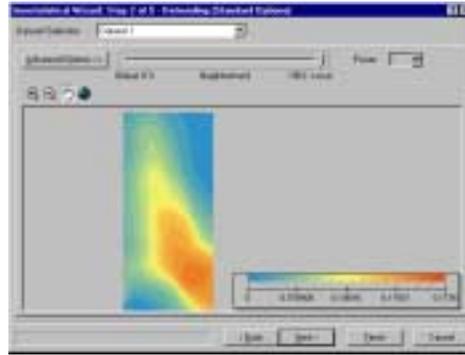
The yellow lines show the semivariogram models for many different directions. The models are a theoretical "best fit" semivariogram model to the empirical semivariogram. Geostatistical Analyst automatically calculates the optimum parameters (e.g., the major range, minor range, and angle of direction) to account for the anisotropic influence.

Identify directional influences.

Detrending Tools

Sometimes it is useful to remove a surface trend from the data and use kriging or cokriging on the detrended (residual) data. For example, if an ozone data set has an overwhelming directional influence from easterly winds, the user may want to remove this trend for a better understanding of actual ozone levels in the area. Detrending decomposes data into two components, a deterministic trend component and an autocorrelated random component. Once the trend is removed, the user will carry out

kriging on the residuals. Before the final predictions are actually calculated, the trend is added back to the output surface.

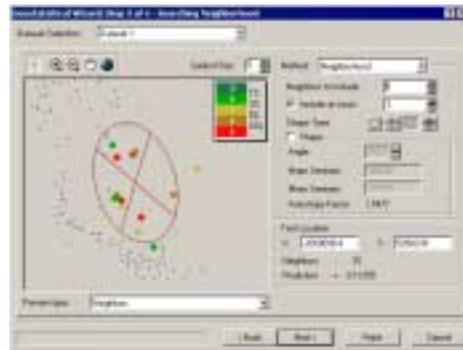


Detrending decomposes the data into a deterministic trend component and an autocorrelated random component.

Detrending

Search Neighborhood Specification

As the data locations become farther away from a location where the value is unknown, they may not be as useful when predicting the value at an unmeasured location. At some distance, the points will have no correlation with the prediction location, and it is possible that they may even be located in an area much different than the unknown location. Therefore, it is common practice to specify a search neighborhood that limits the number and the configuration of the points that will be used in the predictions. There are two controlling mechanisms to limit the points used, namely, specifying the shape of the neighborhood and establishing constraints on the points within and outside the shape.



The points highlighted in the data view give an indicator of the weights (absolute value in percent) associated with each point located in the moving window. The weights are used to estimate the value at the unknown location, which is at the center of the crosshair.

Searching Neighborhood

Data Transformations

Transformations can be used to make the data more normally distributed and satisfy assumptions of constant variability. Data transformations are performed before using some geostatistical methods such as disjunctive kriging and for maps that require the normality assumption. Quantile and probability maps from simple, ordinary, and universal kriging methods require the normality assumption. Geostatistical Analyst supports transformations including Box-Cox (also known as power transformations), logarithmic, arcsine, and normal score.

Probability Density



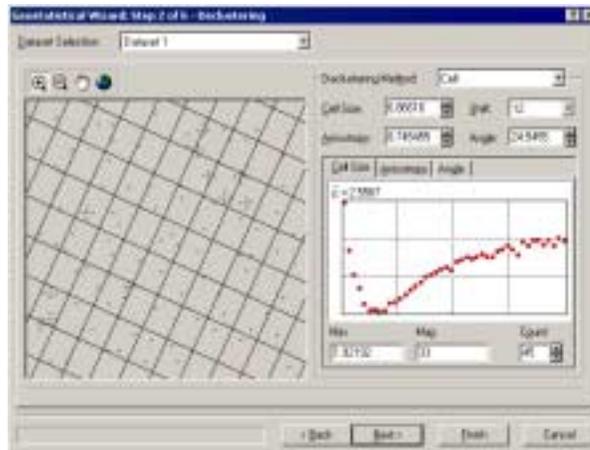
Cumulative Distribution



Using transformations makes variances constant throughout the study area and makes the data more normally distributed.

Declustering Method

Oftentimes the data may have a high density of sample points near one another. This may lead to a spatial preferentiality among the data points. If data is preferentially sampled when it is spatially autocorrelated, the resulting histogram from the sample may not reflect the histogram of the population. Therefore, the declustering method assigns higher weights to the less densely populated sample points and lower weights to the high density areas.

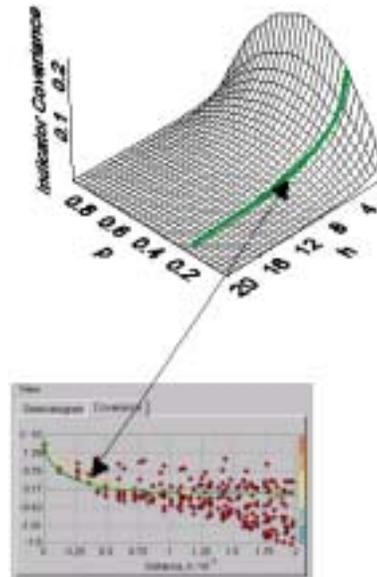


In cell declustering, rectangular cells are arranged over the data locations in a grid, and the weight attached to each data location is inversely proportional to the number of data points in its cell. By adjusting the grid size and orientation, the data points can be effectively declustered.

Correct for spatial preferentiality.

Bivariate Normal Distribution

Disjunctive kriging requires that the data have a bivariate normal distribution. A bivariate distribution describes relative frequencies of occurrence in the population of pairs of values. Also, to develop probability and quantile maps using simple kriging, ordinary kriging, and universal kriging, the user can assume that the data comes from a full multivariate normal distribution. By checking for bivariate normality, it is reasonable to assume that the data comes from a full multivariate normal distribution.



The darker line is the theoretical curve of indicator covariance assuming the data comes from a bivariate normal distribution, and the lighter line is fitted to the observed indicator data. Thus, the darker line and the lighter line should be similar if the data has a bivariate normal distribution.

Bivariate Normal Distribution

Perform Diagnostics

After creating a prediction surface, it is useful to recognize if the model is optimal for the data set in question. Geostatistical Analyst provides cross-validation and validation tools that allow the analyst to evaluate the surface. The tools quantify the "accuracy" of the model. The user can either accept the model and its parameters or refine the parameters of the model to create a better surface.

Cross-Validation and Validation

Cross-validation and validation help the analyst make an informed decision as to which model provides the best predictions. The calculated statistics serve as diagnostics that indicate whether the model and/or its associated parameter values are reasonable.

Cross-validation and validation use the following idea—withhold one or more data locations and then predict their associated data using the data at the rest of the locations. In this way, the analyst can compare the predicted value to the observed value and from this, acquire useful information about some of the analyst's previous decisions on the kriging model (e.g., the semivariogram parameters, the searching neighborhood, and so on).

■ Cross-Validation

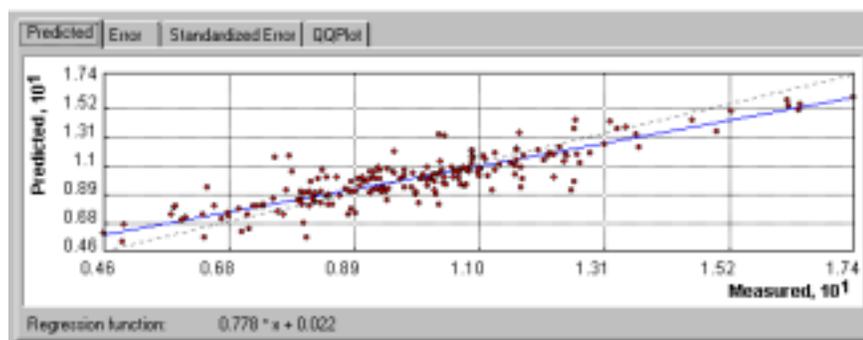
Cross-validation uses all of the data to estimate the trend and autocorrelation models. Then it removes each data location, one at a time, and predicts the associated data value. For all points, cross-validation compares the measured and predicted values. After completing cross-validation, some data locations may be set aside as unusual, requiring the trend and autocorrelation models to be refitted.

■ Validation

Validation uses part of the data to develop the trend and autocorrelation models to be used for prediction. Then predictions to the known locations are compared with the measured data. Validation checks whether a "protocol" of decisions is valid (for example, choice of semivariogram model, choice of lag size, and choice of search neighborhood).

■ Graphs and Summaries for Cross-Validation and Validation

Geostatistical Analyst gives several graphs and summaries of the measurement values versus the predicted values.

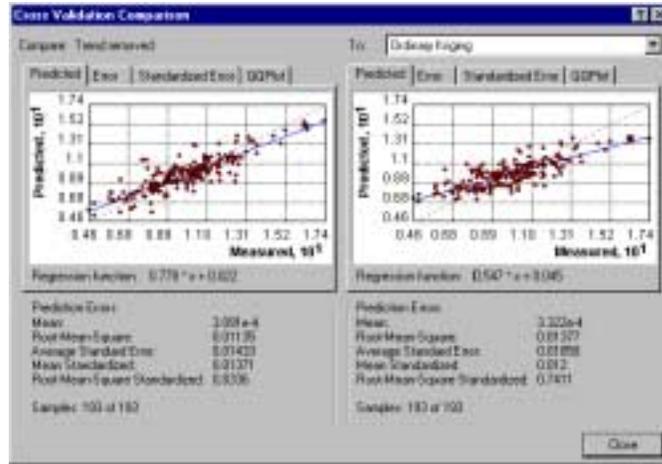


The **Predicted** plot has a blue fitted line through the scatter of points and an equation is given just below the plot. The **Error** plot is the same as the prediction plot, except here the true values are subtracted from the predicted values. For the **Standardized Error** plot, the true values are subtracted from the predicted values and then divided by the estimated kriging standard errors. All three of these plots help to show how well kriging is predicting. With autocorrelation and a good kriging model, the blue line should be closer to the 1:1 (black dashed) line.

The **QQPlot** shows the quantiles of the difference between the standardized errors and the corresponding quantiles from a standard normal distribution. If the errors of the predictions from their true values are normally distributed, the points should lie roughly along the dashed line. If the errors are normally distributed, the user can be confident of using methods that rely on normality (e.g., quantile maps in ordinary kriging).

Model Comparison

A Comparison dialog box uses the cross-validation technique and allows the analyst to examine the statistics and the plots side by side. Generally, the best model is the one that has the standardized mean nearest to zero, the smallest root-mean-squared prediction error, the average standard error nearest the root-mean-squared prediction error, and the standardized root-mean-squared prediction error nearest to one.



The comparison dialog box allows the user to examine statistics and plots of different models side by side.

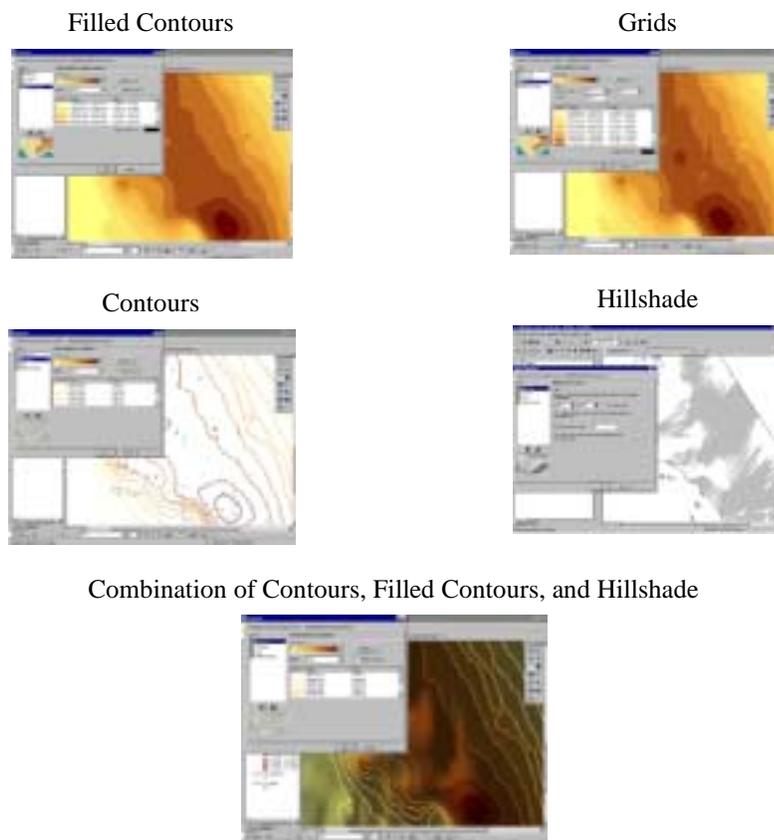
Display Options

A geostatistical layer's functionality is similar to all ArcMap layers. The user can add it to ArcMap, remove it, display it, and alter the symbology in countless ways. However, a geostatistical layer differs from other layers because of the way it is created and stored. A geostatistical layer can only be created by Geostatistical Analyst. Most ArcMap layer types store the reference to the data source, the symbology for displaying the layer, and other defining characteristics. A geostatistical layer stores the source of the data from which it was created (usually a point feature layer), the symbology, and other defining characteristics, but it also stores the model parameters from the interpolation, which can be refined at any time. From the Properties dialog for a geostatistical layer, the user can view both the original data source and the model parameters.

Not only can the user identify the source of the input points and the model parameters, but the user can also retrieve general information with the general tab, see and alter the layer's map extent with the extent tab, change the symbology with the symbology tab, and set the transparency and whether to show map tips with the display tab.

Display Formats

A geostatistical layer can be viewed in four different formats: filled contours, contours, grid, or hillshade. The analyst can also combine multiple formats in a single display of the layer to achieve various effects. A full range of symbology and controlling parameters exists for each format.



3. Conclusion

Geostatistical Analyst is much more than a surface creation software package. Geostatistical Analyst is a revolutionary technology that provides a dynamic environment with a wide variety of tools and a friendly wizard interface to explore data, analyze anomalies, and optimally display an interpolated surface with associated uncertainties. Geostatistical Analyst gives the user the power to fully understand the qualitative and quantitative aspects of their data. By providing the user with the freedom to statistically predict and model situations and incorporating powerful exploration and visualization tools, Geostatistical Analyst effectively bridges the gap between geostatistics and GIS.

Data Used in Screen Shots

- Agriculture data was provided by the University of Illinois.
- Radioceasium forest berry contamination data was provided by the Institute of radiation safety "BELRAD", Minsk, Belarus.
- Air quality data was provided by the California Environmental Protection Agency, Air Resource Board.
- Carpathian Mountains data was provided by the USDA Forest Service, Riverside, California.

- Radiocesium soil contamination data was provided by the International Sakharov Environmental University, Minsk, Belarus.