SAS/STAT® 14.2 User's Guide
The KRIGE2D Procedure

This document is an individual chapter from *SAS/STAT® 14.2 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2016. *SAS/STAT® 14.2 User's Guide*. Cary, NC: SAS Institute Inc.

SAS/STAT® 14.2 User's Guide

Copyright © 2016, SAS Institute Inc., Cary, NC, USA

All Rights Reserved. Produced in the United States of America.

For a hard-copy book: No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

For a web download or e-book: Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication, or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a), and DFAR 227.7202-4, and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

November 2016

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

Chapter 68
The KRIGE2D Procedure

Contents

Overview: KRIGE2D Procedure **5002**
- Introduction to Spatial Prediction 5002

Getting Started: KRIGE2D Procedure **5003**
- Spatial Prediction Using Kriging, Contour Plots 5003

Syntax: KRIGE2D Procedure **5008**
- PROC KRIGE2D Statement 5009
- BY Statement 5015
- COORDINATES Statement 5016
- GRID Statement 5016
- ID Statement 5019
- PREDICT Statement 5019
- MODEL Statement 5021
- RESTORE Statement 5029

Details: KRIGE2D Procedure **5031**
- Theoretical Semivariogram Models 5031
- The Nugget Effect 5038
- Anisotropic Models 5040
  - Geometric Anisotropy 5040
  - Zonal Anisotropy 5043
  - Anisotropic Nugget Effect 5046
- Details of Ordinary Kriging 5047
  - Introduction 5047
  - Spatial Random Fields 5047
  - Ordinary Kriging 5048
- Computational Resources 5050
- Output Data Sets 5051
- Displayed Output 5052
- ODS Table Names 5053
- ODS Graphics 5053

Examples: KRIGE2D Procedure **5054**
- Example 68.1: Spatial Prediction of Pollutant Concentration 5054
- Example 68.2: Investigating the Effect of Model Specification on Spatial Prediction . 5064
- Example 68.3: Data Quality and Prediction with Missing Values 5068

References **5073**

Overview: KRIGE2D Procedure

The KRIGE2D procedure performs ordinary kriging in two dimensions. PROC KRIGE2D can handle anisotropic and nested semivariogram models. Eight semivariogram models are supported: the Gaussian, exponential, spherical, power, cubic, pentaspherical, sine hole effect, and Matérn models. A single nugget effect is also supported. You can specify the correlation model by naming the form and supplying the associated parameters, or by using the contents of an item store file that was previously created by PROC VARIOGRAM.

You can specify the locations of kriging predictions in a GRID statement, or they can be read from a SAS data set. The grid specification is most suitable for a regular grid; the data set specification can handle any irregular pattern of points.

Local kriging is supported through the specification of a radius around a grid point or the specification of the number of nearest neighbors to use in the kriging system. When you perform local kriging, a separate kriging system is solved at each grid point by using a neighborhood of the data point established by the radius or number specification.

The KRIGE2D procedure writes the kriging predictions and associated standard errors for each grid to an output data set. When you perform local kriging, PROC KRIGE2D writes the neighborhood information for each grid point to an additional, optional data set. The KRIGE2D procedure does not produce any displayed output.

The KRIGE2D procedure uses ODS Graphics to create graphs as part of its output. For general information about ODS Graphics, see Chapter 21, "Statistical Graphics Using ODS." For more information about the graphics available in PROC KRIGE2D, see the section "ODS Graphics" on page 5053.

Introduction to Spatial Prediction

Many activities in science and technology involve measurements of one or more quantities at given spatial locations, with the goal of predicting the measured quantities at unsampled locations. Application areas include reservoir prediction in mining and petroleum exploration, in addition to modeling in a broad spectrum of fields (for example, environmental health, environmental pollution, natural resources and energy, hydrology, and risk analysis). Often, the unsampled locations are on a regular grid, and the predictions are used to produce surface plots or contour maps.

The preceding tasks fall within the scope of *spatial prediction*, which, in general, is any prediction method that incorporates spatial dependence. The study of these tasks involves naturally occurring uncertainties that cannot be ignored. Stochastic analysis frameworks and methods are often used to account for these uncertainties. Hence, the terms *stochastic spatial prediction* and *stochastic modeling* are also used to characterize this type of analysis.

A popular method of spatial prediction is *ordinary kriging*, which produces both predicted values and associated standard errors. Ordinary kriging requires the complete specification (the form and parameter values) of the spatial dependence that characterizes the spatial process. For this purpose, models for the spatial dependence are expressed in terms of the distance between any two locations in the spatial domain of interest. These models take the form of a covariance or semivariance function.

Spatial prediction, then, involves two steps. First, you model the covariance or semivariance of the spatial process. These measures are typically not known in advance. This step involves computing an empirical estimate, in addition to determining both the mathematical form and the values of any parameters for a theoretical form of the dependence model. Second, you use this dependence model to solve the kriging system at a specified set of spatial points, resulting in predicted values and associated standard errors.

The KRIGE2D procedure performs the second of these steps by using ordinary kriging of two-dimensional data.

This introduction concludes with a note on terminology. You might commonly encounter the terms *estimation* and *prediction* used interchangeably by experts in different fields; this could be a source of confusion. A precise statistical vernacular uses the term *estimation* to refer to inferences about the value of fixed but unknown parameters, whereas *prediction* concerns inferences about the value of random variables—see, for example, Cressie (1993, p. 106). In light of these definitions, kriging methods are clearly predictive techniques, since they are concerned with making inferences about the value of a spatial random field at observed or unobserved locations. The SAS/STAT suite of procedures for spatial analysis and prediction (VARIOGRAM, KRIGE2D, and SIM2D) follows the statistical vernacular in the use of the terms *estimation* and *prediction*.

Getting Started: KRIGE2D Procedure

Spatial Prediction Using Kriging, Contour Plots

After an appropriate semivariogram model is chosen, a number of choices are involved in producing the kriging surface. In order to illustrate these choices, you use the theoretical semivariogram model that was fitted to the coal seam thickness data empirical semivariogram in "Theoretical Semivariogram Model Fitting" on page 10036 in Chapter 124, "The VARIOGRAM Procedure." This model is Gaussian,

$$\gamma_z(h) = c_0 \left[1 - \exp\left(-\frac{h^2}{a_0^2} \right) \right]$$

with a scale of $c_0 = 7.4599$ (that is, the model sill) and a range of $a_0 = 30.1111$, based on the weighted least squares fitting results in the PROC VARIOGRAM example.

The first choice is whether to use local or global kriging. Local kriging uses only data points in the neighborhood of a grid point, and you choose this type of analysis by specifying a data search radius around the grid point. Global kriging uses all data points.

The most important consideration in this decision is the spatial covariance structure. Global kriging is appropriate when the correlation range ϵ is approximately equal to the length of the spatial domain. The correlation range ϵ is the distance r_ϵ (also known as *effective* or *practical* range) at which the covariance is 5% of its value at zero. That is,

$$C_Z(r_\epsilon) = 0.05 C_z(0)$$

For a Gaussian model, r_ϵ is $\sqrt{3}a_0 \approx 52{,}000$ feet. The data points are scattered uniformly throughout a 100×100 (10^6 ft^2) area. Hence, the linear dimension of the data is nearly double the r_ϵ range. This indicates that local kriging rather than global kriging is appropriate because data that are farther away than r_ϵ essentially add to the computational burden without significant contribution to the prediction. The following DATA step inputs the thickness data set thick, which is available from the Sashelp library. In the thick data set, thickness is represented by the Thick variable.

```
title 'Spatial Prediction With Kriging';

data thick;
   input East North Thick @@;
   label Thick='Coal Seam Thickness';
   datalines;
    0.7  59.6  34.1   2.1  82.7  42.2   4.7  75.1  39.5
    4.8  52.8  34.3   5.9  67.1  37.0   6.0  35.7  35.9
    6.4  33.7  36.4   7.0  46.7  34.6   8.2  40.1  35.4
   13.3   0.6  44.7  13.3  68.2  37.8  13.4  31.3  37.8
   17.8   6.9  43.9  20.1  66.3  37.7  22.7  87.6  42.8
   23.0  93.9  43.6  24.3  73.0  39.3  24.8  15.1  42.3
   24.8  26.3  39.7  26.4  58.0  36.9  26.9  65.0  37.8
   27.7  83.3  41.8  27.9  90.8  43.3  29.1  47.9  36.7
   29.5  89.4  43.0  30.1   6.1  43.6  30.8  12.1  42.8
   32.7  40.2  37.5  34.8   8.1  43.3  35.3  32.0  38.8
   37.0  70.3  39.2  38.2  77.9  40.7  38.9  23.3  40.5
   39.4  82.5  41.4  43.0   4.7  43.3  43.7   7.6  43.1
   46.4  84.1  41.5  46.7  10.6  42.6  49.9  22.1  40.7
   51.0  88.8  42.0  52.8  68.9  39.3  52.9  32.7  39.2
   55.5  92.9  42.2  56.0   1.6  42.7  60.6  75.2  40.1
   62.1  26.6  40.1  63.0  12.7  41.8  69.0  75.6  40.1
   70.5  83.7  40.9  70.9  11.0  41.7  71.5  29.5  39.8
   78.1  45.5  38.7  78.2   9.1  41.7  78.4  20.0  40.8
   80.5  55.9  38.7  81.1  51.0  38.6  83.8   7.9  41.6
   84.5  11.0  41.5  85.2  67.3  39.4  85.5  73.0  39.8
   86.7  70.4  39.6  87.2  55.7  38.8  88.1   0.0  41.6
   88.4  12.1  41.3  88.4  99.6  41.2  88.8  82.9  40.5
   88.9   6.2  41.5  90.6   7.0  41.5  90.7  49.6  38.9
   91.5  55.4  39.0  92.9  46.8  39.1  93.4  70.9  39.7
   55.8  50.5  38.1  96.2  84.3  40.3  98.2  58.2  39.5
;
```

Local kriging is performed by using only data points within a specified radius of each grid point. In this example, a radius of 60,000 feet is used. Other choices involved in local kriging are the minimum and maximum number of data points in each neighborhood (around a grid point). The minimum number is left at the default value of 20; the maximum number defaults to all observations in the data set within the specified radius.

The last step in contouring the data is to define the prediction grid point (node) locations. The prediction grid is typically rectangular, and you decide on the grid points population and spacing based on your available data in addition to your application needs. A convenient area that encompasses all the data points is a square of side length 100,000 feet. In the present analysis, a distance of 2,500 feet between nodes in the prediction grid is selected to obtain a smooth contour plot. Based on this choice, you obtain predictions on a square grid with 41 nodes on each side, which yields a total of 1681 grid points.

You can visualize the outcome of your analysis by using the PLOTS option in the PROC KRIGE2D statement. By default, PROC KRIGE2D produces one plot that displays the kriging prediction and its corresponding standard error at each output grid point. The locations of the Thick observations are displayed too, as outlines in the default plot. You can also ask for a plot of the thick data set observations and their values by specifying the OBSERV option in the PLOTS option.

The kriging analysis with the KRIGE2D procedure requires that you provide the prediction parameters in the PREDICT statement. You use the VAR= option to specify that you want to use the Thick variable in the kriging system, and the RADIUS= option to specify the radius of the local kriging regression. In this scenario you want to consider for your predictions all the neighboring data within a radius of 60,000 feet from each prediction location. You can specify more than one PREDICT statements; for example, you can do this when you want predictions for different variables in your DATA= data set.

The coordinates of your variable are specified in the COORDINATES statement. The MODEL statement contains the parameters that describe your data spatial correlation. Namely, the FORM= option specifies the model type, based on its mathematical form. The SCALE= and RANGE= options specify the model sill and range, respectively. You can specify more than one MODEL statement for the same PREDICT statement in order to obtain predictions based on different correlation models.

When you use the RADIUS= option to perform local kriging, as in the present example, it is suggested that the radius parameter be at least as large as your model range, so that you include data points that can contribute to your prediction.

Eventually, you specify the region of predictions with the GRID statement. The following SAS statements compute the kriged surface by using the preceding options and grid choice:

```
ods graphics on;

proc krige2d data=thick;
   coordinates xc=East yc=North;
   predict var=Thick radius=60;
   model scale=7.4599 range=30.1111 form=gauss;
   grid x=0 to 100 by 2.5 y=0 to 100 by 2.5;
run;

ods graphics off;
```

The table in Figure 68.1 shows the number of observations read and used in the kriging prediction. This table provides you with useful information in case you have missing values in the input data.

Figure 68.1 Number of Observations for the thick Data Set

Spatial Prediction With Kriging

The KRIGE2D Procedure

Dependent Variable: Thick

Number of Observations Read	75
Number of Observations Used	75

Figure 68.2 shows some general information about the kriging analysis. This includes the count of the output grid points. You have specified the RADIUS= option; therefore you also see that local kriging is requested. Because this is a local analysis, the table also displays the parameters related to the neighborhood search around the grid points.

Figure 68.2 Kriging Analysis Information

Kriging Information	
Prediction Grid Points	1681
Type of Analysis	Local
Neighborhood Search Radius	60
Minimum Neighbors	20
Maximum Neighbors	All Within Radius

The covariance model parameters, including the effective range of the Gaussian model you specified, are shown in Figure 68.3.

Figure 68.3 Kriging Covariance Model Information

Spatial Prediction With Kriging

The KRIGE2D Procedure

Dependent Variable: Thick
Prediction: Pred1, Model: Model1

Covariance Model Information	
Type	Gaussian
Sill	7.4599
Range	30.1111
Effective Range	52.153955
Nugget Effect	0

Figure 68.4 shows a map of the kriging prediction contours based on the Thick observations in the specified spatial domain. The prediction error is displayed as a surface in the background.

Figure 68.4 Contour Plot of Kriged Coal Seam Thickness

Note the locations of the observed data in Figure 68.4. The figure suggests that the Thick sampling locations are not ideally spread around the prediction area; however, there are no extended areas lacking measurements.

Based on the spatial distribution of the Thick data and the range r_ϵ of your covariance model, you can roughly see that for each prediction location there are at least several neighboring data points that contribute to the prediction value. Except perhaps for the nodes close to the boundaries of the prediction grid, you can then expect the prediction errors to be reasonably low compared to the predicted Thick values.

The kriging outcome in Figure 68.4 indicates that the standard errors are smaller in the neighborhoods where data are available. The size of these neighborhoods depends on the range of the specified covariance model that characterizes the spatial continuity of the domain, and on the prediction radius, if one is specified as in this example. The standard errors tend to increase toward the borders of the prediction area, beyond which no observations are available.

Syntax: KRIGE2D Procedure

The following statements are available in the KRIGE2D procedure:

```
PROC KRIGE2D options ;
    BY variables ;
    COORDINATES | COORD coordinate-variables ;
    GRID grid-options ;
    ID variable ;
    PREDICT | PRED | P predict-options ;
    MODEL model-options ;
    RESTORE restore-options ;
```

The PREDICT and MODEL statements are hierarchical; the PREDICT statement is followed by a MODEL statement. If more than one MODEL statement is given, only the last one is used for the analysis. The MODEL statement following a PREDICT statement uses the variable and neighborhood specifications in that PREDICT statement.

You must specify at least one PREDICT statement and one MODEL statement. You must supply a single COORDINATES statement to identify the *x* and *y* coordinate variables in the input data set. You must also specify a single GRID statement to include the grid information.

Table 68.1 outlines the *options* available in PROC KRIGE2D classified by function.

Table 68.1 Options Available in the KRIGE2D Procedure

Task	Statement	Option
Data Set Options		
Specifies input data set	PROC KRIGE2D	DATA=
Specifies grid data set	GRID	GDATA=
Specifies labels for individual grid points or in 1-D	GRID	LABEL
Specifies model data set	MODEL	MDATA=
Writes kriging predictions and standard errors	PROC KRIGE2D	OUTEST=
Writes neighborhood information for each grid point	PROC KRIGE2D	OUTNBHD=
Specifies plot display and options	PROC KRIGE2D	PLOTS
Declaring the Role of Variables		
Specifies variables to define analysis subgroups	BY	
Specifies variable with observation labels	ID	
Specifies the variables to be predicted (kriged)	PREDICT	VAR=
Specifies the *x* and *y* coordinate variables in the DATA= data set	COORDINATES	XC= YC=
Specifies the *x* and *y* coordinate variables in the GDATA= data set	GRID	XC= YC=
Controlling the Prediction		
Specifies the number of grid points in one-dimensional cases	GRID	NPTS=

Table 68.1 *continued*

Task	Statement	Option
Controlling Kriging Neighborhoods		
Specifies the radius of a neighborhood for all grid points	PREDICT	RADIUS=
Specifies the number of neighbors for all grid points	PREDICT	NUMPOINTS=
Specifies the maximum of neighbors for all grid points	PREDICT	MAXPOINTS=
Specifies the minimum of neighbors for all grid points	PREDICT	MINPOINTS=
Specifies the action when maximum not met	PREDICT	NODECREMENT
Specifies the action when minimum not met	PREDICT	NOINCREMENT
Controlling the Semivariogram Model		
Specifies an angle for an anisotropic model	MODEL	ANGLE=
Specifies a type with a functional form	MODEL	FORM=
Specifies an item store with correlation information	RESTORE	IN=
Specifies a nugget effect	MODEL	NUGGET=
Allows power exponent values outside [0,2)	MODEL	POWNOBOUND
Specifies a range parameter	MODEL	RANGE=
Specifies a minor-major axis ratio for an anisotropic model	MODEL	RATIO=
Specifies a scale parameter	MODEL	SCALE=
Specifies model and parameters from an item store	MODEL	STORESELECT

PROC KRIGE2D Statement

PROC KRIGE2D *options* **;**

The PROC KRIGE2D statement invokes the KRIGE2D procedure. Table 68.2 summarizes the *options* available in the PROC KRIGE2D statement.

Table 68.2 PROC KRIGE2D Statement Options

Option	Description
DATA=	Specifies input data set
IDGLOBAL	Uses ascending observation numbers as observation labels across BY groups
IDNUM	Uses observation number as observation labels
NOPRINT	Suppresses the normal display of results
OUTEST=	Writes kriging predictions and standard errors
OUTNBHD=	Writes neighborhood information for each grid point
PLOTS	Specifies plot display and options
PREDICTION	Produces the kriging prediction plot
SEMIVARIOGRAM	Produces the semivariogram used for the kriging prediction
SINGULARMSG=	Controls the number of warning messages displayed for a singular matrix

You can specify the following *options* in the PROC KRIGE2D statement.

DATA=*SAS-data-set*
specifies a SAS data set that contains the *x* and *y* coordinate variables and the VAR= variables in the PREDICT statement.

IDGLOBAL
specifies that ascending observation numbers be used across BY groups for the observation labels in the appropriate output data sets and the OBSERVATIONS plot, instead of resetting the observation number in the beginning of each BY group. The IDGLOBAL option is ignored if no BY variables are specified. Also, if you specify the ID statement, then the IDGLOBAL option is ignored unless you also specify the IDNUM option in the PROC KRIGE2D statement.

IDNUM
specifies that the observation number be used for the observation labels in the appropriate output data sets and the OBSERVATIONS plot. The IDNUM option takes effect when you specify the ID statement; otherwise, it is ignored.

NOPRINT
suppresses the normal display of results. The NOPRINT option is useful when you want only to create one or more output data sets with the procedure. **NOTE:** This option temporarily disables the Output Delivery System (ODS); see the section "ODS Graphics" on page 5053 for more information.

OUTEST=*SAS-data-set*

OUTE=*SAS-data-set*
specifies a SAS data set in which to store the kriging predictions, standard errors, and grid location. For details, see the section "OUTEST=*SAS-data-set*" on page 5051.

OUTNBHD=*SAS-data-set*

OUTN=*SAS-data-set*
specifies a SAS data set in which to store the neighborhood information for each grid point. Information is written to this data set only if one or more PREDICT statements have options that specify local kriging. For details, see the section "OUTNBHD=*SAS-data-set*" on page 5051.

PLOTS **<** (*global-plot-option*) **>** **<** **=** *plot-request* **<** (*options*) **>** **>**

PLOTS **<** (*global-plot-option*) **>** **<** **=** (*plot-request* **<** (*options*) **>** **<** **...** *plot-request* **<** (*options*) **>** **>**) **>**
controls the plots produced through ODS Graphics. When you specify only one plot request, you can omit the parentheses around the plot request. Here are some examples:

```
plots=none
plots=observ
plots=(observ(outl) prediction)
plots=(prediction(fill=pred line=se obs=grad) prediction(fill=se))
```

ODS Graphics must be enabled before plots can be requested. For example:

```
ods graphics on;

proc krige2d data=thick;
   coordinates xc=East yc=North;
   predict var=thick r=60;
   model scale=7.4599 range=30.1111 form=gauss;
   grid x=0 to 100 by 10 y=0 to 100 by 10;
run;

ods graphics off;
```

For more information about enabling and disabling ODS Graphics, see the section "Enabling and Disabling ODS Graphics" on page 607 in Chapter 21, "Statistical Graphics Using ODS."

If ODS Graphics is enabled but you omitted the PLOTS option or have specified PLOTS=ALL, then PROC KRIGE2D produces a default plot for each MODEL statement of every PREDICT statement that you specify. The default PROC KRIGE2D plot displays a contour plot of the kriging prediction and the gradient of the kriging prediction standard error at every location of the prediction grid, in addition to empty circles that indicate the observation locations. See Figure 68.4 for an example of the default KRIGE2D plot.

The following *global-plot-option* is available:

ONLY
suppresses the default plot. Only plots that are specifically requested are displayed.

The following individual *plot-requests* and plot *options* are available:

ALL
produces all appropriate plots. You can specify other *options* with ALL. For example, to request the default plot and an additional plot of the predictions, specify PLOTS=(ALL PREDICTION).

EQUATE
specifies that all appropriate plots be produced in a way that the axes coordinates have equal size units.

NONE
suppresses all plots.

OBSERVATIONS < (*observations-plot-options*) >

OBSERV < (*observations-plot-options*) >

OBS < (*observations-plot-options*) >
produces the observed data plot. Only one observations plot is created if you specify the OBSERVATIONS option more than once within a PLOTS option.

The OBSERVATIONS option has the following suboptions:

GRADIENT
specifies that observations be displayed as circles colored by the observed measurement.

LABEL < (*label-option*) >
labels the observations. The label is the ID variable if the ID statement is specified; otherwise, it is the observation number. The *label-option* can be one of the following:

EQ=*number*
specifies that labels show for any observation whose value is equal to the specified *number*.

MAX=*number*
specifies that labels show for observations with values smaller than or equal to the specified *number*.

MIN=*number*
specifies that labels show for observations with values equal to or greater than the specified *number*.

If you specify multiple instances of the OBSERVATIONS option and you specify the LABEL suboption in any of those, then the resulting observations plot displays the observations labels. If more than one *label-option* is specified in multiple LABEL suboptions, then the prevailing *label-option* in the resulting OBSERVATIONS plot emerges by adhering to the choosing order: MIN, MAX, EQ.

OUTLINE
specifies that observations be displayed as circles with a border but with a completely transparent fill.

OUTLINEGRADIENT
is the same as OBSERVATIONS(GRADIENT) except that a border is shown around each observation.

SHOWMISSING
specifies that observations with missing values be displayed in addition to the observations with nonmissing values. By default, missing values locations are not shown on the plot. If you specify multiple instances of the OBSERVATIONS option and you specify the SHOWMISSING suboption in any of those, then the resulting observations plot displays the observations with missing values.

If you omit any of the GRADIENT, OUTLINE, and OUTLINEGRADIENT suboptions, the OUTLINEGRADIENT is the default suboption. If you specify multiple instances of the OBSERVATIONS option or multiple suboptions for OBSERVATIONS, then the resulting observations plot honors the last specified GRADIENT, OUTLINE, or OUTLINEGRADIENT suboption.

PREDICTION < (*prediction-plot-options*) >
PRED < (*prediction-plot-options*) >
specifies that the kriging prediction plot be produced. You can specify the PREDICTION option multiple times in the same PLOTS option to request instances of plots with the following *prediction-plot-options*:

ALPHA=*number*

specifies a parameter to obtain the confidence level for constructing confidence limits based on the prediction standard error. The value of *number* must be between 0 and 1, and the confidence level is $1-$*number*. The default is ALPHA=0.05; this corresponds to the confidence level of 95%, or about 1.96 times the prediction standard error. The ALPHA= suboption is used only for prediction plots in one dimension, and it is incompatible with the FILL and LINE suboptions.

CLONLY

specifies that only the confidence limits be shown in a prediction plot without the predicted values. This suboption can be useful for identifying confidence limits when the prediction standard error is small at the prediction locations. CLONLY is used only for prediction plots in one dimension, and it is incompatible with the FILL and LINE suboptions.

CONNP

specifies that grid points that you provide as individual prediction locations be connected with a line on the area map. This suboption is ignored when you have a single grid point, a prediction grid in two dimensions, or when you also specify the NOMAP suboption. The CONNP suboption is incompatible with the FILL and LINE suboptions.

FILL=NONE | PRED | SE

produces a surface plot for either the predicted values or the standard errors. FILL=SE is the default. However, if you omit the FILL suboption, the behavior depends on the LINE suboption as follows: If you specify LINE=NONE or entirely omit the LINE suboption, then the FILL suboption is set to its default value. If LINE=PRED or LINE=SE, then the FILL suboption is set to the same value as the LINE suboption.

LINE=NONE | PRED | SE

produces a contour line plot for either the predicted values or the standard errors. LINE=PRED is the default. However, if you omit the LINE suboption the behavior depends on the FILL suboption as follows: If you specify FILL=NONE or entirely omit the FILL suboption, then the LINE suboption is set to its default value. If FILL=PRED or FILL=SE, then the LINE suboption is set to the same value as the FILL suboption.

NOMAP

specifies that the prediction plot be produced without a map of the domain where you have observations. The NOMAP suboption is used in the case of prediction in one dimension or at individual points. It is incompatible with the FILL and LINE suboptions.

OBS=*obs-options*

produces an overlaid scatter plot of the observations in addition to the specified contour plots. The following *obs-options* are available:

GRAD

specifies that observations be displayed as circles colored by the observed measurement. The same color gradient displays the prediction surface and the observations. Observations where the prediction is close to the observed values have similar colors—the greater the contrast between the color of an observation and the surface, the larger the prediction standard error is at that point.

LINEGRAD
is the same as OBS=GRAD except that a border is shown around each observation. This option is useful for identifying the location of observations where the standard errors are small, because at these points the color of the observations and the color of the surface are indistinguishable.

NONE
specifies that no observations be displayed.

OUTL
specifies that observations be displayed as circles with a border but with a completely transparent fill.

OBS=NONE is the default when you specify a grid in two dimensions, and OBS=LINEGRAD is the default used in the area map when you have a grid in one dimension. However, the default PROC KRIGE2D plot for a surface grid displays the observations locations as outlines.

SHOWD
specifies that the horizontal axis in scatter plots of linear prediction grids show the distance between grid points instead of the grid points' coordinates. When the area map is displayed, the prediction locations are also connected with a line. In all other grid configurations the SHOWD suboption is ignored, and it is incompatible with the FILL and LINE suboptions.

SHOWP
specifies that the grid points in band plots of linear prediction grids be shown as marks on the band plot. In all other grid configurations the SHOWP suboption is ignored, and it is incompatible with the FILL and LINE suboptions.

TYPE=BAND | SCAT
requests a particular type of plot when you have a linear grid, regardless of the default PREDICTION plot behavior in this case. The TYPE suboption is incompatible with the FILL and LINE suboptions.

If you specify multiple instances of the ALPHA, FILL, LINE, OBS, or TYPE suboptions in the same PREDICTION option, then the resulting predictions plot honors the last value specified for any of the suboptions. Any combination where you specify FILL=NONE and LINE=NONE is not available. When the prediction grid is in two dimensions, only the FILL, LINE, and OBS suboptions apply. If you specify incompatible suboptions in the same PREDICTION plot, then the plot instance is skipped.

The PREDICTION option produces a surface or contour line plot for grids in two dimensions and a band plot or scatter plot with error bars for grids in one dimension or individual points. In two dimensions the plot illustrates the predicted values and prediction error at each grid point. By default, when you specify a linear grid with fewer than 10 points, PROC KRIGE2D produces a PREDICTION scatter plot for each one of the prediction grid points. For 10 or more points in a linear grid, the PREDICTION plot is a band plot of the predicted means and the confidence limits at the 95% confidence level. You can override the default behavior in linear grids with the TYPE suboption. Prediction at individual locations always produces a PREDICTION scatter plot.

In cases of prediction in one dimension or at individual points, an area map is produced that shows the observations and the grid points. Band plots of linear grids display the grid points as a

line on the map. When you specify individual prediction locations, the grid points are indicated with marks on the area map. The area map appears on the side of the prediction band plot or scatter plot, unless you specify the NOMAP suboption. You can also label the individual grid points or the ends of linear grid segments with the LABEL option of the GRID statement.

SEMIVARIOGRAM < (*semivar-plot-option*) >

SEMIVAR < (*semivar-plot-option*) >

specifies that the semivariogram used for the kriging prediction be produced. You can use the following *semivar-plot-option*:

MAXD=*number*

specifies a positive value for the upper limit of the semivariogram horizontal axis of distance. The SEMIVARIOGRAM plot extends by default to a distance that depends on the correlation model range. You can use the MAXD= option to adjust the default maximum distance value for the plot.

The SEMIVARIOGRAM option produces a plot for each correlation model that you specify for your prediction tasks. In an anisotropic case, the plot is not produced if you assign different anisotropy angles for different model components. The only exception is when you specify zonal components at right angles with the nonzonal model components. Also, the SEMIVARIOGRAM option is ignored for models that consist of purely zonal components.

SINGULARMSG=*number*

SMSG=*number*

controls the number of warning messages displayed for a singular matrix. When local kriging is performed, a separate kriging system is solved for each grid point. Anytime a singular matrix is encountered, a warning message is displayed up to a total of *number* times. The default is SINGULARMSG=10.

BY Statement

BY *variables* **;**

You can specify a BY statement with PROC KRIGE2D to obtain separate analyses of observations in groups that are defined by the BY variables. When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables. If you specify more than one BY statement, only the last one specified is used.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data by using the SORT procedure with a similar BY statement.
- Specify the NOTSORTED or DESCENDING option in the BY statement for the KRIGE2D procedure. The NOTSORTED option does not mean that the data are unsorted but rather that the data are arranged in groups (according to values of the BY variables) and that these groups are not necessarily in alphabetical or increasing numeric order.
- Create an index on the BY variables by using the DATASETS procedure (in Base SAS software).

For more information about BY-group processing, see the discussion in *SAS Language Reference: Concepts*. For more information about the DATASETS procedure, see the discussion in the *Base SAS Procedures Guide*.

COORDINATES Statement

COORDINATES | COORD *coordinate-variables* **;**

The following two options specify the names of the variables in the DATA= data set that contains the values of the *x* and *y* coordinates of the data.

Only one COORDINATES statement is allowed, and it is applied to all PREDICT statements. In other words, it is assumed that all the VAR= variables in all PREDICT statements have the same *x* and *y* coordinates.

This is not a limitation. Since each VAR= variable is processed separately, observations for which the current VAR= variable is missing are excluded. With the next VAR= variable, the entire data are read again, this time excluding missing values in this next variable. Hence, a single run of PROC KRIGE2D can be used for variables measured at different locations without overlap.

XCOORD= (*variable-name***)**
XC= (*variable-name***)**
specifies the name of the variable that contains the *x* coordinate of the data locations in the DATA= data set.

YCOORD= (*variable-name***)**
YC= (*variable-name***)**
specifies the name of the variable that contains the *y* coordinate of the data locations in the DATA= data set.

GRID Statement

GRID *grid-options* < / *option* > **;**

The GRID statement specifies the grid of spatial locations for kriging predictions. The grid specification is applied to all PREDICT and MODEL statements. Specify the grid in one of the following three ways:

- Specify the *x* and *y* coordinates explicitly for a grid in two dimensions.
- Specify the NPTS= option in addition to the *x* and *y* coordinates to define a grid of individual points or in one dimension.
- Specify the coordinates by using a SAS data set for a grid of individual points or in one dimension.

The GRID statement has the following *grid-options*:

NPTS=*number* **| ALL**
controls specification of a grid in one dimension or a grid of individual prediction locations.

When you specify the NPTS=*number* option and the coordinates of two points in the GRIDDATA= data set or in both the X= and Y= options, you request a linear prediction grid. Its direction is across

the line defined by the specified points. The grid size is equal to the *number* of points that you specify in the NPTS= option, where *number* ≥ 2.

When you specify the NPTS=ALL option and the coordinates for any number of points in the GRIDDATA= data set or in each of the X= and Y= options, the KRIGE2D procedure performs prediction only at the specified individual locations. Use the NPTS=ALL option to examine a set of individual points anywhere on the XY plane or to specify a custom grid in one dimension.

If the number of x coordinates and the number of y coordinates in the X= and Y= options, respectively, are different, then the NPTS= option is ignored; in that case, a two-dimensional grid is used according to the specified X= and Y= options.

If you specify a prediction grid with any number of points other than two in the GRIDDATA= data set, then the option NPTS=ALL has the same effect as omitting the NPTS= option.

X=*number*
X=$x_1, \ldots, x_m$
X=x_1 **TO** x_m
X=x_1 **TO** x_m **BY** δx
specifies the x coordinate of the grid locations.

Y=*number*
Y=$y_1, \ldots, y_m$
Y=y_1 **TO** y_m
Y=y_1 **TO** y_m **BY** δy
specifies the y coordinate of the grid locations.

Use the X= and Y= options of the GRID statement to specify a grid in one or two dimensions, or a grid of individual prediction locations.

For example, the following two GRID statements are equivalent.

```
grid x=1,2,3,4,5  y=0,2,4,6,8,10;

grid x=1 to 5     y=0 to 10 by 2;
```

In the following example, the first GRID statement produces a grid in two dimensions. The second statement produces predictions only for the four individual points at the locations (1,0), (2,5), (3,7), and (4,10) on the XY plane.

```
grid x=1 to 4 y=0,5,7,10;

grid x=1 to 4 y=0,5,7,10 npts=all;
```

In the next example, the first GRID statement specifies a 2-by-2 grid in two dimensions. The second GRID statement specifies a linear grid of eight points. The grid is in the direction of the line defined by the specified points (2,8) and (3,5) on the XY plane and it extends between these two points.

```
grid x=2,3 y=8,5;

grid x=2,3 y=8,5 npts=8;
```

The last example shows a GRID statement that specifies a linear grid made of seven points across the Y axis. In this case, the syntax is sufficient to fully define a linear grid without the NPTS= option.

```
grid x=5 y=3 to 9;
```

To specify grid locations from a SAS data set, you must provide the name of the data set and the variables that contain the values of the *x* and *y* coordinates.

GRIDDATA=*SAS-data-set*

GDATA=*SAS-data-set*

specifies a SAS data set that contains the *x* and *y* grid coordinates. Use the GRIDDATA= option of the GRID statement to specify a grid in one dimension or a grid of individual prediction locations.

XCOORD= (*variable-name***)**

XC= (*variable-name***)**

specifies the name of the variable that contains the *x* coordinate of the grid locations in the GRIDDATA= data set.

YCOORD= (*variable-name***)**

YC= (*variable-name***)**

specifies the name of the variable that contains the *y* coordinate of the grid locations in the GRIDDATA= data set.

You can specify the following *option* in the GRID statement after a slash (/):

LABEL < (*suboption***) > = (***character-list***)**

specifies labels to tag grid points in prediction plots when you use grids in one dimension. You can specify one or more such labels as quoted strings in the *character-list*.

When the number of labels in the *character-list* exceeds the number of points in your grid, the labels in the list are used sequentially and any labels in excess are ignored. When the number of labels in the *character-list* is smaller than the number of points in your grid, the behavior is as follows:

- If an area map is included in the prediction plot, then blank labels are assigned to the remaining nonlabeled grid points on the map.
- For the prediction band and scatter plots, the coordinates of nonlabeled grid points are automatically assigned as their labels.

If the grid points are collinear and the horizontal axis displays distance, then two labels appear by default in the prediction plot. These are assigned to the first and the last points of the grid to help identify the ends of the linear grid segment on the plot map. This label pair is shown only when the plot includes an area map. Specifically, the two labels appear when you request prediction band plots, or prediction scatter plots for which you specify the PREDICTION(SHOWD) suboption, if applicable. The two labels do not appear if you specify explicitly the NOMAP suboption in the PLOTS=PRED option.

The two labels have default values, unless you choose to specify your own labels with the LABEL= option. If you specify more than two labels in the *character-list* under these conditions, then only the first and last labels in the list are used; any additional labels in between are ignored.

The LABEL= option has the following *suboption*:

ALL

specifies that all individual points in the grid be assigned sequentially the labels you specify in the LABEL(ALL)= option when the PREDICTION(SHOWD) suboption is applicable and specified in a prediction scatter plot. In all other cases, the ALL suboption is ignored.

The ALL suboption enables you to override the default behavior when the PREDICTION(SHOWD) suboption is specified (the default behavior is to display labels only for the first and last grid points). As a result, you can use the ALL suboption to label grid points regardless of whether you specify the NOMAP suboption in the PLOTS=PRED option.

The LABEL= option is ignored when you produce prediction plots of grids in two dimensions.

ID Statement

ID *variable* **;**

The ID statement specifies which variable to include for identification of the observations in the OUTNBHD= output data set. The ID statement variable is also used for the labels and tool tips in the OBSERVATIONS plot and the tool tips in the PREDICTION plot.

In the KRIGE2D procedure you can specify only one ID variable in the ID statement. If no ID statement is given, then PROC KRIGE2D uses the observation number in the data sets and the plots.

PREDICT Statement

PREDICT | PRED | P *predict-options* **;**

You can specify the following *options* in a PREDICT statement.

MAXPOINTS=*number*

MAXP=*number*

MAX=*number*

specifies the maximum number of data points in a neighborhood. You specify this option in conjunction with the RADIUS= option. When the number of data points in the neighborhood formed at a given grid point by the RADIUS= option is greater than the MAXPOINTS= value, the RADIUS= value is decreased just enough to honor the MAXPOINTS= value unless you specify the NODECREMENT option. The default is to include all data points within the specified RADIUS= value. Neighborhoods with very large numbers of data points might lead to unnecessarily slow execution times and potential lack of memory issues, depending on the problem setup and your computational resources. In that case, you could use the MAXPOINTS= option to set a cap for your neighborhood size. For details about numerical considerations, see the section "Computational Resources" on page 5050. Unless the RADIUS= option is also specified, when the MAXPOINTS= and NUMPOINTS= options are specified in the same PREDICT statement the MAXPOINTS= option is ignored.

MINPOINTS=*number*

MINP=*number*

MIN=*number*

specifies the minimum number of data points in a neighborhood. You specify this option in conjunction with the RADIUS= option. When the number of data points in the neighborhood formed at a given grid point by the RADIUS= option is less than the MINPOINTS= value, the RADIUS= value is increased just enough to honor the MINPOINTS= value unless you specify the NOINCREMENT option. The default is MINPOINTS=20. When enough data are available, you might improve prediction if you increase this value. When the MINPOINTS= and NUMPOINTS= options are specified in the same PREDICT statement, the MINPOINTS= option is set to the value of NUMPOINTS=.

NODECREMENT | NODECR

requests that the RADIUS= value not be decremented when the MAXPOINTS= value is exceeded at a grid point. This option is relevant only when you specify both a RADIUS= value and a MAXPOINTS= value. In this case, when the number of points in the neighborhood constructed from the RADIUS= specification is greater than the MAXPOINTS= value, the RADIUS= value is decremented enough to honor the MAXPOINTS= value, and the kriging system is solved for this grid point. If you specify the NODECREMENT option, no decrementing is done, prediction is skipped at this grid point, and a message is written to the log.

NOINCREMENT | NOINCR

requests that the RADIUS= value not be incremented when the MINPOINTS= value is not met at a grid point. This option is relevant only when you specify both a RADIUS= value and a MINPOINTS= number. In this case, when the number of points in the neighborhood constructed from the RADIUS= specification is less than the MINPOINTS= value, the RADIUS= value is incremented enough to honor the MINPOINTS= value, and the kriging system is solved for this grid point. If you specify the NOINCREMENT option, no incrementing is done, prediction is skipped at this grid point, and a message is written to the log.

NUMPOINTS=*number*

NPOINTS=*number*

NPTS=*number*

NP=*number*

specifies the exact size of a neighborhood. This option is incompatible with all other PREDICT statement options that control the neighborhood; it must appear by itself. In particular, if you specify both NUMPOINTS= and the RADIUS= option in the same PREDICT statement, then RADIUS= is honored, instead. In this event the value of the MINPOINTS= option is set to NUMPOINTS=, and the value of the MAXPOINTS= option is set to default, regardless of whether these options have been specified or not. If you specify any of the MINPOINTS= or MAXPOINTS= option without the RADIUS= option in the same PREDICT statement as NUMPOINTS=, then the NUMPOINTS= option is honored.

RADIUS=*number*

R=*number*

specifies the radius to use in a local kriging regression. When you specify this option, a separate kriging system is solved at each grid point by finding the neighborhood of this grid point that consists of all data points within the distance specified by the RADIUS= value. Thus, you can avoid unnecessary computational burden in your analysis by specifying the RADIUS= value to include data points situated

within the extent of your problem's spatial correlation. For additional control on the neighborhood, see the MAXPOINTS= and MINPOINTS= options. When you specify the RADIUS= and NUMPOINTS= options in the same PREDICT statement, then RADIUS= is honored.

VAR= *variable-name*
specifies the single numeric variable used in the kriging system.

MODEL Statement

MODEL *model-options* **;**

The MODEL statement specifies details about the correlation model that you use in the kriging system for prediction. The specified model is used in the kriging system defined by the most previous PREDICT statement. You can specify a semivariogram or covariance model in three ways:

- You specify the required parameters SCALE, RANGE, FORM, and SMOOTH (if you specify the MATERN form), and possibly the optional parameters NUGGET, ANGLE, and RATIO, explicitly in the MODEL statement.
- You specify an MDATA= data set. This data set contains variables that correspond to the required parameters SCALE, RANGE, FORM and SMOOTH (if you specify the MATERN form), and optionally variables for the NUGGET, ANGLE, and RATIO parameters.
- You can specify an input item store in the RESTORE statement. The item store contains one or more correlation models for one or more direction angles. You can specify these models in the STORESELECT option of the MODEL statement to perform a prediction task.

The three methods are mutually exclusive: you specify all parameters explicitly, they are all are read from the MDATA= data set, or you select a model and its parameters from an input item store.

Table 68.3 summarizes the *options* available in the MODEL statement.

Table 68.3 MODEL Statement Options

Option	Description
ANGLE=	Specifies the angle of the major axis
FORM=	Specifies the functional form (type)
MDATA=	Specifies the input data set containing parameter values
NUGGET=	Specifies the nugget effect for the model
POWNOBOUND	Allows values for the power model exponent parameter outside the range of $[0, 2)$
RANGE=	Specifies the range parameter
RATIO=	Specifies the ratio of the length of the minor axis
SCALE=	Specifies the scale parameter
SINGULAR=	Gives the singularity criterion for solving kriging systems
SMOOTH=	Specifies the smoothness parameter
STORESELECT	Uses the information from an input item store

You can use the following *model-options* with the MODEL statement:

ANGLE=*angle* | (*angle1, . . . , anglek*)
specifies the angle of the major axis for anisotropic models, measured in degrees clockwise from the N-S axis. The default is ANGLE=0.

In the case of a nested semivariogram model with k nestings, you have the following two ways to specify the anisotropy major axis: you can specify only one *angle* which is then applied to all nested forms, or you can specify one angle for each of the k nestings.

NOTE: The syntax makes it possible to specify different angles for different forms of the nested model, but this practice is rarely used.

FORM=*form* | (*form1, . . . , formk*)
specifies the functional form (type) of the semivariogram model. Use the syntax with the single *form* to specify a non-nested model. Use the syntax with forms *formi*, $i = 1, \ldots, k$, to specify a nested model with k structures. Each of the forms can be any of the following:

CUBIC | EXPONENTIAL | GAUSSIAN | MATERN |
PENTASPHERICAL | POWER | SINEHOLEEFFECT | SPHERICAL
CUB | EXP | GAU | MAT | PEN | POW | SHE | SPH
Usage examples follow.

For example, the syntax

```
FORM=GAU
```

specifies a model with a single Gaussian structure. Also, the syntax

```
FORM=(EXP,SHE,MAT)
```

specifies a nested model with an exponential, a sine hole effect, and a Matérn structure. Finally

```
FORM=(EXP,EXP)
```

specifies a nested model with two structures both of which are exponential.

NOTE: In the documentation, models are named either by using their full names or by using the first three letters of their structures. Also, the names of different structures in a nested model are separated by a hyphen (-). According to this convention, the previous examples illustrate how to specify a GAU, an EXP-SHE-MAT, and an EXP-EXP model, respectively, with the FORM= option.

All the supported model forms have two parameters specified by the SCALE= and RANGE= options, except for the MATERN model which has a third parameter specified by the SMOOTH= option. A FORM= value is required, unless you specify the MDATA= option or the STORESELECT option.

Computation of the MATERN covariance is numerically demanding. As a result, predictions that use Matérn covariance structures can be time-consuming.

See the section "Theoretical Semivariogram Models" on page 5031 for details about how the FORM= forms are determined.

MDATA=*SAS-data-set*

specifies the input data set that contains parameter values for the covariance or semivariogram model. The MDATA= option cannot be combined with any of the FORM= or STORESELECT options.

The MDATA= data set must contain variables named SCALE, RANGE, and FORM, and it can optionally contain variables NUGGET, ANGLE, and RATIO. If you specify the MATERN form, then you must also include a variable named SMOOTH in the MDATA= data set.

The FORM variable must be a character variable, and it can assume only the values allowed in the explicit FORM= syntax described previously. The RANGE, SCALE and SMOOTH variables must be numeric. The optional variables ANGLE, RATIO, and NUGGET must also be numeric if present.

The number of observations present in the MDATA= data set corresponds to the level of nesting of the covariance or semivariogram model. For example, to specify a non-nested model that uses a spherical covariance, an MDATA= data set might be given by the following statement:

```
data md1;
   input scale range form $;
   datalines;
   25 10 SPH
run;
```

The PROC KRIGE2D statement to use the MDATA= specification is of the form shown in the following:

```
proc krige2d data=...;
   predict var=....;
   model mdata=md1;
run;
```

This is equivalent to the following explicit specification of the covariance model parameters:

```
proc krige2d data=...;
   predict var=....;
   model scale=25 range=10 form=sph;
run;
```

The following MDATA= data set is an example of an anisotropic nested model:

```
data md1;
   input scale range form $ nugget angle ratio;
   datalines;
   20 8 SPH  5 35 0.7   .
   12 3 MAT  5 0  0.8 2.8
   4  1 GAU  5 45 0.5   .
;
```

This is equivalent to the following explicit specification of the covariance model parameters:

```
proc krige2d data=...;
   predict var=....;
   model scale=(20,12,4) range=(8,3,1) form=(SPH,MAT,GAU)
         angle=(35,0,45) ratio=(0.7,0.8,0.5) nugget=5 smooth=2.8;
run;
```

This example is somewhat artificial in that it is usually hard to detect different anisotropy directions and ratios for different nestings by using an empirical semivariogram. **NOTE:** The NUGGET variable value is the same for all nestings. This is always the case; the nugget effect is a single additive term for all models. For further details, see the section "The Nugget Effect" on page 5038.

The example also shows that if you specify a MATERN form in the nested model, then the SMOOTH variable must be specified for all nestings in the MDATA= data set. You simply specify the SMOOTH value as missing for nestings other than MATERN.

NUGGET=*number*

specifies the nugget effect for the model. The nugget effect is due to a discontinuity in the semivariogram as determined by plotting the sample semivariogram. For details, see the section "The Nugget Effect" on page 5038 and Chapter 124, "The VARIOGRAM Procedure." For models without any nugget effect, this option is left out; the default is NUGGET=0.

POWNOBOUND

specifies that values for the power model exponent parameter outside the range of $[0, 2)$ be allowed. The POWNOBOUND option applies only when you specify a power form in the MODEL statement.

Power models yield permissible covariance models only when the exponent parameter is nonnegative and less than 2. By default, PROC KRIGE2D produces an error if you specify a negative power exponent or one that is equal to or larger than 2 in the RANGE= option of the MODEL statement.

See the section "The Power Semivariogram Model" on page 5036 for more details about the power model form and its exponent parameter.

RANGE=*range* | (*range1, . . . , rangek*)

specifies the range parameter in semivariogram models. If you have anisotropy, you must specify the range of the major anisotropy axis, or the range of the minor anisotropy axis for any zonal components. In the case of a nested semivariogram model with k nestings, you must specify a range for each nested structure.

The range parameter has units of distance, and it is related to the correlation scale for the underlying spatial process.

NOTE: If you specify this parameter for a power model, then it does not correspond to a range. For power models, the parameter you specify in the RANGE option is a dimensionless power exponent whose value must range within [0,2) so that the power model is a valid semivariance function. See also the POWNOBOUND option of the MODEL statement.

See the section "Theoretical Semivariogram Models" on page 5031 for details about how the RANGE= values are determined.

RATIO=*ratio* | (*ratio1, . . . , ratiok*)

specifies the ratio of the length of the minor axis to the length of the major axis for anisotropic models. The value of the RATIO= option must be between 0 and 1. An exception is the case of zonal anisotropy, where the ratio of zonal components must be designated by a very large number for the RATIO= option. For further details, see the section "Zonal Anisotropy" on page 5043.

In the case of a nested semivariogram model with k nestings, you can specify a ratio for each nesting. The default is RATIO=1.

SCALE=*scale* | (*scale1, . . . , scalek*)

specifies the scale parameter in semivariogram models. In the case of a nested semivariogram model with k nestings, you must specify a scale for each nesting.

The scale parameter is the multiplicative factor in all supported models; it has the same units as the variance of the VAR= variable in the preceding PREDICT statement.

In power models the SCALE= parameter does not correspond to a sill because the power model has no sill. Instead, PROC KRIGE2D uses the SCALE= option to designate the slope (or scaling factor) in power model forms. The power model slope has the same variance units as the VAR= variable.

See the section "Theoretical Semivariogram Models" on page 5031 for details about how the SCALE= values are determined.

SINGULAR=*number*

gives the singularity criterion for solving kriging systems. The larger the value of the SINGULAR= option, the easier it is for a kriging system to be declared singular. The default is SINGULAR=1E–7. See the section "Ordinary Kriging" on page 5048 for more detailed information.

SMOOTH=*smooth* | (*smooth1, . . . , smoothm*)

specifies the smoothness parameter $\nu > 0$ in the Matérn type of semivariance structures. The special case $\nu = 0.5$ is equivalent to the exponential model, whereas $\nu \to \infty$ gives the Gaussian model.

When you specify m different MATERN forms in the FORM= option, you must also provide m smoothness values in the SMOOTH option. If you must specify more than one smoothness value, the values are assigned sequentially to the MATERN nestings in the order the nestings are specified. If you specify more smoothness values than necessary, then values in excess are ignored.

STORESELECT(*ssel-options***)**

SSEL(*ssel-options***)**

specifies that information from an input item store be used for the prediction. You cannot combine the STORESELECT option with any of the FORM= or MDATA= options. The STORESELECT option has the following *ssel-options*:

TYPE=*field-type*

specifies whether to perform isotropic or anisotropic prediction. You can choose the *field-type* from one of the following:

ISO

specifies an isotropic field for the prediction.

ANIGEO | GEO
specifies a field with geometric anisotropy for the prediction.

ANIZON(*zonal-form1, . . . , zonal-formn***)**
ZON(*zonal-form1, . . . , zonal-formn***)**
specifies a field with zonal anisotropy for the prediction. Each *zonal-formi*, $i = 1, \ldots, n$, can be any of the following:

CUB | EXP | GAU | MAT | PEN | POW | SHE | SPH
Each *zonal-formi*, $i = 1, \ldots, n$, is a structure in the purely zonal component of the correlation model in the direction angle of the minor anisotropy axis. For this reason, when you specify the TYPE=ANIZON suboption you must also specify the nonzonal component of the correlation model in the MODEL= suboption of the STORESELECT option.

Assume the nonzonal component has k structures; these are common across all directions and each one has the same scale in all directions. In that sense, you use the TYPE=ANIZON suboption to specify only the n zonal anisotropy structures of an input store $(k + n)$-structure nested model in the direction angle of the minor anisotropy axis.

Given this specification, $k+n$ must be up to the maximum number of nested model structures that is supported by the item store. See also the MODEL= suboption of the STORESELECT option.

In conclusion, you can use an input item store for prediction with zonal anisotropy if you know that every structure in the nonzonal model component has the same scale across all directions. When this condition does not apply for the item store models, specify the model parameters explicitly in the MODEL statement. For more details, see the examples in the section "Zonal Anisotropy" on page 5043.

Computation of the MATERN covariance is numerically demanding. As a result, predictions that use Matérn covariance structures can be time-consuming.

If you omit the TYPE= option, the default behavior is TYPE=ISO when the input item store contains information for only one angle or for the omnidirectional case. If you specify an item store with information for more than one direction, then the default behavior is TYPE=ANIGEO.

When you specify TYPE=ISO to request isotropic analysis in the presence of an item store with information for multiple directions, you must specify the ANGLEID= suboption of the STORESELECT option with one argument. This argument specifies which of the direction angles information to use for the isotropic analysis.

When you indicate the presence of anisotropy with the TYPE=ANIGEO or TYPE=ANIZON suboptions of the STORESELECT option, the following conditions apply:

- You must specify the ANGLEID= suboption of the STORESELECT option to designate the major and minor anisotropy axes. See the ANGLEID= suboption of the STORESELECT option for details.
- – For TYPE=ANIGEO, ensure that you have the same scale in all anisotropy directions.
  – For TYPE=ANIZON, ensure that the nonzonal component scale is the same in all anisotropy directions.

  If you import a nested model, these rules also apply to each one of the nested structures.

- Model ranges in the major anisotropy axis must be longer than ranges in the minor anisotropy axis.
- Any Matérn covariance structure must maintain its smoothness parameter value in all anisotropy directions.

ANGLEID=*angleid1* | (*angleid1, angleid2*)

specifies which direction angles in the input item store be used for prediction. The angles are identified by the corresponding number in the AngleID column of the "Store Models Information" table, or by the AngleID parameter in the table title when you specify the INFO(DETAILS) option in the RESTORE statement.

If you request isotropic prediction in the TYPE= suboption of the STORESELECT option and the item store has omnidirectional contents or information about only one angle, then the ANGLEID= option is ignored. The prediction input comes from the omnidirectional information. In the case of a single angle, you still perform isotropic prediction and the model parameters are provided by the model in the single direction angle in the item store. However, if the item store contains information for more than one angle, then you must specify one angle ID in *angleid1*. The model information from the corresponding angle is then used in your isotropic prediction.

When you specify an anisotropic prediction in the TYPE= option of the STORESELECT option, you need to have information about two perpendicular direction angles. One of them is the major and the other is the minor anisotropy axis. You must always specify the major anisotropy axis angle ID in *angleid1* and the minor anisotropy axis angle ID in *angleid2*. This means that the range parameters of the model forms in the angle designated by the *angleid1* need to be larger than the corresponding ranges of the forms in the angle designated by the *angleid2*. Conveniently, if the item store has only two angles, then you only need to specify the ID *angleid1* of the major anisotropy axis angle. If the item store has only one angle, then you cannot perform anisotropic prediction with input from the item store.

NOTE: You can perform geometric anisotropic analysis even if the item store does not contain information about a direction that is perpendicular to the one specified by *angleid1*. This is possible due to the geometry of the ellipse. In particular, when you specify the major axis with *angleid1* and an angle ID for a second direction with a corresponding smaller range, then PROC KRIGE2D automatically computes the minor anisotropy axis range and the necessary range ratio parameter.

Anisotropic analysis is not possible when you specify instances of the same angle in the input item store. It is possible that PROC VARIOGRAM produces an item store in which two or more directions can be the same if their corresponding correlation models were obtained for different angle tolerances or bandwidths in the VARIOGRAM procedure. Consequently, you cannot specify anisotropic prediction if the input store contains only two angles that are the same or if you specify *angleid1* and *angleid2* that correspond to equal angles.

MODEL=*form* | (*form1, . . . , formk*)

specifies the theoretical semivariogram model selection to use for the prediction. Use any combination of one, two, or three forms to describe a model in the input item store because up to three nested structures are supported. Each *formi*, $i = 1, \ldots, k$, can be any of the following:

CUB | EXP | GAU | MAT | PEN | POW | SHE | SPH

Computation of the MATERN covariance is numerically demanding. As a result, predictions that use Matérn covariance structures can be time-consuming.

All fitted models that are stored in the input item store contain information about their component parameters and also about the nugget effect if any. The KRIGE2D procedure retrieves this information when you make a model selection in the MODEL= option, and you do not need to individually specify a nugget effect or any other parameter of the model.

By default, the model that is ranked first among the models for a given angle in the item store is used for the prediction task. If more than one model is available in the item store, then you can specify the MODEL= option to use a different model for the prediction.

In an anisotropic prediction, the default selection is the model that is ranked first in the direction angle of the major anisotropy axis. If you specify the TYPE=ANIGEO option, then a model that consists of identical structures needs to be present in the selected minor anisotropy axis angle in the item store. If you specify the TYPE=ANIZON option, then a model with the exact same first k structures must be present in the selected minor anisotropy axis angle, and it must feature at least one more structure as a zonal component. The zonal component is specified separately in the TYPE=ANIZON suboption of the STORESELECT option. Consequently, remember that in zonal anisotropy the MODEL= suboption designates only the nonzonal component of the correlation model in the minor anisotropy axis direction. In all, if there are k common structures and n structures in the purely zonal component, then $k + n$ must be up to the maximum number of nested model structures that is supported by the item store.

In comparison to the other two ways of specifying a correlation model in PROC KRIGE2D, the STORESELECT option is quite different because you can avoid explicit specification of all parameter values of a model. When you specify the STORESELECT option, then the corresponding scale, range, nugget effect, and smoothness (if appropriate) parameter values are invoked as saved attributes of the model that you select from the item store.

In the case of anisotropy, you specify the angles indirectly with the ANGLEID= option of the STORESELECT option, and the ratios are computed implicitly by using the selected model ranges. Explore how to specify valid anisotropical models imported from an input item store with the two examples that follow.

In the first example, assume the input item store InStoreGeo contains exponential models in the angles $\theta_1 = 0°$, $\theta_2 = 45°$, and $\theta_3 = 90°$. You know in advance that all models have the same scale $c_1 = c_2 = c_3$ across these directions and that the respective ranges are $a_1 = 15$, $a_2 = 20$, and $a_3 = 25$ in distance units. Hence, you have a case of geometric anisotropy where the major anisotropy axis is in the direction of angle θ_3 and the minor anisotropy axis is in the direction of angle θ_1. The following statements in PROC KRIGE2D use the information in the item store InStoreGeo to perform simulation under the assumption of geometric anisotropy:

```
proc krige2d data=...;
   restore in=InStoreGeo;
   predict var=....;
   model storeselect(model=exp type=anigeo angleid=(3,1));
run;
```

For the second example, assume a case of zonal anisotropy. Consider the input item store InStoreZon, which contains models in the two angles, $\theta_1 = 30°$ and $\theta_2 = 120°$. Specifically, in θ_1 you have an exponential-spherical model: the exponential structure has scale $c_{1E} = 3$ and range $a_{1E} = 10$; the spherical structure has scale $c_{1S} = 1$ and range $a_{1S} = 6$. In direction θ_2 you have an exponential model with scale $c_{1E} = 3$ and range $a_{1E} = 12$. Hence, the zonal anisotropy major axis is in the direction of the lowest total variance, which is in angle θ_2; then, the minor axis is in the direction of angle θ_1. The following statements in PROC KRIGE2D use the information in the store InStoreZon to perform prediction under the assumption of zonal anisotropy:

```
proc krige2d data=...;
   restore in=InStoreZon;
   predict var=....;
   model storeselect(model=exp type=anizon(sph) angleid=(2,1));
run;
```

RESTORE Statement

RESTORE IN=*store-name* < / *option* > **;**

The RESTORE statement specifies an item store that provides spatial correlation model input for the PROC KRIGE2D prediction tasks. An item store is a binary file defined by the SAS System. You cannot modify the contents of an item store. The KRIGE2D procedure can use only item stores created by PROC VARIOGRAM.

Item stores enable you to use saved correlation models without having to repeat specification of these models in the MODEL statement. In principle, an item store contains the chosen model from a model fitting process in PROC VARIOGRAM. If more than one model form is fitted, then all successful fits are included in the item store. In this case, you can choose any of the available models to use for prediction with the STORESELECT(MODEL=) option in the MODEL statement. Successfully fitted models might include questionable fits, which are so flagged when you specify the INFO option to display model names.

The *store-name* is a usual one- or two-level SAS name, as for SAS data sets. If you specify a one-level name, then the item store resides in the WORK library and is deleted at the end of the SAS session. Since item stores are often used for postprocessing tasks, typical usage specifies a two-level name of the form *libname.membername*.

When you specify the RESTORE statement, the default output contains some general information about the input item store. This information includes the store name, label (if assigned), the data set that was used to create the store, BY group information, the procedure that created the store, and the creation date.

You can specify the following *option* in the RESTORE statement after a slash (/):

INFO < (*info-options*) >
specifies that additional information about the input item store be printed. This information is provided in two ODS tables. One table displays the variables in the item store, in addition to the mean and standard deviation for each of them. These statistics are based on the observations that were used to produce the store results. The second table shows the model on top of the list of all fitted models for each direction angle in the item store. The INFO option has the following *info-options*:

DETAILS
DET
specifies that more detailed information be displayed about the input item store. This option produces the full list of models for each direction angle in the item store, in addition to the model equivalence class. For more information about classes of equivalence, see the section "Classes of Equivalence" on page 10099 in Chapter 124, "The VARIOGRAM Procedure." The DETAILS option is ignored if the input item store contains information about a single fitted model.

ONLY
specifies that only information about the input item store without any prediction tasks be displayed.

When you specify an input item store with the RESTORE statement in PROC KRIGE2D, all the DATA= input data set variables must match input item store variables. If there are BY groups in the input DATA= set or in the input RESTORE variables, then PROC KRIGE2D handles the different cases as follows:

- If both PROC KRIGE2D has BY groups and the RESTORE statement has BY groups, then the analysis variables must match. This matching assumes implicitly that in each BY group of PROC KRIGE2D and the item store, the corresponding set of observations and correlation model comes from the same random field. This assumption is valid if you use the same data set, first in PROC VARIOGRAM to fit a model and save it in the item store, and then in PROC KRIGE2D to perform predictions with the resulting correlation models.
- If PROC KRIGE2D has BY groups but the item store does not, then the item store is accepted only if the procedure and the item store analysis variables match. In this case, the same item store model choice iterates across the BY groups of the input data. You are advised to proceed with caution: each BY group in the input DATA= set corresponds to a different realization of a random field. Hence, by using the same correlation model for prediction purposes, you implicitly assume that all these different realizations are instances of the same random field.
- If PROC KRIGE2D has no BY groups but the item store does, then the item store is rejected.

Details: KRIGE2D Procedure

Theoretical Semivariogram Models

Consider a stochastic spatial process represented by the stationary spatial random field (SRF) $\{Z(\boldsymbol{s}), \boldsymbol{s} \in D \subset \mathcal{R}^2\}$ (Christakos 1992). The VARIOGRAM procedure computes the empirical (also known as sample or experimental) semivariance of $Z(\boldsymbol{s})$. Prediction of the spatial process $Z(\boldsymbol{s})$ at unsampled locations by techniques such as ordinary kriging requires a theoretical semivariogram or covariance.

When you use PROC VARIOGRAM and PROC KRIGE2D to perform spatial prediction, you must determine a suitable theoretical semivariogram based on the sample semivariogram. Various methods exist to fit semivariogram models, such as least squares, maximum likelihood, and robust methods (Cressie 1993, section 2.6). You can use PROC VARIOGRAM to perform automated fitting of a semivariogram model with weighted or ordinary least squares. A different approach is manual fitting, in which a theoretical semivariogram is chosen based on visual inspection of the empirical estimate; see, for example, Hohn (1988, p. 25).

In some cases, a plot of the experimental semivariogram suggests that a single theoretical model is inadequate. Nested models, anisotropic models, and the nugget effect increase the scope of theoretical models available. All of these concepts are discussed in this section. The specification of the final theoretical model is provided by the syntax of PROC KRIGE2D.

Figure 68.5 shows the general flow of investigation. The empirical semivariogram is computed after a suitable choice is made for the LAGDISTANCE= and MAXLAGS= options in PROC VARIOGRAM, and possibly the NDIR= option or the DIRECTIONS statement for computations in more than one directions. Potential theoretical models (which can also incorporate nesting, anisotropy, and the nugget effect) are then plotted against the empirical semivariogram and evaluated. A suitable theoretical model is found by using the methodology presented in the section "Examples: VARIOGRAM Procedure" on page 10112 in Chapter 124, "The VARIOGRAM Procedure."

Eight theoretical models are supported by PROC KRIGE2D: the Gaussian, exponential, Matérn, spherical, cubic, pentaspherical, sine hole effect and power models. See also the section "Theoretical Semivariogram Models" on page 10072 in Chapter 124, "The VARIOGRAM Procedure." These eight model forms are now examined in more detail: the Gaussian, exponential, and Matérn forms are examined as one group; the spherical, cubic, and pentaspherical as a second group; and the remaining power and sine hole effect models are examined individually. For comparison purposes, the axes in the forms' illustrations are kept the same across the plots, and the corresponding parameters of the different forms have the same values.

In PROC KRIGE2D the parameters a_0 and c_0 for all forms correspond to the RANGE= and SCALE= options, respectively, in the MODEL statement. For all model forms, the dimension of c_0 is the same as the dimension of the variance of the spatial process $Z(\boldsymbol{s})$. For all forms but the power model, the dimension of a_0 is length with same units as the distance h in the semivariance $\gamma_z(h)$. See the section "The Power Semivariogram Model" on page 5036 for more details about interpretation of the power model a_0 parameter.

Figure 68.5 Flowchart for Semivariogram Selection

The Gaussian Semivariogram Model

The form of the Gaussian model is

$$\gamma_z(h) = c_0 \left[1 - \exp\left(-\frac{h^2}{a_0^2} \right) \right]$$

The shape is displayed in Figure 68.6, using range $a_0 = 1$ and scale $c_0 = 4$.

The vertical line at $h = r_\epsilon = \sqrt{3}a_0$ shows the *effective* (or *practical*) *range* as defined by Deutsch and Journel (1992) or the *range* ϵ defined by Christakos (1992). The effective range is the h-value where the covariance is approximately 5% of its value at zero. Alternatively, the stationarity assumption implies that the effective range is the h value where the semivariance is approximately 5% of the sill value, as shown in Figure 68.6.

In the Gaussian model the semivariance $\gamma_z(h)$ approaches the sill asymptotically at c_0.

The Exponential Semivariogram Model

The form of the exponential model is

$$\gamma_z(h) = c_0 \left[1 - \exp\left(-\frac{h}{a_0} \right) \right]$$

The shape is displayed in Figure 68.6, using range $a_0 = 1$ and scale $c_0 = 4$.

The vertical line at $h = r_\epsilon = 3a_0$ is the *effective* (or *practical*) *range* or the *range* ϵ (that is, the h-value where the covariance is approximately 5% of its value at zero).

As in the Gaussian model, the sill in this example is at 4.0 variance units (corresponding to $c_0 = 4$) and is approached asymptotically.

The major distinguishing feature of the Gaussian and exponential forms is the shape in the neighborhood of the origin $h = 0$, as Figure 68.6 illustrates. In general, small lags are important in determining an appropriate theoretical form based on an empirical semivariogram.

The Matérn Semivariogram Model

The form of the Matérn model is

$$\gamma_z(h) = c_0 \left[1 - \frac{2}{\Gamma(\nu)} \left(\frac{h\sqrt{\nu}}{a_0} \right)^{\nu} K_\nu \left(2\frac{h\sqrt{\nu}}{a_0} \right) \right]$$

where $\nu > 0$ is the smoothness factor parameter. Figure 68.6 shows an example of the Matérn form, where range $a_0 = 1$, scale $c_0 = 4$, and $\nu = 1.5$.

The Matérn semivariance $\gamma_z(h)$ is a class of semivariance models that emerge for different values of the smoothing parameter ν. The Matérn form reaches its sill value c_0 asymptotically.

The Gaussian and exponential semivariances are two frequently used members of the Matérn class of semivariances. In particular, the exponential semivariance model is derived from the Matérn class of models for $\nu = 0.5$. Also, when $\nu \to \infty$ then the Matérn semivariance gives the Gaussian model. In Figure 68.6 the selected value of $\nu = 1.5$ places the Matérn form in between the Gaussian and the exponential. The Matérn semivariance typically begins to look and behave as the Gaussian for values of $\nu > 10$.

Figure 68.6 Gaussian, Exponential, and Matérn Semivariograms with Parameters $a_0 = 1$, $c_0 = 4$, and $\nu = 1.5$

The Spherical Semivariogram Model

The form of the spherical model is

$$\gamma_z(h) = \begin{cases} c_0 \left[\frac{3}{2}\frac{h}{a_0} - \frac{1}{2}(\frac{h}{a_0})^3 \right] & \text{for } h \le a_0 \\ c_0 & \text{for } h > a_0 \end{cases}$$

The shape is displayed in Figure 68.7, using range $a_0 = 1$ and scale $c_0 = 4$.

The vertical line at $h = 1$ shows the range a_0 of the model.

In the case of the spherical model, $\gamma_z(h)$ actually reaches the sill value at c_0, unlike the Gaussian and exponential types where the sill is a horizontal asymptote.

The Cubic Semivariogram Model

The form of the cubic model is

$$\gamma_z(h) = \begin{cases} c_0 \left[7(\frac{h}{a_0})^2 - \frac{35}{4}(\frac{h}{a_0})^3 + \frac{7}{2}(\frac{h}{a_0})^5 - \frac{3}{4}(\frac{h}{a_0})^7 \right] & \text{for } h \le a_0 \\ c_0 & \text{for } h > a_0 \end{cases}$$

The cubic form shape is displayed in Figure 68.7, using range $a_0 = 1$ and scale $c_0 = 4$.

The vertical line at $h = 1$ shows the range a_0 of the model.

Similarly to the spherical model, the cubic model, $\gamma_z(h)$ reaches the sill value at c_0 and maintains this value after a distance h equal to the model range.

The Pentaspherical Semivariogram Model

The form of the pentaspherical model is

$$\gamma_z(h) = \begin{cases} c_0 \left[\frac{15}{8}\frac{h}{a_0} - \frac{5}{4}(\frac{h}{a_0})^3 + \frac{3}{8}(\frac{h}{a_0})^5 \right] & \text{for } h \le a_0 \\ c_0 & \text{for } h > a_0 \end{cases}$$

The pentaspherical form shape is displayed in Figure 68.7, using range $a_0 = 1$ and scale $c_0 = 4$.

The vertical line at $h = 1$ shows the range a_0 of the model.

The pentaspherical semivariance behaves like the spherical and cubic semivariances, in that $\gamma_z(h)$ increases with distance until it reaches the sill value c_0 at the distance h equal to the model range a_0.

Figure 68.7 accents the differences in the behavior of the featured semivariances. Specifically, the cubic and pentaspherical forms reach the sill value faster than the spherical form. Also, the spherical and pentaspherical forms exhibit a more linear behavior at distances close to the origin $h = 0$.

Figure 68.7 Spherical, Cubic, and Pentaspherical Semivariograms with Parameters $a_0 = 1$ and $c_0 = 4$

The Sine Hole Effect Semivariogram Model

The form of the sine hole effect model is

$$\gamma_z(h) = c_0 \left[1 - \frac{\sin(\pi h/a_0)}{\pi h/a_0} \right]$$

Figure 68.8 shows an example of the sine hole effect form, where range $a_0 = 1$ and scale $c_0 = 4$.

The vertical line at $h = 1$ shows the range a_0 of the model.

The sine hole effect semivariance $\gamma_z(h)$ increases with distance. It has the distinct characteristic that it reaches the sill at a distance $h = a_0$ equal to the model range and then it oscillates around the sill value with a decreasing amplitude as it moves to higher values of h.

The Power Semivariogram Model

The form of the power model is

$$\gamma_z(h) = c_0 h^{a_0}$$

For this model, the parameter a_0 is known as the power exponent. This is a dimensionless quantity which must range within $0 \leq a_0 < 2$ so that the power model is a permissible semivariance model.

The KRIGE2D procedure enables you to specify power exponent values that are outside this range when you also explicitly specify the POWNOBOUND option in the MODEL statement. However, parameter values equal to or greater than 2 can result in singular covariance matrices or negative prediction errors.

For the special case of $a_0 = 1$ the form yields a straight line. In this case the power model reduces to the linear model. The parameter c_0 designates the slope of the power form and has dimensions of the variance as in the other models.

The power model has no sill; this differentiates it from the rest of the models presented earlier. Spatial correlation that is described by a power model indicates that the stochastic process variance increases constantly with distance. The shape of the power model with $a_0 = 0.4$ and $c_0 = 4$ is displayed in Figure 68.8.

Figure 68.8 Sine Hole Effect Semivariogram with Range $a_0 = 1$ and Scale $c_0 = 4$, and Power Semivariogram with Exponent $a_0 = 0.4$ and Slope $c_0 = 4$

For comparison purposes, Figure 68.9 displays all eight semivariance forms that you can use with PROC KRIGE2D. The figure displays a composition of the different forms with the parameter values selected earlier throughout this section. Depending on the empirical semivariogram, these models provide you with flexibility to select an appropriate theoretical semivariance model for prediction.

Figure 68.9 Semivariogram Forms Used in PROC KRIGE2D

Nested Models

For a given set of spatial data, a plot of an experimental semivariogram might not seem to fit any of the individual theoretical models. In such a case, you might obtain a more accurate fit if you consider your covariance model to be the sum of two or more covariance structures. Such covariance models are called *nested* models. Nesting is common in geologic applications where correlations can exist at different length scales. At small lag distances h, the smaller scale correlations dominate, while the large scale correlations dominate at larger lag distances.

Nested models are permissible covariances if they are the sum of permissible models. Therefore, you can include in a sum any combination of the models presented in the preceding subsections and produce permissible covariance models. As an illustration, consider two semivariogram models: an exponential and a spherical,

$$\gamma_{z,1}(h) = c_{0,1} \exp(-\frac{h}{a_{0,1}})$$

and

$$\gamma_{z,2}(h) = \left\{ \begin{array}{ll} c_{0,2}\left[\frac{3}{2}\frac{h}{a_{0,2}} - \frac{1}{2}(\frac{h}{a_{0,2}})^3\right], & \text{for } h \le a_{0,2} \\ c_{0,2}, & \text{for } h > a_{0,2} \end{array} \right\}$$

with $c_{0,1} = 1, a_{0,1} = 2.5, c_{0,2} = 2$, and $a_{0,2} = 1$. If both of these correlation structures are present in a spatial process $\{Z(\boldsymbol{s}), \boldsymbol{s} \in D\}$, then the semivariance $\gamma_z(h)$ of this process can be expressed as

$$\gamma_z(h) = \gamma_{z,1}(h) + \gamma_{z,2}(h)$$

In this case, the experimental semivariogram $\gamma_z(h)$ for the process $Z(\boldsymbol{s})$ resembles the semivariogram of the sum of $\gamma_{z,1}(h)$ and $\gamma_{z,2}(h)$. This is illustrated in Figure 68.10.

The sum of $\gamma_{z,1}(h)$ and $\gamma_{z,2}(h)$ in Figure 68.10 does not resemble any *single* theoretical semivariogram; however, its shape at $h = 1$ is similar to a spherical form. The asymptotic approach to a sill at three variance units, along with the shape around $h = 0$, indicates an exponential structure. The sill value c_0 of the sum is the sum of the individual sills $c_{0,1} = 1$ and $c_{0,2} = 2$. In general, a nested model has a sill equal to the sum of the sills of its nested structures plus the nugget effect, if present.

See Hohn (1988, p. 38ff) for further examples of nested correlation structures.

Figure 68.10 Sum of Exponential and Spherical Structures at Different Scales

The Nugget Effect

For all the semivariogram models considered previously, the following property holds:

$$\gamma_z(0) = \lim_{h \downarrow 0} \gamma_z(h) = 0$$

However, a plot of the experimental semivariogram might indicate a discontinuity at $h = 0$; that is, $\gamma_z(h) \rightarrow c_n > 0$ as $h \rightarrow 0$, while $\gamma_z(0) = 0$. The quantity c_n is called the *nugget effect*; this term is from mining

geostatistics where nuggets literally exist, and it represents variations at a much smaller scale than any of the measured pairwise distances—that is, at distances $h \ll h_{min}$, where

$$h_{min} = \min_{i,j} h_{ij} = \min_{i,j} | \boldsymbol{s}_i - \boldsymbol{s}_j |$$

Nonzero nugget effects have been associated with conceptual and theoretical difficulties; see Cressie (1993, section 2.3.1) and Christakos (1992, section 7.4.3) for details. There is no *practical* difficulty, however; you simply visually extrapolate the experimental semivariogram as $h \to 0$. The importance of availability of data at small lag distances is again illustrated.

As an example, an exponential semivariogram with a nugget effect c_n has the form

$$\gamma_z(h) = c_n + {\sigma_0}^2 \left[1 - \exp\left(-\frac{h}{a_0} \right) \right], h > 0$$

and

$$\gamma_z(0) = 0$$

where the factor ${\sigma_0}^2$ is called the *partial sill* and the sill $c_0 = c_n + {\sigma_0}^2$.

This is illustrated in Figure 68.11 for the parameters $a_0 = 1$, ${\sigma_0}^2 = 4$, and nugget effect $c_n = 1.5$.

You can specify the nugget effect in PROC KRIGE2D with the NUGGET= option in the MODEL statement. It is a separate, additive term independent of direction; that is, it is isotropic. The way to approximate an anisotropic nugget effect is described in the following section.

Figure 68.11 Exponential Semivariogram Model with a Nugget Effect $c_n = 1.5$

Anisotropic Models

In all of the theoretical models considered previously, the lag distance h is entered as a scalar value. This implies that the correlation between the spatial process at two point pairs P_1, P_2 is dependent *only* on the separation distance $h = \mid P_1 P_2 \mid$, not on the orientation of the vector $\boldsymbol{h}$. A spatial process described by an SRF $\{Z(\boldsymbol{s}), \boldsymbol{s} \in D \subset \mathcal{R}^2\}$ with this property is called isotropic, as is the associated covariance or semivariogram.

However, real spatial phenomena often show directional effects. Particularly in geologic applications, measurements along a particular direction might be highly correlated, while typically the perpendicular direction shows little or no correlation. Such processes are called anisotropic; see, for example, Journel and Huijbregts (1978, section III.B.4).

When the correlation structure varies across different directions, you need different models for each direction so that you can account correctly for the continuity within the SRF. The following subsections describe how techniques are applied to override the anisotropy effects for computational purposes. First, characteristics of anisotropy are examined.

The semivariogram sill is a measure of the process variability; hence the direction of the highest continuity is perpendicular to the direction where the highest sill occurs. If the sill is the same in all directions, then the direction with the highest range indicates highest continuity. The directions in which the spatial process $\{Z(\boldsymbol{s}), \boldsymbol{s} \in D\}$ is most and least correlated are called the *major* and *minor* axis of anisotropy, respectively.

In some cases, these directions are known a priori. This can occur in mining applications where the geology of a region is known in advance. In most cases however, nothing is known about possible anisotropy. Depending on the amount of data available, using several directions is usually sufficient to determine the presence of anisotropy and to find the approximate major and minor axis directions; see the discussion in the section "Anisotropy" on page 10080 in Chapter 124, "The VARIOGRAM Procedure," documentation. You can find a detailed example of anisotropy investigation in the section "Example 124.2: An Anisotropic Case Study with Surface Trend in the Data" on page 10122 in Chapter 124, "The VARIOGRAM Procedure," documentation.

After you explore an anisotropic process and you identify the minor and major axis directions, you can compute the *anisotropy factor* parameter R which is defined as

$$R = \frac{{a_0}^{min}}{{a_0}^{max}}$$

where ${a_0}^{min}$ is the semivariogram range in the direction of the minor axis and ${a_0}^{max}$ is the semivariogram range in the direction of the major axis.

There are two types of anisotropy, depending on which semivariogram characteristics change in different directions. These types are the *geometric* and the *zonal* anisotropy, and either or both can be present. Both are examined in detail in the following subsections.

Geometric Anisotropy

Geometric anisotropy is the simplest type of anisotropy. It occurs when the same sill (or scale) parameter c_0 is present in all directions but the range a_0 changes with direction. In geometric anisotropy the covariance model uses the same forms in all directions.

Therefore, geometric anisotropy features one single sill value, and depending on the direction the semivariogram reaches the sill within a different distance. This is illustrated in Figure 68.12, where an anisotropic

exponential semivariogram is plotted. Assume that the two curves displayed in this figure have the same sill $c_0 = 1.5$ and are generated using the ranges $a_{0,1} = 3$ in the direction $\theta_1 = 30°$ (effective range is $r_{\epsilon,1} = 9$) and $a_{0,2} = 1$ in the direction $\theta_2 = 120°$ (effective range is $r_{\epsilon,2} = 3$).

As you can see from the figure, the ratio of the shorter to longer range is $R = 1/3$. The anisotropy factor R is the value to use in the RATIO= parameter in the MODEL statement in PROC KRIGE2D. When you model geometric anisotropy $R \leq 1$. In fact, isotropy is a partial case of geometric anisotropy for which ${a_0}^{min} = {a_0}^{max}$ and $R = 1$.

Figure 68.12 Geometric Anisotropy with Major Axis in the Direction $\theta_1 = 30°$

The values of the RANGE= and ANGLE= parameters in the MODEL statement in PROC KRIGE2D are set based on the major anisotropy axis characteristics. Specifically, the RANGE= parameter is the value of the major axis range ${a_0}^{max} = a_{0,1}$, and the ANGLE= parameter is the angle θ_1 of the major axis measured clockwise from north (angles measured in this way are also known as *azimuths*). You can then specify the following MODEL statement in PROC KRIGE2D to approximate the covariance structure:

```
model form=exp range=3 scale=1.5 angle=30 ratio=0.3333;
```

If you use a nested model, provide the type for each one of the nested structures with the FORM= option, and assign the individual SCALE= parameters so that they add up to the total sill (include in the sum the nugget effect, if present). In the typical case, all of your nested structures have the same anisotropy axes. This means that you specify the same ANGLE= parameter value for all structures. Each structure likely has its own values for the RANGE= and RATIO= parameters depending on the degree of its contribution to the nested model.

The terminology associated with geometric anisotropy is that of ellipses. To see how this comes about, consider the following hypothetical set of calculations. Let $\{Z(\boldsymbol{s}), \boldsymbol{s} \in D \subset \mathcal{R}^2\}$ be a geometrically anisotropic process, and assume sufficient data points are present to calculate an experimental semivariogram at a large number of angle classes $\theta \in \{0, \delta\theta, 2\delta\theta, \ldots, 180°\}$. At each of these angles θ, the experimental semivariogram is plotted and the range a_0 is recorded. A diagram in polar coordinates (a_0, θ) yields an

ellipse with the major axis ${a_0}^{max}$ in the direction of the largest a_0 and the minor axis ${a_0}^{min}$ perpendicular to it. For the example in Figure 68.12, the ellipse is shown in Figure 68.13(a). Its major axis has size ${a_0}^{max}$ situated at angle θ_1 clockwise from north, and the minor axis has size ${a_0}^{min}$ oriented at angle θ_2 clockwise from north.

The KRIGE2D procedure handles geometric anisotropy by applying a reversible transformation in two steps that converts geometric anisotropy into isotropic conditions.

The first step is to align your coordinates axes with the anisotropy ellipse axes. Specifically, you choose to rotate by an angle φ the standard Cartesian orientation of the (x, y) coordinates system shown in Figure 68.13(a) so that the Y axis coincides with the ellipse minor axis. The rotation result is illustrated in Figure 68.13(b). The second step is to elongate the minor axis so its length equals that of the major axis of the ellipse. You can see the result in Figure 68.13(c). The computational details are shown in the following.

Figure 68.13 Transformation Applied to Geometric Anisotropy

The transformation angle φ is measured in standard Cartesian orientation counterclockwise from the X axis (east). If the major axis azimuth is θ_1, then the Cartesian system of (x, y) needs to be rotated by $\varphi = 90° - \theta_1$ so that the Y axis can coincide with the ellipse minor axis; see Figure 68.13(a).

Let us call the ellipse major axis X' and the minor axis Y'. The transformation that converts any coordinates in the (x, y) system into (x', y') coordinates in terms of φ is given by the matrix:

$$\mathbf{H} = \begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -\sin(\varphi) & \cos(\varphi) \end{pmatrix}$$

The elongation of the minor axis in the second step is performed with the matrix:

$$\mathbf{D}_R = \begin{pmatrix} 1 & 0 \\ 0 & 1/R \end{pmatrix}$$

NOTE: These two steps are sequential and their order cannot be reversed. For any point pair P_1 and P_2 with respective coordinates $\boldsymbol{s}_1 = (x_1, y_1)$ and $\boldsymbol{s}_2 = (x_2, y_2)$ in the (x, y) axes, their distance is given by

$$| \, P_i P_j \, |_{(x,y)} = h = \sqrt{(\delta x)^2 + (\delta y)^2}$$

where the distance components $\delta x = x_2 - x_1$ and $\delta y = y_2 - y_1$. Based on the previous, the corresponding distances $\delta x'$ and $\delta y'$ in the (x', y') coordinates system are given by the vector:

$$\begin{pmatrix} \delta x' \\ \delta y' \end{pmatrix} = \mathbf{D}_R \, \mathbf{H} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = \begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -\sin(\varphi)/R & \cos(\varphi)/R \end{pmatrix} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix}$$

The transformed interpair distance is then:

$$| \, P_i P_j \, |_{(x',y')} = h' = \sqrt{(\delta x')^2 + (\delta y')^2}$$

As a result, the original anisotropic semivariogram in Figure 68.12 that was a function $\gamma(\boldsymbol{h}) = \gamma(h, \theta)$ of both h and θ is then transformed to an equivalent function $\hat{\gamma}(h')$ only of h':

$$\hat{\gamma}(h') = \gamma(\boldsymbol{h})$$

This single isotropic semivariogram $\hat{\gamma}(h')$ is then used for kriging purposes.

The two steps used by PROC KRIGE2D in the previous analysis can also be performed in a different manner. For instance, you might equivalently choose to rotate the (x, y) Cartesian coordinates so that the Y axis coincides with the ellipse major axis, rather than with the minor axis as was shown earlier. Also, you might prefer to compress the major axis rather than elongating the short one. In any case, you need to perform the appropriate computations for the transformation of your choice.

Zonal Anisotropy

In zonal anisotropy, the sill (or scale) parameter c_0 is different for different directions. It is not possible to transform such a structure into an isotropic semivariogram. Instead, nesting and geometric anisotropy are used together to approximate zonal anisotropy.

When the scale varies with direction, the lowest scale (that is, the lowest variance) naturally corresponds to the maximum continuity direction. The same direction has the longest range, as also discussed in the section "Geometric Anisotropy" on page 5040.

A varying scale with direction can be interpreted as having one or more model components whose individual contributions to the total variance differ with direction. For each such component, its contribution (scale) ranges between zero and a maximum value. This makes it unlikely that you can describe a natural process with a pure zonal model, because doing so would imply zero continuity in the direction of zero contribution; see also Chilès and Delfiner (1999, p. 96).

In a simple case of zonal anisotropy, a model includes one zonal component. The zonal component makes its highest contribution in a direction perpendicular to the maximum continuity direction, and it contributes zero to the maximum continuity direction. This is necessary; otherwise, there would be a direction with a total scale less than the scale in the maximum continuity direction. Following a similar reasoning, the zonal component's direction of maximum contribution cannot coincide with the one of maximum continuity. In the general case, there can be multiple zonal components, each making its highest contribution in a different direction.

The following describes how to deal with zonal anisotropy in your analysis; see also Goovaerts (1997, p. 96) and Deutsch and Journel (1992, pp. 27–32). If you start with an empirical semivariogram, you can investigate zonal anisotropy by identifying whether a maximum and a minimum scale exist in two specific directions.

If they exist, typically these two directions might be perpendicular. Then proceed to identify the zonal component that causes the difference in scale by fitting the empirical semivariogram. You represent zonal component as an additional nested structure in the direction of maximum total scale.

If the minimum and maximum sills are not in perpendicular directions, then you might be seeing the combined effects of multiple zonal components in different directions. In that case you might be able to approximate the continuity behavior by assuming a single zonal component in the direction that is perpendicular to the one of maximum continuity. Alternatively, you might decide to investigate a more elaborate configuration for the model components. In this case, you need to maintain a geometrical anisotropy part across all directions and add zonal components in an appropriate way to match your empirical semivariance in different directions.

After you have a theoretical semivariance model with zonal anisotropy, the next step is to include zonal components in your prediction or simulation analysis. In PROC KRIGE2D you can specify zonal components either explicitly or with the use of results previously saved in item stores produced by the VARIOGRAM procedure.

Specifying a zonal component explicitly in the MODEL statement has the following implications:

- The RANGE= parameter for the zonal component refers to the range value in the direction of maximum zonal contribution, unlike the case of ranges specified for nonzonal components that refer to the direction of maximum continuity.
- The anisotropy factor *R* in the RATIO= parameter for the zonal component should be specified as a large positive value to designate zero contribution in the perpendicular direction.

To explain the previous point, remember that *R* is defined as $R = {a_0}^{min}/{a_0}^{max}$. Its value specifies how much to elongate the minor anisotropy axis to make it equal to the major anisotropy axis, in order to transform geometric anisotropy into isotropy. Intuitively, an infinite *R* value makes it impossible for the minor axis to become as large as the major axis. This is equivalent to having a very large major anisotropy axis; hence, it indicates a very large range across the major axis direction. Indeed, you can consider a zero zonal contribution in the major anisotropy axis as a very large range of the zonal component along this direction. The particular range is so large that the zonal component practically never reaches its scale along this direction, and this is interpreted as zero contribution.

In the case where you specify zonal anisotropy by using the contents of an item store, you only need to specify the geometric anisotropy components in the SSEL(MODEL=) option, and the zonal components as suboptions of the SSEL(TYPE=ANIZON) option. Then, the KRIGE2D or SIM2D procedure checks whether the item store contains models that are suitable to use, based on your specifications.

The following two examples illustrate different instances of zonal anisotropy and how to specify the corresponding covariance model parameters in PROC KRIGE2D.

Example 1

The first example shows that if you can model the direction with the highest sill as a nested model, then you can treat the case as a composition of geometric anisotropy and an additional structure that acts only in the direction of the increased sill.

Consider a spatial process in which the fitting of theoretical models in your experimental semivariogram produces a correlation structure like the one shown in Figure 68.14. In the direction $\theta_1 = 40°$, the covariance model has a single exponential structure $\gamma_1(\boldsymbol{h}) = \mathrm{Exp}(a_{0,1E}, c_{0,1E})$ with range $a_{0,1E} = 2$ and sill $c_{0,1E} = 1.5$. In the direction $\theta_2 = 130°$, the covariance model $\gamma_2(\boldsymbol{h}) = \mathrm{Exp}(a_{0,2E}, c_{0,2E}) + \mathrm{Sph}(a_{0,2S}, c_{0,2S})$ has

two nested structures: an exponential structure with range $a_{0,2E} = 0.5$ and sill $c_{0,2E} = 1.5$ and a spherical structure with range $a_{0,2S} = 1$ and sill $c_{0,2S} = 0.5$.

Figure 68.14 Zonal Anisotropy in Two Directions

The total sill in the direction θ_2 of highest variance is the sum of the nested structures' sills $c_{0,2E} + c_{0,2S} = 2$. You can consider that your process is characterized by a geometrically anisotropic exponential structure with common sill $c_{0,E} = 1.5$ across all directions and major axis range $a_{0,1E} = 2$, and by a spherical structure which is a zonal anisotropy component that contributes only in the θ_2 direction. Based on the remarks in this section, the RATIO= parameter for the exponential structure is $R_E = 0.5/2 = 0.25$, whereas for the spherical structure you choose a large value, such as $R_S = 10^8$.

Then, you can approximate this structure in PROC KRIGE2D by specifying the two structures with the following MODEL statement:

```
model form=(exp,sph) range=(2,1) scale=(1.5,0.5)
      angle=(40,130) ratio=(0.25,1e8);
```

You can handle more elaborate cases in a similar way, where the covariance models in different directions might all be nested models. Your goal is to model the continuity by starting with a sum of isotropic or geometrically anisotropic structures whose total sill is the lowest sill in all directions. Then, in each of the directions with higher sills you add a zonal anisotropy component to the corresponding sum to compensate for the increased variability in that direction.

Example 2

The second example provides important perspective about the physics of zonal anisotropy analysis. It is an extreme case of the general guidelines for zonal anisotropy. You examine what happens when each of the directions is modeled with a single-form, non-nested model, and the sills for these models are clearly different.

Consider a spatial process with a continuity description almost identical to the one in the previous example. In the direction $\theta_1 = 40°$, the covariance model has again a single exponential structure $\gamma_1(\boldsymbol{h}) = \mathrm{Exp}(a_{0,1E}, c_{0,1E})$ with range $a_{0,1E} = 2$ and sill $c_{0,1E} = 1.5$. However, this time in the di-

rection $\theta_2 = 130°$ you have fit the experimental semivariogram by using a single exponential structure $\gamma_2(\boldsymbol{h}) = \mathrm{Exp}(a_{0,2E}, c_{0,2E})$ with range $a_{0,2E} = 1$ and sill $c_{0,2E} = 2$. These models are shown in Figure 68.15.

Figure 68.15 Zonal Anisotropy in Two Directions

In this case you have a simplified situation with a single covariance structure in each direction, and the two structures have different scale parameter values. This is a case of zonal anisotropy in which all directions have no shared component. Hence, you have a case with two pure zonal components, where both structures can be practically approximated by specifying two models with large RATIO= values. You could then use the following MODEL statement in PROC KRIGE2D to describe the covariance in this example:

```
model form=(exp,exp) range=(2,1) scale=(1.5,2)
      angle=(40,130) ratio=(1e8,1e8);
```

The semivariogram of the specified model is accurately shown in Figure 68.15, because the angles θ_1 and θ_2 are perpendicular and each component has a contribution to all directions except for the one that is perpendicular to its angle. In the general case, θ_1 and θ_2 might not be perpendicular; hence the maximum and minimum scale values can be different from those displayed in Figure 68.15.

In general, avoid configurations with pure zonal components. Correlation models with pure zonal components might imply zero continuity along some direction, which is a very unlikely occurrence in natural processes. For that reason, in similar cases try to use the analysis illustrated in the previous example. In particular, try to model the highest sill direction as a nested structure (such that it contains a geometrical anisotropy component whose cumulative sill is equal to the lower sill) and a zonal anisotropy component that accounts for the sill difference.

Anisotropic Nugget Effect

Isotropic nugget effects can be approximated with nested models, where one of the nested structures has a very small range. Applying a geometric anisotropy specification to this nested structure results in an anisotropic nugget effect.

Details of Ordinary Kriging

Introduction

Three common characteristics are often observed with spatial data (that is, data indexed by their spatial locations):

(i) slowly varying, large-scale variations in the measured values

(ii) irregular, small-scale variations

(iii) similarity of measurements at locations close together

As an illustration, consider a hypothetical example in which an organic solvent leaks from an industrial site and spreads over a large area. Assume the solvent is absorbed and immobilized into the subsoil above any groundwater level, so you can ignore any time dependence.

To find the areal extent and the concentration values of the solvent, you need measurements. Although the problem is inherently three-dimensional, if you measure total concentration in a column of soil or take a depth-averaged concentration, it can be handled reasonably well with two-dimensional techniques.

You usually assume that measured concentrations are higher closer to the source and decrease at larger distances from the source. On top of this smooth variation, measured concentrations typically have small-scale variations, due perhaps to the inherent variability of soil properties.

You also tend to suspect that measurements made close together yield similar concentration values, while measurements made far apart can have very different values.

These physically reasonable qualitative statements have no explicit probabilistic content. A number of numerical smoothing techniques, such as inverse distance weighting and splines, make use of large-scale variations and "close distance-close value" characteristics of spatial data to interpolate the measured concentrations for contouring purposes.

While characteristics (i) and (iii) are handled by such smoothing methods, characteristic (ii), the small-scale residual variation in the concentration field, is not accounted for.

There can be situations, due to the use of the prediction map or the relative magnitude of the irregular fluctuations, where you cannot ignore these small-scale irregular fluctuations. In other words, the smoothed or predicted values of the concentration field alone are not a sufficient characterization; you also need the possible spread around these contoured values.

Spatial Random Fields

One method of incorporating characteristic (ii) into the construction of a contour map is to model the concentration field as a spatial random field (SRF). The mathematical details of SRF models are given in a number of texts, such as Cressie (1993) and Christakos (1992). The mathematics of SRFs is formidable. However, under certain simplifying assumptions, it produces classical linear predictors with very simple properties, enabling easy implementation for prediction purposes. These predictors, primarily ordinary kriging (OK), give both a prediction and a standard error of prediction at unsampled locations. This allows the construction of a map of both predicted values and level of uncertainty about the predicted values.

The key assumption in applying the SRF formalism is that the measurements come from a single realization of the SRF. However, in most geostatistical applications, the focus is on a single, unique realization. This is unlike most other situations in stochastic modeling in which there will be future experiments or observational activities (at least conceptually) under similar circumstances. This renders many traditional ideas of statistical inference ambiguous and somewhat counterintuitive.

Additional logical and methodological problems could stand in the way of applying a stochastic model to a unique but partly unknown natural process; see the introduction in Matheron (1971) and Cressie (1993, section 2.3). These difficulties have resulted in attempts to frame the prediction problem in a completely deterministic way (Isaaks and Srivastava 1988; Journel 1985). Also, some issues with kriging, and with spatial prediction methods in general, are related to the necessary assumption of ergodicity of the spatial process. This assumption is required to estimate the covariance or semivariogram from sample data. Details are provided in Cressie (1993, pp. 52–58).

Despite these difficulties, ordinary kriging remains a popular and widely used tool in modeling spatial data, especially in generating surface plots and contour maps. An abbreviated derivation of the OK predictor for point prediction and the associated standard error is discussed in the following section. Full details are given in Journel and Huijbregts (1978); Christakos (1992); Cressie (1993).

Ordinary Kriging

Denote the SRF by $Z(\boldsymbol{s}), \boldsymbol{s} \in D \subset \mathcal{R}^2$. Following the notation in Cressie (1993), the following model for $Z(\boldsymbol{s})$ is assumed:

$$Z(\boldsymbol{s}) = \mu + \varepsilon(\boldsymbol{s})$$

Here, μ is the fixed, unknown mean of the process, and $\varepsilon(\boldsymbol{s})$ is a zero mean SRF, which represents the variation around the mean.

In most practical applications, an additional assumption is required in order to estimate the covariance C_z of the $Z(\boldsymbol{s})$ process. This assumption is second-order stationarity:

$$C_z(\boldsymbol{s}_1, \boldsymbol{s}_2) = \mathrm{E}[\varepsilon(\boldsymbol{s}_1)\varepsilon(\boldsymbol{s}_2)] = C_z(\boldsymbol{s}_1 - \boldsymbol{s}_2) = C_z(\boldsymbol{h})$$

This requirement can be relaxed slightly when you are using the semivariogram instead of the covariance. In this case, second-order stationarity is required of the differences $\varepsilon(\boldsymbol{s}_1) - \varepsilon(\boldsymbol{s}_2)$ rather than $\varepsilon(\boldsymbol{s})$:

$$\gamma_z(\boldsymbol{s}_1, \boldsymbol{s}_2) = \frac{1}{2}\mathrm{E}[(\varepsilon(\boldsymbol{s}_1) - \varepsilon(\boldsymbol{s}_2))^2] = \gamma_z(\boldsymbol{s}_1 - \boldsymbol{s}_2) = \gamma_z(\boldsymbol{h})$$

By performing local kriging, the spatial processes represented by the previous equation for $Z(\boldsymbol{s})$ are more general than they appear. In local kriging, at an unsampled location $\boldsymbol{s}_0$, a separate model is fit using only data in a neighborhood of $\boldsymbol{s}_0$. This has the effect of fitting a separate mean μ at each point, and it is similar to the *kriging with trend* (KT) method discussed in Journel and Rossi (1989).

Given the N measurements $Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_N)$ at known locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N$, you want to obtain a prediction of Z at an unsampled location $\boldsymbol{s}_0$. When the following three requirements are imposed on the predictor $\hat{Z}$, the OK predictor is obtained:

(i) $\hat{Z}$ is linear in $Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_N)$

(ii) $\hat{Z}$ is unbiased

(ii) $\hat{Z}$ minimizes the mean square prediction error $\mathrm{E}[(Z(\boldsymbol{s}_0) - \hat{Z}(\boldsymbol{s}_0))^2]$

Linearity requires the following form for $\hat{Z}(\boldsymbol{s}_0)$:

$$\hat{Z}(\boldsymbol{s}_0) = \sum_{i=1}^{N} \lambda_i Z(\boldsymbol{s}_i)$$

Applying the unbiasedness condition to the preceding equation yields

$$\mathrm{E}[\hat{Z}(\boldsymbol{s}_0)] = \mu \Rightarrow \sum_{i=1}^{N} \lambda_i \mathrm{E}[Z(\boldsymbol{s}_i)] = \mu \Rightarrow \sum_{i=1}^{N} \lambda_i \mu = \mu \Rightarrow \sum_{i=1}^{N} \lambda_i = 1$$

Finally, the third condition requires a constrained linear optimization that involves $\lambda_1, \ldots, \lambda_N$ and a Lagrange parameter $2m$. This constrained linear optimization can be expressed in terms of the function $L(\lambda_1, \ldots, \lambda_N, m)$ given by

$$L = \mathrm{E}\left[\left(Z(\boldsymbol{s}_0) - \sum_{i=1}^{N} \lambda_i Z(\boldsymbol{s}_i)\right)^2\right] - 2m\left(\sum_{i=1}^{N} \lambda_i - 1\right)$$

Define the $N \times 1$ column vector $\boldsymbol{\lambda}$ by

$$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)'$$

and the $(N+1) \times 1$ column vector $\boldsymbol{\lambda}_0$ by

$$\boldsymbol{\lambda}_0 = (\lambda_1, \ldots, \lambda_N, m)' = \begin{pmatrix} \boldsymbol{\lambda} \\ m \end{pmatrix}$$

The optimization is performed by solving

$$\frac{\partial L}{\partial \boldsymbol{\lambda}_0} = \mathbf{0}$$

in terms of $\lambda_1, \ldots, \lambda_N$ and m.

The resulting matrix equation can be expressed in terms of either the covariance $C_z(\boldsymbol{h})$ or semivariogram $\gamma_z(\boldsymbol{h})$. In terms of the covariance, the preceding equation results in the matrix equation

$$\mathbf{C}\boldsymbol{\lambda}_0 = \mathbf{C}_0$$

where

$$\mathbf{C} = \begin{pmatrix} C_z(\mathbf{0}) & C_z(\boldsymbol{s}_1 - \boldsymbol{s}_2) & \cdots & C_z(\boldsymbol{s}_1 - \boldsymbol{s}_N) & 1 \\ C_z(\boldsymbol{s}_2 - \boldsymbol{s}_1) & C_z(\mathbf{0}) & \cdots & C_z(\boldsymbol{s}_2 - \boldsymbol{s}_N) & 1 \\ & & \ddots & & \\ C_z(\boldsymbol{s}_N - \boldsymbol{s}_1) & C_z(\boldsymbol{s}_N - \boldsymbol{s}_2) & \cdots & C_z(\mathbf{0}) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix}$$

and

$$\mathbf{C_0} = \begin{pmatrix} C_z(\boldsymbol{s}_0 - \boldsymbol{s}_1) \\ C_z(\boldsymbol{s}_0 - \boldsymbol{s}_2) \\ \vdots \\ C_z(\boldsymbol{s}_0 - \boldsymbol{s}_N) \\ 1 \end{pmatrix}$$

The solution to the previous matrix equation is

$$\hat{\boldsymbol{\lambda}}_0 = \mathbf{C}^{-1}\mathbf{C_0}$$

Using this solution for $\boldsymbol{\lambda}$ and m, the ordinary kriging prediction at r_0 is

$$\hat{Z}(\boldsymbol{s}_0) = \lambda_1 Z(\boldsymbol{s}_1) + \cdots + \lambda_N Z(\boldsymbol{s}_N)$$

with associated prediction error the square root of the variance

$${\sigma_z}^2(\boldsymbol{s}_0) = C_z(\mathbf{0}) - \boldsymbol{\lambda}'\mathbf{c_0} + m$$

where $\mathbf{c_0}$ is $\mathbf{C_0}$ with the 1 in the last row removed, making it an $N \times 1$ vector.

These formulas are used in the best linear unbiased prediction (BLUP) of random variables (Robinson 1991). Further details are provided in Cressie (1993, pp. 119–123).

Because of possible numeric problems when solving the previous matrix equation, Deutsch and Journel (1992) suggest replacing the last row and column of 1s in the preceding matrix $\mathbf{C}$ by $C_z(0)$, keeping the 0 in the $(N+1, N+1)$ position and similarly replacing the last element in the preceding right-hand vector $\mathbf{C_0}$ with $C_z(0)$. This results in an equivalent system but avoids numeric problems when $C_z(0)$ is large or small relative to 1.

Computational Resources

To generate a predicted value at a single grid point by using N data points, PROC KRIGE2D must solve the kriging system

$$\mathbf{C}\boldsymbol{\lambda}_0 = \mathbf{C_0}$$

where the dimensions of $\mathbf{C}$ are $(N+1) \times (N+1)$ and the right-hand-side $\mathbf{C_0}$ has one column.

Holding the matrix and vector associated with this system in core requires approximately $8N^2/2$ bytes. The CPU time used in solving the system is proportional to N^3. For large N, this time dominates the $O(N^2)$ time to compute the elements of the covariance matrix $\mathbf{C}$ from the specified covariance or semivariogram model.

For local kriging, the kriging system is set up and solved for each grid point. Part of the setup process involves determining the neighborhood of each grid point. A fast K-D tree algorithm determines neighborhoods. For G grid points, the dominant CPU time factor is setting up and solving the G kriging systems. The N in the algorithm of the section "Ordinary Kriging" on page 5048 is the number of data points in a given neighborhood, and it can differ for each grid point.

In global kriging, the entire input data set and all grid points set up and solve the single system

$$\mathbf{C}\boldsymbol{\lambda}_{\mathbf{0}} = \mathbf{C}_{\mathbf{0}}$$

Again $\mathbf{C}$ has dimensions $(N+1) \times (N+1)$, but $\boldsymbol{\lambda}_{\mathbf{0}}$ and $\mathbf{C}_{\mathbf{0}}$ now have G columns, where G is the number of grid points. Memory requirements are approximately $8[(N^2/2) + GN]$ bytes. The CPU time used in solving the system is still dominated by the N^3 factorization of the left-hand side.

Output Data Sets

The KRIGE2D procedure produces two data sets: the OUTEST=*SAS-data-set* and the OUTNBHD=*SAS-data-set*. These data sets are described as follows.

OUTEST=*SAS-data-set*

The OUTEST= data set contains the kriging predictions and the associated standard errors. The OUTEST= data set contains the following variables:

- ESTIMATE, which is the kriging prediction for the current variable.
- GXC, which is the *x* coordinate of the grid point at which the kriging prediction is made.
- GYC, which is the *y* coordinate of the grid point at which the kriging prediction is made.
- LABEL, which is the label for the current PREDICT/MODEL combination that produces the kriging prediction. If you do not specify a label, default labels of the form Pred*j*.Model*k* are used.
- NPOINTS, which is the number of points used in the prediction. This number varies for each grid point if local kriging is performed.
- STDERR, which is the standard error of the kriging predict.
- VARNAME, which is the variable name.

OUTNBHD=*SAS-data-set*

When you specify the RADIUS= option or the NUMPOINTS= option in the PREDICT statement, local kriging is performed. Local kriging is simply ordinary kriging at a given grid location, using only those data points in a neighborhood defined by the RADIUS= value or the NUMPOINTS= value.

The OUTNBHD= data set contains one observation for each data point in each neighborhood. Hence, this data set can be large. For example, if the grid specification results in 1,000 grid points and each grid point has a neighborhood of 100 points, the resulting OUTNBHD= data set contains 100,000 points.

The OUTNBHD= data set contains the following variables:

- GXC, which is the *x* coordinate of the grid point.
- GYC, which is the *y* coordinate of the grid point.

- ID, which is the ID variable value or observation. number of the current data point
- LABEL, which is the label for the current PREDICT/MODEL combination. If you do not specify a label, default labels of the form Pred*j*.Model*k* are used.
- NPOINTS, which is the number of points used in the prediction.
- RADIUS, which is the radius used for each neighborhood.
- VALUE, which is the value of the variable at the current data point.
- VARNAME, which is the variable name of the current variable.
- XC, which is the *x* coordinate of the current data point.
- YC, which is the *y* coordinate of the current data point.

If no ID statement is specified, then the corresponding observation number is assigned to the variable ID, instead.

Displayed Output

In addition to the output data sets, the KRIGE2D procedure produces output objects as well. The KRIGE2D procedure output objects are the following:

- a default "Number of Observations" table that displays the number of observations read from the input data set and the number of observations used in the analysis.
- a map that shows the spatial distribution of the observations of the current VAR= variable in the PREDICT statement. The observations are displayed by default with circled markers whose color indicates the VAR= value at the corresponding location.
- a default table for each PREDICT statement that sums up basic information about the kriging analysis.
- a default table for each MODEL statement that shows the covariance model parameters for the corresponding PREDICT statement.
- plots of the kriging prediction and the prediction standard error at each point of the specified output grid or at specified individual locations. The KRIGE2D procedure produces by default a plot of the kriging prediction and the corresponding prediction error for each MODEL statement of every PREDICT statement that you specify. You can produce more of these plots with styles that you can specify by using the available suboptions of the PLOTS=PREDICTION option.
- a "Store Info" table with basic information about the input item store. This table is produced by default when you specify the RESTORE statement.
- a "Store Variables Information" table that describes the analysis variables of an input item store. The table is produced by default when you specify an item store with the RESTORE statement.
- a "Store Models Information" table with detailed information about the models and direction angles that are contained in an input item store. The table is produced by default when you specify an item store with the RESTORE statement.

ODS Table Names

Each table created by PROC KRIGE2D has a name associated with it, and you must use this name to reference the table when using ODS Graphics. These names are listed in Table 68.4.

Table 68.4 ODS Tables Produced by PROC KRIGE2D

ODS Table Name	Description	Statement	Option
KrigInfo	Kriging analysis general information	PROC	Default output
ModelInfo	Parameters of the covariance model used in current kriging analysis	PROC	Default output
NObs	Number of observations read and used	PROC	Default output
StoreInfo	Input item store identity information	RESTORE	Default output
StoreModelInfo	Input item direction angles and models information	RESTORE	INFO
StoreVarInfo	Input item store variables and their statistics	RESTORE	INFO

ODS Graphics

Statistical procedures use ODS Graphics to create graphs as part of their output. ODS Graphics is described in detail in Chapter 21, "Statistical Graphics Using ODS."

Before you create graphs, ODS Graphics must be enabled (for example, by specifying the ODS GRAPHICS ON statement). For more information about enabling and disabling ODS Graphics, see the section "Enabling and Disabling ODS Graphics" on page 607 in Chapter 21, "Statistical Graphics Using ODS."

The overall appearance of graphs is controlled by ODS styles. Styles and other aspects of using ODS Graphics are discussed in the section "A Primer on ODS Statistical Graphics" on page 606 in Chapter 21, "Statistical Graphics Using ODS."

For additional control of the graphics that are displayed, see the PLOTS option in the section "PROC KRIGE2D Statement" on page 5009.

ODS Graph Names

PROC KRIGE2D assigns a name to each graph it creates by using ODS Graphics. You can use these names to reference the graphs when using ODS Graphics. You must also specify the PLOTS= option indicated in Table 68.5.

Table 68.5 Graphs Produced by PROC KRIGE2D

ODS Graph Name	Plot Description	Statement	Option
ObservationsPlot	Scatter plot of observed data and colored markers indicating observed values	PROC	PLOTS=OBSERV

Table 68.5 *continued*

ODS Graph Name	Plot Description	Statement	Option
PredictionPlot	Contour plots of the kriging prediction, surface of the prediction error, and outlines of the observation locations	PROC	PLOTS=PREDICTION
Semivariogram	Plots of the semivariogram models used for all prediction tasks	PROC	PLOTS=SEMIVAR

Examples: KRIGE2D Procedure

Example 68.1: Spatial Prediction of Pollutant Concentration

The example in the section "Example 124.1: Aspects of Semivariogram Model Fitting" on page 10112 in Chapter 124, "The VARIOGRAM Procedure," investigates fitting of a theoretical model to describe spatial correlation in a study of 138 simulated arsenic logarithm concentration (logAs) observations. These observations form the logAsData data set, which is treated as actual data for illustration in the examples.

In this example, you use the logAsData data set and the semivariogram analysis results to predict the logAs variable values across space in a specified square region of size 500 km × 500 km. Your goal is to answer scientific questions in your analysis by means of your prediction results. This application highlights the impact of the correlation model choice on predictions. The example in section "Example 68.2: Investigating the Effect of Model Specification on Spatial Prediction" on page 5064 examines additional aspects of this impact.

The World Health Organization (WHO) standard for maximum arsenic concentration in drinking water is 10 μg/lt. Assume that you want to answer the following question: In what percentage of the study area does the Arsenic concentration exceed the WHO regulatory standard?

First, you read the logAsData data set with the following DATA step:

```
title 'Spatial Prediction of Log-Arsenic Concentration';

data logAsData;
   input East North logAs @@;
   label logAs='log(As) Concentration';
   datalines;
   193.0 296.6 -0.68153   232.6 479.1  0.96279   268.7 312.5 -1.02908
    43.6   4.9  0.65010   152.6  54.9  1.87076   449.1 395.8  0.95932
   310.9 493.6 -1.66208   287.8 164.9 -0.01779   330.0   8.0  2.06837
   225.7 241.7  0.15899   452.3  83.4 -1.21217   156.5 462.5 -0.89031
    11.5  84.4 -0.24496   144.4 335.7  0.11950   149.0 431.8 -0.57251
   234.3 123.2 -1.33642    37.8 197.8 -0.27624   183.1 173.9 -2.14558
   149.3 426.7 -1.06506   434.4  67.5 -1.04657   439.6 237.0 -0.09074
    36.4 175.2 -1.21211   370.6 244.0  3.28091   452.0  96.5 -0.77081
   247.0  86.8  0.04720   413.6 373.2  1.78235   253.5 291.7  0.56132
   129.7 111.9  1.34000   352.7  42.1  0.23621   279.3  82.7  2.12350
   382.6 290.7  0.86756   188.2 222.8 -1.23308   382.8 154.5 -0.94094
```

```
   304.4 309.2 -1.95158   337.5 387.2 -1.31294   490.7 189.8  0.40206
   159.0 100.1 -0.22272   245.5 329.2 -0.26082   372.1 379.5 -1.89078
   417.8  84.1 -1.25176   173.9 407.6 -0.24240   121.5 107.7  1.54509
   453.5 313.6  0.65895   143.5 346.7 -0.87196   157.4 125.5 -1.96165
   371.8 353.2 -0.59464   358.9 338.2 -1.07133     8.6 437.8  1.44203
   395.9 394.2 -0.24144   149.5  58.9  1.17459   453.5 420.6 -0.63951
   182.3  85.0  1.00005    21.0 290.1  0.31016    11.1 352.2 -0.88418
   131.2 238.4 -0.57184   104.9   6.3  1.12054   247.3 256.0  0.14019
   428.4 383.7  0.92448   327.8 481.1 -2.72543   199.2  92.8 -0.05717
   453.9 230.1  0.16571   205.0 250.6  0.07581   459.5 271.6  0.93700
   229.5 262.8  1.83590   370.4 228.6  2.96611   330.2 281.9  1.79723
   354.8 388.3 -3.18262   406.2 222.7  2.41594   254.4 393.1  2.03221
    96.7  85.2 -0.47156   407.2 256.8  0.66747   498.5 273.8  1.03041
   417.2 471.4 -1.42766   368.8 424.3 -0.70506   303.0  59.1  1.43070
   403.1 264.1  1.64554    21.2 360.8  0.67094   148.2  78.1  2.15323
   305.5 310.7 -1.47985   228.5 180.3 -0.68386   161.1 143.3  1.07901
    70.5 155.1  0.54652   363.1 282.6 -0.43051    86.0 472.5 -1.18855
   175.9 105.3 -2.08112    96.8 426.3  1.56592   475.1 453.1 -1.53776
   125.7 485.4  1.40054   277.9 201.6 -0.54565   406.2 125.0 -1.38657
    60.0 275.5 -0.59966   431.3 494.6 -0.36860   399.9 399.0 -0.77265
    28.8 311.1  0.91693   166.1 348.2 -0.49056   266.6  83.5  0.67277
    54.7 356.3  0.49596   433.5 460.3 -1.61309   201.7 167.6 -1.40678
   158.1 203.6 -1.32499    67.6 230.4  1.14672    81.9 250.0  0.63378
   372.0  50.7  0.72445    26.4 264.6  1.00862   300.1  91.7 -0.74089
   303.0 447.4  1.74589   108.4 386.2  1.12847    55.6 191.7  0.95175
    36.3 273.2  1.78880    94.5 298.3 -2.43320   366.1 187.3 -0.80526
   130.7 389.2 -0.31513    37.2 324.2  0.24489   295.5 211.8  0.41899
    58.6 206.2  0.18495   346.3 142.8 -0.92038   484.2 215.9  0.08012
   451.4 415.7  0.02773    58.9  86.5  0.17652   212.6 363.9  0.17215
   378.7 407.6  0.51516   265.9 305.0 -0.30718   123.2 314.8 -0.90591
    26.9 471.7  1.70285    16.5   7.1  0.51736   255.1 472.6  2.02381
   111.5 148.4 -0.09658   440.4 375.0  1.23285   406.4  19.5  1.01181
   321.2  65.8 -0.02095   466.4 357.1 -0.49272     2.0 484.6  0.50994
   200.9 205.1  0.43543    30.3 337.0  1.60882   297.0  12.7  1.79824
   158.2 450.7  0.05295   122.8 105.3  1.53936   417.8 329.7 -2.08124
;
```

For prediction of the logAs values in the specified area, assume a rectangular grid of nodes with an equal spacing of 5 km between neighboring nodes in the north and east directions. This produces a total of $101 \times 101 = 10201$ prediction locations.

In the section "Example 124.1: Aspects of Semivariogram Model Fitting" on page 10112 in Chapter 124, "The VARIOGRAM Procedure," you saved the selected fitted model that resulted from the correlation analysis into the SemivAsStore item store as shown in the following statements:

```
ods graphics on;

proc variogram data=logAsData plots=none;
   store out=SemivAsStore / label='LogAs Concentration Models';
   compute lagd=5 maxlag=40;
   coord xc=East yc=North;
   model form=auto(mlist=(exp,gau,mat) nest=1 to 2);
   var logAs;
run;
```

In the KRIGE2D procedure you specify the name of the item store you want to use for prediction input in the IN= option of the RESTORE statement. You request use of the selected model for prediction by specifying the STORESELECT option in the MODEL statement.

The INFO option of the RESTORE statement produces a table with information about the selected fitted model in the item store. To review all models in the input item store, specify the two INFO option suboptions. In particular, specify the DET suboption to request details about all additional fitted models that are included in the item store and the ONLY suboption to suppress prediction and produce only the tables about the item store, as shown in the following statements:

```
proc krige2d data=logAsData outest=pred plots=none;
   restore in=SemivAsStore / info(det only);
   coordinates xc=East yc=North;
   predict var=logAs;
   model storeselect;
   grid x=0 to 500 by 5 y=0 to 500 by 5;
run;
```

PROC KRIGE2D produces a table with general information about the input item store identity, as shown in Output 68.1.1.

Output 68.1.1 PROC KRIGE2D and Input Item Store General Information

Spatial Prediction of Log-Arsenic Concentration

The KRIGE2D Procedure

Correlation Model Item Store Information	
Input Item Store	WORK.SEMIVASSTORE
Item Store Label	LogAs Concentration Models
Data Set Created From	WORK.LOGASDATA
By-group Information	No By-groups Present
Created By	PROC VARIOGRAM
Date Created	21OCT16:21:28:43

The second table in Output 68.1.2 itemizes the variables in the item store and displays the sample mean and standard deviation of their data set of origin. Hence, the values shown in Output 68.1.2 refer to the observations in the logAsData data set.

Output 68.1.2 Variables in the Input Item Store

Item Store Variables		
Variable	**Mean**	**Std Deviation**
logAs	0.084309	1.527707

The table in Output 68.1.3 presents all the correlation models fitted to the arsenic logarithm logAs empirical semivariance that are saved in the SemivAsStore item store.

Output 68.1.3 Angle and Models Information in the Input Item Store

Item Store Models For logAs	
Class	**Model**
1	Gau-Gau
	Gau-Mat
2	Exp-Gau
3	Exp-Mat
4	Mat
5	Gau
6	Exp
	Exp-Exp
	Mat-Exp
	Gau-Exp

According to Output 68.1.3, the Gaussian-Gaussian model is the selected model for the empirical semivariance fit based on the specific weighted least squares fit and ranking criteria. In the section "Example 124.1: Aspects of Semivariogram Model Fitting" on page 10112 in Chapter 124, "The VARIOGRAM Procedure," it is noted that all fitted models in the first five equivalence classes produce very similar semivariograms, and this is likely to lead to similar results in prediction analysis. For comparison purposes, you choose to examine the selected model, in addition to the exponential model in the SemivAsStore item store. As shown in Output 68.1.3, the exponential model is one of the least well-fit models based on the criteria used for the specific fit. You are interested in comparing the predictions from each one of these two models, and you examine their impact on your analysis.

The default item store model selection is the model on top of the list in Output 68.1.3. Hence, you specify the STORESELECT option in the MODEL statement without any suboptions, and it invokes the Gaussian-Gaussian model from the SemivAsStore item store. You assign the label SELMODEL to the corresponding MODEL statement.

You also specify a second MODEL statement with the label EXPMODEL to request prediction based on the exponential correlation form. In this case you specify the STORESELECT(MODEL=) option in the MODEL statement to request the desired form.

You omit the INFO option from the RESTORE statement. You specify the PRED and the SEMIVAR options in the PLOTS option of the PROC KRIGE2D statement to produce plots of the predicted values and the semivariance model, respectively, for each MODEL statement. You request that the prediction output be saved in the Pred output data set.

You satisfy the preceding requests by specifying the following statements:

```
proc krige2d data=logAsData outest=Pred plots(only)=(pred semivar);
   restore in=SemivAsStore;
   coordinates xc=East yc=North;
   predict var=logAs;
   SelModel: model storeselect;
   ExpModel: model storeselect(model=exp);
   grid x=0 to 500 by 5 y=0 to 500 by 5;
run;
```

When you run these statements, in addition to the input item store information table, PROC KRIGE2D also produces the number of observations table and general kriging process information, as shown in Output 68.1.4.

Output 68.1.4 Number of Observations and Kriging Information Tables

Spatial Prediction of Log-Arsenic Concentration

The KRIGE2D Procedure

Dependent Variable: logAs

Number of Observations Read	138
Number of Observations Used	138

Kriging Information	
Prediction Grid Points	10201
Type of Analysis	Global

PROC KRIGE2D first uses the Gaussian-Gaussian model. The table in Output 68.1.5 shows the saved parameter values of the fitted Gaussian-Gaussian model in the SemivAsStore item store. PROC KRIGE2D uses these parameters for the prediction based on the selected model.

Output 68.1.5 Information about the Gaussian-Gaussian Model

Spatial Prediction of Log-Arsenic Concentration

The KRIGE2D Procedure

Dependent Variable: logAs
Prediction: Pred1, Model: SelModel

Covariance Model Information for SelModel	
Nested Structure 1 Type	Gaussian
Nested Structure 1 Sill	0.3276646
Nested Structure 1 Range	62.312728
Nested Structure 1 Effective Range	107.92881
Nested Structure 2 Type	Gaussian
Nested Structure 2 Sill	1.261545
Nested Structure 2 Range	21.459563
Nested Structure 2 Effective Range	37.169053
Nugget Effect	0.0830758

The semivariogram of the Gaussian-Gaussian model with the parameters shown in Output 68.1.5 is depicted in Output 68.1.6.

Output 68.1.6 Gaussian-Gaussian Semivariogram Model Used in Kriging Predictions

Output 68.1.7 is a map of the kriging prediction of the arsenic concentration values logAs in the specified domain. The prediction error surface shows a naturally increasing error as you move farther away from the observation locations. Interestingly, kriging predicts a small area of increased arsenic concentration values located in the central-eastern part of the domain. The WHO threshold of 10 μg/lt for the maximum allowed arsenic concentration in water translates into about 2.3 in the log scale, and the particular area exhibits values in excess of 3. Due to the suggested violation of the WHO standard, this particular area is very likely to be the focus of further environmental risk analysis.

Output 68.1.7 Predicted Arsenic Logarithm Values with Gaussian-Gaussian Covariance

Next, PROC KRIGE2D performs prediction with the exponential model. The model parameters are also read from the SemivAsStore item store and are shown in Output 68.1.8.

Output 68.1.8 Information about the Exponential Model

Spatial Prediction of Log-Arsenic Concentration

The KRIGE2D Procedure

Dependent Variable: logAs
Prediction: Pred1, Model: ExpModel

Covariance Model Information for ExpModel	
Type	Exponential
Sill	1.6779788
Range	24.537294
Effective Range	73.611882
Nugget Effect	0

Output 68.1.9 illustrates the semivariogram of the nested exponential model where its parameter values are those shown in Output 68.1.8.

Output 68.1.9 Exponential Semivariogram Model Used in Kriging Predictions

The prediction plot for the exponential model is shown in Output 68.1.10. Prediction values and spatial patterns are similar overall to those of the Gaussian-Gaussian case. Clearly, although both models predict the same basic characteristics for the arsenic logarithm concentration distribution, the exponential model suggests a more limited spatial variability in closely neighboring locations. The lack of a nugget effect in the exponential model justifies this behavior. Also, the exponential model predictions seem less inclined to deviate farther away from the near-zero mean than the Gaussian-Gaussian model predictions.The prediction error reaches about the same upper values for both models, though its low values are slightly smaller in the exponential model.

Output 68.1.10 Predicted Arsenic Logarithm Values with Exponential Covariance

In the following two-step computation, you proceed to compute the percentage of the study area where the arsenic concentration exceeds the WHO regulatory standard according to your predictions. First, a DATA step marks the arsenic predicted values in excess of the WHO concentration threshold of 10 μg/lt and saves the outcome into an indicator variable OverLimit. The DATA step input is the prediction Pred output data set, where the logarithm arsenic prediction is stored in the estimate variable. The DATA step also transforms the arsenic logarithm values back into arsenic concentration values to compare them to the threshold value. You use the following statements:

```
data AsOverLimit;
   set Pred;
   OverLimit = (exp(estimate) > 10) * 100;
run;
```

The second step uses the MEANS procedure to express the selected nodes population, where the WHO arsenic concentration limit violation occurs, as a percentage of the entire domain area. You study the results of each correlation model separately by specifying the BY statement in the PROC MEANS. The BY variable is the Label variable in the AsOverLimit and Pred data sets. You need to sort the AsOverLimit data prior to using PROC MEANS. You run the following statements:

```
proc sort data=AsOverLimit;
   by Label;
run;
proc means data=AsOverLimit mean;
   var OverLimit;
   by label;
   label Overlimit="Percent above WHO threshold";
run;

ods graphics off;
```

The Gaussian-Gaussian model prediction produces the result in Output 68.1.11. The analysis suggests a minimal occurrence of excessive arsenic concentration in drinking water in about 0.43% of the study region.

Output 68.1.11 Violation of Arsenic Concentration Threshold Using Gaussian-Gaussian Model

Spatial Prediction of Log-Arsenic Concentration

The MEANS Procedure

Label for the PREDICT/MODEL combination=Pred1.SelModel

Analysis Variable : OverLimit Percent above WHO threshold
Mean
0.4313303

The exponential model predicts that the WHO arsenic concentration threshold is exceeded in about 0.27% of the domain, as shown in Output 68.1.12. Although this is still a minimal occurrence of the threshold violation across the region, the exponential model estimates the impact to be at about two thirds of the Gaussian-Gaussian model percentage.

Output 68.1.12 Violation of Arsenic Concentration Threshold Using Exponential Model

Spatial Prediction of Log-Arsenic Concentration

The MEANS Procedure

Label for the PREDICT/MODEL combination=Pred1.ExpModel

Analysis Variable : OverLimit Percent above WHO threshold
Mean
0.2744829

The results in Output 68.1.11 and Output 68.1.12 suggest that it might not be possible to provide a unique answer about the area percentage that is affected by increased arsenic concentration. You chose to examine two different correlation models whose performance is relatively similar, and they provide impact estimates that differ by about 37%.

You might conclude that the answer to the initial question about the percentage value lies in the neighborhood of the results given by the two correlation models. Further analysis with more models is necessary to validate this assumption. It is important to note that apart from the continuity model choice, additional factors contribute to this investigation. Such factors could be the use of local instead of global kriging, or even going back to the empirical semivariogram computation stage and repeating the analysis for different possible spatial continuity empirical estimates. A sensible approach to tackle this analysis would be to investigate the range of the impact suggested by all candidate correlation models and to proceed by defining the best and worst case scenarios for the size of the affected area.

Eventually, when it comes to using your findings, it is important to account for the subjective nature of stochastic analysis and multiple possible answers to your questions. In that sense, some scientific questions might be more sensible than others to interpret your results correctly. For instance, you might want to investigate only whether the adversely affected domain percentage is below 1%, rather than attempting to provide a specific value for it. Then, you might consider the preceding findings sufficient, despite any fluctuations in the estimated percentage. In a different scenario, the areas with high pollutant concentration could be populated. Hence, any local health standard violation is probably unacceptable, and it can be crucial that you provide solid and more detailed assessment in that case.

The section "Example 105.3: Risk Analysis with Simulation" on page 8727 in Chapter 105, "The SIM2D Procedure," investigates a different aspect of this study and offers additional perspective about spatial analysis.

Example 68.2: Investigating the Effect of Model Specification on Spatial Prediction

It is generally believed that spatial prediction is robust against model specification, while the standard error computation is not so robust. This example investigates the effect of using these different models on the prediction and associated standard errors.

In the section "Theoretical Semivariogram Model Fitting" on page 10036 in Chapter 124, "The VARIOGRAM Procedure," a particular theoretical semivariogram is fitted to the coal seam thickness data empirical semivariogram. The chosen semivariogram is Gaussian with a scale (sill) of $c_0 = 7.2881$ and a range of $a_0 = 30.6239$.

Another possible model choice could be the spherical semivariogram. First, use a DATA step to input the thickness data:

```
title 'Effect of Model Specification on Prediction';

data thick;
   input East North Thick @@;
   label Thick='Coal Seam Thickness';
   datalines;
    0.7  59.6  34.1   2.1  82.7  42.2   4.7  75.1  39.5
    4.8  52.8  34.3   5.9  67.1  37.0   6.0  35.7  35.9
    6.4  33.7  36.4   7.0  46.7  34.6   8.2  40.1  35.4
   13.3   0.6  44.7  13.3  68.2  37.8  13.4  31.3  37.8
   17.8   6.9  43.9  20.1  66.3  37.7  22.7  87.6  42.8
   23.0  93.9  43.6  24.3  73.0  39.3  24.8  15.1  42.3
   24.8  26.3  39.7  26.4  58.0  36.9  26.9  65.0  37.8
   27.7  83.3  41.8  27.9  90.8  43.3  29.1  47.9  36.7
   29.5  89.4  43.0  30.1   6.1  43.6  30.8  12.1  42.8
```

```
   32.7  40.2  37.5  34.8   8.1  43.3  35.3  32.0  38.8
   37.0  70.3  39.2  38.2  77.9  40.7  38.9  23.3  40.5
   39.4  82.5  41.4  43.0   4.7  43.3  43.7   7.6  43.1
   46.4  84.1  41.5  46.7  10.6  42.6  49.9  22.1  40.7
   51.0  88.8  42.0  52.8  68.9  39.3  52.9  32.7  39.2
   55.5  92.9  42.2  56.0   1.6  42.7  60.6  75.2  40.1
   62.1  26.6  40.1  63.0  12.7  41.8  69.0  75.6  40.1
   70.5  83.7  40.9  70.9  11.0  41.7  71.5  29.5  39.8
   78.1  45.5  38.7  78.2   9.1  41.7  78.4  20.0  40.8
   80.5  55.9  38.7  81.1  51.0  38.6  83.8   7.9  41.6
   84.5  11.0  41.5  85.2  67.3  39.4  85.5  73.0  39.8
   86.7  70.4  39.6  87.2  55.7  38.8  88.1   0.0  41.6
   88.4  12.1  41.3  88.4  99.6  41.2  88.8  82.9  40.5
   88.9   6.2  41.5  90.6   7.0  41.5  90.7  49.6  38.9
   91.5  55.4  39.0  92.9  46.8  39.1  93.4  70.9  39.7
   55.8  50.5  38.1  96.2  84.3  40.3  98.2  58.2  39.5
;
```

Fitting of the Gaussian model is performed in the section "Theoretical Semivariogram Model Fitting" on page 10036 in Chapter 124, "The VARIOGRAM Procedure," and the fitting parameters are saved in the SemivStoreGau item store with the following statements:

```
ods graphics on;

proc variogram data=thick noprint;
   store out=SemivStoreGau / label='Thickness Gaussian Model';
   compute lagd=7 maxlag=10;
   coord xc=East yc=North;
   model form=gau;
   var Thick;
run;
```

For prediction with the saved Gaussian model, you use the following statements to run the KRIGE2D procedure with input from the SemivStoreGau item store. You invoke the item store with the RESTORE statement. The STORESELECT option in the MODEL statement that specifies that you want to use the selected model in the item store as input for your prediction.

```
proc krige2d data=thick outest=pred1 noprint;
   restore in=SemivStoreGau;
   coordinates xc=East yc=North;
   predict var=Thick r=60;
   model storeselect;
   grid x=0 to 100 by 10 y=0 to 100 by 10;
run;
```

Then, you run the KRIGE2D procedure by using a spherical model. Start by using the VARIOGRAM procedure to fit a spherical model to the thick data set empirical semivariogram. You specify the STORE statement again in PROC VARIOGRAM to save the spherical model estimated parameters in an item store with the name SemivStoreSph. You use the following statements:

```
proc variogram data=thick plots(only)=fit;
   store out=SemivStoreSph / label='Thickness Sph Model';
   compute lagd=7 maxlag=10;
   coord xc=East yc=North;
   model form=sph;
   var Thick;
run;
```

The VARIOGRAM procedure fits the spherical model successfully, and the estimated parameters for this fit are shown in Output 68.2.1.

Output 68.2.1 Spherical Model Fitting Parameter Estimates

Effect of Model Specification on Prediction

The VARIOGRAM Procedure

Dependent Variable: Thick
Angle: Omnidirectional
Current Model: Spherical

Parameter Estimates

Parameter	Estimate	Approx Std Error	DF	t Value	Approx Pr > |t|
Nugget	0	0	8	.	.
Scale	7.1914	0.2827	8	25.44	<.0001
Range	63.2351	4.1050	8	15.40	<.0001

The fit summary is displayed in Output 68.2.2. When compared to the corresponding result in the section "Theoretical Semivariogram Model Fitting" on page 10036 in Chapter 124, "The VARIOGRAM Procedure," the goodness-of-fit criteria indicate a worse statistical fit for the spherical model compared to the Gaussian.

Output 68.2.2 Spherical Model Fit Summary

Fit Summary

Model	Weighted SSE	AIC
Sph	52.26791	23.14336

Output 68.2.3 suggests an acceptable fit of the spherical model to the thick data set. Obviously, the fit of the spherical model in the sensitive area near the semivariogram origin is less faithful to the empirical semivariance than the Gaussian model. The following analysis explores the consequence in the kriging prediction of this discrepancy.

Output 68.2.3 Fitted Spherical and Empirical Thick Semivariogram

For the next step, you run the KRIGE2D procedure by using the spherical model parameters stored in the SemivStoreSph item store. You use the following statements:

```
proc krige2d data=thick outest=pred2 noprint;
   restore in=SemivStoreSph;
   coordinates xc=East yc=North;
   predict var=Thick r=60;
   model storeselect;
   grid x=0 to 100 by 10 y=0 to 100 by 10;
run;
```

Eventually, you compare the prediction results and errors of the two models. You use a DATA step to compute the relative difference of the predicted values and the prediction error for each one of the Gaussian and the spherical models. You store the prediction relative difference in the prdRelDiff variable and the prediction relative error in the stdRelDiff variable. You save the output in the compare data set with the following statements:

```
data compare;
   merge pred1(rename=(estimate=g_prd stderr=g_std))
         pred2(rename=(estimate=s_prd stderr=s_std));
   prdRelDif = ((g_prd-s_prd) / s_prd) * 100;
   stdRelDif = ((g_std-s_std) / s_std) * 100;
run;
```

The MEANS procedure uses the compare data set to produce statistics about the prediction relative difference and error for each one of the prdRelDiff and stdRelDiff variables with the following statements:

```
proc means data=compare;
   var prdRelDif stdRelDif;
run;

ods graphics off;
```

Output 68.2.4 shows that on average the predicted values are very close for the two semivariogram models. The mean relative difference in the prediction values is close to zero with a low standard deviation, whereas the relative difference values fluctuate with an absolute maximum of about 5%.

However, note that the mean relative standard error is about -96%. According to the definition of the stdRelDiff variable, the high negative value indicates that the prediction error difference between the two models is very close to the spherical model prediction error. Hence, the prediction standard error of the spherical model is substantially larger than that of the Gaussian model. In fact, the prediction relative error never gets smaller than about 66% for the two models, where the negative sign in the Minimum and Maximum columns in Output 68.2.4 means that the prediction error is always greater for the spherical model.

Output 68.2.4 Comparison of Gaussian and Spherical Models

Effect of Model Specification on Prediction

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
prdRelDif	121	-0.0544593	1.3384023	-5.0751449	5.1926236
stdRelDif	121	-96.2515099	5.9400029	-99.8974418	-65.9275907

Example 68.3: Data Quality and Prediction with Missing Values

Kriging methods depend primarily on your data. The quantity and quality of your observations are important factors in minimizing prediction errors and increasing accuracy in your prediction analysis.

A typical aspect of data quality is measurement accuracy. In principle, the accuracy level of your data is not a parameter in kriging prediction; kriging assumes by definition that your data are perfectly accurate (hard) measurements. Whether you accept this assumption depends on your application. For example, an instrumentation error of $\pm 1\%$ in the data values might be regarded as considerable in one case, whereas the same level of uncertainty might be trivial within a different framework. Your experience and judgment are crucial when you consider whether observations in a data set might be too noisy for kriging predictions to be useful.

A second aspect of data quality involves the spatial arrangement of your observations. You need to have a sufficient number of observations in order to perform spatial prediction. Also, a key element in minimizing

prediction errors is an adequate sampling density. Interpretation of the expressions "sufficient number" and "adequate sampling" is again case-specific. In any event, you want enough measurements so that you can deduce the underlying spatial correlation in the working domain; see also the discussion in the section "Choosing the Size of Classes" on page 10087 in Chapter 124, "The VARIOGRAM Procedure."

This example focuses on the effects of different sampling densities on the prediction analysis. The demonstration is a slight variation of the example in the section "Getting Started: KRIGE2D Procedure" on page 5003. Specifically, you use the same correlation structure and prediction grid. However, the thick data set, is modified as follows: three values in the central area of the grid are assumed missing, namely the observation values at locations $\boldsymbol{s}_1 = (x_1, y_1) = (55.8, 50.5)$, $\boldsymbol{s}_2 = (x_2, y_2) = (52.8, 68.9)$, and $\boldsymbol{s}_3 = (x_3, y_3) = (52.9, 32.7)$. These locations have been selected so that an extended area without observations is created in the domain. The following DATA step is the input for the modified thick data set:

```
title 'Kriging Prediction in the Presence of Missing Values';

data thick;
   input East North Thick @@;
   label Thick='Coal Seam Thickness';
   datalines;
    0.7  59.6  34.1   2.1  82.7  42.2   4.7  75.1  39.5
    4.8  52.8  34.3   5.9  67.1  37.0   6.0  35.7  35.9
    6.4  33.7  36.4   7.0  46.7  34.6   8.2  40.1  35.4
   13.3   0.6  44.7  13.3  68.2  37.8  13.4  31.3  37.8
   17.8   6.9  43.9  20.1  66.3  37.7  22.7  87.6  42.8
   23.0  93.9  43.6  24.3  73.0  39.3  24.8  15.1  42.3
   24.8  26.3  39.7  26.4  58.0  36.9  26.9  65.0  37.8
   27.7  83.3  41.8  27.9  90.8  43.3  29.1  47.9  36.7
   29.5  89.4  43.0  30.1   6.1  43.6  30.8  12.1  42.8
   32.7  40.2  37.5  34.8   8.1  43.3  35.3  32.0  38.8
   37.0  70.3  39.2  38.2  77.9  40.7  38.9  23.3  40.5
   39.4  82.5  41.4  43.0   4.7  43.3  43.7   7.6  43.1
   46.4  84.1  41.5  46.7  10.6  42.6  49.9  22.1  40.7
   51.0  88.8  42.0  52.8  68.9     .  52.9  32.7     .
   55.5  92.9  42.2  56.0   1.6  42.7  60.6  75.2  40.1
   62.1  26.6  40.1  63.0  12.7  41.8  69.0  75.6  40.1
   70.5  83.7  40.9  70.9  11.0  41.7  71.5  29.5  39.8
   78.1  45.5  38.7  78.2   9.1  41.7  78.4  20.0  40.8
   80.5  55.9  38.7  81.1  51.0  38.6  83.8   7.9  41.6
   84.5  11.0  41.5  85.2  67.3  39.4  85.5  73.0  39.8
   86.7  70.4  39.6  87.2  55.7  38.8  88.1   0.0  41.6
   88.4  12.1  41.3  88.4  99.6  41.2  88.8  82.9  40.5
   88.9   6.2  41.5  90.6   7.0  41.5  90.7  49.6  38.9
   91.5  55.4  39.0  92.9  46.8  39.1  93.4  70.9  39.7
   55.8  50.5     .  96.2  84.3  40.3  98.2  58.2  39.5
;

ods graphics on;
```

NOTE: Here you assume prior knowledge of the correlation structure model, because its parameters are based on the complete thick data set. A covariance model extracted from the incomplete set with the missing values would be a covariance model coming from a different data set; hence, it is likely to have different parameters.

After you define the modified data set, you run PROC KRIGE2D and request the OBSERVATIONS plot with the SHOWMISSING suboption. You also request two instances of the PREDICTION plot: one that displays the prediction surface and contours, and another that plots the kriging standard error surface and contours. In both of these PREDICTION plots you specify that the observations be shown as gradient markers with outlines. The following statements compute the kriged predictions and produce the requested graphics:

```
proc krige2d data=thick outest=predictions
             plots(only)=(observ(showmissing)
                          pred(fill=pred line=pred obs=linegrad)
                          pred(fill=se line=se obs=linegrad));
   coordinates xc=East yc=North;
   predict var=Thick r=60;
   model scale=7.4599 range=30.1111 form=gauss;
   grid x=0 to 100 by 2.5 y=0 to 100 by 2.5;
run;

ods graphics off;
```

The number of observations table indicates the three missing values in Output 68.3.1.

Output 68.3.1 Number of Observations for the Modified thick Data Set

Kriging Prediction in the Presence of Missing Values

The KRIGE2D Procedure

Dependent Variable: Thick

Number of Observations Read	75
Number of Observations Used	72

Output 68.3.2 is a scatter plot of the modified observed data. The SHOWMISSING suboption produces marks in the observations plot that conveniently indicate the locations s_1, s_2, and s_3 of the missing values. Consequently, Output 68.3.2 displays an extended area with no observed Thick values in the central part of the domain.

Output 68.3.2 Scatter Plot of the Observations Spatial Distribution

Predictions at grid points with few neighboring data points rely heavily on the underlying covariance structure. The covariance model has a range of about 30,000 feet, which suggests that within this range a grid point might have no data neighbors at all and still obtain a prediction value on the basis of the correlation structure alone. This type of behavior is demonstrated in the Output 68.3.3, which shows a circular region in the center of the plot that has no data points. Predictions at the nodes in this area are mostly influenced by the covariance structure.

You can see the impact of this effect on the predictions if you compare the prediction contours in the Output 68.3.3 to the ones in Figure 68.4. Despite the contribution of the neighboring Thick data values to the predictions within the area of no observations, the outcome is clearly altered by the absence of observations at the locations s_1, s_2, and s_3.

Output 68.3.3 Surface Plot and Contours of Kriged Coal Seam Thickness

A noticeable difference is also apparent in the plot of the prediction standard errors. Output 68.3.4 displays these errors, and you can compare it to the standard error surface in Figure 68.4. The comparison shows a slight difference in the color gradient within the area of the missing data values. Output 68.3.4 uses standard error contours to enhance the effect of this difference.

The lack of information from the removed data results in an increase of the prediction uncertainty at the grid nodes that are most remotely situated from any observation in the central part of the domain. According to Output 68.3.4, the standard error at these nodes is almost comparable to the error observed near the borders of the domain, where the nodes of the prediction grid have relatively fewer data neighbors than other nodes in the domain.

Output 68.3.4 Surface Plot and Contours of Prediction Standard Errors

On a side note, PREDICTION plots display only observations with nonmissing values, as the plots in Output 68.3.3 and Output 68.3.4 demonstrate.

References

Chilès, J. P., and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. New York: John Wiley & Sons.

Christakos, G. (1992). *Random Field Models in Earth Sciences*. New York: Academic Press.

Cressie, N. (1993). *Statistics for Spatial Data*. Rev. ed. New York: John Wiley & Sons.

Deutsch, C. V., and Journel, A. G. (1992). *GSLIB: Geostatistical Software Library and User's Guide*. New York: Oxford University Press.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press.

Hohn, M. E. (1988). *Geostatistics and Petroleum Geology*. New York: Van Nostrand Reinhold.

Isaaks, E. H., and Srivastava, R. M. (1988). "Spatial Continuity Measures for Probabilistic and Deterministic Geostatistics." *Mathematical Geology* 20:313–341.

Journel, A. G. (1985). "The Deterministic Side of Geostatistics." *Mathematical Geology* 17:1–15.

Journel, A. G., and Huijbregts, C. J. (1978). *Mining Geostatistics*. New York: Academic Press.

Journel, A. G., and Rossi, M. (1989). "When Do We Need a Trend Model in Kriging?" *Mathematical Geology* 21:715–739.

Matheron, G. (1971). *The Theory of Regionalized Variables and Its Applications.* Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau. Paris: École Nationale Supérieure des Mines de Paris.

Robinson, G. K. (1991). "That BLUP Is a Good Thing: The Estimation of Random Effects." *Statistical Science* 6:15–51.

Subject Index

anisotropic
  models (KRIGE2D), 5040, 5041, 5043, 5046
  nugget effect (KRIGE2D), 5046
anisotropy
  factor (KRIGE2D), 5040
  geometric (KRIGE2D), 5040
  zonal (KRIGE2D), 5040
azimuth
  KRIGE2D procedure, 5041

correlation
  range (KRIGE2D), 5003
cubic semivariance model
  KRIGE2D procedure, 5022, 5034

estimation
  KRIGE2D procedure, 5003
exponential semivariance model
  KRIGE2D procedure, 5022, 5033

Gaussian semivariance model
  KRIGE2D procedure, 5022, 5032
geometric anisotropy
  KRIGE2D procedure, 5040, 5041
global kriging
  KRIGE2D procedure, 5003

KRIGE2D procedure
  anisotropic models, 5040, 5041, 5043, 5046
  azimuth, 5041
  best linear unbiased prediction (BLUP), 5050
  correlation range, 5003
  cubic semivariance model, 5022, 5034
  discontinuity, 5038
  effective range, 5003, 5032, 5033
  estimation, 5003
  examples, 5054, 5064, 5068
  exponential semivariance model, 5022, 5033
  Gaussian semivariance model, 5022, 5032
  geometric anisotropy, 5040, 5041
  global kriging, 5003
  input data set, 5010
  kriging with trend, 5048
  local kriging, 5002, 5003
  Matérn semivariance model, 5022, 5033
  modeling, 5002
  nested models, 5037, 5038
  nugget effect, 5024, 5038, 5039
  ODS graph names, 5053
  ODS Graphics, 5010
  ODS table names, 5053
  ordinary kriging, 5003, 5047–5050
  OUTEST= data sets, 5051
  OUTNBHD= data set, 5051, 5052
  output data sets, 5010, 5051, 5052
  pentaspherical semivariance model, 5022, 5035
  power semivariance model, 5022, 5024, 5036
  practical range, 5003, 5032, 5033
  prediction, 5003
  sill, 5033–5035
  sine hole effect semivariance model, 5022, 5035
  spatial continuity, 5003
  spatial covariance, 5003
  spatial data, 5047
  spatial prediction, 5002
  spatial random fields, 5047, 5048
  spherical semivariance model, 5022, 5034
  standard errors, 5003
  stochastic analysis, 5002
  uncertainty, 5002
  zonal anisotropy, 5043
KRIGE2D procedure, plots
  Observations, 5053
  Prediction, 5053
  Semivariogram, 5053
KRIGE2D procedure, tables
  Kriging Information, 5052, 5053
  Model Information, 5053
  Number of Observations, 5052
  Store Information, 5030, 5053
  Store Model Information, 5030, 5053
  Store Variables Information, 5030, 5053
kriging
  ordinary kriging (KRIGE2D), 5003, 5048
  with trend (KRIGE2D), 5048

local kriging
  KRIGE2D procedure, 5003

Matérn semivariance model
  KRIGE2D procedure, 5022, 5033
modeling
  KRIGE2D procedure, 5002

nested models
  KRIGE2D procedure, 5037, 5038
nugget effect
  KRIGE2D procedure, 5024, 5038, 5039

ODS graph names
KRIGE2D procedure, 5053
ODS Graphics
KRIGE2D procedure, 5010
ODS table names
KRIGE2D procedure, 5053
ordinary kriging
KRIGE2D procedure, 5047–5050
OUTEST= data sets
KRIGE2D procedure, 5051
OUTNBHD= data set
KRIGE2D procedure, 5051, 5052
output data sets
KRIGE2D procedure, 5010, 5051, 5052

pentaspherical semivariance model
KRIGE2D procedure, 5022, 5035
plots (KRIGE2D procedure)
Observations, 5053
Prediction, 5053
Semivariogram, 5053
power semivariance model
KRIGE2D procedure, 5022, 5024, 5036
prediction
at individual locations (KRIGE2D), 5016
correlation model (KRIGE2D), 5003, 5005, 5021, 5031, 5056
KRIGE2D procedure, 5003
on one-dimensional grid (KRIGE2D), 5016

range
correlation (KRIGE2D), 5003
effective (KRIGE2D), 5003, 5032, 5033
practical (KRIGE2D), 5003, 5032, 5033
range ϵ
KRIGE2D procedure, 5032

sill
KRIGE2D procedure, 5033–5035
SIM2D procedure, tables
Store Information, 5056
Store Model Information, 5056
Store Variables Information, 5056
sine hole effect semivariance model
KRIGE2D procedure, 5022, 5035
spatial continuity
KRIGE2D procedure, 5003
spatial dependence, *see* spatial continuity
spatial prediction
KRIGE2D procedure, 5002
spatial structure, *see* spatial continuity
spherical semivariance model
KRIGE2D procedure, 5022, 5034
standard errors
KRIGE2D procedure, 5003
stochastic analysis
KRIGE2D procedure, 5002
stochastic modeling, *see* modeling
stochastic spatial prediction, *see* spatial prediction

tables (KRIGE2D procedure)
Kriging Information, 5052, 5053
Model Information, 5053
Number of Observations, 5052
Store Information, 5030, 5053, 5056
Store Model Information, 5030, 5053, 5056
Store Variables Information, 5030, 5053, 5056

uncertainty
KRIGE2D procedure, 5002

zonal anisotropy
KRIGE2D procedure, 5043

Syntax Index

ANGLE= option
MODEL statement (KRIGE2D), 5022

BY statement
KRIGE2D procedure, 5015

COORDINATES statement
KRIGE2D procedure, 5016

DATA= option
PROC KRIGE2D statement, 5010

FORM= option
MODEL statement (KRIGE2D), 5022

GRID statement
KRIGE2D procedure, 5016
GRIDDATA= option
GRID statement (KRIGE2D), 5018

ID statement
KRIGE2D procedure, 5019
IDGLOBAL option
PROC KRIGE2D statement, 5010
IDNUM option
PROC KRIGE2D statement, 5010
INFO DETAILS option
RESTORE statement (KRIGE2D), 5030
INFO ONLY option
RESTORE statement (KRIGE2D), 5030
INFO option
RESTORE statement (KRIGE2D), 5030

KRIGE2D procedure, 5002
syntax, 5008
KRIGE2D procedure, BY statement, 5015
KRIGE2D procedure, COORDINATES statement, 5016
XCCORD= option, 5016
YCCORD= option, 5016
KRIGE2D procedure, GRID statement, 5016
GRIDDATA= option, 5018
LABEL= option, 5018
NPTS= option, 5016
X= option, 5017
XCOORD= option, 5018
Y= option, 5017
YCOORD= option, 5018
KRIGE2D procedure, ID statement, 5019
KRIGE2D procedure, MODEL statement, 5021
ANGLE= option, 5022
FORM= option, 5022
MDATA= option, 5023
NUGGET= option, 5024
POWNOBOUND option, 5024
RANGE= option, 5024
RATIO= option, 5025
SCALE= option, 5025
SINGULAR= option, 5025
SMOOTH= option, 5025
STORESELECT ANGLEID= option, 5027
STORESELECT MODEL= option, 5027
STORESELECT option, 5025
STORESELECT TYPE= option, 5025
KRIGE2D procedure, PREDICT statement, 5019
MAXPOINTS= option, 5019
MINPOINTS= option, 5020
NODECREMENT option, 5020
NOINCREMENT option, 5020
NUMPOINTS= option, 5020
RADIUS= option, 5020
VAR= option, 5021
KRIGE2D procedure, PROC KRIGE2D statement, 5009
DATA= option, 5010
IDGLOBAL option, 5010
IDNUM option, 5010
NOPRINT option, 5010
OUTEST= option, 5010
OUTNBHD= option, 5010
PLOTS option, 5010
PLOTS(ONLY) option, 5011
PLOTS=ALL option, 5011
PLOTS=EQUATE option, 5011
PLOTS=NONE option, 5011
PLOTS=OBSERVATIONS option, 5011
PLOTS=PREDICTION option, 5012
PLOTS=SEMIVARIOGRAM option, 5015
SINGULARMSG= option, 5015
KRIGE2D procedure, RESTORE statement, 5029
INFO DETAILS option, 5030
INFO ONLY option, 5030
INFO options, 5030

LABEL= option
GRID statement (KRIGE2D), 5018

MAX= option

PREDICT statement (KRIGE2D), 5019
MDATA= option
MODEL statement (KRIGE2D), 5023
MINPOINTS= option
PREDICT statement (KRIGE2D), 5020
MODEL statement
KRIGE2D procedure, 5021

NODECREMENT option
PREDICT statement (KRIGE2D), 5020
NOINCREMENT option
PREDICT statement (KRIGE2D), 5020
NOPRINT option
PROC KRIGE2D statement, 5010
NPTS= option
GRID statement (KRIGE2D), 5016
NUGGET= option
MODEL statement (KRIGE2D), 5024
NUMPOINTS= option
PREDICT statement (KRIGE2D), 5020

OUTEST= option
PROC KRIGE2D statement, 5010
OUTNBHD= option
PROC KRIGE2D statement, 5010

PLOTS option
KRIGE2D procedure, PROC KRIGE2D statement, 5010
PLOTS(ONLY) option
KRIGE2D procedure, PROC KRIGE2D statement, 5011
PLOTS=ALL option
KRIGE2D procedure, PROC KRIGE2D statement, 5011
PLOTS=EQUATE option
KRIGE2D procedure, PROC KRIGE2D statement, 5011
PLOTS=NONE option
KRIGE2D procedure, PROC KRIGE2D statement, 5011
PLOTS=OBSERVATIONS option
KRIGE2D procedure, PROC KRIGE2D statement, 5011
PLOTS=PREDICTION option
KRIGE2D procedure, PROC KRIGE2D statement, 5012
PLOTS=SEMIVARIOGRAM option
KRIGE2D procedure, PROC KRIGE2D statement, 5015
POWNOBOUND option
MODEL statement (KRIGE2D), 5024
PREDICT statement (KRIGE2D), 5019
PROC KRIGE2D statement, *see* KRIGE2D procedure

RADIUS= option
PREDICT statement (KRIGE2D), 5020
RANGE= option
MODEL statement (KRIGE2D), 5024
RATIO= option
MODEL statement (KRIGE2D), 5025
RESTORE statement (KRIGE2D), 5029

SCALE= option
MODEL statement (KRIGE2D), 5025
SINGULAR= option
MODEL statement (KRIGE2D), 5025
SINGULARMSG= option
PROC KRIGE2D statement, 5015
SMOOTH= option
MODEL statement (KRIGE2D), 5025
STORESELECT ANGLEID= option
MODEL statement (KRIGE2D), 5027
STORESELECT MODEL= option
MODEL statement (KRIGE2D), 5027
STORESELECT option
MODEL statement (KRIGE2D), 5025
STORESELECT TYPE= option
MODEL statement (KRIGE2D), 5025

VAR= option
PREDICT statement (KRIGE2D), 5021

X= option
GRID statement (KRIGE2D), 5017
XCOORD= option
COORDINATES statement (KRIGE2D), 5016
GRID statement (KRIGE2D), 5018

Y= option
GRID statement (KRIGE2D), 5017
YCOORD= option
COORDINATES statement (KRIGE2D), 5016
GRID statement (KRIGE2D), 5018