

# Chapter 34

## The KRIGE2D Procedure

### Chapter Table of Contents

---

<b>OVERVIEW</b> . . . . .	1709
Introduction to Spatial Prediction . . . . .	1709
<b>GETTING STARTED</b> . . . . .	1710
Spatial Prediction Using Kriging, Contour Plots . . . . .	1710
<b>SYNTAX</b> . . . . .	1713
PROC KRIGE2D Statement . . . . .	1715
COORDINATES Statement . . . . .	1715
GRID Statement . . . . .	1716
PREDICT Statement . . . . .	1717
MODEL Statement . . . . .	1718
<b>DETAILS</b> . . . . .	1721
Theoretical Semivariogram Models . . . . .	1721
The Nugget Effect . . . . .	1727
Anisotropic Models . . . . .	1729
Details of Ordinary Kriging . . . . .	1733
Output Data Sets . . . . .	1737
Computational Resources . . . . .	1738
<b>EXAMPLE</b> . . . . .	1739
Example 34.1 Investigating the Effect of Model Specification on Prediction . . . . .	1739
<b>REFERENCES</b> . . . . .	1743



# Chapter 34

## The KRIGE2D Procedure

---

### Overview

The KRIGE2D procedure performs ordinary kriging in two dimensions. PROC KRIGE2D can handle anisotropic and nested semivariogram models. Four semivariogram models are supported: the Gaussian, exponential, spherical, and power models. A single nugget effect is also supported.

You can specify the locations of kriging estimates 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 using a neighborhood of the data point established by the radius or number specification.

The KRIGE2D procedure writes the kriging estimates 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.

---

### Introduction to Spatial Prediction

Spatial prediction, in general, is any prediction method that incorporates spatial dependence. A simple and popular spatial prediction method is ordinary kriging.

Ordinary kriging requires a model of the spatial continuity, or dependence. This is typically in the form of a covariance or semivariogram.

Spatial prediction, then, involves two steps. First, you model the covariance or semivariogram of the spatial process. This involves choosing both a mathematical form and the values of the associated parameters. Second, you use this dependence model in solving 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 using ordinary kriging of two-dimensional data.

---

## Getting Started

---

### Spatial Prediction Using Kriging, Contour Plots

After an appropriate variogram model is chosen, there are a number of choices involved in producing the kriging surface. In order to illustrate these choices, the variogram model in the “Getting Started” section of Chapter 70, “The VARIOGRAM Procedure,” is used. 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  $c_0 = 7.5$  and a range of  $a_0 = 30$ .

The first choice is whether to use local or global kriging. Local kriging uses only data points in the neighborhood of a 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  $\epsilon$  is approximately equal to the length of the spatial domain. The correlation range  $\epsilon$  is the distance  $r_\epsilon$  at which the covariance is 5% of its value at zero. That is,

$$C_Z(r_\epsilon) = .05C_Z(0)$$

For a Gaussian model,  $r_\epsilon$  is  $\sqrt{3}a_0 \approx 52$  (thousand ft). The data points are scattered uniformly throughout a  $100 \times 100$  ( $10^6$  ft<sup>2</sup>) area. Hence, the linear dimension of the data is nearly double the  $\epsilon$  range. This indicates that local kriging rather than global kriging is appropriate.

Local kriging is performed by using only data points within a specified radius of each grid point. In this example, a radius of 60 (thousand ft) 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.

The last step in contouring the data is to decide on the grid point locations. A convenient area that encompasses all the data points is a square of length 100 (thousand ft). The spacing of the grid points depends on the use of the contouring; a spacing of five distance units (thousand ft) is chosen for plotting purposes.

The following SAS code inputs the data and computes the kriged surface using these parameter and grid choices. The kriged surface is plotted in Figure 34.1, and the associated standard errors are plotted in Figure 34.2. The standard errors are smaller where more data are available.

```

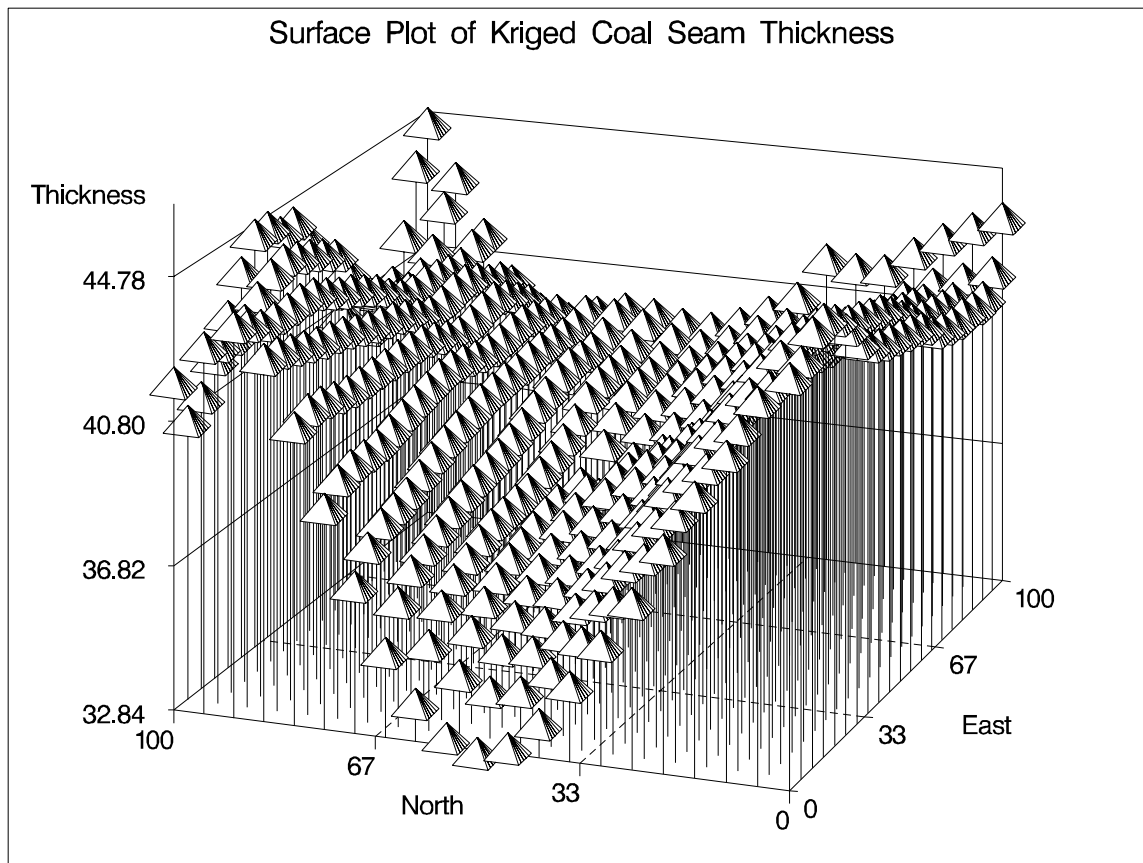
data thick;
  input east north thick @@;
  datalines;
    0.7 59.6 34.1 2.1 82.7 42.2 4.7 75.1 39.5
    4.8 52.8 34.3 5.9 67.1 37.0 6.0 35.7 35.9
    6.4 33.7 36.4 7.0 46.7 34.6 8.2 40.1 35.4
    13.3 0.6 44.7 13.3 68.2 37.8 13.4 31.3 37.8
    17.8 6.9 43.9 20.1 66.3 37.7 22.7 87.6 42.8
    23.0 93.9 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
    94.8 71.5 39.7 96.2 84.3 40.3 98.2 58.2 39.5
  ;

proc krige2d data=thick outest=est;
  pred var=thick r=60;
  model scale=7.5 range=30 form=gauss;
  coord xc=east yc=north;
  grid x=0 to 100 by 5 y=0 to 100 by 5;
run;

proc g3d data=est;
  title 'Surface Plot of Kriged Coal Seam Thickness';
  scatter gyc*gxc=estimate / grid;
  label gyc      = 'North'
        gxc      = 'East'
        estimate = 'Thickness'
  ;
run;

proc g3d data=est;
  title 'Surface Plot of Standard Errors of Kriging Estimates';
  scatter gyc*gxc=stderr / grid;
  label gyc      = 'North'
        gxc      = 'East'
        stderr   = 'Std Error'
  ;
run;

```



**Figure 34.1.** Surface Plot of Kriged Coal Seam Thickness

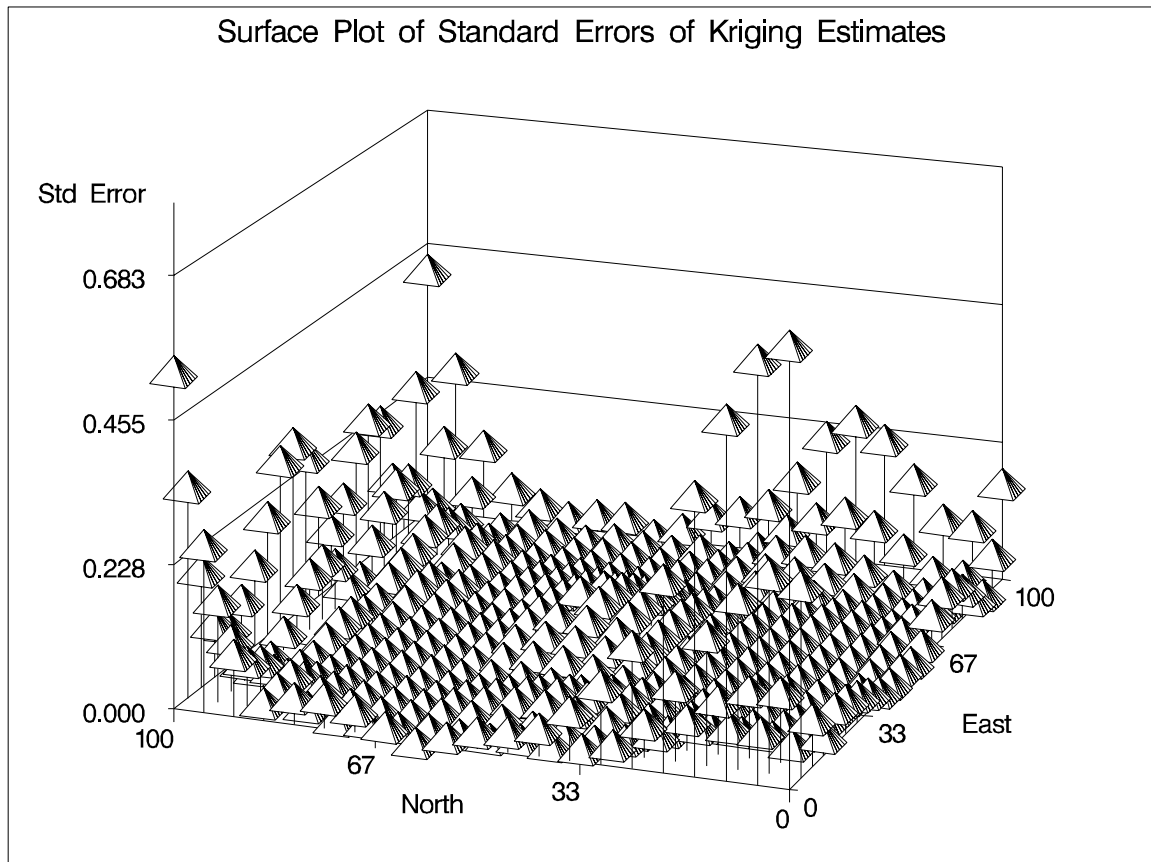


Figure 34.2. Surface Plot of Standard Errors of Kriging Estimates

## Syntax

The following statements are available in PROC KRIGE2D.

```

PROC KRIGE2D options ;
  COORDINATES | COORD coordinate-variables ;
  GRID grid-options ;
  PREDICT | PRED | P predict-options ;
  MODEL model-options ;

```

The PREDICT and MODEL statements are hierarchical; the PREDICT statement is followed by one or more MODEL statements. All the MODEL statements following a PREDICT statement use 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.

The following table outlines the options available in PROC KRIGE2D classified by function.

**Table 34.1.** Options Available in the KRIGE2D Procedure

Task	Statement	Option
<b>Data Set Options</b>		
specify input data set	PROC KRIGE2D	DATA=
specify grid data set	GRID	GDATA=
specify model data set	MODEL	MDATA=
write kriging estimates and standard errors	PROC KRIGE2D	OUTEST=
write neighborhood information for each grid point	PROC KRIGE2D	OUTNBHD=
<b>Declaring the Role of Variables</b>		
specify the variables to be estimated (kriged)	PREDICT	VAR=
specify the x and y coordinate variables in the DATA= data set	COORDINATES	XC= YC=
specify the x and y coordinate variables in the GDATA= data set	GRID	XC= YC=
<b>Controlling Kriging Neighborhoods</b>		
specify the radius of a neighborhood for all grid points	PREDICT	RADIUS=
specify the number of neighbors for all grid points	PREDICT	NUMPOINTS=
specify the maximum of neighbors for all grid points	PREDICT	MAXPOINTS=
specify the minimum of neighbors for all grid points	PREDICT	MINPOINTS=
specify action when maximum not met	PREDICT	NODECREMENT
specify action when minimum not met	PREDICT	NOINCREMENT
<b>Controlling the Semivariogram Model</b>		
specify a nugget effect	MODEL	NUGGET=
specify a functional form	MODEL	FORM=
specify a range parameter	MODEL	RANGE=
specify a scale parameter	MODEL	SCALE=
specify an angle for an anisotropic model	MODEL	ANGLE=
specify a minor-major axis ratio for an anisotropic model	MODEL	RATIO=



---

## PROC KRIGE2D Statement

### PROC KRIGE2D *options* ;

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

**DATA=SAS-data-set**

specifies a SAS data set containing the  $x$  and  $y$  coordinate variables and the VAR= variables in the PREDICT statement.

**OUTEST=SAS-data-set**

**OUTE=SAS-data-set**

specifies a SAS data set in which to store the kriging estimates, standard errors and grid location. For details, see the section “OUTEST=SAS-data-set” on page 1737.

**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 specifying local kriging. For details, see the section “OUTNBHD=SAS-data-set” on page 1737.

**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 SINGULARMSG= $n$  times. The default is SINGULARMSG=10.

---

## COORDINATES Statement

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

The following two options specify the names of the variables in the DATA= data set containing 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 containing the  $x$  coordinate of the data locations in the DATA= data set.

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

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

specifies the name of the variable containing the  $y$  coordinate of the data locations in the DATA= data set.

## GRID Statement

**GRID** *grid-options* ;

You can use the following options to specify the grid of spatial locations for the kriging estimates. The grid specification is applied to all PREDICT and MODEL statements.

There are two basic methods for specifying the grid. You can specify the  $x$  and  $y$  coordinates explicitly, or they can be read from a SAS data set. The options for the explicit specification of grid locations are as follows.

**X=number**

**X=** $x_1, \dots, x_m$

**X=** $x_1$  to  $x_m$

**X=** $x_1$  to  $x_m$  by  $\delta x$

specifies the  $x$  coordinate of the grid locations.

**Y=number**

**Y=** $y_1, \dots, y_m$

**Y=** $y_1$  to  $y_m$

**Y=** $y_1$  to  $y_m$  by  $\delta y$

specifies the  $y$  coordinate of the grid 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;
```

To specify grid locations from a SAS data set, you must give the name of the data set and the variables containing the values of the  $x$  and  $y$  coordinates.

**GRIDDATA=***SAS-data-set*

**GDATA=***SAS-data-set*

specifies a SAS data set containing the  $x$  and  $y$  grid coordinates.

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

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

specifies the name of the variable containing the  $x$  coordinate of the grid locations in the GRIDDATA= data set.

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

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

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

---

## PREDICT Statement

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

You can specify the following options in a PREDICT statement.

**MAXPOINTS=***number*

**MAXPOINTS=***number*

**MAXP=***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.

**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. If enough data are available, this value should be increased to 30 to improve estimation.

**NODECREMENT | NODECR**

requests that the RADIUS= value not be decremented when the MAX= 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, estimation 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 MIN= 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, estimation 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 controlling the neighborhood; it must appear by itself.

**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 consisting of all data points within the distance specified by the RADIUS= value. See the MAXPOINTS= and MINPOINTS= options for additional control on the neighborhood.

**VAR**= *variable-name*

specifies the single numeric variable used in the kriging system.

---

## MODEL Statement

**MODEL** *model-options* ;

You can use the following options to specify a semivariogram or covariance model. The specified model is used in the kriging system defined by the most previous PREDICT statement.

There are two ways to specify a semivariogram or covariance model. In the first method, you specify the required parameters SCALE, RANGE, and FORM, and possibly the optional parameters NUGGET, ANGLE, and RATIO, explicitly in the MODEL statement.

In the second method, you specify an MDATA= data set. This data set contains variables corresponding to the required SCALE, RANGE, and FORM parameters, and, optionally, variables for the NUGGET, ANGLE, and RATIO parameters.

The two methods are exclusive; either you specify all parameters explicitly, or they all are read from the MDATA= data set.

**ANGLE**=*angle*

**ANGLE**=(*angle*<sub>1</sub>, . . . , *angle*<sub>*k*</sub>)

specifies the angle of the major axis for anisotropic models, measured in degrees clockwise from the N-S axis. In the case of a nested semivariogram model, you can specify an angle for each nesting. The default is ANGLE=0.

**FORM=SPHERICAL | EXPONENTIAL | GAUSSIAN | POWER**  
**FORM=SPH | EXP | GAUSS | PW**

specifies the functional form of the semivariogram model. All the supported models are two-parameter models (SCALE= and RANGE=). A FORM= value is required; in the case of a nested semivariogram model, you must specify a form for each nesting.

See the section “Theoretical Semivariogram Models” beginning on page 1721 for details on 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= data set must contain variables named SCALE, RANGE, and FORM, and it can optionally contain variables NUGGET, ANGLE, and RATIO.

The FORM variable must be a character variable, assuming only the values allowed in the explicit FORM= syntax described previously. The RANGE and SCALE 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 semivariogram model. For example, to specify a nonnested model using a spherical covariance, an MDATA= data set might look like

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

The PROC KRIGE2D statement to use the MDATA= specification is of the form

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

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

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

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

```
data mdl;
  input scale range form $ nugget angle ratio;
  datalines;
  20 8 S 5 35 0.7
  12 3 G 5 0 0.8
  4 1 G 5 45 0.5
  ;
```

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

```
proc krige2d data=...;
  pred var=...;
  model scale=(20,12,4) range=(8,3,1) form=(S,G,G)
        angle=(35,0,45) ratio=(0.7,0.8,0.5) nugget=5;
run;
```

This example is somewhat artificial in that it is usually hard to detect different anisotropy directions and ratios for different nestings using an experimental semivariogram. Note that the NUGGET 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 “Theoretical and Computational Details of the Semivariogram” on page 3664 in the chapter on the VARIOGRAM procedure.

**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 (see the chapter on the VARIOGRAM procedure for details). For models without any nugget effect, this option is left out; the default is NUGGET=0.

**RANGE=range**

**RANGE=(range<sub>1</sub>,...,range<sub>k</sub>)**

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

The range parameter is the divisor in the exponent in all supported models except the power model. It has the units of distance or distance squared for these models, and it is related to the correlation scale for the underlying spatial process. See the section “Theoretical Semivariogram Models” beginning on page 1721 for details on how the RANGE= values are determined.

**RATIO=ratio**

**RATIO=(ratio<sub>1</sub>,...,ratio<sub>k</sub>)**

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

**SCALE=***scale*

**SCALE=**(*scale*<sub>1</sub>, . . . , *scale*<sub>*k*</sub>)

specifies the scale parameter in semivariogram models. In the case of a nested semivariogram model, 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. See the section “Theoretical Semivariogram Models” beginning on page 1721 for details on how the SCALE= values are determined.

**SINGULAR=***number*

gives the singularity criteria 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 “Details of Ordinary Kriging” beginning on page 1733 for more detailed information.

---

## Details

---

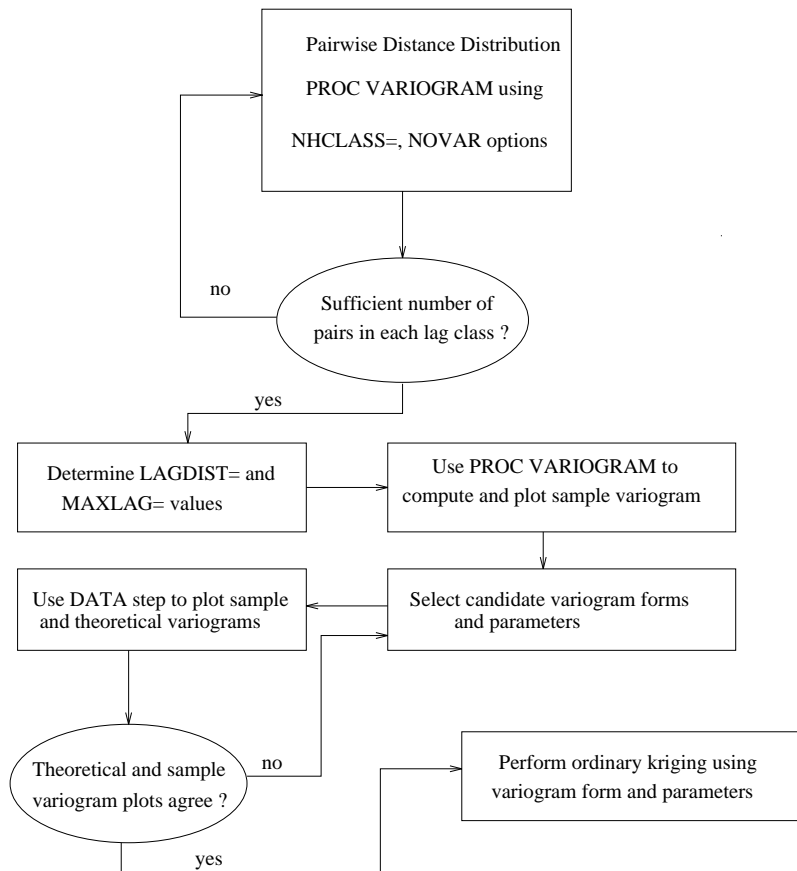
### Theoretical Semivariogram Models

PROC VARIOGRAM computes the sample, or experimental semivariogram. Prediction of the spatial process 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. While there are various methods of fitting semivariogram models, such as least squares, maximum likelihood, and robust methods (Cressie 1993, section 2.6), these techniques are not appropriate for data sets resulting in a small number of variogram points. Instead, a visual fit of the variogram points to a few standard models is often satisfactory. Even when there are sufficient variogram points, a visual check against a fitted theoretical model is appropriate (Hohn 1988, p. 25ff).

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.

Note the general flow of investigation. After a suitable choice is made of the LAGDIST= and MAXLAG= options and, possibly, the NDIR= option (or a DIRECTIONS statement), the experimental semivariogram is computed. Potential theoretical models, possibly incorporating nesting, anisotropy, and the nugget effect, are computed by a DATA step, then they are plotted against the experimental semivariogram and evaluated. A suitable theoretical model is thus found visually, and the specification of the model is used in PROC KRIGE2D. This flow is illustrated in Figure 34.3; also see the “Getting Started” section on page 3644 in the chapter on the VARIOGRAM procedure for an illustration in a simple case.



**Figure 34.3.** Flowchart for Variogram Selection

Four theoretical models are supported by PROC KRIGE2D: the spherical, Gaussian, exponential, and power models. For the first three types, the parameters  $a_0$  and  $c_0$ , corresponding to the RANGE= and SCALE= options in the MODEL statement in PROC KRIGE2D, have the same dimensions and have similar effects on the shape of  $\gamma_z(h)$ , as illustrated in the following paragraph.

In particular, the dimension of  $c_0$  is the same as the dimension of the variance of the spatial process  $\{Z(r), r \in D \subset \mathcal{R}^2\}$ . The dimension of  $a_0$  is length with the same units as  $h$ .

These three model forms are now examined in more detail.

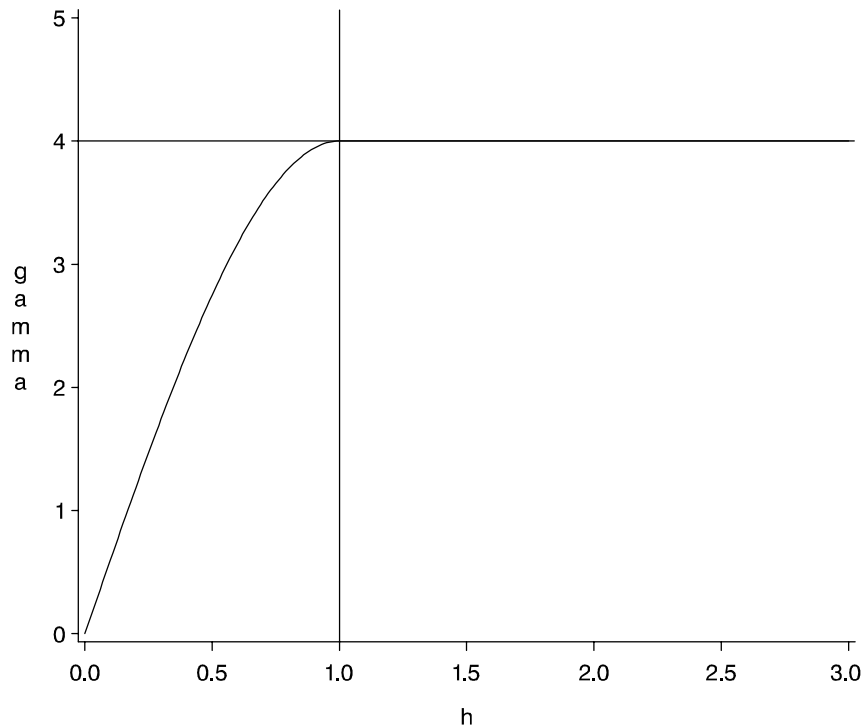


### The Spherical Semivariogram Model

The form of the spherical model is

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

The shape is displayed in Figure 34.4 using range  $a_0 = 1$  and scale  $c_0 = 4$ .



**Figure 34.4.** Spherical Semivariogram Model with Parameters  $a_0 = 1$  and  $c_0 = 4$

The vertical line at  $h = 1$  is the “effective range” as defined by Duetsch and Journel (1992), or the “range  $\epsilon$ ” defined by Christakos (1992). This quantity, denoted  $r_\epsilon$ , is the  $h$ -value where the covariance is approximately zero. For the spherical model, it is *exactly* zero; for the Gaussian and exponential models, the definition of  $r_\epsilon$  is modified slightly.

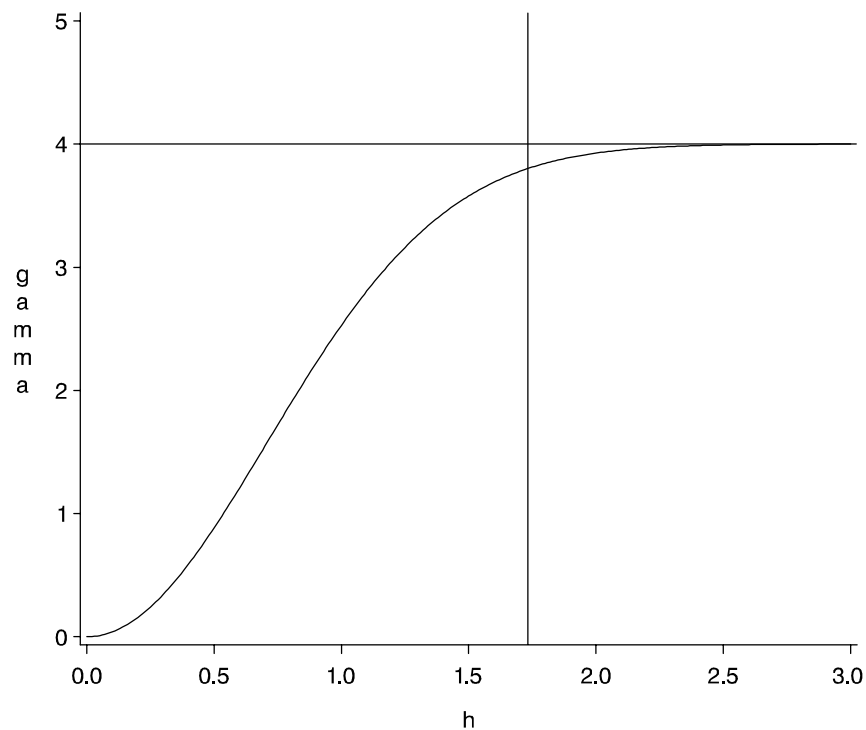
The horizontal line at 4.0 variance units (corresponding to  $c_0 = 4$ ) is called the “sill.” In the case of the spherical model,  $\gamma_z(h)$  actually reaches this value. For the other two model forms, the sill is a horizontal asymptote.

### 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 34.5 using range  $a_0 = 1$  and scale  $c_0 = 4$ .



**Figure 34.5.** Gaussian Semivariogram Model with Parameters  $a_0 = 1$  and  $c_0 = 4$

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

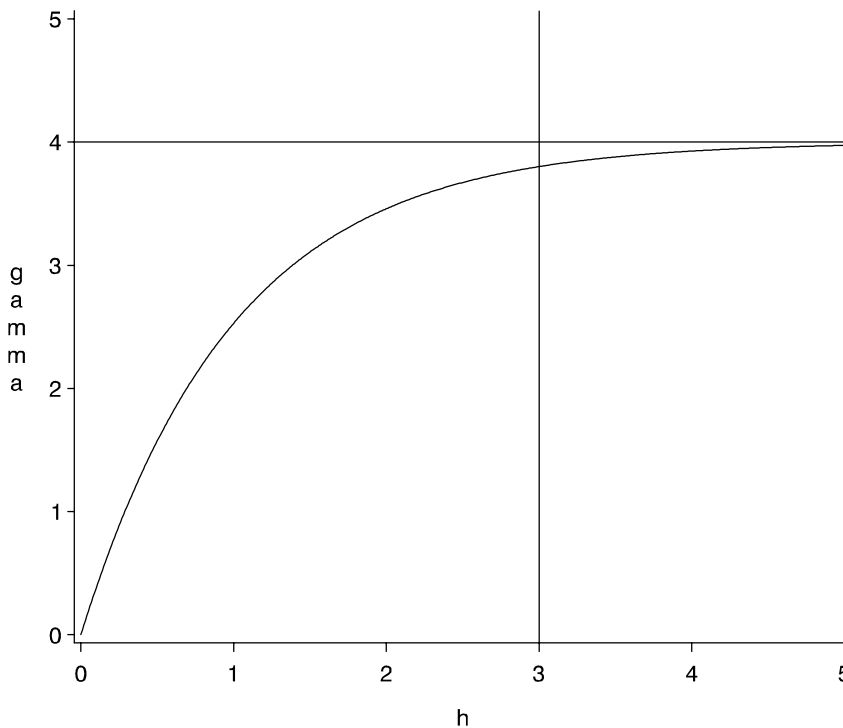
The horizontal line at 4.0 variance units (corresponding to  $c_0 = 4$ ) is the sill;  $\gamma_z(h)$  approaches the sill asymptotically.

### **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 34.6 using range  $a_0 = 1$  and scale  $c_0 = 4$ .



**Figure 34.6.** Exponential Semivariogram Model with Parameters  $a_0 = 1$  and  $c_0 = 4$

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

The horizontal line at 4.0 variance units (corresponding to  $c_0 = 4$ ) is the sill, as in the other model forms.

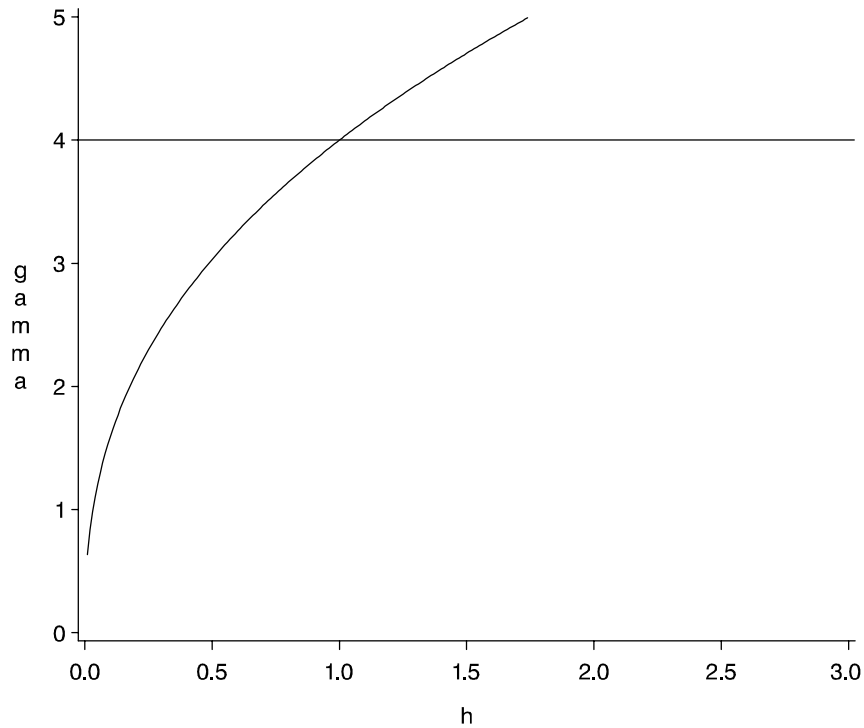
It is noted from Figure 34.5 and Figure 34.6 that the major distinguishing feature of the Gaussian and exponential forms is the shape in the neighborhood of the origin  $h = 0$ . In general, small lags are important in determining an appropriate theoretical form based on a sample semivariogram.

### **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 a dimensionless quantity, with typical values  $0 < a_0 < 2$ . Note that the value of  $a_0 = 1$  yields a straight line. The parameter  $c_0$  has dimensions of the variance, as in the other models. There is no sill for the power model. The shape of the power model with  $a_0 = 0.4$  and  $c_0 = 4$  is displayed in Figure 34.7.



**Figure 34.7.** Power Semivariogram Model with Parameters  $a_0 = 0.4$  and  $c_0 = 4$

### **Nested Models**

For a given set of spatial data, a plot of an experimental semivariogram may not seem to fit any one of the theoretical models. In such a case, the covariance structure of the spatial process may be a sum of two or more covariances. This is common in geologic applications where there are correlations at different length scales. At small lag distances  $h$ , the smaller scale correlations dominate, while the large scale correlations dominate at larger lag distances.

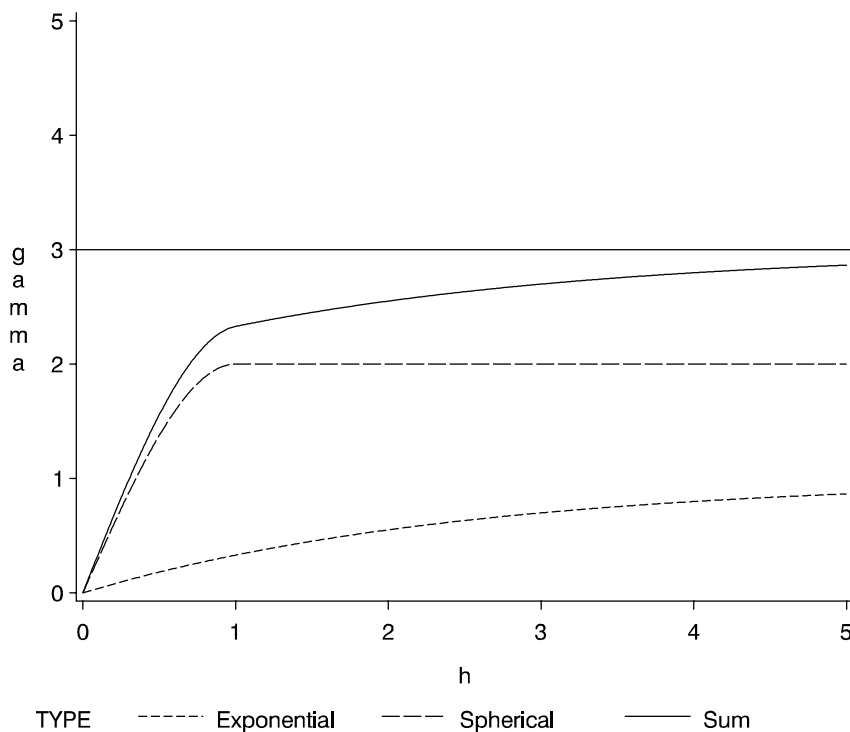
As an illustration, consider two semivariogram models, an exponential and a spherical.

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

and

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

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  $\{Z(r), r \in D\}$ , then a plot of the experimental semivariogram would resemble the sum of these two semivariograms. This is illustrated in Figure 34.8.



**Figure 34.8.** Sum of Exponential and Spherical Structures at Different Scales

This sum of  $\gamma_1(h)$  and  $\gamma_2(h)$  in Figure 34.8 does not resemble any *single* theoretical semivariogram; however, the shape at  $h = 1$  is similar to a spherical. The asymptotic approach to a sill at three variance units, along with the shape around  $h = 0$ , indicates an exponential structure. Note that the sill value is the sum of the individual sills  $c_{0,1} = 1$  and  $c_{0,2} = 2$ .

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

---

## The Nugget Effect

For all the variogram 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 may 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

$$h_{min} = \min_{i,j} h_{ij} = \min_{i,j} |r_i - r_j|$$

There are conceptual and theoretical difficulties associated with a nonzero nugget effect; refer to 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  $h \rightarrow 0$ . The importance of availability of data at small lag distances is again illustrated.

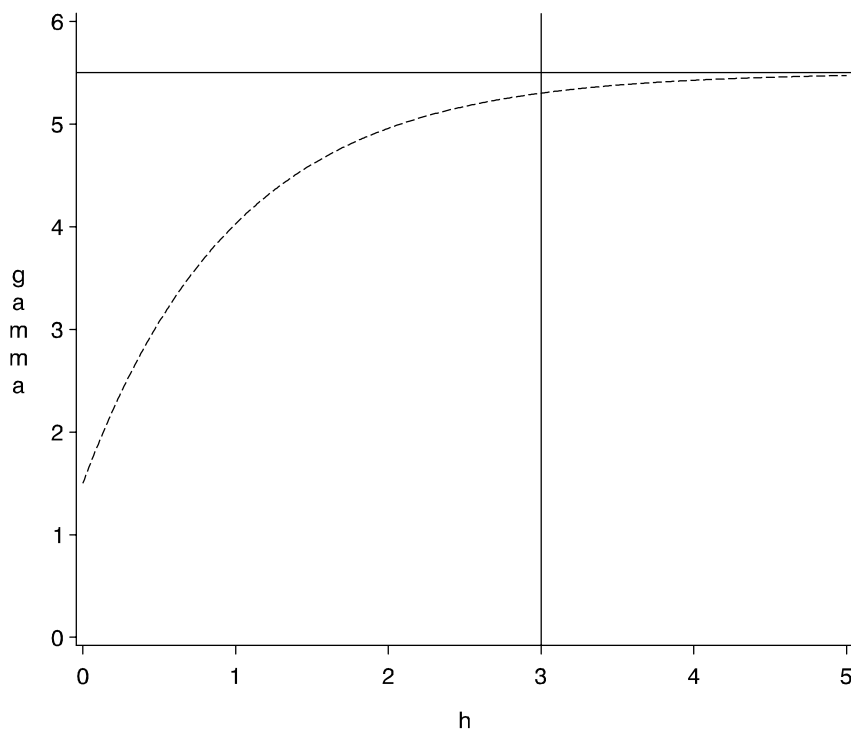
As an example, an exponential semivariogram with a nugget effect  $c_n$  has the form

$$\gamma_z(h) = c_n + c_0 \left[ 1 - \exp\left(-\frac{h}{a_0}\right) \right], h > 0$$

and

$$\gamma_z(0) = 0$$

This is illustrated in Figure 34.9 for parameters  $a_0 = 1$ ,  $c_0 = 4$ , and nugget effect  $c_n = 1.5$ .



**Figure 34.9.** Exponential Semivariogram Model with a 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. There is a way to approximate an anisotropic nugget effect; this is described in the following section.

---

## Anisotropic Models

In all the theoretical models considered previously, the lag distance  $h$  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  $h = |P_1 P_2|$ , not on the orientation of the two points. A spatial process  $\{Z(r), r \in D\}$  with this property is called isotropic, as is the associated covariance or semivariogram.

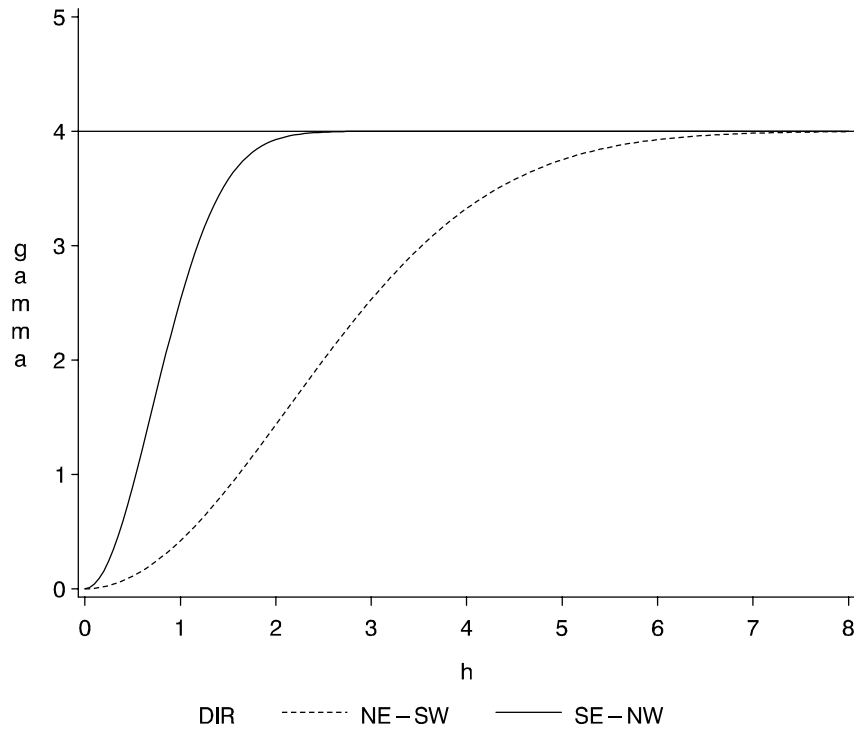
However, real spatial phenomena often show directional effects. Particularly in geologic applications, measurements along a particular direction may be highly correlated, while the perpendicular direction shows little or no correlation. Such processes are called anisotropic. Refer to Journel and Huijbregts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

There are two types of anisotropy. The simplest type occurs when the same covariance *form* and scale parameter  $c_0$  is present in all directions but the range  $a_0$  changes with direction. In this case, there is a single sill, but the semivariogram reaches the sill in a shorter lag distance along a certain direction.

This type of anisotropy is called “geometric” and is discussed in the following section.

### **Geometric Anisotropy**

Geometric anisotropy is illustrated in Figure 34.10, where an anisotropic Gaussian semivariogram is plotted. The two curves displayed in this figure are generated using  $a_0 = 1$  in the NE–SW direction and  $a_0 = 3$  in the SE–NW direction.



**Figure 34.10.** Geometric Anisotropy with Major Axis along E–W Direction

As you can see from the figure, the SE–NW curve gets “close” to the sill at approximately  $h = 2$ , while the NE–SW curve does so at  $h = 6$ . The ratio of the shorter to longer distance is  $\frac{2}{6} = \frac{1}{3}$ . This is the value to use in the `RATIO=` parameter in the `MODEL` statement in `PROC KRIGE2D`. Since the longer, or major, axis is in the NE–SW direction, the `ANGLE=` parameter in the `MODEL` statement in `PROC KRIGE2D` should be  $45^\circ$  (all angles are measured clockwise from north).

The terminology associated with geometric anisotropy is that of ellipses. To see how this comes about, consider the following hypothetical set of calculations.

Let  $\{Z(r), r \in D\}$  be a geometrically anisotropic process, and assume that there are sufficient data points to calculate an experimental semivariogram at a large number of angle classes  $\theta \in \{0, \delta\theta, 2\delta\theta, \dots, 180^\circ\}$ . At each of these angles  $\theta$ , the experimental semivariogram is plotted and the effective range  $r_\epsilon$  is recorded. A diagram, in polar coordinates, of  $(r_\epsilon, \theta)$  yields an ellipse, with the major axis in the direction of the largest  $r_\epsilon$  and the minor axis perpendicular. Denote the largest  $r_\epsilon$  by  $r_\epsilon^{max}$ , the smallest by  $r_\epsilon^{min}$ , and their ratio by

$$R = \frac{r_\epsilon^{min}}{r_\epsilon^{max}}$$



By a rotation, a new set of axes are aligned along the major and minor axis. Then, a rescaling elongates the minor axis so its length equals that of the major axis of the ellipse.

First, the angle  $\theta$  of the major axis of the ellipse (measured clockwise from north) is transformed to standard Cartesian orientation or counter-clockwise from the x-axis (east). Let  $\varphi = 90^\circ - \theta$  denote the transformed angle. The matrix to transform the distance  $h$  is in terms of  $\varphi$  and the ratio  $R$  and it is given by

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

For a given point pair  $P_1P_2$ , with coordinates  $(x_1, y_1), (x_2, y_2)$ , the transformed interpair distance is computed by first transforming the components  $\delta x = x_1 - x_2$  and  $\delta y = y_1 - y_2$  by

$$\begin{pmatrix} \delta x' \\ \delta y' \end{pmatrix} = \mathbf{H} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix}$$

The transformed interpair distance is then

$$h' = \sqrt{(\delta x')^2 + (\delta y')^2}$$

The original semivariogram, a function of *both*  $h$  and  $\theta$ , is then transformed to a function only of  $h'$ :

$$\hat{\gamma}(h') = \gamma(h, \theta)$$

This single semivariogram is then used for kriging purposes.

The interpretation of the major/minor axis in the case of geometric anisotropy is that the direction of the major axis is the direction in which the spatial process  $\{Z(r), r \in D\}$  is most highly correlated; the process is least correlated in the perpendicular direction.

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 four to six directions is usually sufficient to determine the presence of anisotropy and find the approximate major/minor axis directions.

The most convenient way of performing this is to use the NDIR= option in the COMPUTE statement in PROC VARIOGRAM to obtain a separate experimental semivariogram for each direction. After determining the direction of the major axis, use a DIRECTIONS statement on a subsequent run of PROC VARIOGRAM with this direction and its perpendicular direction. For example, if the initial run of PROC VARIOGRAM with NDIR=6 in the COMPUTE statement indicates that  $\theta = 45^\circ$  is the major axis (has the largest  $r_\epsilon$ ), then rerun PROC VARIOGRAM with

```
DIRECTIONS 45,135;
```

Then, determine the ratio of  $r_\epsilon$  for the minor and major axis for the RATIO= parameter in the COMPUTE statement of PROC KRIGE2D. This ratio is  $\leq 1$  for modeling geometric anisotropy. In the other type of anisotropy, *zonal* anisotropy, the RATIO= parameter is set to a large number for reasons explained in the following section.

### Zonal Anisotropy

In zonal anisotropy, either the *form* covariance structure or the parameter  $c_0$  (or both) is different in different directions. In particular, the sill is different for different directions. In geologic applications, this is the more common type of anisotropy. 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. For example, suppose the spatial process has a correlation structure in the N–S direction described by  $\gamma_{z,1}$ , a spherical model with sill at  $c_0 = 6$  and range  $a_0 = 2$ , while in the E–W direction the correlation structure, described by  $\gamma_{z,2}$ , is again a spherical model but with sill at  $c_0 = 3$  and range  $a_0 = 1$ .

You can approximate this structure in PROC KRIGE2D by specifying two nested models with large RATIO= values. In particular, the appropriate MODEL statement is

```
MODEL FORM=(S,S) ANGLE=(0,90) SCALE=(6,3)
      RANGE=(2,1) RATIO=(1E8,1E8);
```

The large values of the RATIO= parameter for each nested structure have the effect of an “infinite” range parameter in the direction of the minor axis. Hence, there is no variation in  $\gamma_{z,1}$  in the E–W direction and no variation in  $\gamma_{z,2}$  in the N–S direction.

### Anisotropic Nugget Effect

Note that an isotropic nugget effect can be approximated by using nested models, with one of the nested models having a small range. Applying a geometric anisotropy specification to this nested structure results in an anisotropic nugget effect.

---

## Details of Ordinary Kriging

### Introduction

There are three common characteristics 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 ground-water level, so you can ignore any time dependence.

For you to find the areal extent and the concentration values of the solvent, measurements are required. 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, there are small-scale variations in the measured concentrations, 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, and there are a number of numerical smoothing techniques, such as inverse distance weighting and splines, that 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 may be situations, due to the use of the prediction map or due to the relative magnitude of the irregular fluctuations, where you cannot ignore these small-scale irregular fluctuations. In other words, the smoothed or estimated 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, for example, Cressie (1993) and Christakos (1992). The mathematics of SRFs are formidable. However, under certain simplifying assumptions, they produce classical linear estimators with very simple properties, allowing easy implementation for prediction purposes. These estimators, 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.

There are additional logical and methodological problems in applying a stochastic model to a unique but partly unknown natural process; refer to the introduction in Matheron (1971) and Cressie (1993, section 2.3). These difficulties have resulted in attempts to frame the estimation problem in a completely deterministic way (Isaaks and Srivastava 1988; Journel 1985).

Additional problems with kriging, and with spatial estimation 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 estimator for point estimation and the associated standard error is discussed in the following section. Full details are given in Journel and Huijbregts (1978), Christakos (1992), and Cressie (1993).

### **Ordinary Kriging**

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

$$Z(r) = \mu + \varepsilon(r)$$

Here,  $\mu$  is the fixed, unknown mean of the process, and  $\varepsilon(r)$  is a zero mean SRF representing the variation around the mean.

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

$$C_z(r_1, r_2) = E[\varepsilon(r_1)\varepsilon(r_2)] = C_z(r_1 - r_2)$$

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  $\varepsilon(r_1) - \varepsilon(r_2)$  rather than  $\varepsilon(r)$ :

$$\gamma_Z(r_1, r_2) = \frac{1}{2} E[\varepsilon(r_1) - \varepsilon(r_2)]^2 = \gamma_Z(r_1 - r_2)$$

By performing local kriging, the spatial processes represented by the previous equation for  $Z(r)$  are more general than they appear. In local kriging, at an unsampled location  $r_0$ , a separate model is fit using only data in a neighborhood of  $r_0$ . This has the effect of fitting a separate mean  $\mu$  at each point, and it is similar to the “kriging with trend” (KT) method discussed in Journel and Rossi (1989).

Given the  $N$  measurements  $Z(r_1), \dots, Z(r_N)$  at known locations  $r_1, \dots, r_N$ , you want to obtain an estimate  $\hat{Z}$  of  $Z$  at an unsampled location  $r_0$ . When the following three requirements are imposed on the estimator  $\hat{Z}$ , the OK estimator is obtained.

- (i)  $\hat{Z}$  is linear in  $Z(r_1), \dots, Z(r_N)$ .
- (ii)  $\hat{Z}$  is unbiased.
- (iii)  $\hat{Z}$  minimizes the mean-square prediction error  $E(Z(r_0) - \hat{Z}(r_0))^2$ .

Linearity requires the following form for  $\hat{Z}(r_0)$ :

$$\hat{Z}(r_0) = \sum_{i=1}^N \lambda_i Z(r_i)$$

Applying the unbiasedness condition to the preceding equation yields

$$E\hat{Z}(r_0) = \mu \Rightarrow \mu = \sum_{i=1}^N \lambda_i E Z(r_i) \Rightarrow$$

$$\sum_{i=1}^N \lambda_i \mu = \mu \Rightarrow \sum_{i=1}^N \lambda_i = 1$$

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

$$L = E \left( Z(r_0) - \sum_{i=1}^N \lambda_i Z(r_i) \right)^2 - 2m \left( \sum_{i=1}^N \lambda_i - 1 \right)$$

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

$$\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)^T$$

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

$$\boldsymbol{\lambda}_0 = (\lambda_1, \dots, \lambda_N, m)^T = \begin{pmatrix} \boldsymbol{\lambda} \\ m \end{pmatrix}$$

The optimization is performed by solving

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

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

The resulting matrix equation can be expressed in terms of either the covariance  $C_z(r)$  or semivariogram  $\gamma_z(r)$ . In terms of the covariance, the preceding equation results in the following matrix equation:

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

where

$$\mathbf{C} = \begin{pmatrix} C_z(0) & C_z(r_1 - r_2) & \cdots & C_z(r_1 - r_N) & 1 \\ C_z(r_2 - r_1) & C_z(0) & \cdots & C_z(r_2 - r_N) & 1 \\ & & \ddots & & \\ C_z(r_N - r_1) & C_z(r_N - r_2) & \cdots & C_z(0) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix}$$

and

$$\mathbf{C}_0 = \begin{pmatrix} C_z(r_0 - r_1) \\ C_z(r_0 - r_2) \\ \vdots \\ C_z(r_0 - r_N) \\ 1 \end{pmatrix}$$

The solution to the previous matrix equation is

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

Using this solution for  $\lambda$  and  $m$ , the ordinary kriging estimate at  $r_0$  is

$$\hat{Z}(r_0) = \lambda_1 Z(r_1) + \cdots + \lambda_N Z(r_N)$$

with associated prediction error

$$\sigma_z(r_0) = C_z(0) - \lambda' \mathbf{c}_0 + m$$

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, Duetsch 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.

---

## 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 estimates and the associated standard errors. The `OUTEST=` data set contains the following variables:

- `ESTIMATE`, which is the kriging estimate for the current variable
- `GXC`, which is the x-coordinate of the grid point at which the kriging estimate is made
- `GYC`, which is the y-coordinate of the grid point at which the kriging estimate is made
- `LABEL`, which is the label for the current `PREDICT/MODEL` combination producing the kriging estimate. If you do not specify a label, default labels of the form `Predj.Modelk` are used.
- `NPOINTS`, which is the number of points used in the estimation. This number varies for each grid point if local kriging is performed.
- `STDERR`, which is the standard error of the kriging estimate
- `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
- LABEL, which is the label for the current PREDICT/MODEL combination. If you do not specify a label, default labels of the form Predj.Modelk are used.
- NPOINTS, which is the number of points used in the estimation
- 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

---

## Computational Resources

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

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

where  $\mathbf{C}$  is  $(N + 1) \times (N + 1)$ , and the right-hand side vector  $\mathbf{C}_0$  is  $(N + 1) \times 1$ .

Holding the matrix and vector associated with this system in core requires approximately  $\frac{N^2}{2}$  doubles (with typically eight bytes per double). The CPU time used in solving the system is proportional to  $N^3$ . For large  $N$ , this time dominates the time to compute the  $\frac{(N+1)(N+2)}{2}$  elements of the covariance matrix  $\mathbf{C}$  from the specified covariance or variogram model. This latter computation is proportional to  $N^2$ .

For local kriging, the kriging system is set up and solved for each grid point. Part of the set up process involves determining the neighborhood of each grid point. A fast K-D tree algorithm is used to determine neighborhoods. For  $G$  grid points, the dominant CPU time factor is setting up and solving the  $G$  kriging systems. The  $N$  in the preceding algorithm 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 are used to set up and solve the single system

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



Again  $\mathbf{C}$  is  $(N + 1) \times (N + 1)$ , but  $\boldsymbol{\lambda}_0$  is now  $(N + 1) \times G$ , where  $G$  is the number of grid points, and  $N$  is the number of nonmissing observations in the input data set. The right-hand side matrix  $\mathbf{C}_0$  is  $(N + 1) \times G$ . Memory requirements are approximately  $\frac{N^2}{2} + GN$  doubles. The CPU time used in solving the system is still dominated by the  $N^3$  factorization of the left-hand side.

---

## Example

---

### Example 34.1. Investigating the Effect of Model Specification on Prediction

In the “Getting Started” section of the chapter on the VARIOGRAM procedure, a particular variogram is chosen for the coal seam thickness data. The chosen variogram is Gaussian with a scale (sill) of  $c_0 = 7.5$ , and a range of  $a_0 = 30$ . This choice of the variogram is based on a visual fit—a comparison of the plots of the regular and robust sample variograms and the Gaussian variogram for various scale (sill) and range values.

Another possible choice of model is the spherical variogram with the same scale (sill) of  $c_0 = 7.5$  but with a range of  $a_0 = 60$ . This choice of range is again based on a visual fit; while not as good as the Gaussian model, the fit is reasonable.

It is generally held 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.

```

data thick;
  input east north thick @@;
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
94.8  71.5  39.7  96.2  84.3  40.3  98.2  58.2  39.5
;

```

```

/*- Run KRIGE2D on original Gaussian model -----*/
proc krige2d data=thick outest=est1;
  pred var=thick r=60;
  model scale=7.5 range=30 form=gauss;
  coord xc=east yc=north;
  grid x=0 to 100 by 10 y=0 to 100 by 10;
run;

/*- Run KRIGE2D using Spherical Model, modified range -*/
proc krige2d data=thick outest=est2;
  pred var=thick r=60;
  model scale=7.5 range=60 form=spherical;
  coord xc=east yc=north;
  grid x=0 to 100 by 10 y=0 to 100 by 10;
run;

data compare ;
  merge est1(rename=(estimate=g_est stderr=g_std))
        est2(rename=(estimate=s_est stderr=s_std));
  est_dif=g_est-s_est;
  std_dif=g_std-s_std;
run;

proc print data=compare;
  title 'Comparison of Gaussian and Spherical Models';
  title2 'Differences of Estimates and Standard Errors';
  var gxc gyc npoints g_est s_est est_dif g_std s_std
      std_dif;
run;

```

**Output 34.1.1. Comparison of Gaussian and Spherical Models**

Comparison of Gaussian and Spherical Models Differences of Estimates and Standard Errors									
Obs	GXC	GYC	NPOINTS	g_est	s_est	est_dif	g_std	s_std	std_dif
1	0	0	23	43.9408	42.6700	1.27087	0.68260	2.05947	-1.37687
2	0	10	28	41.6828	41.6780	0.00483	0.55909	2.03464	-1.47554
3	0	20	31	38.9601	39.7285	-0.76843	0.30185	1.93478	-1.63293
4	0	30	32	36.1701	37.3275	-1.15739	0.12705	1.54844	-1.42139
5	0	40	39	33.8376	35.4320	-1.59440	0.04872	1.37821	-1.32949
6	0	50	38	32.8375	34.3930	-1.55550	0.02983	1.22584	-1.19602
7	0	60	35	33.9576	34.3155	-0.35785	0.00195	0.54122	-0.53927
8	0	70	30	36.9502	37.6669	-0.71673	0.04006	1.20451	-1.16444
9	0	80	31	41.1097	41.1016	0.00812	0.04705	0.99544	-0.94839
10	0	90	28	43.6671	42.5216	1.14546	0.10236	1.57357	-1.47121
11	0	100	23	41.9443	42.6511	-0.70681	0.53646	2.20792	-1.67146
12	10	0	25	44.6795	44.1959	0.48355	0.07833	1.09743	-1.01910
13	10	10	31	42.8397	42.7496	0.09008	0.10982	1.46686	-1.35703
14	10	20	34	40.3120	40.3634	-0.05140	0.05315	1.54889	-1.49574
15	10	30	39	37.7593	37.7648	-0.00544	0.00889	0.94136	-0.93247
16	10	40	44	35.6365	35.5471	0.08940	0.00595	0.75920	-0.75325
17	10	50	44	35.0603	34.7042	0.35612	0.01564	1.05033	-1.03469
18	10	60	41	36.0716	35.4737	0.59794	0.01321	1.18277	-1.16957
19	10	70	36	38.1196	38.1040	0.01565	0.00315	0.89157	-0.88842
20	10	80	33	41.2799	41.0734	0.20644	0.02446	1.22772	-1.20326
21	10	90	30	43.2193	42.8904	0.32890	0.05988	1.49438	-1.43450
22	10	100	26	41.0358	43.1350	-2.09918	0.19050	1.93434	-1.74384
23	20	0	29	44.4890	44.4359	0.05317	0.06179	1.23618	-1.17439
24	20	10	35	43.3391	43.2938	0.04531	0.00526	0.95512	-0.94986
25	20	20	39	41.1293	40.9885	0.14079	0.00675	1.18544	-1.17870
26	20	30	43	38.6060	38.5300	0.07598	0.00898	1.08973	-1.08075
27	20	40	49	36.5013	36.5275	-0.02623	0.03037	1.33620	-1.30583
28	20	50	49	36.1158	35.7959	0.31990	0.02535	1.31986	-1.29451
29	20	60	49	36.8115	36.5397	0.27182	0.00835	1.11490	-1.10656
30	20	70	39	38.4308	38.5182	-0.08746	0.00257	0.89419	-0.89162
31	20	80	36	41.0601	41.0449	0.01511	0.00766	1.18548	-1.17781
32	20	90	33	43.1788	43.1073	0.07144	0.00613	0.94924	-0.94311
33	20	100	27	42.7757	43.4689	-0.69313	0.06770	1.52094	-1.45324
34	30	0	35	43.3601	43.9579	-0.59779	0.04662	1.32306	-1.27644
35	30	10	39	43.1539	43.1448	0.00912	0.00245	0.72413	-0.72167
36	30	20	44	41.2400	41.2166	0.02336	0.00528	1.10234	-1.09706
37	30	30	52	38.9296	39.0178	-0.08816	0.00489	1.04501	-1.04012
38	30	40	57	37.2813	37.3412	-0.05992	0.00804	0.89242	-0.88438
39	30	50	57	36.7198	36.7558	-0.03597	0.00652	0.83517	-0.82865
40	30	60	55	37.2047	37.3407	-0.13597	0.00682	1.00330	-0.99648
41	30	70	48	38.8856	38.8919	-0.00628	0.00285	1.01430	-1.01145
42	30	80	43	41.0627	41.0663	-0.00359	0.00260	0.97336	-0.97077
43	30	90	36	43.0969	43.0465	0.05038	0.00194	0.51312	-0.51118
44	30	100	29	44.5840	43.3474	1.23663	0.13593	1.57267	-1.43674
45	40	0	36	42.8186	43.5157	-0.69706	0.01976	1.25689	-1.23713
46	40	10	40	42.8970	42.9168	-0.01984	0.00301	0.95163	-0.94862
47	40	20	52	41.1025	41.1824	-0.07989	0.00193	0.96204	-0.96012
48	40	30	60	39.3288	39.2992	0.02960	0.00451	1.05561	-1.05111
49	40	40	67	38.2096	37.9680	0.24161	0.01791	1.29139	-1.27349
50	40	50	68	37.3139	37.5055	-0.19150	0.04039	1.51095	-1.47056
51	40	60	64	37.3353	37.9400	-0.60462	0.02973	1.45391	-1.42418
52	40	70	58	39.2288	39.2541	-0.02528	0.00271	0.93775	-0.93503
53	40	80	53	41.0334	41.0063	0.02715	0.00081	0.72274	-0.72193
54	40	90	43	42.6291	42.4154	0.21372	0.02307	1.25552	-1.23246
55	40	100	33	44.1642	42.7534	1.41071	0.27397	1.76406	-1.49010
56	50	0	35	42.5825	43.0164	-0.43392	0.02145	1.19943	-1.17798
57	50	10	43	42.5996	42.5198	0.07972	0.00374	0.95597	-0.95223
58	50	20	52	41.0230	41.0736	-0.05060	0.00190	0.80091	-0.79901
59	50	30	64	39.5184	39.5140	0.00449	0.00460	0.99050	-0.98590
60	50	40	71	38.3804	38.4002	-0.01977	0.02814	1.41467	-1.38654

61	50	50	72	37.1603	38.0278	-0.86749	0.07057	1.69401	-1.62344
62	50	60	68	37.6008	38.3635	-0.76274	0.04500	1.50710	-1.46210
63	50	70	58	39.4703	39.4391	0.03119	0.00467	0.94172	-0.93705
64	50	80	52	40.9501	40.8713	0.07884	0.00418	1.11901	-1.11483
65	50	90	44	42.2058	42.1254	0.08044	0.01048	0.71614	-0.70566
66	50	100	35	43.5303	42.4478	1.08245	0.25062	1.62033	-1.36971
67	60	0	35	42.2662	42.4700	-0.20384	0.02755	1.15384	-1.12629
68	60	10	42	42.2378	42.1038	0.13400	0.00956	1.01338	-1.00382
69	60	20	51	40.9834	40.9730	0.01036	0.00760	1.13076	-1.12316
70	60	30	61	39.5977	39.6665	-0.06880	0.00558	0.99476	-0.98918
71	60	40	66	37.9681	38.6875	-0.71935	0.02987	1.48938	-1.45951
72	60	50	70	37.1422	38.2826	-1.14040	0.06112	1.70395	-1.64284
73	60	60	68	37.9842	38.5250	-0.54087	0.03955	1.57370	-1.53415
74	60	70	54	39.5706	39.5196	0.05102	0.01233	1.10246	-1.09013
75	60	80	46	40.5708	40.7059	-0.13513	0.01457	1.09420	-1.07963
76	60	90	42	41.6046	41.7339	-0.12934	0.03151	1.16079	-1.12929
77	60	100	35	41.4345	42.0550	-0.62052	0.09857	1.60124	-1.50267
78	70	0	35	41.7605	42.1236	-0.36312	0.08049	1.52418	-1.44368
79	70	10	38	41.7842	41.7844	-0.00018	0.00461	0.66583	-0.66122
80	70	20	47	40.7629	40.8773	-0.11440	0.01291	1.15745	-1.14454
81	70	30	52	39.7303	39.7416	-0.01127	0.00205	0.71282	-0.71077
82	70	40	57	38.5335	38.8522	-0.31867	0.01477	1.37830	-1.36353
83	70	50	62	37.9375	38.4673	-0.52985	0.02498	1.45962	-1.43464
84	70	60	56	38.6802	38.7377	-0.05750	0.02250	1.50287	-1.48037
85	70	70	47	39.6669	39.5180	0.14887	0.01535	1.21800	-1.20265
86	70	80	42	40.5276	40.5466	-0.01904	0.00726	0.87303	-0.86577
87	70	90	37	41.2246	41.3097	-0.08508	0.04701	1.30379	-1.25678
88	70	100	33	39.9290	41.6639	-1.73498	0.20448	1.77135	-1.56686
89	80	0	31	41.6827	41.8330	-0.15024	0.05229	1.32478	-1.27249
90	80	10	35	41.6503	41.6131	0.03723	0.00202	0.70805	-0.70604
91	80	20	43	40.8009	40.7935	0.00746	0.00375	0.72766	-0.72391
92	80	30	47	40.0556	39.8526	0.20295	0.01814	1.34578	-1.32764
93	80	40	52	39.2875	39.0467	0.24085	0.01159	1.22214	-1.21055
94	80	50	50	38.5870	38.5990	-0.01203	0.00074	0.65595	-0.65521
95	80	60	49	38.9292	38.8683	0.06096	0.00258	1.03199	-1.02941
96	80	70	45	39.6483	39.5615	0.08682	0.00349	1.12472	-1.12123
97	80	80	37	40.6906	40.3853	0.30529	0.01635	1.22567	-1.20932
98	80	90	33	41.1603	41.0230	0.13723	0.06477	1.39154	-1.32678
99	80	100	31	39.9106	41.3872	-1.47665	0.21764	1.51630	-1.29866
100	90	0	28	41.6452	41.5506	0.09467	0.01214	0.79513	-0.78299
101	90	10	31	41.3929	41.3776	0.01531	0.00213	0.73545	-0.73333
102	90	20	36	40.4533	40.7600	-0.30663	0.02865	1.43345	-1.40480
103	90	30	41	40.0628	39.9885	0.07429	0.04942	1.67601	-1.62659
104	90	40	41	39.4289	39.2936	0.13531	0.02642	1.37536	-1.34895
105	90	50	44	38.8618	38.8703	-0.00850	0.00042	0.51538	-0.51496
106	90	60	39	39.1550	39.0936	0.06138	0.00418	0.98673	-0.98255
107	90	70	32	39.6165	39.6119	0.00467	0.00080	0.79697	-0.79618
108	90	80	27	40.1824	40.2622	-0.07974	0.01010	0.89490	-0.88480
109	90	90	26	41.0182	40.7950	0.22323	0.03405	1.18818	-1.15413
110	90	100	25	41.6405	41.1315	0.50896	0.05499	0.77000	-0.71500
111	100	0	26	43.4372	41.2850	2.15225	0.16375	1.82802	-1.66427
112	100	10	27	42.6488	41.1598	1.48896	0.09281	1.74294	-1.65013
113	100	20	31	41.5685	40.7558	0.81271	0.21441	1.93836	-1.72394
114	100	30	33	41.7093	40.1598	1.54955	0.20921	1.99653	-1.78732
115	100	40	34	39.9971	39.6565	0.34063	0.08372	1.71559	-1.63187
116	100	50	34	39.3376	39.4252	-0.08764	0.05489	1.33611	-1.28122
117	100	60	34	39.5622	39.5883	-0.02604	0.01056	0.92205	-0.91149
118	100	70	27	39.4602	39.7773	-0.31713	0.03231	1.24455	-1.21223
119	100	80	24	39.3618	40.1209	-0.75906	0.03926	1.24930	-1.21005
120	100	90	23	41.4052	40.4980	0.90718	0.12795	1.43988	-1.31192
121	100	100	23	44.5381	40.7383	3.79975	0.41616	1.80688	-1.39072

The predicted values at each of the grid locations do not differ greatly for the two variogram models. However, the standard error of prediction for the spherical model is substantially larger than the Gaussian model.

---

## References

- Christakos, G. (1992), *Random Field Models in Earth Sciences*, New York: Academic Press.
- Cressie, N.A.C. (1993), *Statistics for Spatial Data*, New York: John Wiley & Sons, Inc.
- Duetsch, C.V. and Journel, A.G. (1992), *GSLIB: Geostatistical Software Library and User's Guide*, 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 (4), 313–341.
- Journel, A.G. (1985), "The Deterministic Side of Geostatistics," *Mathematical Geology*, 17 (1), 1–15.
- Journel, A.G. and Huijbregts, Ch.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 (7), 715–739.
- Matheron, G. (1971), *The Theory of Regionalized Variables and Its Applications*, Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau.
- Robinson, G.K. (1991), "That BLUP is a Good Thing: The Estimation of Random Effects," *Statistical Science*, 6, 15–51.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS/STAT® User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999.

**SAS/STAT® User's Guide, Version 8**

Copyright © 1999 by SAS Institute Inc., Cary, NC, USA.

ISBN 1-58025-494-2

All rights reserved. Produced in the United States of America. 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.

**U.S. Government Restricted Rights Notice.** Use, duplication, or disclosure of the software and related documentation by the U.S. government is subject to the Agreement with SAS Institute and the restrictions set forth in FAR 52.227-19 Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999

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 registered trademarks or trademarks of their respective companies.

The Institute is a private company devoted to the support and further development of its software and related services.