

Variogram Tutorial

Randal Barnes
Golden Software, Inc.

Table of Contents

| | |
|---|----|
| 1 – Introduction | 3 |
| 2 – What does a variogram represent? | 4 |
| 3 – What is a variogram? | 6 |
| 4 – The variogram grid | 7 |
| 5 – Modeling the omni-directional variogram | 9 |
| 6 – Modeling the variogram anisotropy | 12 |
| 7 – Rules of thumb | 15 |
| 8 – Frequently asked questions | 16 |
| 9 – Some geostatistical references | 21 |

1 Introduction

The variogram characterizes the spatial continuity or roughness of a data set. Ordinary one-dimensional statistics for two data sets may be nearly identical, but the spatial continuity may be quite different. Refer to Section 2 for a partial justification of the variogram.

Variogram analysis consists of the *experimental variogram* calculated from the data and the *variogram model* fitted to the data. The experimental variogram is calculated by averaging one-half the difference squared of the z-values over all pairs of observations with the specified separation distance and direction. It is plotted as a two-dimensional graph. Refer to Section 3 for details about the mathematical formulas used to calculate the experimental variogram.

The variogram model is chosen from a set of mathematical functions that describe spatial relationships. The appropriate model is chosen by matching the shape of the curve of the experimental variogram to the shape of the curve of the mathematical function.

Refer to the **Surfer** User's Guide and the topic **Variogram Model Graphics** in the **Surfer** Help for graphs illustrating the curve shapes for each function. To account for geometric anisotropy (variable spatial continuity in different directions), separate experimental and model variograms can be calculated for different directions in the data set.

2 What does a variogram represent?

Consider two synthetic data sets; we will call them **A** and **B**. Some common descriptive statistics for these two data sets are given in Table 1.1.

| | A | B |
|--------------------|--------|--------|
| Count | 15251 | 15251 |
| Average | 100.00 | 100.00 |
| Standard Deviation | 20.00 | 20.00 |
| Median | 100.35 | 100.92 |
| 10 Percentile | 73.89 | 73.95 |
| 90 Percentile | 125.61 | 124.72 |

Table 1.1 Some common descriptive statistics for the two example data sets.

The histograms for these two data sets are given in Figures 1.1 and 1.2. According to this evidence the two data sets are almost identical.

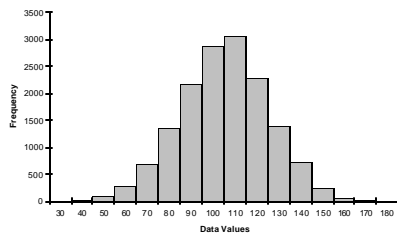


Figure 1.1 Data Set A Histogram

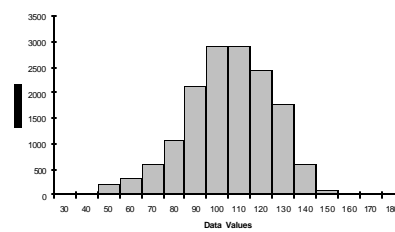


Figure 1.2 Data Set B Histogram

However, these two data sets are significantly different in ways that are not captured by the common descriptive statistics and histograms. As can be seen by comparing the associated contour plots (see Figures 1.3 and 1.4), data set **A** is rougher than data set **B**. Note that we can not say that data set **A** is "more variable" than data set **B**, since the standard deviations for the two data sets are the same, as are the magnitudes of highs and lows. The visually apparent difference between these two data sets is one of texture and not variability.

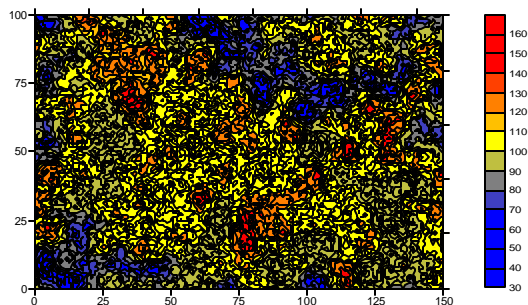


Figure 1.3 Data Set A Contour Plot

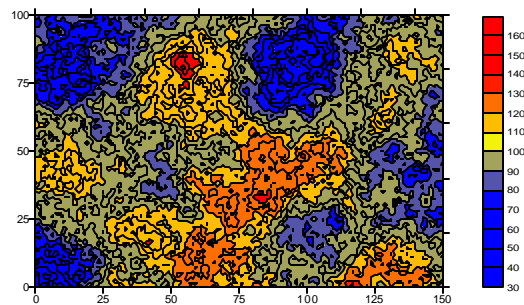


Figure 1.4 Data Set B Contour Plot

In particular, data set **A** changes more rapidly in space than does data set **B**. The continuous high zones (red patches) and continuous low zones (blue patches) are, on the average, smaller for data set **A** than for data set **B**. Such differences can have a significant impact on sample design, site characterization, and spatial prediction in general.

It is not surprising that the common descriptive statistics and the histograms fail to identify, let alone quantify, the textural difference between these two example data sets. Common descriptive statistics and histograms do not incorporate the spatial locations of data into their defining computations.

The variogram is a quantitative descriptive statistic that can be graphically represented in a manner which characterizes the spatial continuity (i.e. roughness) of a data set. The variograms for these two data sets are shown in Figures 1.5 and 1.6. The difference in the initial slope of the curves is apparent.

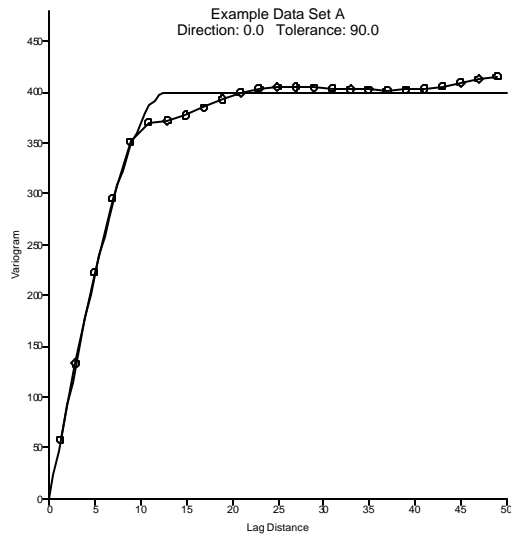


Figure 1.5 Data Set A
Variogram and Model

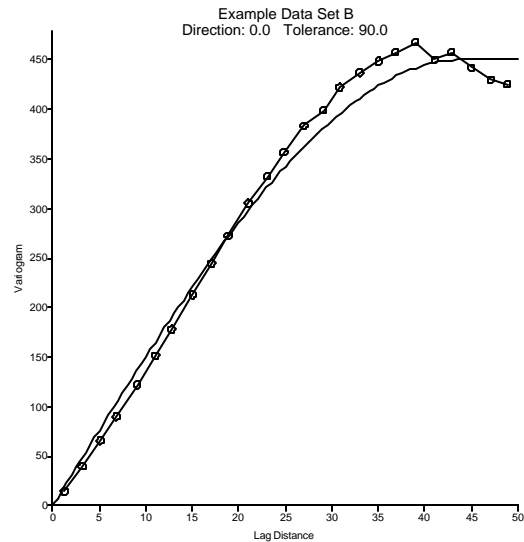


Figure 1.6 Data Set B
Variogram and Model

3 What is a variogram?

The mathematical definition of the variogram is

$$\gamma(\Delta x, \Delta y) = \frac{1}{2} \mathcal{E} \left[[Z(x + \Delta x, y + \Delta y) - Z(x, y)]^2 \right] \quad (3.1)$$

where $Z(x,y)$ is the value of the variable of interest at location (x, y) , and $\mathcal{E} []$ is the statistical expectation operator. Note that the variogram, $\gamma()$, is a function of the separation between points (Dx, Dy) , and not a function of the specific location (x, y) . This mathematical definition is a useful abstraction, but not easy to apply to observed values.

Consider a set of n observed data: $\{(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_n, y_n, z_n)\}$, where (x_i, y_i) is the location of observation i , and z_i is the associated observed value. There are $n(n - 1)/2$ unique pairs of observations. For each of these pairs we can calculate the associated separation vector:

$$(Dx_{i,j}, Dy_{i,j}) = (x_i - x_j, y_i - y_j) \quad (3.2)$$

When we want to infer the variogram for a particular separation vector, (Dx, Dy) , we will use all of the data pairs whose separation vector is approximately equal to this separation of interest:

$$(Dx_{i,j}, Dy_{i,j}) \approx (Dx, Dy) \quad (3.3)$$

Let $S(Dx, Dy)$ be the set of all such pairs:

$$S(Dx, Dy) = \{ (i,j) \mid (Dx_{i,j}, Dy_{i,j}) \approx (Dx, Dy) \} \quad (3.4)$$

Furthermore, let $N(Dx, Dy)$ equal the number of pairs in $S(Dx, Dy)$. To infer the variogram from observed data we will then use the formula for the *experimental variogram*.

$$\hat{\gamma}(\Delta x, \Delta y) = \frac{1}{2N(\Delta x, \Delta y)} \sum_{(i,j) \in S(\Delta x, \Delta y)} (z_i - z_j)^2 \quad (3.5)$$

That is, the experimental variogram for a particular separation vector of interest is calculated by averaging one-half the difference squared of the z -values over all pairs of observations separated by approximately that vector.

4 The variogram grid

If there are n observed data, there are $n(n - 1)/2$ unique pairs of observations. Thus, even a data set of moderate size generates a large number of pairs. For example, if $n = 500$, $n(n - 1)/2 = 124,745$ pairs. The manipulation of such a large number of pairs can be time consuming, even for a fast computer. **Surfer** pre-computes all of the pairs and stores the necessary sums and differences in the *variogram grid*. (Note: a *variogram grid* is not the same format as a grid used in creating a map.)

To create a new variogram, choose the **Grid | Variogram | New Variogram** menu command, specify the data file name in the **Open** dialog box, and click the **Open** button. Specify the X, Y, and Z columns, *Duplicates* settings, *Data Exclusion Filter* (if any), and review the *Data Statistics*.

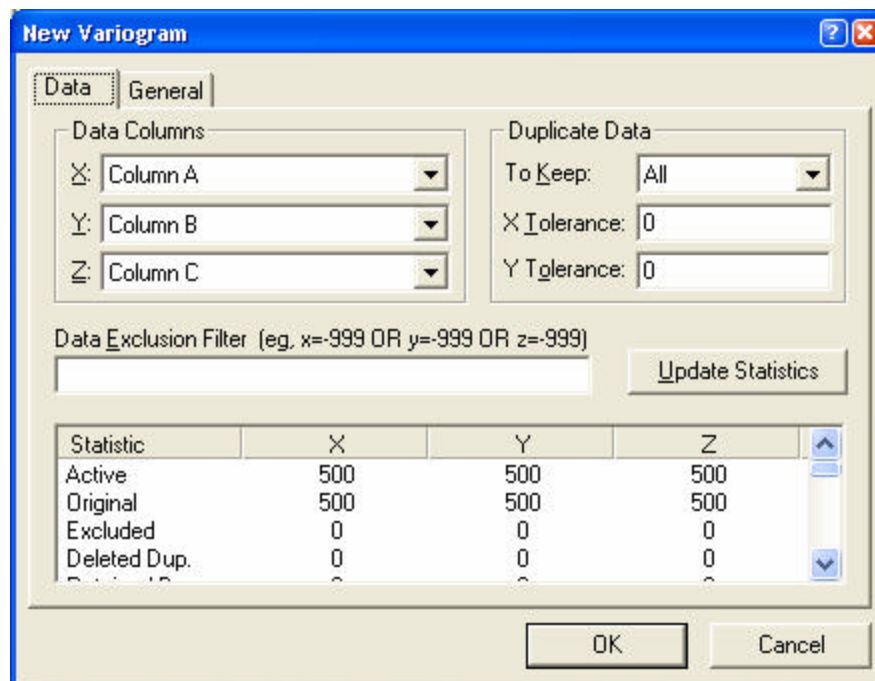


Figure 4.1 Choose the **Grid | Variogram | New Variogram** menu to display the **Data** tab of the **New Variogram** dialog box.

Click the **General** tab to view the *Variogram Grid* and *Detrending* options. The *Max Lag Distance* is the maximum separation distance to be considered during variogram modeling. By default, this is approximately one-third the diagonal extent of the observed data. The *Angular divisions* of 180 and the *Radial divisions* of 100 are adequate for almost any setting.

The *Detrend* options offer advanced data handling options for *universal kriging*. Typically, the appropriate option is *Do not detrend the data*. However, if you know that a strong trend exists in the data, you may want to consider *Linear* detrending. Choose the *Generate Report* option to create a list of the *Data Filter Settings* and *Data Statistics*.

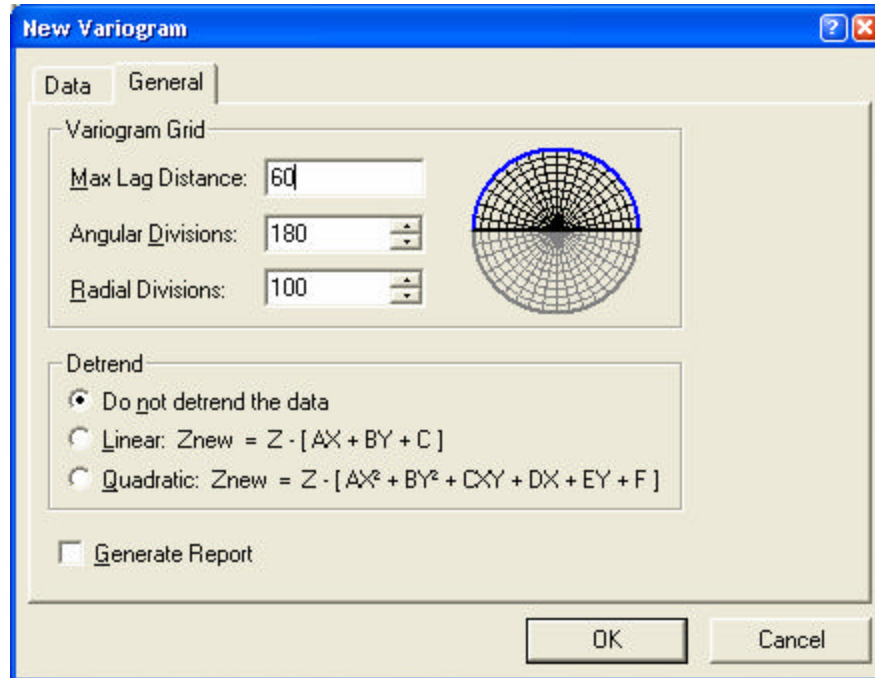


Figure 4.2 Click the **Options** tab of the **New Variogram** dialog box to display the Variogram Grid, Detrend, and Report options.

Without changing any of the settings, select **OK**. Figure 4.3 is displayed.

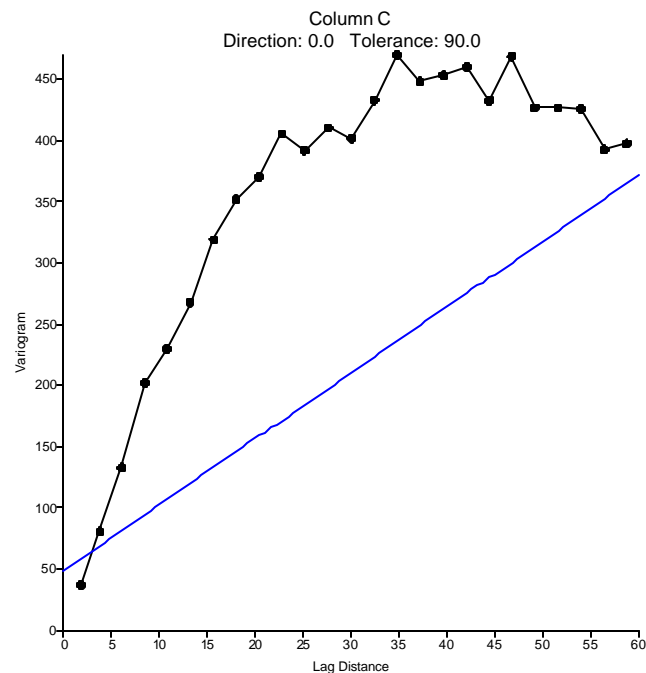


Figure 4.3 Resulting variogram with default variogram settings using ExampleDataSetC.xls.

The black line with the dots is the omni-directional experimental variogram, while the blue line is a first pass (albeit a poor one) at a fitted variogram model.

5 Modeling the omni-directional variogram

By default, this first plot is the omni-directional variogram (the directional tolerance is 90 degrees). Choose the model type, the sill, and the nugget effect based upon the omni-directional variogram.

5.1 Selecting the variogram model type

There are infinitely many possible variogram models. **Surfer** allows for the construction of thousands of different variogram models by selecting combinations of the ten available component types. When combined with a *nugget effect*, one of three models is adequate for most data sets: the *linear*, the *exponential*, and the *spherical* models. Examples of these three models are shown in Figure 5.1.

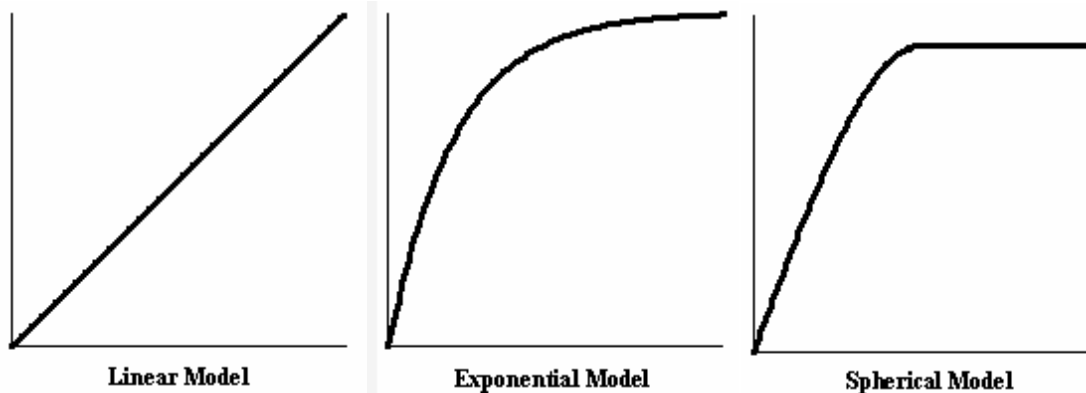


Figure 5.1 Variogram Models

If the experimental variogram never levels out, then the linear model is usually appropriate. If the experimental variogram levels out, but is "curvy" all the way up, then the exponential model should be considered. If the experimental variogram starts out straight, then bends over sharply and levels out, the spherical model is a good first choice.

For the data in ExampleDataSetC.xls, a spherical model appears appropriate (one could also try an exponential model). Double click on the variogram plot and select the **Model** tab.

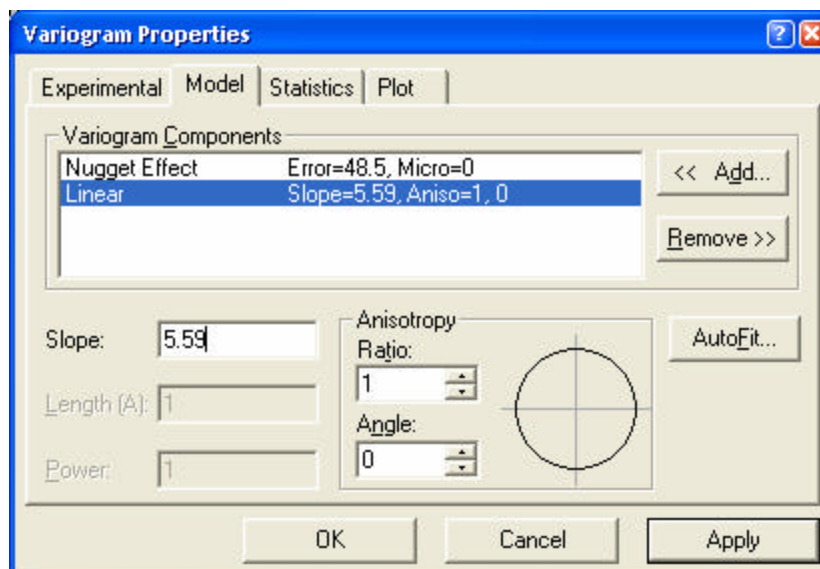


Figure 5.2 The **Model** tab of the **Variogram Properties** dialog box.

Press the **Remove>>** button twice to remove the inappropriate default model. Then press the **<<Add** button, select the *Spherical* model and press **OK**.

5.2 Selecting the variogram model scale and length parameters

We must now set the *Scale* and the *Length (A)* parameters using an iterative approach (i.e. *guess and check*). The *Scale* is the height on the y-axis at which the variogram levels off. By simply looking at the plot, a value between 400 and 450 seems reasonable: enter 425. The *Length (A)* for a spherical model is the lag distance at which the variogram levels off. Again, from the plot a value between 30 and 40 seems reasonable: enter 35. Press the **Apply** button and the new candidate variogram model is drawn.

This is not a bad first guess, but upon examination of the redrawn curve, it appears that the *Length (A)* is a little bit too long since the model (blue line) lies to the right of the experimental variogram plot (black line and dots). Reset the *Length (A)* to 30 and press the **Apply** button. This is still a little bit too long. Try 29 for the *Length (A)*. This is a good fit for a variogram.

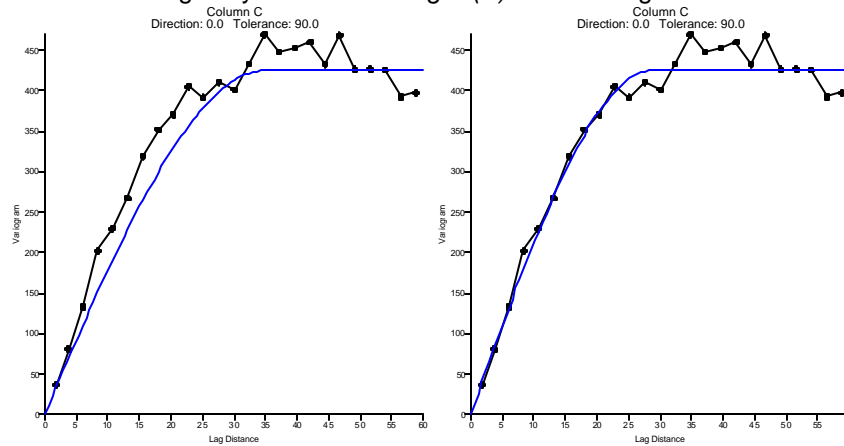


Figure 5.3 Variogram model with initial assumptions
 Left: Scale = 425, Length (A) = 35. Right: Scale = 425, Length (A) = 29.

5.3 Selecting the variogram nugget effect

If the experimental variogram appears to have a non-zero intercept on the vertical axis, then the model may need a nugget effect component. The variance of *Delta Z* in the *Nearest Neighbor Statistics* section of the *Variogram Grid Report* offers a quantitative upper-bound for the nugget effect in most circumstances.

In **Surfer** the nugget effect is partitioned into two sub-components: the *error variance* and the *micro variance*. Both of these sub-components are non-negative, and the sum of these two sub-components should equal the apparent non-zero intercept.

The *error variance* measures the reproducibility of observations. This includes both sampling and assaying (analytical) errors. The error variance is best selected by computing the variance of differences between duplicate samples.

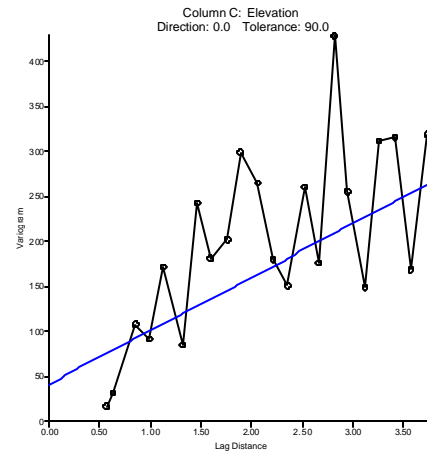


Figure 5.4 Linear Variogram model with nugget effect for data set Demogrid.dat.

The *micro variance* is a substitute for the unknown variogram at separation distances of less than the typical sample spacing. This is best selected by taking the difference between the apparent non-zero intercept of the experimental variogram and the error variance.

The model for our example appears to intersect the vertical axis at 0, so we will not apply a nugget effect.

6 Modeling the variogram anisotropy

Often, the experimental variogram shows different length scales in different directions. This is called *geometric anisotropy*. For a linear variogram model this would appear as a different slope in different directions, while a spherical model manifests geometric anisotropy as different apparent *Length (A)* parameters in different directions.

To investigate the variogram anisotropy select the **Variogram Properties | Experimental** tab. The omni-directional experimental variogram averages the behavior over all directions. To look in particular directions change the *Direction* and the *Tolerance* parameters. Start by setting the *Tolerance* to 30 degrees. Notice how the small circular diagram in the *Lag Direction* portion of the property sheet now shows a blue sector centered at 0 degrees (east in this model), with an angular tolerance (plus or minus) of 30 degrees.

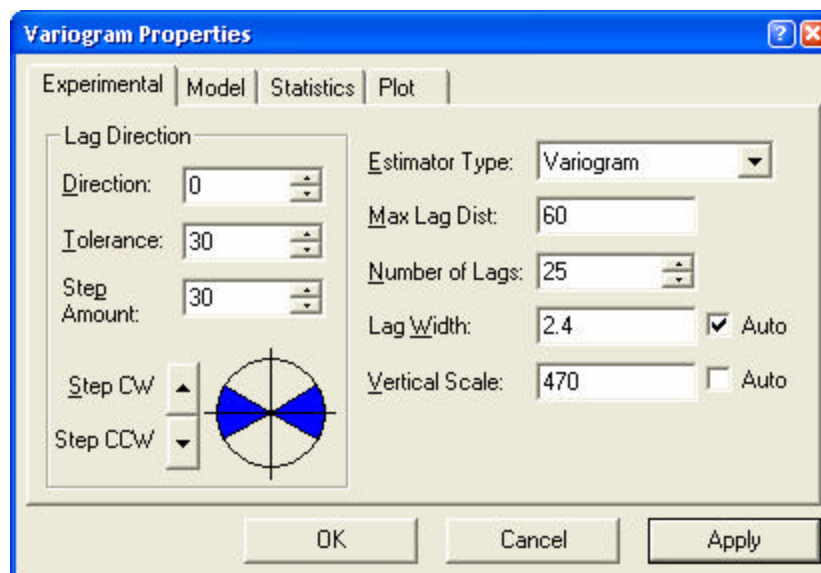


Figure 6.1 Set the *Tolerance* to 30 in the **Experimental** tab of the **Variogram Properties** dialog box.

Drag the dialog box to the side to make the variogram plot visible. To quickly sweep through all directions press the *Step CCW* button. Press it once and the plot is now set for a direction of 30 degrees. Press it again and you have selected 60 degrees. Press it again and again and again and notice how the small circular diagram helps you keep track of the current direction.

Sweep through all directions a few times. Notice how the scaling of the vertical axis changes. To fix the vertical scale of the plot, set the *Vertical Scale* to 515 and press **Apply**.

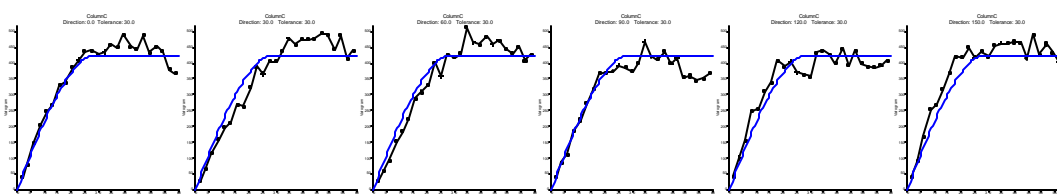


Figure 6.2 Directional variogram plots with *Tolerance* set to 30°. From left to right: *Direction* = 0°, 30°, 60°, 90°, 120°, 150°.

Change the *Step Amount* to 15 degrees and sweep through all directions a couple of times. You can hold the *Step CCW* button down to produce a continuous sweep. It appears that for

directions between 90 and 180 degrees the omni-directional model is consistently to the right of the ascending portion of the experimental variogram. Conversely, the omni-directional model is consistently to the left of the ascending portion of the experimental variogram for directions between 0 and 90 degrees. This is a clear indication of geometric anisotropy.

By pressing the *Step CCW* button move to the direction at which the apparent length and scale of the experimental variogram is maximized: say, 30 degrees. Go to the **Model** tab and change the *Length (A)* parameter to better fit this directional variogram. 40 is too long, 30 is too short, and 37 appears to fit quite well. Change the *Anisotropy Angle* to 30 degrees, and press **Apply** (the plot should not change at this point).

Click on the **Experimental** tab and change the *Direction* to 120 degrees (this is orthogonal to the direction of maximum length, $120 = 30 + 90$). Click on the **Model** tab. Change only the *Anisotropy Ratio* to fit the variogram model to the current directional experimental variogram. (As you change the setting, press the **Apply** button to see the updated plot.) A *Ratio* value of 1.6 appears to be a good fit.

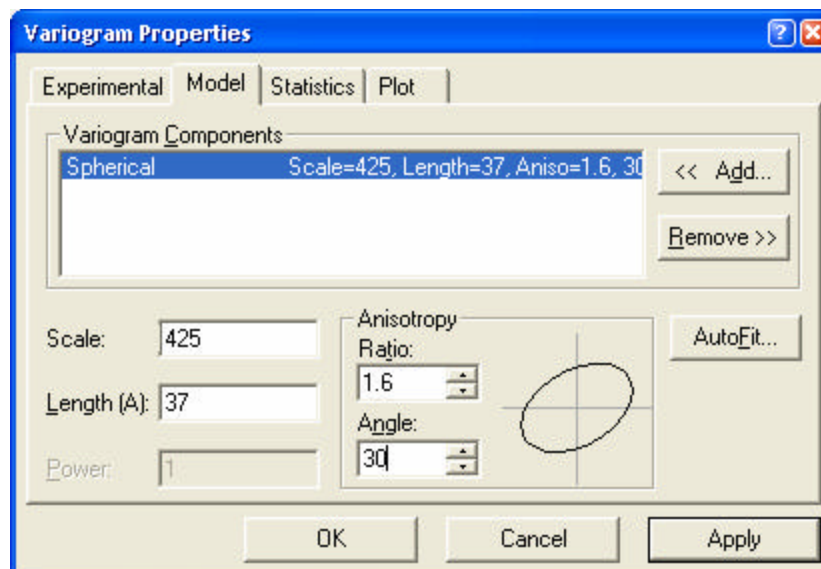


Figure 6.3 Change the variogram model properties to add anisotropy.

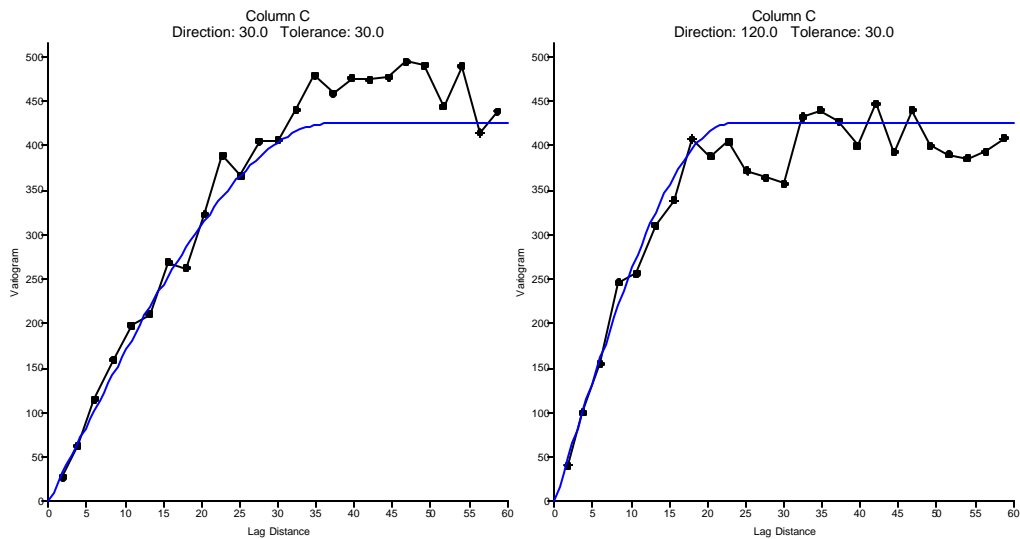


Figure 6.4 Directional variograms with anisotropy settings.
 Left: Direction = 30°, Right: Direction = 120°
 Anisotropy Ratio = 1.6, Angle = 30°

To get a final visual verification of your model, take it for a "spin". Click on the **Experimental** tab and click and hold down the *Step CCW* button. Notice how the variogram model mimics the changing shape of the experimental variogram. You can fine tune the model parameters to better match the model to the experimental variogram, but do not over-model: one or two significant digits is all that you can hope for.

Click on the **AutoFit** button to fine-tune the model after choosing appropriate initial parameters. Click **OK** to accept the defaults, and **AutoFit** returns an *Anisotropy Ratio* of 1.75 and *Angle* of 38.49°.

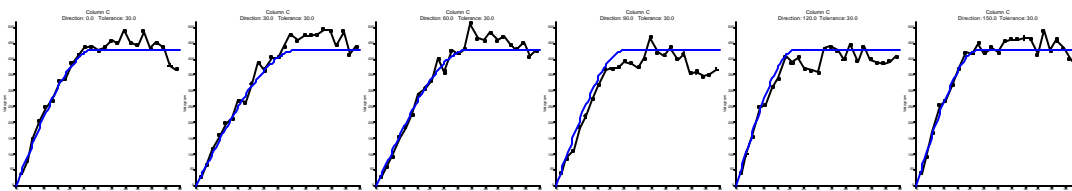


Figure 6.5 Directional Variogram plots with Anisotropy Ratio = 1.75 and Angle = 38.49°.
 From left to right: Direction = 0°, 30°, 60°, 90°, 120°, 150°.

7 Rules-of-thumb

- Know your data! Before calculating the experimental variogram, calculate regular non-spatial statistics. Use the *Variogram Report* to display the data minimum, maximum, median, mean, standard deviation, variance, and skewness. Create a post map or classed post map in **Surfer** to display scatter plots. Use **Grapher** to create histograms and cumulative frequency plots.
- Do not over-model. The simplest model that reproduces the important features of the experimental variogram is the best model.
- When in doubt, use the default variogram model for gridding. A simple linear variogram model usually generates an acceptable grid; this is especially true for initial data analysis. Remember, however, that the kriging standard deviation grid generated using the default variogram is meaningless.
- Unless there is a clear, unambiguous, physical justification, do not use an anisotropy ratio of greater than 3 to 1. If the experimental variogram appears to support an anisotropy of greater than 3 to 1, and there is no unambiguous, physical justification for such a severe anisotropy, there may be a trend in the data. Consider detrending the data before carrying out your variogram analysis.
- Try the model and see how the resulting grid looks in a contour map. If you have competing candidate variogram models, generate a grid and contour map from each. If there are no significant differences, choose the simplest variogram model.
- The range of the variogram is often close to the average size of physical anomalies in the spatial fluctuation of the Z values. In the absence of a reliable experiment variogram, this rule-of-thumb may be used to postulate a variogram range.
- An experimental variogram that fluctuates around a constant value is not an ill-behaved variogram. It is an indication that the Z values are uncorrelated at the scale of the typical sample spacing. In such a situation a contour map, regardless of the gridding method used, is an unreliable representation of the data, and more data at closer sample spacing are needed for detailed local characterization.
- If the following three conditions are met, then the sample variance is a reasonable approximation for the variogram sill:
 1. The data are evenly distributed across the area of interest, as displayed in a post map.
 2. There is no significant trend in the data across the area of interest.
 3. The dimension of the area of interest is more than three times the effective variogram range.

8 Frequently asked questions

Q. What is a variogram?

A. A variogram is a statistically-based, quantitative, description of a surface's roughness. A variogram is a function of a separation vector: this includes both distance and direction, or a Δx and a Δy . The variogram function yields the average dissimilarity between points separated by the specified vector (dissimilarity is measured by the squared difference in the Z-values).

Q. What is the difference between a variogram and a semi-variogram?

A. The difference is simply a factor of 2. The variogram was originally defined as

$$2\gamma(\Delta x, \Delta y) = \mathcal{E} \left[\{Z(x + \Delta x, y + \Delta y) - Z(x, y)\}^2 \right] \quad (8.1)$$

where $2\gamma(\)$ was given the name *variogram*. The function of practical interest was $\gamma(\)$ which was given the name *semi-variogram* since it is one-half the variogram function. Since only the function $\gamma(\)$ is used in kriging, the prefix *semi-* is regularly dropped, and the function $\gamma(\)$ is interchangeably called the variogram and the semi-variogram in the geostatistical literature.

Q. What are the units of the variogram plot?

A. Let (X, Y) represent the location of an observation, and let Z represent the associated observed value. The abscissa (lag distance axis) has the same units as X and Y. The ordinate (variogram axis) has the units equal to the units of Z squared (like the variance of Z). For example, let

- X represent the distance east of a benchmark measured in [m],
- Y represent the distance north of a benchmark also measured in [m], and
- Z represent rainfall intensity measured in [cm/hr].

The abscissa subsequently has units of [m], and the ordinate has units of [(cm/hr)²].

Q. Why can't I specify the sill in the variogram parameters? How is the scale related to the sill?

A. The *sill* is the value of the ordinate at which the variogram levels off. A variogram model may be comprised of multiple components (e.g. a spherical component plus a nugget effect). While the individual components may be said to have a sill, such usage often leads to confusion between the sill of the overall model and the various sills of the individual components. A further source of confusion is that some components, and therefore models including such components, do not have a sill (e.g. the linear model).

To reduce confusion, the word *sill* was used to describe the leveling-off of the entire model, and not of the individual components. When discussing the leveling-off of an individual component, we use the term *scale*.

Q. What does the variogram direction do? Why does the variogram change when I change the direction?

A. This behavior is called *anisotropy*. The variogram measures dissimilarity as a function of separation distance and direction. In many physical settings it is possible for the variable of interest to change more rapidly in one direction than in another. For example, the distribution of grain size changes more rapidly in a direction perpendicular to the shoreline than it does parallel

to the shoreline. Similarly, in an arid climate the prevalence of certain species of plants changes more rapidly as one moves in a direction perpendicular to a river than it does as one moves along the river.; i.e. **different** behavior in different directions. The direction parameter allows you to investigate the variogram in different directions. Thus, you can identify and quantify the anisotropy.

Q. Are the various variogram directions used when gridding with **Surfer**?

A. Yes. The variogram anisotropy is explicitly entered when specifying the variogram for gridding with the kriging algorithm.

Q. Does **Surfer** have a multi-directional gridding algorithm so that more than just the anisotropy can be changed (different sills in different directions)?

A. The *variogram model* contains the necessary information to describe the behavior of the variogram in all directions. The spherical model, for example, is a funnel-like surface. The variogram plot on which a single line is drawn to represent the variogram model is a vertical slice through the variogram model.

The *anisotropy ratio and angle* describe the ratio of the *Lengths (A)* in different directions (*geometric anisotropy*). Sometimes the apparent sill of the experimental variogram differs in different directions (*zonal anisotropy*). For example, an experimental variogram appears to be well described by a spherical model with a *Length (A)* of 100 meters. It levels off (sill) at 2.0 in the East/West direction (Angle = 0 degrees), while it levels off at 3.0 in the North/South direction (Angle = 90 degrees). **Surfer** can reproduce such behavior using a nested model. Make the first component an isotropic spherical model with a length of 100 and a scale of 2. Make the second component be an anisotropic model with length 100, ratio 1000, direction 90, and scale of 1. Adding the two components results in a compound model with a sill of 2 in the East/West direction and a sill of 3 in the North/South direction.

Q. Why are some of the cells blank when exporting variogram data?

A. A cell is blank when there are no data pairs separated by the specified lag distance interval and angular tolerance represented by the cell.

Q. How do I perform kriging cross-validation, jack-knifing (leave-out-one modeling), and such.

A. The **Grid | Data** command in **Surfer** offers an extensive set of cross-validation tools. The **Cross Validation Report** (using the Kriging gridding method) offers numerous quantitative measures that can be used as a goodness-of-fit statistic for the variogram. We recommend three particular statistics as the most consistently useful: the *median absolute deviation* of the cross validation residuals, the *standard deviation* of the cross validation residuals, and the *rank correlation* between the measurements and the estimates.

Q. How do I plot the 3D variogram surface?

A. This option is not available in **Surfer**.

Q. Does **Surfer** calculate S^2 (variance of the data) and how do we find this value in the program outputs?

A. **Surfer** uses the formula in the Help and Manual for calculating the sample variance in the experimental variogram. This value is reported in the **Statistics** tab of the **Variogram Properties** dialog box and in the *Variogram Report*.

Q. Why doesn't the grid change when using a linear variogram with different slope values?

A. This occurs if there is no nugget effect. Changing the slope of a linear variogram is equivalent to changing only the scale of the observation locations; thus, the ratio of the variogram values between any two pairs of points is unchanged by changing the slope. It is important to note, however, that the kriging variance is sensitive to the slope even if the interpolation weights and interpolated value are not.

Q. How can I get information on the distance between data points?

A. Plot your data as a Post Map (see the **Map | Post Map | New Post Map** menu). **Surfer** reports some useful quantitative statistics in the *Variogram Report* and at the bottom of the table on the **Statistics** tab of the **Variogram Properties** dialog box. For example, the *Nearest Neighbor (NN)* values for the ExampleDataSetC.xls are:

- Avg. distance to nearest neighbor: 2.90403
- Min. distance to nearest neighbor: 1
- Max. distance to nearest neighbor: 9.48683
- Gamma for nearest neighbors: 64.735

In addition, the **Grid | Data** menu with the **Data Metrics** gridding method offers an extensive suite of tools for characterizing the data geometry. This includes a neighborhood count, data density, distance to the nearest neighbor, etc.

Q. When using an exponential variogram model for kriging, is the input parameter *Length (A)* the correlation length (i.e. the length at which the variogram value reaches 63 percent of the sill) or the practical range (i.e. the length at which the variogram value reaches 95 percent of the sill)?

A. Different geostatistical software packages use various definitions for the length parameter of an exponential model. **Surfer** defines the *Length (A)* parameter, as the correlation length. Graphically, the correlation length corresponds to the lag distance at which the tangent line to the curve at lag distance zero intersects the sill of the exponential model. The exponential model achieves about 63 percent of the sill at *Lag Distance (A)*. At a lag distance of 3A the exponential model reaches about 95 percent of the sill.

Q. When applying the Gaussian model, I seem to be getting a mismatch between what I see on the plot of the variogram and the length reported by the software after **AutoFit** has finished. The length at which the sill is reached seems to be much larger than the value reported under the **Model** tab when **AutoFit** has finished.

A. As with the exponential model, different geostatistical software packages use various definitions for the *Length (A)* parameter of a Gaussian model. **Surfer** defines the *Length (A)* parameter as the correlation length. Graphically, the correlation length corresponds to the lag distance of the inflection point of the curve. The Gaussian model achieves about 63 percent of the sill at *Lag Distance (A)*. At a lag distance of 3A the Gaussian model reaches about 99.99 percent of the sill.

Q. Is there a function in the variogram procedure in the new **Surfer** that can actually tell me the exact nugget, range, and sill?

A. To view the nugget, range, and sill, choose the **Grid | Variogram | New Variogram** menu command. Specify the data file, click on the **Open** button, then click **OK** to use the defaults. Double-click on the variogram graph, click on the **Variogram Properties | Model** tab to view the *Nugget Error* and *Micro Variance* and the default linear model *Slope* and *Anisotropy* parameters.

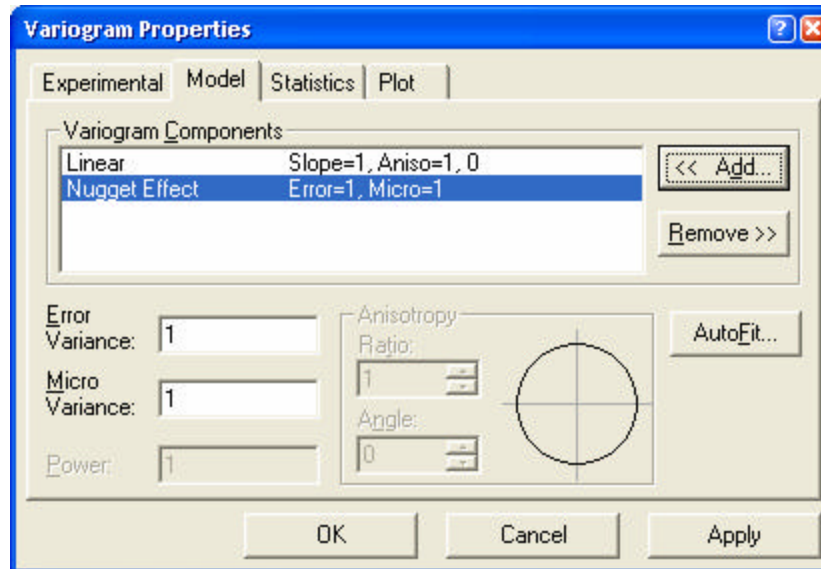


Figure 8.1 Viewing the Variogram properties Slope and Anisotropy

Q. I tried to fit a spherical variogram using **AutoFit** and all the standard settings. The fit is poor. Could you explain how **Surfer** does the fitting? How can I make an appropriate fit using the spherical function?

A. The **AutoFit** function requires reasonable initial parameters to arrive at a good fit. Examine the plot of the experimental variogram and estimate the *Length (A)* and *Scale* visually. If it is difficult to estimate the scale from your data set, use the *Z* variance listed in the *Variogram Report* and on the **Statistics** tab of the **Variogram Properties** dialog box.

Q. I set the *Lag Width* to a particular value. Why aren't the points on the variogram plotted at multiples of this value?

A. The *Lag Width* defines the midpoint of a range plus or minus half the lag width. The point is plotted at the average separation distance for all points within that range.

Q. How can I get a list of the average separation distances?

A. Export the variogram to a data file. This file includes the average lags, half the average variance, and the number of pairs in the range.

Q. Is it possible to create a grid file from two juxtaposed areas on which different variogram models are applied? The areas have spatially contrasted morphologies.

A. If the areas are defined by rectangles, you may be able to grid each area with a separate model, convert the grid to *XYZ* data files, merge the data files in the worksheet and regrid using the *Nearest Neighbor* method to prevent further interpolation.

Q. How is the drift used within the kriging algorithm? Should the modeling be done on transformed data (data minus drift) or on non-transformed data? Kriging with prior removal of the trend is different than Universal Kriging.

A. We recommend calculating the variogram model of the detrended data prior to applying drift during kriging, after Kitanidis, 1997, p. 143, section 6.12.

Q. I haven't found any procedure in **Surfer** that can help me to tell HOW well the theoretical variogram models are fitted to the experimental data in my variogram. I understand that I can use **AutoFit** but I would like a quantitative number that measures the goodness-of-fit.

A. The best measure of "how well" the theoretical model fits the data is to try it out and see. Grid the data with the selected variogram model. Create a contour map and overlay a post map of the data file. Judge the quality of the model by examining the contour map. The **Cross Validation Report** (using the Kriging gridding method) offers numerous quantitative measures that can be used as a goodness-of-fit statistic for the variogram. We recommend three particular statistics as the most consistently useful: the *median absolute deviation* of the cross validation residuals, the *standard deviation* of the cross validation residuals, and the *rank correlation* between the measurements and the estimates.

Q. Can **Surfer** import a variogram model from Variowin, GeoEAS, GSLIB, or other software?

A. **Surfer** does not have this feature. Submit a request to surferupport@goldensoftware.com with the name of the software you would like to support.

Q. I've read somewhere that the scale should not exceed the variance. Is this true? For the data set ExampleDatSetC.xls, the variance of Z is 405.371, but you recommend a scale of 425.

A. Under a reasonable and commonly applicable set of assumptions, the sill of the variogram should be equal to the variance of the underlying population; however, the sample variance is often an inappropriate measure of the population variance (the standard formula for estimating the variance assumes independent data, which is invalid in most geological settings). Thus, a rule like "the scale should not exceed the variance" is not a useful generalization.

The critical result is

$$\mathcal{E} [S^2] = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \gamma(x_i - x_j, y_i - y_j)$$

This result states that the expected value of the sample variance, S^2 , is equal to the average value of the variogram between all n^2 pairs of sample values. If the n sample values are evenly distributed over an areal extent many times larger than the range of the variogram, then there will be significantly more pairs at long separation distances and fewer pairs at short separation distances. Thus, the average value of the variogram between all pairs will be the average of many values equal to the sill and a few values less than the sill. In this case, the sample variance is a reasonable first estimate for the variogram sill.

9 Some geostatistical references

Armstrong, M. [ed.], 1989, *Geostatistics*, volumes 1 and 2, Kluwer Academic Publishers, 1027 pp.

Armstrong, M., 1997, *Basic Linear Geostatistics*, Springer-Verlag, New York, ISBN: 3540618457, 220 pp.

Baafi, E. Y., 1997, *Geostatistics Wollongong '96 – Quantitative Geology and Geostatistics*, Kluwer Academic Publishers, ISBN: 0792302044, 1333 pp.

Bras, R., and Rodriguez-Iturbe, R., 1985, *Random Functions and Hydrology*, Addison-Wesley, 559 pp.

Chiles, J., and Delfiner, P., 1999, *Geostatistics: Modeling Spatial Uncertainty*, John Wiley and Sons, New York, ISBN: 0471083151, 695 pp.

Christakos, G., 2001, *Modern Spatiotemporal Geostatistics*, Oxford University Press; ISBN: 0195138953, 312 pp.

Clark, I., 1979, *Practical Geostatistics*, Applied Science Publishers, 129 pp.

Clark, I., and W. V. Harper, 2000, *Practical Geostatistics 2000*, Ecosse North America, LLC; ISBN: 0970331703, 442 pp.

Cressie, N., 1991, *Statistics for Spatial Data*, John Wiley and Sons, New York, 900 pp.

David, M., 1977, *Geostatistical Ore Reserve Estimation*, 2nd edition, Elsevier, 364 pp.

David, M., 1988, *Handbook of Applied Advanced Geostatistical Ore Reserve Estimation*, Elsevier, 216 pp.

Davis, J., 1973, *Statistics and Data Analysis in Geology*, Wiley, 550 pp.

deMarsily, G., 1986, *Quantitative Hydrogeology*, Academic Press, 440 pp.

Deutsch, Clayton, V., Journel, Andre G., 1992, *GSLIB - Geostatistical Software Library and User's Guide*, Oxford University Press, New York, 338 pp.

Draper, N., and Smith, H., 1981, *Applied Regression Analysis*, second edition, Wiley-Interscience, 709 pp.

Dubrule, O. 1998, *Geostatistics in Petroleum Geology*, American Association of Petroleum Geologists; ISBN: 0891811877.

Gelhar, L. W., 1993, *Stochastic Subsurface Hydrology*, Prentice Hall, Englewood Cliffs, New Jersey, 390 pp.

Geostatistics, 1980, McGraw-Hill, 168 pp.

Gilbert, R.O., 1987, *Statistical Methods for Environmental Pollution Monitoring*, Van Nostrand Reinhold Company, 320 pp.

Gomez-Hernandez, J., A. Soares, and R. Froidevaux [eds.], *Geoenv II -- Geostatistics for Environmental Applications*, Kluwer Academic Publishers, 562 pp.

Goovaerts, P. 1997, *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York, 483 pp.

Guarascio, M., David, M., and Huijbregts, C. [eds.], 1976, *Advanced Geostatistics in the Mineral Industry*, Reidel, 491 pp.

- Gy, P., 1979, *Sampling of Particulate Materials*, Elsevier, 431 pp.
- Houlding, S. W., 2000, *Practical Geostatistics - Modeling and Spatial Analysis*, Springer-Verlag, New York, 160 pp.
- Hohn, M. E., 1998, *Geostatistics and Petroleum Geology*, 2nd edition, Kluwer Academic Publishers; ISBN: 041275780X, 238 pp.
- Isaaks, E. H., and Srivastava, R. M., 1989, *An Introduction to Applied Geostatistics*, Oxford University Press, New York, 561 pp.
- Johnson, M. E., 1987, *Multivariate Statistical Simulation*, John Wiley and Sons, New York, 230 pp.
- Journel, A.G., and Huijbregts, C., 1978, *Mining Geostatistics*, Academic Press, 600 pp.
- Journel, A.G., 1988, *Geostatistics for the Environmental Sciences*, EPA Project Report, Project No. CR 811893, EPA/EMSL, Las Vegas, Nevada.
- Journel, A.G., 1989, *Fundamentals of Geostatistics in Five Lessons*, American Geophysical Union, Washington D.C.
- Kitanidis, P. K., 1997, *Introduction to Geostatistics: Applications in Hydrogeology*, ISBN 0-521-58747-6.
- Koch, G., and Link, R., 1971, *Statistical Analysis of Geological Data*, volumes I and II, Wiley, 375 pp. now available through Dover reprints.
- Krige, D., 1978, *Lognormal DeWijsian Geostatistics for Ore Evaluation*, South African I.M.M., 55 pp.
- Knudsen, H.P, 1988, *A Short Course on Geostatistical Ore Reserve Estimation*, Division of Mining and Minerals Engineering, Montana Tech.
- Manoukian, E.B., 1986, *Modern Concepts and Theorems of Mathematical Statistics*, Springer-Verlag, 156 pp.
- Matheron, G., 1965, *Les Variables Regionalisees et leur Estimation*, Editions Masson, Paris.
- Matheron, G., and Armstrong, M. [eds.], 1987, *Geostatistical Case Studies*, Reidel, 248 pp.
- Monestiez, P., D. Allard, R. Froidevaux [eds.], 2001, *Geoenv III -- Geostatistics for Environmental Applications*, Kluwer Academic Publishers, ISBN: 0792371070, 555 pp.
- Olea, R., 1999, *Geostatistics for Engineers and Earth Scientists*, Kluwer Academic Publishers, Boston, MA, 303 pp.
- Pannatier, Y., 1996, *VARIOWIN: Software for Spatial Data Analysis in 2D*, Springer-Verlag, New York, 91 pp.
- Rendu, J., 1978, *An Introduction to Geostatistical Methods of Mineral Evaluation*, South African I.M.M., 100pp.
- Ripley, B., 1981, *Spatial Statistics*, John Wiley and Sons, New York, 252 pp.
- Rouhani, S., and M. Srivastava, 1997, *Geostatistics for Environmental and Geotechnical Applications*, ASTM Special Technical Publication 1283, American Society for Testing and Materials, ISBN: 0803124147.

Rivoirard, J., J. Simmonds, K. G. Foote, P. Fernandes, and N. Bez [eds.], *Geostatistics for Estimating Fish Abundance*, Blackwell Science Inc; ISBN: 0632054441, 206 pp.

Soares, A. [ed.], 1992, *Geostatistics Troia '92*, volumes 1 and 2, Reidel, 1088 pp.

Soares, A., Gomez-Hernandez, J., and R. Froidevaux [eds.], *Geoenv I -- Geostatistics for Environmental Applications*, Kluwer Academic Publishers, ISBN: 0792345908, 516 pp.

Thompson, S. K., 1992, *Sampling*, John Wiley and Sons, New York, 343 pp.

Verly, G., David, M., Journel, A.G., and Marachel, A. [eds.], 1984, *Geostatistics for Natural Resources Characterization*, volumes 1 and 2, Reidel, 1092 pp.

Webster, R. and M. A. Oliver, 2001, *Geostatistics for Environmental Scientists*, John Wiley and Sons, Ltd., Chichester, England, 271 pp.