

Kansas Geological Survey
Open File Report 2007-27

S-GeMS Tutorial Notes

presented

26 June 2007

in

Hydrogeophysics: Theory, Methods, and Modeling
Boise State University, Boise, Idaho

Geoff Bohling
Assistant Scientist
Kansas Geological Survey
geoff@kgs.ku.edu
864-2093

These notes, tutorial, and data available at:

<http://people.ku.edu/~gbohling/BoiseGeostat>

Introduction

These notes provide a brief introduction to geostatistical data analysis with the S-GeMS software, which is available from <http://sgems.sourceforge.net/>. For Windows, you can download and automatic installer. The download also includes an S-GeMS manual, `sgems_manual.pdf`, that ends up in the folder `C:\Program Files\SGeMS\src\doc` (if you install to the default directory under Windows).

You can read background information on the software on that page. As mentioned there, S-GeMS provides a (fairly) comprehensive collection of geostatistical estimation and simulation algorithms and also provides a nice 3D visualization environment. It provides a more limited selection of options for standard statistical data analysis and essentially no facilities for data management (subsetting data sets, etc.). So, you would probably want to run S-GeMS in tandem with some other data analysis & management software, such as Excel or (my preference) R. R is an open-source data analysis package built on the S language and is available from <http://www.r-project.org/>.

There are a few spatial & geostatistical data analysis packages available for R. My favorite is the `gstat` package by Edzer Pebesma (www.gstat.org). Some other software options are listed at <http://people.ku.edu/~gbohling/geostats/>.

Note that S-GeMS does not handle screen real estate very well. You often have to scroll around in the various windows and panels in order to access all the controls.

Loading Data

S-GeMS can read data from files in its own format and can also read files in GSLIB format, which is pretty much a standard format for geostatistical data. We will start by loading data from a file named ZoneA.dat, a GSLIB-format data file containing porosity and permeability data from 85 wells in Zone A of the Big Bean Oil Field (fictitious, but based on characteristics of a real field, 20 x 16 km in extent). You can open ZoneA.dat in a text editor, such as WordPad, to see what is in it. The header lines and first six data lines look like:

```
Zone A Data, Big Bean Field
8
X m                meters east of origin
Y m                meters north of origin
Thk m              zone thickness in meters
Por %              porosity in percent
Perm md            permeability in millidarcies
LogPerm -          base 10 log of permeability
LogPermPrd -       LogPerm predicted by regression
against Por
LogPermRsd -       Residual from LogPerm-Por regression
12100 8300 37.1531 14.6515 2.8547 0.4556 0.1357 0.3198
5300 8700 31.4993 14.5093 -999.9999 -999.9999 -999.9999 -999.9999
3500 13900 36.9185 14.0639 -999.9999 -999.9999 -999.9999 -999.9999
5100 1900 24.0156 15.1084 1.1407 0.0572 0.2268 -0.1696
9900 13700 35.0411 13.919 -999.9999 -999.9999 -999.9999 -999.9999
2900 900 28.4249 13.1304 0.3897 -0.4093 -0.1674 -0.2419
```

The -999.9999's represent missing values (permeability information is available only at 42 of the 85 wells).

First, fire up S-GeMS. Once it is open, I suggest maximizing it on your screen, which may require dragging the S-GeMS window to the left so that you can reach the maximize button in the upper right corner. (As I said, S-GeMS does not handle screen real estate very well.)

To load the Zone A data, select **Load Object...** from the **Objects** menu. Use the resulting Open dialog box to navigate to and open the file ZoneA.dat.

You will then be presented with an **Import from GSLIB** dialog box with a preview of the file. Set the object type to *point set* in the **Select Object type** dropdown box:



Then click **Next**> to get:

Import from Gslib

Additional Parameters

Pointset name:

Points Coordinates

Specify the columns containing the coordinates of the points

Col. # for X coordinates:

Col. # for Y coordinates:

Col. # for Z coordinates:

File Preview

| | |
|--|---|
| Zone A Data, Big Bean Field | |
| 8 | |
| X m | meters east of origin |
| Y m | meters north of origin |
| Thk m | zone thickness in meters |
| Por % | porosity in percent |
| Perm md | permeability in millidarcies |
| LogPerm - | base 10 log of permeability |
| LogPermPrd - | LogPerm predicted by regression against Por |
| LogPermRsd - | Residual from LogPerm-Por regression |
| 12100 8300 37.1531 14.6515 2.8547 0.4556 0.1357 0.3198 | |

No Data Value

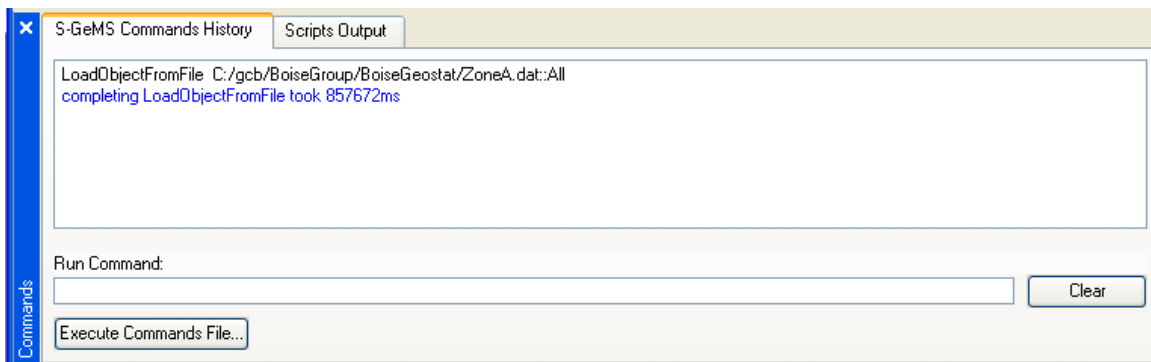
Use No-Data-Value

No-Data-Value:

Help < Back Finish Cancel

As shown above, set the **Pointset name** to ZoneA (or whatever you like), check **Use No-Data-Value**, and set the **No-Data-Value** to -999.9999. The default entries regarding the coordinate columns is correct – X is in column 1, Y is in column 2, and there is no Z (data are 2D). Click Finish and the data will be loaded.

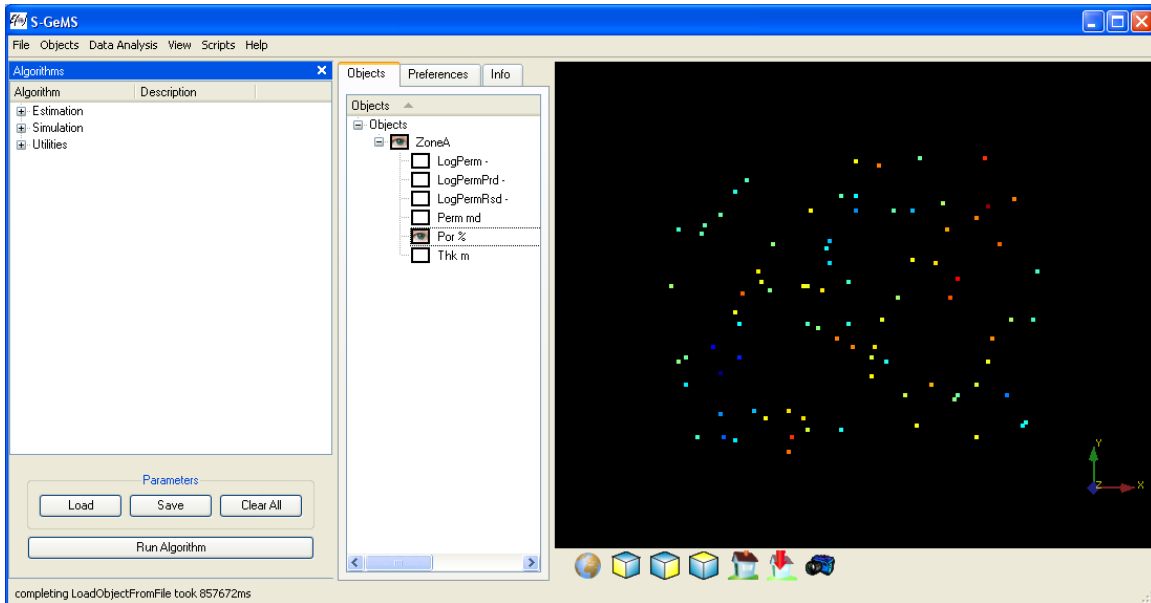
Note that if you have the **Commands Panel** displayed (at the bottom of the S-GeMS window) with the **S-GeMS Commands History** tab selected, you will get feedback on what S-GeMS is doing. This window will display the S-GeMS scripting language command associated with every action you perform and will tell you how long it took to complete each action. After loading the above file, it should show something like:



This feature is very handy, as it lets you learn the commands associated with each action, allowing you to subsequently enter individual commands into the **Run Command** text box or transfer a sequence of commands to a script and run the script (possibly after editing) using the **Execute Commands File...** button.

Viewing Data

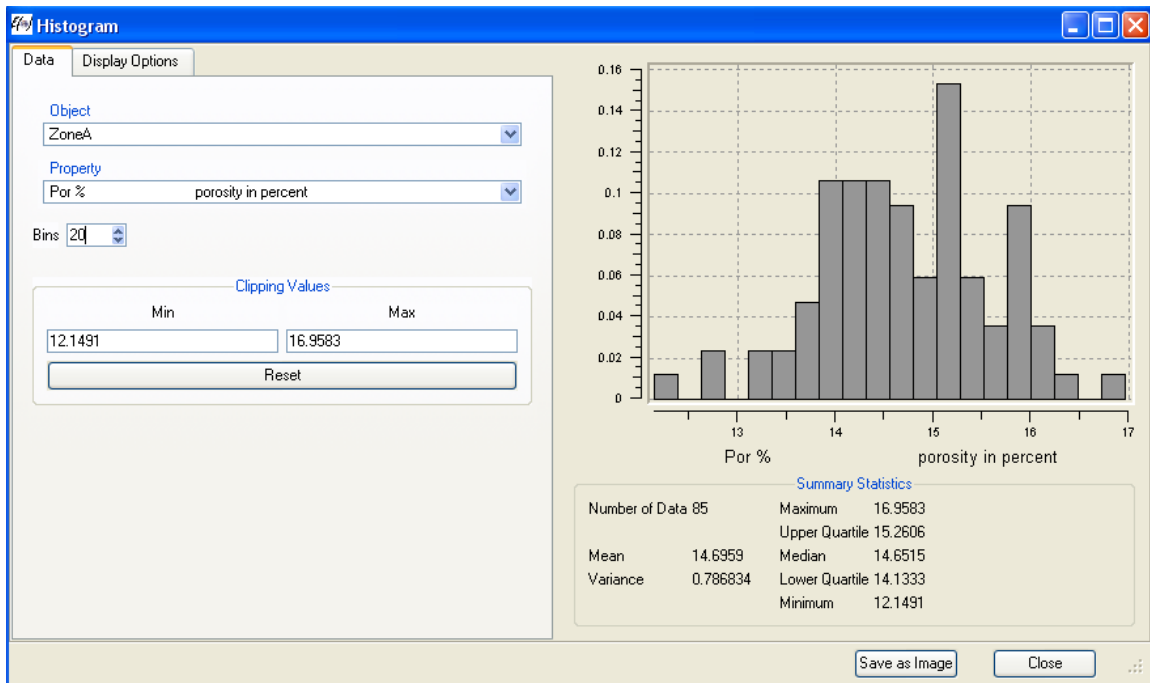
In between the Algorithms panel on the left and the data display window on the right, there is a panel containing **Objects**, **Preferences**, and **Info** tabs – actually this is part of the Visualization panel. To view the Zone A porosity data, expand **Objects** and then **ZoneA** on the **Objects** tab and check the boxes next to *both* **ZoneA** and **Por %**:



You may also need to click on the globe below the visualization window to get the data to display. The initial view will be a map view, looking straight down on the data, which is appropriate for this 2D dataset. However, you can rotate the data in 3D space, if you care to, by clicking in the visualization window and moving the mouse around. I won't elaborate on the visualization controls here – they are easy enough to figure out. Note however that holding down Shift+Ctrl allows you to use the mouse to zoom the data and holding down just Shift or Ctrl allows you to pan the data (without rotating). To get a color scale, go to the **Preferences** tab and check **Show colorbar of**, then select the object ZoneA and the property Por %. You may need to scroll to reach those controls.

Basic Data Analysis/Display

The **Data Analysis** menu contains a few options for examining data distributions. For example, to get a histogram of porosity, select Histogram from this menu, then select ZoneA in the **Object** dropdown box and Por % in the **Property** dropdown box. The result should look something like:

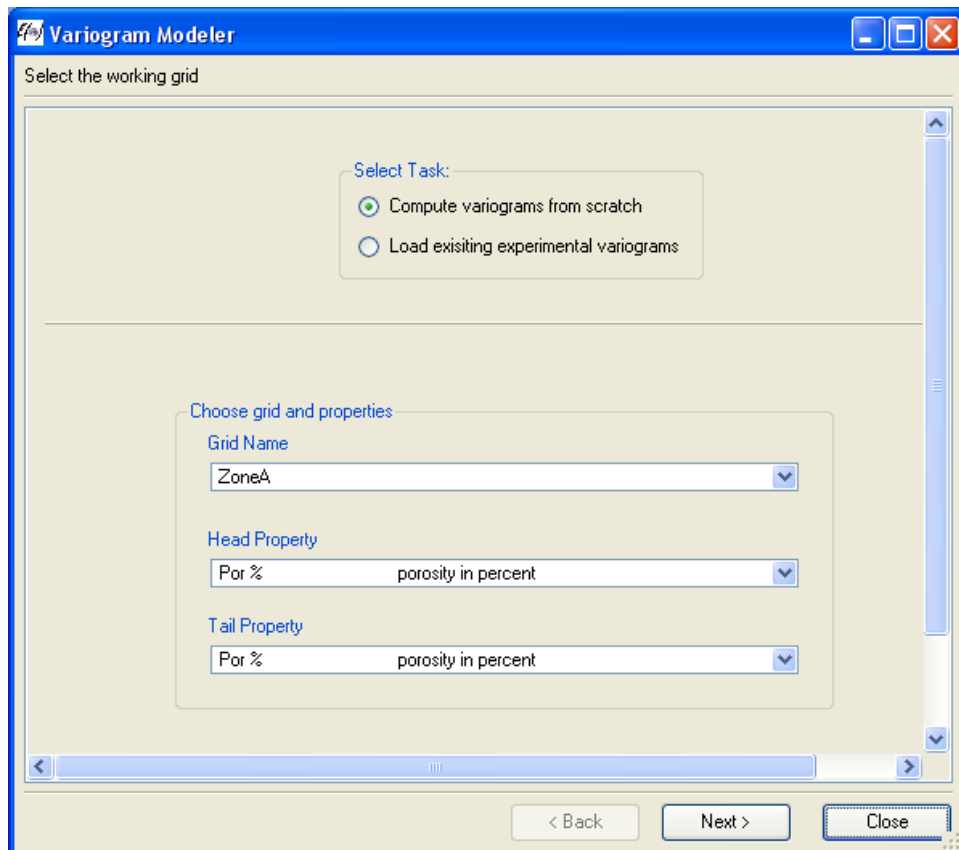


Note that this and other displays (such as the variogram) comes up in a new window – which may get hidden at some point, and needs to be re-displayed by selecting it on the task bar (or Alt-Tab). Also note that the control panel on the left side may need to be widened to allow access to all the controls.

As mentioned in the lecture notes, geostatistical methods for continuous variables are optimal when the data are normally distributed, so it is good to check for significant deviations from normality. The Zone A porosity data are not terribly non-normal, so we will continue analyzing them without transformation. (Be more careful in real work!)

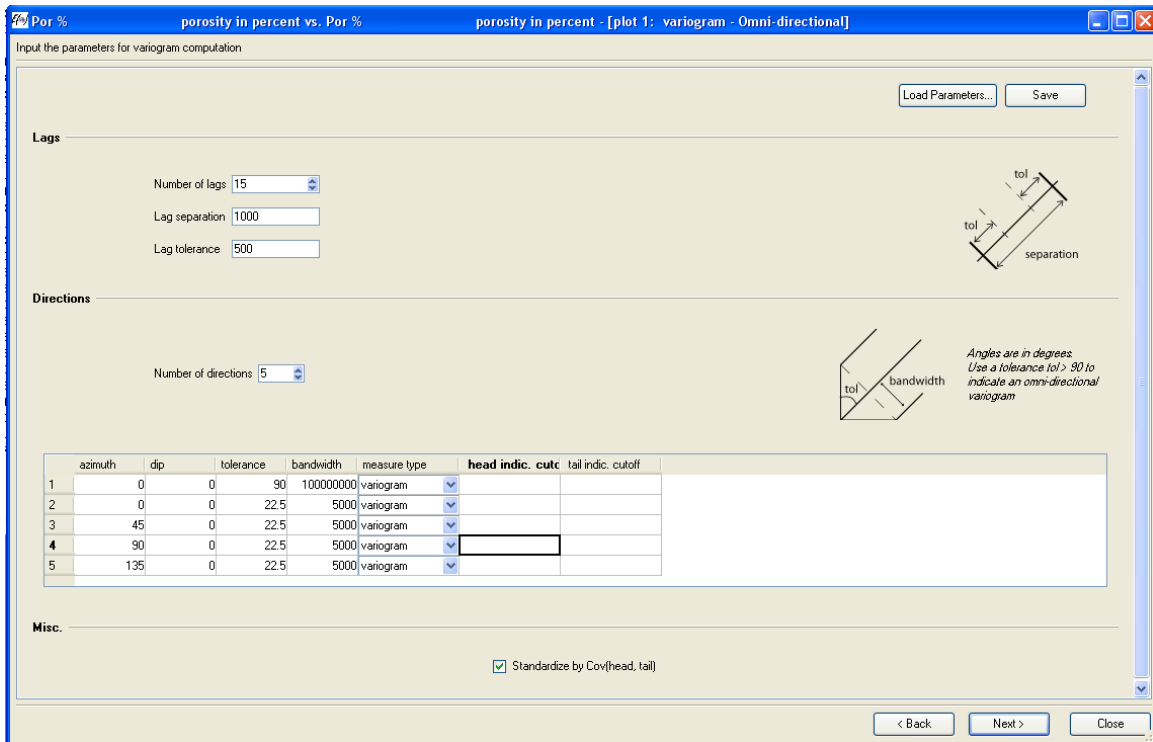
Variogram Analysis – Continuous Variable

To compute a variogram for the Zone A porosity data, select **Variogram** on the **Data Analysis** menu. On the **Variogram Modeler** dialog, leave **Select Task** at its default setting (compute variogram from scratch), select ZoneA under **Grid Name** (this really should be labeled **Object**), and then select Por % as both the head and tail properties:



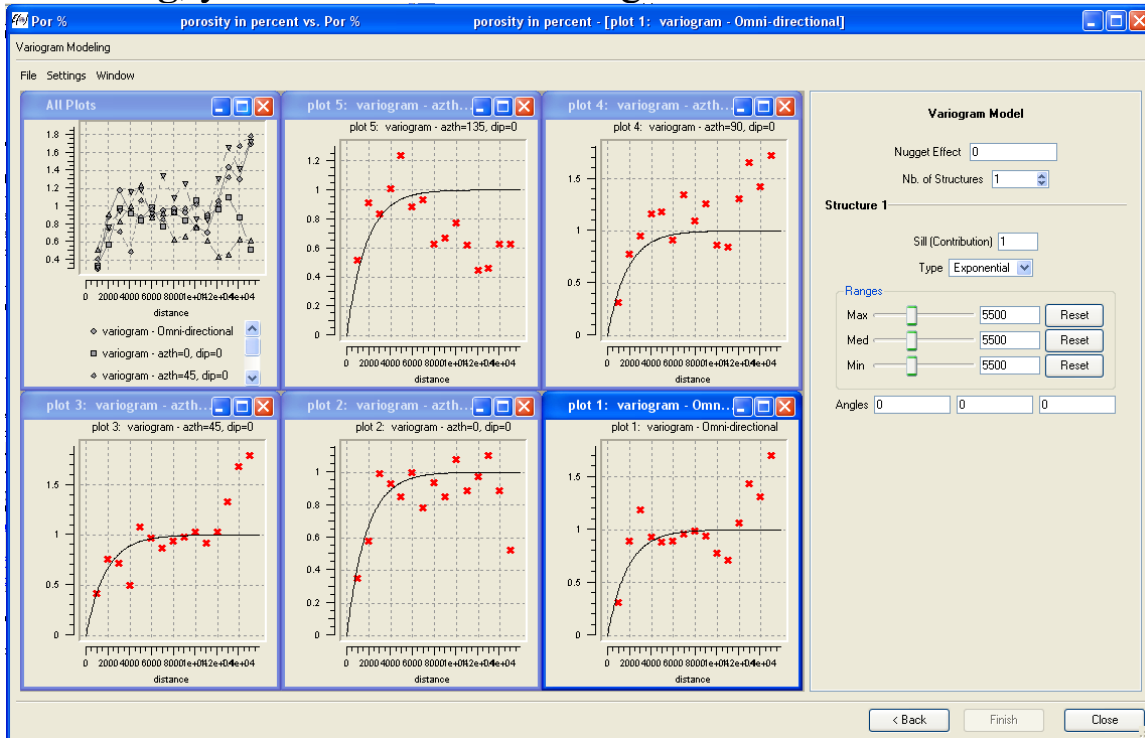
Then click next . . .

On the next dialog box, set the parameters as follows to compute one omnidirectional variogram (with an angular tolerance of 90 degrees and a very large bandwidth) and four directional variograms (at 0, 45, 90 and 135 degrees, each with an angular tolerance of 22.5 degrees and a bandwidth of 5000 m), computing 15 lags with a nominal lag spacing of 1000 m and a lag tolerance of 500 m. Also check **Standardize by Cov(head,tail)** to produce a scaled variogram with a sill of about 1. This gives us one less parameter to estimate (the sill) and matches expectations of the sequential Gaussian simulation we will do later.



These settings will compute lag results out to about 15 km, almost the full extent of the field in the north-south direction. In accordance with suggested rules of thumb, we will not pay too much attention to results beyond 10 km, about half the maximum separation distance between wells.

After setting up the variogram computation parameter above, clicking **Next>**, and then tiling the individual variogram plots by selecting **Tile** from the **Window** menu under **Variogram Modeling**, you should see something like:



The omnidirectional variogram is in the lower right, and all the plots together are in the upper left, with the four directional variograms in between. **Note:** The direction convention in S-GeMS is “Cartesian” (degrees counterclockwise from east), not geographic (degrees clockwise from north) like most other software. The directional variograms maybe show some hint of trend. Specifically, the 90 degree variogram looks fairly “trendy” while the 0 degree variogram looks fairly trend-free. Nevertheless, we will use a single omnidirectional model, exponential with a (practical) range of 5500 meters and a sill of 1. The sill has been imposed by our scaling by Cov(head,tail). Without this scaling, we would have seen the same variogram, but with a sill of about 0.8 – the variance of the porosity values. The **File** menu offers various options for saving and loading variograms.

Kriging a Continuous Variable

Before we krig the porosity, we need to create a grid for holding the kriged values. If you click on the Info tab and select ZoneA under Information on, you will see that the ZoneA dataset has a bounding box running from (100, 100) to (19500,15700) in meters. We will create grid a 100 x 80 grid running from X = 100 to 19900 meters in increments of 200 m and from Y = 100 to 15900 meters in increments of 200 m. Those specifications are for grid cell centers, so the grid will span 0 to 20 km in X and 0 to 16 km in Y, edge to edge. To do this, select **New Cartesian Grid** on the **Objects** menu, fill out the resulting dialog box as follows . . .

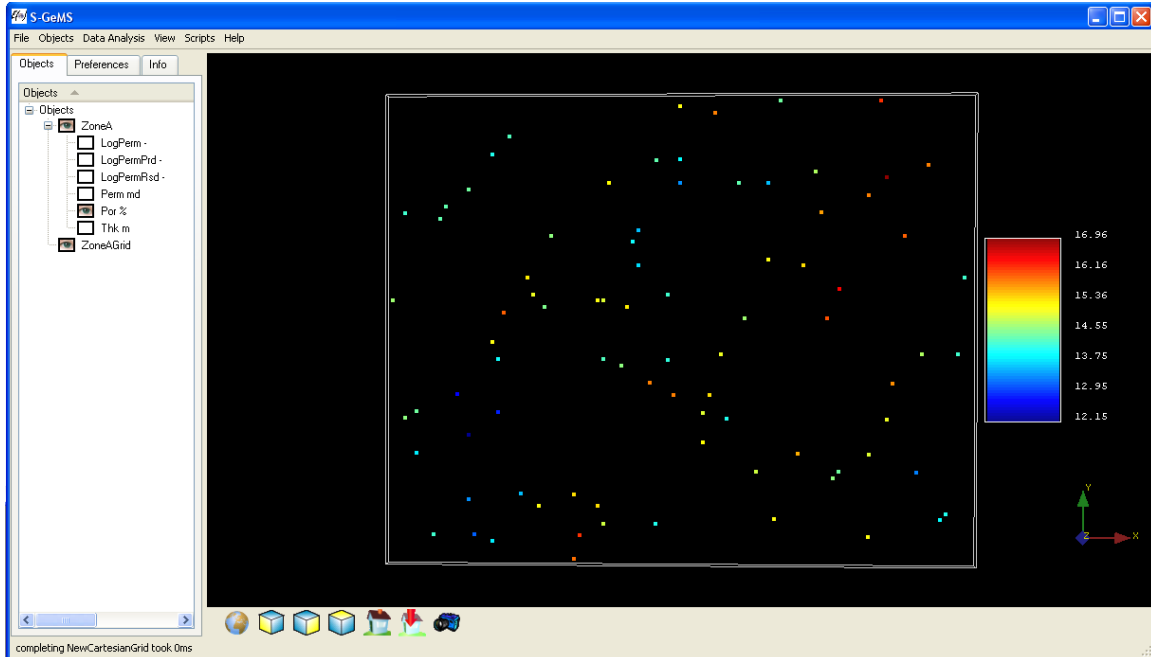
The screenshot shows a dialog box titled "Create new cartesian grid". It has a blue title bar with a question mark icon and a close button. The dialog box is divided into several sections:

- Grid name:** A text box containing "ZoneAGrid".
- Grid Dimensions:** Three spinners for "Number of cells in X" (100), "Number of cells in Y" (80), and "Number of cells in Z" (1).
- Cell Size:** Three text boxes for "Size in X" (200), "Size in Y" (200), and "Size in Z" (200). The label "The dimensions of a single cell" is above these boxes.
- Origin Coordinates:** Three text boxes for "Ox" (0), "Oy" (0), and "Oz" (0). The label "The coordinates of the lower left corner of the grid" is above these boxes.

At the bottom of the dialog box are two buttons: "Create Grid" and "Cancel".

and click **Create Grid**.

ZoneAGrid will then be added to the list of objects available for visualization. If you check its entry on the Objects tab, the grid's bounding box will be added to the display:



(I added the color scale to the display through the **Preferences** tab and close the **Algorithms** and **Commands** panels to maximize display space.)

To do the kriging thing, expand the **Estimation** entry on the **Algorithms** panel and select the entry **kriging**. The middle part of the **Algorithms** panel should then display two tabs, General and Data and Variogram, for specifying the parameter controlling the kriging process.

Fill out the General and Data tab as follows:

General and Data Variogram

Kriging Grid

Grid Name
ZoneAGrid

New Property Name
PorOK

Ordinary Kriging (OK)

Hard Data

Object
ZoneA

Property
Por % porosity in percent

Search Ellipsoid

Conditioning data

Min 2 Max 12

| | Max | Med | Min |
|---------------|-------|-------|-------|
| Ranges | 10000 | 10000 | 10000 |
| Angles | 0 | 0 | 0 |

This will use ordinary kriging to add a new property, PorOK, to ZoneAGrid, conditioned on the porosity values in the ZoneA point set and using a circular “search ellipsoid” with a radius of 10 km, requiring a minimum of 2 data points to be found within 10 km of each grid point and using no more than 12 nearest neighboring points in the estimation for each grid point.

Fill out the **Variogram** tab with the information on the porosity variogram we developed above. Here I have used the “raw” sill of 0.8, rather than the scaled sill of 1.0, since we are kriging the raw data (but this really makes no difference, except in changing the magnitude of the estimated kriging variance at each grid node):

General and Data Variogram

Variogram

Load existing model...

Nugget Effect 0

Nb of Structures 1

Structure 1

Contribution 0.8

Type Exponential

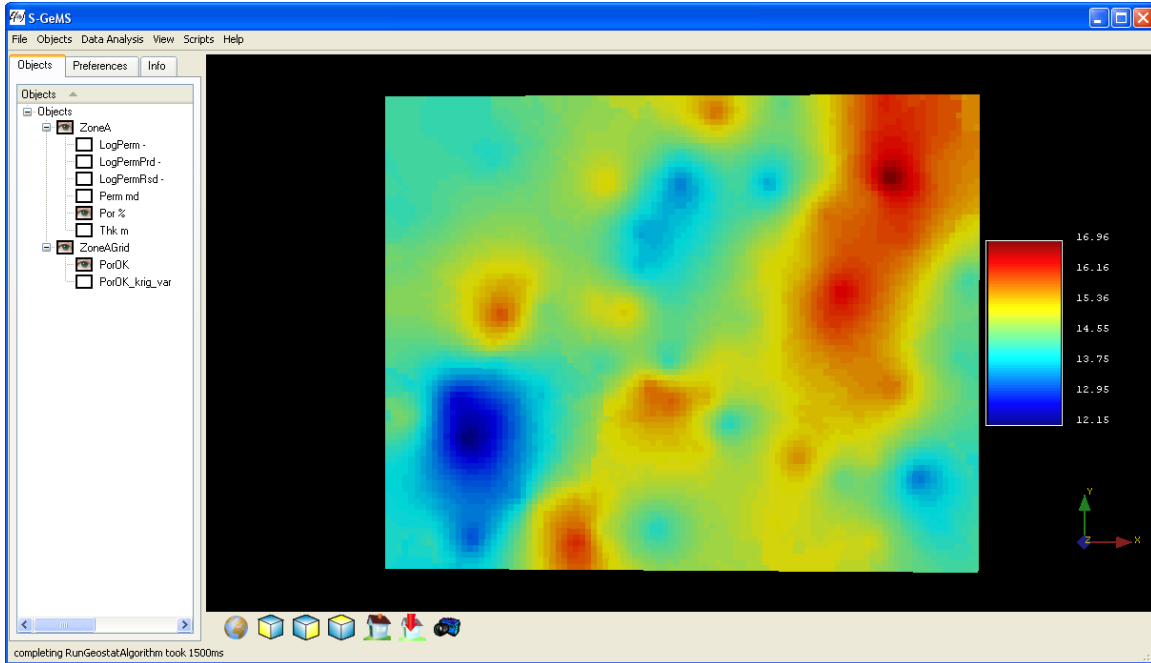
| | Max | Med | Min |
|---------------|------|------|------|
| Ranges | 5500 | 5500 | 5500 |

| | Azimu | Dip | Strike |
|---------------|-------|-----|--------|
| Angles | 0 | 0 | 0 |

If you saved the variogram model to a file earlier, you can reload that information here using the **Load existing model...** button.

After you have filled in all the parameters, click the **Run Algorithm** button down below to make it go . . .

Then, on the Objects tab for the visualization window, check the property PorOK under ZoneAGrid (which must also be checked) to display the kriged porosity grid:

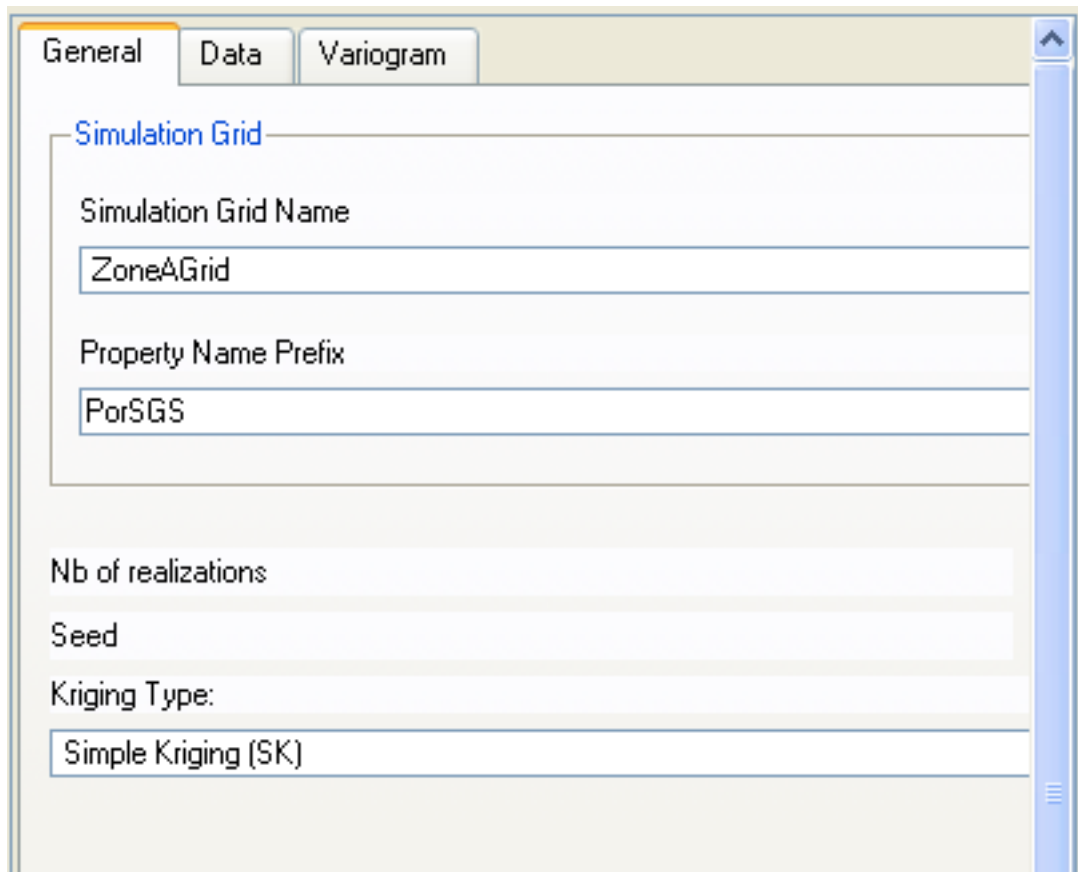


You can toggle the display of PorOK off and on, leaving the data values (Por % in the ZoneA object) on to see that the kriging has reproduced the data values.

If you wish, you can take a look at the kriging variance map, PorOK_krig_var. You will see that it looks very much like a map of the distance to the nearest well (control point).

Sequential Gaussian Simulation of Porosity

Next we will generate five realizations of porosity using sequential Gaussian simulation. To do this, expand **Simulation** on the **Algorithms** tab and set up the **General** tab as follows:



The image shows a software dialog box with three tabs: "General", "Data", and "Variogram". The "General" tab is selected. The dialog is titled "Simulation Grid" and contains the following fields:

- Simulation Grid Name: ZoneAGrid
- Property Name Prefix: PorSGS
- Nb of realizations: (empty)
- Seed: (empty)
- Kriging Type: Simple Kriging (SK)

What is hidden off to the right is the fact that the number of realizations is set to 5. You can leave the random number seed (“Seed”) at its default value. The sequential Gaussian simulation will actually use a normal score transform to turn the porosity values at the wells into a set of values that perfectly follow a standard normal distribution (zero mean, unit standard deviation) and will then generate grids of simulated values whose univariate distribution is also standard normal. Therefore we are using Simple Kriging – the spatially constant mean will be assumed to be zero.

Set up the Data tab more or less as follows:

The screenshot shows the 'Data' tab of a software interface. It is divided into several sections:

- Hard Data:**
 - Object: ZoneA
 - Property: Por % (porosity in percent)
 - Assign hard data to simulation grid
- Search Ellipsoid:**
 - Max conditioning data: 16
- Angles Table:**

| | Max | Med | Min |
|---------------|---------|-------|-------|
| Ranges | 10000 | 10000 | 10000 |
| | Azimuth | Dip | Rake |
| Angles | 0 | 0 | 0 |
- Target Histogram:**
 - Use Target Histogram
 - Reference distribution from:
 - From Object
 - From File
 - Property with ref. distribution:
 - Object: ZoneA
 - Property: Por % (porosity in p)
 - Lower Tail Extrapolation:
 - Function: Power
 - Min: 12
 - omega: 3
 - Upper Tail Extrapolation:
 - Function: Power
 - Max: 17
 - omega: 0.333

We have checked Use Target Histogram so that the standard normal simulations will be backtransformed to match the histogram of the actual porosity values after they are generated. This backtransformation involves extrapolation in the tails of the distribution, for which the user has to set minimum and maximum allowed limits for the backtransformed values. I have used 12% and 17%, just outside the limits of the actual data.

Set up the Variograms tab to represent the “unit-sill” version of the variogram we developed earlier:

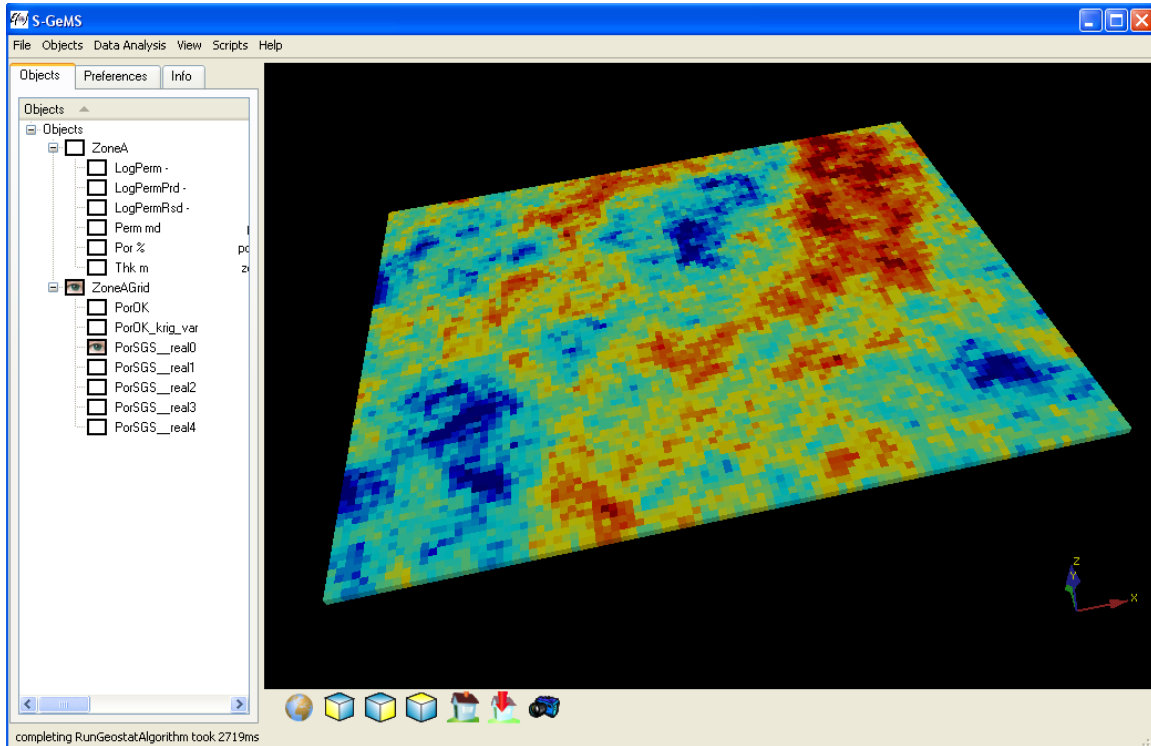
The screenshot shows the 'Variogram' tab in a software interface. At the top, there are three tabs: 'General', 'Data', and 'Variogram'. Below the tabs, there is a 'Variogram' section with a 'Load existing model...' button. Underneath, there are input fields for 'Nugget Effect' (set to 0) and 'Nb of Structures' (set to 1). Below this, there is a section for 'Structure 1' with a 'Contribution' field (set to 1.0) and a 'Type' dropdown menu (set to 'Exponential'). At the bottom, there are two tables. The first table has columns 'Max', 'Med', and 'Min', with a row labeled 'Ranges' containing the value 5500 in all three columns. The second table has columns 'Azimuth', 'Dip', and 'Rake', with a row labeled 'Angles' containing the value 0 in all three columns.

| | Max | Med | Min |
|---------------|------|------|------|
| Ranges | 5500 | 5500 | 5500 |

| | Azimuth | Dip | Rake |
|---------------|---------|-----|------|
| Angles | 0 | 0 | 0 |

What the SGS algorithm really wants is the variogram of the normal-score transformed data, which we have not computed. We are taking a short-cut, assuming that the variogram of the normal-score transformed data would look very much like the variogram of the raw data scaled to a unit sill. This is the case for these data, since the shape of the univariate porosity distribution is in fact reasonably normal.

Click on **Run Algorithm** to make it go and the software should generate five equiprobable realizations of porosity, as properties in ZoneAGrid:



Sequential Indicator Simulation of Facies

To illustrate sequential indicator simulation of facies, we will use data from the Green Goblin Gas Field, a Permian gas field somewhere in the US mid-continent. The facies values have been generated from geophysical well logs using a neural network trained on observed log-facies associations in cored wells. The headers and first few lines of data look like:

```
Facies data, Green Goblin Gas Field
11
WellNo
Depth
X   ft
Y   ft
Z   ft
F05
Ind01
Ind02
Ind03
Ind04
Ind05
12 3031.5 65826.02 17091.19 -176.5 2 0 1 0 0 0
12 3029.75 65826.02 17091.19 -174.75 4 0 0 0 1 0
12 3027.75 65826.02 17091.19 -172.75 3 0 0 1 0 0
12 3025.75 65826.02 17091.19 -170.75 4 0 0 0 1 0
```

As indicated, the coordinates, X, Y and Z are in feet. X and Y are measured from an arbitrary origin. Z is elevation relative to sea level. F05 is an integer running from 1 to 5 representing each of the five different facies. Ind01 – Ind05 are indicator representations of the facies occurrence. That is Ind01 is 1 if F05 = 1 and 0 otherwise. The indicators can be thought of as a set of “hard” probabilities, representing certainty of which facies occurs at each data point. (In truth we are really not certain, and a “soft” or probabilistic representation of the neural net predictions would really be more appropriate.)

The facies sequence (and this is only a piece of it) was developed as the seas rolled in and the seas rolled out across a very broad, gently dipping shelf. The facies codes represent:

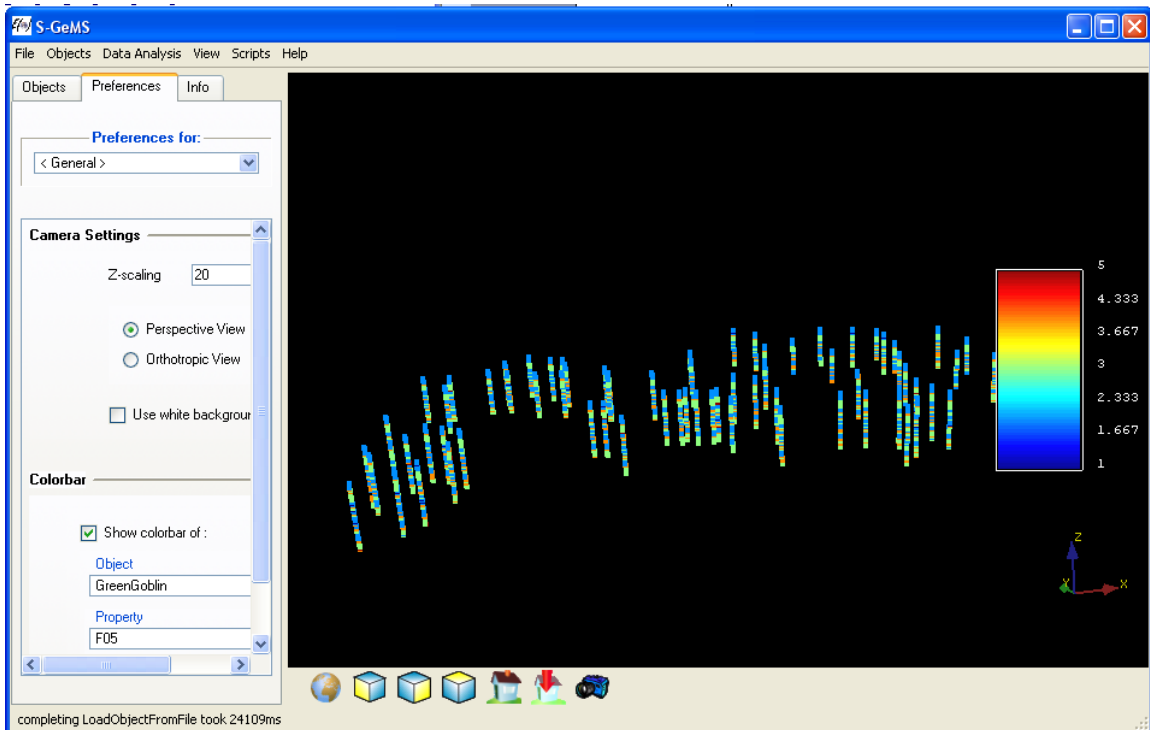
- 1 – continental sandstone
- 2 – continental siltstone
- 3 – mud-supported limestone
- 4 – grain-supported limestone & dolomite
- 5 – marine sandstone

Late we will need the proportions of each facies in the dataset. These are (in decimal):

1: 0.008, 2: 0.437, 3:0.368, 4: 0.183, 5: 0.004

To get started, use the **Load Object** on the **Objects** menu to read in the file GreenGoblin.dat, creating a point set named GreenGoblin. The coordinates are in columns 3, 4, and 5 and you do not need to specify a No-Data-Value because there are no missing values. After the file is loaded you can use the Info tab to see that there are 8553 data points and to get information on the bounding box of the data.

On the **Objects** tab, check both GreenGoblin and F05 to display them in the visualization window. On the preferences tab, set the Z-scaling to 20 (the maximum allowed) and add the colorbar of F05 to the display to get a rendition of the facies values at the 74 wells:



Then use **New Cartesian Grid** on the **Objects** menu to create a 101 x 66 x 128 grid with spacings of 660 feet in the X and Y directions and 2 feet in the Z direction, with origin coordinates of 0, 0 in X and Y and -253.75 in Z. You can name it something like GreenGoblinGrid. Check the new grids entry on the **Object** tab to display its bounding box. Make sure that the grid encompasses the data.

Variogram analysis for indicator variables

For SIS, we need to specify a variogram for the indicator variable representing each category. We will go very briefly through the indicator variogram development for facies 3 (the most abundant facies), represented by Ind03. From the **Data Analysis** menu, choose **Variogram**, and then choose Ind03 from GreenGoblin as the variable to analyze. First we will look at the vertical variogram, setting up the lag and direction parameters as follows:

Input the parameters for variogram computation

Lags

Number of lags: 75
Lag separation: 2
Lag tolerance: 1

Directions

Number of directions: 1

| | azimuth | dip | tolerance | bandwidth | measure type | head indic. cutc | tail indic. cutoff |
|---|---------|-----|-----------|-----------|--------------|------------------|--------------------|
| 1 | 0 | 90 | 1 | 1 | variogram | | |

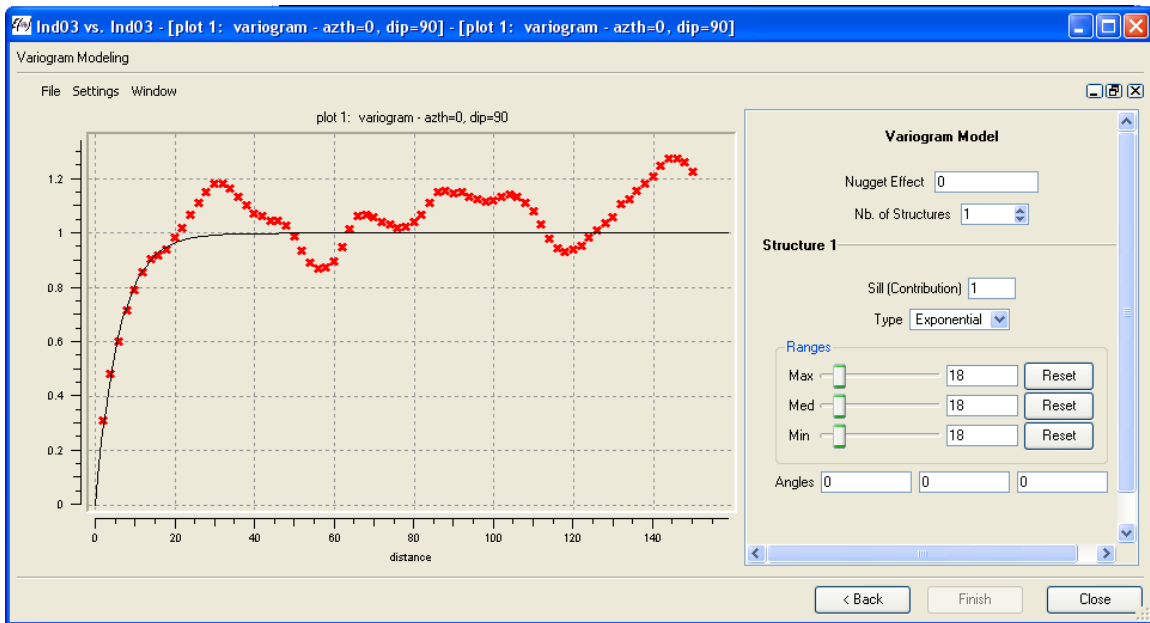
Misc.

Standardize by Cov(head, tail)

< Back Next > Close

Also check **Standardize by Cov(head,tail)** to generate a unit-sill variogram. (It will be scaled by $p*(1-p)$, where p is the overall proportion of the facies. This is the variance of a Bernoulli distributed variable.)

The early lags of the vertical variogram can be reasonably fit with an exponential structure with a range of 18 feet:



The horizontal variogram modeling gets tricky . . .

I have done the rest of the variogram modeling in R using the gstat add-in and the results are encapsulated in the S-GeMS SIS parameter file GreenGoblinF05sis.par, which you can read into S-GeMS.

Select **sisim** under **Simulation** on the **Algorithms** tab and then click load to open up the parameter file GreenGoblinF05sis.par. This should fill in the General, Data, and Variogram tabs with the appropriate settings, which you can explore. Then click Run Algorithm and wait a while to get something like . . .

