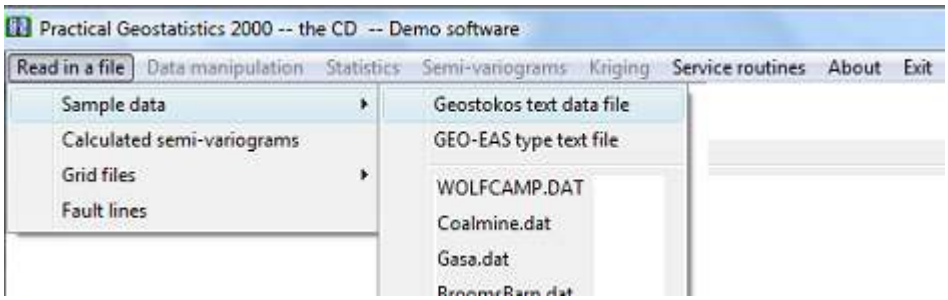
Software Tutorial Session – Three dimensional Semi-variograms

The example session with **PG2000** which is described below is intended as an example run to familiarise the user with the package. This documented example illustrates one possible set of analyses which may be carried out on data which has been collected from a three dimensional space. This tutorial takes you through the following sequence of analyses:

- Reading in a data file
- Calculating and interpreting a semi-variogram for 3D data

There are many other facilities within the package, which are given as alternative options on the menus. To start the tutorial, choose **PG2000** from your Start menu. See Tutorial One for starting up and specifying your ghost file output.

Reading in a data file



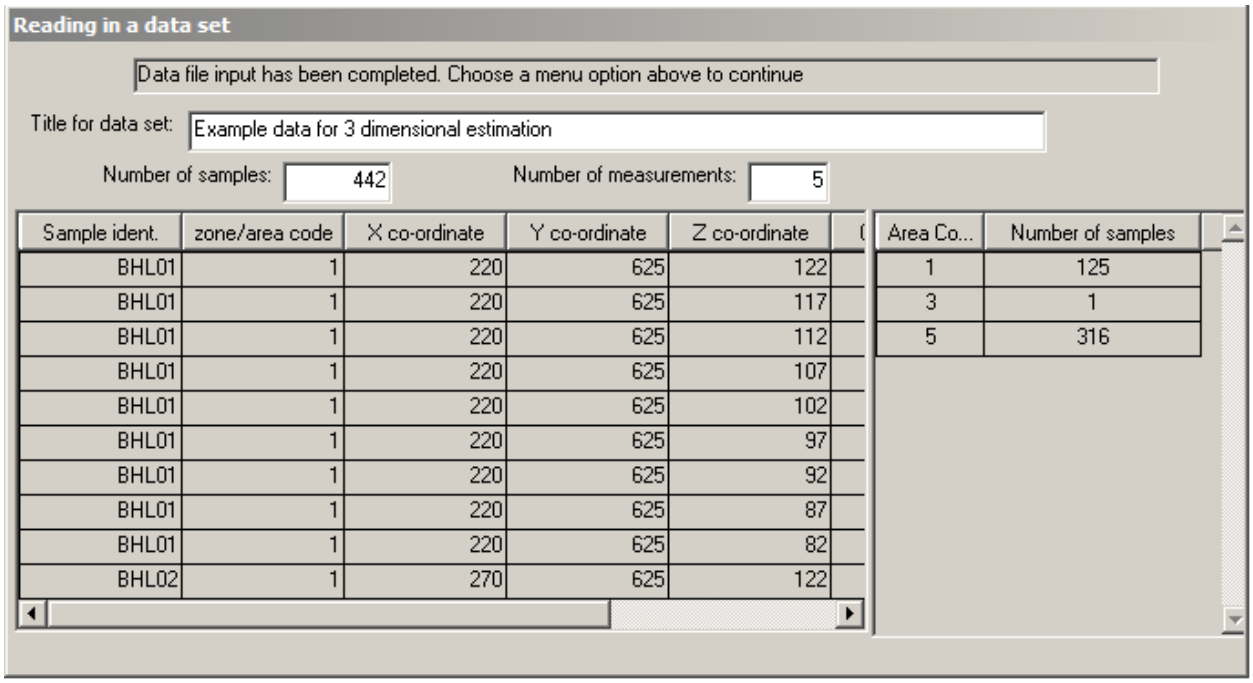
As you can see from the above I have elected to read in a set of sample data by clicking on the Read in a file option and selecting Sample data from the menu which appears. **PG2000** will remember the last five data files accessed and include these in your options. Three input file types can be read in. I will read in a standard Geostokos data file.

The layout of such files is described in detail in the main **PG2000** documentation. The routine which reads in the data shows the first 10 lines of your data file so that you can check it is going in OK. The routine also checks whether we actually had the correct number of samples on the file and informs you if there is any discrepancy.

Even if you select a file from the list of previously analysed data files, **PG2000** will ask you to confirm your choice. This is actually a quick way of getting back to your working directory, since you can change your choice at this point. Be warned, though, that if you change which file you want to read it must be the same type of file – that is, if you are reading a standard Geostokos data file, you cannot change your mind at this point and read in a CSV type file.

For this illustration, I have selected `copper.dat` for my input data file. This is a set of 442 borehole samples drilled into an unspecified mine dump in Eastern Europe. Copper grade measurements are made on each sample and the length of each core section is recorded. All co-ordinates are in metres.

As your data is read in, it is stored on a working binary file. A progress bar will indicate how far the process has gone. When data input is complete, your Window should look like this:



Sample ident.	zone/area code	X co-ordinate	Y co-ordinate	Z co-ordinate
BHL01	1	220	625	122
BHL01	1	220	625	117
BHL01	1	220	625	112
BHL01	1	220	625	107
BHL01	1	220	625	102
BHL01	1	220	625	97
BHL01	1	220	625	92
BHL01	1	220	625	87
BHL01	1	220	625	82
BHL02	1	270	625	122

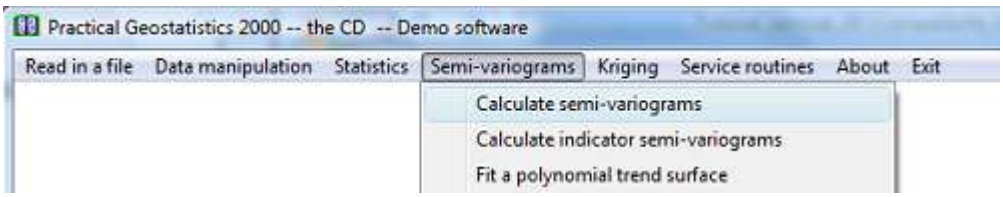
Area Co...	Number of samples
1	125
3	1
5	316

Note that this data set also includes ‘geology’ or ‘zone’ coding. A column on the data file includes an integer code which was of some meaning to the original logger. Perhaps a lithology or a dating code (since this is a dump). This enables you to select data by code for each analysis and has the effect of ‘separating’ samples in analyses such as semi-variogram calculation.

NOTE: for this data you will need to switch into 3 dimensions. Use the service menu to do this before continuing.

Semi-variogram calculation

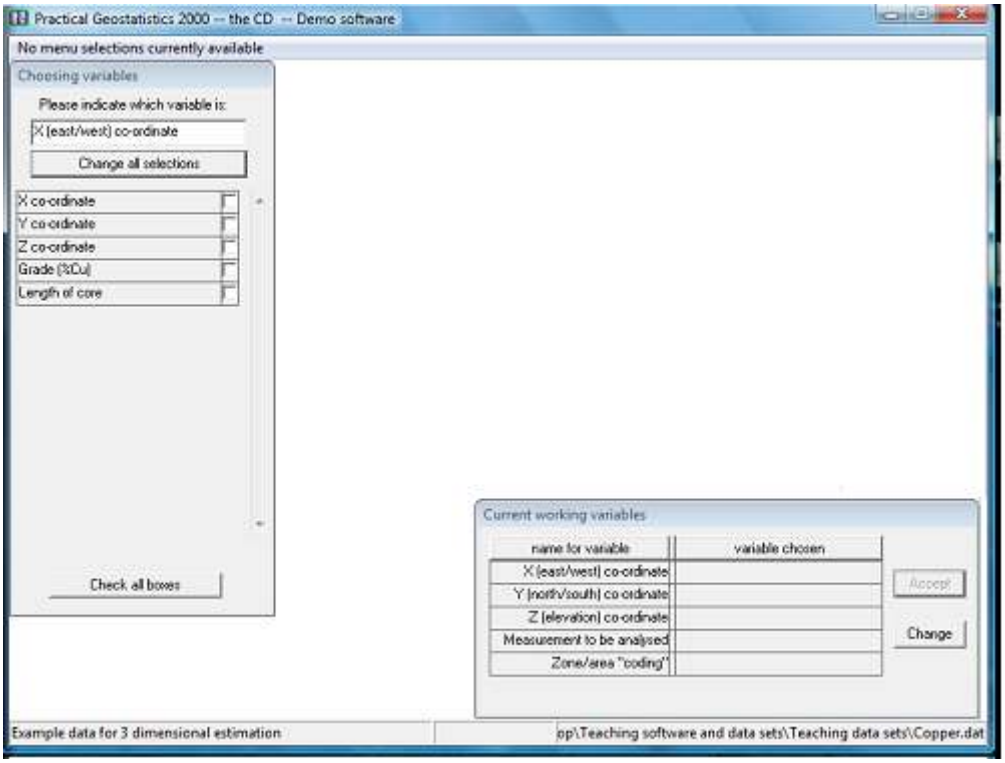
When the data has been read in, you will see that the “greyed out” options on the main menu bar will be activated. We use the menu bar to select an option, say:



The screen will prompt you to choose the four variables for the analysis – X, Y and Z co-ordinates and the measurement which is to be analysed.

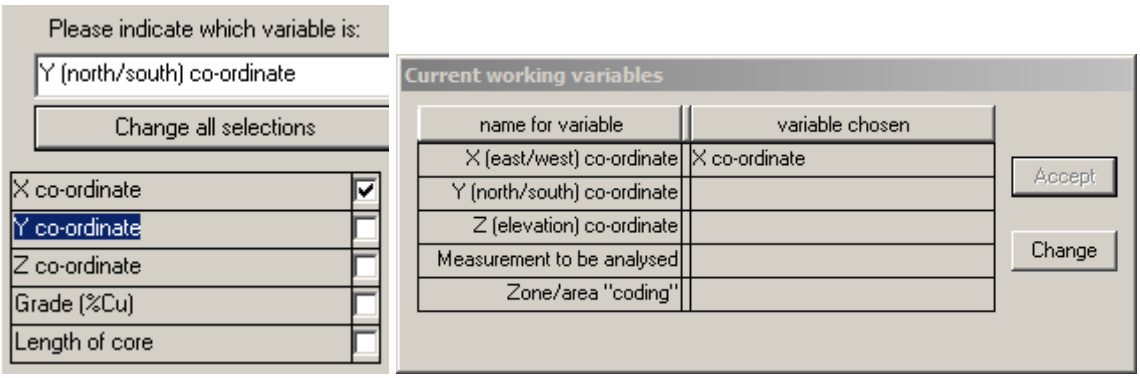
The routine, needs to have information on the position of the samples and on the value at each sample location. This particular data file only contains three variables. However, at this stage, **PG2000** does not know which of these variables is which.

You will see two dialog boxes. The one in the top left hand corner lists the variables available for analysis in your data file and the bottom right box shows the variables already chosen (at this point, none!).

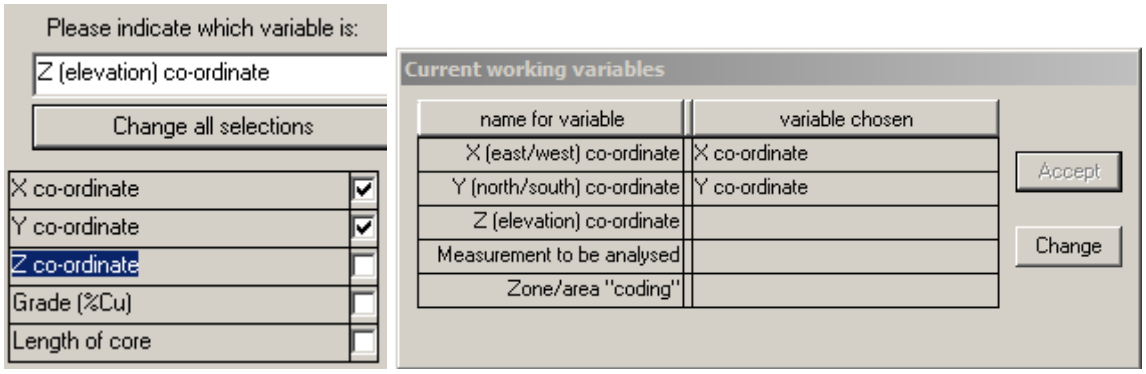


There is a lot of information on the screen. At the bottom of the Window, you see the "status bar" which shows the name of the current data file and the title read from that file. The "already chosen" dialog box shows you that you are expected to select variables to be the "X (east/west) co-ordinate", "Y (north/south) co-ordinate" and "Measurement to be analysed" for your semi-variogram.

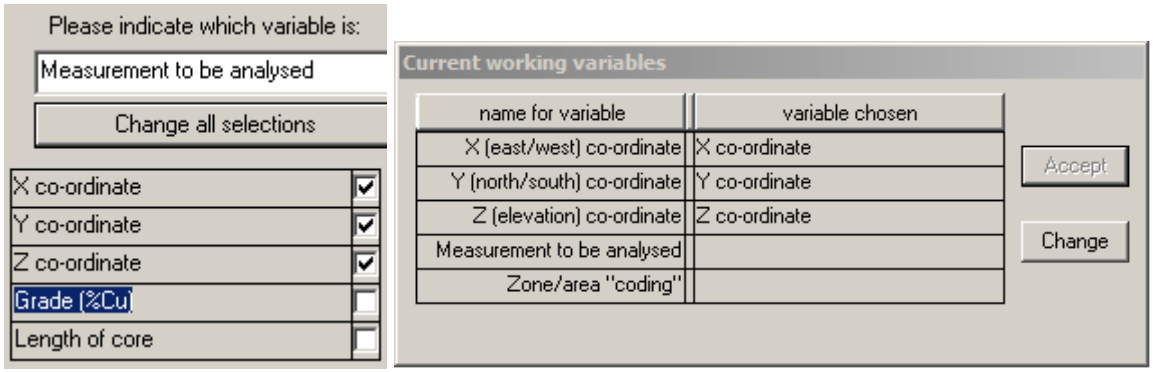
The upper left dialog box lists the variable names as they appeared in the data file and is prompting you to choose the variable which will be the "X co-ordinate" on the graph. For this example, let us choose X co-ordinate for the X co-ordinate:



We may then choose "Y co-ordinate" for the Y co-ordinate:

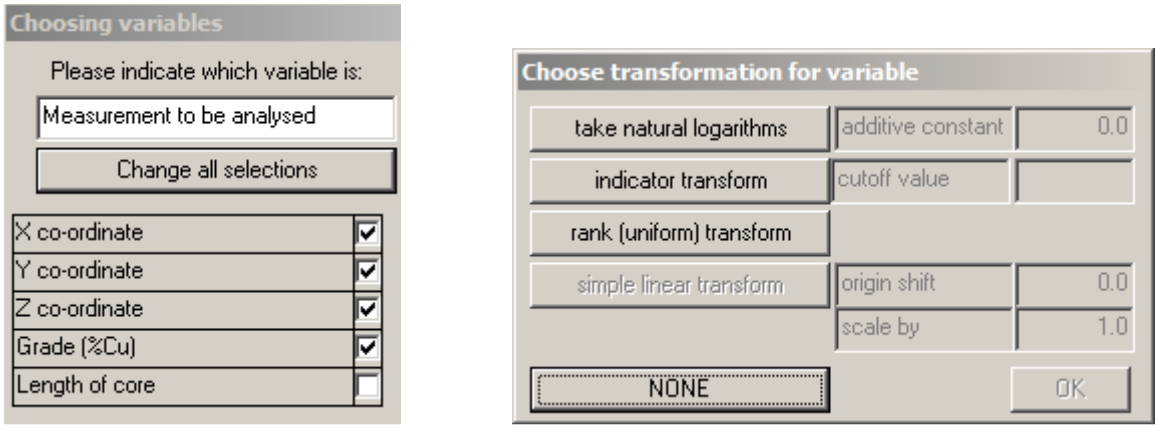


Since we are working in three dimensions, we need a third co-ordinate:



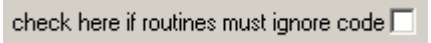
We must choose the variable to be analysed and state any relevant transformations to be made. For this data we require no transformation of the variable “Grade (%Cu)”, so click on 

.

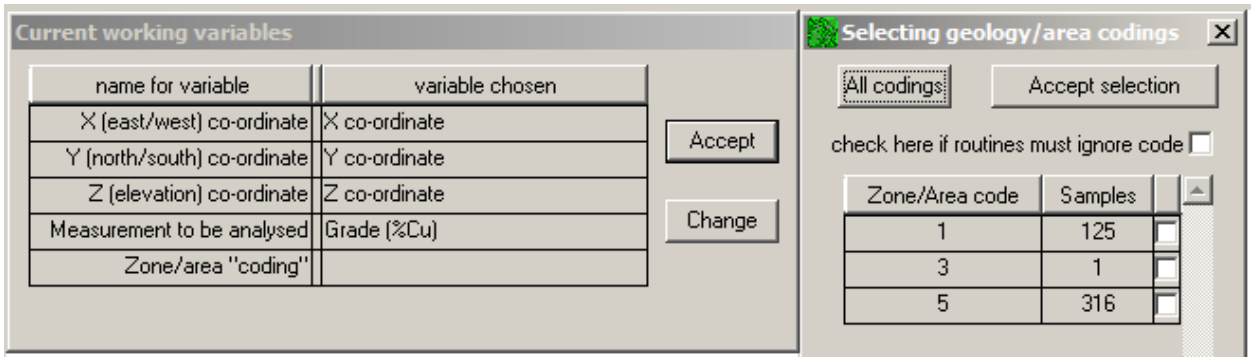


Since this data file includes ‘zone’ codes, you will need to specify which codes are to be included in the analysis and whether the routine should distinguish between them.

You may select any combination of zone codes by clicking in the check boxes.

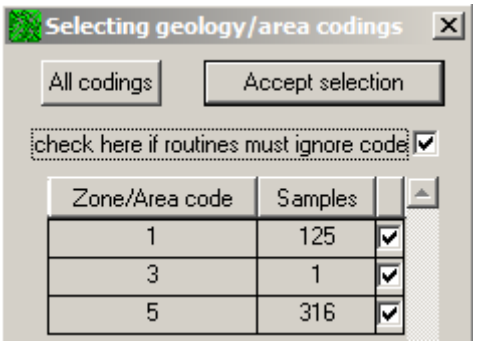
Once selected, use the 

to tell the routine whether or not to distinguish between samples with different zone codes. The semi-variogram calculation will not include pairs of samples with different codes unless you check this box.

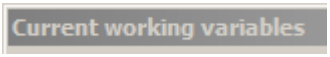


For this illustration, we will select all codes by clicking on 

and check the 'ignore code' box:

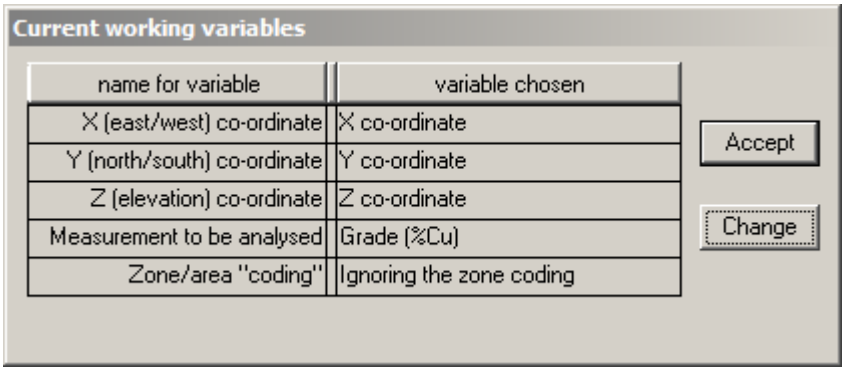


Once we click on 

, the 

dialog will show the complete set of chosen variables. You still have the option to change your mind here by clicking on 

.

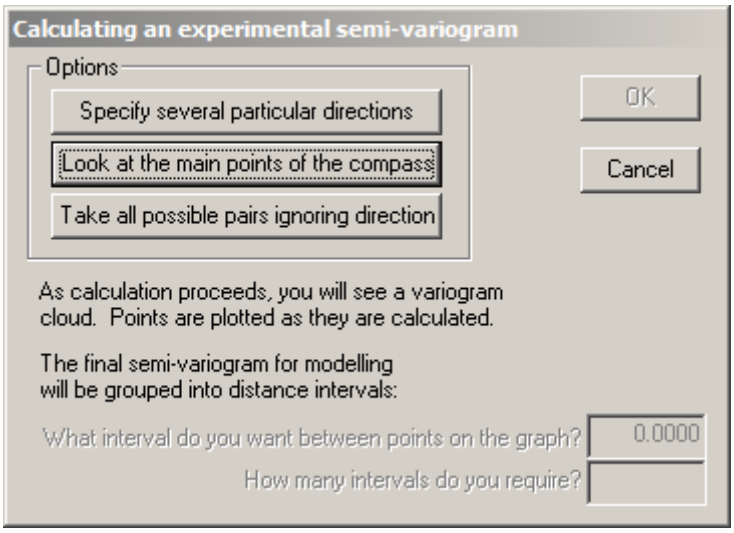


This choice of variables is acceptable, so click on 

to proceed. This may seem tedious to you at the moment, but (later) try running the program with another set of data with more variables. Or try a data set where the columns are in a different order. The **PG2000** input routine has been written to allow you this flexibility in building your data files.

Now, we may finally proceed to calculating a semi-variogram. For the complete data set, samples are paired up. The difference between the values of the two samples is calculated and squared. Plotting each of these points on a graph □ squared difference versus distance □ results in a "variogram cloud".

For the semi-variogram interpretation and modelling routines the "differences" are grouped together into "distance" intervals. That is, all pairs of samples which are more or less the same distance apart are grouped together and the differences averaged. To do this, you must choose a distance interval and a number of groups. The maximum distance considered will be the product of these two values.



You have the opportunity to specify your own directions, in which case you will need to define direction as *azimuth clockwise from North* and *dip down from horizontal*. For user defined directions, you must also specify a tolerance angle to be allowed on either side of your specified azimuth and dip directions.

You can simply make a graph which ignores direction entirely and groups all possible pairs of samples into one semi-variogram.

Alternatively, you may accept the thirteen default directions: 0°, 45°, 90° and 135° (North, Northeast, East and Southeast) horizontal, 0°, 45°, 90°, 135°, 180°, 225°, 270° and 315° at a dip of 45° down from horizontal plus a 'vertical' semi-variogram for all directions dip 90°. If you choose the "main points of the compass" you will also get the "omni-directional" semi-variogram. The default directions allow 22.5° either side, so that the four directions cover all possibilities.

For borehole data, accepting the default directions may not be appropriate, since you have to select a single interval width to be used in all directions. For this illustration, we will choose our own directions and specify azimuth, dip and tolerances as well as interval width and number of intervals.

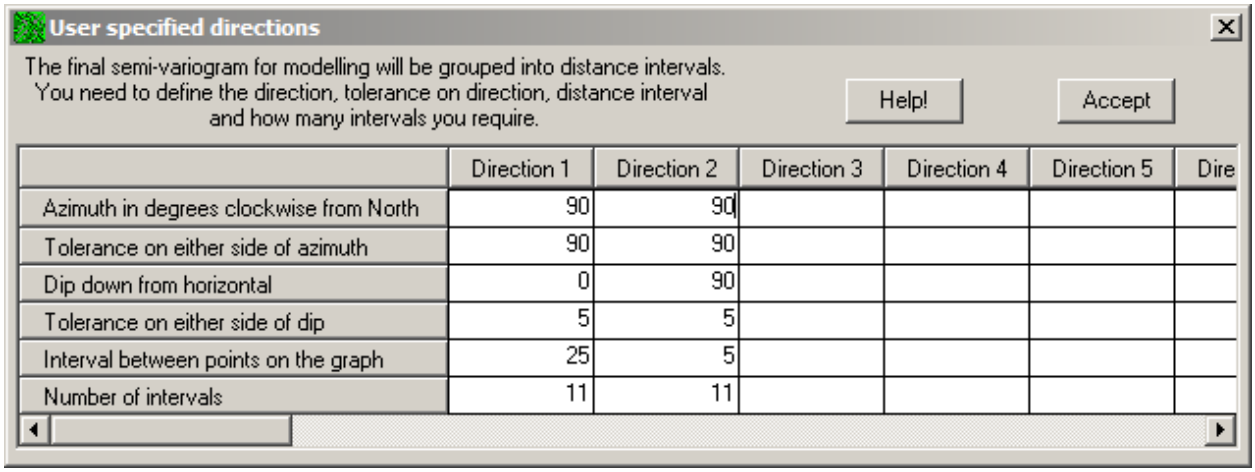
It is a good idea to have run an exercise such as that described in Tutorial Two, to have a good idea of your inherent sample spacing and the extent of your study area before selecting interval widths and maximum distances.

Click on 

and a new dialog will appear.

You can define up to 24 different directions by scrolling along the bottom of the dialog. I will specify two directions only for this example:

1. all possible pairs in the horizontal directions: azimuth 90° ± 90°, dip 0° ± 5°
2. the vertical (down-hole) semi-variogram: azimuth 90° ± 90°, dip 90° ± 5°

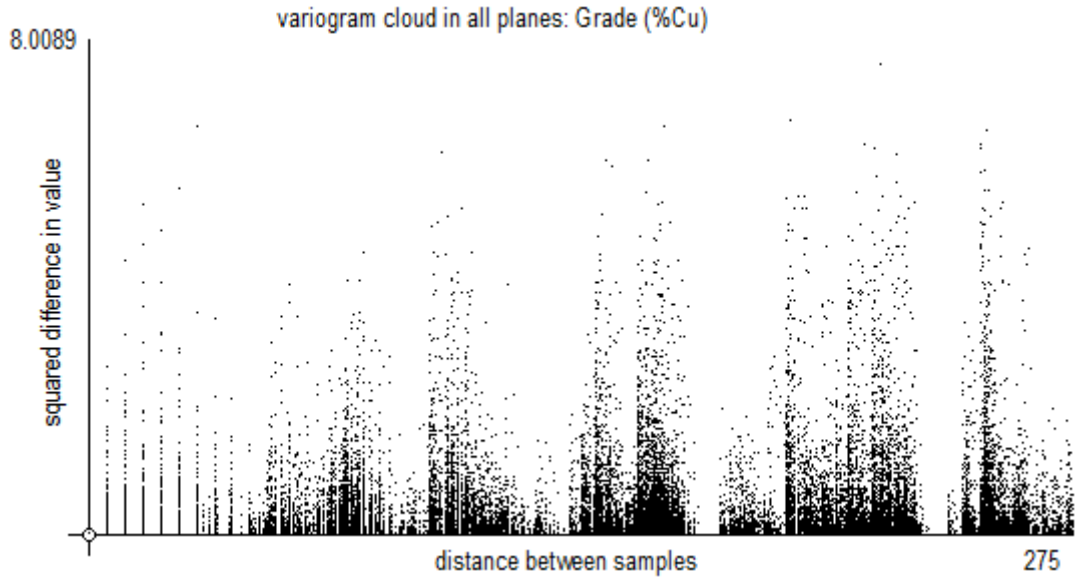


	Direction 1	Direction 2	Direction 3	Direction 4	Direction 5	Dire
Azimuth in degrees clockwise from North	90	90				
Tolerance on either side of azimuth	90	90				
Dip down from horizontal	0	90				
Tolerance on either side of dip	5	5				
Interval between points on the graph	25	5				
Number of intervals	11	11				

For the copper data, the average inter-borehole spacing is around 100 metres. For the 'between hole' semi-variogram I have chosen an interval of 25 metres. For the down-the-hole semi-variogram, I have chosen the length of the core sections, 5 metres.

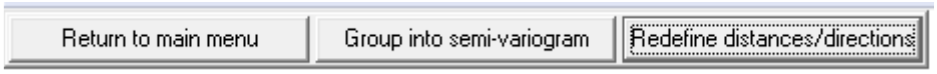
You also have a choice as to how much graphical output should occur. The graphics slow down the calculation process. If you have a very large data sets, you should minimise the amount of graphical activity to speed up the routine. In the following illustration, we use full graphics.

A progress bar and four plots will appear on your screen to let you know that the calculation is proceeding.



The variogram cloud shows the complete calculation process. For each pair of samples a pixel is plotted showing the distance between them and the square of the difference in value squared. In addition to this plot, three plots show a pixel illustrating how far apart the samples are in the X direction and in the Y direction respectively (top left), X and Z directions (top right) and Y and Z directions (bottom right).

When all the pairs have been identified, semi-variogram maps are plotted for the three main directions, and an options bar will appear:

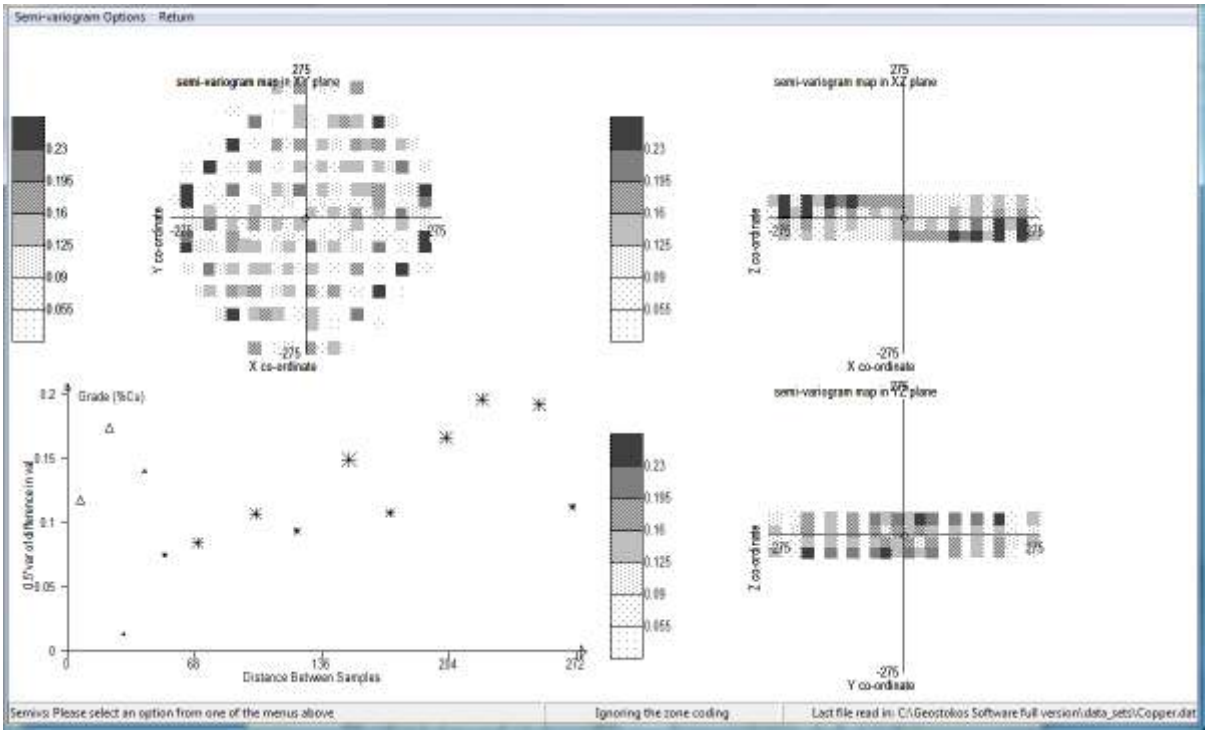


For large data sets, the grouping process can take considerable time. Because of this, you are offered the opportunity to scrutinise the semi-variogram maps and decide whether your directional parameters should be re-specified before grouping. In particular, reducing the maximum distance considered will speed up the grouping process considerably.

Click on 

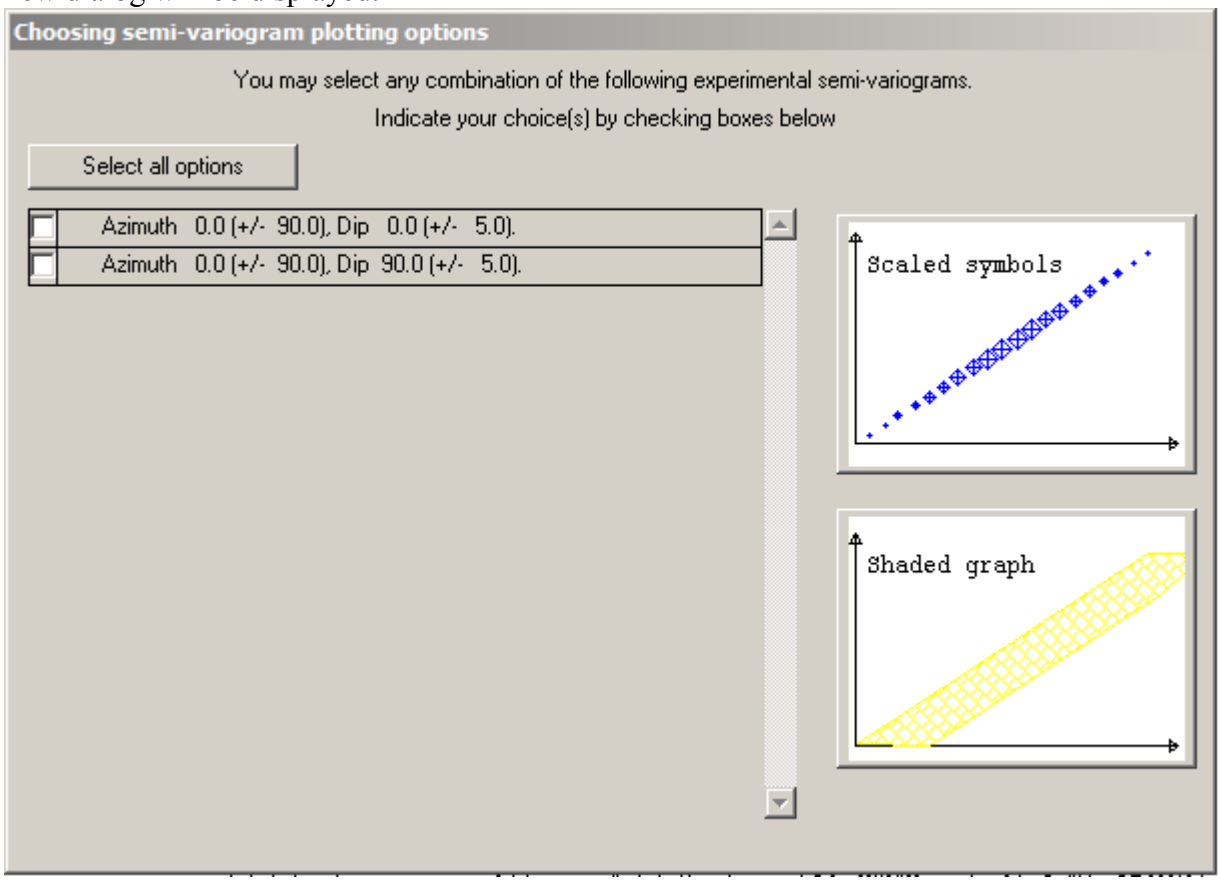
to continue with the calculations.
With our choices, all pairs of samples between 2.5 and 7.5 [5 ± 2.5] metres apart will be grouped together in the vertical (down-hole) direction. For each of these pairs, the difference in value will be calculated and squared. All of these values will be added together and divided by twice the number of pairs. This calculation will result in one point to be plotted on our final semi-variogram graph. This process will be repeated for all pairs of samples between 7.5 and 12.5 metres apart, and so on.

When the calculation is finished, the variogram cloud will be replaced by the traditional semi-variogram point plot.



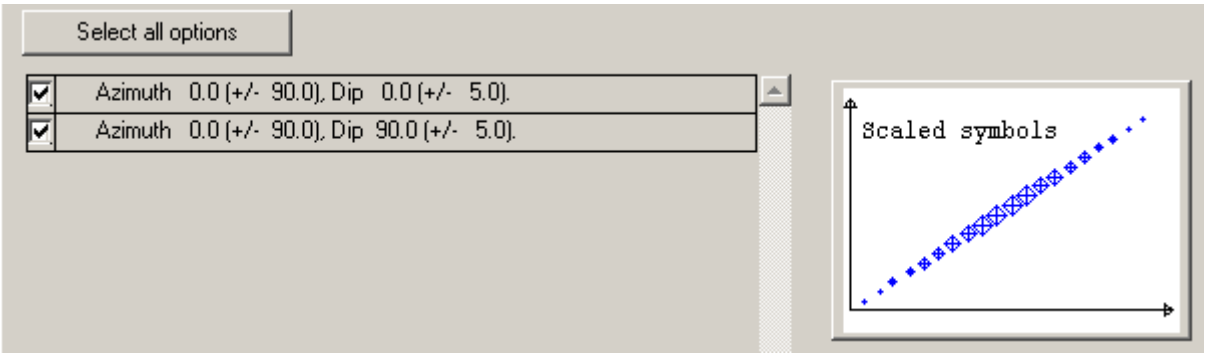
For display and modelling of the semi-variograms, you will find it easier if you choose the Full screen display of semi-variograms option. If you do, this option will switch to Display the semi-variogram cloud maps so that you can return to the 4-part screen if you wish.

When you select Plot experimental semi-variograms from the Semi-variogram Options Return menu, a new dialog will be displayed.

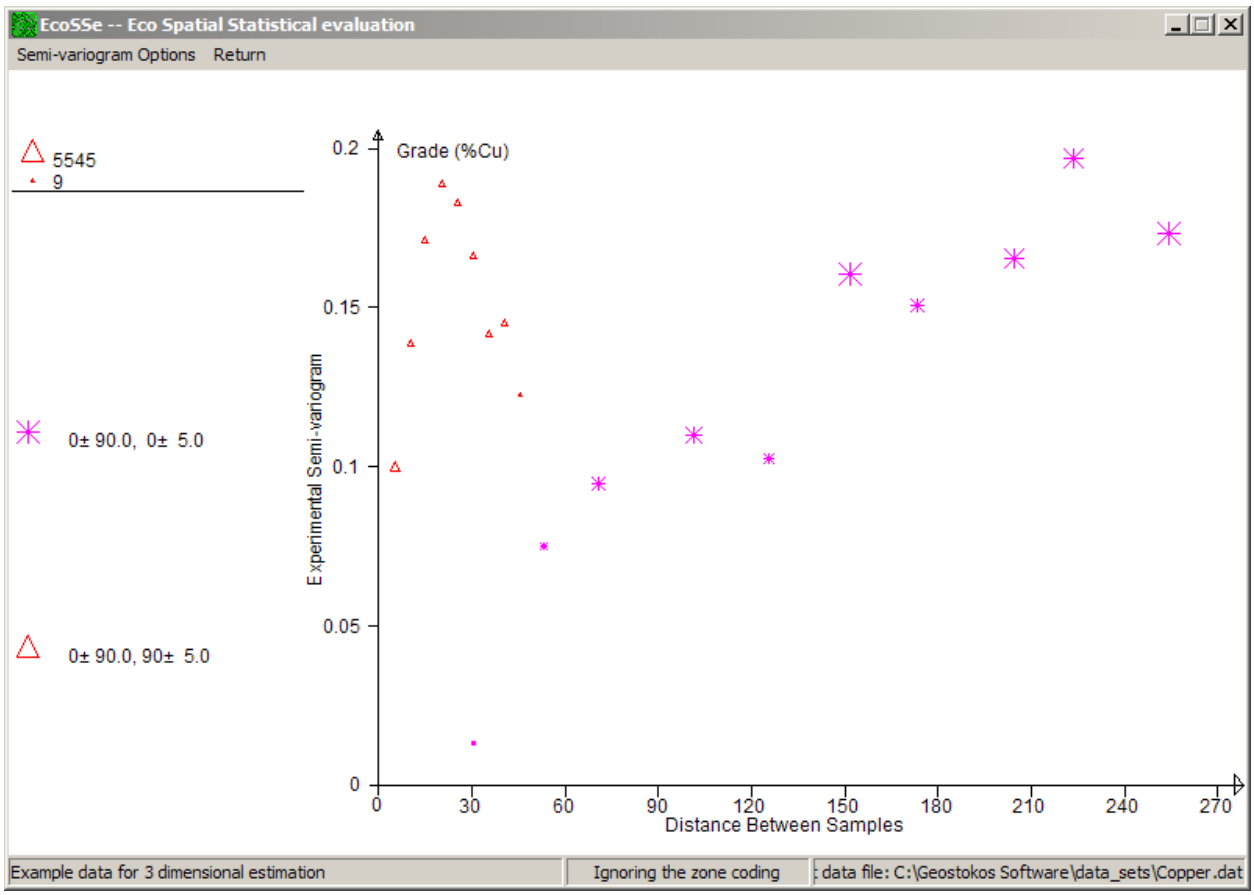


This dialog lists all of the semi-variograms which have been calculated (and for which the routine found pairs of samples). You may plot any combination of these calculated semi-variograms on the screen at once. Next to each calculated semi-variogram is a check box. At the top of the dialog, you will see the message "You may select one or more of these at one time". Check the boxes for the ones you want to plot.

At the right hand side of the dialog, you will see two small graphs. These indicate the type of graph you can plot. There are two ways in which the graph can be plotted. Firstly as a symbol for each calculated point on the graph. The symbol size is proportional to the number of pairs of samples which have been averaged to obtain that point.

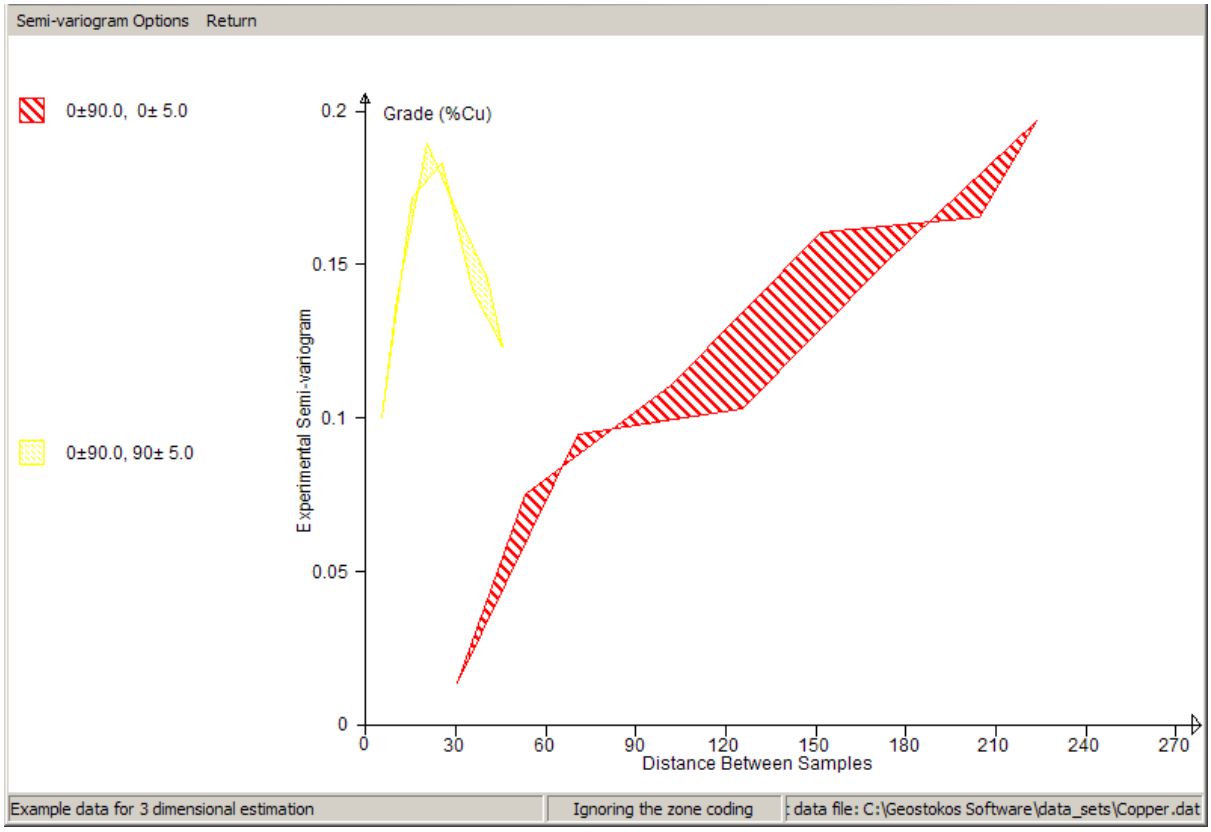


Choosing the two directions and clicking on the upper of the plotting options – the scaled symbols plot – results in the following graph.



The symbols in this graph are scaled to illustrate the number of pairs of samples which were found in that interval. The largest symbol in this graph has 5545pairs grouped together into one interval. The smallest point has only 9 pairs in its calculation and is, one would think, somewhat less reliable.

If we had chosen the shaded graph option, we would get the following graph.



This display is produced as follows:

- For each calculated semi-variogram, we join the first point to the third, the third to the fifth and so on.
- The second point is joined to the fourth, the fourth to the sixth and so on.
- The area between these two lines is shaded.

This display may be easier to interpret, especially for beginners. It does not, however, give any information about the number of pairs of samples in each interval. One single pair of samples giving an erratic high point can seriously distort the shaded graph. The best combination, perhaps, is to use the shaded graph to get an idea of shape, possible anisotropy etc, then use the symbol graphs to do the actual modelling.

When you have looked at the graphs to your heart's content, you can choose to fit a model to the calculated (or *experimental*) semi-variogram.

Note in particular the third option on the semi-variogram menu, which enables you to store the *experimental* semi-variograms 🕮 not the model 🕮 on a text file for input to (say) a report quality graphics package. An **PG2000** option (read in experimental semi-variograms) exists, which allows you to read this file back in and continue with the modelling stage.

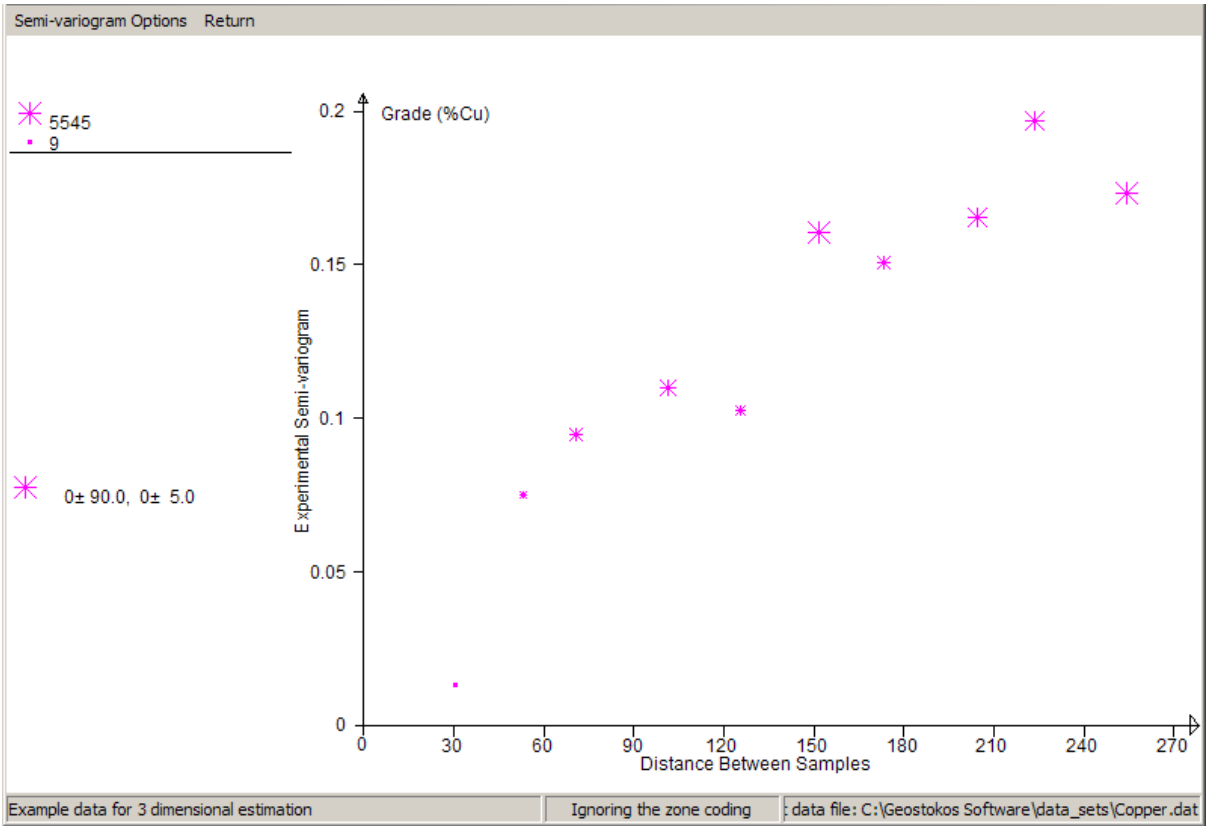
We see that the two semi-variograms look very different in shape. They appear to have similar nugget effects and final sills, but widely different ranges of influence. We can, therefore, assume that we have obvious *geometric anisotropy*.

In point of fact, we constructed the semi-variograms in these particular directions because we had a good prior idea that the dump is stratified horizontally. If you have no preconceptions

on which directions anisotropy might take (if any), then the default directions are as good a place to start as any.

Fitting a semi-variogram model

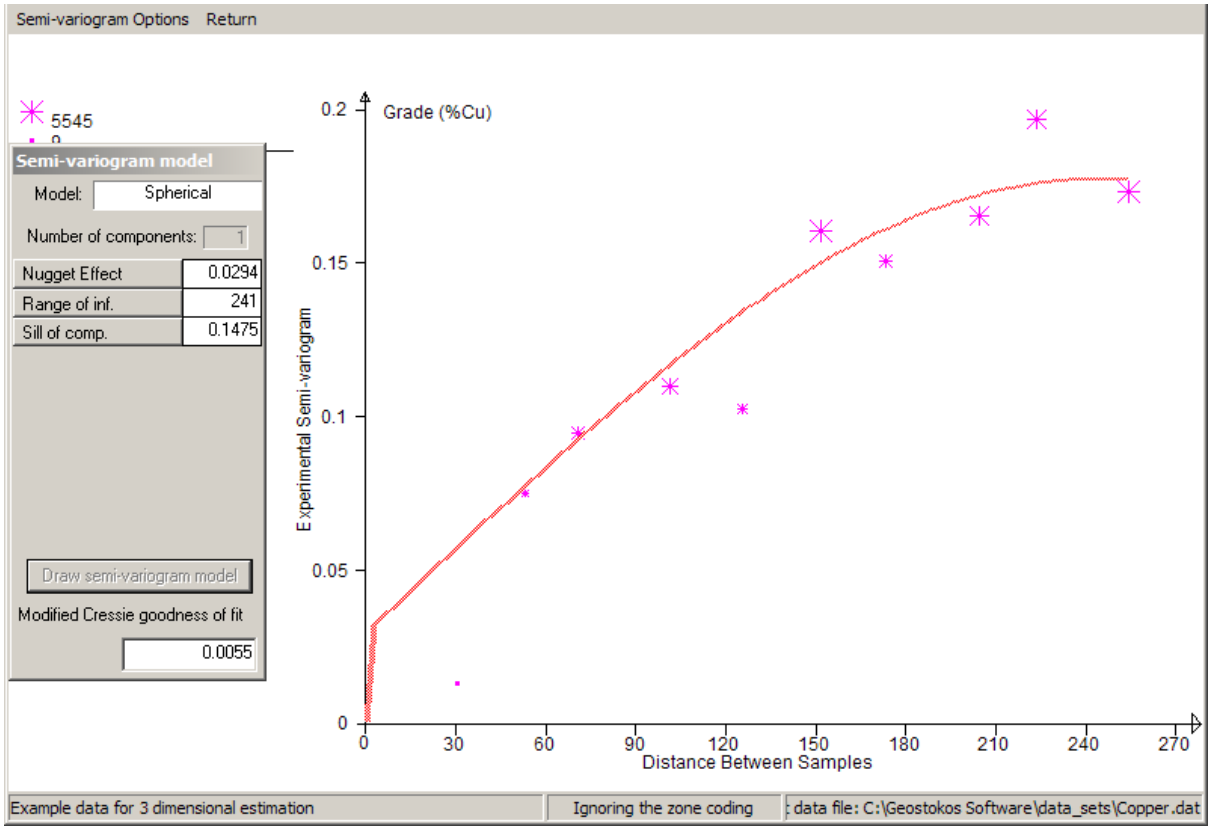
It would now be appropriate to fit a model to the experimental graph, so that we can proceed to estimating unsampled locations. We will need to fit semi-variograms in our major directions of anisotropy. Consider firstly the 'horizontal' semi-variogram:



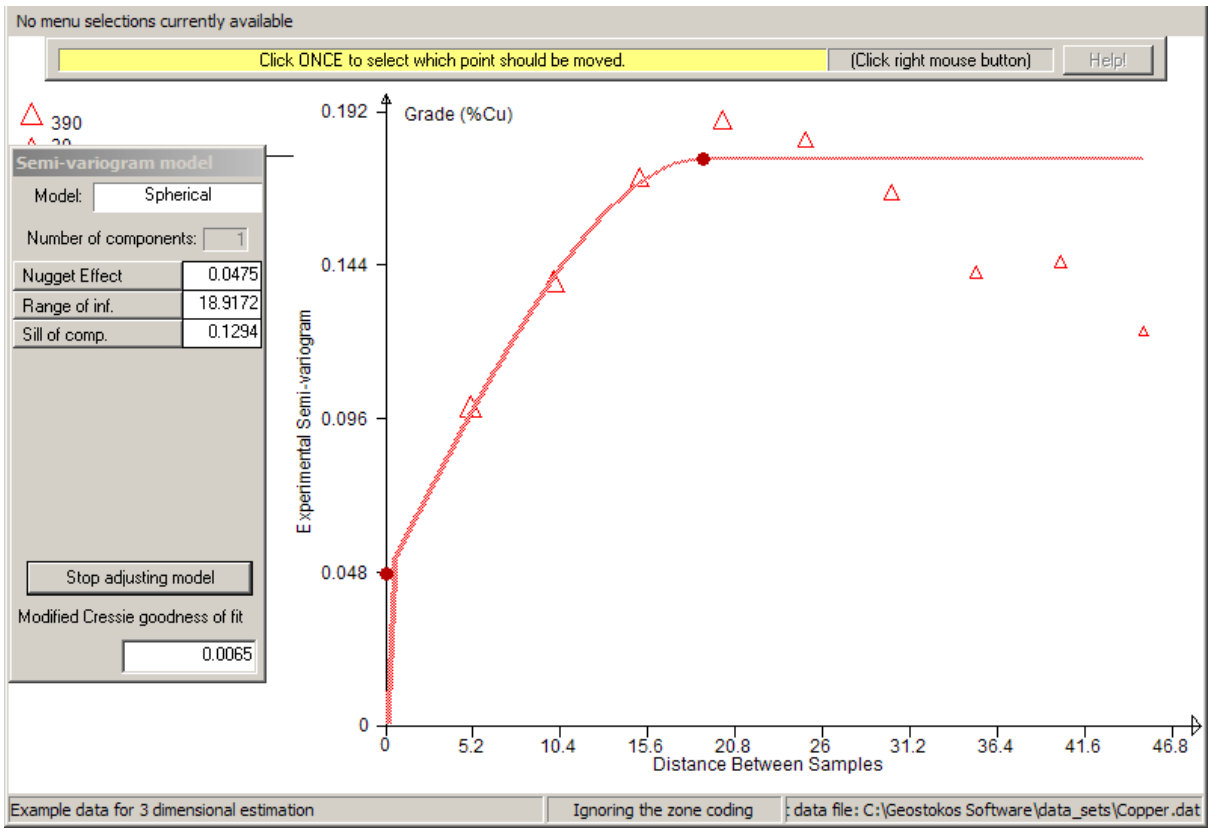
This looks pretty Spherical, although the nugget effect is difficult to pin down, since we have few samples at the shorter distances on the graph.

On the other hand, we have enormous numbers of pairs of samples in the longer distance points, enabling us to get a pretty good handle on the range of influence and sill for the semi-variogram model.

Using the 'point and click' modelling facilities, described earlier in Tutorial Three (Part 1) and Tutorial Four, we came up with the following model for the 'all horizontal' semi-variogram:



Turning to the vertical model, we bear in mind the sill and nugget effect found in this direction and use that to guide our choices for the shorter range direction:

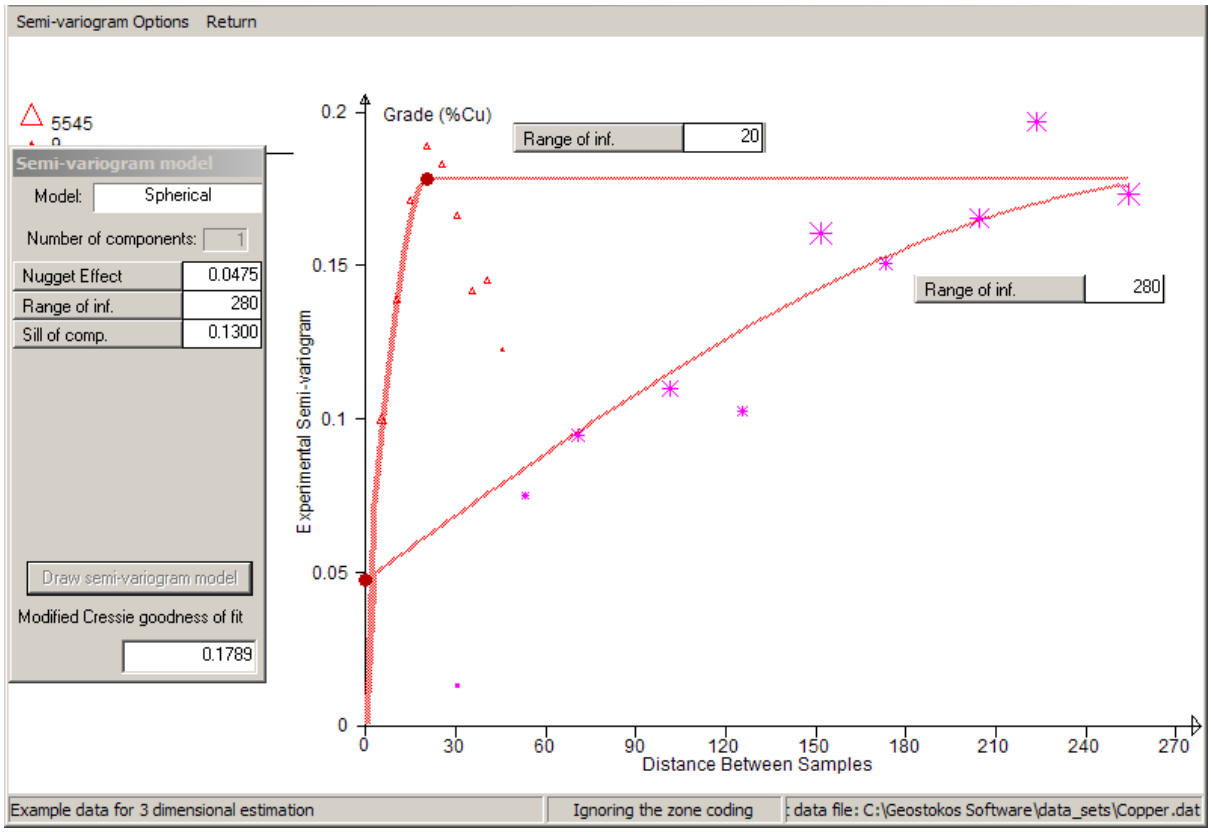


As a measure of the goodness of fit of the model to the calculated points, the Cressie goodness of fit statistic is quoted at the bottom of the dialog. We use a modified version of this statistic which is standardised by the total number of pairs included in the graph. This gives a figure which is not influenced by the number of pairs and can, perhaps, be more objectively interpreted. The statistic is calculated as follows:

$$\frac{\sum_h N(h)\left(\frac{\gamma^*(h)-\gamma(h)}{\gamma(h)}\right)^2}{\sum_h N(h)}$$

It is a good idea to get this as low as possible, but not at the expense of a good visual fit.

Given that the above model is an attractive fit to the points in the vertical direction, we return to the horizontal direction and adjust the model to find the longer range of influence.



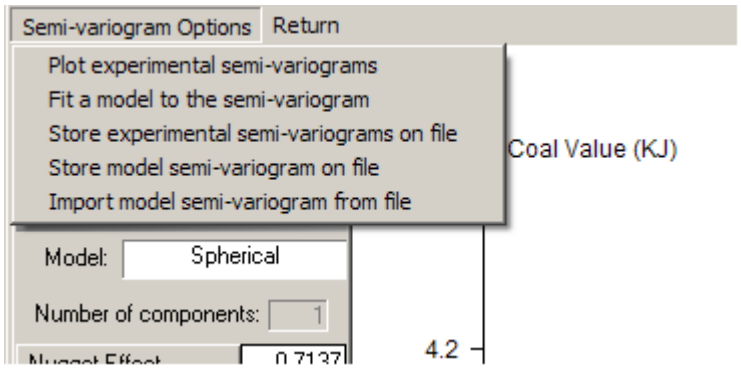
By the way, you cannot get this plot with the software. This was produced by capturing the screens, pasting into MSPaint™ and editing for clarity!

In the **PG2000** software, only geometric anisotropy is allowed. That is, you have a 'standard' semi-variogram in one direction. The models for the two other directions orthogonal to this must have the same shape of semi-variogram model but can differ in range of influence. When you begin kriging, you will specify the 'anisotropy factors' for the change in range of influence with direction.

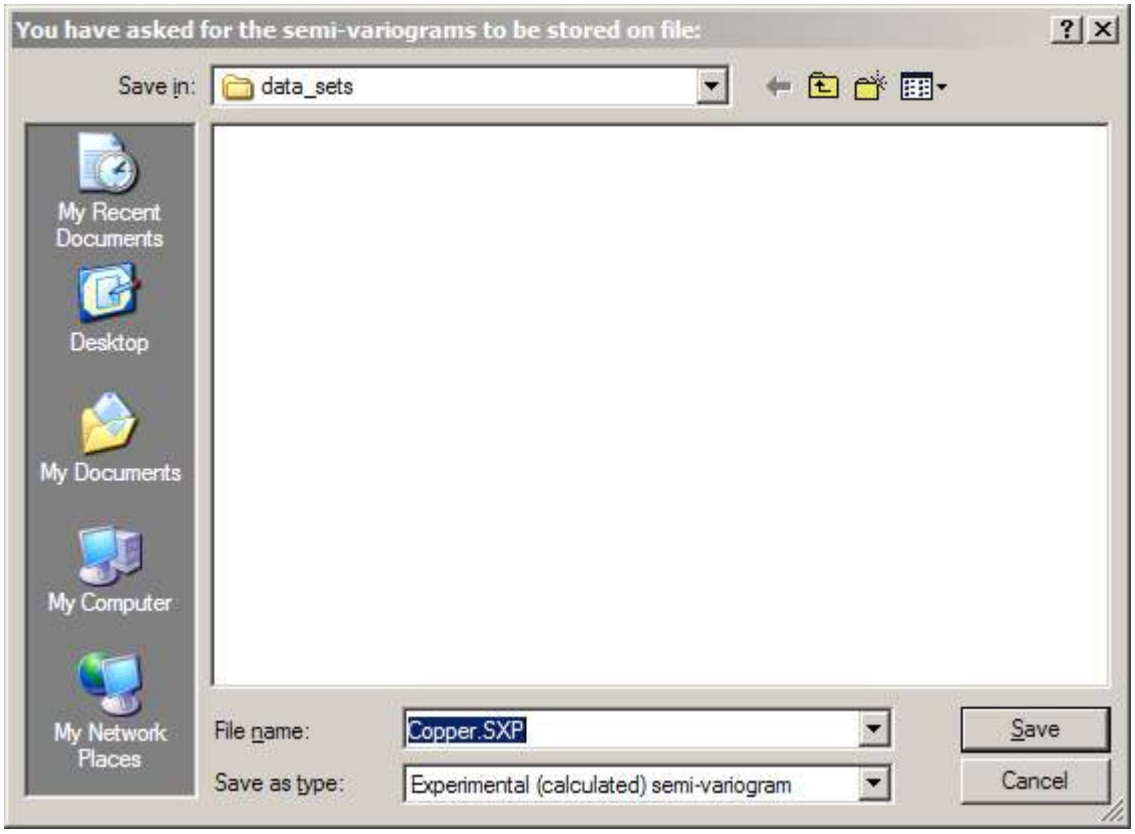
The important thing to note here is that the horizontal direction has to have a range 14 times that of the vertical direction, if we are to keep the nugget effect and sills the same in all directions.

If you store a model on file, at this stage it will not contain anisotropy information.

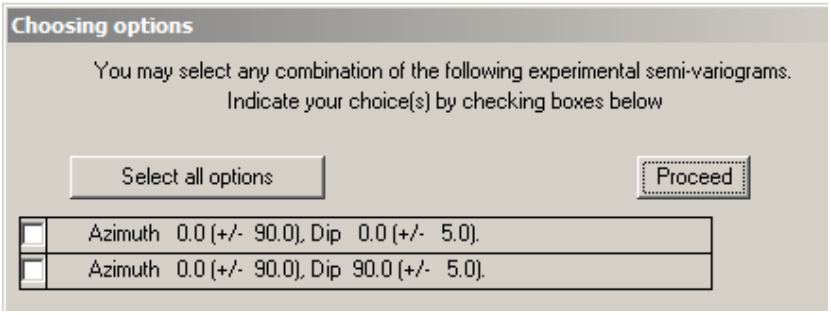
Remember to update your model once you enter these – possibly after cross validation.



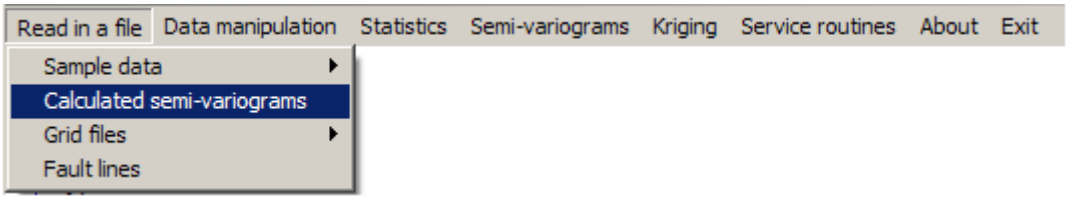
Before quitting this routine, we might want to store the calculated semi-variograms in case we want to look at them again later. You can store any combination of the calculated semi-variograms on a file. **PG2000** will prompt you for the name of the file. The default name is the original data file name with an extension of `.SXP`. You can change the default extension simply by typing in a new one. Alternatively you can change the whole name to something entirely different.



Choosing which semi-variograms to store is identical to choosing which ones to plot:

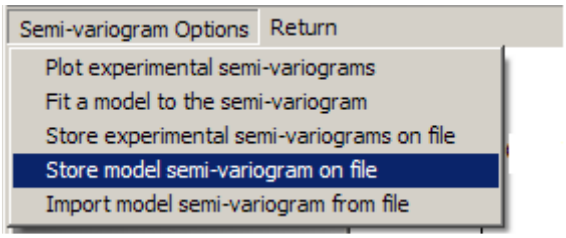


The stored semi-variograms can be read back in at any time using the option on the main menu:



Storing a semi-variogram model

You may also want to store your model on a file for future use in modelling or in the kriging routines:



You will be prompted for the name of the output file on which the model will be stored. This has a default name the same as your original data file and an extension of **`.par`**.

The model file is a flat text file listing all the possible semi-variogram parameters and can be accessed by Wordpad, Notepad or some such for reporting or editing.

Having done all we need to do with the semi-variogram calculation and modelling:

which returns you to the main menu bar.

This Tutorial is continued in Software Tutorial Session – 3D Ordinary Kriging.