Software Tutorial Session -- Ordinary Kriging

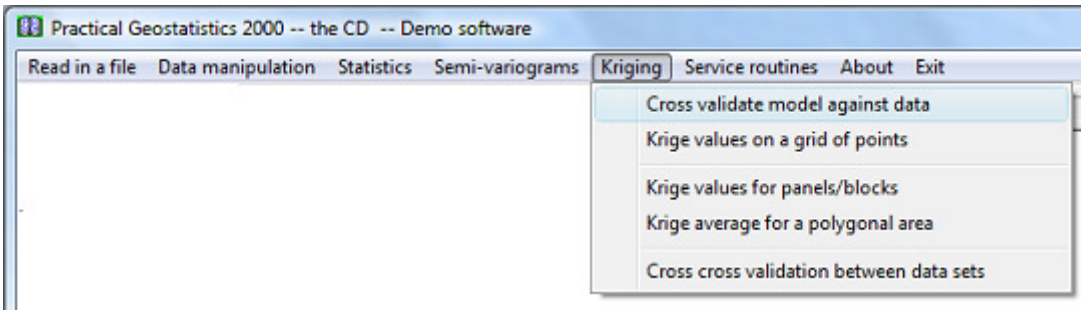
The example session with **PG2000** which is described in this and Part 1 is intended as an example run to familiarise the user with the geostatistical facilities within the package. This documented example illustrates one possible set of analyses which may be carried out. It takes you through the following sequence of analyses:

- Cross-validation of the semi-variogram model
- Kriging a grid of point values for mapping

There are many other facilities within the package, which are given as alternative options on the menus. This part of the documentation assumes that you have worked through Tutorial Three (Part 1) where we constructed and modelled semi-variograms on the calorific values listed in the **`coalmine`** data set.

Cross validation of the semi-variogram model

For this Tutorial, we have decided to continue with some geostatistical estimation using the model which we have fitted.



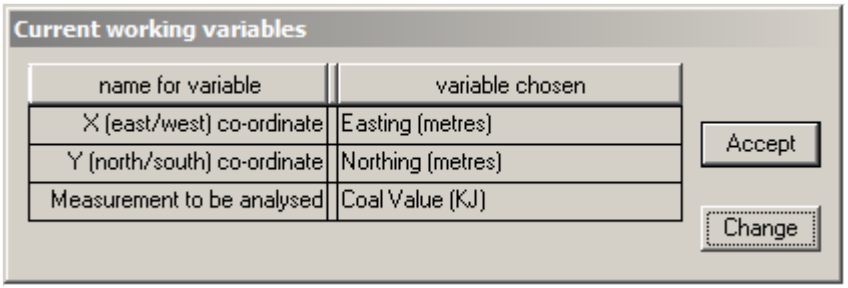
As you can see, I have chosen the option to 

. There is a bit of confusion in the literature in the naming of the process. Some authors call this *jack-knifing*. This nomenclature is misleading, since the procedure bears little relationship to what statisticians would expect by jack-knifing. Other authors use the two words hyphenated or as a single word. We have chosen the above form to emphasis the meaning of the procedure. We attempt to *validate* our semi-variogram by dropping out each sample value and (cross) estimating the value at that location from the neighbouring samples. We then compare the estimated value with the actual value, and the difference between them with the supposed geostatistical error.

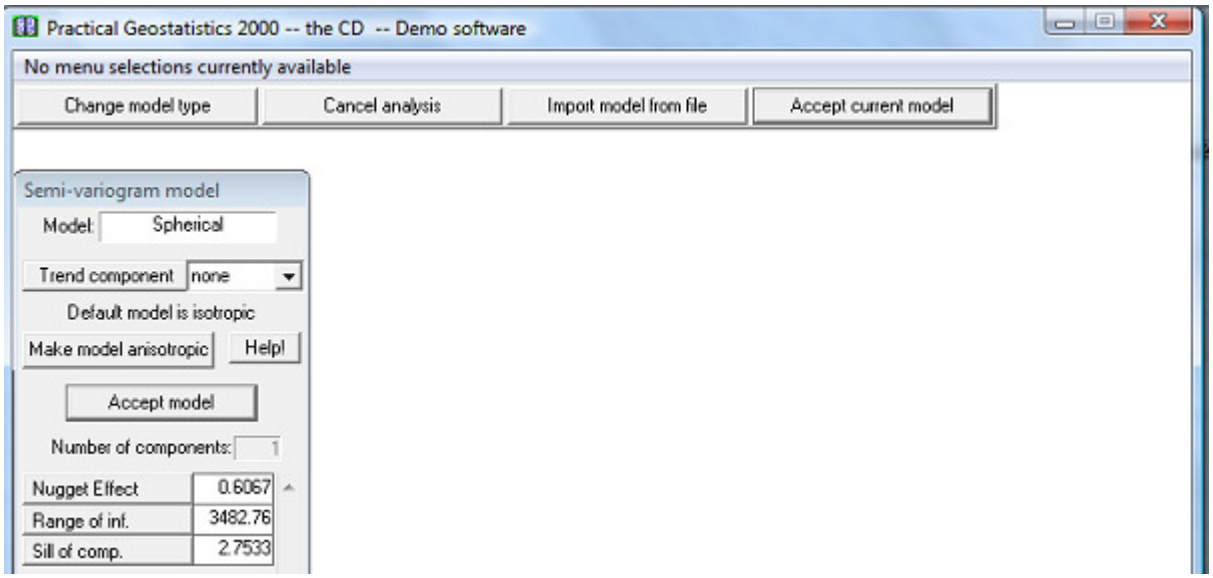
There are different ways of comparing the actual error with the Kriging error. We have chosen a simple method by calculating the ratio between the two – i.e. actual error divided by Kriging standard deviation. If certain basic assumptions are satisfied, *and we have chosen the correct semi-variogram model*, these (error) statistics should average zero and have a standard deviation of one. We use the mnemonic `XVAL` for cross validation. Running this option will also give you an idea of how long Kriging is likely to take on your computer.

PG2000 remembers which variables you were studying for the duration of a single run. If you model your semi-variogram and do not continue immediately to kriging, you should store

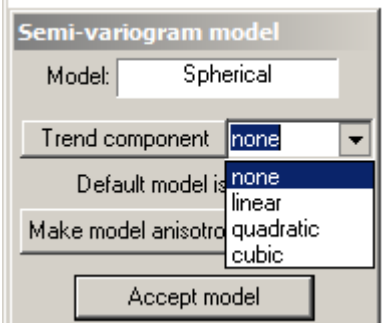
your semi-variogram model on file for later access. If you start a new run of **PG2000** you can come to this routine directly. There is no need to go through the whole procedure to get this far! In that case you will have to select the variables in the same way that you did for the semi-variogram analysis.



To carry out cross validation – which includes kriging the estimates – you need a semi-variogram model. In this Tutorial we have already fitted a model, so **PG2000** will remember this and offer it to you as the base model. However, for kriging there are slight changes and additions in the semi-variogram model definition:

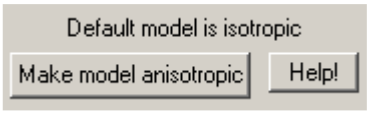


If you have a significant trend that should be defined so that the kriging can allow for it. In the presence of trend, we would use Universal Kriging (see Tutorial Five). The level of trend is specified in a dropdown menu list:

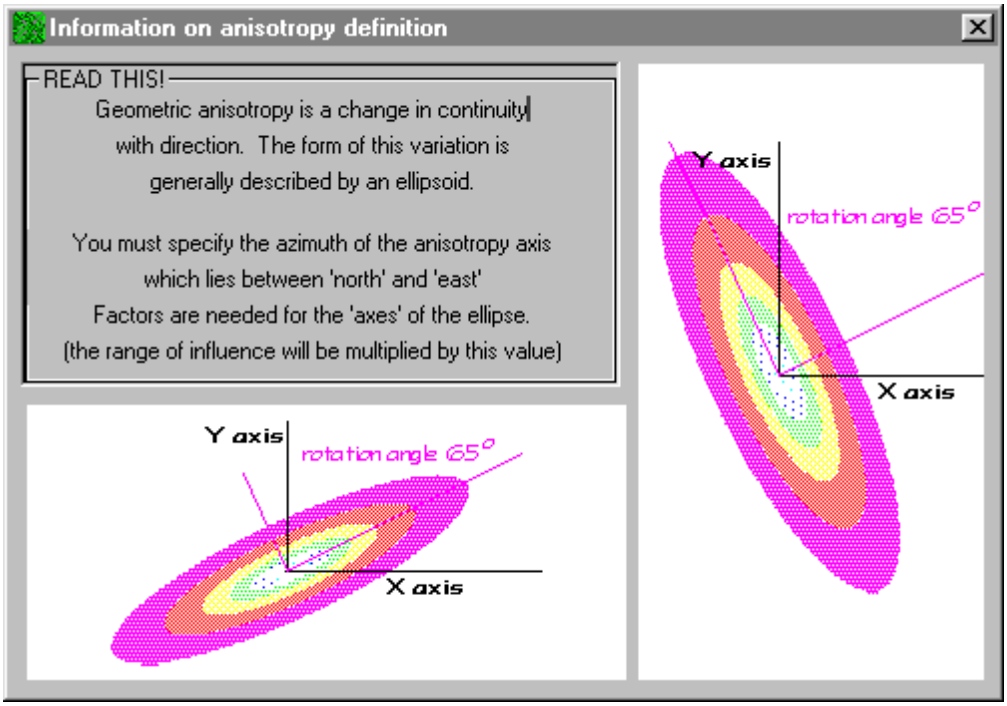


Always bear in mind that when there is a trend in the original sample data, all parameters refers to the semi-variogram of the residuals.

Another button in the dialog allow you to specify simple geometric anisotropy.



Click on the Help! button and you will get the following information:



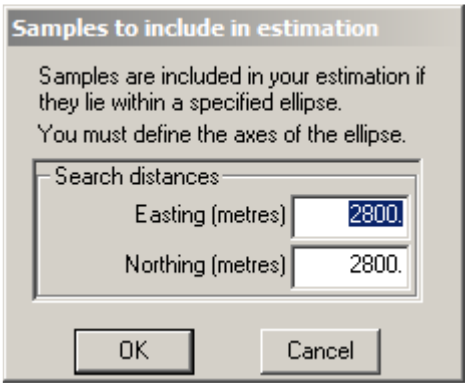
Close the box ()when you have seen enough.

Click on Accept model or Accept current model to get the routine to accept the semi-variogram model for kriging.

Until you do this, you still have the full option list:

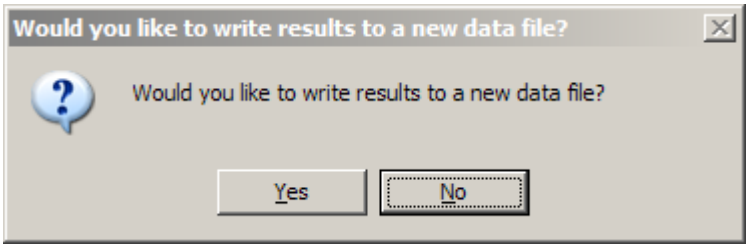
Now we have variables to study and a semi-variogram model which tells the software how the values are related to one another. For cross validation, we will take each sample in turn and remove it from the data set. The neighbouring samples will be used to produce an estimate at this location. We can then compare estimated value with the actual value found in the sample at that location.

Before we can go any further, we need to define the “neighbourhood”. That is, how far do we want the software to search for samples to be included in the estimation process. Since we have an “isotropic” semi-variogram model, it seems sensible to select an isotropic search radius. Since we have a Spherical model, **PG2000** will suggest that we use the range of influence of the model as a default search radius.



The default search radius is always the range of influence of the first component fitted – providing the model is Spherical. For Exponential and Gaussian models the search radius is adjusted to a realistic distance. For models without a sill, **PG2000** cannot guess what an appropriate search radius would be and uses a similar default to that in the inverse distance interpolation routines.

The default search radius, given our semi-variogram model, is 2,800 metres. The software will then enquire as whether you wish the results stored on file for later analysis:

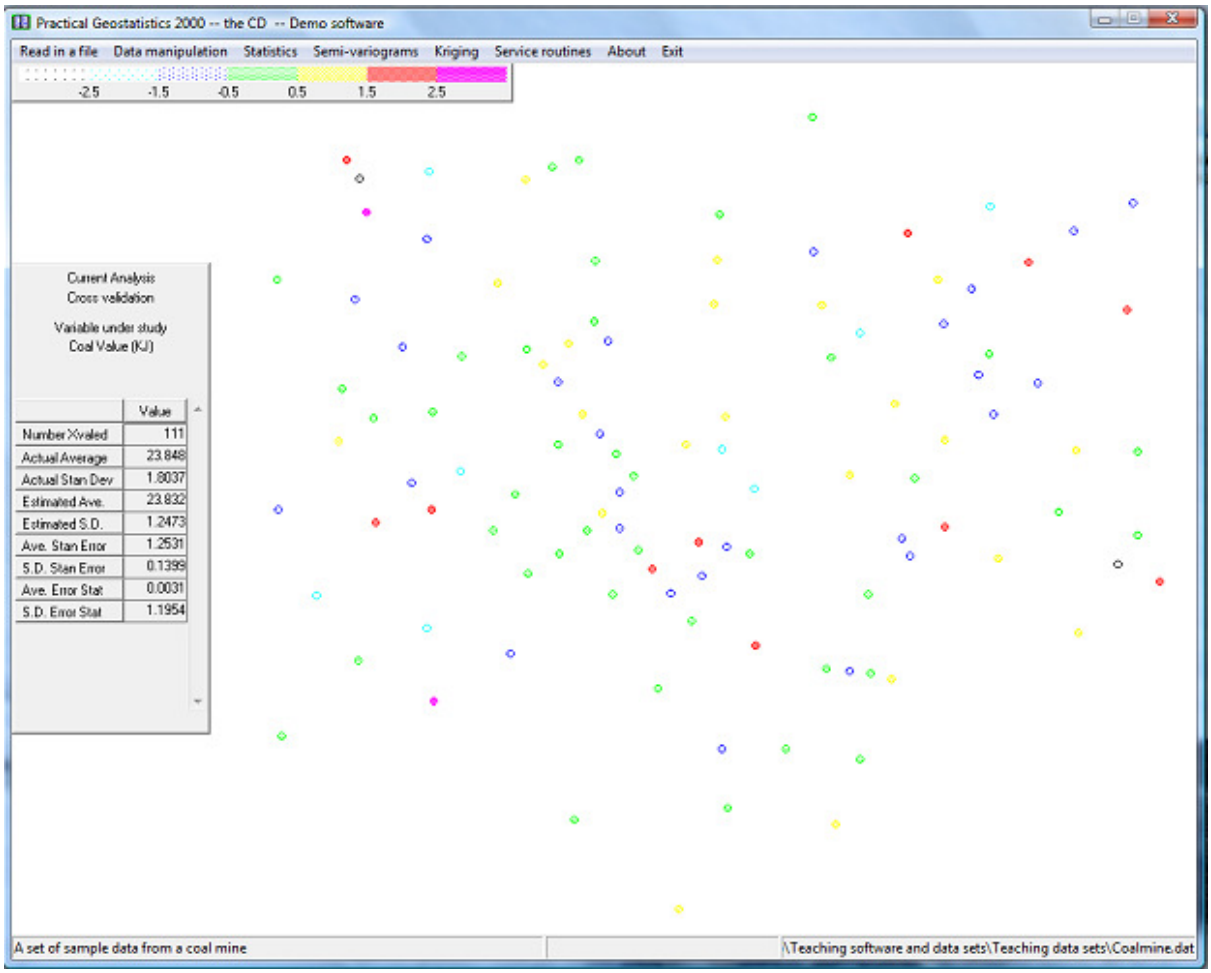


If you answer "yes" to this question, you will be prompted for an output file name. The default name is that of the original data file with the extension `.XVL`. You can change the data file name, the extension or both if you so wish. This file will be written in the correct format to be read back in as a data file. The cross validation outputs a table of values on the `ghost.lis` file as the estimation proceeds. The final column "error statistic" is the ratio of the actual error to the Kriging standard error.

As the cross validation is carried out, a post plot will be drawn of the "error statistics". The contour levels for this graph have been chosen so that a value in the highest (+2.5) and the lowest (−2.5) contour bands should occur one time in one hundred. This plot is an excellent device for visually spotting outliers in the sample data. These need not be outliers in the usual statistical sense. That is, they may be quite acceptable *values* as such. What the cross validation will show is whether they are acceptable values *in the context of the neighbouring samples*.

The left hand box on the screen summarises the various calculated values. A direct comparison can be made between the average actual value (23.85 KJ) and the average Kriged value (23.83 KJ). The standard deviation of actual calorific values is 1.80 and that of the estimated values is 1.25 KJ. This is illustrative of the smoothing which takes place when ordinary kriging is used as an estimation technique. A weighted average of a set of samples will have a smaller standard deviation than values when taken singly.

The average (typical?) Kriging standard error is 1.25 KJ, although the individual standard errors vary around this value. Finally, looking at the all-important "error statistics", we find an average of 0.0031 and a standard deviation of 1.1954. Ideally, we are looking for zero and one. It is your decision as to whether 1.19 is close enough to 1.0 to be accepted (sic). Please refer to full documentation for further discussion of this point.



Note from the post plot above that the north-west corner of the study area contains three samples which differ wildly from one another.

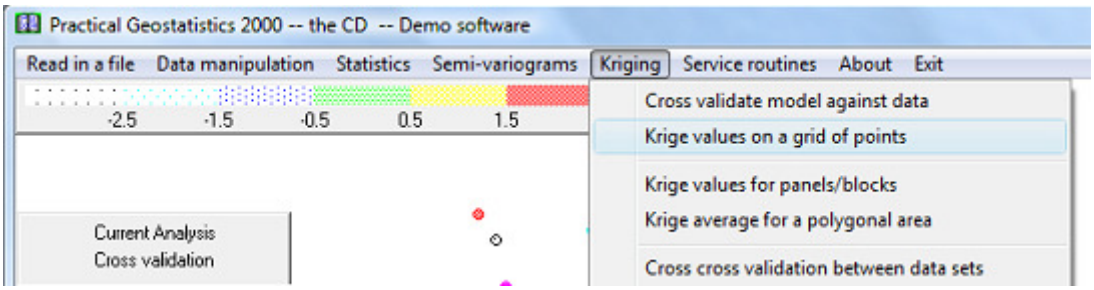
- A point which is black in colour and 'empty' denotes a sample with an error statistics below -2.5. That is, a low value sample which has been massively over-estimated by its neighbours.
- The purple point has an error statistic above +2.5 and is a high value which has been massively under-estimated.
- The red point has an error statistic between 1.5 and 2.5 and is a high-ish value which has been under-estimated.

The contours are chosen so that black and purple points should appear about 1% of the time. With 116 samples, one black point and one purple are not a concern. What *IS* of concern is the fact that they are right next to one another. In real life, we would want to go back and query these samples and why we have such large variation in such a short space.

For the purposes of this Tutorial we will accept the above model. If you do not wish to accept, you will need to change the semi-variogram model. To reduce the standard deviation of the errors you will need to raise the Sill or nugget effect, or reduce the range of influence.

If you choose to store your cross validation results on a file, this file can be read back into **PG2000**. You can produce scattergrams of, for example, estimated values versus actual values to see how well the kriging is performing. You can also do a probability plot of the 'error statistics' to see if Normal confidence intervals would be appropriate.

Interpolating a map with kriging

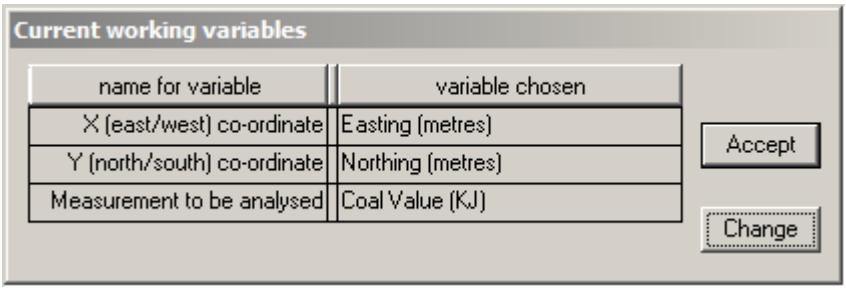


Interpolating a grid of points with kriging will produce an estimated (or "predicted") map of the values over the study area. This map reflects the actual values measured at the actual sample locations and uses a weighted average estimator for grid points which have not been sampled. Weights are determined by a set of equations which combine:

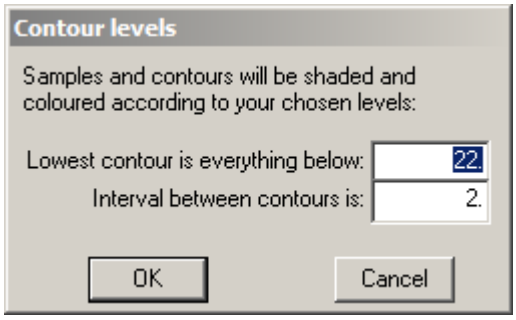
- the spatial continuity as modelled by the semi-variogram
- any anisotropy identified and modelled in the semi-variogram
- any trend component identified and defined by the user
- the spatial layout of the samples relative to the points being estimated
- the spatial layout amongst the samples themselves (clusters, irregularities etc)

The chosen weights will minimise the "estimation variance", which may be interpreted as a measure of the estimation error.

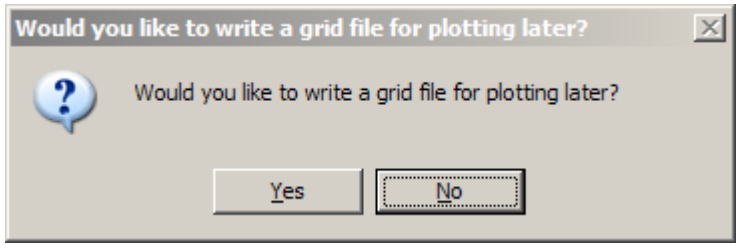
PG2000 will remember everything which has been defined during this run. We have already defined which variables we have been analysing:



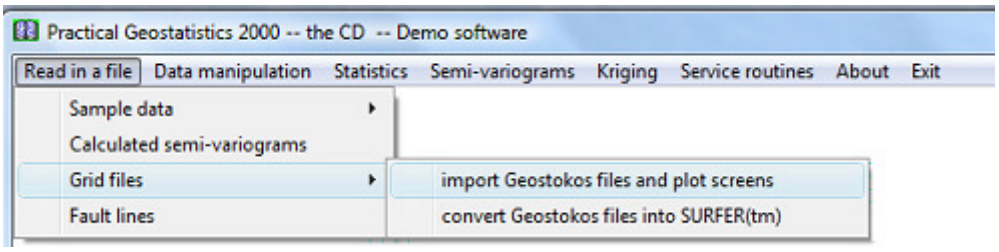
Click on  to proceed. The routine also needs contour levels:



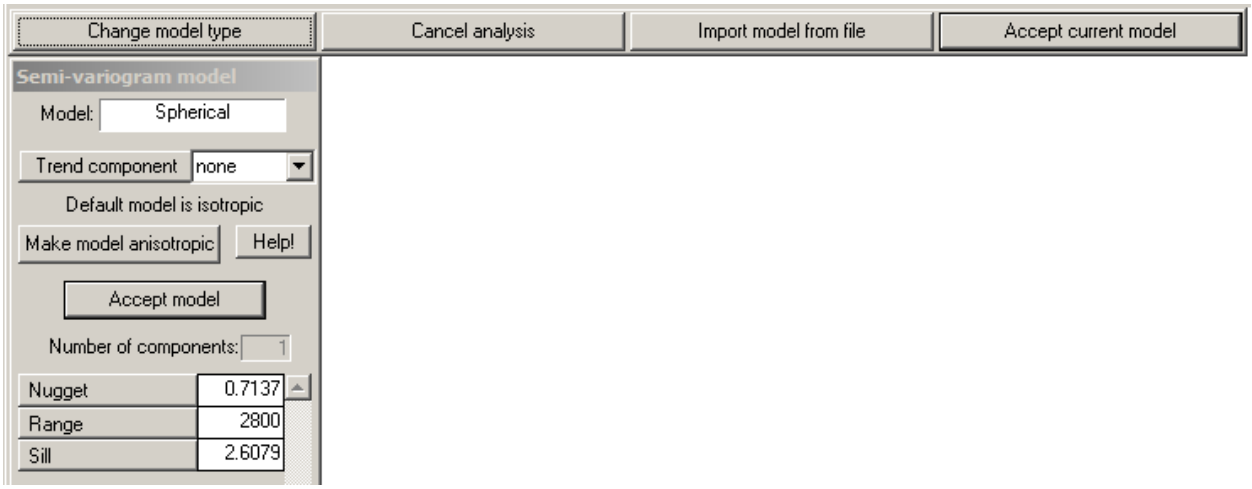
and to know whether you want the results stored on a "grid" file:



PG2000 will suggest contour levels based on the variability of the sample values. You can change these if you so desire. Alternatively you can run with the default contours and draw prettier maps by reading the grid files back in. The default name for a grid file is the original data file name with the extension `.GEA`. Please note that "grid" files are not in the same format as "data" files. If you want to read them back in, you must use the option:



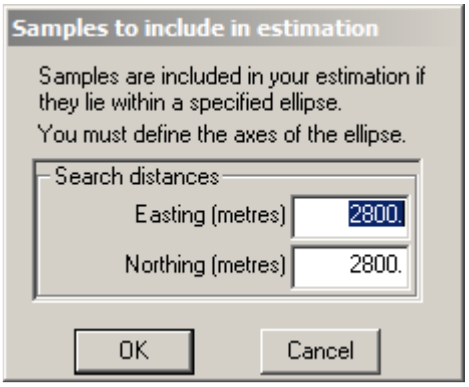
You need to confirm semi-variogram model, search parameters and the area which is to be studied.



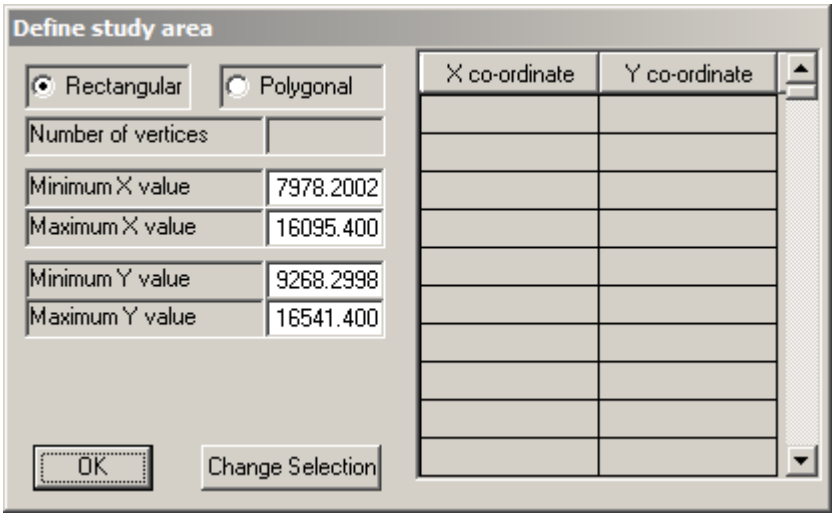
The semi-variogram we defined previously in the cross validation section. If you come directly to the kriging routines without passing through any others, you will need to respecify (or import) your semi-variogram model including the local trend component.

The neighbouring samples will be used to produce an estimate at each unsampled grid point. Before we can go any further, we need to define the "neighbourhood". That is, how far do we want the software to search for samples to be included in the estimation process.

PG2000 cannot guess what an appropriate search radius would be. As a simple default, with models such as the Spherical, a default based on the semi-variogram parameters is offered. When the value at a specified grid point is being estimated, all samples within this circle of the point will be used in the Kriging process. If there are too many samples within this circle, those closest to the "unsampled" location will be selected. The default search radius, given our semi-variogram model, is 2800 metres.

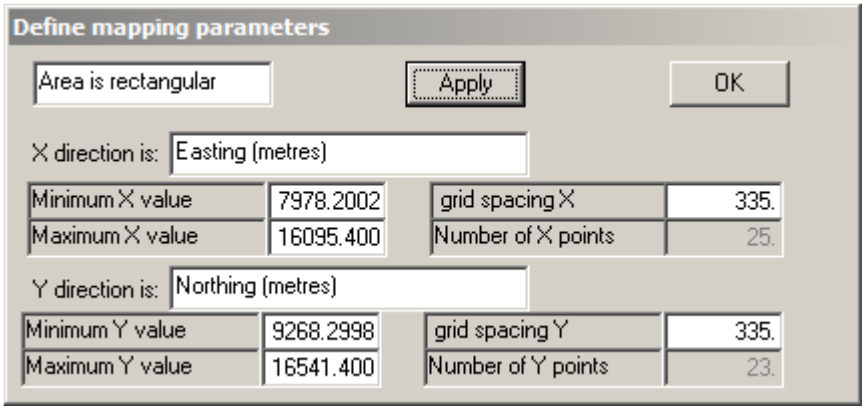


We need to define the area which is to be mapped. Firstly we define the limits of the area and, if appropriate, a polygonal boundary within which values are to be mapped. The default area offered will be the extent of the sample data:

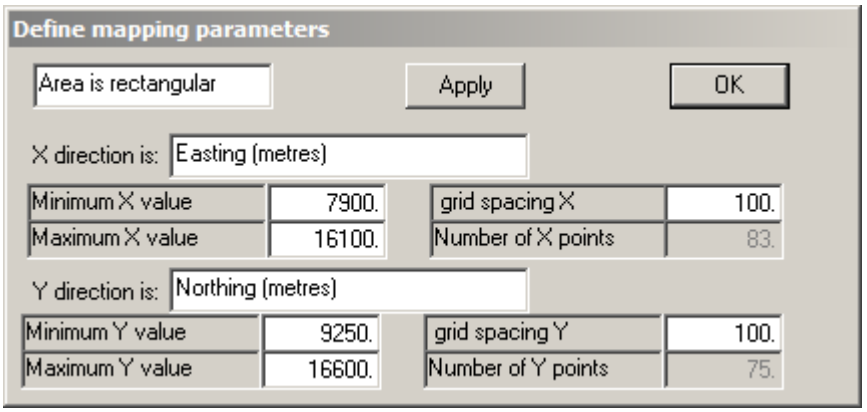


You can manually change the rectangular limits of the area or you can read in a file containing the polygonal boundary. The default name for a boundary is always the original data file name with the extension **`.bln`**.

Once you have chosen the area to be studied, you must define the grid spacing to be used. Points will be calculated at each grid node and represented on the screen as a shaded rectangle of the appropriate size. The program offers a default which will give you about 25 grid points in the 'X' direction:



We can alter the grid spacing or the limits of the map by changing the number in the relevant box.

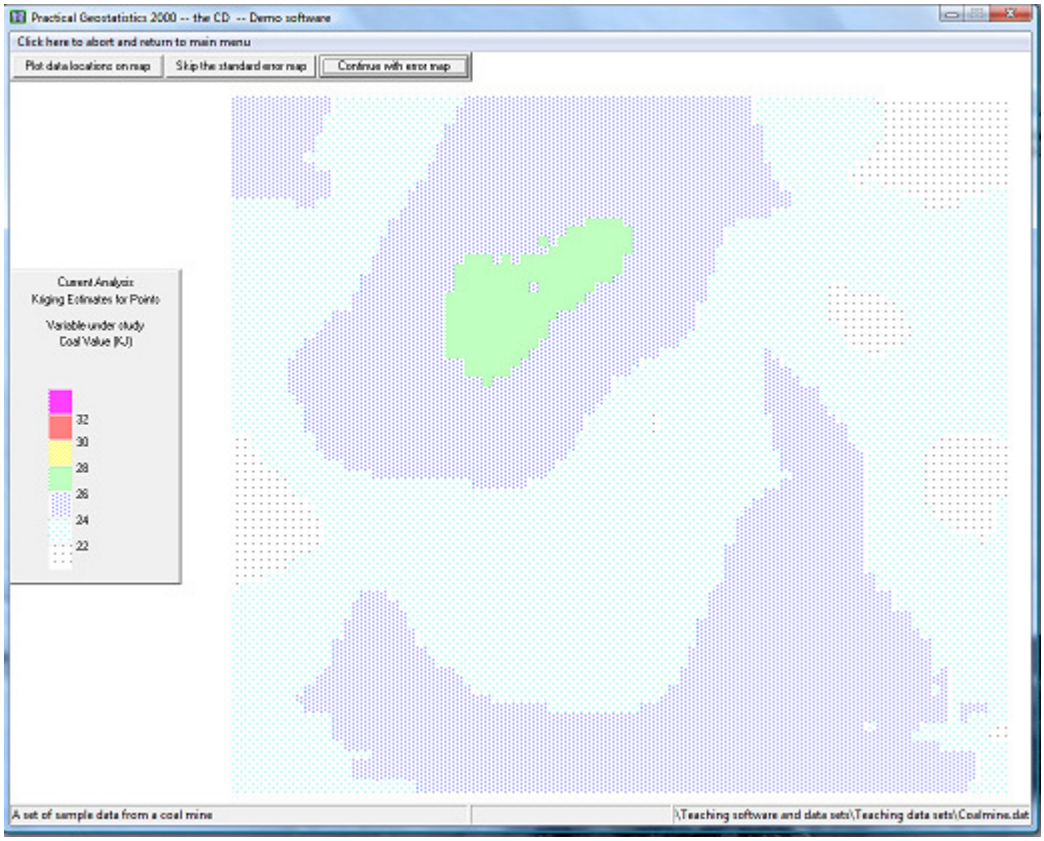


If you make a change and want to check how many grid points you have before proceeding, click on 

and the rest of the parameters will be updated. You may also change minimum and maximum X and Y values at this stage. Once you click on 

the map parameters will be defined.

Interpolating a grid of points produces a sketch map on the screen. The shading information for the contour levels will appear in the left hand box and the map itself in the right. A shaded square will be displayed on the map to show you which point is being estimated in addition to the information in the prompt box. You may copy the screen to your printer at any stage during the estimation process.



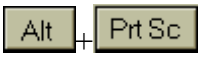
When the Kriging has been completed, you have the following options:

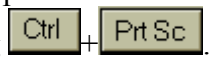
To display the data locations, click on 

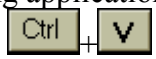
.

If you click 

the "error" map will be displayed showing the standard errors associated with the estimated grid points.

You can copy the plots with 

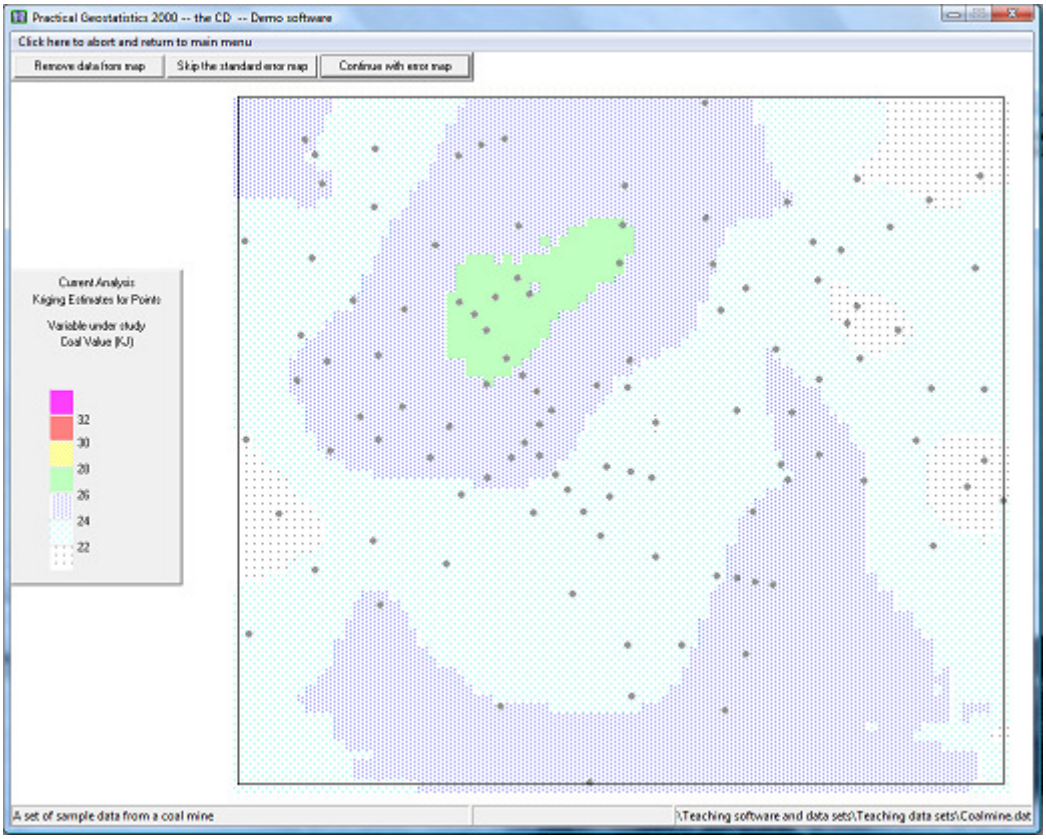
and paste them into another application. Some systems (notably Windows NT) require pressing 

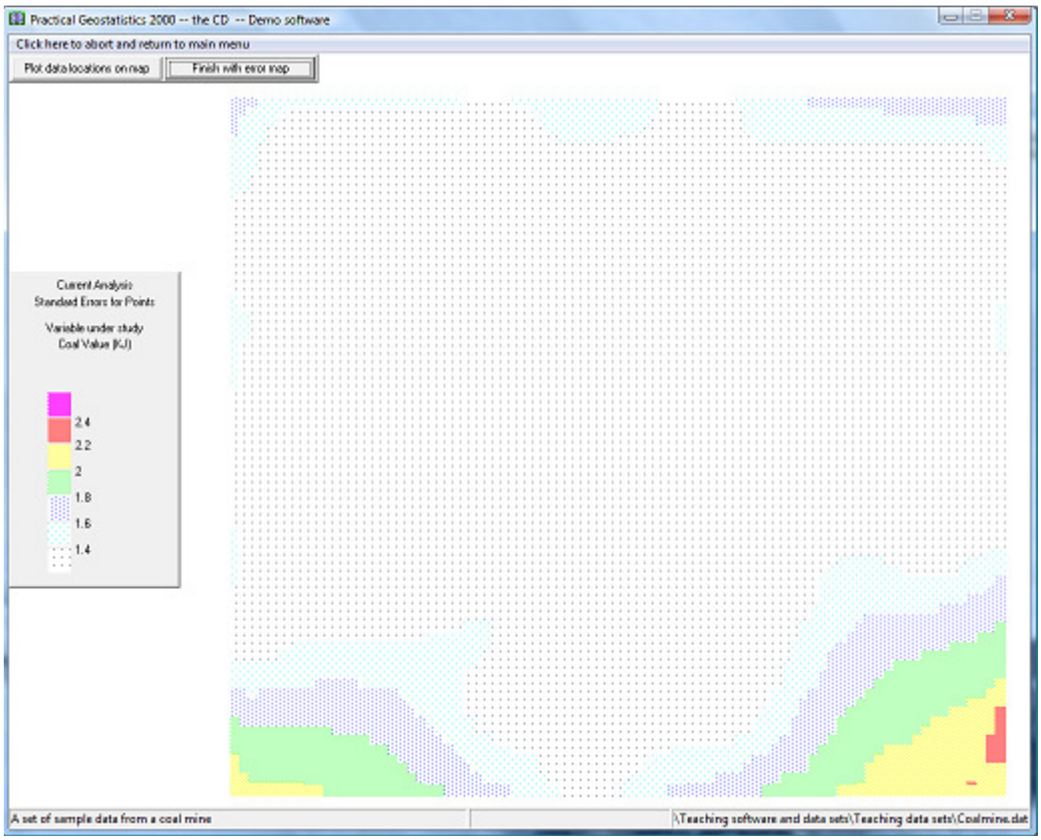
. This will place a copy of the Window in the clipboard. You can *import* the picture into a Word processing application such as Microsoft Word, a spreadsheet application like Lotus or Excel, or paste 

into many applications, such as MSPaint.

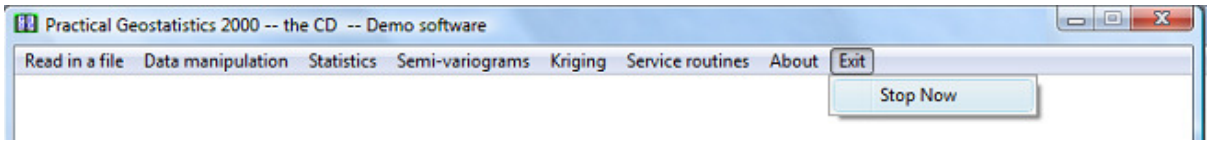
See overleaf for both of the above mentioned plots.

If you have elected to write a grid file, these values are also stored on the `.GEA` file so that you can redraw the maps with different contours by reading back the grid file. You can see from the default error map overleaf that this option can be rather useful.

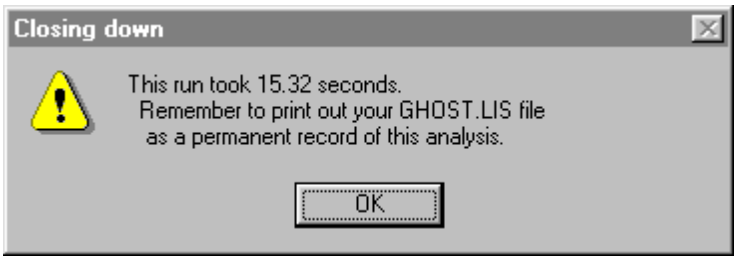

Practical Geostatistics 2000 -- the CD -- Demo software
Click here to abort and return to main menu
Remove data from map
Skip the standard error map
Continue with error map
Current Analysis:
Kriging Estimates for Points
Variable under study
Coal Value (KJ)
32
30
28
26
24
22
A set of sample data from a coal mine
\Teaching software and data sets\Teaching data sets\Coalmine.dat


Practical Geostatistics 2000 -- the CD -- Demo software
Click here to abort and return to main menu
Plot data locations on map
Finish with error map
Current Analysis:
Standard Errors for Points
Variable under study
Coal Value (KJ)
2.4
2.2
2
1.8
1.6
1.4
A set of sample data from a coal mine
\Teaching software and data sets\Teaching data sets\Coalmine.dat

Finishing up



Clicking on this menu item or on  will end your run with the software. You will see the closing down dialog box:



The above Tutorial session should serve only to illustrate a possible use of the various routines from **PG2000**. Try running the program again, choosing your own responses. try looking at reef width instead of grade. This variable has a standard two parameter lognormal distribution. Try reading in one of the other data files which are provided, say, `samples.dat`.

General Notes

There are a few points which you may have noted in following the Tutorial session above. Most of the routines communicate between themselves, without you having to worry about getting the right information from one to the other. For example, after you read in the complete contents of the data file, the routines ask which of the variables you actually want to analysis. This information is then stored internally and may be accessed by any of the other routines. This is a feature of most of **PG2000**, in that it will recall what you chose previously and ask whether this is to change or not. You should bear this in mind if you are analysing more than one data file in a single run. In particular, the boundary used in mapping will be remembered. If you change data file or even which variables you analyse this will not automatically update.

A copy of this run should have been made on a file called `GHOST.LIS` unless you changed the name at the beginning of the run. Send this file to your printer if you want a record of the analysis or look at it with Wordpad or Notepad.

PG2000 — like any computer software — is not completely error-free. Neither is it fool-proof. You can always get out of the software by right clicking on the Taskbar. This will invoke the 'End Task' facility to close the Window without damaging the rest of your system. If you cannot figure out what went wrong, note down as much information as you can about the program you were running, the data you were using and exactly where it broke down. Contact your supplier locally or Geostokos direct for assistance, software@kriging.com. Send us the ghost.lis file and (if you can) the data you were analysing at the time.