Spatial data analysis: geostatistical tools

Geostatistics studies spatial variability of regionalized variables. Regionalized variables are variables that have an attribute value and a location in a two or three dimensional space. Geostatistics are also described sometimes as a set of techniques/tools used to analyze and predict values of a variable distributed in space or in time. With geostatistics you can explore your sample data, construct variogram models and produce interpolated surfaces.

A point interpolation (also known as gridding) performs an interpolation on randomly distributed point values and returns regularly distributed point values. The input for point interpolation is mostly a point map with the domain type value or a point map with domain type class or identifier that is linked to an attribute table, in which the attribute values are stored in a column with domain type value. The output of a point interpolation is a raster map in which each pixel has a value calculated by an interpolation on the input point values. The interpolation methods that will be treated in section 11.3 are: Nearest Point, Moving Average, Trend Surface and Moving Surface.

An alternative to the above mentioned straightforward deterministic methods is kriging. Kriging is a statistical method based on the theory of regionalized variables. Before you can use kriging, you must make a semi-variogram model, which will determine the interpolation function. Furthermore kriging is the only interpolation method available in ILWIS that gives you an interpolated output map and an error map with the standard errors of the estimates. In this manner, kriging allows you to quantify the quality of your predictions. In section 11.4 you will successively become familiar with the kriging operations Ordinary Kriging, Simple Kriging, Indicator Kriging, Universal Kriging, Anisotropic Kriging and CoKriging.

Sometimes the input for interpolation is already a raster map itself. Further interpolation on the raster values, for example to fill up gaps in the raster map or to get a more dense network of sample points can be performed using the Densify operation (to obtain more dense pixels), the Resample operation with bilinear or bicubic interpolation (mainly used to obtain raster maps that all use the same georeference), or the Kriging from Raster operation. These raster interpolation methods will however not be treated in this chapter.

Point statistics may help you to get an impression of the nature of your point data prior to for instance an interpolation and to find necessary input parameters for kriging operations. Various other operations that might be useful, which are treated in

this chapter are Pattern Analysis (11.1), omnidirectional and bidirectional Spatial Correlation (11.2 and 11.4), Cross Variogram (11.4), Variogram Surface (11.4) and AutoCorrelation and SemiVariogram (11.5).

Before you can start with the exercises, you should start up ILWIS and change to the subdirectory C:\ILWIS 3.0 Data\Users Guide\Chapter11, where the data files for this chapter are stored.

☞

- Double-click the ILWIS icon on the desktop to open ILWIS.
- Use the Navigator to go the directory: C:\ILWIS 3.0 Data\Users Guide\ Chapter11.

Visualization of the input data

The data set used in this exercise belongs to synthetic drillhole data from an oil field that is under exploration. The vertical distances towards the top and the bottom of sandstone unit are given (Figure 11.1); the sandstone unit is considered to be a potential reservoir rock for both oil and gas. The exploration geologist is asked to provide his judgement on the spatial variability of the thickness of the reservoir, which would intrinsically be important to determine the success ratio to drill for oil in the area.

The following point maps are available:
- `Sandstone`: point map linked to an attribute table `Sandstone` with the attribute columns `Top`, `Bottom` and `Thickness`.
- `Top`: point map with vertical distance (m) from sea level to the top of a sandstone unit.
- `Bottom`: point map with vertical distance (m) from sea level to the bottom of the sandstone unit.
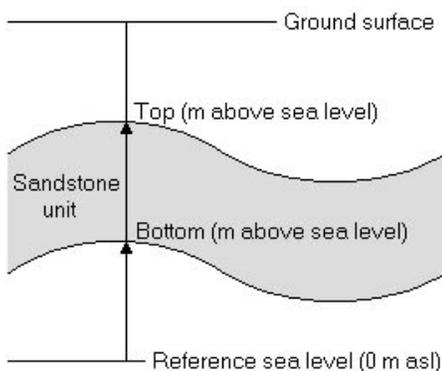

Figure 11.1: Schematic representation of a potential oil reservoir. Given are the vertical distances towards the top and the bottom of sandstone unit.

The spatial distribution of the points in the maps can be examined to acquire more knowledge about the phenomenon and the responsible process.

☞
- Double-click point map `Top` in the Catalog. The Display Options – Point Map dialog box is opened.
- In the Display Options – Point Map dialog box accept all defaults by clicking OK. The map is displayed in a map window.

Visually inspect the point map. Move with your mouse pointer over the map window and keep the left mouse button pressed. Notice that the point values are displayed if you position the mouse pointer near a point in the map. In the Status bar, you see the map coordinates changing while moving the mouse.

☞
- Open the Layers menu in the map window and choose Add Grid Lines.
- Accept the defaults and click OK.

Estimate the size of the map (X- and Y-extent) and verify this using the Measure Distance tool.

☞
- Press the pair of compasses in the toolbar of the map window and measure some distances.

Press the pair of compasses in the toolbar of the map window and measure some distances.

Measure the length of both the smallest and the largest point-pair vector i.e. the shortest distance between two points and the longest distance between two points. Find also the angle between each of these two point pairs.

☞
- Double-click Properties in the Layer Management pane. The Properties sheet of point map `Top` appears.
- In the Properties sheet, press the right mouse button on the text Point Map "Top" to open the context-sensitive menu.
- On the context-sensitive menu, select the Open as Table command.

! How many different point pairs can be made theoretically?

☞
- Close all windows and dialog boxes that are open.
- Repeat the previously described procedure to evaluate map `Bottom`.

11.1 Pattern analysis

Point interpolations assume spatial randomness of the input points. To examine the arrangement of the point data in space and to investigate whether your points are *randomly distributed* you can use the *Pattern analysis* operation prior to a point interpolation.

Spatial arrangement of point data in a point map can be one of the following three fundamental types (Figure 11.2): *Complete Spatial Randomness* (CSR), *Clustered pattern* and *Regular pattern*.

- In case of CSR, no correlation exists between locations of points.
- In a clustered pattern, subgroups of points tend to be significantly closer to each other than to other subgroups of points.
- In a regular pattern, individual point items tend to repel each other and distances between adjacent points tend to be further apart than for CSR.

> **!** In case your points are regularly distributed, e.g. as a regular grid, it is advised to directly rasterize the points with the Points to Raster operation and further inter-polate the raster map with the Densify or Resample operation or with the Kriging from Raster operation.
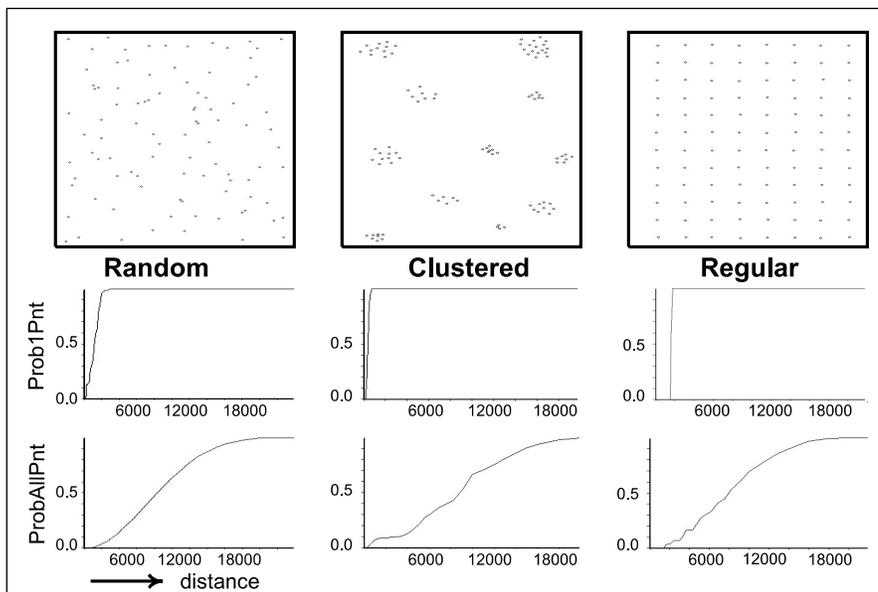

Figure 11.2: Fundamental types of patterns: Complete Spatial Randomness (CSR), Clustered pattern and Regural pattern.

There are two two basically different techniques that are used in ILWIS to examine the spatial pattern of points in a point map:

- *Distance to Nearest Neighbours*: this method involves analyzing spatial distribution of the points using distance characteristics between individual points in the map. The pattern is analyzed by calculating distances between individual points and its first to sixth nearest neighbour points in the pattern. The results are tested against the expected distances in CSR. If the individual points are closer than they would be for CSR, this indicates a clustered pattern. If, on the other hand, individual points are further apart than they would be in CSR, a more regular pattern is assumed.

- *Reflexive Nearest Neighbours* (*RNN*): this method considers two points as first order reflexive nearest neighbours if these points are each other's nearest neighbours. This definition can be extended to higher orders; second order RNNs are points that are each other's second-nearest neighbours, etc. The frequencies are calculated for RNNs of the first up to the sixth order. Most researchers suggest that higher order values in excess of CSR expectations indicate measure of regularity in arrangement of points, whereas lower empirical values imply elements of grouping in the pattern. If more (first order) RNNs occur than is expected in CSR, it can be concluded that isolated and relatively uniformly arranged couples exist.

The output of the pattern analysis operation is a table in which measured frequencies of occurrences of RNNs are compared with expected frequencies of occurrences in a situation of CSR. The CSR is simulated for the same area and the same number of points. If the number of observed RNNs are almost the same as the expected one in a situation of CSR, the point pattern is CSR. The points have a clustered pattern, when the number of observed RNNs are less than the expected ones in a situation of CRS. They have a regular pattern when the number of observed RNNs are higher than the expected ones in a situation of CSR.

The output is a table with eight columns and additional information. Column `Distance` lists distances that should be considered as distance from any point in the input point map. Column `Prob1Pnt` lists the probability that, within the specified distance (column `Distance`) of any point, another point will be found. `Prob2Pnt` lists the possibility that, within that distance, from any point, two other points will be found. The same goes for columns `Prob3Pnt`, `Prob4Pnt`, `Prob5Pnt` and `Prob6Pnt`. For a data set of *n* points `ProbAllPnt` is the summation of `Prob1Pnt` + `Prob2Pnt` + .. + `ProbnPnt-1`, divided by (*n - 1*). Additional information on pattern analysis is displayed in a separate window with three columns. First to sixth order nearest neighbours are listed in the column `Order`. Observed values for each order calculated using reflexive nearest neighbours and mean distance to nearest neighbour, are displayed in the column `Observed values`. The assumed values with CSR, for each order calculated using the above mentioned techniques, are listed in the column `Assumed with CSR`. The data present in the additional information window together with the data in the output table can be used to define the type of spatial distribution of the points in the map.

In this exercise the Pattern Analysis operation is performed on point map `Top`. To start the operation:

☞
- Double-click Pattern Analysis in the Operation-list. The Pattern Analysis dialog box is opened.
- Select point map `Top` in the list box Point Map.
- Type `Top_pattern_analysis` in the text box Output Table and click the Show button. A table window containing the result of the analysis is opened.

Display the table data in graphs and compare them with the graphs in Figure 11.2, to define a pattern for the point map. You can also use additional information to define how the points are distributed in the point map. To create a graph from the output table data:

☞
- From the Graphs menu in the table window, select the New Graph Window command. The Graph dialog box appears.
- Select column `Distance` for the X-Axis and column `ProbAllPnt` for the Y-Axis and click OK. The Graph Options sheet is opened.
- Accept the defaults by clicking the OK button.

Compare the shape of the graph with Figure 11.2. From the graph, you can see that beyond distances of about 2200 m always all other points will be found.

☞
- Repeat the exercise: select this time column `Distance` for the X-Axis and column `Prob1Pnt` for the Y-Axis and click OK. The Graph Options sheet is opened.
- On the X-axis tab, type Min – Max `0–750`, Grid Step `50` and Text Rotation `90°` and click OK.
- Accept all other defaults by clicking the OK button.

From the graph, you can see that for any point in the point map at distance 225 it is certain that at least 1 other point is found.

! In the table, find the distance at which at least 6 points will be found for any point. Are 6 points, which will form 6 point vectors for any point, enough to perform an interpolation?

To view the additional information:

☞
> • From the View menu of the table window, select Additional Info. The additional information is displayed in a separate window.
>
> • Close the additional information window, the graph windows and the table window when you have finished the exercise.

Summary: Pattern Analysis

- Pattern analysis is an operation to investigate whether your sample points are randomly distributed and should be used prior to a point interpolation.

- Three fundamental types of patterns exist: Complete Spatial Randomness (CSR), Clustered Pattern, and Regular Pattern.

- ILWIS uses two techniques to examine the spatial pattern of the points in a map: Distance to Nearest Neighbours, Reflexive Nearest Neighbours (RNN).

- If in the Distance to Nearest Neighbours the individual points are closer to each other than they would be for CSR than the points are clustered. If this is not the case the pattern is more regular.

- In RNN higher order values in excess of CSR expectations are a measure of regularity whereas lower empirical values imply elements of grouping in the pattern.

- If more (first order) RNNs occur than is expected in CSR, it can be concluded that isolated and relatively uniformly arranged couples exist.

11.2 Spatial autocorrelation of point data

Point interpolations assume also a certain degree of spatial correlation between input point values. The Spatial Correlation operation can be used to investigate whether point values are randomly distributed or whether the values have a relation with distance. Many variables that have discrete values measured at several specific geographic positions (i.e. individual observations can be approximated by dimensionless points) can be considered as random processes and can thus be analyzed using *spatial autocorrelation analysis*. Examples of such phenomena are: total amount of rainfall, toxic element concentrations, grain sizes, elevation at triangulated points, etc.

In ILWIS, a point map can be used for spatial correlation analysis. The input point map should either be a value map itself, or a class or ID map which has a linked attribute table with one or more value columns.

In the Spatial Correlation operation you can choose to use either the Omnidirectional or the Bidirectional method. The *omnidirectional* method simply determines all distances between point pairs, regardless of any direction. Thus all point pairs that have a certain distance towards each other will be counted in a distance class. Then, spatial autocorrelation (as *Moran's I*), spatial variance (as *Geary's c*), and experimental semi-variogram values are calculated for all point pairs within each distance class. The *bidirectional* method first counts, just like the omnidirectional method, all pairs of points that have a certain distance to each other, and then calculates the Moran's *I* and Geary's *c* for these point pairs within each distance class. Furthermore, all point pairs are counted with a certain distance to each other and with a certain *direction* towards each other. For point pairs in a certain distance class and in the correct direction, experimental semi-variogram values are calculated. Then also for the direction perpendicular to the specified direction, point pairs are counted and experimental semi-variogram values are calculated.

In this exercise the top and the bottom of the sandstone unit will be examined using the omnidirectional spatial correlation operation. The bidirectional option, that takes anisotropy into account, will be extensively treated in the part about kriging interpolation.

☞
- Double-click the Spatial Correlation operation in the Operation-list. The Spatial Correlation dialog box is opened.
- Select point map Top in the list box Point Map.
- Check if the Omnidirectional option is selected.
- Choose as Lag spacing 125 meter.
- Type Top_Spatcorr in the text box Output Table and click Show. The spatial correlation is calculated and the result of the operation is displayed in a table window.

Inspect the first two columns in the output table: i.e. the distance values represent the middle of a distance class.
- 70 point pairs are found of which the points are 62.5-187.5 m apart
- 147 point pairs are found of which the points are at least 187.5 m but less than 312.5 m apart, etc.

! Do you think these number of point pairs are sufficient to determine a suitable semi-variogram model later on?

Calculate the total number of point pairs and compare with the answer on page 407.

The next two columns, `I` and `c` give the calculated values for the point pairs in a specific distance class. Column `AvgLag` gives the calculated average distance between points of the point pairs in a certain distance class. Column `SemiVar` gives the semi-variogram values. This will be explained in section 11.4 (kriging).

Interpretation of the data in the output table will be done by means of graphs. The spatial autocorrelation function, shown in a graph is referred to as *spatial autocorrelogram*, showing the correlation between a series of points or a map and itself for different shifts in space or time. It visualizes the spatial variability of the phenomena under study. In general, point pairs that are close to each other have on average a lower variance (i.e. are better correlated), than point pairs at greater distances to each other. The autocorrelogram quantifies this relationship and allows to gain insight into the spatial behavior of the phenomenon under study.

Two graphs will be prepared for the depth to the top of the reservoir rock. One graph will show the change in correlation with distance and the other graph will show the change in variance with distance.

☞
- In the table window, open the Graphs menu and select the New Graph Window command. The Graph dialog box appears.
- In the Graph dialog box select the column Distance for the X-axis and the column I for the Y-axis and click OK. The Graph Options sheet is opened.
- On the Distance x I tab, select the Point option, the Color Red and click OK. The graph is displayed, showing the distance against the spatial autocor-relation (as Moran's *I*).
- Click the New Graph Window button in the Toolbar of the table window. The Graph dialog box shows up.
- Select column c for the Y-axis and click OK. The Graph Options sheet is opened.
- On the Distance x c tab, select the Point option, the Color Green and click OK. The graph is displayed, showing the distance against the spatial variance (as Geary's *c*).

Both graphs are based on the omnidirectional spatial correlation table, which also contains the total number of point pairs, used to calculate for each of the individual distance classes, the variance and correlation. Thus the points represent an average value for a large number of pairs of points, in order to filter out some of the erratic behavior of the variables. Note that the pairs are selected, based only on the distance separating them and not on their orientation, thus the curves are representative for the entire area rather than for a particular direction. This implies that directional components or anisotropy's in the data set cannot be detected.

In the distance versus variance (Geary's *c*) graph, you can see that at short distance the variance is low. This indicates that points close to each other are identical. When moving to larger pair separations, you may observe an increase in variance to a distance of approximately 500 meters. From 500 meters onward, the imaginary curve that can be fitted through the points (i.e. this curve is known as the *variogram*) tends to level off to a constant value. Thus, points separated over distances larger than 500 meter are no longer related or correlated to each other (notice this also in the graph showing the correlation versus the distance). This critical distance in geostatistics is known as the *range of influence* or *correlation length* of the data and is a measure of spatial variability: a long range implies homogenous data whereas a short range implies high spatial variability. Therefore, to ensure a proper sampling of the top of the reservoir rock, we need to have at least points with an average sample spacing of less than 500 meter, to capture the spatial variability. In general, a distance in the order of 1/4 of the range values (i.e. 125 meter spacing) is recommended. Note also that when extrapolating the value of the variance for a distance of zero, you find a small but non-zero value for the variance. This implies that there is variability for points separated by a distance of zero meters! This hypothetical variance is known as *nugget* variance and can be attributed to measurement errors and to small-scale variability occurring at a distance smaller than the average sampling interval.

☞
- Close all windows when you have finished the exercise.

- Repeat the procedure for point map `Bottom`. Evaluate the results and close the table and graph windows afterwards.

Summary: Spatial Correlation

- Spatial Correlation examines the correlation of a random process with itself in space.

- The spatial autocorrelation function shown in a graph and referred to as *spatial autocorrelogram* shows the correlation between a series of points or a map and itself for different shifts in space or time, and visualizes the *spatial variability* of the phenomena under study.

- A point map can be used for spatial correlation analysis if the points in the map are values or if the point map has a class or ID domain and the values are stored in a column of the attribute table that is linked to the map.

- The imaginary curve that can be fitted through the points of the variance versus distance graph is known as the variogram.

- The output of the omnidirectional spatial correlation operation is a table consisting of six columns: `Distance`, `NrPairs`, `I`, `c`, `AvgLag` and `SemiVar`.
    - Column `Distance` lists the middle values of the distance intervals;
    - Column `NrPairs` lists for each distance interval, the number of point pairs found at these distances towards each other;
    - Column `I` lists for each distance interval, the spatial autocorrelation of the point pairs in this distance interval;
    - Column `c` lists for each distance interval, a statistic for spatial variance of the point pairs in this distance interval;
    - Column `AvgLag` lists for each distance interval, the average distance between points of point pairs in this distance interval;
    - Column `Semivar` lists for each distance interval, the experimental semi-variogram value of the point pairs in this distance interval.

- The output of the bidirectional spatial correlation operation (the method that takes anisotropy into account and that will be explained in section 11.4.4.) is a table consisting of ten columns: `Distance`, `I`, `c`, `AvgLag1`, `NrPairs1`, `SemiVar1`, `AvgLag2`, `NrPairs2`, `SemiVar2`.

11.3 Point interpolation

A point interpolation (also known as gridding) performs an interpolation on randomly distributed point values and returns regularly distributed point values. The input for interpolation is mostly a point map with the domain type value or domain type class or identifier. If a point map has a domain type class or identifier, it is necessary that the map is linked to an attribute table, in which the attribute values are stored in a column with the domain type value. The output of a point interpolation is a raster map in which each pixel has a value calculated by an interpolation on the input point values. The various interpolation methods that will be treated in this section are: *Nearest Point*, *Moving Average*, *Trend Surface* and *Moving Surface*.

You will create the *isochore* map (i.e. equal thickness map) for the sandstone reservoir, by interpolating the values for the top and the values for the bottom and subtracting these two maps. Make sure that you use the same cell sizes for all calculations. When interpolating the point values, it is, for time efficiency reasons, strongly advised to choose a rather large pixel size for the output map. Further interpolation on the raster map values can be performed using the Densify, Resample or Kriging from Raster operations.

Nearest Point

In the *Nearest Point* method (also called *Nearest Neighbour*, *Voronoi Tesselation* or *Thiessen Polygons*) the value, identifier, or class name of the nearest point is assigned to the pixels according to the *Euclidean distance*. This method offers a quick way to obtain a Thiessen map from point data (Figure 11.3).

Figure 11.3: (A) An input point map. (B) The output map obtained as the result of the interpolation operation applying the Nearest Point interpolation method.

To interpolate the point map Top with the Nearest Point interpolation method:

☞
- Select point map Top in the Catalog. Use the right mouse button and select Interpolation, Nearest Point from the context-sensitive menu. The Nearest Point dialog box is opened.
- Type Top_np in the Output Raster Map text box.
➡

➥
- Select Wells as GeoReference.
- Type Nearest point interpolation in the text box Description.
- Accept all other defaults and click the Show button. The output map will be calculated and the Display Options - Raster Map dialog box will be opened.
- Accept the defaults by clicking OK. The interpolated map is shown.
- Check the values in the map and close it afterwards.
- Repeat the procedure for point map Bottom. Call the Output Map Bottom_np.

Now that you know the depth of the top and the depth of the bottom of the sandstone unit, you can calculate the thickness of the unit.

☞
- Type the following formula on the Command line of the Main window: Thickness_np = Top_np – Bottom_np ↵
- Accept the defaults in the Raster Map Definition dialog box and click Show. The map is calculated and the Display Options – Raster Map dialog box is opened.
- Click OK and check the values of the map and those of the input maps using the pixel information window. If you like, you can open the Display Options of map Thickness_np again and Stretch from e.g. 0-400 m.

Get an idea of where the sandstone unit is thick and where it is thin.

☞
- Drag and drop point map Sandstone from the Catalog in to the raster map Thickness_np. The Display Options – Point Map dialog box appears.
- In the Display Options – Point Map dialog box select Attribute Thickness, select the Text check box and click the Font button
- In the Font dialog box select Font Arial, Font Style Regular and Size 8. Click OK.
- Click the Symbol button and select Symbol +, Size 4 in the Symbol dialog box.
- Click OK in Display Options – Point Map dialog box. The point map Sandstone with attribute Thickness is now displayed on top of raster map Thickness_np.

Notice that the thickness observed in the isochore map diminishes towards an area stretching from the middle lower left-hand side of the mapped area to the upper right-hand side of the area. This is caused by a normal fault dipping south, cutting out part of the reservoir sandstone in the central part of the area. From the interpolated map

you can conclude that the *nearest neighbour interpolation* method is very quick, but far too rough for this specific application.

As additional geological exercise, try to sketch a geological cross section from the top (north) to the bottom (south) of the mapped area. Notice that the thickness of the sandstone diminishes in the central part of the area, due to a south dipping normal fault and that the reservoir is also folded.

☞
- Additional exercise: create a cross section through the area from the north to the south, with the method explained in section 10.6.

❗ See the ILWIS Help topic How to calculate a cross section through the terrain for more information about creating a cross section through the terrain.

Moving Average

The Moving Average method performs a weighted averaging on point values of a point. The output value for a pixel is calculated as the sum of the products of weights and point values, divided by the sum of weights. Weight values are calculated in such a way that points close to an output pixel obtain large weights and points further away obtain small weights. Thus, the values of points close to an output pixel are of greater importance to the output pixel value, than the values of points that are further away. You have the *inverse distance* and *linear* option to specify the method to calculate weight values (see the ILWIS Help topic Moving average Algorithm for more information). Furthermore, you also have to specify a *limiting distance*. Points that are further away from an output pixel than the limiting distance, obtain weight zero and thus have no influence on the output value for that pixel. This speeds up the calculation and prevents artifacts. It is however important to choose a limiting distance that contains enough data points.

☞
- Double-click the Moving Average item in the Operation-list to start the operation. The Moving Average dialog box is opened.
- Select point map Top in the list box Point Map.
- Type Top_ma in the Output Raster Map text box.
- Accept Inverse Distance as Weight Function.
- Select Wells as GeoReference.
- Type Moving average interpolation in the text box Description.
- Accept the other defaults by clicking the Show button. After the calculation, the Display Options - Raster Map dialog box will be opened.
- Accept the defaults and click OK. The map is displayed. ➡

➥

- Repeat the procedure for point map `Bottom` and call the Output Map `Bottom_ma`.
- Check the values of the maps and those of the input maps using the pixel information window.
- Close all windows

Now that you know the depth of the top and of the depth of the bottom of the unit, you can calculate the sandstone thickness in the same way as during the nearest point interpolation.

☞

- Type the following formula on the Command line of the Main window:
  `Thickness_ma = Top_ma – Bottom_ma` ↵
- Accept the defaults in the Raster Map Definition dialog box and click Show. The map will be calculated and the Display Options - Raster Map dialog box will be opened.
- Accept the defaults by clicking OK. The map will be displayed.
- Drag point map `Sandstone` with attribute column `Thickness` on top of the map and compare the map with the result from the nearest point interpolation (map `Thickness_np`).
- Close the map windows when you have finished the exercise.

From the output map you can conclude that the Moving Average interpolation method is a useful method for this specific problem. The erratic behavior near the fault trace can be due to natural variation/inhomogeneity. A more continuous area of low values could be achieved by directional interpolation (i.e. using a search ellipse instead of a search radius). Anisotropic Kriging, an interpolation method that will be treated in section 11.4, is a method that takes spatial dependencies into account.

Trend Surface

In the *Trend Surface* method, pixel values are calculated by fitting a surface through all point values in the map. The surface may be of the first up to the sixth order. A trend surface may give a general impression of the data. Surface fitting is performed by a *Least Squares Fit* (*LSF*).

☞

- Double-click the Trend Surface operation in the Operation-list. The Trend Surface dialog box is opened.
- Select point map `Top` in the Point Map list box.
- Accept `2nd degree` in the Surface list box.
- Type `Top_ts` in the Output Raster Map text box.

➥

☞
> - Select `Wells` as GeoReference.
> - Type `Trend surface interpolation` in the Description text box.
> - Accept all other defaults and click Show. After the map has been calculated, the Display Options - Raster Map dialog box will be opened.
> - Accept the defaults by clicking the OK button.

The result of the interpolation is a map consisting of surfaces representing homogeneous areas.

☞
> - Close the map window.
> - Repeat the procedure for point map `Bottom` and name the Output Map `Bottom_ts`.
> - Calculate the thickness of the sandstone unit by subtracting the map `Bottom_ts` from the map `Top_ts`. Call the Output Map `Thickness_ts`.
> - Compare the map `Thickness_ts` with the maps `Thickness_np` and `Thickness_ma`.
> - Close all map windows when you have finished the exercise.

From the output map you can conclude that the Trend Surface method is not useful for this particular data set. The result only shows a very general trend, without any local detail.

Moving Surface

In the *Moving Surface* method pixel values are calculated by fitting a surface through weighted point values. Weights for all points are calculated by a user-specified weight function. Weights may, for instance, equal the inverse distance. The weight functions are implemented in such a way that points which are further away from an output pixel than the user-specified limiting distance obtain weight zero. Surface fitting is performed by a least squares fit.

☞
> - Double-click the Moving Surface operation in the Operation-list.
> - Select point map `Top` in the Point Map list box
> - Accept `Inverse Distance` as Weight Function.
> - Select `2nd degree linear` in the Surface drop-down list box.
> - Type `Top_ms` in the Output Raster Map text box.
> - Select `Wells` as GeoReference.
> - Type `Moving surface interpolation` in the Description text box.

➡
> - Accept all other defaults and click the Show button. The calculation will start and after the calculation the Display Options - Raster Map dialog box will be opened.
> - Accept the defaults by clicking the OK button.

The result of this interpolation is a map consisting of surfaces representing homogeneous areas.

☞
> - Close the map window.
> - Repeat the procedure for point map Bottom and name the Output Map Bottom_ms.
> - Calculate the thickness of the sandstone unit. Call the Output Map Thickness_ms.

Compare the output map with the other interpolated maps.

☞
> - Display the thickness map of the Moving Average operation and the thickness map of the Moving Surface operation next to each other
> - Overlay both maps with point map Sandstone.
> - Furthermore open a pixel information window, and add both raster maps.
> - Zoom in and check the values of the maps.

! What is the best interpolation method?

☞
> - Close all map windows when you have finished the exercise.

Summary: Point interpolation

- Point interpolation also known as gridding performs an interpolation on randomly distributed point values and returns regularly distributed point values.

- The input for interpolation is either a value map itself or a class or ID map, which has a linked attribute table with one or more value columns.

- The output of a point interpolation, is a raster map in which each pixel has a value calculated by an interpolation on the input point values.

- The various interpolation methods are Nearest Point (also called Nearest Neighbour, Voronoi Tesselation or Thiessen Polygons), Moving Average, Trend Surface and Moving Surface.

- The ILWIS Help topic Point interpolation gives a useful overview of the different point interpolation techniques and their advantages and disadvantages.

11.4 Kriging estimation

Kriging is an alternative to many other point interpolation techniques. Unlike straightforward deterministic methods, such as Nearest Point, Trend Surface, Moving Average or Moving Surface, it is a statistical method based on the *theory of regionalized variables*. Before you can use kriging, you must make a semi-variogram model, which will determine the interpolation function. Furthermore kriging is the only interpolation method available in ILWIS that gives you an interpolated map and output *error map* with the standard errors of the estimates. In this manner, kriging allows you to quantify the quality of your predictions. As kriging is an advanced technique that heavily relies on statistical theory, the novice GIS user may need to ask advice from an expert.

Before you are going to use the kriging method you should have thought about things like:
- Do I really need the kriging interpolation method? When estimates with their errors are required, you should use kriging instead of another interpolation technique. Examples of situations where kriging could be very helpful are the mining industry, environmental research where decisions could have major economical and juridical consequences (e.g. is the area under study polluted or not) and so on.
- Is kriging the most appropriate interpolation method for my sample set? Before using an interpolation technique, first the assumptions of the method(s) should be considered carefully. As a user, you should choose the most appropriate method for your job. For example, when you want to do some modelling it is better to choose a straightforward interpolation method. When you are interested in the estimation errors however you should use kriging.

Once you have decided that kriging is the method you want to use, you should continue with the following steps:
- Step 1: Examining the input data
- Step 2: Calculating experimental variograms
- Step 3: Modelling variograms
- Step 4: Kriging interpolation

11.4.1. Ordinary Kriging

Visualization of the input data

In the Catalog, a map with Carbon 14 (^{14}C) water measurements in Botswana, Southern Africa is given and named C14. The map is linked to an attribute table C14.

Step 1 in the kriging process is the visual and statistical inspection of the data set:

☞
- Double-click point map C14 in the Catalog
- In the Display Options – Point Map dialog box select the Attribute check box, accept the default attribute column C14 and click OK. The map is displayed in a map window. ➥

➡
- Open the Layers menu in the map window and choose Add Grid Lines.
- Accept the defaults and click OK. The grid layer is displayed on top of the point map.
- Estimate the size of the map and verify this using the Distance tool.

Measure the length of both the shortest and the longest point-pair vector. A point-pair vector is the distance (and the angle) between two points. Zoom in if needed. Find also the angle and the difference in ^{14}C percentages for each of these 2 point pairs.

☞
- Double-click the word Properties in the Layer Management pane. The Properties of Point Map "C14" sheet appears.
- In the Properties sheet, press the right mouse button on the text Point Map "C14" to get the context-sensitive menu.
- From the context-sensitive menu, select Open as Table. The point map is opened as a table.

! How can you verify that the map is from an area in Botswana?

☞
- Also open table C14. In Layer Management pane, expand Properties and double-click the table C14 to open it.

In the C14 attribute column you will find the percentages of ^{14}C content, measured in well water at all locations appearing in the point map. Percentages are given with respect to a stabilized ^{14}C/^{12}C ratio (p.m.c. = percents 'modern carbon').

☞
- Close all windows.

Spatial Correlation and Empirical Semi-Variogram

Kriging assumes a certain degree of spatial correlation between the input point values. To investigate whether your point values are spatially correlated and until which distance from any point this correlation occurs, you can use the Spatial Correlation operation (as discussed in section 11.2) prior to a kriging interpolation. This is the second step in the kriging process.

☞

- Select point map C14 in the Catalog, press the right mouse button and select Statistics, Spatial Correlation. The Spatial Correlation dialog box appears.
- Specify a Lag spacing of 10 km.
- Call the Output Table Omnidirectional_10.
- Type in the Description box Point pair statistics with a lag-interval of 10 km. Remember that this text is optional and can always be changed later on.
- Accept all other defaults and click Show. The spatial correlation is calculated and the output table appears.

Inspect the different columns and their properties by double-clicking the column names and viewing the details, i.e. the Column Properties. Explain the contents of these columns, and compare in particular the Distance column with the AvgLag column.

What does 37 in the first field of the NrPairs column mean precisely?
How many point-pair "vectors" are longer than 55 km?
Do the six longest vectors show the expected semi-variance?

To check this, you should:

☞

- Open the C14 map with attribute C14 selected and square the ^{14}C differences between the most remote point pairs.

The next step is to create a point graph, i.e. a *discrete* experimental *semi-variogram* from the columns Distance and SemiVar of the Spatial Correlation table.

☞

- Choose the option New Graph Window from the Graphs menu in the table window.
- In the Graph dialog box, choose the Distance column for the X-axis and the SemiVar column for the Y-axis and click OK. The Graph Options sheet appears.
- On the Distance x SemiVar tab, select Point and change the Color into Red.
- Accept all other defaults and click OK.

What is the unit used for distance (X-axis) and what is the unit used for the semi-variogram values (Y-axis)?

☞ • Perform the Spatial Correlation operation again, now with a lag-interval of 5 km. Use `Omnidirectional_5` as Output Table name.

! Compare both output tables and explain the differences in the `NrPairs` column. To which distance interval refers the value 20,000 in the `Distance` column in the tables `Omnidirectional_10` and `Omnidirectional_5`?

Modelling the Semi-Variogram

The third step in the kriging process is to model the discrete values of your experimental semi-variogram by a continuous function, which will give an expected value for any desired distance. Figure 11.4 below shows a semi-variogram depicting a spherical model.
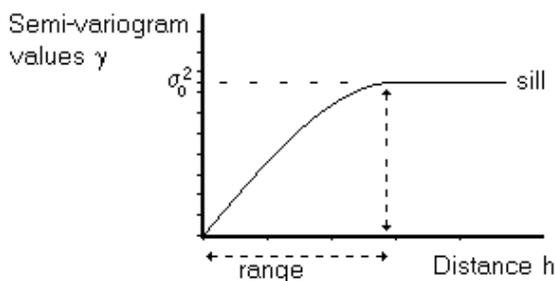

Figure 11.4: Example of a spherical semi-variogram model. σ_0^2 is the total variance of input data. g are semi-variogram values.

When the distance between sample points is 0, the differences between sampled values is also expected to be 0. Thus, the semi-variogram value γ at distance 0 equals 0, i.e. $\gamma(0)=0$. Samples at a very small distance to each other are expected to have almost the same values; thus, the squared differences between sample values are expected to be small positive values at small distances. With increasing distance between point pairs, the expected squared differences between point values will also increase. At some distance the points that are compared are so far apart that they are not any more related to each other, i.e. the sample values will become independent of one another. Then, the squared differences of the point values will become equal in magnitude to the variance of the variable. The semi-variogram no longer increases and the semi-variogram develops a flat region, called the *sill*. The distance at which the semi-variogram approaches the variance is referred to as the *range* or the span of the variable.

A semi-variogram with a *nugget* effect is a semi-variogram that goes from 0 to the level of the nugget effect in a distance less than the sampling distance. The semi-variogram model shows the semi-variogram value 0 at distance 0 and a discontinuity (jump) to a semi-variogram value at an extremely small distance. A nugget effect

indicates that the variable is erratic over very short distances, and/or that the variable is highly variable over distances less than the specified lag spacing or the sampling interval.

The theoretical or estimated semi-variogram model can be made as follows:

☞
- In the graph window of the `Omnidirectional_10` table, click the right mouse button to open the context-sensitive menu. Select Add Graph, Semivariogram Model.
- In the Add Graph Semi Variogram dialog box, choose `Spherical Model` for the SemiVariogram model, use values `200 m`, `800 m` and `40 km` for the Nugget, Sill and Range respectively and click OK. The Graph Options sheet appears.
- Accept the defaults and click OK.

❗ At which distance do you find the greatest discrepancy between the semi-variogram model and the measured values?

☞
- In the graph window, click the right mouse button to open the context-sensitive menu and select the Graph Options command. The Graph Options sheet appears.
- On the Spherical Model tab, modify the Sill until you obtain a good fit between curve and plotted points until distances ≤ 40000 m.

Make a list of all available semi-variogram models (Spherical, Circular, Exponential, Gaussian, etc.), their numeric parameters and the units in which these parameters are measured.

☞
- In the Graph Options sheet, click the Graph Management tab, click the Add Graph button, and select Semivariogram Model.
- Add a `Circular` semi-variogram model using the same values for Nugget, Sill and Range as before. Use a different color to display the model.

Variogram Models: Goodness of Fit

You can use the Column SemiVariogram operation to find which model fits your experimental semi-variogram values best. The Column SemiVariogram operation calculates semi-variogram values according to a user-specified model and stores the calculated values in an output table column. To create new columns for both the rounded distances and the real mean distances per distance class:

☞
- In the `Omnidirectional_10` table, open the Columns menu and choose the SemiVariogram operation. The Column SemiVariogram dialog box appears.

- Select `Distance` for the Distance column, choose `Spherical Model` for the SemiVariogram model and use values `200`, `750` and `40000` for the Nugget, Sill and Range respectively.

- Call the Output Column `SemiCol_Dist` and click OK. The Column Properties dialog box is opened.

- Accept all defaults and click OK.

- Repeat the procedure but use this time `AvgLag` as Distance column and call the Output Column `SemiCol_Avg`.

- Make graphs of these 2 output columns. Plot respectively `SemiCol_ Dist` against `Distance`, and `SemiCol_Avg` against `AvgLag`.

! Visually inspect the difference between experimental and model outcomes.

You can also test which semi-variogram model fits your data best by calculating the *Goodness of Fit* (R^2).

$$R^2 = 1 - \frac{\sum (\hat{\gamma}(h_i) - \gamma(h_i))^2}{\sum (\hat{\gamma}(h_i) - \sum \hat{\gamma}(h_i)/N)^2}$$

where:

$\hat{\gamma}$ are the experimental semi-variogram values calculated by the Spatial Correlation operation;

γ are the semi-variogram values calculated by the Column SemiVariogram operation;

N is the total number of distance classes/intervals.

The numerator of the fraction gives the sum of the squared differences between the experimental semi-variogram values and the semi-variogram values calculated by a user-specified semi-variogram model.

The denominator of the fraction gives the sum of the squared differences between the experimental values and the average experimental semi-variogram values of all distance classes/intervals.

The Goodness of Fit indicator can be calculated using TabCalc statements. The terms in the numerator are computed by first making a new column in the `Omnidirectional_10` table with an expression as:

Outputcol = SQ(*SemiVar-SemiCol*)
where:
Outputcol   is the name of the out put column;
SemiVar   is the name of the column with the experimental semi-variogram values;
SemiCol   is the name of the column with the semi-variogram values calculated with the Column SemiVariogram operation.

☞

> • Type the following formula on the Command line of the table window:
>   `G_Gam_Sqr = SQ(SemiVar-SemiCol_Avg)` ↵

The terms in the denominator first require the value of Σ (h$_i$)/N (i.e. the average experimental semi-variogram values) in a new column:

☞

> • Type for example the following formula on the Command line of the table window: `Mean_G = AVG(SemiVar)` ↵

This average must be used in a new column with squared terms:

☞

> • Type the following formula on the Command line of the table window:
>   `G_Avg_Sqr = SQ(SemiVar – Mean_G)` ↵

The "Goodness of Fit" indicator R^2 can now be obtained by:

R2 = 1 – SUM(*G_Gam_Sqr*) / SUM(*G_Avg_Sqr*)

☞

> • Compute the "Goodness of Fit" indicator.

The maximum value of R^2 is 1, meaning an exact match of calculated semi-variogram values and experimental semi-variogram values.

! One can obtain a better fit by excluding the distances beyond the range. See the ILWIS Help for more information.
The TabCalc statements mentioned above are also available in an ILWIS script. You can run the Goodness of Fit script by expanding the Script item in the Operation-tree, then double-click the R2 script.

Once you have decided which semi-variogram model and which values for sill, range and nugget fit your data best, you can continue with the kriging interpolation.

Kriging estimation and Error map

The final step in the kriging process is the calculation of the kriging estimations and the estimation errors.

☞
- Double-click the Kriging operation in the Operation-list. The Kriging dialog box is opened.
- Select point map C14 with attribute column C14 as input Point Map.
- Accept the default Spherical Model and for Nugget, Sill and Range fill out the values as used before.
- Specify a Limiting Distance of 25 km.
- Take 4 for the Minimum and 14 for the Maximum number of points.
- Call the Output Raster Map Ord_krig25.
- Use GeoReference C14 and select the Error Map check box.
- Accept all other defaults and click Show. The kriging map with its error map is calculated and after you click OK in the Display Options dialog box, it is displayed.

! Which parameters influence the calculation speed? Explain the numbers in the progress bar.

☞
- Use the right mouse button in the Catalog to view the Properties of raster maps Ord_krig25 and Ord_krig25_Error.
- Find the range of output values in both maps.

! What are the units in which the values of the output kriging map and the output error map are expressed?

Perform another kriging estimation on the C14 point map.

☞
- This time use, a Limiting Distance (search radius) of 40 km and specify all other parameters as used before.
- Call the Output Raster Map with the kriging estimates Ord_krig40.
- Open both error maps Ord_krig25_Error and Ord_krig40_Error and close all other map windows.
- Drag the input point map C14 on top of each of the error maps.
- In the Display Options – Point Map dialog box select Attribute C14, select the Text check box and press the Font button.
- In the Font dialog box use the Font Arial, Font Type Regular, Size 8pt and click OK.
- Click the Symbols button and use Crosses for the Point Symbols
- Find the point with the best and the point with the worst kriging estimate in both error maps and close the window afterwards.

Confidence interval maps

The output maps of the kriging operation (the output kriging map with the estimates and the output error map) can be used to create (3-valued) confidence interval maps.

In this section a 3-valued confidence interval map will be made using the maps `Ord_krig25` and `Ord_krig25_Error`. As reference value we take the value 30.

The ILWIS `IFF` function enables us to make confidence interval maps that show the upper and lower limits of confidence intervals.

☞
- Type the following formulae on the Command line of the Main window:
  ```
  Over30 = IFF(Ord_krig25 – Ord_krig25_Error > 30,1,0) ↵
  Under30 = IFF(Ord_krig25 + Ord_krig25_Error < 30,1,0) ↵
  ```

These expressions produce raster maps with values 1 and 0, and give with a probability of about 84% (the 'sigma' limit), that the ^{14}C p.m.c. is over respectively under the reference value of 30.

Combining both maps produces a 3-valued occurrence map.

☞
- Type the following formula on the Command line of the Main window:
  ```
  Occurrence_3_valued = Over30 + Under30 ↵
  ```

For more information about creating confidence interval maps, see the ILWIS Help topic "How to calculate confidence interval maps".

11.4.2 Simple Kriging and Indicator Kriging (additional exercise)

Simple Kriging

Perform a Simple Kriging interpolation similar to the previous exercise. List the parameters which are not needed in Simple Kriging compared to Ordinary Kriging.

☞
- Call the Output Raster Map `Simple_krig40`.
- Drag the input point map `C14` over raster map `Simple_krig40_Error` in the same way as before, and compare the result with the previously calculated error maps.
- Find the point with minimum and the point with maximum error variance ("uncertainty").
- Close all windows after finishing the exercise.

Indicator Kriging

☞
> • Open table `C14` that contains the attributes of the point map `C14`.

Column `Over50` has values 0 and 1 which "indicate" whether a ^{14}C sample surpasses a cutoff of 50 p.m.c. or not.

Indicator Kriging can be carried out by considering the attribute `Over50` as point measurements with values 0 and 1, which have spatial continuity and are modelled by a semi-variogram.

☞
> • Make an Omnidirectional Spatial Correlation table and find a semi-variogram model, which reasonably fits the experimental values.
>
> • Perform an Ordinary Kriging operation with these model settings, call the Output Raster Map `Indicator_kriging` and make sure that also an error map is created.

The output map is called an *indicator map*.

11.4.3 Universal Kriging

Universal Kriging is a variant of Ordinary Kriging. Universal Kriging is kriging with a local trend or drift. This local trend or drift is a continuous and slowly varying trend surface on top of which the variation to be interpolated is superimposed. The local trend is recomputed for each output pixel and the operation is therefore more similar to the Moving Surface operation than to the Trend Surface operation. In ILWIS this supposed trend can be of degree 1 (tilted trend plane) or of degree 2 (quadratic trend surface).

☞
> • Interpolate the `C14` point map using Universal Kriging. Use the linear option as well as the quadratic option.
>
> • Use the same Variogram Model with the same parameter values as in the `Ord_krig25` interpolation.
>
> • Call the prediction maps respectively `Universal1` and `Universal2` and select in both cases the Error Map check box.

! Can the results of this operation be compared with the outputs of the Ordinary Kriging operation? Explain!

Remark:
Extrapolation outside the default value range causes that many pixels in the output map get an undefined value. This can be avoided by setting a relatively large value range for the output kriging map, either on the Command line or in the dialog box.

11.4.4 Anisotropic Kriging

Investigating anisotropy with variogram surfaces

Sometimes, the variation in data may not simple have a relation to distance, but the relation to distance may occur in a certain direction only. If that is the case, your data is said to be anisotropic. To krige such data you should use the Anisotropic Kriging operation.

In the following exercises, you will investigate with the *Variogram Surface* operation whether data is anisotropic or not. The Variogram Surface operation produces a 'raster' plot (with an origin) in which each cell contains the semi-variogram value for the specific distance class and the specific direction of the cell in relation to the origin of the plot where the distance and the direction are zero. The output plot of the Variogram Surface operation, helps you to visualize whether any anisotropy occurs, and to determine parameters as direction, tolerance and band width of the anisotropy.

Then from the output table of the Spatial Correlation operation, using the bidirectional method, graphs can be constructed (distance against semi-variogram values) in the same way as from the output table of the Spatial Correlation operation omnidirectional method as before.

From the graph, you can determine the appropriate semi-variogram model. When the parameters for the semi-variogram model are clear, you can finally krige your anisotropic data with the Anisotropic Kriging operation.

For the exercises, you will use point map `Artificial` which has attribute `Samples` in an attribute table. Attribute `Samples` is expected to show anisotropy when using lags smaller than 500 m.

Variogram Surface

☞
- Double-click the VarioSurface operation in the Operation-list. The Variogram Surface dialog box is opened.
- Select point map `Artificial` with attribute column `Samples` as Input Map. Accept the proposed defaults: Lag spacing 55.1 and Number of lags 10.
- Call the Output Raster Map `Var_surf55`, accept the other defaults and press the Show button. The variogram surface is calculated and the Display Options – Raster Map dialog box appears.
- Accept the defaults in the Display Options – Raster Map dialog box by clicking OK. The variogram surface is displayed.

The output of the Variogram Surface operation is a plot with the origin in the center. Each cell in the plot (each pixel in the output map) has the size of the user-specified Lag spacing. The number of cells of the output surface (from the central

cell at the origin towards positive and negative X and Y axis) is defined by the user-specified Number of lags. The cells in the output surface represent directional distance classes: each cell contains the semi-variogram value of all point pairs of which the separation vector ends up in that cell/pixel (see Figures 1-4 in the ILWIS Help topic "Variogram surface: Functionality").

The output surface can be best viewed with a representation Pseudo. Semi-variogram values close to the origin of the output map are expected to be small (blue in representation Pseudo), as values of points at very short distances to each other are expected to be similar. When there is no anisotropy, semi-variogram values will gradually increase from the origin into all directions. You will then find circle-like shapes from the origin outwards, while the color gradually changes from blue (at the origin) to green and red (away from the origin). Your data is supposed to be anisotropic when you find an ellipse-like shape of low semi-variogram values in a certain direction going through the origin. In this direction, the semi-variogram values do not increase much. However, in the perpendicular direction, you find a clear increase in semi-variogram values.

! How many cells does the output variogram surface contain?
What is the distance between the origin and one of the corner cells in the variogram surface?
Explain the meaning of a semi-variogram value in a corner cell.

☞ • Make another Variogram Surface; use 20 lags and a lag spacing of 10 km.

• Call the output map Var_surf55_20

! Do you discover a geometric anisotropy in the spatial continuity of the measured regionalized variable?
In which direction?

Change of support and/or change of scale

Repeat the Variogram Surface operation:

☞ • For the Input Map, use point map Artificial, attribute Samples.

• For the Lag spacing, use 100, 200, 500 and 900 m respectively.

• For the Number of lags, use 10. Use of 100 m and 10.

• Name the Output Raster Maps: Var_surf_100, Var_surf_200, Var_surf_500 and Var_surf_900.

! You can use the history of the Command line to retrieve previously used commands.

Display the surfaces using the same stretch range and the same representation (Pseudo). Compare the output surfaces; determine the directions (azimuth) of the principle axis of anisotropy.

Spatial Correlation, bidirectional

Now we will use the bidirectional method of the Spatial Correlation operation. semi-variogram values, will be calculated for the principle axis of anisotropy and the perpendicular 'direction'. You need to specify the parameters Direction, Tolerance and Band width. These parameters should be found from the variogram surfaces as created above. The parameters for the bidirectional method are schematically presented in Figure 11.5.
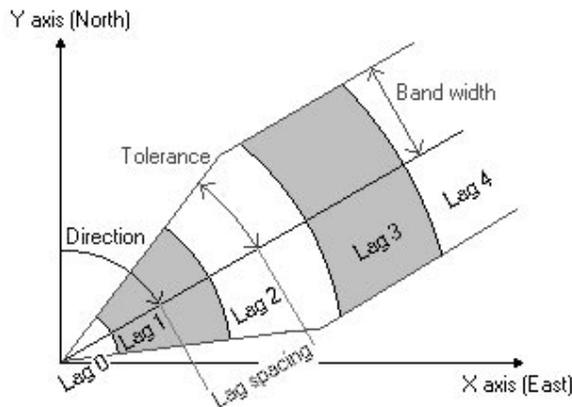

Figure 11.5: Schematic explanation of the input parameters for Spatial Correlation bidirectional method: lag spacing, nr of lags, direction, tolerance and band-width.

The *direction* angle is measured clockwise from the Y-axis and defines the direction in which points should be located relative to each other. When you use a direction angle of 90°, it means that only point pairs for which the points are located in west-east or in east-west direction will be considered (i.e. +90° clockwise from the Y-axis). The *tolerance* angle is a parameter with which you can limit the number of point pairs. When a tolerance of 45° is used, all point pairs in the map will contribute to calculated semi-variogram values. When using a tolerance of 10°, the direction of every 2 points may differ -10° or +10° from the specified direction (90°). So, in fact, all points that are found in a position within 80° to 100° to one another are valid pairs. Then, for the valid point pairs, the distance class to which they belong will be determined.

Optionally, you can set a third parameter, the *band width* (m), to limit the tolerance to a certain width.

You should find the values for these parameters from your variogram surfaces as created in the previous exercise. To calculate the spatial correlation table for Point Map Artificial:

☞
- Select Point Map `Artificial`, press the right mouse button and choose Statistics, Spatial Correlation from the context-sensitive menu. The Spatial Correlation dialog box appears.
- Use parameter `Samples` as attribute column again.
- Select the option Bidirectional.
- Use a Lag spacing of `100` m and use the Direction angle as found in the previous exercises (Variogram surfaces).
- Call the Output Table `Artificial_bidirect`, accept all other defaults and click Show. The Bidirectional Spatial Correlation table is calculated and shown.

Inspect the columns of the table `Artificial_bidirect` and explain the main differences between the columns `NrPairs1` and `NrPairs2` and the columns `SemiVar1` and `SemiVar2`.

Semi-variogram models

In the table window, open the Graphs menu and select New Graph Window. The Graph dialog box is shown.

☞
- Plot `SemiVar1` (Y-column) against `AvgLag1` (X-column). Use the Point option with `Red` points and take as X-Axis bounds `0` and `1000` and as Y-Axis bounds `0` and `10`.
- Do the same for `SemiVar2` against `AvgLag2`. Use this time the Color `Green` as point color.
- Create also a third graph. Plot the `NrPairs` against `Distance`. Take now Needle as graph presentation and select again a different color. Set the maximum X-axis bound at `1000` and leave the Y bounds unchanged.

❗ Do you conclude a trend or anisotropy from these figures? Why?

Note that the first distance class contains very few point pairs. When this first point in the graphs is neglected, one can fit Power models through both graphs.

☞
- In the graph of `Direction` against `SemiVar1` add a semi-variogram model. Choose the `Power` model and use for instance Nugget `1.2`, Slope `0.000015` and Power `1.78`.

Furthermore, we need to now the relation between semi-variogram values on the principle axis of anisotropy and in the perpendicular direction. When you inspect the variogram surface output 'map' you can use the Distance tool and measure the

distance from the origin to a certain semi-variogram value on the principle axis of anisotropy (i.e direction 20°). Then try to find the same semi-variogram value in the perpendicular direction (i.e. 110°). The ratio of these figures is called the scale factor.

For `SemiVar2` the Slope could be `0.00002` but in case of geometric anisotropy we

Scale Factor = A/B

A

B

A = Principle axis of anisotropy
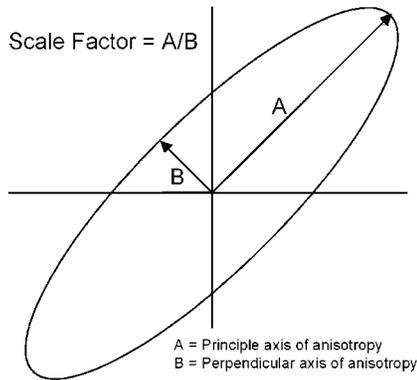B = Perpendicular axis of anisotropy

Figure 11.6: Explanation of the scale factor and the principle axis and perpendicular axis of anisotropy.

only need one semi-variogram model and the scale factor. If we consider a Range or Limiting Distance of `600` meter, a Ratio of 6/5 (= `1.2`) can be taken.

Anisotropic Kriging

We are now going to use the anisotropy found in the Variogram Surface operation of point map `Artificial` in the Anisotropic Kriging interpolation. Anisotropic Kriging is a variant of the Ordinary Kriging operation. Anisotropic Kriging incorpotates the influence of direction dependency, i.e. anisotropy in the kriging process.

☞

- Perform an Anisotropic Kriging operation on attribute `Samples` with the above mentioned findings.

- Use an Angle of `20`°, a Ratio of `1.2` and a Limiting Distance of `600` m.

- Call the Output Raster Map `Anisotropic`.

- Create a Georeference. Use Coordinate Systems `Unknown`, accept the default boundary values but change the Pixel size into `20`m.

- Select the Error Map checkbox and click the Show button to start the calculation. After the calculation, the Display Options dialog box shows up. Accept all defaults and click OK to display the map.

- Close the map window when you have finished the exercise.

11.4.5 CoKriging

Often, data may be available for more than one attribute per sampled location. One set of samples may be expensive or difficult to measure and is therefore sampled infrequently while another variable may be cheap or easy to measure and has more observations or more accurate ones. If the correlation between the two variables is high (positive or negative), then it may be possible to use the information about the spatial variation of a well-sampled variable (the *covariable*) to help to interpolate a sparsely sampled variable (the *predictand*). This estimation method, in fact the multivariate variant of Ordinary Kriging, is known as *CoKriging*.

Cross Variogram

Before Cokriging, you should perform the *Cross Variogram* operation. The Cross Variogram operation calculates experimental semi-variogram values for two variables (the predictand and the covariable) and cross-variogram values for the combination of both variables. As two variables are handled simultaneously, the Cross Variogram operation can be seen as the multivariate form of the Spatial Correlation operation. From the output table of the Cross Variogram operation, you can create semi-variogram models for both variables, and a cross-variogram model for the combination of the variables. The models do not need to be identical but should obey the *Cauchy-Schwarz* inequality. For more information, see the ILWIS Help.

In this exercise, point map Mander with moisture deficit observations near Mander, Overijssel, the Netherlands will be used. The map is linked to an attribute table Mander. The table has 3 columns: column Moist76 (with moisture deficit values for the year 1976), column Sub76 (with a subset of column Moist76) and column Avg30 (with moisture deficit values for a 30 year average).

☞
- Open point map Mander.

The coordinates, which you see on the Status bar when moving the mouse pointer, are meters in the Dutch national grid system (RD) with stereographic projection.

☞
- Open table Mander and notice that column Sub76 is a subset of column Moist76.
- In the table window select Columns, Statistics. The Column Statistics dialog box is opened.
- In the Column Statistics dialog box select the Correlation function and select the attribute columns Sub76 and Avg30.
- Click OK to show the result.

Write down the correlation between Sub76 and Avg30.

☞
- Perform Spatial Correlation operations on the attributes `Sub76`, `Moist76` and `Avg30`. Use in all three cases a Lag spacing of `200` meter.
- Call the new tables respectively `Spat_corr_Sub76`, `Spat_corr_Moist76` and `Spat_corr_Avg30`.
- Also make a Cross Variogram table (Lag spacing `200` meter) for the correlated attributes `Sub76` and `Avg30`.
- Use TabCalc commands to check the Cauchy-Schwarz condition: `Cauchy_Schwarz = SQ(CrossVarAB)< SemiVarA * SemiVarB`.
- Display the graphs of the 2 semi-variograms and the cross-variogram and try to fit them with `Power` models. Hint: the slopes are small values (about `0.02` or less) and the `Power` exponents are around `1.5`.
- In the Graph window, use the context-sensitive menu to modify the way the different variograms are displayed. Use `Red` for the Predictand `Sub76`, `Green` for the Covariable `Avg30` and `Blue` for the Cross-Variogram.

Write down the parameters for the Nugget, Slope and Power from the 'best' fitting models. Use these models in the Cokriging operation such that the interpolation of the Predictand `Sub76` values is supported by the Covariable `Avg30`.

☞
- Start the CoKriging operation and fill out the parameters as found above.
- Use for the Limiting Distance `800` meter and for the Minimum and Maximum number of sample points `1` and `8` respectively.
- Call the Output Raster Map `Cokriging`.
- Create a GeoReference. Use a Pixel size of `10` or `20` m and call the new georeference `Mander10` or `Mander20`.
- Select the Error Map check box, and click the Show button to execute the operation.

Subsequently, interpolate attribute `Sub76` via the Ordinary Kriging method.

☞
- Use the same Variogram Model with the same values for the parameters as given above and call the Output Raster Map `Ord_krig_Sub76`.
- Compare the error maps of both the CoKriging and the Ordinary Kriging operation and observe the error reduction due to the use of a correlated covariable.
- Close all windows after you have finished the exercise.

Summary: Kriging estimation

- The four major steps in a kriging interpolation are: (1) Examining the input data, (2) Calculating experimental variograms, (3) Modelling variograms and (4) Kriging interpolation.

- Kriging assumes a certain degree of spatial correlation between the input point values. To investigate whether your point values are spatially correlated and until which distance from any point this correlation occurs, you can use the Spatial Correlation operation.

- In the Spatial Correlation operation you can either choose the *omnidirectional* or the *bidirectional method*. The omnidirectional method simply determines semi-variogram values in all directions. The bidirectional method on the other hand calculates semi-variogram in a certain direction and perpendicular to that direction.
  - When the variation of the variable under study is not the same in all directions, then anisotropy is present and you must use the bidirectional method. In case of suspected anisotropy you should calculate a variogram surface with the Variogram Surface operation beforehand.
  - When you want to use Cokriging, you should investigate the input point values with the Cross Variogram operation.

- From the `Distance` and `Semivar` columns of the Spatial Correlation or Cross Variogram output table you can create point graphs, i.e. *experimental semi- and cross-variogram(s).*You must try to fit a semi-variogram model through the experimental semi-variogram. The model will be used as interpolation function in the kriging operation.
  - Several *variogram models* exist e.g. Spherical, Gaussian, Exponential, Power, Wave, Rational quadratic, Circular.
  - Each type of semi- or cross-variogram model has a *sill*, a *range* and a *nugget* (or when using the Power model values for the *sill*, *slope* and *power*).
  - You can check which model fits your data best by using the Column SemiVariogram operation in combination with the *Goodness of Fit test*.

- Kriging can be seen as a point interpolation that requires a point map as input and returns a raster map with estimations and optionally an error map.

- In ILWIS several kriging methods are available: Simple Kriging, Ordinary Kriging, Anisotropic Kriging, Universal Kriging and Cokriging.
  - Simple Kriging is the simplest kriging method; all input points are used to calculate each output pixel value
  - Ordinary Kriging is the kriging method in which you can influence the number of points that should be taken into account in the calculation of an output pixel value by specifying a limiting distance and a minimum and maximum number of points.
  - Universal Kriging is variant of the Ordinary Kriging operation; it is kriging with a local trend. The local trend or drift is a continuous and slowly varying trend surface on top of which the variation to be interpolated is superimposed.

- Anisotropic Kriging is another variant of the Ordinary Kriging operation. It incorporates the influence of direction dependency i.e. anisotropy in the kriging process.
- Co-Kriging is a multivariate extension of the Ordinary Kriging technique. CoKriging calculates estimates or predictions for a poorly sampled variable (the predictand) with help of a well-sampled variable (the covariable). Both variables should however be highly correlated (positive or negative).

- From the combination of a kriged output map containing the estimates and its output error map, you can create *confidence interval maps*.

11.5 Spatial autocorrelation of raster maps

The *AutoCorrelation and SemiVariogram* operation calculates the autocorrelation and the semi-variance, between pixel values of a raster map and pixel values of the same raster map for different shifts (lags), in horizontal or vertical directions. This method is based on the assumption that a regional variable becomes random at a great distance, i.e. the correlation of a regional variable with other pixels will normally decrease with increasing distance. The AutoCorrelation – Semivariance operation should be used prior to a raster interpolation operation like Densify, Resample or Kriging from Raster

The results of auto-correlation and semi-variance in vertical and horizontal directions are stored in a table. The data of this table can also be represented as graphs. The output table contains the following columns:
- `Distance`: distance values
- `HorCorr`: horizontal autocorrelation
- `VertCorr`: vertical autocorrelation
- `HorSemVar`: horizontal semi-variance
- `VertSemVar`: vertical semi-variance.

To assess the effect of spatial resolution on image statistics and the spatial structure of the image, variograms can be used.

The data set available in this section contains parts (250 by 250 pixels) of Landsat TM scenes, with three different characteristic lithologies:
- `Gneiss`: Raster map of gneissic units, extracted from a Landsat TM scene acquired on the 25th of March 1989 of an area near Málaga in southern Spain.
- `Schist`: Raster map of schist, extracted from a Landsat TM scene acquired on the 19th of July 1984 of an area near Montalbán in central Spain.
- `Carbonate`: Raster map of a folded carbonate sequence, extracted from a Landsat TM scene acquired on the 19th of July 1984 of an area near Montalbán in central Spain.

☞
- Double-click raster map `Gneiss` in the Catalog. The Display Options – Raster Map dialog box is opened.
- Accept all default settings and press OK. The map is shown.
- Check the values in this map by clicking some pixels. Close the map window.
- Repeat the procedure for the maps `Schist` and `Carbonate`.

Now you will calculate and display semi-variograms for the three lithologies.

☞

- Double-click the Autocorrelation-Semivariance operation in the Operation-list. The AutoCorrelation and SemiVariogram dialog box is opened.

- Select raster map Gneiss in the Raster Map list box.

- Type Autocorr_Gneiss in the text box Output Table. Accept the default Maximum Pixel Shift and click the Show button. The result of the operation is displayed in a table window.

- From the Graphs menu select the New Graph Window command. The Graph dialog box appears.

- Select column Distance for the X-axis, select column HorCorr for the Y-axis and click OK. The Graph Options sheet is opened.

- On the Distance x HorCorr tab, select Point, the Color Red and click OK.

- In the graph window, click the right mouse button to open the context-sensitive menu. Select the Add Graph, from Columns command. The Graph dialog box shows up.

- Select the column VertCorr for the Y-axis and click OK. The Graph Options sheet appears.

- On the Distance x VertCorr tab, select Point, choose Color Green and click OK again.

- Click the New Graph Window button. The Graph dialog box is opened.

- In the Graph dialog box, select column Distance for the X-axis, select column HorSemVar for the Y-axis and click OK.

- On the Distance x HorSemVar tab, select the Point option, the Color Red and click OK.

- In the graph window, click the right mouse button to open the context-sensitive menu. Select the Add Graph, from Columns command. The Add Graph dialog box appears.

- Select the column VertSemVar for the Y-axis and click OK. The Graph Options sheet appears.

- On the Distance x VertSemVar tab select the Point option, choose Color Green and click OK again.

- Repeat the previous described procedure for the maps Schist and Carbonate.

In general, the semi-variograms show low values for short lag spacings, indicating that pixels that are close or adjacent are relatively similar (i.e. the variance is low). In some cases the curve does not seem to approach the origin of the plot, but tends to reach a small Y-intercept known as nugget. This indicates that there is variability at very short (almost zero pixel) distances. In remote sensing, this variance can be attributed to noise in the images. At larger lag spacings, the increase in semi-variance tends to level off to a constant value, known as the sill of the semi-variogram. The distance at which this sill is reached is known as the range of the variogram.

Specific features that can be observed in the semi-variograms for the different lithologies are:

- *Gneiss*: The semi-variogram has a very short range due to the fact that the gneisses lack bedding or other kind of structurization. The vertical and horizontal semi-variograms are identical, due to the homogenous and isotropic landforms on the gneissic terrain lacking any preferred orientation. The sill is high, due to strong differences in absolute digital numbers between the sunlit slopes and the shadowed slopes.
- *Schist*: The range of the semi-variograms of the schist is very long pointing to very continuous data with less spatial variability. The sill value is very low due to small spectral variation observed in the image.
- *Carbonate*: The carbonate image has again a long range and a low sill value. Note the differences between the vertical and horizontal semi-variograms. The long range in the horizontal semi-variogram points to more continuity and the short range in the vertical semi-variogram to more variability in the respective direction. This can be attributed to the bedding that predominantly strikes east-west and is pronouncedly clear, due to the alternation of carbonates and marls.

☞
- Close the graph and the table windows when you have finished the exercise.

If the Gneiss, Schist and Carbonate maps would not have been surfaces yet but would have been real sample point surrounded by undefined values, you could have:

- Performed the Variogram Surface operation to determine the anisotropy for Gneiss, Schist and Carbonate.
- Perform AutoCorrelation – Semivariance operation again for Gneiss, Schist and Carbonate.
- Displayed the experimental semi-variograms for Gneiss, Schist and Carbonate.
- Chosen an appropriate model, to determine the parameters for sill, nugget and range.
- Performed the Kriging from Raster operation.

For more information, see the ILWIS Help topic "Kriging from Raster".

Summary: Autocorrelation

- AutoCorrelation – Semivariance calculates autocorrelation and semi-variogram values for a raster map and is more or less similar to the bidirectional Spatial Correlation operation for point maps.

- The AutoCorrelation – Semivariance operation should be used prior to a raster interpolation operation like Densify, Resample or Kriging from Raster.