10. Variogram Analysis

ACE 492 SA - Spatial Analysis
Fall 2003

1 Objectives

The purpose of this lab is to learn to apply ESRI's *Geostatistical Analyst Extension* to carry out exploratory variography. Only a limited subset of the the functionality of the extension will be covered. You are welcome to explore this extension in more detail using the Tutorial in Chapter 2 of ESRI's *Using ArcGIS Geostatistical Analyst* manual. More technical information can be found in Chapters 4 and 6 of the manual. The manual is available in digital form on the lab server and in the readings directory.

The data you will use are ozone measurements for July 1996, obtained at monitoring stations in the Los Angeles basin. A compressed file `labasin96.zip` is available in the labs directory of the course web site. It contains two shape files: a polygon shape file, `labasin.shp`, with the outline of the four counties (Los Angeles, Orange, Riverside and San Bernadino), and a point shape file, `oz96.shp` with the ozone measurements at 32 monitoring stations.

2 Exploratory Data Analysis

Start ArcMap with a layer containing the point feature for the monitoring stations and the polygon feature with the county boundaries. Make sure the Geostatistical Analyst is active, `Tools > Extensions > Geostatistical Analyst`, and that its toolbar is visible, `View > Toolbars`.

2.1 Histogram Analysis

- in the `Geostatistical Analyst` toolbar, go to `Explore Data` and select the item `Histogram` from the drop down menu
- don't be fooled by the first graph, by default it is a histogram for the first variable, in this case `STATION`
- to get a more meaningful histogram, change the `Attribute` to `MAXDAY` and make sure the `Layer` refers to the point layer (`oz96`)
- note how the `Statistics` check box is checked by default: this gives a summary of the data in the upper right corner of the graph; if you uncheck the box, the summary will disappear
- you can fine tune the appearance of the histogram, by changing the number of `Bars` in the graph or by choosing a data transformation to check how that affects the distribution of the variable

- change the number of bars to 11, and experiment with a `log` transformation
- the `Histogram` and map are *linked*. To see this, turn back to the original scale (set `Transformation` to `None`), with 8 bars in the histogram, and select (click, followed by shift-click) the two left-most bars. Note where the corresponding monitors are located on the map.
- using the `Select Features` tool from the toolbar, select the two Northern-most locations in San Bernadino county (stations 5213 and 5181, make sure you have the selectable layer set to `oz96`). Note where they rate on the histogram; how do you think this will this affect the analysis of the semi-variogram later on?
- check on the distribution of `MAXDAY` for the four monitoring stations in Orange County
- as you have done during EDA in GeoDa, you can use the linking facility to find the locations of subsets of the data (according to the matching bar in the histogram) or to view a sub-histogram for selected points

2.2 Assess the Normality of the Data

- in the `Geostatistical Analyst` toolbar, select `Explore Data` and choose `Normal QQPlot` from the drop down list
- make sure to reset the `Attribute` to `MAXDAY` (and **not** to `STATION`)
- if the distribution of the data is normal, the plot will follow the diagonal line (but this is only based on "visual inspection," not on a formal test)
- note how the selected data points are highlighted in the `QQ-plot`
- similar to the histogram, the `QQ-plot` has the facility to link and find the matching points on the plot for subsets of the data

2.3 Practice

Use the `Histogram` and `Normal QQ-Plot` tools to assess the distribution of the `AV8TOP` variable. Alternatively, you can load the `BALTIM` sample data set from the SAL site and explore the distribution of the `Price` variable.

3 Modeling Spatial Trends

One of the fundamental assumptions underlying variogram modeling is the absence of a spatial trend. In other words, the mean value of the process should be constant throughout space. The Geostatistical Analyst contains functionality to assess the presence of a spatial trend by visual inspection. Note that in this exploratory phase, you cannot save a fitted trend surface (any trend surface needed in the variogram model is fit within the "Geostatistical Wizard").

3.1 Trend Analysis

Select `Trend Analysis` from the `Explore Data` drop down list to assess spatial trends in the `MAXDAY` variable.

- make sure to set the `Attribute` correctly
- a three-dimensional graph will appear in the window, with the point locations on the horizontal plane, the values floating in the vertical dimension, spikes connecting the point locations to their values and a projection on the plane to the right and to the back
- check out the various options by unchecking and checking `Sticks` (takes out the spikes), `Projected Data` (takes out the projected data points on the side and back panel), `Input Data Points` (takes out the floating points in 3-D space) and `Trend on Projections` (takes out the projected trends in the Y and X plane)
- note the trend in the pollution measure away from the coast and from North to South
- you can rotate the view, but it takes some practice to realize what is being rotated, it can be either the points (Locations) or the graph itself; the latter gives you a better idea of how the projected lines reflect a trend in the data
- to rotate the graph, select `Graph` from the drop down list and click on the arrows directly to the right of that item (the vertical bar with arrows allows you to rotate the graph in the vertical dimension)
- when you rotate the locations, you can see how the shape of the projected trend changes as you rotate the points (the "angle" of the planes on which you project changes as you move the points around); you can see the green trend change from inverse u-shaped to an almost straight line, for example
- as with the `Histogram` and `QQPlot`, you can select items on the map and see their "spikes" highlighted in the graph (note that the points are not highlighted in the projected trend surface); you can also select spikes in the graph and check their locations on the map

3.2 Practice

Assess the spatial trend in the `AV8TOP` variable, or of the `Price` variable in the `BALTIM` data set.

4 Exploring the Semivariogram

The data exploration can be continued by taking a closer look at the variogram structure. This is implemented in the `Semivariogram/Covariance` item in the `Explore`

Data menu. All the analyses will be for the MAXDAY variable. Make sure to set Attribute to that variable, since the default is to use STATION, which is rather meaningless.

4.1 Creating a Semivariogram

- after invoking Semivariogram/Covariance, keep the Tab on Semivariogram
- the scatterplot-like graph is the *semivariogram cloud plot*, it shows for each distance *bin* (horizontal axis) the squared difference between each *pair* of observations on the vertical axis
- change the Lag Size to 0.2 and the Number of Lags to 20. You will see that the cloud plot does not contain any values beyond a distance of about 2.4 units on the x-axis.
- a good rule of thumb is to constrain the variogram to 1/2 the maximum distance and to make sure at least 30 pairs are in each bin
- change the Number of Lags to 10 and the Lag Size to 0.12 and note how the cloud plot is truncated
- experiment with some other combinations of Lag Size and Number of Lags combinations to get a good sense for what is going on

4.2 Exploratory Variograpy – Linking and Selection

- use the semivariogram settings with Number of Lags to 10 and the Lag Size to 0.12 and select the three points in the upper left corner of the semivariogram
- locate the corresponding *pairs* in the map and note how all three have a common *origin* (station 5213)
- relate this finding to what you saw in the histogram in Section 2.1
- in variogram terms, this means that the squared difference of the MAXDAY variable observed for pairs this distance apart is much higher than for the others, suggesting that station 5213 may be a true outlier, or, that it may represent a *pocket of local nonstationarity*
- select the two top most points in the semi-variogram and assess their relative location; how would you interpret this finding?
- select a point in the map (use the selection tool) and see what happens in the variogram (nothing); select a pair of points and a line is drawn between them as well as a point highlighted in the semi-variogram cloud plot (not always visible if it is in the bottom part)
- check the semivariogram for other possible outliers

4.3 Directional Effects

You can check the semi-variogram for directional effects using the `Semivariogram Surface`. The surface is shown in the lower left corner of the window and contains the average squared difference for all pairs that match a given distance/direction combination. These values are contained in square grid cells. Asymmetries in the surface suggest directional effects.

- activate the directional analysis by checking the `Show Search Direction` check box
- note the angle and parallel lines on the surface plot; these can be manipulated with the cursor to select a directional range
- note how the number of points in the semivariogram is a lot less than in the original plot (for all the data pairs): only those pairs matching the directional criterion are kept
- select some of the points in the semivariogram and check on the map how the lines connecting the pairs are all in the same general angle
- move the direction selection tool, reselect points in the semivariogram and check how the pairs correspond to a different angle
- experiment with some different angles and suggest a directional effect

4.4 Practice

Carry out the exploratory variography for the `AV8TOP` variable and compare the outliers and other distinct patterns (directional effect) to those found for `MAXDAY`. Alternatively, explore the semivariogram structure for the `Price` variable in Baltimore.