11. Kriging

ACE 492 SA - Spatial Analysis
Fall 2003

1 Objectives

The goal of this lab is to further familiarize yourself with ESRI's *Geostatistical Analyst*, extending the exploratory approach covered in the previous lab (Lab 10) to spatial prediction, or *kriging*. Refer to Chapters 4 and 6 in ESRI's *Using ArcGIS Geostatistical Analyst* manual for more extensive technical information.

As in Lab 10, 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 Basics of Spatial Prediction

Kriging, or, spatial interpolation based on the spatial information in the semivariogram is implemented in a *Geostatistical Wizard*. This guides you step-by-step through the process. You can leave all options to their default settings and the Geostatistical Analyst will produce a predicted surface map. However, you should exercise some judgment in the choice of variogram function, data transformations, directional effects, etc.

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`.

In the Geostatistical Analyst toolbar, select `Geostatistical Wizard` and make sure the `Input Data` is set to the point layer (`oz96`) and the `Attribute` to `MAXDAY`. Follow the steps in the Wizard with all the defaults left as they are, including the `Method` as `Kriging`. Start the Wizard by clicking on `Next`.

2.1 Geostatistical Wizard

- **Step 1** in the interactive process is the choice of `Method`; for now, leave the default to `Ordinary Kriging, Prediction Map`, without transformations or trend removal (Ordinary Kriging will take out an estimate for the mean value), click on `Next` to move on

- **Step 2** is `Semivariogram modeling`, based on the same two graphs as in the exploratory part, a semivariogram cloud plot and a surface

- note how a variogram function has been fit through the cloud plot (the yellow line) with as default the `Spherical` model
- note the values for the parameters (`Major Range`, `Partial Sill`, and `Nugget`) and the choices for the `Lag Size` and `Number of Lags` (later, you will change these), move on by clicking on `Next`
- **Step 3** specifies the `Searching Neighborhood`; ignore this for now and click on `Next`
- **Step 4** provides some indication of model fit by systematically dropping an observation from the data set, and refitting the model, or *cross-validation*
- make a note of some of the measures of fit for future comparison
- click `Finish` and a summary of the model parameters will be presented in a window

Complete the process by clicking on `OK` in the last dialog to obtain a predicted surface in ArcMap, covering a bounding rectangle around the sample points. Rearrange the layers so that you can see the original monitor locations on top of the predicted surface.

2.2 Practice

Go through the steps in the Geostatistical Wizard to create a predicted surface for `AV8TOP`, using the default settings. Alternatively, use the `Price` variable in the `BALTIM` data set.

3 Customizing Spatial Prediction

There are several ways in which you can customize the presentation of the predicted surface and the application of various geostatistical methods to obtain the interpolation. Refer back to ESRI's *Using ArcGIS Geostatistical Analyst* manual for specifics.

3.1 Fine Tuning Presentation

There are several ways in which you can customize the presentation of the predicted surface. For example, you can change the extent of the rectangle containing the predicted surface:

- in the `Ordinary Kriging` layer (i.e., the one containing the predicted surface), right click and select `Properties`; select the `Extent` tab
- check out the options in the `Set the extent to` drop down list. For example, choose the rectangular extent of `labasin` (the LA county boundary layer); click on `OK` to see the effect
- note how bad the predictions get as you move out of the core window used for estimation

You can also *clip* the rectangle with the predicted surface such that it matches the irregular outline of another shape. For example, you may want to clip it to match the boundaries of one of the LA Basin counties. First you need to create a separate layer (shape file) that contains only the selected county:

- start by selecting Los Angeles county, using the `Select Features` item from the toolbar and click anywhere in the county to select it
- right click on the `labasin` layer item in the legend and select `Data > Export Data` from the menu
- leave the option to `Export selected features` and enter an appropriate file name for the shape file to be created (e.g., `LAcounty`); click on `OK` to save the new shape file
- reply `Yes` when asked to add the exported data to the map as a layer

To limit the interpolated surface map to the extent of Los Angeles county, you need to change the `Properties` of the `Layers`:

- right click on `Layers`, select `Properties` and click on the `Data Frame` tab
- in the `Data Frame Properties` dialog, focus on the `Clip to Shape` item and check the `Enable` box; specify the LA county layer as the shape to use as a clip
- click `OK`; now the predicted map coincides with the outline of the single county

The interpolated surface can be portrayed in a number of different ways. This is set in the `Symbology` Property of the layer. The default is `Filled Contours`, but you can also choose `Hillshade`, `Contours`, or `Grid`:

- right click on the `Ordinary Kriging` layer and select `Properties`
- select the `Symbology` tab
- uncheck `Filled Contours` and check `Contours` instead; click `OK` to see the effect
- experiment with the other types of symbology for the predicted surface

The predicted surface can be turned into a permanent feature class or shape file by means of the `Data Export` function. This will allow you to then use the predicted surface in other analyses. For example, an exported grid surface (or raster) can be used to compare the predicted values to those obtained with another technique using the ArcGIS Spatial Analyst Extension raster calculator:

- right click on the `Ordinary Kriging` layer; in the `Data` item, select `Export to Raster`
- choose a file name for the output file and click `OK`; add the raster to the layers if you wish
- alternatively, use the `Data > Export to Vector` function to create a shape file with the interpolated contours

3.2 Fine Tuning Analysis

So far, the predicted surface was based on using the defaults in the Wizard. However, this is by no means necessarily the best way to proceed. It is useful to reanalyze the spatial correlation structure in the MAXDAY variable using different variogram models and/or by including trend removal. Also, the fitted model is sensitive to the choice of the number of bins and their range.

Create some new predicted surfaces by using different settings:

- start the Geostatistical Wizard, and make sure the Input Data are set to the point layer (oz96), and the Attribute to MAXDAY
- keep the Methods setting to Kriging and click on Next
- in the following dialog, set the Order of Trend Removal to second and click Next to continue
- the window will show the predicted second order trend surface; click Next again
- in the semivariogram dialog, set the Lag Size to 0.1 and the Number of Lags to 12, and observe the change in the parameter estimates
- run the analysis for a spherical variogram and repeat for an exponential variogram

3.3 Practice

Experiment with model selection and changing lag size for the AV8TOP variable. Present the results in different ways, as filled contours, contour lines, and raster surfaces. Try using various clips to portray the predicted value surface for administrative units. Alternatively, experiment using the Price variable in the BALTIM data set.

4 Assessing and Comparing Model Fit

In order to assess how good the predicted surface is, its fit needs to be quantified. There are basically two methods to accomplish this. In one, an observation point is left out and then predicted using the surface fit for the remaining observations. This is done for each observation in turn. The overall fit can then be summarized as a mean squared error or other indicator. This method is referred to as *cross-validation*. The other approach extends this idea to a specific subset of the data. For example, several observation points are left out of the analysis (the model fitting) and the estimated surface is used to predict their values. Those predictions are then compared to the actual values.

4.1 Cross-Validation

The cross-validation statistics are generated by the Geostatistical Wizard:

- make sure you have at least two predicted surfaces for MAXDAY in your table of contents, such as the ordinary kriging one obtained previously (no trend removal, spherical model), and ordinary kriging using an exponential model with second order trend removed
- right click on one of the predicted surfaces in the table of contents and select Compare...
- a dialog will open with on the left hand side the cross-validation statistics for the current surface, below Compare
- on the right you can select the surface to compare it to in the To: drop down list
- when comparing two predicted surfaces in this way, focus on the mean prediction error (should be close to 0), compare the root mean square (smaller is better) and check the root-mean-square standardized (closest to 1 is best)

To illustrate the second method, first create two new shape files by using the selection tool and Data > Export Data. One file should contain only the four monitors for Orange County, the other should contain all monitors *except* those in Orange County. Now proceed with creating an interpolated surface based on the incomplete data set (all points except those in Orange County). To compare the results:

- in the layer for the predicted surface, right-click and select Validation
- in the dialog, select the layer with the left out points for Orange county as the Input Data
- make sure the attribute is set to MAXDAY and specify a shape file to save the output
- the new layer will be created with the original attributes for the point layer as well as new variables, including Measured, Predicted and Error
- use the Statistics on the Error field to summarize the prediction error (right click on Error and select Statistics, **not** Summarize)

4.2 Practice

Assess the fit of your predicted surface for AV8TOP using the standard cross-validation and leaving out points methods. Compare different models and how their performance may vary for subareas in the data. Note that the Geostatistical Wizard also contains non-geostatistical methods, such as Inverse Distance Weighting and Polynomial interpolation. You can also use cross-validation to assess the effect of the "outliers" identified in the exploratory analysis. Alternatively, carry out a similar exercise for the Price variable in the Baltimore housing data.