GEO 423/EVS 523
Exercise 14: Surfaces and Geostatistics

The field sciences use geospatial data from numerous sources. However, many of these data sources are inherently raster – such as aerial photographs and satellite images; others are inherently vector – such as environmental data from specific sample-site points or transect lines or polygons. It is often necessary to put these vector and raster layers together in a single map that provides information derived from all of the layers. This map represents a model of the system comprising the issues included in the various layers. In addition, it is often useful to assess various layers' statistical attributes to determine how best to maximize the validity of the composite model.

In this exercise, you will convert a point file of measurements of atmospheric ozone concentration across the State of California into a raster surface. You will then use several statistical methods to improve the model of statewide ozone concentration.

Copy the ozone.gdb geodatabase from the P: drive to the X: drive. Make sure that you have turned on both the spatial analyst and the geostatistical analyst extensions are turned on. Add the layers in the geodatabase to your layout. Change the color of the State of California to "no color" so that only the outline remains. Adjust the color ramp of the 03_Sep06_3pm layer's ozone field so that you can see the spatial distribution of ozone across the state. Where are the highest and lowest ozone concentrations?

Add the geostatistical analyst toolbar to your layout; click on the Geostatistical Analyst arrow on the toolbar, and choose the Geostatistical Wizard. Choose the Kriging/CoKriging in the Methods list box. Make sure that the Source Dataset is 03_Sept06_3pm, and the Data Field is OZONE. Click Next to go to the next step. Choose Ordinary Kriging, and make sure that Prediction Map is selected as the output type. Click Next to open the semivariogram/covariance model. The semivariogram is a graph of the difference in ozone-concentration values vs. distances apart among pairs of sample sites. The blue line is the model.

Click Next. The crosshair shows an area with no measured value. The colors of the pixels in the image indicate the relative weighting of the estimation, with red points being weighted more than the green points, since they are closer to the unknown pixel than points that are further away.

Click Next. The cross-validation diagram gives you an idea of how well the model predicts the values at the unknown locations. Click Finish, then OK. The predicted ozone map appears. You will want to save this map. To do that, double-click Layer in the Table-of-Contents window to open the Layer Properties dialog box. On the General tab, change the layer's name to something that suggests that this is the result of default kriging. Click OK. Next, you will restrict the prediction surface to the land area of the state. Right-click the name of the default-kriging model, and choose the Extent tab. Click the Set the Extent arrow, and choose the rectangular extent of ca_outline, and click OK. Right-click the Layers data frame in the Table of Contents, click the Data Frame tab. Click the Clip Options arrow, choose Clip to shape, and click the Specify Shape button. On the Data Frame Clipping dialog box, click the Outline of Features button, click the Lawyer button, and again choose ca_outline. Click OK twice, and adjust the positioning of the various layers so that they show what you want.

It is possible to visualize how well the default layer actually reflects the distribution of ozone concentrations, but it is possible to be more quantitative. Right-click the default-kriging layer, and choose Validation/Prediction. In the GA Layer to Points tool, which opens, the input layer is set to the default-kriging layer. Set the ca_cities layer as the input point observation locations, and give your output statistics layer an appropriate name. When you click OK, the tool runs and generates a point file, whose attribute table includes a predicted ozone concentration and standard error of estimate. Print a map showing as much of the information you've generated so far as you can.

Now, in the Table-of-Contents window, select the 03_Sep06_3pm layer. On the Geostatistical Analyst toolbar, click Geostatistical Analyst -> Explore Data -> Histogram. Choose to explore the Ozone data points (not the Elevation data points). Note that in order to read the histogram, actual values have been multiplied by 10. The skewness of the distribution shows that very high values are less common than very low values. Select the two histogram bars with ozone-concentration values larger than 0.1 ppm. Where are the points with these concentrations?

Now click Geostatistical Analyst -> Explore Data -> Normal QQPlot to generate the quantile-quantile plot. This shows the degree to which the data set corresponds to a standard normal distribution. A standard normal distribution would have a QQ plot lying precisely along the 45° diagonal. If you want to see which data points are most different from the normal distribution (e.g. the low ozone concentrations), you can select points from the QQ plot by selecting the rectangle containing them. The significance of the QQ Plot is that if a dataset does not fit the normal distribution, it may be necessary to transform it in order to use some standard statistical techniques.

In order to determine the trend in the dataset (if a trend exists), we can calculate a mathematical formula for the trend surface. This surface may be only a rough indication of the actual trend, but it may be better than nothing. Click Geostatistical Analyst -> Explore Data -> Trend Analysis, chose 03_Sep06_3pm as the Data Source Layer, and choose Ozone as the attribute to be considered. You may wish to turn some of the details of the resulting graph off and on, and you can turn the graph around using the upper scroll wheel. Note that the most meaningful angles are 0°, 90°, 180°, and 270°. The green line shows the trend along the XZ plane; the blue line shows the trend along the YZ plane. Both are U-shaped, suggesting that a second-order polynomial is an appropriate representation of the trend. In this particular case, the trend may be caused by the fact that large human populations exist in the Central Valley and that they tail off both over the Pacific Ocean and over Nevada.

Now you will create a new map eliminating the trend. As you did before, click Geostatistical Analyst -> Geostatistical Wizard. Choose Kriging/Cokriging in the Methods list box. The input is 03_Sep06_3pm, and the Attribute is Ozone. Click Next, and Orginary Kriging. From the Order of Trend Removal dropdown, choose Second, and click Next. The resulting image shows the global trend in the dataset. In general, one should eliminate a trend only when there is a concrete justification. That fits here, so click Next.

The semivariogram/covariance modeling window opens. Again, the average value for each cell of the semivariogram surface is plotted on the graph as a red point. The average for each bin (encompassing many cells) is shown as the blue cross. The x-axis of the graph is the distance from the center of the cell to the center of the semivariogram surface. For these ozone data, the semivariogram starts slow at short distances – ozone values measured at close-together locations are similar to each other – and increases as distances increase – the ozone values are more different as they get further apart. Earlier, you removed a coarse-scale trend, but there is still a directional component to the autocorrelation that you need to incorporate in your analysis. The Lag Size is a measure of the size of the bin into which data have been put. Change it to 15000 to zoom in, in order to model the details of local spatial variation. Change the Show Search Direction option from False to True. Grab the center blue line in the semivariogram map, and drag the line around the circle. Only the semivariogram surface values within the search are plotted on the graph. See how they change as the search direction changes!

Change the Anisotropy option from false to true. Again, drag the blue line around the resulting ellipse. You can change the search direction angle to 61.35 or 241.35 to make the directional pointer coincide with the ellipse's minor axis or 151.35 or 331.35 to make the pointer coincide with the ellipse's major axis. Note that as you go out on the X axis, the blue line rises fairly continuously until it gets to around 180 km when pointing toward the major axis, but only about 110 km when pointing to the minor axis. Click Next.

The Cross-Validation window opens. This provides information that will enable you to judge the validity of your spatial model. If your model is valid, the various prediction errors in the tables will be near 0, indicating that your predictions are unbiased. The root-mean-square standardized prediction error will be close to 1, indicating that the standard errors are accurate. Finally, in the main table of values, the measured and predicted values are close. If you click the QQPlot tab, the QQ Plot will appear. Is it close to the 45° diagonal? If so, click Finish and then OK. The new model predicting ozone concentrations will be generated. Again, change the extent to the values of ca_outline, as you did previously. Right-click the trend-removed prediction layer, and choose to show the prediction standard error. Note that the prediction standard error is lowest near areas of a lot of data points. Note that if you right-click the trend-removed prediction-standard-error layer, you can choose to change it back to the prediction layer.

Print both the prediction layer and the prediction standard error in a single layout.

You will now compare your two models. Right-click the layer name for one of the models, and choose Compare. The Cross-Validation Comparison box will open. Since there are only two models to compare, they should both be available in the comparison. Compare the statistical information presented to you. Which model is preferable?

The final exercise will be to map the probability of ozone exceeding a critical threshold. For purposes of this exercise, we will assume that 0.09 ppm is the critical value. Click Geostatistical Analyst -> Geostatistical Wizard, and choose Kriging/Cokriging in the Methods list box. The input dataset is 03_Sep06_3pm, and the Attribute is Ozone. Click Next, and choose Indicator Kriging. Note that the output type is Probability Map. Insure that the Threshold is set to Exceed, and set the Primary Threshold to 0.09. Click Next, and set the Lag Size to 15000. Change Anisotropy to True.

Click Next to move to the Searching Neighborhood box and then Next again to move to the Cross-Validation box. The blue line represents the threshold. Points to the left are below that value; points to the right are above. The indicator column indicates which are the former (i.e. = 0) and which are the latter (i.e. = 1). You can click on any datapoint in the main table, and the corresponding point will show up in the indicator-prediction window. Note that because of the standard error, some points near the cutoff threshold may or may not be what they would appear to be. Click Finish to exit the Cross-Validation box and then OK. The probability map will be generated. It shows the probability that the threshold value of 0.09 ppm was exceeded between 3 and 4 PM on 6 September, 2007. Again, change the extents to encompass the entire State of California. Again, adjust the order of the layers so that the basic dataset is on top, the indicator layer is next, and the trend-removed layer is next.

Right-click on the indicator layer, and choose Symbology. Un-check the Filled Contours checkbox, and check the Contours checkbox. This will change the layer from a map to a contour map. Change the classification scheme of the contour map so that it emphasizes the likelihood of exceeding the threshold value. You should now be looking at the trend-removed model with contours showing the degree to which an area is likely to exceed 0.09 ppm ozone concentration and all of the datapoints on which the models are based. Print this image.

Portfolio

13-1 Map showing the distribution of estimated ozone concentrations across the State of California using the default kriging model, along with a table showing estimated ozone concentrations and standard errors of the estimate for at least 6 cities in the State.

13-2 Map showing the distribution of estimated ozone concentrations across the State of California using the kriging model with the trend removed, as well as a map showing the prediction standard error, along with the points showing the sample sites on which these maps are based.

13-3 Map showing the distribution of estimated ozone concentrations across the State of California using the kriging model with the trend removed, overlain by a map showing contour lines indicating the probability of ozone concentrations exceeding the 0.09 ppm threshold, along with the points showing the sample sites on which these maps are based.