James K. Pringle
140.663 Geostatistics
Dr. Curriero
Assignment 3
April 6, 2014

Assignment 3

Questions 5

5 Kriging the EPAs AQS Ozone Data.

This data set corresponds to the EPAs AQS average annual daily 8 hour maximum ozone for 2007. The questions and code provided run through a kriging analysis of this data based on several different approaches. All the data you will need for this problem are in the two zipped folders `Ozone_Monitors_2007_reproj` and `States_reproj_lower48` posted on Courseplus. Unzip these and save them in your R working directory.

Exploratory Spatial Data Analysis of the Ozone Data

a. Produce a map of ozone with symbols signifying magnitude.

b. Produce a 2×2 display of 4 descriptive plots using the `plot(geodata object)` command for the ozone data. This is a large area to consider, so to better see possible spatial trends across the US, plot the data separately versus the x and y coordinates.

c. Estimate and plot the semivariogram of the ozone data using the default binning in `variog`. Actually estimate and plot the semivariogram with and without restricting the distances to be within half the maximum inter-point distance (so estimate two semivariograms). In the future though whenever asked to estimate a semivariogram/variogram always restrict it to be within half the maximum inter-point distance.

d. For the ozone semivariogram based on distances restricted to be within half the maximum inter-point distance use the `eyefit` command and select a semivariogram function and set of parameter estimates that appear to fit the semivariogram well. For Mac users who have problem getting the `eyefit` command to work just eyeball a set of parameter estimates.

e. Following the results from (b) go ahead a fit a large scale spatial trend based on the easting (or x) coordinate, ignoring and trend in the northing (or y) coordinate. The code I provide fits a natural spline of the easting coordinate with 4 degrees of freedom

to try and match the apparent trend. Now using the residuals from this model, estimate and plot the semivariogram (residual semivariogram) and use the `eyefit` command and select a semivariogram function and set of parameter estimates that appear to fit the semivariogram well. Select the same semivariogram function as you did in (d).

f. With the information generated (a) - (e) address the following and reference specific plots in your answers/interpretations.

(i) Does the ozone data appear to be Normally distributed?

(ii) Argue for the existence of a large scale spatial trend in the ozone data.

(iii) Describe the difference in the two estimated semivariograms from (c) and what might be influencing the pattern seen in the semivariogram estimated based on all pairwise distances.

(iv) Specify the spatial regression model (its either ordinary or universal kriging) for what the semivariogram estimated in (c) is for and what the semivariogram estimated in (e) is for. So two models need to be specified. Also for each describe what data the semivariogram is estimating spatial dependence of.

(v) Describe any difference in the fitted semivariogram functions arrived at in (d) and (e). How have the total sills changed and provide an interpretation for this?

Kriging the ozone data

g. Using weighted least squares, fit the semivariogram function from (d) to the ozone data using the initial values selected in (d). Again using weighted least squares fit the semivariogram function from (e) to the residuals of the model used in (e) using the initial values selected.

h. Produce a map of IDW predicted ozone.

i. Produce a map of trend surface model ozone predictions and a map of predicted standard errors. Specify the trend using the natural spline (with 4 degrees of freedom) of the easting coordinate as utilized previously.

j. Produce a map of ordinary kriged ozone predictions and a map of corresponding prediction standard errors.

k. Produce a map of universal kriged ozone predictions and a map of corresponding prediction standard errors. For the trend use the same natural spline on the easting coordinate as in the trend surface model predictions.

l. With the information generated (g) - (k) address the following and reference specific plots in your answers/interpretations.

(i) Write out the statistical regression models used for generating the predictions in (h) through (k). If a statistical model doesnt exist just say so. Level of detail for the written models should be commensurate with that found in the lecture notes.

(ii) For each of the spatial prediction approaches considered (IDW, trend surface, ordinary and universal kriging) describe the behavior of the predictions and prediction standard errors as prediction locations get further away from the sampled data.

(iii) Spend some time studying the difference between the spatial prediction approaches presented with this data. There is nothing to write down or hand in for this, but Im hoping it might generate some questions.