

## ASSIGNMENT 5

(Due on Tuesday, April 14)

In this study you will carry out a **geostatistical kriging** of the England and Wales Temperature data from the month of August, 1981, discussed in B&G (p.147). [A similar data set is studied by Upton and Fingleton (Item 18 in the References on the Class Web Page, pp. 325-331).] For additional information, try Googling topics like “England Climate”, etc.

Since the only data available here is the coordinate data (X,Y) for each measurement location, the spatial trend function will involve only the variables (X,Y,X<sup>2</sup>). A more suitable universal kriging model would of course involve other relevant explanatory variables (altitude, etc.). So the present model should be considered simply for demonstration purposes in your pedagogical study. The data can be viewed in ARCMAP by opening the file **..projects/eng\_temp/eng\_temp.mxd**.

- (1) As in previous assignments, make a subdirectory in your home directory, **S:/home/eng\_temp**, and then copy all files in **..projects/eng\_temp** of the form (**eng\_temp\_81.\***, **mask.\***, **eng\_temp\_bnd.\***). Use these to make a map document similar to **eng\_temp.mxd**. The order of the layers from top to bottom should be as above from left to right. Unlike the South American data in Assignment 4, the true map coordinates here are in “Kilometers” (*km*). So for later purposes, you should reset the displayed map coordinates to “Kilometers” [using the same procedure as in part (b).3.a of Problem 2 in Assignment 4.].
- (2) To kriging this data, first open MATLAB and import the text file **eng\_temp\_81.txt** (which can be found in the directory **..sys502/matlab**). As usual, use **File** → **Import Data**, and check the workspace to be sure that the matrix **eng\_temp\_81** appears.
- (3) The kriging will be performed using the MATLAB program **geo\_krige.m**. Open this program by writing:

» **edit geo\_krige**

Observe that the inputs require the explanatory data **X0** and coordinates **L0** at the measurement points, along with explanatory data **X1** and coordinates **L1** at the points where a kriging prediction is to be made. You will begin with the measurement points, as defined by the matrix **eng\_temp\_81**.

- (a) If you look at the first few rows of this matrix [**» eng\_temp\_81(1:3,:)**] you will see that the coordinates [X,Y] are the first two columns, and the temperature data is the last column. So to construct the matrix of explanatory variables, write:

```
» x = eng_temp_81(:,1);
» y = eng_temp_81(:,2);
```

» **X0 = [x,y,x.^2];**

The third command defines the matrix of variables (X,Y,X<sup>2</sup>) at the measurement points. Note in particular that the period in the term **x.^2** specifies component-wise operations, so that each element of the vector **x** is squared.

(b) The set of locations **L0** is then given simply by

» **L0 = [x,y];**

(c) Before proceeding, define the temperature values **y** at the measurement points by

» **y0 = eng\_temp\_81(:,3);**

(4) Next you will construct a grid of points at which kriging predictions are to be made. If you examine the coordinates on the England map (in ARCMAP) you will see that the boundary is contained in a box with base interval [160,580] and height interval [120,580]. To construct a grid, you will use the MATLAB program, **grid\_form.m**. To see this program open MATLAB and write

» **edit grid\_form**

(a) The **box limits** and **cell size** are required as inputs to this program. To define the box, set **Xmin = 160, Xmax = 580, Ymin = 120, Ymax = 580**.

(b) Also observe from the map that a reasonable **cell size** here is 20x20 *km*. So in the program you will set **Xcell = 20, Ycell = 20**. To construct the grid, write

» **G = grid\_form(Xmin,Xmax,Xcell,Ymin,Ymax,Ycell);**

You will see the upper left corner of the grid displayed, just as a visual consistency check. The matrix **G** contains the grid point coordinates at which the kriging is to be made.

(c) To construct the appropriate inputs **X1** and **L1** for **geo\_krige** write

» **x = G(:,1);**  
» **y = G(:,2);**  
» **X1 = [x,y,x.^2];**  
» **L1 = [x,y];**

(Note that **x** and **y** will automatically be redefined by this operation, so there is no need to clear the old values of **x** and **y** above.)

(5) You are now ready to do the kriging. If the number of explanatory variables is denoted by **k** (= 3), then the usual rule of thumb here is that the kriging for each point should

involve at least  $k+1$  ( $= 4$ ) data points. Observe from the map displayed in ARCMAP that a bandwidth of  $h = 50$  km is big enough to ensure that neighborhoods of this size will contain at least 4 data points for most grid points. [The program automatically chooses the  $k+1$  closest data points when this condition is not met.] So to run the desired krige, write:

» **DAT = geo\_krige(y0,X0,L0,X1,L1,50);**

- (a) When completed, you can view (and save) the vectors of GLS-regression coefficients and corresponding P-Values as a matrix, **M**, by writing (be sure to set **Numerical Format** = “short g” using **File** → **Preferences**):

» **M = DAT{1}**

Using this information, comment on the overall spatial pattern of 1981 temperatures in England, and in particular on the nonlinear effect,  $\mathbf{x} \cdot \mathbf{x}^2$ .

- (b) Also, you can view (and save) the vector, **p**, of estimated variogram parameters by writing:

» **p = DAT{2}**

Using this information, what can you conclude about the nature of *spatial dependence* among the residuals in the GLS regression?

- (c) Recall that the **sill value**,  $s$ , should be an estimate of the true variance of the residuals in this regression. Hence it is of interest to compare this estimate with the standard (OLS) estimate of variance, namely *Mean Square Error (MSE)*:

$$MSE = \frac{1}{n-(k+1)} \sum_{i=1}^n \hat{e}_i^2$$

where  $(\hat{e}_i : i = 1, \dots, n)$  are the OLS regression residuals. To do so, you can construct this quantity in MATLAB as follows:

1. First to construct the appropriate data matrix, now add the *unit vector*, **u**, to the data matrix, **X0**, by writing:

» **u = ones(48,1);**

» **XX0 = [u,X0];**

2. In terms of this augmented data, recall that the OLS estimates of the betas are given by

» **b0 = inv(XX0'\*XX0)\*XX0'\*y0;**

3. Hence the OLS residuals above can be calculated as

$$\gg \mathbf{res} = \mathbf{y0} - \mathbf{XX0} * \mathbf{b0};$$

[so that  $\mathbf{res} = (\hat{e}_i : i = 1, \dots, n)'$ ].

4. One can compute the **MSE** for this regression (with  $k = 3$  explanatory variables) by writing:

$$\gg \mathbf{MSE} = (1/(48 - 4)) * \text{sum}(\mathbf{res}.^2);$$

5. As one final comparison, consider the simple *sample variance* estimate

$$\text{VAR}_n = \frac{1}{n-1} \sum_{i=1}^n (\hat{e}_i - \bar{e})^2, \text{ where } \bar{e} = \frac{1}{n} \sum_{i=1}^n \hat{e}_i$$

6. This can be computed in MATLAB by the single command:

$$\gg \mathbf{VAR} = \text{var}(\mathbf{res});$$

7. How do these **MSE** and **VAR** estimates compare with the sill value? How does this relate to the discussion of variance estimation in Section 4.10 of Part II (Continuous Data Analysis) in the NOTEBOOK?

8. What additional information about this relation is provided by your discussion of *spatial dependence* in part (b) above?

(6) The next objective is to import the Kriging and Std Error results (**DAT{3}** and **DAT{4}**) along with the 528 grid points to ARCMAP. To do so, write:

```
\gg K1 = DAT{3};
\gg S1 = DAT{4};
\gg K(:,1:2) = L1;
\gg K(:,3) = K1;
\gg K(:,4) = S1;
```

The (528x4) matrix **K** now contains all the desired data (**L1** contains the grid points). To save this as a text file, **krig\_data.txt**, in your home directory, write

```
\gg save S:\home\krig_data.txt K -ascii
```

- (7) To be sure that the data is in proper format, it is best to import it to EXCEL.
- (a) In EXCEL you can change the data format to a more visually appropriate form by first selecting all four columns, clicking **Format** → **Cells** → **Number**, and then setting **Decimal places** = 3.
  - (b) Now select the top row and insert a new row by clicking **Insert** → **Rows**.
  - (c) Add column labels, **X, Y, Krige, Std\_Err**, in this new (top) row.
  - (d) To save this data in appropriate form, click **Save as** → **Text(tab delimited)**.
  - (e) Finally, change the extension of this file to **.tab** (which ARCMAP recognizes as a text file in tab format). You are now ready import to import **krg\_data.tab** to ARCMAP.
- (8) In ARCMAP click on the **Add Data** button and add **krg\_data.tab** from your home directory. It will appear as a new layer in the Table of Contents.
- (a) Right click on the layer “krg\_data.tab” and click **Display XY Data**.
  - (b) The **X Field** and **Y Field** should already be set to “X” and “Y”.
  - (c) Click **OK** and the grid of points should now appear on the map and as a new layer “krg\_data.tab Events”.
  - (d) Save as a shape file, **krg\_data.shp** using **Data** → **Export Data**, and add it to the data frame. (You can now remove “krg\_data.tab” and krg\_data.tab Events”.)
  - (e) If you drag this layer to a position just below the “mask” layer then only the points inside the England boundary should be visible.
  - (f) Before proceeding further, it is best to turn off the display of these points.
- (9) Next, you will interpolate these kriged values using a **raster spline**.
- (a) First set the appropriate extent for the display by clicking **Spatial Analyst** → **Options** → **Extent** and setting the extent equal to that of the layer “eng\_temp\_bnd.shp”.

- (b) Then click **Spatial Analyst** → **Interpolate to Raster** → **Spline**.
- (c) In the **Spline** window be sure that **Input points** = “krge\_data”, and set **Z value field** = “Krige”. Also set **Number of points** = 4, and leave other values at defaults. [The small number of points together with the small value of **Weight** (default = 0.1) will yield an interpolation that closely fits the kriged data points.]
- (d) In **Output raster** save the result as a grid file, **krige\_spline**, in your home directory.
- (e) Click **OK**, and a new layer “krige\_spline” will appear in the Table of Contents. To view this result properly, you should drag the layer “krige\_spline” to a position between “mask” and “eng\_temp\_bnd”.
1. Here you will need to increase the “Unique Value” setting in ARCMAP in order to edit your raster outputs. Follow the procedure in Part IV of the NOTEBOOK, Section 1.2.20, on the class web page.
  2. If the image shows colors, proceed to the next step. If the image appears in black-and-white and looks “smooth”, you can alter this by right clicking on “krige\_spline”, and then clicking **Properties** → **Symbology** and changing the setting of **Show** to “Classified”.
  3. You can now choose the interval size (classification number) and color ramp to be whatever you like. (To reverse the color order, right click on any symbol in the **Layer Properties** window, and select **Flip Colors**.)
- (10) Finally, you will construct a contour map of this spline interpolation.
- (a) First click **Spatial Analyst** → **Surface Analysis** → **Contour**.
- (b) In the **Contour** window be sure that **Input surface** = “krige\_spline”.
- (c) Leave all numerical values as defaults, and in **Output features** save this result as a shape file, **krige\_cntrs**, in your home directory.
- (d) Click **OK**, and the result will appear as a new layer “krige\_cntrs”. Drag this layer to a position just above “krige\_spline” in order to view it properly.
- (e) If you turn off the display of “krige\_spline”, you will see that this contour map yields a suitable form for printing the krige result in black-and-white.
- (11) Now, repeat steps (9) and (10) for the standard error values, **Std\_Err**, in layer

“krge\_data”.

- (a) Save the raster spline file as **stderr\_spline** in your home directory.
  - (b) When you display these results using the default classification settings, you will see that the relevant standard deviation values lie in a *narrow range*, so the most of the map has the *same color*.
    1. To improve this representation, on the right-hand side of the **Layer Properties** window, click **Classify...** and you will see that the default classification scheme (usually “Natural Breaks” or “Equal Intervals”) leaves most of the color histogram in the same interval.
    2. Try the “Quantile” representation in this case. You will see that the representation is improved.
    3. Also try setting **Classes** to a higher value.
  - (c) When making contours of **stderr\_spline** use **Contour interval** = 1, and save the results as a shape file, **stderr\_cntrs**.
- (12) Given this construction, you are now ready to do some **ANALYSIS**.
- (a) First, zoom in on the area around the point (338,402), which is just about the center of Liverpool.
    1. Using the Identify tool, with **krige\_spline** displayed, determine the kriged estimate of (Celsius) temperature in Liverpool. (In the “Identify” window you may need to set **Identify from** = <all layers> in order to see the value for **krige\_spline**.)
    2. Now repeat this procedure with **stderr\_spline** displayed and determine the standard error of this kriged estimate.
    3. Using these values, determine a 95% Confidence Interval for the mean 1981 temperature in Liverpool.
    4. Given that Fahrenheit ( $F$ ) is related to Celsius ( $C$ ) by  $F = (9/5) \cdot C + 32$ , determine a 95% confidence interval for Liverpool temperature in Fahrenheit.
    5. The actual mean temperature for August in Liverpool is about  $61^\circ F$ .<sup>1</sup> How does this compare with your result above?

---

<sup>1</sup> This is taken from the *Washington Post Archives*. See the file **Liverpool\_Historical\_Data.pdf** in the **Eng\_Temp** directory.

- (b) To compare this with an *exact* estimate, return to MATLAB and redo the kriging for this single point.
1. Here the only change required is the definition of **X1** and **L1** in (4).(c) above.
  2. Let  $\mathbf{x} = 338$ ,  $\mathbf{y} = 402$ , and again set **L1** = [**x,y**] and **X1** = [**x,y,x^2**] (no period required for the scalar case).
  3. Now repeat the command in (5) above and examine **DAT{3}** and **DAT{4}**.
  4. How might you account for difference between this **standard error** value and the value obtained by the spline interpolation? (HINT: Look at the raster spline interpolation near Liverpool with the “mask” layer turned off.)
- (c) Next, display only the layers “eng\_temp\_81”, “mask”, “krige\_spline”, “eng\_temp\_bnd” and zoom in on the area around the **measurement sample point** (422,380). You will notice that this point is exactly on the line between the kriged points, (420,380) and (440,380).
1. Now compare these two kriged values with the measured value at (422,380). Does anything seem strange to you?
  2. If so, how might you explain this apparent anomaly?