

ASSIGNMENT 3

(1) In this study you will use some of the **Cobalt data** from “Vancouver Island Geochemistry” data set in B&G (pp.150, 202) to carry out essentially the same type of analysis done in class for the Nickel data. This is basically a two-part study. In the first part [sections (a) and (b) below] you will analyze the pattern of *spatial dependencies* among these cobalt values in terms of variograms and their derived covariograms. In the second part [sections (c) and (d) below] you will use the covariances estimated in the first part to carry out *stochastic interpolation* of cobalt values in terms of a *simple kriging model*. To gain further perspective on these statistical methods of *geochemical prospecting*:

- I have included two short references, *Geochemical_Propecting.pdf* and *Geochemistry_in_BC.pdf* in **T:\sys502\arcview\projects\Cobalt_2**. Both were obtained by a Google search. You can find further useful information on Google (including good maps of Vancouver Island).
- In the Reference Materials, I have also included a standard reference on *Geochemical Mapping*, namely Chapter 5 in Howarth, R.J. (ed.) *Statistics and Data Analysis in Geochemical Prospecting*, 1984, Elsevier: Amsterdam. This is intended simply to give you some idea of the many mapping techniques used in geochemical prospecting. (It should be sufficient to just scroll through this material online.)

To view this data, open ARCMAP and then open the class file **T:\sys502\arcview\projects\Cobalt_2\Cobalt_2.mxd**. [Note that the interpolation grids saved in this assignment (and later assignments) can be large. Since space is limited on personal ENIAC accounts, you may need to store these on some form of diskette, depending on your available space.]

- (a) Your first task is to recreate this map document in your home directory so that you can edit it.
1. In “This PC” navigate to **T:\sys502\arcview\projects\Cobalt_2** and copy the entire directory to your home directory, say **S:\home\Cobalt_2**. [Note: At this point you could simply take my version of **Cobalt_2.mxd** and change all the “source” paths to your home directory. Hence the purpose of this exercise is partly to give you experience in constructing such a map document (**.mxd**) file from scratch.]
 2. Now return to ARCMAP and clear **Cobalt_2.mxd** by clicking **File > New**, and in the “New” window, clicking **My Templates > Blank Map > OK**. In this new document click “Add Data” on the toolbar (icon = +). Then, in the “Add Data” window, click on the “Connect to Folder” icon on the Toolbar, and open your home directory, **S:\home\Cobalt_2**. (Ignore any error messages about “spatial references”.)

3. Next add the three shape files (**geo_bnd.shp**, **mask.shp**, **geo_dat.shp**) to the open data frame “Layers”. [Be sure to add them exactly in this order.] They will now appear as Layers (subheadings) in the Table of Contents).
4. Change the name of the data frame to “Cobalt Region” by right clicking on “Layers”, and clicking **Properties > General**. Similarly, rename the layers “geo_bnd” and “geo_dat” as “Vancouver Island” and “Cobalt Data”, respectively, by right clicking on each of their labels. Now save this document in your home directory as **Cobalt_2**. (You might first *rename* the existing copy of this file in your directory as **Cobalt_2_old** – if you do not wish to overwrite it.)
5. You should now see “Vancouver Island” together with the locations of the “Cobalt Data”.
 - a. To display the “mask” properly, double click on its colored square in the Table of contents, and set **Fill Color** = “White”, **Outline Color** = “Black”, and **Outline Width** = “2”.
 - b. You can change the color of the Vancouver map as you like.
 - c. Be sure to *re-save* the file every time you make additions or changes.
6. To display the **Cobalt data** values:¹
 - a. Right click on “Cobalt Data” in the Table of Contents, and click:

Properties > Symbology > Quantities (on the left side)
 - b. In the “Quantities” window (which should be set to “Graduated colors”) set **Value** = “CO”. You can then edit the number of classes or values symbols for Cobalt as you like. [In particular, to edit *all* symbols at once, be sure to have all value ranges selected, and then right click on “Properties for Selected Symbol(s)”.]
 - c. Click **Apply** and then **OK**.
7. To set the proper distance units to meters, right click on the “Cobalt Region” data frame, click **Properties** (which should open to the “General” tab) and in the **Units** box, change **Map** units to “Meters” and then change **Display** units to “Meters”. Click **Apply** and **OK**.
8. Save one more time, and leave this file *open* for later use in the Assignment.

¹ The values shown are obtained from stream measurements. Stream sediment is filtered through a mesh [typically around 100 microns (millionths of an inch)] and the measurements of cobalt concentration are then in terms of parts per million (*ppm*). So a cobalt value of say 20 means “20 *ppm* in filtered stream sediment”.

- (b) Next you will construct a **variogram** (and its derived **covariogram**) for this Cobalt data in MATLAB. The data for this construction is contained in the workspace **cobalt_2.mat**.
1. The matrix, **cobalt_2**, in this workspace contains the coordinates and cobalt values (x_i, y_i, z_i) for the 286 locations in this region.
 2. To examine possible **maximum distances** for the desired variogram, observe first (by an application of the measurement tool in ARCMAP) that a neighborhood of 20000 meters around typical sites (about $D_{max}/4$) appears to be large enough to contain most positive dependencies with other sites. So try a variogram with this maximum distance (using the options structure, **opts**):


```
» opts.maxdist = 20000; var_spher_plot(cobalt_2,opts);
```

 Hit “Enter” again to observe the **derived covariogram**. What does this covariogram suggest about local dependencies among neighboring cobalt sites?
 3. Next, observe (again by using the measurement tool) that the usual “ $h_{max}/2$ ” value of maximum distance in this case is about 40,000 meters. So try this variogram:


```
» opts.maxdist = 40000; var_spher_plot(cobalt_2,opts);
```
 4. What are the differences between these two derived covariograms? Try to explain these differences in terms of the pattern of Cobalt values displayed in ARCMAP.
- (c) Next you will Krige this data in MATLAB using the program **krige_simple.m**. Here we will use the variogram estimate obtained for the 20,000-meter max distance.

- (i) To do so, rerun **var_spher_plot** and define the parameter vector, **p**, as follows:

```
» opts.maxdist = 20000; OUT = var_spher_plot(cobalt_2,opts);
» p = OUT{1};
```

If you now type **» p** you will see that **p** contains the parameter values (r,s,a) displayed as output from **var_spher_plot**. (To view such numerical displays in the most natural format, remember to set the **Numerical Format** to “short g” as described in the “Opening_MATLAB” section of the class notebook.)

- (ii) You will now predict the Cobalt value and standard error of prediction at the

point location (615000, 586500) using a Kriging bandwidth of 5000 meters as follows:

```
» s = [615000, 586500];  
» h = 5000;  
» krigesimple(h,p,cobalt_2,s)
```

By leaving off the semicolon on the last line, you should now see a display of the **predicted cobalt value** at **s** followed by the **standard error of prediction** at **s**. Determine a **95% prediction interval** for Cobalt at location **s**.

(d) Finally you will use the spherical variogram parameters for 20000-meter max distance case to **Krige** the cobalt data using the **Geostatistical Analyst** extension in ARCMAP.

1. Open the map document **Cobalt_2.mxd** created above. Here you will use the **Geostatistical Analyst** extension in ARCMAP, which should be on the Main Menus in the Lab. If you are using ARCMAP elsewhere, you can open this extension by clicking:

Tools > Extensions > Geostatistical Analyst

and can view it by clicking

View > Toolbars > Geostatistical Analyst

2. Before Kriging the Cobalt Data, calculate the **mean Cobalt value** by opening the attribute table for Cobalt Data, right clicking the Cobalt column (**CO**) and selecting “Statistics”.
3. To Krige the cobalt data, click:

Geostatistical Analyst > Geostatistical Wizard

4. In the “Wizard” window, set **Source Dataset** = “Cobalt Data” and **Data Field** = “CO”. For **Methods** select “Kriging/CoKriging”, and click **Next**. (If the “Handling Coincidental Samples” window opens, just click **OK**.)
5. In the **Kriging step 2** window that opens, be sure that the **Kriging Type** = “Simple” and **Output Type** = “Prediction”. Also be *sure* that **Transformation type** = “None”.
 - a. Notice on the right that the **Mean Value** denotes the value used for μ in the Simple Kriging model. (Remember that μ is assumed to be *known*.)

b. Compare this value to the **mean Cobalt value** you calculated above. What can you conclude?

Now leave all else as is and click **Next**.

6. In the **Step 3** window, you will see the variogram plotted. Under **General** set **Variable** = “semivariogram” to see the (semi) variogram. You must make a number of additional changes. Under **Model Nugget** set **Measurement Error** = 0%. Under **Model #1** set **Type** = “Spherical”. To make this variogram comparable to the variogram estimated in MATLAB above, set **Lag Size** = 1000 and **Number of Lags** = 20 (yielding a total **max distance** of $1000 \times 20 = 20000$). [Also under **View Settings** you should set **Show points** = “Binned” to remove “averaging” symbols.]

a. Now compare the *Major Range* and *Nugget* values with the Range and Nugget computed in MATLAB.

b. Also compare the *Sill* (= *Partial Sill* + *Nugget*) value with that computed in MATLAB. Based on the discussion in class, how might you account for the differences in these values?

NOTE: If you wish to export this variogram frame to WORD for your report, proceed as follows:

- With the variogram frame displayed in ARCMAP, hold down the “Alt” key and press the “Print Screen” key (on the upper right hand corner of your keyboard).

- Next open WORD and click **Edit > Paste**. The variogram image should now be displayed in WORD.

- **NOTE:** Many of the key values are in light gray and hard to read (some can be darkened by setting calculations to “False”, but this doesn’t always work.) So *another option* is to just print the (semi)variogram [right click and use copy-paste into WORD]. For everything else, if you right click on the right hand panel and select “copy all”, you can copy-paste all data into WORD, and organize it as you like.

Next, to kriging the Cobalt data, leave all else as is in the **Step 3** (Semivariogram) window, and click **Next**.

7. In the **Step 4** window, we shall modify the neighborhood search to roughly approximate the computational procedure used in MATLAB, which used a Kriging bandwidth of $h = 5000$ and required that at least 3 points be in the interpolation. To match this:

- a. First set **Sector Type** = “1 Sector” (i.e., “O” with *no subdivisions*).
- b. Next set **Copy from Variogram** = “False”, so that you can set **Angle** = 0 and set the **Major semiaxis** and **Minor semiaxis** both equal to 5000 (this will yield a circle of radius 5000).
- c. Finally, set **Maximum Neighbors** = 50 and **Minimum Neighbors** = 3. The large number, 50, will ensure that this is never a constraint (so that all points in the circle will be included).

There are no more settings to be made. So leave all else as is and click **Finish**.

- 8. In the **Method Report** window that opens you will see a summary of the settings you have made. Click **OK**, and the **Kriged Map** should appear with an entry in the Table of Contents labeled as “Kriging”. (You may have to wait a bit before the map displays – depending on the “mood” of the network. Try hitting the “refresh” button in the lower left corner of the map display window. If all else fails, you can also try saving the file and reopening it to register changes.) To display this krige properly on the map, first be sure that the Table of Contents is in “List by Drawing Order” mode (first icon at the top of the Table of Contents). Then drag the “Kriging” layer to a position just above the “Vancouver Island” layer (and uncheck the “Cobalt Data” layer).
- 9. Right click on **Properties** and rename this as “Simple Kriging”. In the “Layer Properties” window click **Symbology**, and be sure that the **Show** window is set to “Filled Contours”. By clicking on this option, you can now change the color ramp and the number of classes for the Krige as you like. (Again, you may have to wait a bit before the changes are displayed.)
- 10. Finally, you will zero in on the same point location, $s = (615000, 586500)$, and determine the Kriged value of Cobalt at this location:
 - a. To zero in on a given point, the best approach is to use the “Go To XY” option on the **Tool** toolbar [icon = XY under a small circle].
 - b. When the “Go To XY” window opens, first click on the “units” arrow (black) and select **meters**. Now write the coordinates of point s in the boxes provided, and select the “Add Point” option from the window menu.
 - c. A black dot will then appear at this location on the map. To alter the appearance of this point (for your study presentation), open the **Draw** toolbar (**Customize > Toolbars > Draw**), and use the “Select Elements” pointer on this toolbar. If you then right click on this dot and select “Properties”, you can edit the shape and color of this dot to obtain a more satisfactory graphic.

- d. With the Kriged surface active, and the “Cobalt Data” layer inactive, zoom in on this point (with the “plus” magnifying glass on the main toolbar) and use the **Identify** icon on the main toolbar to identify the **Kriged value** at (or very close to) this site. Try clicking points just above and below point **s**, and averaging these. Record your result, and compare it with the predicted value obtained in MATLAB above.
- e. By making the “Cobalt Data” layer active, you will see six data points surrounding this site. Select these six points (using the **Select Features** icon with the shift key held down), open the Attribute Table for “Cobalt Data”, right click on the **CO** column and select “Statistics”. By examining these statistics, comment on whether the Kriged value seems reasonable in relation to these data points. (Remember that Kriging is basically a method of *interpolation*.)
- f. Finally, to obtain the standard error of prediction in ARCMAP, right click on the “Simple Kriging” layer and select **Method Properties**.
- (i) You will now see that the **Geostatistical Wizard** again opens in the **Step 1** window, and contains all the settings that you have used for this Simple Krige. (This also shows you that the Simple Kriging layer is “live” in the sense that it is still linked to the **Geostatistical Wizard**).
- (ii) Now reselect **Output Type** = “Prediction Standard Error”. This will change the displayed map to show standard errors of prediction at each point. Proceed to **Step 3** of the procedure (“Searching Neighborhood”) and stop.
- (iii) This time you will use an alternative method for checking the prediction-error value at point **s**. To do so, observe that one can choose the coordinates of the test point shown in the window. Hence set **X = 615000** and **Y = 586500**, so that point **s** will be displayed, and use the same settings as above. You will now see that the *exact* prediction-error value is shown. You will also see the six neighbors of point **s** that have been used in the Kriging prediction. Now record this **standard error of prediction value**, and compare it with the value obtained in MATLAB above.

NOTE: At this point you may wish to go back and obtain the exact Krige value using the same procedure. This should agree very closely with the approximate value you obtained (by averaging). This second method is obviously superior to the approximation used above. However, that method can be used for evaluating *any raster data* at specific locations. The present method is only possible inside Geostatistical Analyst.

(2) In this second study you will fit a trend surface to the **mean annual temperatures** from the data set in B&G (p.151, 202) on “South American Climate”. This data set is displayed in the ARCMAP file, `..\S_amer\s_amer.mxd`. The data for analysis is contained in the JMP file, `S_amer.jmp`. Additional background materials can be found by Googling topics like “South American Climate”. In particular, you might want to find maps of attributes (like elevation, vegetation, and prevailing winds) that must surely influence the pattern of mean temperatures in South America.

a. Here you will start with a regression analysis to fit simple spatial models to the South American temperature data. First open the file `S_amer.jmp` and save it to your home directory so that you can save changes as you make them. Now open your copy of the file, and use **Analyze > Fit Model** to regress temperature, **Temp**, on the coordinate variables (**X,Y**). [These coordinates are in kilometers, so that in particular the positive and negative **Y** values denote distances above and below the equator.]

i. Examine the residuals of this regression using the **Residual by Predicted Plot**, which can be found under

Response Temp > Row Diagnostics.

What can you conclude about the nature of the fit from this plot?

ii. Now save the predicted values of this regression, which can be found under:

Response Temp > Save Columns > Predicted Values

A new column will appear in the data set. Return to the main data set (leaving the regression window open) and label the new data column as **Pred_1**.

iii. Next, look for possible **spatial autocorrelation** among these residuals, using the following procedure:

1. Return to the regression window again and save the regression residuals by the same procedure (**Save Columns > Residuals**). Rename this column as **Res_1**.

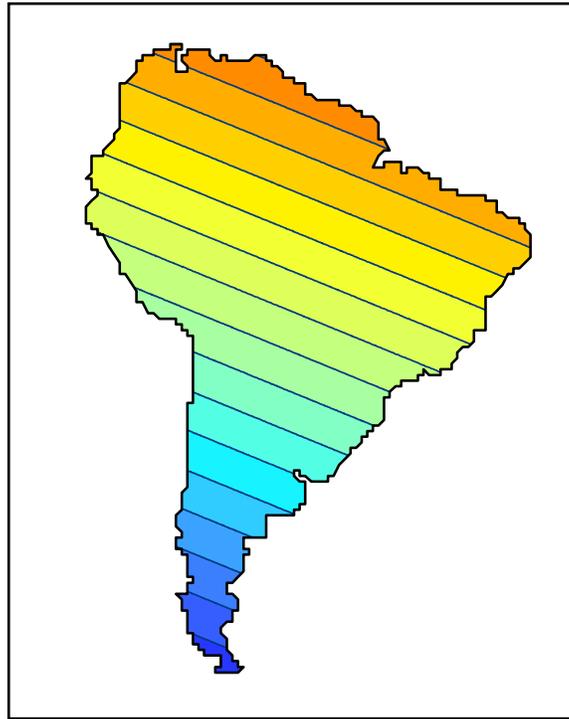
2. Now using the nearest neighbors in the column **nn**, construct the nearest-neighbor residuals [by exactly the same procedure as for the nearest-neighbor “Time” values in Problem 1 [Part (b).3] in Assignment 1]. Label this column as **nn_Res_1**. (The first value should be **4.605**.)

3. Using **Analyze → Fit Y by X**, regress **Res_1** on **nn_Res_1**. What can you conclude about the presence of spatial autocorrelation from the *P-value* of the **nn_Res_1** coefficient?

- b. Based on your observations in part (a).1 above, now do a general quadratic regression of temperature, **Temp**, on the coordinate variables (**X**, **Y**, **XY**, **XX**, **YY**), [already constructed in the file].
- c. Given the results of this quadratic regression:
- i. Observe first that both **X** and **XX** are insignificant. Note the *adjusted R-square* value for this regression. Now remove these two variables, and rerun the regression using only (**Y**, **XY**, **YY**). How does the new *adjusted R-square* value help to justify the removal of **X** and **XX**?
 - ii. What do the *Signs* and *P-values* of the coefficients tell you about this quadratic specification of the temperature trend?
 - iii. What additional information is provided by the new **Residual by Predicted Plot** ?
 - iv. Now repeat step (a).2 above, relabel the new column predicted values as **Pred_2**, and then repeat step (a).3. Name the residuals column as **Res_2** and the nearest-neighbor residuals as **nn_Res_2**.
 - v. What does the new residual regression of **Res_2** on **nn_Res_2** tell you about the presence of **spatial autocorrelation** in the quadratic case? [Be sure to *save* all these results, since you will use them again in the next assignment.]
- d. Finally, you will construct smoothed representations of these trend surfaces in ARCMAP. (These representations are extremely useful for visualizing patterns of spatial data values.)
- i. Start by following the steps in Part (a) of Problem 1 above to make a map document, say **S:\home\s_amer.mxd** in your home directory using the three shape files (**s_am_data.shp**, **mask.shp**, **s_am_bnd.shp**), and be sure that this is the ordering of layers from top to bottom. (If not, check to see that the “List by drawing order” icon is selected at the top of the Table of Contents, and then drag layers to reorder them.) Rename the layer “s_am_data” as “Temperature Data”, and the layer “s_am_bnd” as “South America”. [NOTE: If you want to use all my graphic formats in **s_amer.mxd**, you can copy this file to your home directory and then change the path for each layer to your home directory by right clicking on each layer and using **Properties > Source > Set Data Source.**]
 - ii. Next create a data file in JMP for export to ARCMAP:
 1. First resave **S_amer.jmp** as a new file, **S_amer_Regr_dat.jmp**. In this new file, delete all columns except (**X**, **Y**, **Pred_1**, **Pred_2**), and resave with only these four columns.

2. To export this file to ARCMAP, resave it in EXCEL. (**NOTE:** Depending on the version of ARCMAP you have, you may need to save with extension .xls or .csv, since there appear to be bugs in some ARCMAP versions with respect to its treatment of EXCEL extensions.)
- iii. Now you will create displays of **Pred_1** and **Pred_2** in ARCMAP.
1. Open **S:\home\s_amer.mxd** in ARCMAP and make two new data frames (**Insert > Data Frame**) labeled respectively as “Linear Regression” and “Quadratic Regression”.
 2. Copy the layers “Temperature Data”, “mask”, and “South America” into each of these data frames [right click on each layer, click **Copy**, then right click on each data frame, and click **Paste Layer**]. These layers will now appear in each data frame.
 3. Activate the data frame “Linear Regression” and add the data **S_amer_Regr_dat.xlsx** to it using: **File > Add Data**. The file will appear as layer “S_amer_Regr_dat”.
 4. To display this data, right click on the layer “S_amer_Regr_dat” and click **Display XY Data...** . If you have labeled columns correctly, the choices for **X Field** and **Y Field** should already be displayed as “X” and “Y”, respectively. Click **OK**, and the data will be displayed. It will also be added to the data frame as layer “S_amer_Regr_dat Events”.
 5. To transform this display into a permanent **shape file** (which is more useful for most purposes), again right click on the new layer “S_amer_Regr_dat Events”, click **Data > Export Data**, and save the new shape file in your home directory as **Regr_dat.shp**. (**NOTE:** In some applications, shape file names must be limited to at most 13 characters.)
 6. Add this shape file to the “Linear Regression” data frame, and label the new layer as “Regression Data”. (You can now remove both of the layers “Regr_dat.tab” and “Regr_dat.tab Events” since they will not be needed further.)
- e. Next you will construct an **interpolation** of the linear regression data for display using a **raster spline** in Spatial Analyst.
- i. First, be sure that Spatial Analyst is open. In the main menu click **Customize > Extensions > Spatial Analyst** and then click **Customize > Toolbars > Spatial Analyst** (be sure to scroll to the end of the options).

- ii. Now open **ArcToolbox** and click **Spatial Analyst Tools > Interpolation > Spline**.
 - iii. In the **Spline** window be sure that **Input points** = “Regressions”, and set **value field** = “Pred_1”.
 - iv. It is also important to set a proper extent for the interpolation. At the bottom of the window click **Environments > Processing Extent**. Then set **Extent** = “Same as layer s_am_bnd”. This will cover all of South America.
 - v. Click **OK**, and a new layer “Spline_shp” will appear in the Table of Contents. To *view* this result properly, you should drag the layer “lin_spline” to a position *between* “mask” and “s_am_bnd”.
 - vi. To make this into a permanent file click **Data > Export Data**. At the bottom of the “Export Raster Data” window that opens, set **Name** = “lin_spline”, click **Save**, and export to the map as a layer.
 - vii. Right click on the new layer that appears, and then select **Properties > Symbology**. In the “Layer Properties” window that opens, be sure to set **Show** = “Classified”.
 - viii. In the “Classification” window set **Method** = “Equal Interval” and **Classes** = 20. Finally choose a color ramp that appropriately represents “cold” to “hot” colors. (To reverse the color order, right click on any symbol in the **Layer Properties** window, and select **Flip Colors**.)
- f. Finally, you will construct a **contour map** of this spline interpolation.
- i. First click **ArcToolbox** → **Spatial Analyst Tools** → **Surface** → **Contour**.
 - ii. In the **Contour** window be sure that **Input surface** = “lin_spline”.
 - iii. If the first interval of values in the color chart for “lin_spline” is denoted by values **v1** to **v2**, then set **Base Contour** = **v1** and set **Contour Interval** = **v2 – v1**. [This will yield contours that exactly agree with the equal-interval color bands in lin_spline.] Leave all other numerical values as defaults, and in **Output features** save this result as a shape file, **lin_cntrs**, in your home directory. (Be sure that **Save as Type** = “Shapefile”).
 - iv. Click **OK**, and the result will appear as a new layer “lin_cntrs”. Drag this layer to a position just above “lin_spline” in order to view it properly. The final result should look something like the figure below



- v. If you turn off the display of “lin_spline”, you will see that this contour map yields a suitable form for printing the interpolation result in black-and-white.
- g. Finally, repeat the same procedure in parts d and e above for the Quadratic Regression data, **Pred_2**. Open the “Quadratic Regression” data frame, and follow all steps above with the following changes:
 - i. In part d (iii), set **Z value field** = “Pred_2”.
 - ii. Elsewhere replace “Linear” with “Quadratic”.
- h. Given these smoothed representations of the regressions, your final task is to compare the quality of these two fits to the temperature data, both visually and quantitatively.
 - i. First activate the “Linear Regression” data frame and (with “South America” and “mask” on) display only the “Temperature Data” and “Linear Regression Interpolation”.
 - ii. Now copy this to WORD, using the options in problem (1).(a).6 of Assignment 1 to size and position the image. Leave room for a second image to the right of this image.

- iii. Next activate the “Quadratic Regression” data frame, and repeat this procedure (with “Quadratic Regression Interpolation” replacing “Linear Regression Interpolation”).
- iv. Setting this image to the right of the “Linear Regression Interpolation” image, you can now make a visual comparison of their relative fits. In particular, by looking at the Northern tip of South America, what can you say about the relative quality of the fits for these two regressions? [Notice that the color values don’t quite match up between these interpolations and the original data. But in ARCMAP this is rather difficult to control in simple way.]
- v. Finally, return to the “Quadratic Regression” data frame and be sure that the “Quadratic Regression Interpolation” layer is visible.
 1. First open the Attribute Table for “Temperature Data”, select the row with **FID = 49** (Georgetown, Guyana) and record the Temperature value (Temp).
 2. Next, letting the corresponding point (appearing in red) on the map be denoted by **P**, move the mouse to any location close to **P** and click on this location with the “Identify” tool. This will give you the approximate temperature value (Pixel value) predicted at **P** by the Quadratic Regression.
- vi. Next activate the “Linear Regression” data frame with the “Linear Regression Interpolation” layer visible, and repeat steps v.(1,2) for this case.
- vii. You should now have values for the measured temperature at **P** together with the approximate values predicted by both the Linear and Quadratic Regressions. What does a comparison of these quantitative values add to your visual comparison above? [Again, the relative color values are a bit deceptive here. So it is the *numerical* values that count.]