University of Massachusetts Boston – Department of Environmental, Earth, and Ocean Sciences

EEOS 383 - GISCIENCE FOR WATER RESOURCES RESEARCH – Spring 2010

Instructor: David E. Tenenbaum

Exercise 12: Using a Simple Linear Model to Compare TVDI and API

Introduction

In our previous exercise, we derived a per-pixel description of moisture condition (TVDI) by building a 2-dimensional distribution from surface temperature and vegetation index information derived from remotely-sensed imagery, fitted lines to define the wet and dry lines of the triangular T_s-VI distribution, used an algebraic approach to calculate a given pixel's position in the distribution with respect to the dry and wet lines, and mapped the values back to geographic (as opposed to parameter) space.

In this exercise, we will investigate the extent to which the pattern of surface moisture expressed by TVDI can be explained in terms of patterns of precipitation on the period leading up to the date when the remotely-sensed imagery used to create the TVDI GRID was collected. We will attempt to fit a simple linear model, where TVDI is modeled using an antecedent precipitation index (API) derived from Stage IV radar rainfall data collected by the NEXRAD system of Doppler radars, and then will examine the characteristics of said model and its resulting predictions to learn what we can about the determinants of TVDI in a mixed land-use landscape.

Data

The exercises in the remainder of the course will use remotely-sensed data in the vicinity of North Carolina Climate Division 3. This geographic unit is an agglomeration of 13 counties in north-central North Carolina, which includes the cities and towns of Durham, Chapel Hill, Greensboro, Burlington, and Winston-Salem. This particular exercise will make use of temperature vegetation dryness index (TVDI) data derived from land surface temperature (LST) and normalized difference vegetation index (NDVI) datasets collected by the MODIS Terra sensor on May 24, 2002, along with the NEXRAD based API for the same location, and a specially formatted shapefile to facilitate the extraction of values for use in Microsoft Excel:

- apiids – This shapefile contains square shapes for each of the API pixels in the API GRID for North Carolina Climate Division 3
- api – This GRID contains antecedent precipitation index values in depth equivalent millimetres, derived from NEXRAD gauge corrected precipitation data for North Carolina Climate Division 3
- tvdi – This GRID contains temperature vegetation dryness index values

Procedure and Questions

Downsampling the TVDI GRID

- The first problem we need to tackle in this analysis is to deal with a scale mismatch between the TVDI data and the API data. As you know, the TVDI data is derived from 1km resolution MODIS imagery, yielding 14931 cells for North Carolina Climate Division 3. Unfortunately, the Stage IV NEXRAD gauge-corrected precipitation product (which is the source of our API GRID) has a larger cell size, with roughly 4 km resolution, yielding 850 cells for North Carolina Climate Division 3. To make a reasonable comparison between TVDI and API, we must transform the TVDI into a form more similar to that of the API by downscaling its resolution to match. We can accomplish this using the *Spatial Analyst* (specifically via the *Raster Calculator*) to resample the GRID.
- In the *Spatial Analyst* pull-down menu, under *Options*, use the *Extent* and *Cell Size* tabs to set the analysis characteristics to be equal to those of the API GRID.
- Now, use the *Raster Calculator* to calculate a new GRID that is identical to the original tvdi GRID, except that it will have the extent and cell size characteristics you specified in the previous step (identical to that of the API GRID). Do this by simply calculating a new GRID using the expression [tvdi]. You may wish to use *Make Permanent* to save this new GRID to your H:\ space.

Question 1 – Take a moment to compare the minimum and maximum values of your new, downsampled TVDI GRID with those of the original GRID. Why are these values different / how are the values in the new, downsampled TVDI GRID calculated?

Question 2 – Take a moment to compare the spatial patterns of your new, downsampled TVDI GRID with that of the original GRID. In what respects do they resemble one another, and how do they differ? To supplement your descriptive answer, include a printed map of your new, downsampled TVDI GRID with your exercise, and make sure it has the following characteristics: It should have an appropriate title, north arrow, scale, and legend (be sure to use the Layout View for this).

Collecting the TVDI and API Values

- As was the case in the two previous exercises, we will need to make use of a combination of capabilities of ArcGIS and Excel to accomplish the task. We will once again need to manipulate the values of individual pixels.
- The first capability we must take advantage of is ArcGIS' integration of the vector and raster data models. While we cannot work with the values of individual pixels using *Spatial Analyst*, we can make use of a specially designed shapefile to collect the values of individual pixels in a table. The provided shapefile apiids.shp is such a shapefile: It contains square polygons that each has the same extent as one of the API pixels in the North Carolina Climate Division 3 imagery (there are a total of 850 of these)

- We will use *Spatial Analyst's Zonal Statistics* function (found in the *Spatial Analyst* menu) to collect the required values. The *Zone dataset* will be apiids shapefile, the *Zone field* will be <Value>, the *Value raster* will be the api GRID (at first, we will do this again with the downsampled tvdi GRID), leave *Ignore NoData in Calculations* checked, uncheck *Chart statistic*, and choose an *Output table* location in your H:\ space. Once you click *OK*, the table will be created, opened in ArcGIS, and saved as a .dbf where you specified.
- Repeat the above procedure for the tvdi GRID. You should now have two new .dbf files at specified locations in your H:\ space. They each will contain 850 records. We are now ready to begin working with these two tables in Microsoft Excel.
- Open the api table in Excel. Delete all the fields except VALUE and MEAN. Rename the MEAN field to API. Reduce the number of decimal places shown for the API field to 4.
- Open the tvdi table in Excel. Copy the MEAN field to the api table. Once there, rename it to TVDI, and reduce the number of decimal places shown for the TVDI field to 4.
- You should now have a worksheet that contains 3 fields: Column A should contain VALUE, Column B should contain API, and Column C should contain TVDI. Save this as an .xls file in your H:\ space.
- Create a scatterplot of TVDI vs. API (the convention is to express this in terms of y vs. x, so TVDI should be on the y-axis and API on the x-axis).

Question 3– *What is the shape of the scatterplot formed by this dataset of API and TVDI? Does it look promising for successfully using a linear model to relate these two sets of values? To supplement your descriptive answer, print out and hand in the plot of* TVDI vs. API *(use a scatterplot as described above, make sure the API values are on the x-axis, the TVDI values on the y-axis, with both axes properly labelled).*

Question 4 – *What important characteristic of the TVDI values' range is noticeable in your plot? Do you think that modifying this characteristic would have a significant impact on success in our efforts to use a linear model to analyze this dataset?*

<u>Fitting and Interpreting a Linear Model of API and TVDI</u>

- We are now ready to make use of a simple linear mode to see the extent to which TVDI can be explained by API.
- Click on the *Tools* menu in Excel, and select the *Data Analysis* menu item (if you do not see a menu item entitled *Data Analysis*, click on the *Add-Ins* menu item then check off *Analysis ToolPak – VBA*, and click the *OK* button. Click on the *Tools* menu again, and *Data Analysis* should now appear, and select it).
- Choose the *Regression* Analysis Tool and click the *Ok* button.

- In the *Regression* Dialog that appears, fill in the following values: The *Input Y Range* will be *$C$1:$C$851* and the *Input X Range* will be *$B$1:$B$851* (you can type these out, or use the interface to fill them in instead). Check off the *Labels* item. Leave *New Worksheet Ply* as the selected *Output Option*, naming the worksheet *Linear Model* in the text box provided. Check off the *Residuals* option and click the *OK* button.
- This should result in the creation of a new worksheet entitled *Linear Model* containing two sections: *SUMMARY OUTPUT* should describe the resulting linear model, and *RESIDUAL OUTPUT* should contain the *Predicted TVDI* values and *Residuals* from the model.

Question 5 – What is the Adjusted R Square of the model? What does this tell us about the extent to which we can explain the variation of values of TVDI in terms of API?

<u>Mapping the Residuals Back to Geographic Space</u>

- Make a copy of your original worksheet containing the VALUE, API, and TVDI columns. Then, copy the *Predicted TVDI* values and *Residuals* from the model worksheet to your new worksheet. Rename those two columns two MODEL and RESID respectively, and reduce both to having 4 decimal places.
- Now, use the *Save As* feature to save the worksheet containing the five columns to a comma-delimited (CSV) text file (this a file format that ArcGIS can read easily) in your H:\ space.
- Add the CSV text file to the map document using *Add Data*.
- We will now perform a relational join between the apiids.shp shapefile and the CSV file containing the residual values. Inside the *Data Management Tools* toolbox, you should find *Joins*, and within that sub-section, *Add Join*. The *Layer Name* will be apiids.shp, the *Input Join Field* will be the unique identifier (FID), the *Join Table* will be the CSV file, and the *Output Join Field* will again be the identifier (VALUE).
- Open the attribute table for apiids.shp. You should now find the columns from the CSV file have been joined to it, including the RESID values.
- In the *Spatial Analyst* pull-down menu, use *Convert*, *Feature to Raster* to create a residuals GRID. The *Input features* are apiids.shp, the *Field* is RESID, the *Output cell size* should remain at 4192.15907, and select an appropriate *Output raster* location in your H:\ space.

Question 6 – Can you see any spatial pattern in the residuals that suggests a relationship between TVDI and a spatially distributed phenomenon other than API? To supplement your descriptive answer, include a printed map of your residuals GRID with your exercise, and make sure it has the following characteristics: It should have an appropriate title, north arrow, scale, and legend (be sure to use the Layout View for this).

Question 7 – Create a histogram of the residuals and describe their (aspatial) distribution. What does this suggest about our linear model?