ANTH 583 – GIS and Imagery Analysis

Project 3

This project is designed for you to demonstrate <u>Predictive Modeling.</u> You will employ the following GIS skills:
1) Cumulative viewshed analysis
2) Statistical analysis
3) Use of the map calculator

Background

We will work with the same Wadi Hasa GRASS dataset as for Project 2, and will build on the maps and analyses you conducted for that project. You should make a new mapset for Project 3 to keep your analyses organized. You will be producing several new raster maps for this project, and good organizational skills will be essential. Follow the "best practices" for naming your raster maps, as we have discussed in the past. Use descriptive names that tell you what information is contained in the raster map. Do not use spaces or dashes ("-") in map names. You may use underscores ("_") instead.

While you are doing this project, you may wish to refer to Ullah (2015) *Integrating Older Survey Data into Modern Research Paradigms Identifying and Correcting Spatial Error in "Legacy" Datasets* (available on the course Blackboard site), which discusses a very similar workflow.

Part 1: Cumulative viewshed analysis

1) Use *g.extension* to install the *r.viewshed.cva* GRASS addon module (written by me). Launch it from the Modules tab of the Layer Manager (under "Addons"), or by typing "r.viewshed.cva &" into the terminal (command prompt), and pressing enter.

2) Use *v.random* to create a vector map of 100 random points across the Wadi Hasa region (pro tip: make sure your computational region extents are set to match the area you want these points to appear in!)

3) Use your random points map as the input viewpoints for *r.viewshed.cva*. This will take a while to run, and will produce a cumulative viewshed map of the Wadi Hasa area that estimates the "true" visibility of landscape features in the area.

4) Create an output graphic showing the result of your randomized cumulative viewshed analysis with all the archaeological sites overlain (include all necessary map elements).

Part 2: Statistical Analysis

1) Choose a subset of sites from one of the surveys that you wish to use as a basis for the predictive model (i.e. all WHS sites that are of type "LITH_SCAT" and time period of "PPN"). It is important to only select these sites from <u>ONE</u> of the two surveys (i.e., from either WHS <u>or</u> WHNBS, but not from both). Extract these to a new vector points file using the same procedure we used in Project 2. These will be your <u>training sites.</u>

2) Choose the the kind of data you want to include in your predictive model. Your model should include the walking distance from the streams you made in Project 2, and the cumulative viewshed map you produced in Part 2, above. You should choose at least <u>2</u> additional factors, for a total of at least <u>4</u> factors to include in the predictive model. The choice is up to you, but you should be able to justify your choices. Possibilities could be slope, aspect, elevation, flow accumulation, etc.

3) Using *v.what.rast*, upload values from your 4 chosen raster maps (stream distance, cumulative viewshed, and your other two) into the table of this vector points file of selected sites.

4) Using *v.univar*, find the mean and standard deviation of these four variables in your vector's table.

Part 3: Use the Map Calculator to produce a predictive model

1) Consider the kinds of values in your input maps. How do you want to use these to create your predictive model? Do you want to include values *above* a threshold, or *below* it? Or, do you want to include values *between* an upper and a lower value? Once you decide where your thresholds should be for each variable, you need to come up with a *boolean formula* to enact the thresholding.

2) One way to determine thresholds is to create a histogram of values and to interactively select values. A more automated way is to use the *standard deviation* to create a threshold relative to the *mean.* For an example, if you would like to include values *below* a certain threshold, then *add* the *standard deviation* to the *mean* to derive your threshold. If you want to include values *above* a certain threshold, then *subtract* the *standard deviation* from the *mean* to determine an adequate threshold. If you want to include values *between* an upper and lower threshold, then you have to both subtract *and* add. Keep a note of these calculated values, and whether they are *upper or lower thresholds*. These will be your cutoff points for the locational model you will build in the next steps.

2) Using the map calculator, create a *conditional statement* (boolean formula) that creates an output map coded "1" for positive, and "0" for negative results. use the correct *logical operator* for the type of cutoff. To keep values *below* the threshold, use "<=". It will take the form of:

```
if("input_map" <= "threshold", 1, 0)
```

To keep values *above* the threshold, use ">=". It will take the form of:

```
if("input_map" >= "threshold", 1, 0)
```

To keep values between an upper and a lower boundary, you will need to add an "and" logical operator, which is "&&". Use the following formula:

```
if("input_map" >= "lower threshold" && "input_map" <= "upper threshold", 1, 0)
```

Substitute the real map names for 'input_map' and the real threshold values (from Part 3) for 'threshold.' These new maps are what are called *binary maps* (they contain values of only 1 or 0).

3) Create your *predictive model* by combining your *binary maps* with an *averaging formula*. This mapcalc statement will take the form of:

```
("binary_map1" + "binary_map2" + "binary_map3" + "binary_map4") / 4.0
```

Again, substitute the real map names. Also, *be sure to include the decimal point in the denominator*. This will produce a floating point map with decimal values ranging between 0.0 and 1.0. Areas that have values close to 1 are very likely to have sites, and areas that have values close to 0 are very unlikely to contain sites. If you do not include the decimal point in the denominator, GRASS will calculate an *integer* map, which will therefore remain *binary* (0 or 1).

Part 4: Model Validation

1) Extract a second subset of sites from the *other* survey to test your model against. (i.e., if you had originally extracted WHS sites of type "LITH_SCAT," you now want to extract the WHNBS sites of type "LITH_SCAT" in this step). Again, extract these as a new vector map. These are your *validation sites*.

2) Use *v.what.rast* to upload the values from your predictive model map into the table of this new vector file.

3) Use *v.univar* to calculate the mean and standard deviation in this new column.

4) Create an output map showing your predictive model with the both the *training sites* and the *validation sites* overlain. Make sure the two vector points maps are styled differently from each other. Include all necessary map elements (scale bar, raster legend, vector legend, north arrow).

Part 5: Writeup

Prepare a brief report (1-2 pages) about your analysis. This should be roughly divided into four sections.

1) In the first part, describe the basic patterning visible on your cumulative viewshed map. In general, how much of the Wadi Hasa area seems easily viewable from most areas? Is there any apparent relationship between site locations and more or less easily viewed areas? Why or why not?

2) In the second part, discuss how you set up your predictive model. Which subset of sites did you choose to use for you *training sites*? Why? What other two variables did you choose to include in the analysis (in addition to the cumulative viewshed and the distance to streams)? What do you expect this combination of variables will tell us about why these sites are located where they are (in other words, why did you choose them? What is your expectation for the model?).

3) In the third part, report your findings from the validation procedure. On average, does your predictive model seem to also work for your *validation sites* (values uploaded to the table are all close to 1)? If so, why might that be (how are your training and validation sites related)? If not, why not?

4) In the final part, briefly reflect on the technical procedures you conducted above. How might error propagate through the analysis, and how might that affect the outcome? What other types of analyses would you have liked to do to extend what you did here?

You will turn the writeup in in digital format through TurnItIn on Blackboard. You should include your output graphics as a figures embedded in the writeup.