Lab 10: Raster Analyses

What You’ll Learn: Spatial analysis and modeling with raster data. You will estimate the access costs for all points on a landscape, based on slope and distance to roads. You’ll then apply a threshold to this access cost. There is a mix of old and new functions used in this lab. We’ll explain the new, but you are expected to review old labs if needed to apply the old.

You should read chapter 10 in the GIS Fundamentals textbook before performing this lab.

Data are located in the \L10 subdirectory, including Valley3 and Valley9, three and 9 meter DEMs of a portion of southeastern Minnesota, mar_rdl83.shp, a vector road layer, and mardin, a 30 meter resolution raster elevation grid, and Shasta, a 30m DEM in northern California. All data are in NAD83 UTM coordinates, Z units in meters, the Minnesota files in zone 15, and the California files in zone 11. These instructions were developed with QGIS version 2.6.1, some of the menus are slightly different for QGIS 2.4 and earlier.

What You’ll Produce: Three maps. First you’ll create a mosaic of the valley DEMs. Then you’ll fix the Shasta DEM, and produce a hillshade map. Then you’ll create the second map, a cost surface with an applied threshold, and a third map, some of the intermediate data used for the cost surface.

Project 1: Using the SAGA Grid Calculator, DEM Combination
We often need to conduct raster algebra (see the textbook for a description) on various layers. QGIS provides two raster calculators, one in the top frame menu accessed by Raster -> Raster Calculator (see at right)

This opens a window which displays raster layers, operations, and allows you to build calculator expressions to apply map algebra:
There is a second raster calculator accessed by Processing→Toolbox→SAGA→Grid-Calculus→Raster calculator (see right).

This also opens a window that allows you to build expressions.

Why the complication of two different raster calculators?

It is due to the “commons” nature of QQIS. Some folks felt the original calculator lacked operations, so they integrated a calculator built in another open source project, the SAGA software, into QGIS. As you know by now, QGIS can be extended via plugins, and there are volunteers working on the base code. This extra calculator allows additional functions, improving the value of QGIS, at the added cost of a bit of potential confusion. You need to be familiar with both raster calculators, the base one, which is easier to use, and the SAGA version, which is more complicated but supports a few more key operations.

We’ll use both in this set of exercises to perform raster operations on DEM data. DEM data often come with various errors, different cell sizes, or different extents, or for only parts of our study areas. We may use raster operations, or functions, to combine and improve our DEM data.

In our first project, we will combine data of two different resolutions. We often have data from various sources, for example, DEMs at approximately 30m and 10 m for most of the U.S., and 3m or better resolutions for a small subset of areas.

Our first task here will be combining two DEMs, valley3, at 3 meter cell
size, and valley9, with a 9 meter resolution. We want to use the higher resolution data where we have it, but use the lower resolution data elsewhere.

(Video: Resample)

Start QGIS and add both Valley3 and Valley9 DEM data sets. Remember to add the W001001.adf file within the \L10\valley9 or valley3 subdirectories. Change the DEM name in the TOC/Layers panel to help identify the DEMs. (See right)

Calculate the hillshade for both data sets (Raster→Terrain Analysis→Hillshade) (You may need to load the Raster Terrain Analysis plug-in)

Inspect these hillshades carefully, and note the enhanced detail with the 3 meter DEM, as shown below. The figure on the right is from the hillshade of valley3, on the left the valley9.

Remove the two hillshades from the data view to reduce clutter.

We can use the SAGA raster calculator to combine these two data sets, but first we must do some pre-processing. Note that we want to use the detailed valley3 data where we have it, and the coarser valley9 data everywhere else.

We’ll do several steps. First, we’ll substitute 0 values for No Data values in the Valley3 data set. Remember from the readings, most raster operations return No Data as output if any part of the input is No Data. Since Valley3 has No Data values over most of its area, we can’t combine it with Valley9 until we change them.
Make sure the Valley3 layer is active (selected).

Use the SAGA toolbox to reclassify data,
Processing→Toolbox→SAGA→Grid-Tools→Reclassify grid values

As shown in the example screen below:
- **Specify Valley 3 as input,**
- **Set the Method as “single”** - we can reclassify ranges of values, but won’t here
- **Set both old value and new value to 0,**
- **Make sure to specify “Yes” in the option for replace no data value**
- **Set the new value for no data values to 0**
- **Select “No” in the option to replace other values,** and
- **Specify an output grid name**
There is a box at the bottom you can check to display the output after running the algorithm, check (as shown) to save having to load the data in the display, then click on Run.

After a bit of processing, the new raster layer should display. Change the symbology if you wish to make the categories clearer/prettier, and use the inspection cursor to verify values, and that you have successfully converted the No Data values to zeros.

We’ll now convert our data sets to a common resolution, in this case changing the 9 meter data to a 3 meter cell size. Some GIS operations/systems automatically change resolution during processing, but this depends on the operator and system, and sometimes they coarsen the higher resolution to a lower resolution, or they average or “resample” the data in a way you’d rather not use (again, see Chapter 4 of the GIS Fundamentals textbook if you’re uncertain on what resampling does). It is better to specify the resampling method. Note that resampling does not make the 9 meter data better. Basically, we are making a copy of the data at a finer resolution, resampling to smooth the result.

To resample data:

- **Select Processing** → **Toolbox** → **SAGA** → **Grid-Tools** → **Resampling**
- **Specify the input grid (here, Grid)**
- **Set the Interpolation Methods for Scale Up and Scale Down to [1] Bilinear Interpolation**
- **Change the Cellsize to 3, then click on Run.**
This should create a new raster named Grid. Inspect the properties via a right click in the table of contents, then look at the metadata (bottom topic, left-side of the menu). As you scroll down you should see statistics, including minimum values of zero, and then eventually the cell size, listed as 3,-3 (see image, right):

Now that the Valley 3 has had null values recoded to 0 and the Valley 9 has been resampled to match the 3 m resolution of the Valley 3 layer, we can merge the two layers.

We’ll use the Processing→Toolbox→SAGA→Grid-Calculus→Raster calculator

Note that this is NOT the same as the raster calculator available from the raster menu on the main QGIS window frame. Here we need the full set in the SAGA toolbox.

Also note that the SAGA raster calculator has changed between the 2.4 and the 2.6 versions of QGIS. Both versions let you list a set of input layers. The 2.6 version has an extra menu window for entering the “Main input layer”, as well as an entry for additional layers. The Main input layer is referred to by the letter a in your calculator formulas when using 2.6, and the additional layers by b, c, ... up to as many additional layers as you specify.

The example at right should help.

The main input layer is zerND1 here. It is associated with the value a in the formula. The Additional layers option is opened when clicking on the circled button on the right. We’ve clicked the check box for the layer v9to3v2 (arrow). This is the resampled 9m to 3m data set, and since it is the second selected data set, it is associated with b in our specified raster calculator formula.
Here we have only two files, so we can only enter a and b for files in any formula. If we enter c, or any higher letter, then the raster layer is not specified, and we’ll get erroneous results.

Note two additional things. First, we can add several more layers, e.g., if I had three additional layers, I would check them, and could use up to the letter d when specifying a formula.

Second, note that I don’t want to check the Main input layer in the Additional layers box (here zerND1), as I’ve already specified it as the main layer. I should leave it unchecked, as shown in the graphic above.

This menu is a bit confusing as it is different from some earlier versions. In QGIS 2.4 and earlier there wasn’t a Main input layer selection, and you would just check layers in a window as in the current Multiple selection window. The layer to letter assignment has remained the same, though; it is sequential, with the first selected layer associated with the letter a, the second associated with b, etcetera.

We fill the formula box to indicate what we want done, in the somewhat veiled syntax, using a, b, c, etc., for each of the files we’re using. File names are not used directly. Instead, letters are substituted, according to the order of the files in the selected list.

When we enter the formula: ```ifelse(eq(a, 0), b, a)``` it reads as follows “if zerND1 equals 0 at a cell, then use the value from v9to3v2, else use the value from zerND1 for the output at that cell”. This formula is repeated for each cell in the data layers.

You might wonder what operators you can use, e.g., what might there be besides `ifelse()`, or `eq()`. You can click on the Help tab in the Raster calculator window, and it will list the available operations and functions (see at right). Scroll through them and see what is available.

Look at the description of the `ifelse()` operation.
It tells you that its syntax is a test, then an “if true” layer, then an “if false” layer.

Ifelse does a true/false test on a cell. If the test is true, it outputs the value from the “if true” layer to the output layer. If the test is false, it outputs the value from the ‘if false” layer to the new layer.

In our example above, the eq(a,0) is our test. Read the Help description for the eq() operator.

The formula we’ve specified tests to see if the layer associated with a is equal to zero at the current cell value. Since zerND1 is layer a, basically it looks at each cell in turn for the raster layer zerND1, returns true if that cell equals zero, and returns false if that cell has a non-zero value. It assigns an output value from layer b in a true case, and layer a in the false case.

Switch back to the Parameters window in the Raster Calculator.

You should apply the formula in the SAGA raster calculator, specifying the appropriate files to combine your No Data/fixed 3 meter, valley bottom DEM with the resampled 9 m DEM using the SAGA raster calculator. Remember, you want to substitute the higher resolution data where it exists for the lower resolution data, and keep all the lower resolution data where there are 0 values for the higher resolution data.

You can verify that your processing worked by calculating a hillshade of the output DEM, and verifying the better detail in the valley bottom. You should have something that looks like the image here, with the higher detail in the valley bottom, and the lower detail in the uplands. You may see a “seam” where the two data sets met. There are ways to fix this using raster operations that we’ll leave as a thought experiment, perhaps you’ll know how after completing the next exercise.
Lowpass Filtering
Start a new Project. There is no need to save the earlier practice session.

We’d now like to introduce filtering as a tool to fix “noisy” data. This is often used with interpolated surfaces, particularly LiDAR data, and similar tools are used near edges for mosaiced DEMs and other continuous surfaces. Filters are described in Chapter 10 of the GIS Fundamentals textbook.

Load the DEM named Shasta, remember to use Add Raster and select the w001001.adf file in the Shasta folder. Rename the layer to Shasta DEM.

Now use Raster → Terrain Analysis → Hillshade to calculate a hillshade surface for the Shasta DEM. Leave the Azimuth as is, set the Vertical angle to 25.

Inspect this, and notice the funny artifacts. These are both data “spikes,” white points with a long thin shadow trailing to the southeast, or “pits,” dark areas on the northwest with a white “edge” on the southeast.

How do we remove these? As described in the textbook, we may use a low-pass filter to identify and get rid of this speckle.

We may apply a low-pass filter with the Processing → Toolbox → SAGA → Grid-Filter → Simple filter. (See Video Lowpass)

Specify the Shasta DEM as the input, and a filter type of smooth, and specify an output name, like lowpass. Use a Square Search Mode and a Radius of 3.

Run the filter, and inspect the output. It is probably easiest to see the effects by calculating the hillshade of the output, and looking at the areas that have spikes and pits. The figure at right, corresponding the area above, shows the reduction in the
size of the spikes and pits, although they are still visible.

We could stop here, and just accept the filtered data layer. But if we look carefully at the filtered and unfiltered hillshade, we’ll see we pay a cost for filtering. We lose some of the fine detail, apparent in the image of the unfiltered hillshade (left), compared to the filtered hillshade (right).

We’d like to both keep our detail and remove the biggest errors. We can do that with the raster calculator.

**Video: Calculator**

First we should subtract the filtered layer from the original *Shasta* layer.  
**Raster ➔ Raster Calculator**  
Select the following in the Raster Calculator "shasta" - "lowpass" (note exact names may vary). Name the output *difference*. 
For our next step we want to replace (correct) cells where the difference is large. Here, large is a relative term, but after a few trials, and looking at the difference histogram, a threshold of about 15 works fairly well.

We want to replace the cells in the Shasta image when they are more than 15 meters different from the filtered surface. Otherwise, we leave the Shasta surface alone, and hence, we don’t get any of the degradation in detail in otherwise good data.

We can open the Processing→Toolbox→SAGA→Grid-Calculus→Raster calculator, and apply the following function, shown below.

You should be able to understand the formula, from the description above. If the absolute value of the difference is greater than 15, we write the filtered value to the output. Otherwise, we write the original data value.
Name the output *smoothed*.

We can verify that this is helpful by viewing the hillshade of the output raster, and comparing it to the original, and the filtered surfaces. Note that we have removed most of the speckle, but maintained the detail.

We could apply the filter successively, and average the local points further, and apply the ifelse function to a difference layer again. It would lead to an improved surface, and we would do this if we were using the data in the project. But since
you won’t learn much new with repetition here, we’ll just produce a map and move on to project 2.

Calculate a hillshade of the smoothed raster, name it HillSmooth.

Use an altitude of 25 degrees.

Add the shaded relief (that is the Hillsmooth), making it semi-transparent (50%) over the smoothed elevation data.

Create a map as shown below, add a reasonable legend, name, north arrow, description, and scale bar, and export the map as a .pdf.
Project 2: A Cost Surface

We will create a cost surface for locating a building. Our cost surface will depend on slope and distance to existing roads. In our problem, we will assign a road construction cost of $25 per meter of road required. We have a vector data layer of roads, digitized from USGS maps, and we will use grid functions to convert this to a cost data layer.

Slope also affects access costs, because roads on steeper terrain are more expensive. The cost is nonlinear, increasing slowly at first for low slopes, then more rapidly at steeper slopes. We will derive slope from a DEM data layer. We will modify the tables associated with both the derived slope and distance layers to include a cost column. To reflect the nonlinearity in slope costs, we will apply the trigonometric sine function to model this increase in cost. We will then add these two cost layers. Finally, we wish to apply an upper threshold of $5,000 to consider only those areas that are within our budget.

Before we start the second project, we need to describe a difference between a permanent reclassification you’ll be doing today, and a display reclass you’ve done before and you’ll also do today.

Remember, a reclassification is a conversion from one set of numbers to another. We do this in a raster GIS through a reclass table. This table has a column for input values (Old values in the figure at right) and a column for output values (New values in the figure at right). Each cell value is examined, and input value matched to an entry in the table, and the corresponding output value reassigned according to the table. For example, the table at right specifies that all Old values between 3.385998 and 7.336329 are assigned a new value of 2.

In a permanent reclassification, each output value is saved to a new raster. In a display reclassification, the value is used only to assign symbols for display. No data are changed in the source file, nor are new files saved. In previous lessons we have only performed reclassifications for display. Today we will perform a permanent reclassification. It is easy to get confused, because the classify menus for applying these two classifications are similar.
Part 1: Cost Surface

Start QGIS

- Create a new map project, add the raster *mardem* to the view, and inspect it. Remember to use Add Raster and select the w001001.adf file in the MarDEM folder. Rename the layer to MarDEM

Use the cursor and the layer Raster→Miscellaneous→Information. What are the elevation values? What are the highest and lowest elevation values? Does it make sense?

- Derive the slope for *mardem*. Select Raster→Terrain Analysis→Slope. Name the output file *mar_slope*. Video: Slope & Reclass

- To keep the view uncluttered, remove the *mardem* raster from the map.

- Examine the slope layer. There should be values from 0 to about 33 degrees with decimals.

- For our next step we need to remove the decimal component of the slope values (a crude reclassification). Use Processing→Toolbox→SAGA→Grid-Calculus→Raster calculator and the operator int(a) where (a) is the *mar_slope*. Name the output grid something like *Slpcls*.

You may why work on derived layers, why not just work on the slope layer? We could skip several steps, but the first time it is best that you see the details.

Remove the original slope layer (*mar_slope*), to reduce clutter.

Next apply a formula that determines the cost of building on slopes:
Left click on Raster ➔ Raster calculator.  (Video: Calculator)

Use the mouse to select the following operators and values for the center window: \( \sin\left(\frac{\text{Slpcls}}{57.2958}\right) \times 200. \)

Name the output Slope_cost.

Verify that the cost layer makes sense, that costs are highest where slopes are steepest.

Next, we need to display and generate our distance costs from the roads layer.

Add the Mar_rd83.shp file.

Edit the Mar_rd83 file adding an integer field named Roads and Assigning the value of 1 to all the records.

Convert the vector Mar_rd83 to Raster with

Raster ➔ Conversions ➔ Rasterize (vector to raster). Name the output file raster_rd.  See the figure for settings.
Now we use the road raster to determine the distance from every point on the raster to the nearest road.

Left click **Raster**→**Analysis**→**Proximity (Raster distance)**

Name the output raster *Distance.tif*. See right and below for settings.

Examine the result layer, and make sure it is reasonable.

Now, multiply the distance layer you just produced by the cost per unit distance to estimate distance cost.

**Left click** **Raster**→**Raster calculator** and enter the equation as shown right.

**Name the output raster** *Dist_Cost*. 
Our next step is to combine the two sets of costs. Open the raster calculator again (Raster→Raster calculator), and add the two cost layers, as below:

Name the output **Total_Cost.tif**

Examine the **Total_Cost** layer and make sure it makes sense.

Think a minute about what you've just done. You first calculated a slope, and then a cost associated with building a road per unit distance across the slope. Then you calculated a distance, and then a cost associated with building a road to that distance from an existing road. Both of these were calculated for every grid cell in your study area. You then added these two together for an estimated total cost to build a road to any portion of the mapped area. A real problem would include many other factors, like soils, surface vegetation, slope constraints over minimum segments, etc., but this would only lengthen the analysis, and not change the basic way you are applying the tools.

Our job now is to select those areas below the $5,000 threshold. We will do this by creating a mask grid. This grid will have 1 at all locations where the costs are below $5000 and 0 where the costs are above $5,000. We will then multiply this with our total cost grid, to zero out those areas we don't wish to consider.
Reclassify Total_Cost by Processing→Toolbox→SAGA→Grid-Tools→Reclassify grid values.

Select Simple Table for Method and click the … to build the table.

Remember to check NO for Replace other values.

Make sure you specify the output raster name, here shown as Mask.

After the reclassification, rename the Reclassified Grid to Mask.

Multiply the Total Cost raster by the Mask raster. Left click on Raster→Raster Calculator and multiply Total Cost by Mask, call the output something like Final_Cost.
Display the Final Cost layer in your data view.

Add the roads layer, mar_83.shp, and create a layout with appropriate legend, titles, name, north arrow, etc. Create a pdf of the map.

Also create a composite layout with three separate data frames on the same layout, with 1) a data frame with the mask layer, 2) another data frame with the slope costs layer, and 3) a data frame with the dist_cost layer.
Color the mask as gray and white, and color the distance and slopes costs as graduated colors, with a gray monochromatic color set. Include the appropriate legend for each map.

An example of your composite map of source data.

MAPS TO TURN IN:
- The fixed Shasta DEM shaded relief map
- Potential new road location map
- Three working maps shown above (on one page)