Lab 12: Sampling and Interpolation

**What You’ll Learn:**
- Systematic and random sampling
- Stratified sampling
- Majority filtering
- A few basic interpolation methods

**Data** for the exercise are in the L12 subdirectory.

**What You’ll Produce:** two maps, each with four panels. One map will have panels with i) the shaded relief original DEM surface, with sample points, and ii through iv) panels with the interpolated surface and contours for IDW, trend surface, and spline interpolation.

The second map will have the same panel i) as the first, but will have error surfaces and summary statistics corresponding to each of panel’s ii through iv.

**Background:** Theory is covered in Chapter 12 (Spatial Estimation) and 10 (Raster Analysis) of the GIS Fundamentals textbook.

**Sampling and Interpolation in QGIS**

We’ll practice the mechanics of developing and applying sample points in QGIS, and use them to extract data from a DEM and interpolate a surface. We’ll then compare the interpolated surface to the original DEM to calculate an error surface, and extract error statistics from a different set of sample points from the same DEM.

We’ll apply both systematic and random sampling. We’ll also develop and apply a stratification layer, because sometimes you want to stratify your sample, which means you wish to increase sample density in some portion of your area, using a map of zones, or strata.

Add the DEM layer, `chirdemQ`, a region in southeast Arizona and apply a color scheme to highlight topography in the DEM. Note the variation in topography, where there is a significant amount of change, where there is little change. Create the Hillshade with **Raster ➔ Terrain Analysis ➔ Hillshade**. Use an altitude of 25 degrees. Make the hillshade semi-transparent on top of the `chirdemQ` DEM.

To keep your work organized, “Group” your data layers within the Layers panel. There will be four (4) parts of the lab. Each will have data layers. Within the Layers Panel you can “Group” layers:

- **Select both the chirdemQ and Hillshade of chirdemQ in the TOC** (shift-click (PC) or cmd-click (Mac), then right click, and finally click on “Group Selected” in the dropdown.)
• **Right click on the group and rename it to “Original”**.

As you create layers for each of the three different interpolation methods, select appropriate layers and “Group” them with a logical name. I suggest: Inverse Distance, Nearest Neighbor and Spline Stratified respectively.

Here is what the final output will look like. Maps will be done in Landscape view and arrange the layers as below: You’ll need to set your print canvas to landscape orientation and make sure the map images **DO NOT** overlap.

![Map 1 and Map 2](image-url)
Systematic Sampling and IDW Interpolation
We’ll first perform a systematic (grid) sampling, and then an Inverse Distance interpolation (Video :)

- **Open Vector Tools -> Research Tools -> Regular Points.**

- In the window that appears, navigate to a directory and name your output something like sys1000. Make sure your input is chirdemQ and “Use number of points” is set to 1000 (see at right).

- **Select OK**

Next we need to extract the elevation values at these points.
To sample point values, use

Processing→Toolbox→Models→Example models→Extract raster values (shapefile) (see figure at right)

*Note for QGIS 2.10+: use Saga→Shapes-Grid→Add grid values to points instead.*

Specify input point feature (sys1000, from above), chirdemQ as the input raster, and an output point features of something like sys1000_with_elevation (see below).
Inspect the table for \textit{sys1000\_with\_elevation} noting the attributes and their values. You should have about 1000 rows in the table, one row for each data point. Note the column labeled \textit{chirdemQ1 (or w001001\_1)} is filled with elevation values.

Now, perform an Inverse Distance Weighted interpolation.

- \textit{Start the tool Processing\textgreater SAGA\textgreater Grid-Gridding\textgreater Inverse distance weighted.} Specify the \textit{sys1000\_with\_elevation as your input point file}, the \textit{column of sampled DEM values as the Z value field as input}, a power
of between 1 and 3, a search radius of 100, a Maximum number of Points between 10 and 15, and an output cellsize of 30.

Note that you should name the output raster something that makes sense, but also within the limits on naming conventions. Here we named it IDWSysP2_12.

On Run this should create and load an interpolated data layer.

- Generate contours for the interpolated surface with Raster→Extraction→Contours. Specify the IDW surface as input, a 100 meter contour interval. Name the output contours appropriately.
• **Group** the sample points, contours and IDW surface in the Layers panel appearing similar to the figure at right.

Random Sampling and Nearest Neighbor Interpolation (Videos: )

Generate a set of random sampling points:

• **Vector Tools -> Research Tools -> Random Points**

• **Specify** appropriate file for the output location, name the point feature class something like ran1000, leave the constraining feature class blank, and set the constraining extent to chirdemQ.

• **Specify 1000 as “The Number of Points”**.
Now sample the elevation values from chirdemQ as with the systematic layer:

- **Processing** → **Toolbox** → **Models** → **Example models** → **Extract raster values (shapefile)**.

Note for QGIS 2.10+: use

Saga → Shapes → Grid → Add grid values to points instead.

- **Name the output something like ran1000_with_elevation**.

- **Remove the ran1000 data layer**.

Calculate and interpolation surface, this time using:

**Processing** → **SAGA** → **Grid-Gridding** → **Nearest Neighbor** (see right, clicking on Nearest neighbor opens menu in figure below).

- **Save** to a permanent output data set named something like **NearNeigh**.

- Make sure you use the correct column/attribute for the Z value (**chirdemQ_1**) and a cell size of 30.

Run the interpolation, and then load the output interpolated surface into your data view.
- **Create and add contour lines**, as you did for the IDW interpolation.

- **Group** the sample point, contours, interpolated raster layers within the Layers panel.

**Stratified Random Sampling of a Raster Layer**

(Video :)

Sometimes we want to vary the sampling frequency across a map. Here, we'll place more samples in steeper areas, and also proportional to area. First we'll create three zones, or strata, and then we'll assign samples based on these strata.

Our strata boundaries will be based on calculated slope, filtered to create larger, more generalized areas.

- **Calculate the slope for chirdemQ** (Raster → Terrain Analysis → Slope), set a z-factor set to 1, saving to a permanent dataset named something like “slope_deg.”
Create your strata:

- Reclassify slope_deg using Processing→SAGA→Grid-Tools→Reclassify grid values

- Use **three classes, of 0-1.5, 1.5-18, and 18 and above**. We chose these values to yield acceptably balanced classes. Usually you stratify for some threshold of an attribute, e.g., slopes above which you can’t build, or elevations where you’re unlikely to find a resource of interest.
- **Name the output** something like `rcslope`
Now we want to generalize the strata polygons, removing those from single cells or long, thin areas. Keeping them would complicate sampling needlessly.

- **Use Processing → SAGA → Grid-Filter → Majority filter.** Specify a 4 as the Radius and the 50% threshold.

Notice how the small speckles are removed in the majority filter output, but the long, thin reaches in valleys remain? To generalize further and remove these, you may apply a median filter, via:

- **Processing → SAGA → Grid-Filter → User defined filter.**

- **Specify the majority filtered layer as input, then build a kernel (Fixed 3x3) table** as shown below:
Notice how this substantially generalizes the data, removing the long, thin features.

Convert the final smoothed raster to a vector layer:

- \textit{(Raster} \rightarrow \textit{Conversions} \rightarrow \textit{Polygonize (Raster to vector)}, name it something like “Strata.” Use symbology to something similar to the right.

Now we must create the stratified sampling points.

We would like to have a total of approximately 1000 sample points, with 10 times as many sample points in the steep areas (red) as in the flat (green areas). We’d like three times as many samples in the intermediate areas (yellow) as the flat areas.

So, the relative weightings are 10 for the red, 3 for the yellow, and 1 for the green. The total sums to 14, so that means 10/14ths, 3/14ths, and 1/14\textsuperscript{th}.

This means we want about 1000*10/14 or 714 samples in the red, 1000*3/14 or 214 samples in the yellow, and 1000*1/14 or 72 samples in the green.

We can achieve this by distributing these samples over the polygons, based on the polygon area relative to the total area for the strata. For example, in my layer above, the largest yellow polygon has an area of 116.3 square kilometers, and the total area in yellow is 192.5 square kilometers. So, this largest yellow polygon should get 214 * 116.3/192.5, or 129 sample points.

We multiply the number of points for the strata by the polygon area, and divide it by the total area of the strata.
How do we get the total strata area?

Remember from previous labs, we can calculate the area for each individual polygon, editing the table and using the Field Calculator. Remember the steps are to:

1. Open the data table for editing,
2. add a field called “SQKM” that is float, with a precision of something like18, and a scale of something like 7 or 8 to hold down to sub-meter amounts,
3. calculating the area into this new column (build and expression $\text{area/1000000}$), then
4. save edits and toggle off editing.

We can then use summarize using Vector→Statist by each strata identifier value (1, 2 or 3), using SqKm for the Target Field. Review the previous labs where we used Statist if you have don’t remember how, and see the instructor if you can't figure out its application in this exercise.

You should arrive at calculated areas close to:
- 268.7 square kilometers for the flat (DN = 1) strata
- 191.9 square kilometers for the intermediate (DN = 2) strata, and
- 212.4 square kilometers for the steep (DN = 3) strata.

Your numbers may be slightly different, but should be with a few percent of these areas if you used the methods we described above.
Record your numbers, you will need them later. Samples per polygon may be computed many ways (video), but we'll:

1) Open the strata layer attribute table and add a new long integer field named samp_num
2) Select all polygons for a given strata (DN)
3) Multiply the total number of points for this stratum (e.g., 714 for the steep strata, DN = 3) by the area of the polygon, divided by the total area of the strata (in this case 212.4). Do this for all three strata, substituting the appropriate areas and number of samples for the strata.

Note that your numbers for the strata area and relative number of samples may be different than those shown if you apply a different set of generalization parameters.

Next perform Selection → Clear Selected Features. (Video:)
Now, to create the random points. We use the same tool as before,

- **Vector→Research tools→Random Points**

- Specify an output location and file name, something like strat_ran, and

- “strata” as the Input Boundary Layer”, and

- use the value from the input field samp_num.

This should generate a sample set that looks something like that to the left, with a higher sampling density in the steeper areas.

Now you will use this stratified random sample to produce one last interpolation method.
First, use

- **Processing → Toolbox → Models → Example models → Extract raster values (shapefile)** to assign the chirdemQ elevation values to each sample point (refer to the previous instructions if need be). Name the output file something like `strat_ran_with_elevation`.

**Note for QGIS 2.10+:** use **Saga → Shapes-Grid → Add grid values to points instead.**
Next,

- estimate a surface using a spline interpolation routine, found in the Processing→SAGA→Grid-Spline→Cubic spline approximation.

- Specify the **strat_ran_with_elevation as the input for the point features, chridemQ-1 (your elevation or Z value) for the Attribute**

- **with 3 for the Minimal Number of Points,**
- **20 for the Maximum Number of Points,**
- **5 for the Points per Square,**
- **140 for the Tolerance** and
- set the **cell size to equal 30,** that of chridemQ.

- Specify an **output raster called Spline**

After running the tool, add this output to your data frame. Calculate contours, and “Group” the Spline layers in the Layer panel.

Now display these “Groups” and the sample points on your layout. Arrange the layout so it looks approximately like that in the figure below, with an appropriate title, labels, scale bar, name, uniform elevation legend across all three interpolation layers, and north arrow, and turn this in as a PDF. This is **MAP 1.**
Interpolation and Sampling Methods

Original

Inverse Distance Systematic

Nearest Neighbor Random

Spline Stratified Random

2.5  0  2.5  5  7.5  10 km
Error Map

Now, calculate error surfaces for each of the three interpolations. Do this by using
- **Raster→Raster Calculator**, and **subtract the interpolated raster from chirdemQ** for each of the three methods.

Be sure to use clear output names, for example IDWdiff for the difference raster.

Also note that sometime the raster is added to a different Group in your Layers panel.

If so select it and drag it to the correct Group.

Notice that each of the “difference” rasters has different high and low values, and they may be calculated incorrectly. Remember that the default raster display in QGIS omits the top and bottom 2% of the data, but approximates it for speed. For each each difference layer, use

- **Properties→Style** then **Min/Max and Accuracy→Actual Values (slower)** then select the **LOAD button**.

This will make sure your high and low values are correct.
Now examine the ACTUAL high and low values, and

- **Select each difference layer Properties→Style and change to Singleband pseudocolor.**

- **Type in the highest and lowest values and select Classify**

Array the error surfaces in a second layout, MAP2, similar to that below. Remember you should use a uniform error range for the symbology across all three error surfaces. It also helps to only display your layer then add the layer to the Map Composer and lock the item. Turn this is a PDF.