**VMSG *GIS for Geoscientists* Workshop 2021**

**Session 3 Outline**

Given the amount of material I cover in this course, I can’t produce the same level of exercise guide for Sessions 3 and 4 that I did for 1 and 2. The topics we are going to cover in Sessions 3 and 4 are rich and interesting, but nuanced, and to explain properly I’d need to take a writing sabbatical and write this course into its own book! Instead, this guide serves as an overview of Session 3, and points you to additional resources.

You can think about our first two Sessions together as complements, centered around familiarizing you with GIS principles and basic tools. Exercise #1 - Hazard assessment at Guagua Pichincha – as the main vehicle for teaching these core skills. The second half of this course is meant to expand on these core skills, providing you with depth and breadth in your approach to GIS. However, keep in mind that these sessions are merely introductions – I can’t teach you everything you need to know to be competent with GIS in 8 hours. Take the lessons I impart here and in future lessons as jumping off points! Make your own maps, make mistakes, and experiment. You’ll find your way! And if you get lost these tools are here to help you course correct.

With that in mind, I’m going to outline my plan for Session 3, which includes several smaller exercises aimed at expanding on different elements of our geospatial “toolkit.” In this session (Session 3, 17/11/21), I’ll focus on expanding our **analytical** toolkit. Next time (Session 4, 24/11/21) I’ll touch on **advanced visualization** and more **computational approaches** (e.g., Python, Model Builder, Google Earth Engine**)** to GIS**.**

Below you can find my detailed plan for Session 3. I’ll outline the tasks we will undertake, why we are doing them, and the practical “how to,” for each of them.

**Outline of Session 3:**

1. Introduction and Housekeeping (5 minutes)
2. GIS Ethics: Avoiding Map Colonialism (10 minutes)
3. Wrapping up Exercise #1 (45 minutes)
4. Break and Q&A (5 minutes)
5. **Exercise #2** – Fluid geochemistry at crater summit (25 minutes)
6. **Exercise #3** – Landslide Risk in Quito (25 minutes)
7. Q&A (10 minutes) – This will go past 14:00 GMT but if you have questions, you are welcome to stick around!

**Exercise #2: Fluid Geochemistry @ Guagua Pichincha**

A generic problem all geoscientists using GIS face is the need to import non-georeferenced attribute data, often in the form a spreadsheet like a CSV. How do you display this data on a map and spatially analyze it? We will explore this using another hypothetical scenario, where we are tasked with analyzing the chemistry of fluids measured at Guagua Pichincha (G.P.). This need may arise from concerns about the quality and kinds of hydrothermal activity at the summit, motivating fresh research questions. For instance, are there hotspots of ions of interest like SO4, reflecting increased shallow magmatic activity? These kinds of signals may be useful for us in interpreting where shallow hydrothermal activity is centralized, or it may give us clues as to the spatial variability in the magma feeding these hydrothermal systems. We may be asked to produce a literal “hot spot” map of the region, showing where chemicals of interest are in their highest concentrations. Do these hot spots co-vary? Even if you aren’t a geochemist, these kinds of operations are useful for any geoscientist!

To answer these research questions, we need to:

* Identify relevant geochemical data
* Georeference maps of sampling location
* Extract spatial location of sample points
* Import un-georeferenced geochemical data
* Perform a spatial join of geochemical layer and sample layer
* Interpolate between sample points to generate a “hot spot” map of fluid geochemistry.
* How do data points cluster?

Key terms and concepts:

* *Spatial Join*: operation that takes two input datasets and produces one output dataset sharing the same spatial properties as the inputs, forming a bigger dataset. This increases the number of feature attributes using a condition like a foreign key. This method uses same conceptual and computational framework as SQL databases.
  + See Chapter 3.4 (pg.172) of Huisman and de by (2009): <https://webapps.itc.utwente.nl/librarywww/papers_2009/general/principlesgis.pdf>
* *Geostatistics*: Statistical methods applied to geospatial data. Its own subdiscipline requiring specialized methods, as geographic data violates the independence assumption required by most classical statistical methods. This arises from an observation made about spatial data termed *spatial autocorrelation*. In classical probability theory, the occurrence of event A does not affect the probability of occurrence of event B. However, in geographic data, the probability of event B taking place increases if event A occurs close by.
  + For more, see excellent free resource by de Smith: <https://www.spatialanalysisonline.com/HTML/index.html>
* *Interpolation*: predictive analytical procedure, which seeks to estimate underlying surface connecting discrete points. Generates a fine rectangular grid covering the study region, and then estimates the surface value of each cell in the grid based on spatial distribution and attribute variation in input data. Critically, *geostatistical interpolation* and *deterministic interpolation* are different families of interpolation models. Geostatistical interpolation assumes the source data points for the operation reflect a specific statistical sample e.g., kriging. *Deterministic* methods (more common) differ by having their weighting scheme (how much does X vary with distance?) determined by model choice and parameters used by user, so offers more direct user control e.g., Natural neighbor, Inverse Distance Weighting (IDW), Minimum curvature.
  + See Chapter 6 (Surface and Field Analysis) of de Smith: <https://www.spatialanalysisonline.com/HTML/index.html>
  + See Chapter 11 of QGIS *Gentle Introduction to GIS*: <https://docs.qgis.org/3.22/en/docs/gentle_gis_introduction/spatial_analysis_interpolation.html>

Data source:

Marini, L., Agostini, A., Cioni, R., Guidi, M. and Leon, O., 1991. Guagua Pichincha volcano, Ecuador: fluid geochemistry in volcanic surveillance. Journal of Volcanology and Geothermal Research, 46(1-2), pp.21-35.

<https://www.sciencedirect.com/science/article/abs/pii/0377027391900739>

**Exercise #3: Landslide Risk Near Quito**

Raster data, which provides continuous coverage of data over a region, is helpful for performing powerful spatial calculations at regional or even global scales. Manipulating raster data is therefore one of the most powerful skills that can help you understand how environmental processes impact a particular region. Near Quito, there is substantial relief in the mountains overlooking the city. There is also an active volcano, and substantial precipitation. What risk do landslides pose to particular areas of the city? Are these risks more pronounced in some areas than others? By leveraging the powerful *Raster Calculator*, we can find a way to address these questions in an illustrative exercise.

NOTE: The output file this exercise produces is not meant to be an accurate landslide risk map. Rather, I am showing you how to work with specific tools and datasets and performing a simple Raster Calculation for illustrative purposes. A real landslide risk analysis would be much more mathematically and geologically complex!

To address this question, we need to:

* Import precipitation data
* Calculate Raster Slope
* Use raster calculator to determine compound risk of landslide given probability of landslide is some function of precipitation rate and slope

Key terms and concepts:

*Raster Calculator:* Powerful tool that lets you perform calculations based on existing pixel values in any number of raster layers. Allows you to stack, subtract, multiply, and otherwise mathematically transform raster layers, producing new layers that can further your analysis. Uses a simple syntax that allows you to create complex logical and mathematical methods for calculation.

Data Sources:

Tropical Rainfall Measurement Mission (TRMM) Version 7. I obtained the version used in this practical from NASA GSFC: <https://disc.gsfc.nasa.gov/information/tools?title=Giovanni>

Topography data used to calculate slope is the same SRTM 1-arc second DEM used in other exercises. Sources from NASA EarthData: <https://earthdata.nasa.gov>