    

**Project EDDIE:**

**Correlation of Vegetation and Topographic Indexes at Local and Continental Scale**

**Student Handout**

This module is an adapted version of the module Remote Sensing of Plants and Topography in R with a slightly shifted focus on different landscape indexes by Conrad Jackisch in the EDDIE Faculty Mentoring Network Spring 2021. 16 April 2021. The original module was developed by Kyla M. Dahlin. 8 July 2020. Project EDDIE is supported by NSF 1821567. KMD was supported by NSF 1702379. NEON is a program sponsored by the National Science Foundation and operated under cooperative agreement by Battelle Memorial Institute. This material is based in part upon work supported by the National Science Foundation through the NEON Program.

Overall Question: How do landscape and vegetation indexes correlate at local and continental scale?

Specific Questions: (1) Which topographic indexes are correlated with vegetation height and greenness in the Southern California foothills? (2) How does correlation change in stratified subsets of the data? (3) Do these local relationships persist across the U.S.? (4) Do these relationships follow any macroscale geographic patterns?

Learning objectives:

* Employ methods for terrain delineation and topographic indexes for land surface analyses.
* Visualize multivariate correlations of landscape indexes and vegetation growth and fitness.
* Develop hypotheses about topography-vegetation correlation (and non-correlation) and test these hypotheses in a scripted application to data from other NEON sites.
* Analyze and visualize spatial raster data in R, moving between making maps and doing non-spatial statistical tests.
* Consider macroscale (continental scale) patterns of relationships between topography and vegetation.
* Reflect the perceptual assumptions in the employed analyses.

Why this matters: At the global scale, we know that there is more above ground biomass near the equator than there is towards the poles. At the local scale, however, many different environmental gradients (e.g. topography, geology, hydrology) can influence plant growth. Within each of these environmental factors, there are many possible ways the gradients can be quantified. Understanding these environmental drivers is important because they are likely to change in the face of global change pressures - for example, as temperatures increase, plants may become greener at higher elevations, indicating either an increase in growth of the existing plants or a change in the plant species. Similarly, plants of the same species tend to grow larger under optimal growth conditions. In this module, you will learn how to assess whether plant greenness and height are correlated with topographic variables (elevation, slope, aspect) at one location, and then across many locations in the U.S.

Where did these data come from? The National Ecological Observatory Network (**NEON**) is a U.S. National Science Foundation-funded project with 81 sites across the US. They collect > 170 different types of data at each of the sites (learn more here: <https://www.neonscience.org/data/about-data/getting-started-neon-data>). Of these data types, one set comes from the NEON Airborne Observation Platform (AOP), which flies over most of the sites every year. The NEON AOP includes hyperspectral and LiDAR sensors. These data are processed and then provided in 1 km2 tiles for an approximately 10x10 km box around each site at 1 m resolution. All of the data are available via the NEON data portal here (<https://data.neonscience.org/home>). For each of the sites where data was available for 2018 or 2019, the digital terrain model (**DTM**; meters (m)), digital surface model (**DSM**; m), and Normalized Difference Vegetation Index (**NDVI**; unitless, ranges from -1 to 1) were downloaded and packaged for the tile that includes the NEON eddy covariance tower – essentially the center of each site. At 1 km2, with 1 m data and three data layers, this means we have 3,000,000 data points to start with for each site.

Outline:

1. *Before Lab*: Discussion of papers read for class and PowerPoint presentation. You should have familiarity with DEMs, slope, aspect, and NDVI. You should have some comfort with R/Python and access to the software. See References and Resources for suggested tools.
2. **Activity A:** Calculate slope and aspect from a DEM, map elevation, slope, aspect, NDVI, flow accumulation, TWI, and vegetation height. Calculate northness (cosine of aspect). Based on visualization, generate hypotheses about the most likely driver(s) of vegetation growth.
3. **Activity B:** Calculate correlations between topographic variables and vegetation variables. At the end of this activity, we will decide on one metric and gradient combination to explore across a set of NEON sites to determine the extent to which the local pattern applies at large scales.
4. **Activity C:** Each student chooses a NEON location, accesses data, and calculates correlations as we did in Activity B. Then, the correlations chosen at the end of Activity B are mapped together (by hand or digitally) on a map of the U.S. that is accessible to everyone, and macroscale patterns are considered qualitatively (as a group or as a written assignment).

**Activity A:** Plotting maps of topography and plants and making predictions.

1. Make sure data is downloaded and open RStudio/Jupyter. Make sure the software is set up to your liking and including some topography-driven calculation features.
2. Follow along with the instructor to load packages, set your working directory, load in the DTM, DSM, and NDVI maps.
3. Calculate vegetation height (DSM – DTM), slope, aspect, and northness.
4. Plot these four maps and consider the patterns you observe. Paste the map image here:
5. Use WhiteboxGIS to calculate hydrological catchment indexes like flow accumulation, slope gradient and the topographic wetness index (TWI).
6. Plot these three maps and consider the patterns you observe. Paste the map image here:
7. Based on these maps, make some predictions about the correlations you might find.
   1. Which topographic metrics do you think would have a stronger association with NDVI? Draw a graph of your predicted relationship.
   2. Which topographic metrics do you think have a stronger association with vegetation height? Draw a graph of your predicted relationship.
   3. Overall, does it seem like vegetation patterns are influenced by topography?
   4. What else do you notice about these maps?
8. Discuss, as a class/group, the patterns you see or do not see.

**Activity B:** Plot the relationships between topographic variables and vegetation patterns.

1. Convert the produced raster layers into a data table (if not done so already) and take a random 1% subset to speed along plotting and analysis.
2. Plot the correlations between vegetation variables (y-axes) and topographic variables (x-axes). Paste the figure here:
3. Look at these scatter plots and make some predictions about which will have statistical relationships.
   1. Do any of these plots look like there is a strong relationship?
   2. Is it harder or easier to see relationships with large data sets like this? Why? (Discuss as a class, if possible)
4. Calculate the correlations and record them in the following table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | DTM | Slope | Northness |
| NDVI | R |  |  |  |
| p-value |  |  |  |
| Veg Height | R |  |  |  |
| p-value |  |  |  |

1. How do these relationships compare to what you expected when you looked at the maps and the scatter plots? Discuss with your neighbor and write your answer here.
2. Are all the relationships statistically significant (traditionally defined as p-values lower than 0.05)? What does that make you think about the importance of p-values in large data sets?
3. Reconsider your findings based on an analysis of stratified subsets of your data (e.g. correlation of slope for defined ranges of exposition; e.g. check the assumptions behind regression statistics). What do you find?
4. As a class/group, consider these patterns then decide on one of these relationships (e.g. NDVI compared to northness) to compare for Activity C. What did you choose and why?

**Activity C:** How do these relationships vary across the US?

1. Each student should pick a different NEON site across the U.S. (note that for ease-of-mapping reasons we are using just the contiguous US, but there are NEON sites in Hawai’i, Alaska, and Puerto Rico).
2. What site did you choose?
3. Based on your prior knowledge of this area, what do you expect to be different or similar about this area compared to the California site we started with? How do you expect the vegetation to be different? The topography? Their relationship?
4. Change the path name in the code so it accesses the folder and files where the data for your site are stored. Read these new data in, and use parts of the code to make the following plots and calculations.
5. For your new site, plot the same six panels (DTM, DSM, vegetation height, slope, northness, NDVI) and consider the patterns you observe. Paste the map panel images here:

1. Based on these maps, make some predictions about the correlations you might find.
   1. Which topographic patterns do you think have a stronger association with NDVI?
   2. Which topographic patterns do you think have a stronger association with vegetation height?
   3. Overall, does it seem like vegetation patterns are influenced by topography?
   4. What else do you notice about these maps? Consider the maps alone and in relation to the California maps.
2. As in Activity B, convert these raster data to a data table and take a 1% subset. Plot the correlations between vegetation variables (y-axes) and topographic variables (x-axes). Paste the figure here:
3. Calculate the correlations and record them in the following table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | DTM | Slope | Northness |
| NDVI | R |  |  |  |
| p-value |  |  |  |
| Veg Ht | R |  |  |  |
| p-value |  |  |  |

1. How do these relationships compare to what you expected when you looked at the maps and the scatter plots?
2. Report the value your class decided on in Activity B, step 7 to your instructor so it can be written up on the U.S. map.
3. Once the map is filled in (either in class or later as homework), consider the overall patterns of correlation. Do you notice any patterns? Do they seem to relate to any of the geographic variables we discussed in the lecture? If yes or no, explain why you think this is?
4. What do you think about the scales of our analysis? How might our results be different if we looked at coarser spatial scale data (bigger pixels)? Or if we considered larger spatial extents for each location? How would a temporal change impact these patterns (like a drought in the upper Midwest, for example)?
5. If you were going to ask this type of question again, what would you do differently?