# Exercise 2: Physical environmental controls on land cover pattern

The purpose of this exercise is to familiarize you with data and tools for exploring the relationship between geology, soils, topography and land cover pattern. You will beusing data modeling and visualization tools in ArcGIS and R to examine landscapes of the western Santa Ynez Valley in Santa Barbara County.

## 1) Gettting started

    A. Open ArcGIS (Programs -> ScienceApps -> ArcGIS -> ArcMap 10.3.1)

    B. Open the arcmap project R:\Winter2019\ESM215\data\datavu.mxd (File -> open -> R:\Winter2017\ESM215\data\datavu.mxd). [You should save the file to your own local directory.]

    C. The project legend includes a variety of data files, which can be found in [R:\Winter2017\ESM215\data\](file:///Volumes/courses/Winter2010/ESM%20215/data/datavu.mxd) (To build a local database, use Arc Catalogue to copy these data to your own local working directory).

**airphoto** – 2 m resolution raster: true color air photo mosaic, 2004

**naip2014 –** 1m resolution raster, true color air photo, 2014

**casubsect** - ecological subsections: feature (polygon) data. We will focus on subsection 261Ba, Santa Ynez Hills and Valleys. (<https://databasin.org/datasets/4996c7e61a0e48f2bef646903f51b82b>)

**subgeol**– A subregion of the 1:250,000 scale geologic map of CA. 30 m raster. A more detailed map with legend is available [here](http://www.quake.ca.gov/gmaps/GAM/santamaria/santamaria.html).

**subsoil –** A subregion of the1:24,000 scale soil survey map ([SSURGO](file://localhost/,%20https/::www.nrcs.usda.gov:wps:portal:nrcs:detail:soils:survey:%3Fcid=nrcs142p2_053627)), 28 m raster data. SSURGO maps are the most detailed soil survey maps available for most of the U.S. and are used extensively for landscape-scale analysis.

**subsynezdem** - 28m raster: Shuttle imaging radar topographic data. Values are elevations in meters above sea level.

**subslope** - 28 m raster: slope angle in degrees derived from subsynezdem

**subshad** -  28 m raster: shaded relief image, derived from subsynezdem

**subwintrad -** 28 m the data raster**:** integrated clearsky shortwave radiation, units are watts/sq. m., for December-Feb.

**subwinrad3** – subwintrad reclassified into 3 radiation classes. Use this grid for the exercise.

**subradsum -** 28 m the data raster**:** annualclearsky shortwave radiation, units are watts/sq. m.

**subflocum** –28 m raster: flow accumulation model, derived from synezdem for a subregion corresponding to subsoil30. Pixel values are the drainage area for each pixel. (The data are noisy because errors in the dem propagate to disrupt drainage topology.)

**subflocum3c** – **subflocum** reclassified into 3 accumulation classes. Use this grid for the exercise.

**subfire15** – 28m grid of fires from 1900-1915 (incomplete!), compiled by the CA Dept of Forestry and Fire Protection and available [here](http://frap.fire.ca.gov/data/frapgisdata-sw-fireperimeters_download).

**subveg**   - 28m raster: 1990-2014 vegetation/land cover map produced (mainly) from Landsat Thematic Mapper satellite imagery. California Wildlife Habitat Types are shown here. [Here](http://frap.fire.ca.gov/data/statewide/FGDC_metadata/fveg15_1.xml) is a description of the data.

**subveg15** – a reclassification of subveg merging some agricultural classes. Use this grid for the exercise.

In lab spend some time learning to display the data. Overlay individual layers and combinations on the air photo. Zoom in and out. Play with the symbology. Get the feel for ArcGIS as a visualization environment. In particular, examine apparent land cover pattern (air photo) and vegetation pattern (subveg) in relationship to geology, soils, fire history and topographic factors like elevation, slope, radiation and flow accumulation.

## 2) Quantitative association of thematic (categorical maps)

What controls land use/land cover pattern in the Santa Ynez Hills and Valley subsection? Landscape theory posits that pattern could vary from one landscape to another and reflect interacting local physical controls, disturbance history and population processes such as plant dispersal.

Various techniques exist to quantify the spatial relationship between land cover and environmental factors at different scales (Wagner and Fortin 2005). Here you will learn a method known as “mutual information analysis” useful for measuring the association between categorical maps (e.g., association of land use and soil type). The theoretical underpinnings of the technique are described by Phipps (1981) and example applications to landscape analysis include Davis and Dozier (1990) and Ernoult et al. (2003).

You will be analyzing vegetation pattern in the area covered by the 30 m grid sublandfire30 in your project legend. Here are the steps:

1) Generate a random sample of points at which you will collect information on land cover class (**subveg**), geology (**subgeol**), flow accumulation (**subflocum3c**), and winter radiation (**subwinrad3**). *OR,* you can use the set of 10,000 random points that I generated using the “Create random points” tool in the Data Management folder of Arc toolbox. The sample locations are already in your map legend as **ex2\_sample**.

2) The sample points can be used to extract values at those locations from multiple grids (Spatial Analyst folder, Extraction -> Extract Multi Values to Points) The output will be added to the attribute table for your sample points. *OR*, values already added to ex2\_sample. You can open the tables and export the table to your work directory. *OR*, you can copy the exported table R:\Winter2017\ESM215\data\ex2\_sample\_data.csv.

3) This exercise can be completed using Excel or R, but it is much faster in R. Import the data into R using the function read.csv(). Load the package **entropy**. From here you can use the function table() to cross-tabulate vegetation and other categorical variables for mutual information analysis. Use mi.empirical() to calculate pairwise mutual information of vegetation with geology, winter radiation and flow accumulation.

What is Mutual information? In a nutshell:

a. The spatial heterogeneity (or complexity) of a categorical map can be measured using Shannon's entropy statistic

where pj is the proportion of the map in map class j, j=1,2…u.

b. When the area is jointly categorized by two variables *x* and *y* (for example vegetation and geology), a more complex map will result unless the variables are perfectly associated. The joint entropy of the combined variables is:

where pjk is the proportion of the map where *x* is in class *j* and *y* is in class *k*.

*H(x,y)* is maximized when *x* and *y* are spatially independent. Conversely, a measure of the strength of association or "mutual information" between two mapped categorical variable is the difference between the maximum and the observed joint entropy.

c. For a large sample size *N*, the mutual information between *x* and *y* can be estimated as:

d. Here we are interested in determining which environmental variables are most strongly associated with vegetation pattern in the study area. Calculate *MI* for each environmental variable jointly with land cover.

e. As explained in Phipps (1981) or Davis and Dozier (1990), identify the variable with the highest I and then stratify the samples based on that variable. Then do one more level of the hierarchy by testing the mutual information of each remaining variable within your strata.

4) Report your work by answering the following”

1. **Summarize your results using a tree diagram like that in the papers.**

Attached.

1. **Discuss your results. How strong is the relationship between land cover and physical environmental variables?**

Of the three environmental variables we are looking at (geology, radiation, and flow accumulation), geology is the most strongly associated with the vegetation in the study area, followed by radiation and then flow accumulation. The mutual information analysis found that geology co-occurs with vegetation 16% percent of the time. While 16% may seem like a small amount, given the complexity of the environment with many possible predictor variables, this result indicates a strong relationship between vegetation and environmental variables. In other landscapes that are simpler with fewer explanatory variables, like an expansive tundra plain or homogenous boreal forest, 16% might be considered a weak relationship. Ernoult *et al.*  2003 found a strong relationship between land cover and physical variables —land uses closely followed the topography and geology. In the future, with more detailed data and better methods we will be able to increase the explanatory power of environmental variables.

1. **Compare your findings to those of Davis and Dozier (1990).**

My findings were similar to those of Davis and Dozier (1990). I found that the four geology classes accounted for16% of the information in the vegetation data. Davis and Dozier found that their nineteen land classes accounted for 18.5% of the vegetation data. Furthermore, the hierarchy of explanatory environmental variables was similar to my analysis, with geology being the most powerful explanatory variable followed by radiation (or insolation as it is referred to in the paper). Davis and Dozier did not include flow accumulation in their analysis, likely because it is a weak explanatory variable — as I found in my analysis.

1. **Summarize the strengths and weaknesses of Mutual Information Analysis for establishing land cover-environment associations? What alternatives would you consider?**

Strengths:

* Reduces the amount of intensive field sampling needed, allowing for more practical ability to test with a larger random sample than field methods alone.
* Can be easily compared with existing ecological association studies of the region in question.
* Useful for relating biodiversity to complex landscapes, especially are larger scales.
* Land classes are scalable depending on the level of detail of environmental variables used. For example, a general geology or soils map could be used for general classification, or a finer detail map that includes texture, organic content, depth, etc. can be used to create more specific classes.
* Less subjective than other classification methods.
* With improving remote sensing technology, mutual information is a practical way to study land classes of large and/or difficult to access areas.

Weaknesses:

* Dependent upon the accuracy and detail of the cartographic maps being used for the analysis. Davis and Dozier mention that a 30-m resolution DEM was too course to reliably use microtopography to construct accurate land classes.

Alternatives:

* Controlled experiments could be used to identify causal connections between land cover and environmental variables. However, this would be expensive and time intensive.
* USDA Ecomap is a comprehensive methodology, which if used would allow you to fit your landscape classification into an accepted hierarchy, rather than a unique system just for your area. However, this would only be useful if working at a large scale.
* If you are only interested in a specific species, you could create a simpler classification system. For example, which areas are suitable for species X?

## Literature Cited

Davis, F. W., and J. Dozier. 1990. Information Analysis of a Spatial Database for Ecological Land Classification. Photogrammetric Engineering and Remote Sensing 56:605–613.

Ernoult, A., F. Bureau, and I. Poudevigne. 2003. Patterns of organisation in changing landscapes: implications for the management of biodiversity. Landscape Ecology 18:239–251.

Phipps, M. 1981. Entropy and community pattern analysis. Journal of Theoretical Biology 93:253–273.

Wagner, H. H., and M.-J. Fortin. 2005. Spatial analysis of landscapes: concepts and statistics. Ecology 86:1975–1987.