# Physical environmental controls on landcover pattern

Lab Exercise 2  
ESM 215, Winter 2021  
Due by 4 PM, Wednesday Jan 20. Please submit via GauchoSpace

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 10.8)

    B. Click OK to accept the blank template. Then in the main window, open the arcmap project R:\Winter2021\ESM215\data\sy\_data\datavu.mxd.

(File -> Open -> This PC -> Bren Courses (R:) -> Winter2021 -> ESM215 -> data -> sy\_data -> datavu.mxd)

    C. The project legend includes a variety of data files, which can be found in [R:\Winter2020\ESM215\data\](file:///C:\Volumes\courses\Winter2010\ESM%20215\data\datavu.mxd)sy\_data (Before doing anything else, use Arc Catalog to copy these data to your own local working directory, including a copy of the file **datavu.mxd**. To do this, click the tab “Catalog” on the right side of the main window. The top listed folder should be “Home – Data/Sy-data”. Right click and copy that folder, then navigate in a separate File Explorer window to your H drive (named something like “yourname ([\\babylon\mesm](file://babylon/mesm)) (H:)” which is your local working directory). That way you’ll have a copy saved locally that you can make alterations to, without affecting the original class copy.)

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

**sy\_airphot12** – 1m resolution true color air photo, 2012

**sy\_geol** – A subregion of the 1:250,000 scale geologic map of CA. 94 ft. raster. A more detailed map with legend is available [here](http://www.quake.ca.gov/gmaps/GAM/santamaria/santamaria.html).

**sy\_soil –** A subregion of the1:24,000 scale soil survey map ([SSURGO](https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=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.

**sy\_dem** – Digital elevation model, 28m raster: Shuttle imaging radar topographic data. Values are elevations in meters above sea level.

**sy\_slope** - 28 m raster: slope angle in degrees derived from **sy\_dem**

**sy\_shade** -  28 m raster: shaded relief image, derived from **sy\_dem**

**sy\_allrad -** 28 m the data raster**:** annualclearsky shortwave radiation, units are watts/m2

**sy\_winrad -** 28 m the data raster**:** integrated clearsky shortwave radiation, units are watts/m2, for December-Feb.

**sy\_winrad3** – grid produced by classifying **sy\_winrad** into 3 (low, medium, high) radiation classes. Use this grid for the exercise.

**sy\_flocum** –28 m raster: flow accumulation model, derived from **sy\_dem** 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.)

**sy\_flocum3c** – grid produced by classifying **sy\_flocum** into 3 accumulation classes. Use this grid for the exercise.

**sy\_veg15**   - 28m raster: contemporary vegetation/land cover map produced (mainly) from Landsat Thematic Mapper satellite imagery. California Wildlife Habitat Types are shown here. I merged some agricultural classes to reduce the number of land cover classes to 15. [Here](https://map.dfg.ca.gov/metadata/ds1327.html) is a description of the data.

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 land use/land cover pattern (air photo) and vegetation pattern (sy\_veg15) in relationship to geology, soils, and topographic factors like elevation, slope, radiation and flow accumulation.

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

What environmental features are associated with 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. If more than one landscape is encompassed by our study area, a hierarchical analysis may be appropriate such that the area is first partitioned into separate landscapes and then environmental associations with land use/land cover are evaluated landscape-by-landscape.

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” that is useful for measuring the hierarchical association between a dependent categorical variable - in this case, mapped land use/land cover classes – and categorical independent variables (in this case, geologic classes, radiation classes, and flow accumulation classes. 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 grid **sy\_veg15** 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 (**sy\_veg15**), geology (**sy\_geol**), flow accumulation (**sy\_flocum3c**), and winter radiation (**sy\_winrad3**). *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 available to download from Gauchospace as **ex2\_sample\_dat2.csv**.

2) The sample points can be used to extract values at those locations from multiple grids (You can do so using tools in the ArcToolbox, which is an icon toward the upper right corner of the menu containing a red toolbox. Click ArcToolbox, which will open a smaller window with folders of all the tools. Then Spatial Analyst Tools -> Extraction -> Extract Multi Values to Points.) (Make sure that Customize > Extensions > Spatial Analyst is checked.) The output will be added to the attribute table for your sample points. You can open the tables and export the table to your working directory. *OR*, You can use the values already added to **ex2\_sample\_dat2.csv**.

3) This exercise can be completed using Excel or R, but it is much faster in R using the *entropy* package. One possible workflow in R is included at the end of this document.

4) In R, use the function *mi.empirical*() to calculate pairwise mutual information of landcover with geology, winter radiation and flow accumulation. You should go through at least 2 levels of analysis. The first level will, from the variable with the highest mutual information with landcover, divide the samples into classes. The second level will further subdivide samples, within each class of level 1, into classes based on the next variable (of the 2 remaining) with the highest mutual information with landcover.

4) Report your work by answering the following:

1. Summarize your results using a simple tree diagram like that in the papers.
2. Discuss your results.
   1. Which mapped variable shows the strongest association with mapped land use/land cover?
   2. How strong is the relationship between land cover and physical environmental variables? (Hint: you can measure this using the redundancy measure.)
3. Compare your findings to those of Davis and Dozier (1990).
4. Summarize the strengths and weaknesses of Mutual Information Analysis for establishing land cover-environment associations? What alternatives would you consider?

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 recursively identify additional levels of the hierarchy by testing the mutual information of each remaining variable within each stratum.

Mutual Information Analysis in R

#There are many ways you could perform a mutual information analysis of landscape data in R. #Here is one possibility.

#

#load required packages

library(entropy)

library(dplyr)

#

#read in data downloaded from gauchospace, deleting any lines containing missing data (-9999 in arcgis)

sydat <- read.csv("ex2\_sample\_data2.csv",header=TRUE,na.strings=c(-9999)) %>% na.omit()

#dim(sydat) is 9990 x 8, 10 samples with missing data removed

#

#First level MIA, calculate joint entropy of landcover with geology, flow accumulation and #winter insolation

migeol <- table(sydat$landcover,sydat$geology) %>% mi.empirical() #0.2286847

miflow <- table(sydat$landcover,sydat$flowaccum) %>% mi.empirical() #0.00543883

mirad <- table(sydat$landcover,sydat$winter\_rad) %>% mi.empirical() #0.05124749

#

#Highest association is between landcover and geology (MI=0.2286).

#

#Next, stratify the sample by geology, loop through geology classes 1 to 5,

#and calculate MI for flowaccum and winter\_rad

#

mi\_flow2 <- NULL

for(i in c(1:5)){

y <- filter(sydat,geology == i)

x <- table(y$landcover,y$flowaccum) %>% mi.empirical()

mi\_flow2 <- c(mi\_flow2,x)

}

#

#

mi\_rad2 <- NULL

for(i in c(1:5)){

y <- filter(sydat,geology == i)

x <- table(y$landcover,y$winter\_rad) %>% mi.empirical()

mi\_rad2 <- c(mi\_rad2,x)

}

#Analysis completed. You can now produce a hierarchical dendogram of the first 2 levels of the #MIA

## 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. <https://doi.org/10.1890/04-0914>