**Modeling land-use change in Loudoun County using logistic regression**

**Introduction**

Many cities/counties in the United States have been experiencing accelerated urban growth. Loudoun county has been one of the fastest growing areas in the U.S for the last two decades. Low-density residential and commercial settlements have spilled well beyond recognized urban boundaries. The urban expansion not only consumes forest and agricultural lands, but also breaks the natural landscape into fragments. Such development may cause many negative environmental impacts with respect to water quality, soil erosion, air pollution, local climate, and biodiversity.

For this lab, we’ll predict future land-cover change locations by combining spatial statistical models with spatially explicit data in the R programming environment. We will prepare GIS datasets to build a logistic regression model and then use the logistic model to estimate the probability of land change for each 30m resolution pixel.

**Logistic regression**

For our logistic regression model, each “observation” is a grid cell (30m pixel). The dependent variable is a binary presence or absence event, where 1=forest-to-urban change and 0=stable forest (forest-to-forest), for the period 1992–2001. The following logistic function gives the probability of forest loss as a function of the predictors:

*p̂ = exp(b0 + b1x1 + b2x2 + ... bnxn)/(1 + exp(b0 + b1x1 + b2x2 + ... bnxn))*

Model predictors (xi) can be user selected biophysical (e.g., elevation, slope, distance to roads) and socio-economic (e.g., population, income) variables.

**Data preparation**

*The dependent variable layer*: Please use land cover maps (loudoun1992.tif and loudoun2001.tif) as input to identify *forest-to-urban change* and *stable forest* pixels.

For this dependent variable raster layer, you need to code all *forest-to-urban change* locations as 1 and *stable forest* locations as 0. All other locations should be coded as NoData.

*Predictors: Elevation, Slope, Distance-to-road*

The DEM (30m resolution) can be downloaded through library 'FedData'. Please install this library. A quick demo on how to use this library is available through the instructional video for this lab.

The Road layer is available in the lab6 folder [Roads.shp].

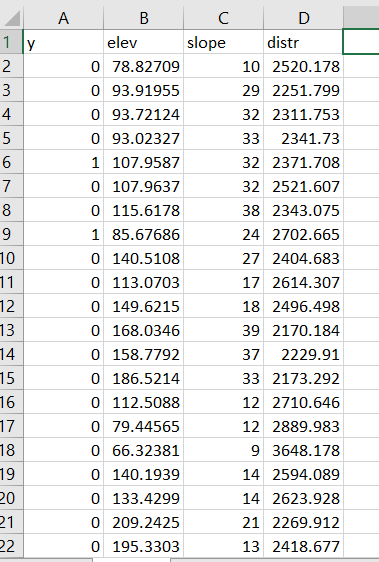
You’ll generate three raster layers (elevation, slope, distance-to-road) in the R environment. These raster layers should have the same spatial extent (col, row number) as those of land cover maps. R raster library is very useful here. Some functions particularly important include: projectRaster(), terrain(), crop(), resample(), rasterize(), and distance().

**Select a subset of total pixels for logistic model development**

Within the *forest-to-urban change* and *stable forest* area mask, identify about 5% of total pixels for logistic model development. You may generate Random Points to support this task. About 5% of pixels can be randomly chosen to support your model development.

sampleRandom() is useful for this step.

For these randomly selected locations or points, please extract pixel values from the dependent variable raster layer (y), elevation layer (x1), slope layer(x2), and distance-to-road raster layer (x3). Combine the extracted pixel values and generate a dataframe with four data columns (y, elev, slope, distr). Save this csv file as alldata.csv. Your csv file should look like the following template.

****

**Using R program to build the logistic regression model**

 In the RStudio console, please use the following command line to import the alldata.csv file (this step is not essential if you already have a dataframe named as alldata):

alldata <- read.csv(file="alldata.csv", header=TRUE, sep=",")

R makes it very easy to fit a logistic regression model. The function to be called is glm(). In the Rstudio console, use the following command line to build the logistic regression model:

glm<-glm(formula = y ~ elev + slope + distr, family = "binomial", data = alldata)

Use summary command to see the results from the logistic regression model:

summary(glm)

From the glm output, you can see that the coefficients are estimated for model *intercept* and predictors of *elev*, *slope*, and *distr*. Now you now predict the ‘probability’ of forest loss for all pixels.

Please search online resource on how we generate predictions for all observations, given that glm model is ready.

**Generate the final probability/suitability map**

The previous step provides a map showing the ‘probability’ of forest loss for all pixels. Clearly, we need to mask out all non-forest pixels, because the probability estimates are not applicable to existing urban, water, and agricultural lands. For your final probability map, please make sure the locations of non-forest pixels in the 2001 land cover map have probability value of 0.

**R coding**

All above GIS analytical steps can be streamlined in R environment. You need **raster** and **rgdal** libraries for data import and analysis. Please implement this land change modelling project in R.

**Deliverables**

* upload your R script to the Canvas
* upload your final probability map (tiff format) to Canvas