Workshop 4_Species Distribution Models

Shalimar G. Moreno

2/7/2020

Objectives

The primary objectives of this analysis is to fit a Species Distribution Model to our data set by:
1. Installing packages for species distribution modeling.
2. Running species distribution models using `bioclim` approach.
3. Visualizing model predictions on a map.

Methods

Species Information

This project is looking at species distribution modeling to estimate the similarity of the conditions at any site to the conditions of the known locations of a phenomenon, in this case looking at species range. This is important so as to predict ranges of our model species and creating a range map which is shown. The species we are looking at is saguaro (*Carnegiea gigantean*), a tree-like cactus species (Fig 1.). It is native in US States Arizona and California as well as in the Mexican State of Sonora. This species is listed as IUCN least concern, therefore not a focus of species conservation but for this project it is our model species. Factors that may affect saguaro distribution is elevation, which is a limiting factor to its environment, and temperature, which saguaros are sensitive to extended frost or cold temperatures. It is considered the northernmost columnar cacti in the Americas which is interesting to see if our species distribution will predict accurate range ranges of our cactus.

Fig 1. The saguaro (*Carnegiea gigantean*), arborescent cactus species. (C) William Clarke Casey

Statistical Analysis (Bioclim)

To get a range map of our species we will be using data that is a provided at:
`https://tinyurl.com/saguaro-obs`
which provides coordinates (latitude/longitude) of observed and known range of saguaro. To begin, we created a workspace on R and installed R packages needed: `dismo`, `maptools`, `rgdal`, `raster`, and `sp`. We then create the map using the `bioclim` algorithm. Bioclim is highly used for species distribution modeling as it is easy to use and understand. It computes the similarity of a location by comparing the values of environmental variables at any location to a percentile distribution of the values at known locations of occurrence. After creating the map we can then use a predict function to create predicted values of where we expect to see our species. To begin, we need to clean our data: remove any NA values that will create errors, switching to read longitude and then latitude order, and checking to map the plotted points. To build our model we use the bioclim function and cropping it to the area we want to work with, where our species is present. We used the 'predict.presence' function to generate an object that has the model's probability of occurrence for saguaros. We can then plot our predicted data on the map and see what the map creates. The map at first didn't seem accurate because we have data showing more observed values, and so we created a presence/absence approach: creating a set of background points at random with as many points as we have observations. We use plot to plot the map with the observed points and points to add the background points. We then build the model using part of our data and reserve 20% of the data for testing, using the `kfold` function in the dismo package. To test the model, we establish a cutoff of occurrence probability to determine. The next step is to separate our observations into a training data set and a testing data set and rebuilding the model to evaluate it using the observation data and the pseudo-absence points. We created a threshold of 0 and 1, 1 being where saguaro can be found and then plot against the threshold to get a range of where they have been found and can be found.

Results

Using the bioclim function we were able to plot the observed values of the saguaro cactus in our region. We then did a presence and pseudo-absence approach to plot additional points where the saguaro either is or isn't there. From this data we created our treshold and then plotted the range of where the saguaro is able to be found which is highlighted in the green area (Fig. 2).
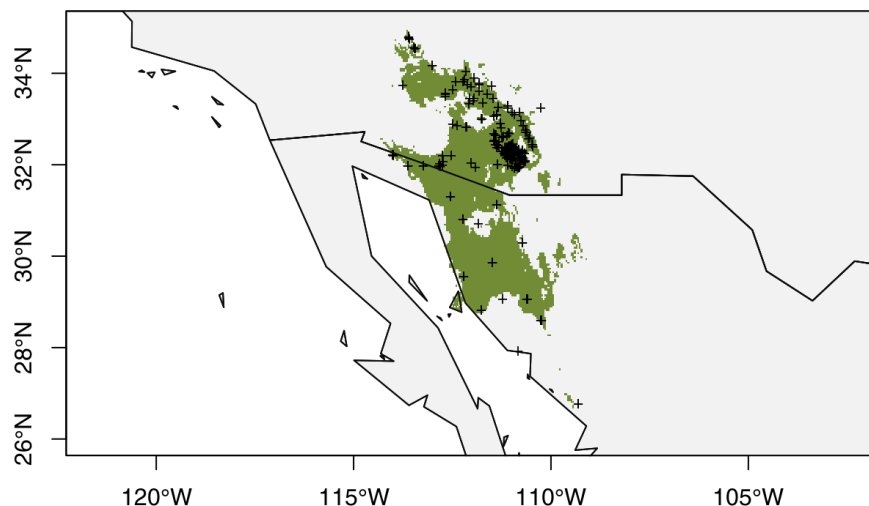

Fig 2. The current known range for the saguaro (*Carnegiea gigantean*) cactus.

In plotting our range we created the predicted area and the threshold of our predicted area to narrow the green range. The data of these raster layer is shown in table 1. Our final map is drawning from our categorical classification as to whether the landscape will be suitable or not for our species of interest, the saguaro. In

conclusion we generated the model, determined the threshold and visualized the results.

Table 1: Predicting presence threshold

RasterLayer Dimensions	Output
$class$:	$RasterLayer$
$dimensions$:	$216, 144, 31104 (nrow, ncol, ncell)$
$resolution$:	$0.04166667, 0.04166667 (x, y)$
$extent$:	$-115, -109, 26, 35 (xmin, xmax, ymin, ymax)$
$coord.ref.$:	$+proj = longlat + datum = WGS84 + ellps = WGS84 + towgs84 = 0, 0, 0$
$datasource$:	$inmemory$
$names$:	$layer$
$values$:	$0, 1 (min, max)$

Discussion

Using the `bioclim` function is interesting because although it has its limitations it was pretty streamlined to plot our observed values and to create predicted values of our data. We were able to easily plot these maps as we worked our way through the coding. Following the above methods we were able to get our final map of predicting the range of the saguaro cactus using observed and data that was easily accessible online.