Workshop 4: Species Distribution Models

Kristy Sullivan

2/4/2020

Objectives:

The objectives of this exercise are to construct a species distribution model of saguro cactus (*Carnegiea gigantea*) using the bioclim approach, and then visualize predicted distribution on a map.

Methods:

Species information:

Saguro is arborescent cactus species native to the Sonora dessert in Arizona. These cacti have a very long lifespan, usually exceeding 150 years. They can serve as nesting habitats for many dessert bird species as well as a food source when fruiting. High predation of the fruits and harsh dessert conditions prevent most seeds from germinating successfully, limiting the dispersal of saguro.

Statistical Analysis (Bioclim):

Required R packages dismo, maptools, rgdal, raster, and sp were installed and loaded to the library. Bioclimatic variable (bioclim) data were then downloaded using the getData function, along with the saguro presence data using read.csv. The boundaries of these data were then stored within geographic.extent and the occurence data test plotted on a map of the area. The bioclim data were then constrained to the geographic extent of saguaro and the species distribution model built over this area with predicted presence. Pseudo-absence data were then created at random to offset the presence-only data (*Figure 1*). The kfold function was then used to group model training and testing data and a model was constructed from these data. A final map was created with the model results (*Figure 2*).

Results:

Saguro presence data appears most localized to Southern Arizona at a minimum latitude of 26.78 and a maximum of 34.80 (*Table 1*). Pseudo-absence points distributed in the area exhibit a randomized pattern to optimize model training and testing (*Figure 1*). The final results of this assignment suggest that there is a larger than observed population of saguro cacti in Northern Mexico than expressed in the observed data (*Figure 2*).

Table 1: Summary statistics of GBIF ID, latitude, and longitude coordinates for observed Carnegiea gigantea.

Statistic	GBIFID	Latitude	Longitude
Minimum	8.910e+08	26.78	-114.0
1st Quartile	1.453e+09	32.17	-111.4
Median	1.571e+09	32.28	-111.1
Mean	1.575e+09	32.16	-111.3
3rd Quartile	1.677e+09	32.38	-111.0
Maximum	1.806e+09	34.80	-109.3

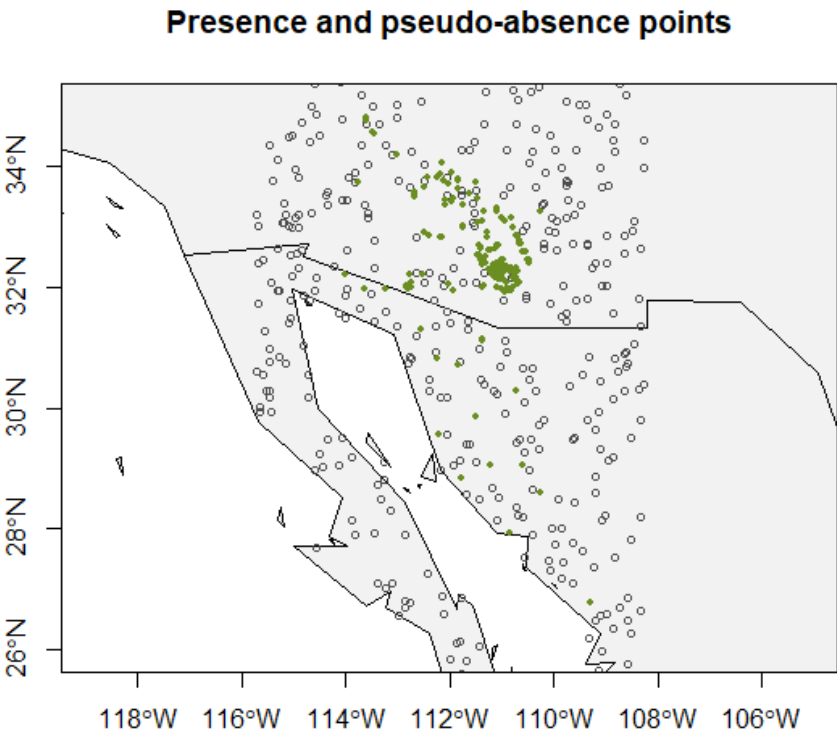**Presence and pseudo-absence points**

Figure 1: Pseudo-absence data (white dots) generated at random to offset observed presence-only data (green dots).
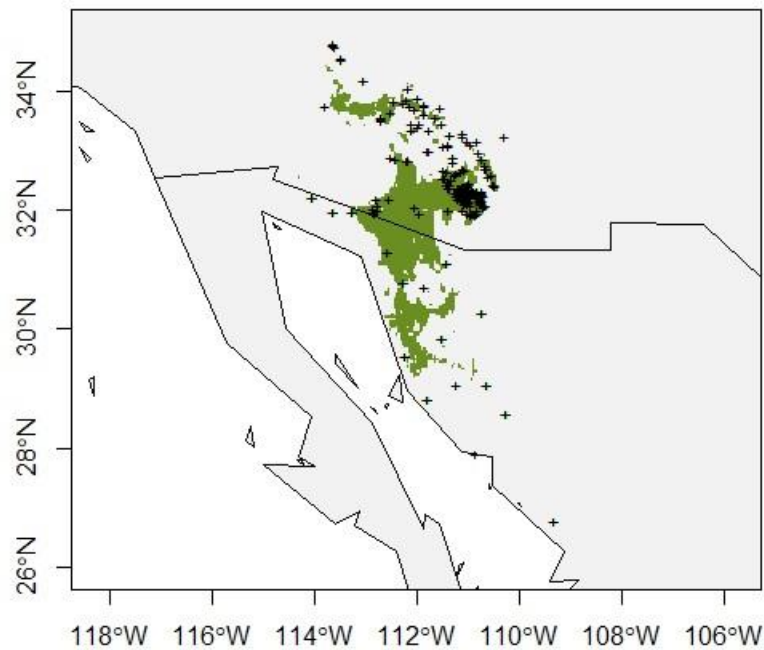
Figure 2: Finalized observed saguro distribution (black crosses) and predicted distribution (green).

Discussion:

The ranges of saguro cactus (*Carnegiea gigantea*) were visualized on a map of the Southwestern United States using the bioclim R packages. The final predicted distribution map suggests there may be more saguro cacti in Northern Mexico than observed for this dataset. Due to the use of pseudo-absence points, the predicted range of saguro may shift each time the model is run and the points are regenerated. This is something that may be neglegable depending on the desired resolution of the data, if better resolution is needed, another approach may be better suited to this dataset.