Species Distribution Model on Saguaro (*Carnegiea gigantea*)

Katherine Castrillon

01/31/2020

Objectives

Species distribution models (SDM) are utilized to estimate the similarity of conditions at any site to the conditions at the locations of known occurrence, or of non-occurrence, of a phenomemon. A SDM was created with available data from GBIF, the Global Biodiversity Information Facility, on observations of Saguaro utilzing the algorithim: BIOCLIM, known as a classic climate-envelope-model (Booth et al.,2004). The Saguaro observation dataset is composed of latitude and longitude coordinate data. Utilizing the coordinate data, a range map of Saguaro was created in RStudio to visualize model predictions on a map. A random sampling method, known as pseudo-absence was used, and a post hoc evaluation was catalyzed to analyze the model.

Methods

Species Information:

The Saguaro, also known as Giant Saguaro, belongs to the Cactaceae family. The Saguaro is a perennial cactus that grows up to 50 feet tall and can weight up to 9 tons. It distributed in the southwest region of North America, and Central America, in Arizona- Yavapai & Mohave to Graham, Santa Cruz, Pine & Yuma Cos., California, and Mexico.

Statistical Analysis:

1. Two folders were created to keep organization in workspace.
    - data
    - output
2. RStudio packages were installed and loaded to the workspace:
    - `dismo`
    - `maptools`
    - `rgdal`
    - `raster`
    - `sp`
3. Used `getData` function to download the bioclimatic variable data.
4. Read Saguaro observations from data folder and check if data loaded correctly.

5. Find latitudinal and lognitudinal boundaries by plotting data points on map.
6. Used bioclimatic variables to create a model of Saguaro to a specific geographic area using the `predict` model from the `dismo` package.
7. Utilized psuedo-absence point approach on `bioclim` data files.
    - Determine spatial resolution of points.
    - Restrict the sampling area to the general region of the observations of Saguaros.
8. Initiate post hoc evaluation of the model
9. Create training data and testing datasets utilizing the `kfold` function in the `dismo` package.
10. Use `group.presence` vector with observed data to seperate observations into a training data set and a testing data set
11. Repeat step 10 for psuedo-absence points.
12. Rebuild model using `bioclim` function using observations in training data stored in `presence.train` .
13. Evaluate model created utilizing the observation data and the psudeo-absence points that were reserved for testing.
14. Use `threshold` function to determine minimum threshold for Saguaro presence.
15. Give map noticable range by giving the two values 0, 1, from the `values` field a corresponding color.
16. Use threshold created from `threshold` function to create a map with predicted range of Saguaro.

Results

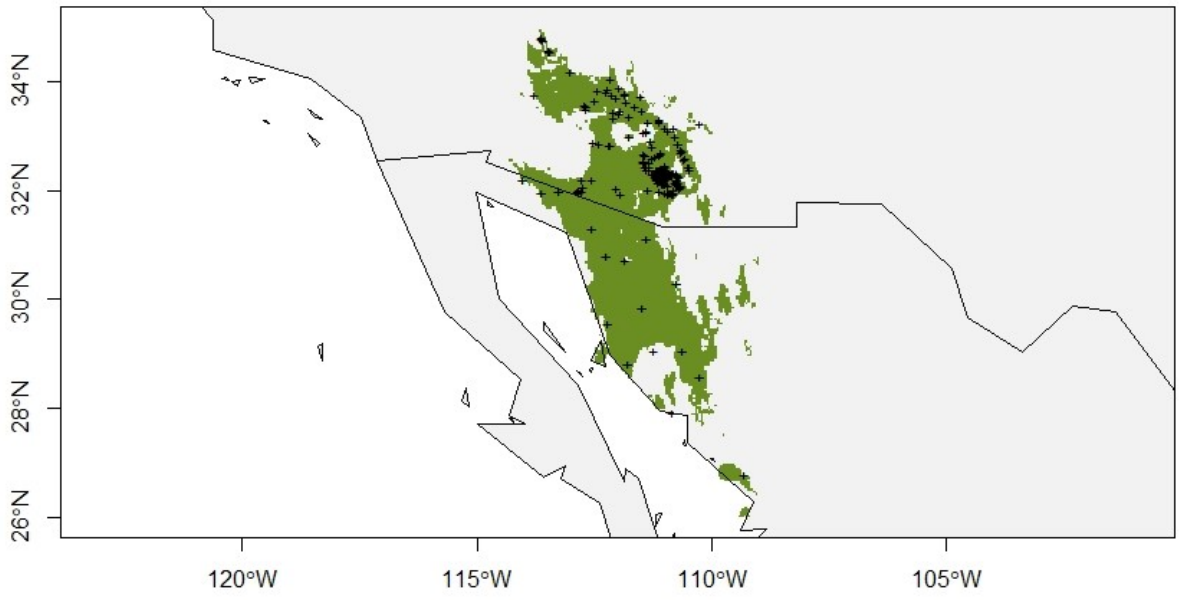

Figure 1. Species Model Distribution of Saguaro cactus along the southwest region of North America, and Central America

Discussion

The map presents a categorical classification of whether a particular point on the landscape will be suitable or not for the Saguaro cactus, shown in **Figure 1**. The green indicates probability of Saguaro occurrence across the map, the black points on the map displays a known Saguaro observation.

In creation of the SDM for the Saguaro cactus population, there were a few issues that required troubleshooting when working with the code. In the methodology, issues arose at #6. Data that was not in correct format in the `obs.data` dataframe requried modification. The argument required the `p` to be composed of two columm matrix. `Obs.data` was modified to two columns, being latitude and longitude dropping the inital first column that was the GBIF identifier. Another error based on plotting latitude and longitude was troubleshot, as in R x-axis refers to the latitude and y-axis refers to longitude, in reality with modeling Saguaro observations, longitude should reflect on the x-axis and latitude should refer to the y-axis.

Annotated Bibliography

Booth, T. H., Nix, H. A., Busby, J. R., & Hutchinson, M. F. (2014). BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. Diversity and Distributions, 20(1), 1-9.

Plant Database. (n.d.). Retrieved from https://www.wildflower.org/plants/result.php?id_plant=cagi10 (https://www.wildflower.org/plants/result.php?id_plant=cagi10)