Exercise 7: Maxent

Adam B. Smith, Danielle Svehla, & Camilo Sanín

July 19, 2017

This tutorial is a brief aside that explains:

1. **How** Maxent does its calculations which in turn are important for understanding
2. **What kind of data** Maxent can use (i.e., presence/absence versus "presence-only" data") and
3. **Tuning Maxent** to get the "best" model.

Species distribution models have been called "black boxes" into which you feed your data and out of which come a magic prediction. This is unfortunate, since the algorithms were created by humans and are therefore understandable.

I recommend these resources for a better understanding of Maxent:

[Phillips, S.J., Anderson, R.P., and Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231-259.](http://dx.doi.org/10.1016/j.ecolmodel.2005.03.026)  
The original paper introducing the Maxent species distribution model. Does not actually explain the inner mathematics of the model in great detail!

[Phillips, S.J. and Dudík, M. 2008. Modeling species distributions with Maxent: New extensions and a comprehensive evaluation. Ecography 31:161-175.](http://dx.doi.org/10.1111/j.0906-7590.2008.5203.x)  
Description of the Maxent tuning process that is now bundled with the algorithm.

[Warren, D.L. and S.N. Siefert. 2011. Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. Ecological Applications 21:335-342.](http://dx.doi.org/10.1890/10-1171.1) How to tune Maxent using AICc (we'll do this later).

[Elith, J., S.J. Phillips, T. Hastie, M. Dud?k, Y.E. Chee, and C.J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17:43-57.](http://dx.doi.org/10.1111/j.1472-4642.2010.00725.x) A technical explanation of how Maxent works.

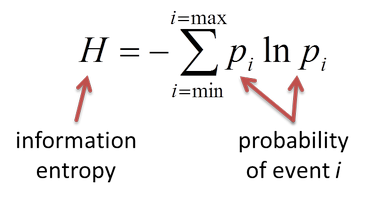
[Merow, C., Smith, M.J., and Silander, J.A. Jr. 2013. A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. Ecography 36:1058-1069.](http://dx.doi.org/10.1111/j.1600-0587.2013.07872.x) The best explanation of how Maxent works and how to use its user-tunable parameterizations.

[Phillips, S.J., Anderson, R.P., Dudík, M. Schapire, R.E., and Blair, M.E. 2017. Opening the black box: An open-source release of Maxent. Ecography 40:887-893.](http://dx.doi.org//10.1111/ecog.03049) The new, open-source version of Maxent--"Maxnet".

[Harte, J. 2011](https://global.oup.com/academic/product/maximum-entropy-and-ecology-9780199593422?cc=us&lang=en&). Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics. Oxford Univ Press. An engaging text on the use of maximization of information entropy in macroecology--covers much more than the Maxent distribution/niche model.

## What is "information entropy"?

Information entropy, akin to physical entropy, is a measure of order in a packet of information. Given a set of probabilities, its calculation is the same as the Shannon diversity index:



Shannon diversity/information entropy

The Maxent algorithm maximizes *H* for a given set of probabilities it attempts to calculate. These probabilities are related to the environmental suitability scores Maxent provides as output.

In this context entropy maximization is a kind of equation solver--i.e., it is akin to reducing the sum of the squared residuals in normal linear regression or maximization of likelihood for many kinds of models. We will not explore the explicit mathematics here, but just as it is important for you to understand that linear regression minimizes the sum of residual squares, understanding the process by which Maxent estimates environmental suitability will help you interpret its output and troubleshoot problems.

The Maxent algorithm proceeds in this fashion:

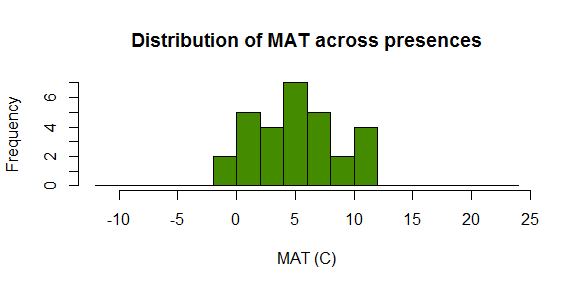
1. Calculate the distribution of each environmental variable across the species' presences. This is the probability that, given the species is present, a variable has a given value.
2. Invert this probability using Bayes' Theorem to estimate the environmental suitability of a site given its environmental values.

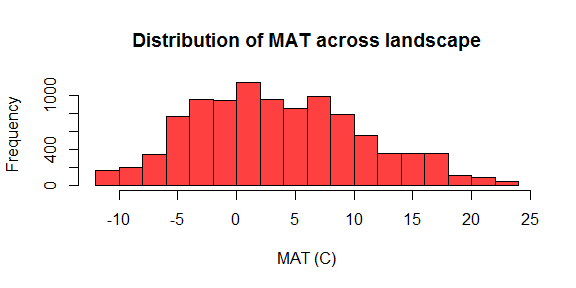
You probably do Maxent-like estimates in your head yet don't recognize it. For example, what if I said I am standing near a polar bear in the wild... and then asked you to guess the mean annual temperature of where I was? You may not have a good idea of the answer, but you know it's probably not 30C or higher. In this case you:

1. First thought about the kind of environments polar bears live in (cold)... i.e., you informally estimated the distribution of temperature across places where polar bears tend to live.
2. Informally inverted this information to estimate the temperature where I was given that a polar bear apparently loved near there. (Careful proofreaders have told us they liked the typo in the previous sentence and that they recommend keeping it.)

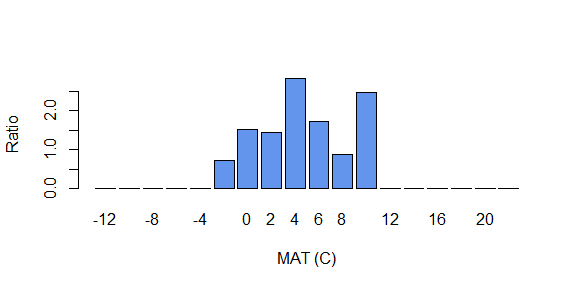
How does Maxent do this? Again, we'll use very informal language to explain.

In step #1 Maxent first looks across all the presence sites you supplied it and in a sense generates a histogram for each predictor variable across these sites. Here is the distribution of MAT across the presence sites of the Columbian ground squirrel:



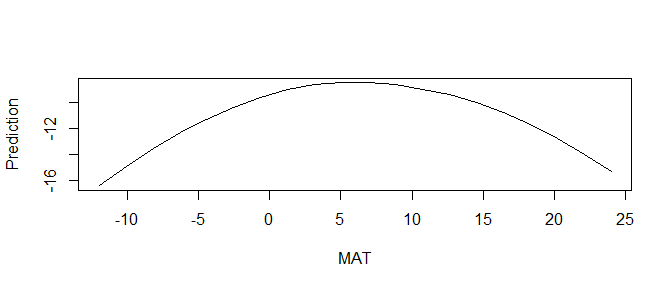
It then compares this distribution to the distribution of the same variable across the landscape (across the background sites). Here is the distribution of MAT across the study region used in the Columbian Ground Squirrel models: 

Then, figuratively speaking, Maxent divides the values in the first histogram by the values in the second histogram. This gives a relative measure of the probability of a presence given a specific value of the environmental variables at a site. Here are the values in the first histogram above divided by the values in the second:



This ratio is *proportionate* to the predicted probability of presence (or rather, the environmental suitability of a sites) given the site's particular environmental values (the x-axis values). The ratio is estimated using a set of "feature" functions that act like polynomial terms (linear, quadratic, 2-way interaction) plus two more that create either a step-like response or a flat-then-linear or linear-then-flat (hinge) response. Multiple features can be applied to any one predictor, enabling Maxent to estimate very complex shapes.

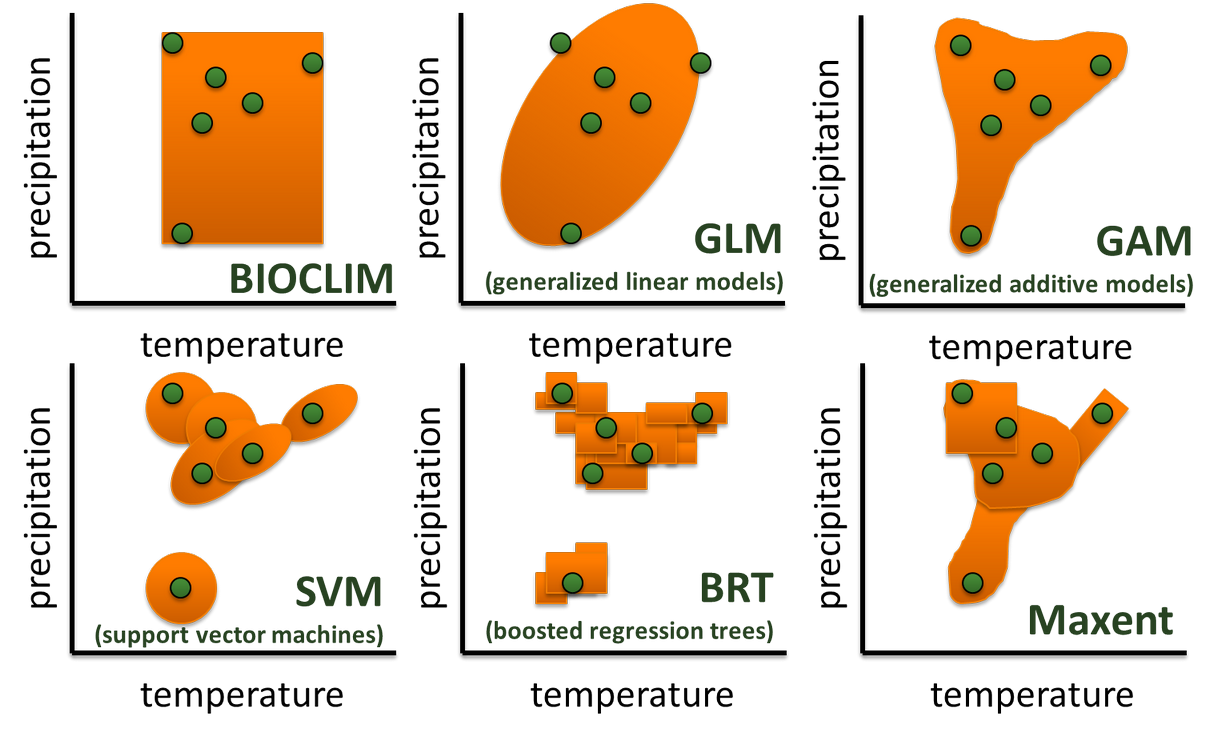
Below is the Maxent prediction for this example. You can see that the peak of the prediction is at the peak of the ratio of MAT at presence to background sites.



Left to its own devices, Maxent would fit the data as exactly as possible. However, this is considered bad form because it reduces the ability of the model to predict to data not used to train the model. As a result, Maxent's fitting algorithm is held somewhat at bay through a process called **regularization** which enables Maxent to fit its functions somewhat approximately. The regularization process has been tuned using a massive data set for hundreds of species (see second reference above).

Along with the free software bundling Maxent, these aspects have led it to be the most popular species distribution modeling algorithm ever.

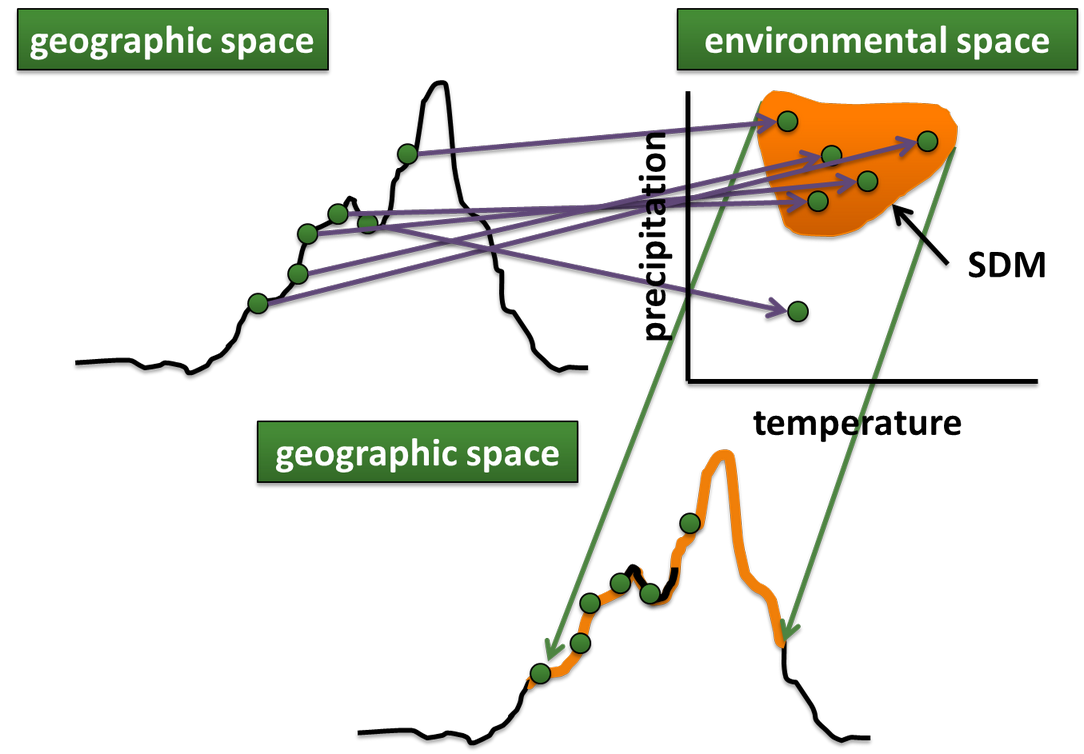
It is important to remember that all species distribution models (Maxent included) do their operations in environmental space, not geographic space. The maps these models produce are simply manifestations of the model fitting that is done in the space defined by the predictors you chose. Again, in a simplified way, these models are using different methods to draw a shape that encompasses the presences in environmental space. The form of this shape depends on the model algorithm. Some models (like Maxent) are fairly flexible and some are not.



How SDMs work

"Shapes" drawn around presence sites in environmental space (here, temperature and precipitation) by different modeling algorithms. Maxent is one of the more flexible approaches.

Once the algorithm comes to a solution in environmental space, predictions can be made. Predictions are the process of mapping the environment from a site into environmental space, seeing where it lies in relation to the algorithm's solution, then mapping this back onto geographic space.



Mapping from geographic to environmental to geographic space by an SDM

## Reflection

1. What would happen if you (erroneously) used absence sites instead of background sites to train a Maxent model?
2. Look at the figure immediately above--specifically, the mapping between geographic space and environmental space (the top of the figure) and then environmental space back to geographic space (arrows pointing from the top right to the bottom part of the figure). Is there a one-to-one mapping between these spaces (i.e., is a geographic site represented by one and only one site in environmental space, and vice versa)? How would this affect models where there is strong local adaptation among populations of the species?