# ENM Settings Overview

## University of Florida

Created by Shelly Gaynor

FLORIDA MUSEUM

UF | UNIVERSITY of FLORIDA

iDigBio
Integrated Digitized Biocollections

NSF

BiotaPhy

# Ecological Niche Modeling

**MaxEnt**

- uses the **principle of maximum entropy** on presence-only data to predicted the species' potential geographic distribution (or niche)

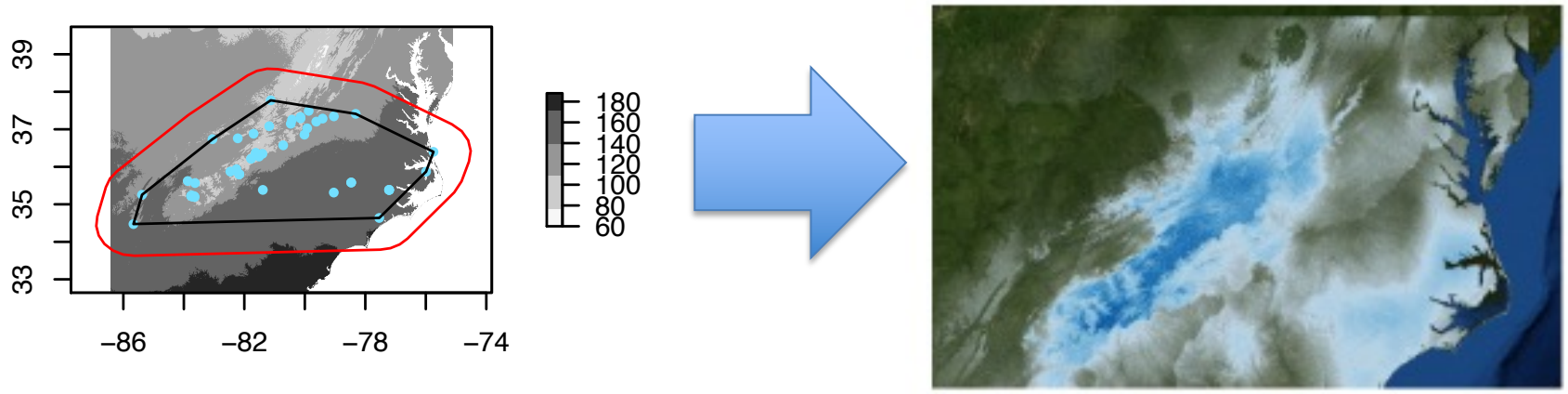



Figure from Gaynor et al. 2018

# MaxEnt

- MaxEnt **used only presence and pseudo-absences** for model development
- No ”best” algorithm for modeling (Qiao et al. 2015, Methods Ecol Evol), but MaxEnt generally performs well across evaluation criteria (Aguirre-Gutierrez et al. 2013, PLoS One)

**What does it do?**

- Makes predictions about the probability that conditions are suitable for a given taxa to occur (does not calculate a true probability of occurrence!)
- Models should be chosen that are as similar as possible to prior expectations while also being consistent with the data.

**Important notes:**

The grids must all have the same geographic bounds and cell size (i.e. all the ascii file headings must match each other perfectly)

# A Brief Tutorial on Maxent

By Steven J. Phillips, AT&T Research

This tutorial gives a basic introduction to use of the MaxEnt program for maximum entropy modelling of species' geographic distributions, written by Steven Phillips, Miro Dudik and Rob Schapire, with support from AT&T Labs-Research, Princeton University, and the Center for Biodiversity and Conservation, American Museum of Natural History. For more details on the theory behind maximum entropy modeling as well as a description of the data used and the main types of statistical analysis used here, see:

**https://biodiversityinformatics.amnh.org/open_source/maxent/Maxent_tutorial2017.pdf**

# MAXENT SETTINGS SUMMARY

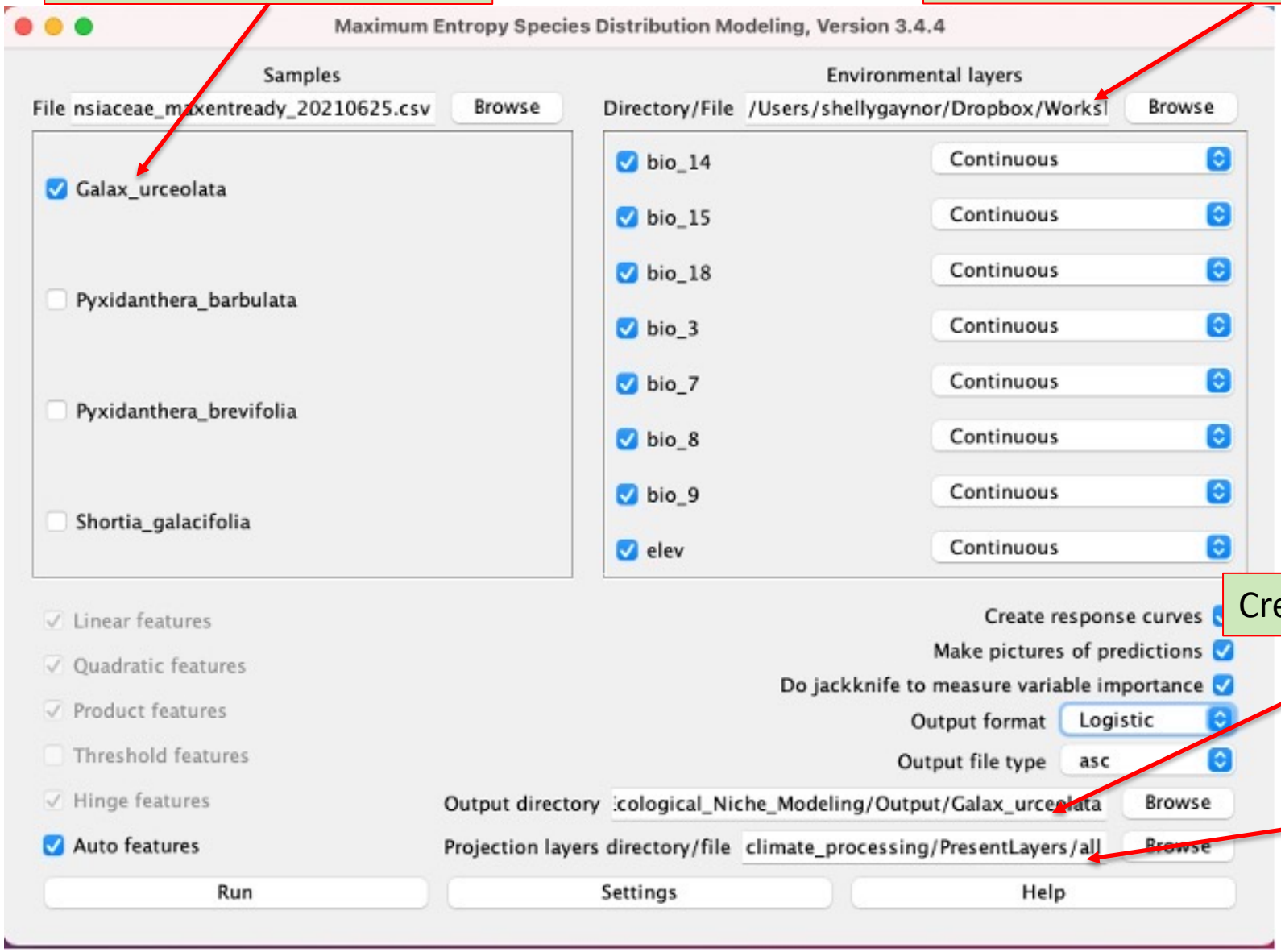
# INPUT

Occurrence Records

Only one species at a time

Training layers

Select species specific training layers




Create output directory

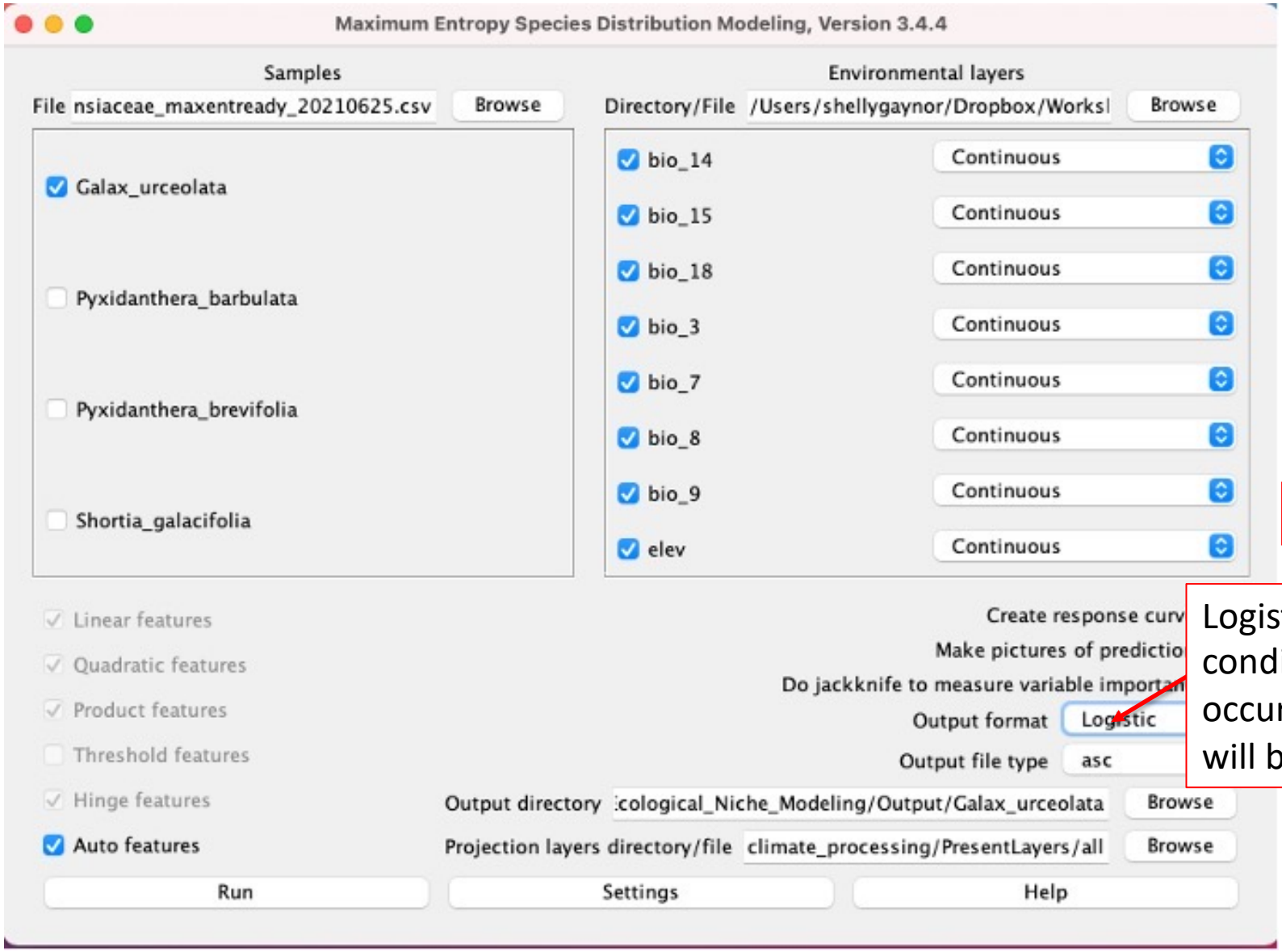
Shared layers

Select projection layers

# OUTPUT SETTINGS

| Output Type | Math | Meaning |
|---|---|---|
| Raw | Entropy value = h | Not interpretable |
| Cumulative | % of max raw value = r | Not interpretable |
| Logistic | $\frac{\exp(h) * r}{1 + (\exp(h) * r)}$ | Conditional probability of occurrence. |
| Cloglog | $1 - \exp(-\exp(h) * r)$ | Probability of presence. |

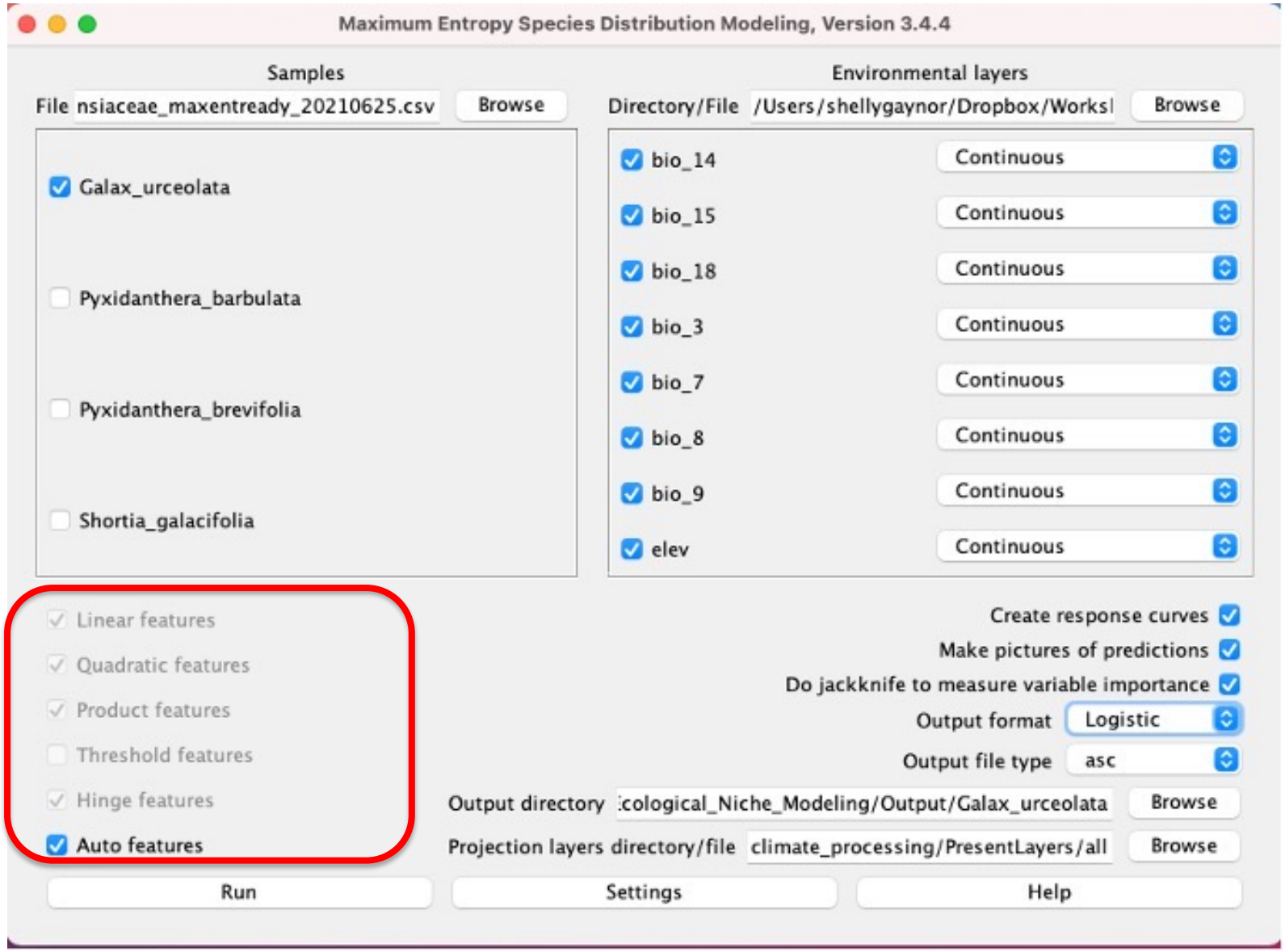
# OUTPUT SETTINGS




Select logistic

Logistic calculates the conditional probability of occurrence. The probability will be between 0 - 1.

# BASIC SETTINGS



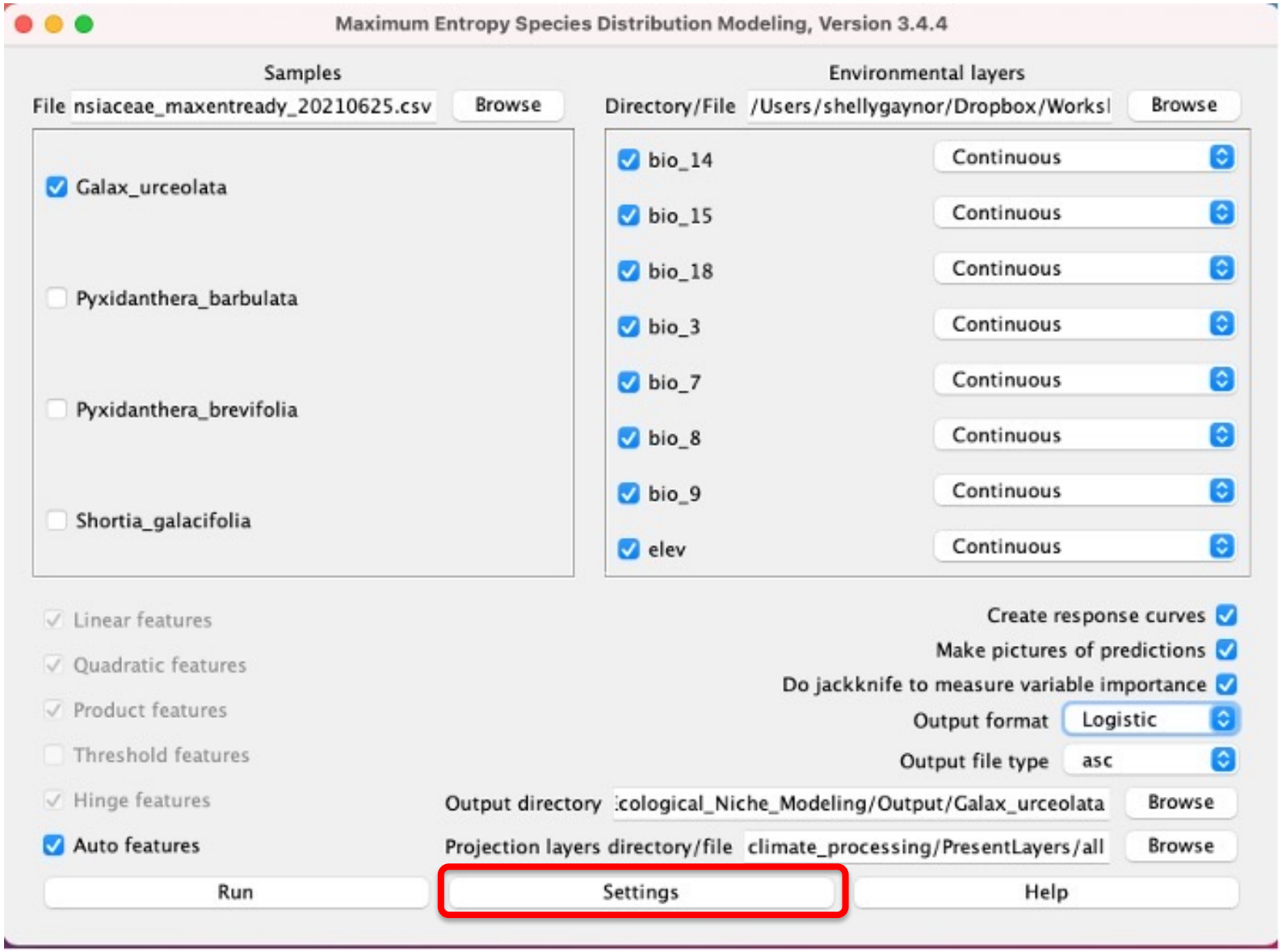
# FEATURE CLASS

= used to transform the original predictor

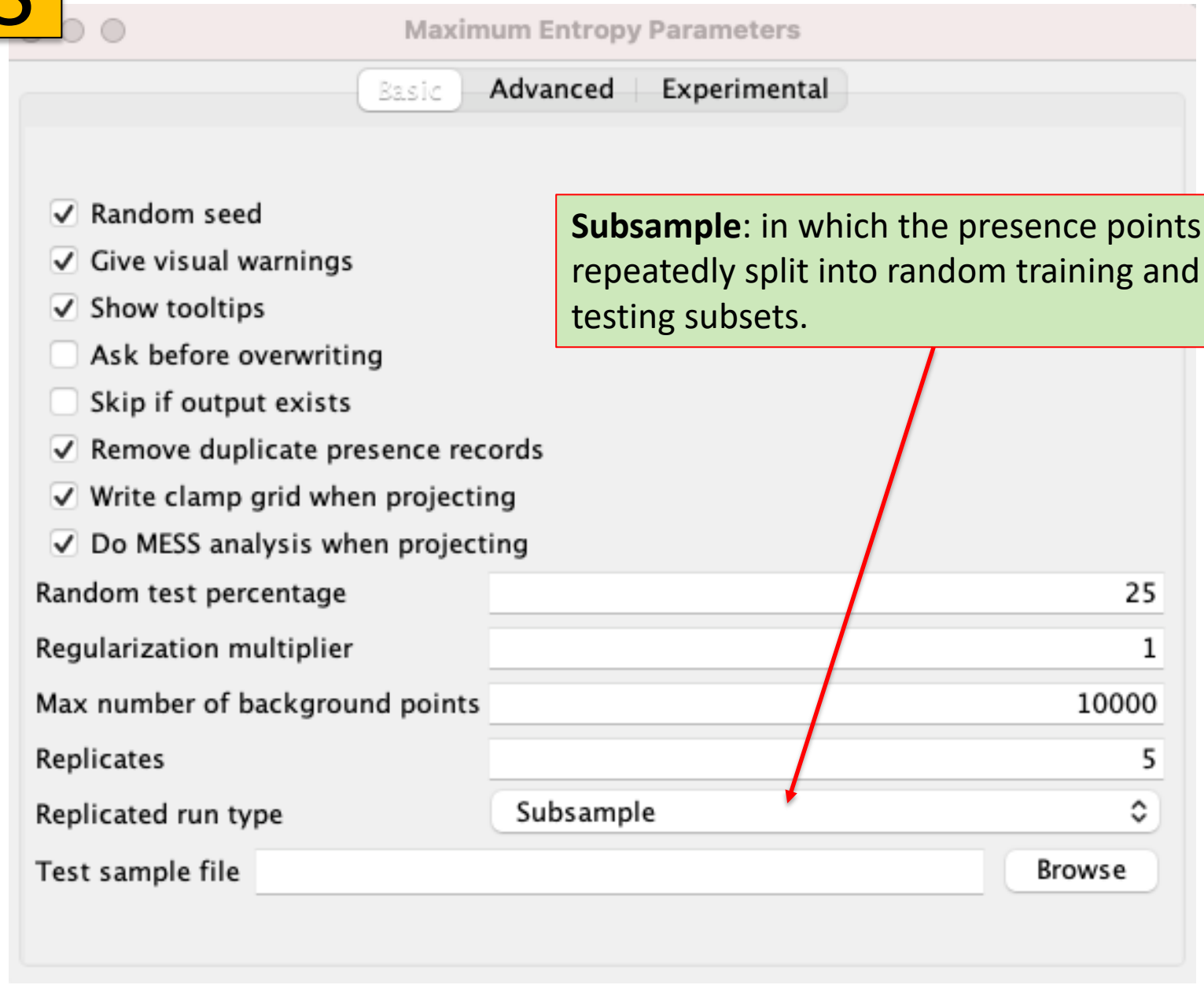
| Feature type | Interpretation | Constraint | Shape |
|---|---|---|---|
| Linear | Continuous variable | The *mean* of each environmental variable at an unknown location should be close to the mean of that variable in known occurrence locations. |  |
| Quadratic | Square of the variable | The *variance* of each environmental variable at an unknown location should be close to the variance of that variable in known occurrence locations. |  |
| Product | Pairs of continuous variables – allows for interactions | The *co-variance* of two environmental variables at an unknown location should be close to the co-variance of those variables in known occurrence locations. |  |
| Hinge | As threshold type, but response after the threshold (knot) is linear | The mean above the knot of each environmental variable at an unknown location should be close to the mean above the knot of that variable in known occurrence locations. |  |

Slide modified from M. Mabry

# BASIC SETTINGS



# SETTINGS



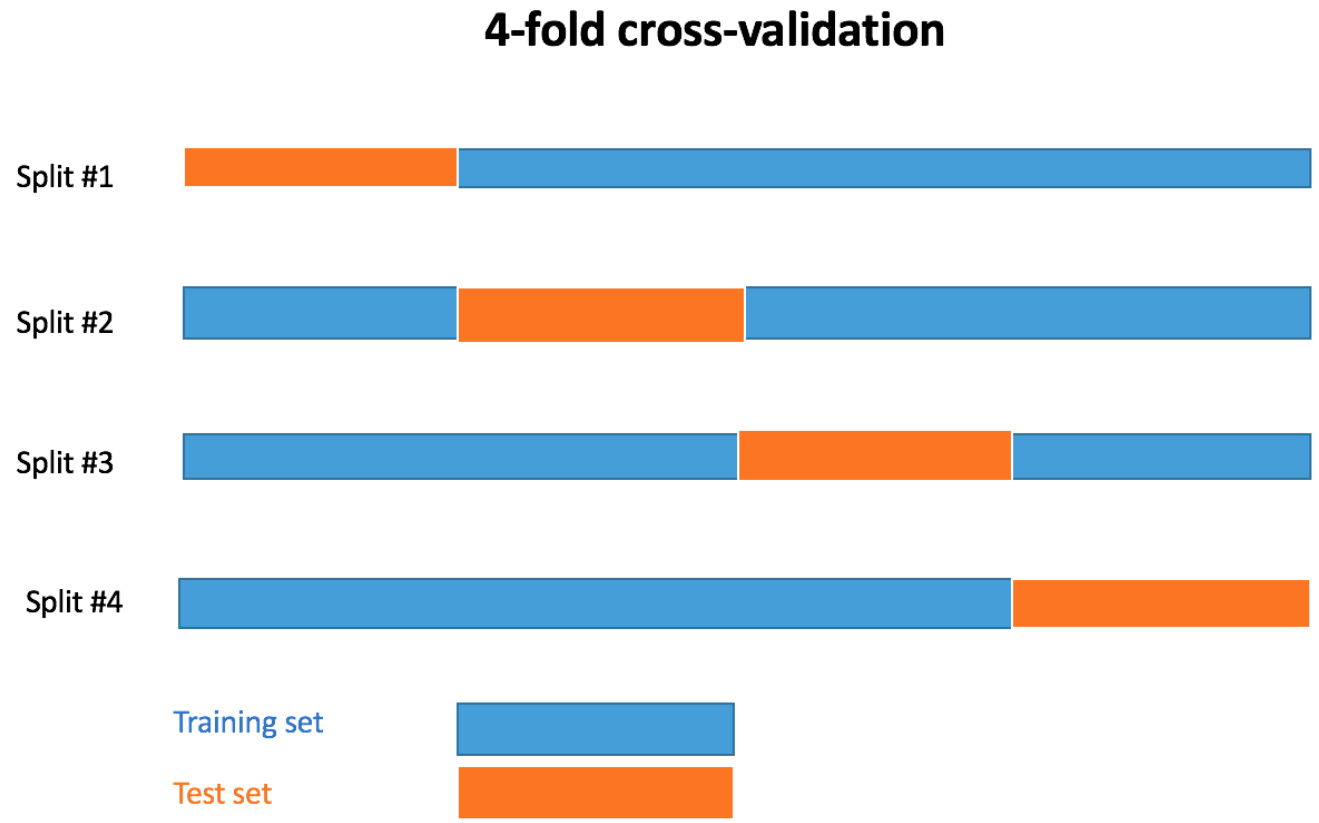

**Subsample**: in which the presence points are repeatedly split into random training and testing subsets.

# REPLICATION SETTINGS

| Replication Type | Definition |
| --- | --- |
| Subsample | Presence points are repeatedly split into random training and testing subsets. |
| Bootstrap | Presence points are sampled with replacement for equal size training and testing subsets. |
| Crossvalidate | Presence points are repeatedly split into equal-sized ”folds” of random training and testing subsets. |

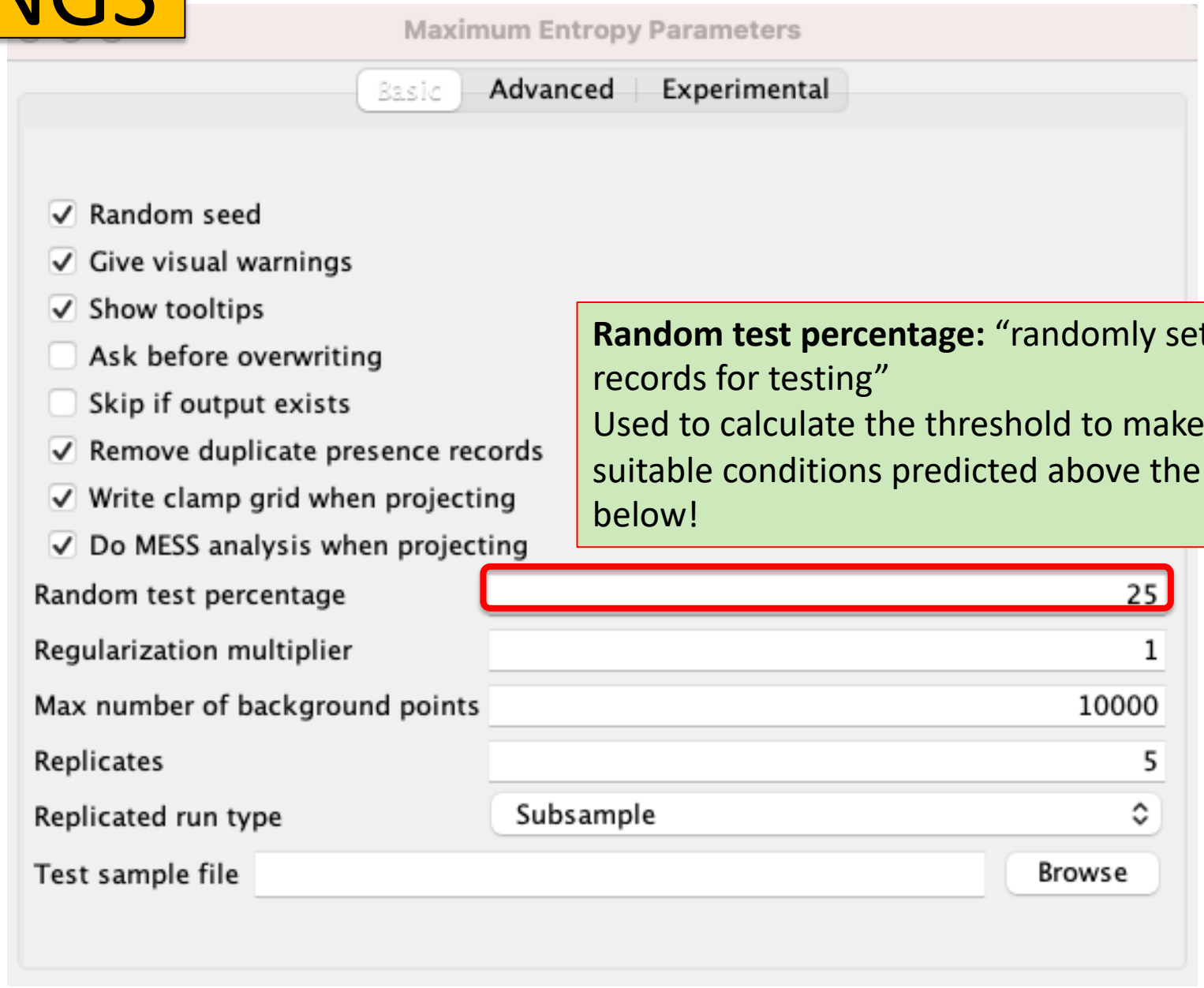
# REPLICATION SETTINGS

- Cross validation = resampling to evaluate model's performance with limited data



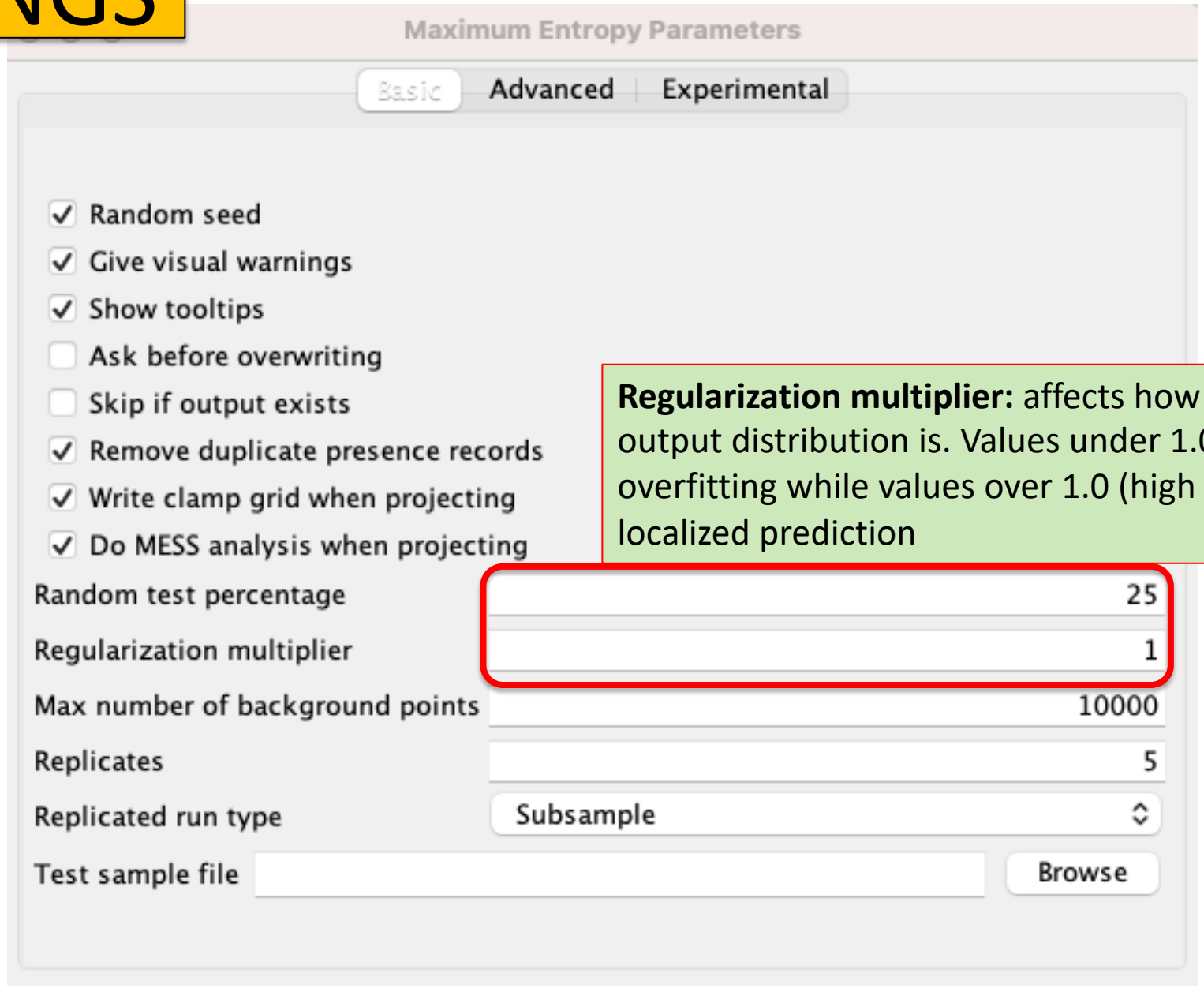

Slide modified from M. Mabry

# SETTINGS




**Random test percentage:** "randomly set aside 25% of the sample records for testing"
Used to calculate the threshold to make a binary prediction, with suitable conditions predicted above the threshold and unsuitable below!

# SETTINGS



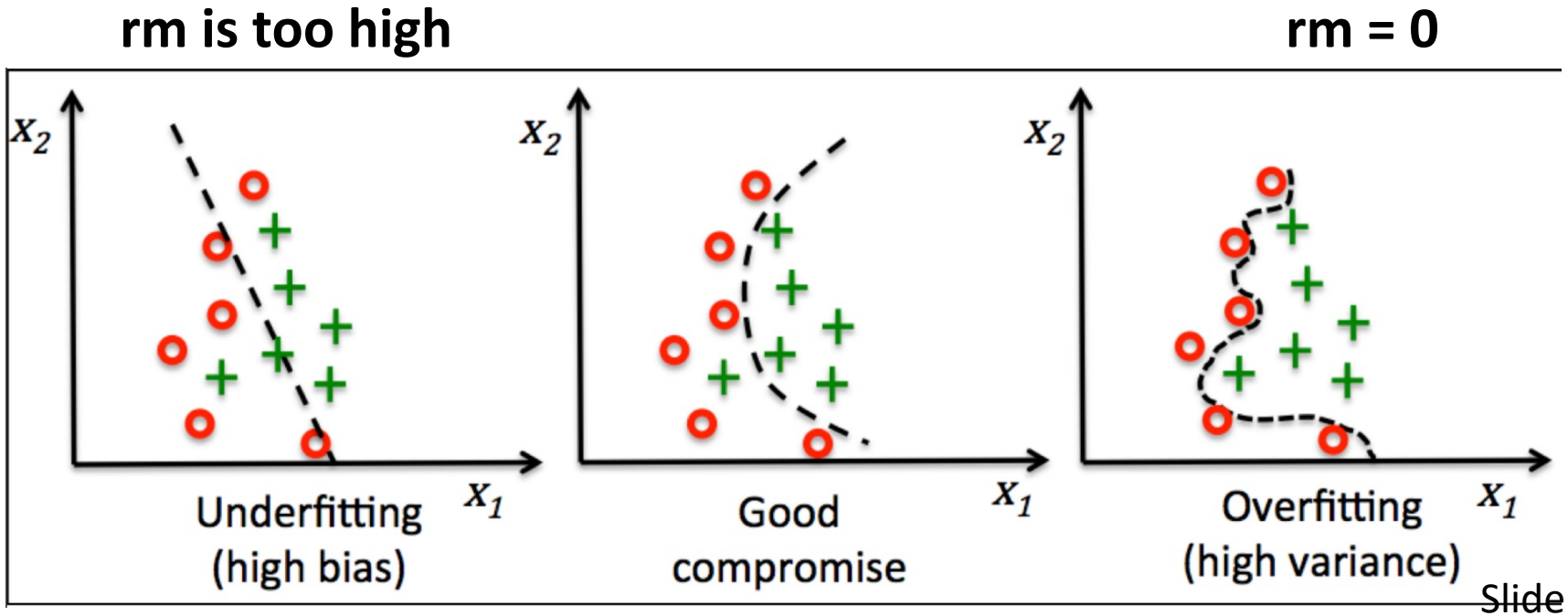

**Regularization multiplier:** affects how focused or closely-fitted the output distribution is. Values under 1.0 (low values) may result into overfitting while values over 1.0 (high values) tend to yield less localized prediction

# REGULARIZATION MULTIPLIER

- Regularization multiplier (rm) then aims to reduce overfitting of the model:

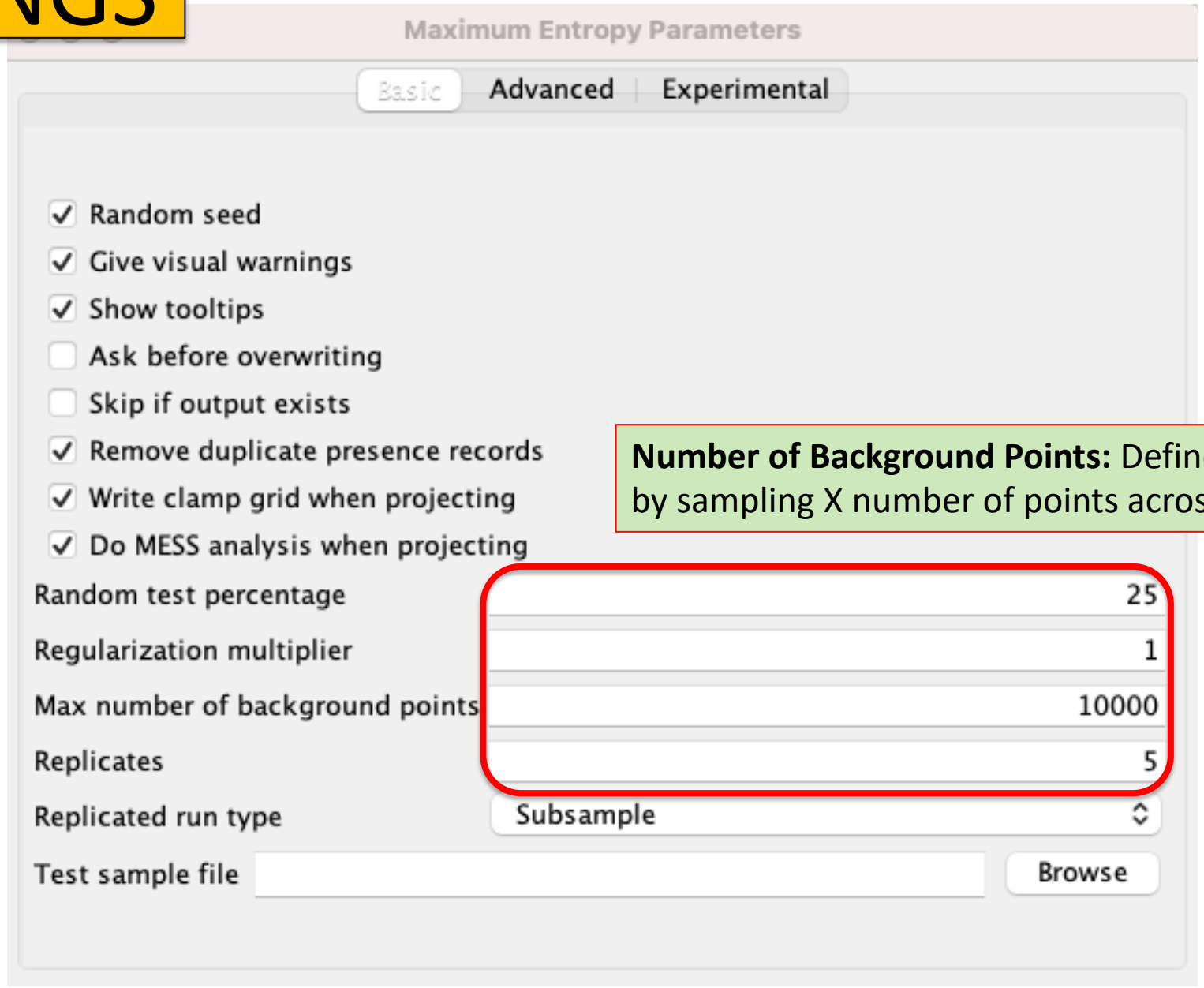
| Regularization multiplier | 1 |
|---|---|

- Determines the penalty associated with including variables or their transformations in the model.



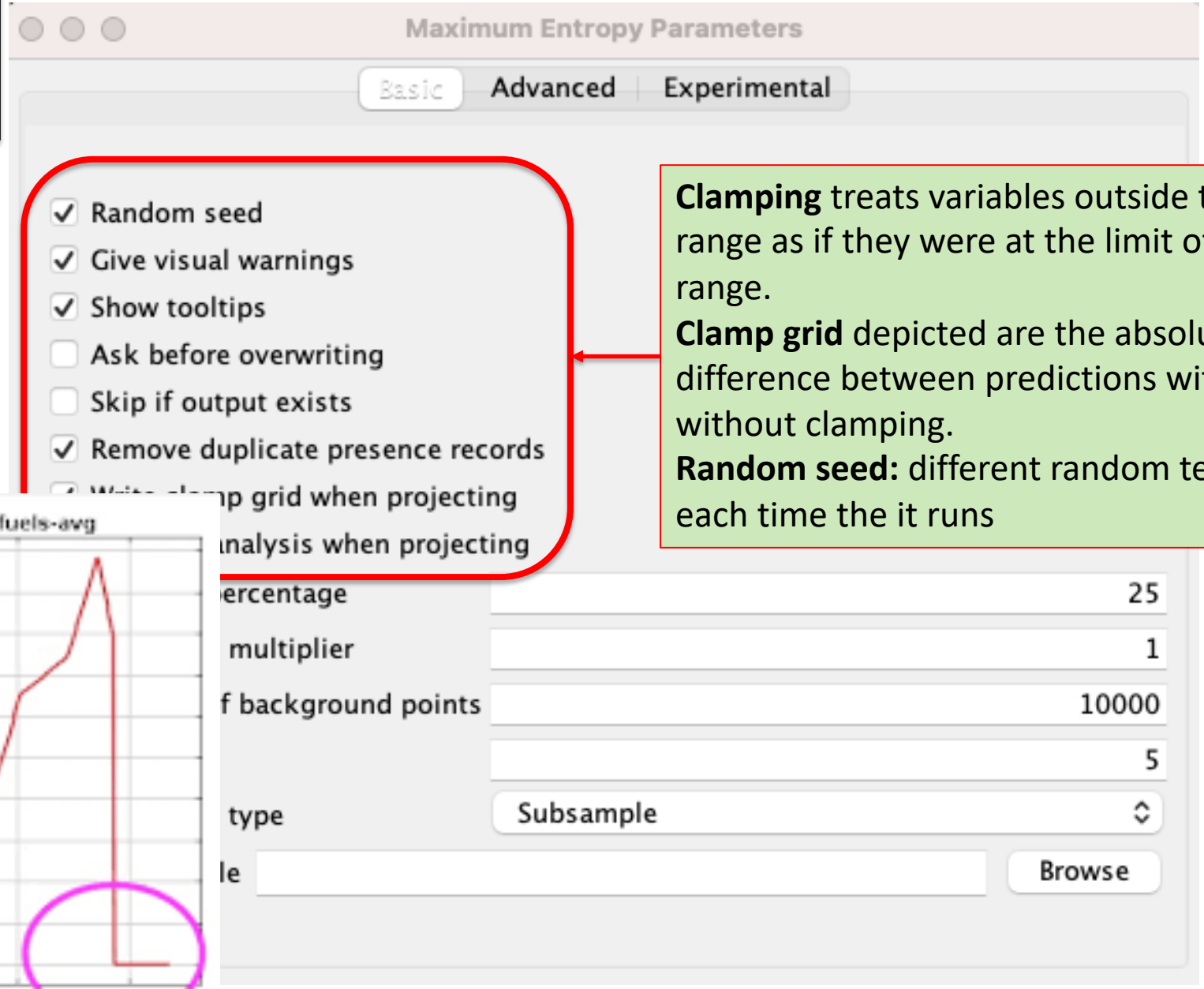

Slide modified from M. Mabry

# SETTINGS

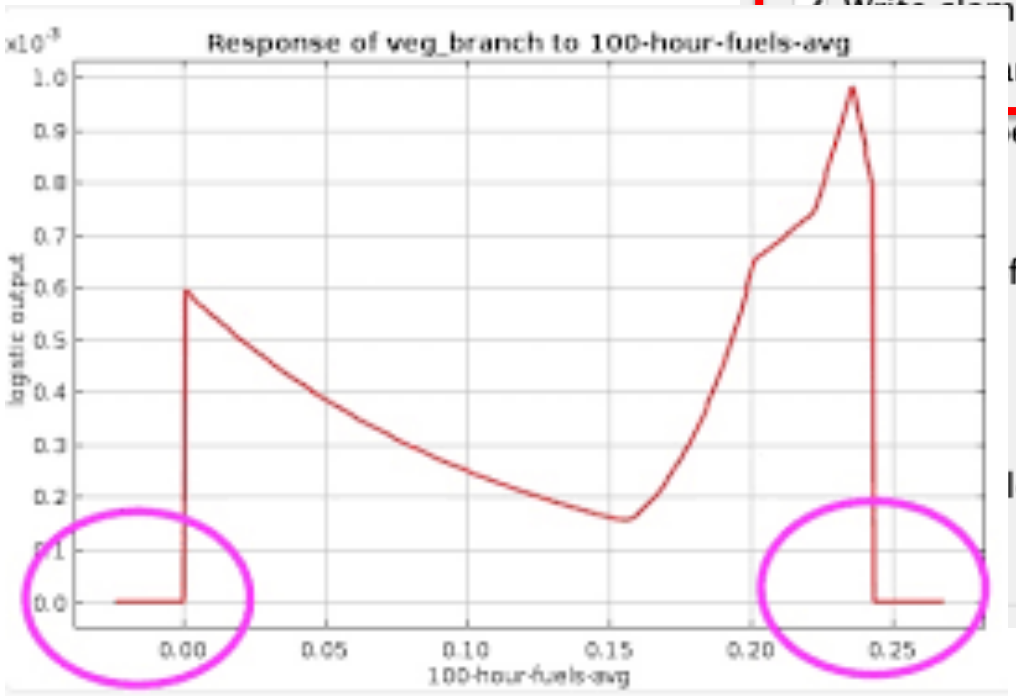



**Number of Background Points:** Defines the available environmental by sampling X number of points across the study area

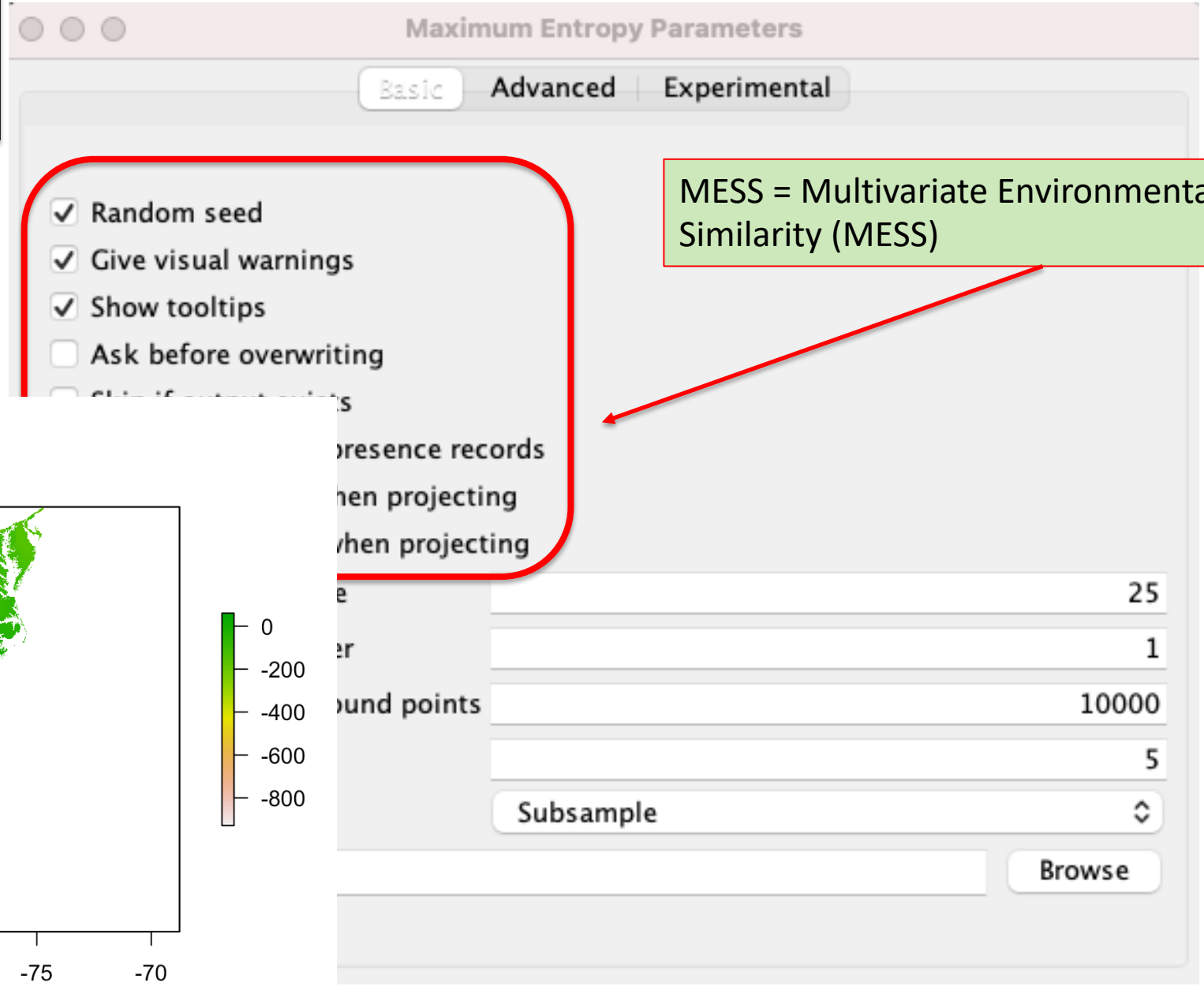
# PROJECTION SETTINGS



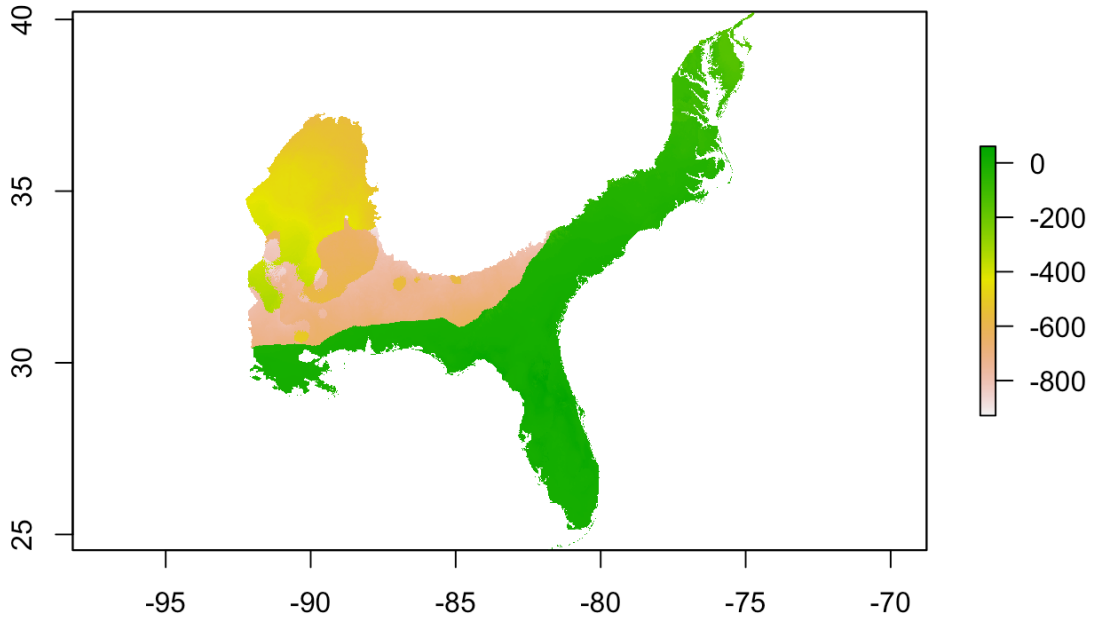

**Clamping** treats variables outside the training range as if they were at the limit of the training range.
**Clamp grid** depicted are the absolute difference between predictions with and without clamping.
**Random seed:** different random test points each time the it runs



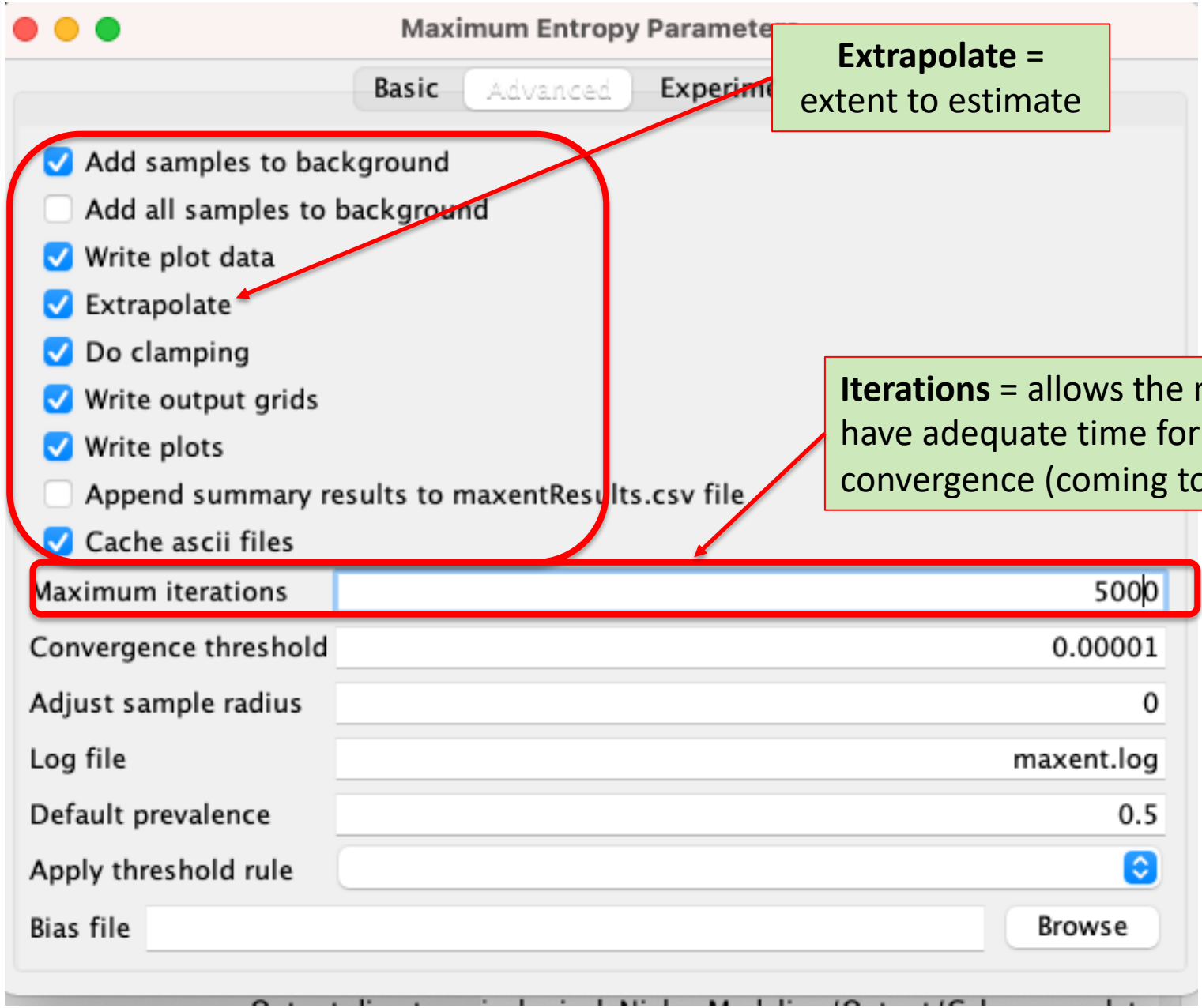
# PROJECTION SETTINGS




MESS = Multivariate Environmental Surface Similarity (MESS)



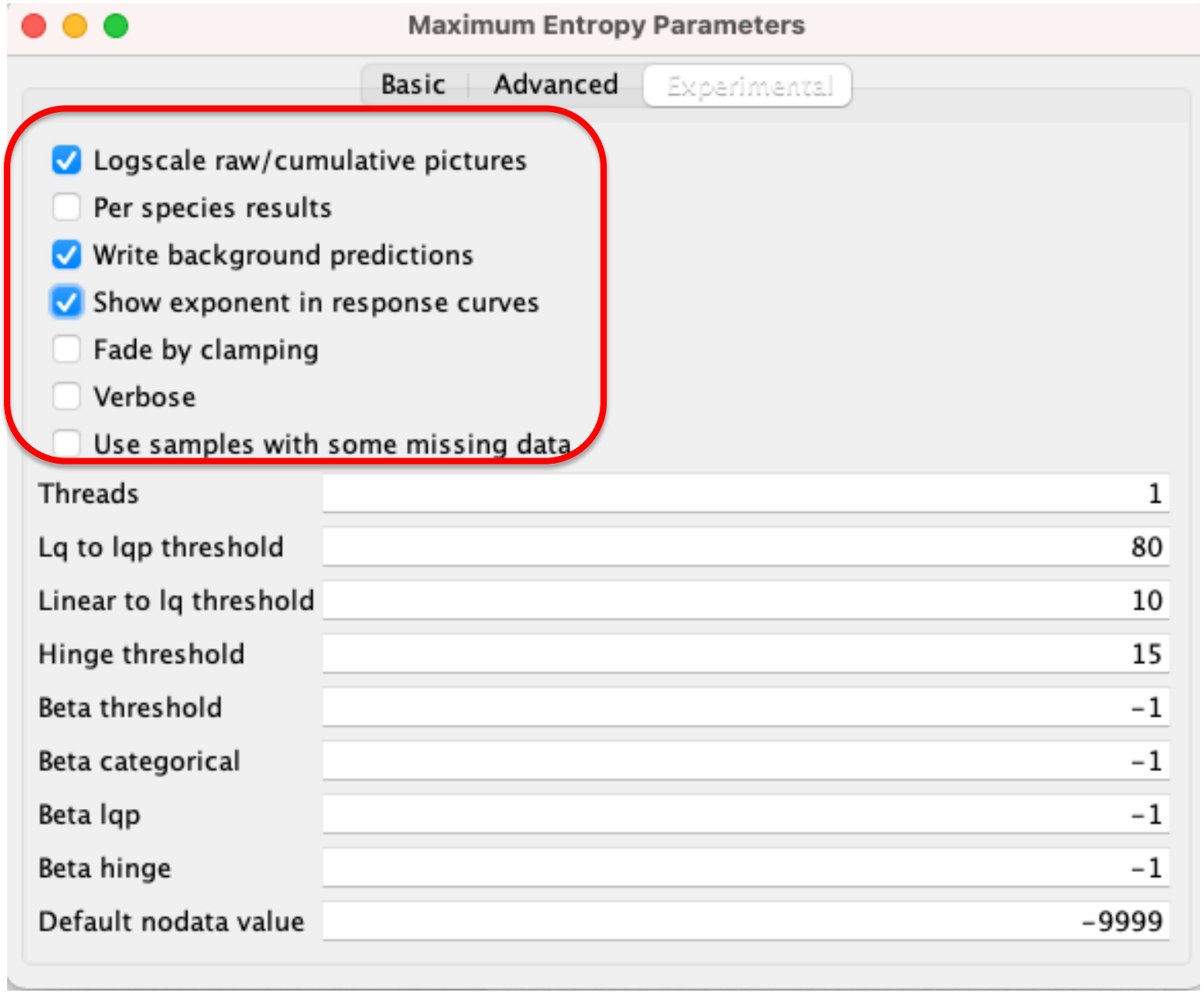
# PROJECTION SETTINGS



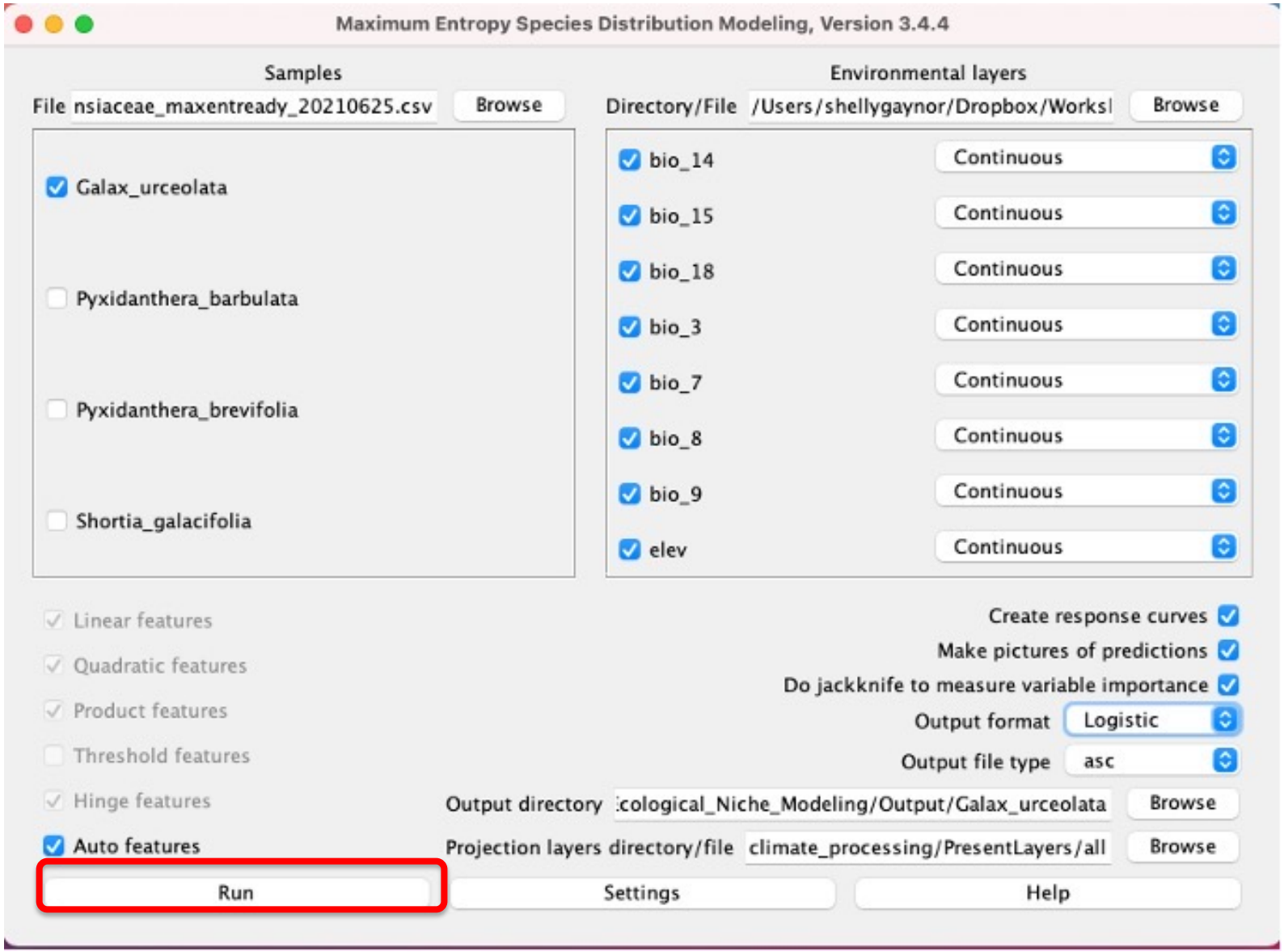

**Extrapolate** = extent to estimate

**Iterations** = allows the model to have adequate time for convergence (coming together)

# SETTINGS


Maximum Entropy Parameters
Basic
Advanced
Experimental
Logscale raw/cumulative pictures
Per species results
Write background predictions
Show exponent in response curves
Fade by clamping
Verbose
Use samples with some missing data
Threads 1
Lq to lqp threshold 80
Linear to lq threshold 10
Hinge threshold 15
Beta threshold -1
Beta categorical -1
Beta lqp -1
Beta hinge -1
Default nodata value -9999

# SETTINGS


Maximum Entropy Species Distribution Modeling, Version 3.4.4
Samples
File nsiaceae_maxentready_20210625.csv
Browse
Galax_urceolata
Pyxidanthera_barbulata
Pyxidanthera_brevifolia
Shortia_galacifolia
Environmental layers
Directory/File /Users/shellygaynor/Dropbox/Works
Browse
bio_14 Continuous
bio_15 Continuous
bio_18 Continuous
bio_3 Continuous
bio_7 Continuous
bio_8 Continuous
bio_9 Continuous
elev Continuous
Linear features
Quadratic features
Product features
Threshold features
Hinge features
Auto features
Create response curves
Make pictures of predictions
Do jackknife to measure variable importance
Output format Logistic
Output file type asc
Output directory cological_Niche_Modeling/Output/Galax_urceolata
Browse
Projection layers directory/file climate_processing/PresentLayers/all
Browse
Run
Settings
Help

# MaxEnt Demo

- Files for this activity can be found in **“Demo/Manual/Ecological_Niche_Modeling/”**
- Try out the demo!
- R-based ENM
    - Cross-validation and data partitioning
    - Advance model comparisons
    - Evaluating ENMs