

06 - Ecological Niche Modeling

Shelly Gaynor

University of Florida



R based Ecological Niche Models

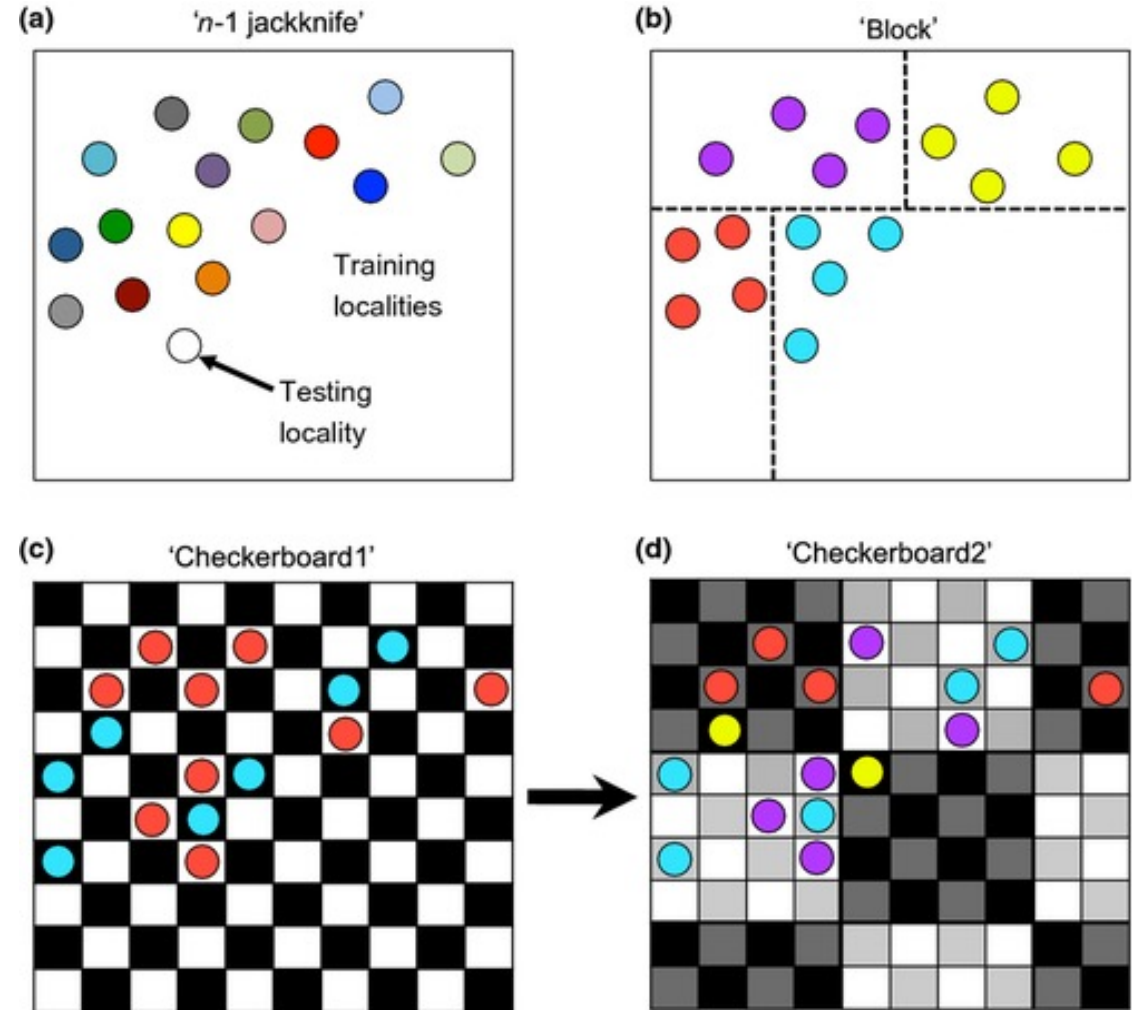
dismo and ENMeval

ENMeval Improvements

1. Additional methods for partitioning occurrence and background data into subsets for training and validation.
2. Builds models with different algorithmic settings and evaluates each model.

Data Partitioning

- partition occurrence localities into testing and training bins (folds) for k -fold cross-validation
- Previous options:
 - Bootstrap
 - Subsample
 - Crossvalidate



Muscarella et al. 2014

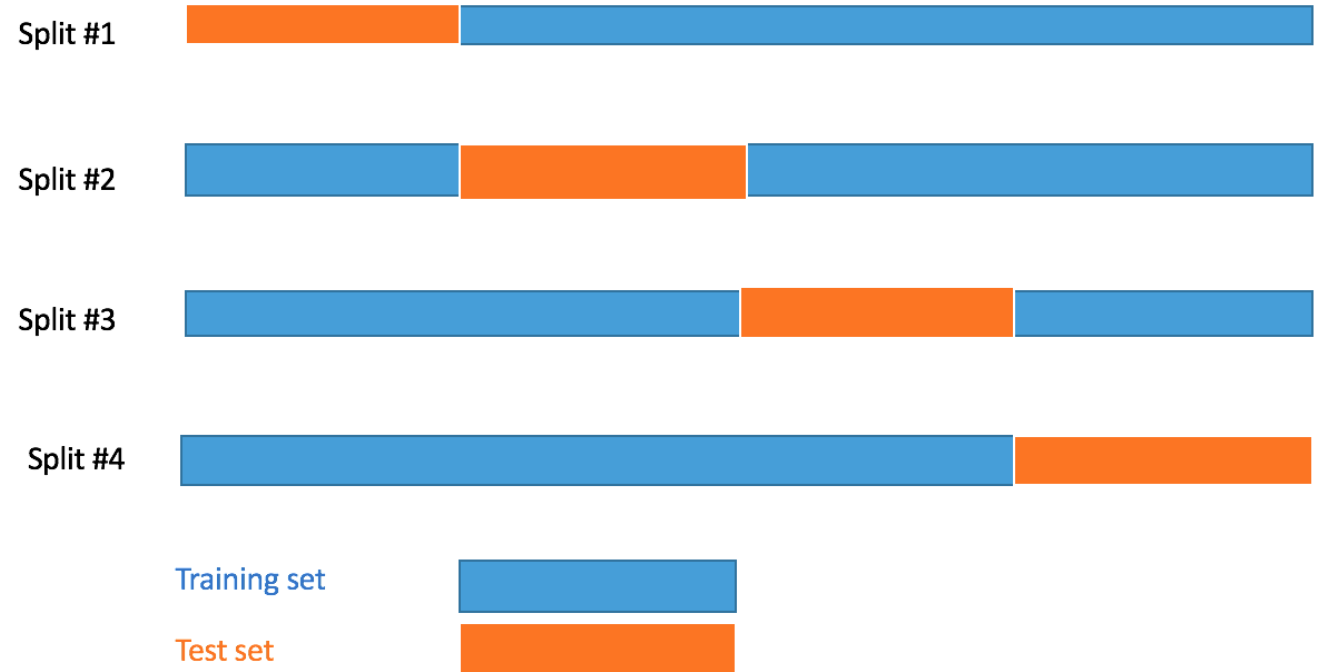
Cross Validation

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

ENMeval

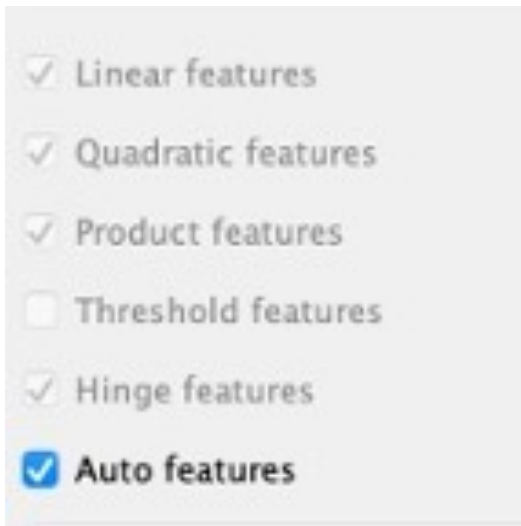
1. [Spatial Block](#)
2. [Spatial Checkerboard](#)
3. [Spatial Hierarchical Checkerboard](#)
4. [Jackknife \(leave-one-out\)](#)
5. [Random k-fold](#)
6. [Fully Withheld Testing Data](#)
7. [User](#)

4-fold cross-validation



Different algorithmic settings

- ENMeval allow the comparison of models with multiple regularization multipliers and different feature classes included






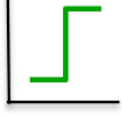


- Feature class is used to transform the original predictor
 - Determines the shape of the marginal responses curve
- MaxEnt Feature Classes:
 - linear (L)
 - quadratic (Q)
 - product (P)
 - threshold (T)
 - hinge (H)

Regularization multiplier

1

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

Feature Class (fc)

Feature type	Interpretation	Constraint	Shape
Linear	Continuous variable	The <i>mean</i> 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 <i>variance</i> 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 <i>co-variance</i> of two environmental variables at an unknown location should be close to the co-variance of those variables in known occurrence locations.	
Threshold	Conversion into binary response based on a threshold	The proportion of predicted occurrences with values above the threshold (binary response = 1) should be close to the proportion of known occurrences.	
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.	
Categorical	Categorical variable	The proportion of predicted occurrences in each category should be close to the proportion of observed occurrences in each category.	

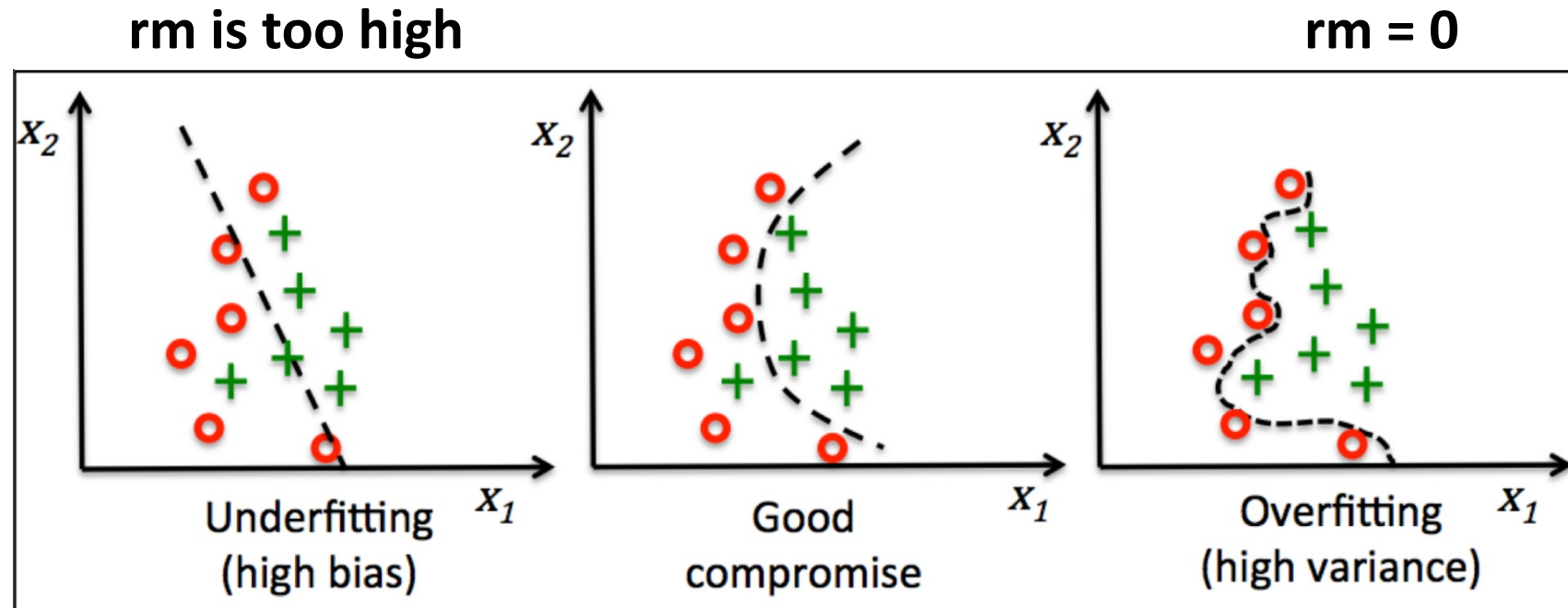
Regularization Multiplier (rm)

- 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.



Background Points (n.bg)

- Defines the available env. by sampling a large number of points across the study area.
- Used to define relative env. suitability.

Load Packages, Functions, & Java

Set up java memory

```
options(java.parameters = "-Xmx16g") # increase memory that can be used
```

Load packages

```
library(raster)  
library(gtools)  
library(dplyr)  
library(dismo)  
library(ENMeval)  
library(ggplot2)  
library(viridis)
```

Load function

```
source("functions/ENMevaluation.R")
```

Load Data Files

Load data file

```
alldf <- read.csv("data/cleaning_demo/maxent_ready/diapensiaceae_maxentready_20230605.csv")
```

Subset for each species

```
Galax_urceolata <- dplyr::filter(alldf, species == "Galax urceolata")  
Pyxidanthera_barbulata <- dplyr::filter(alldf, species == "Pyxidanthera barbulata")  
Pyxidanthera_brevifolia <- dplyr::filter(alldf, species == "Pyxidanthera brevifolia")  
Shortia_galacifolia <- dplyr::filter(alldf, species == "Shortia galacifolia")
```

Load Raster Layers

```
list <- list.files("data/climate_processing/PresentLayers/all", full.names = TRUE, recursive = FALSE)
list <- mixedsort(sort(list))
allstack <- stack(list)
```

Load Raster Layers

Read in species training layers

```
gstack <- stack(mixedsort(sort(list.files("data/climate_processing/PresentLayers/Galax_urc  
eolata/", full.names = TRUE))))  
pbastack <- stack(mixedsort(sort(list.files("data/climate_processing/PresentLayers/Pyxidan  
thera_barbulata/", full.names = TRUE))))  
pbrstack <- stack(mixedsort(sort(list.files("data/climate_processing/PresentLayers/Pyxidan  
thera_brevifolia/", full.names = TRUE))))  
sstack <- stack(mixedsort(sort(list.files("data/climate_processing/PresentLayers/Shortia_g  
alacifolia/", full.names = TRUE))))
```

Fix projection

```
projection(allstack) <- "+proj=longlat +ellps=WGS84 +towgs84=0,0,0,0,0,0,0 +no_defs"  
projection(gstack) <- "+proj=longlat +ellps=WGS84 +towgs84=0,0,0,0,0,0,0 +no_defs"  
projection(pbastack) <- "+proj=longlat +ellps=WGS84 +towgs84=0,0,0,0,0,0,0 +no_defs"  
projection(pbrstack) <- "+proj=longlat +ellps=WGS84 +towgs84=0,0,0,0,0,0,0 +no_defs"  
projection(sstack) <- "+proj=longlat +ellps=WGS84 +towgs84=0,0,0,0,0,0,0 +no_defs"
```

WARNING: TIME INTENSIVE

dismo Model

```
evaldis <- dismo::maxent(x = gstack, p = Galax_urceolata[, c("long", "lat")], nbg = 10000,  
  args = c("projectionlayers=data/climate_processing/PresentLayers/all",  
    "responsecurves", "jackknife",  
    "outputformat=logistic", "randomseed",  
    "randomtestpoints=25", "replicates=5",  
    "replicatetype=subsample", "maximumiterations=5000",  
    "writebackgroundpredictions", "responsecurvesexponent",  
    "writeplotdata"),  
  removeDuplicates = TRUE  
  #, path = "data/Ecological_Niche_Modeling/enm_output/Galax_urceolata/"  
)
```

ENMeval Model

```
eval1 <- ENMeval::ENMevaluate(occ = Galax_urceolata[, c("long", "lat")],  
                             env = gstack,  
                             tune.args = list(fc = c("L", "Q"), rm = 1:2),  
                             partitions = "block",  
                             n.bg = 10000,  
                             parallel = FALSE,  
                             algorithm = 'maxent.jar',  
                             user.eval = proc)
```

- For publication worthy comparisons – test more feature-class combinations and a larger range of regularization multipliers:
 - Ex.: `fc = c("L", "H", "LQ", "LHQ", "LQP")`, `rm = c(0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0)`

For example, see: Wang et al. 2021. Potential distributional shifts in North America of allelopathic invasive plant species under climate change models. Plant Diversity.

Continued in
Interpreting ENMs