

# Introdução à modelagem de distribuição de espécies usando a linguagem R

Mastozóólogos Organizados em uma Conferência Online (MOCÓ)

Maurício H. Vancine

25-26/09/2021

## Resumo

A ampla disponibilidade de dados sobre a biodiversidade e variáveis ambientais propiciam o uso de análises espaciais, dentre elas, os Modelos de Distribuição de Espécies (ou do inglês *Species Distribution Modeling* — SDM). Nessa oficina, oferecerei uma introdução teórica e prática à técnica de SDM utilizando a linguagem R. Primeiramente, apresentarei brevemente os principais conceitos da teoria de nicho ecológico (Grinnell, Elton e Hutchinson) e da teoria de SDMs (espaço geográfico (G), espaço ambiental (E) e diagrama Biótico-Abiótico-Movimentação (BAM)). Seguida à parte teórica, apresentarei as principais bases de dados (ocorrências e variáveis ambientais), tipos de algoritmos (BIOCLIM, Mahalanobis, Gower, GLM, GAM, Random Forest, SVM e MaxEnt), avaliação dos modelos (ROC, AUC e TSS), limites de corte (*thresholds*) e consenso de modelos (*ensemble*). A parte prática será focada na construção dos modelos através da linguagem R, onde abordarei: (1) introdução à linguagem R, (2) obtenção e seleção de dados de ocorrências e variáveis ambientais, (3) ajuste e predição dos modelos, (4) métricas de avaliação, (5) consenso dos modelos (*ensemble*) e (6) composição dos mapas.

## Carga horária

06 horas

## Repositório do GitHub

<https://github.com/mauriciovancine/workshop-r-sdm>

## Pré-requisitos

É desejável, mas não obrigatório, o conhecimento em linguagem R (tidyverse), geoprocessamento e conceitos de estatística básica e inferencial.

## Ministrante

Maurício Vancine é Bacharel em Ecologia, Mestre em Zoologia e Doutorando em Ecologia, Evolução e Biodiversidade pela UNESP de Rio Claro. Possui experiência em Ecologia Espacial, Ecologia da Paisagem, Modelos de Distribuição de Espécies e Ecologia de Anfíbios. Mais informações: <https://mauriciovancine.github.io/>.

## Conteúdo

### 1 Apresentação (10 min)

1. Apresentações
2. Oficina
3. Material
4. Slides
5. Conferindo os computadores

### 2 Teoria de nicho ecológico e distribuição de espécies (90 min)

1. Visão geral dos Modelos de Distribuição de Espécies
2. Nicho ecológico e distribuição de espécies
3. Outros tipos de modelos
4. Padronização para criação dos modelos
  - 4.1. Conceituação
  - 4.2. Preparação dos dados
  - 4.3. Ajuste dos modelos
  - 4.4. Avaliação dos modelos
  - 4.5. Predições no tempo e no espaço

### 3 Funcionamento da linguagem R (20 min)

1. Linguagem R
2. RStudio
3. Console
4. Scripts
5. Operadores
6. Objetos
7. Funções
8. Pacotes
9. Ajuda
10. Diretório de trabalho
11. tidyverse (pipe - %>%)

### 4 Dados de entrada (ocorrências e variáveis) (60 min)

1. Dados geoespaciais
2. Dados de ocorrências
3. Bases de dados de ocorrências
4. Download de dados de ocorrência
5. Limpeza de dados de ocorrência
6. Dados ambientais
7. Base de dados ambientais
8. Download de dados ambientais
9. Colinearidade de dados ambientais

### 5 Ajuste dos modelos (30 min)

1. Tipos de métodos ou algoritmos (apenas-presença, presença-background, presença-ausência)

2. Presenças, ausências e pseudo-ausências
3. Apenas-presença (BIOCLIM, DOMAIN e Mahalanobis)
4. Presença-background (MaxEnt)
5. Presença-ausência estatístico (GLM e GAM)
6. Presença-ausência aprendizado de máquina (Random Forest e SVM)

## **6 Avaliação dos modelos (30 min)**

1. Partição dos dados (treino e teste)
2. Tipos de partição dos dados (cross-validation, bootstrap, jackknife)
3. Matriz de confusão
4. Métricas de avaliação
5. ROC e AUC
6. Kappa e TSS

## **7 Predição e consenso dos modelos (30 min)**

1. Predição dos modelos
2. Limiar de corte dos modelos
3. Consenso dos modelos
4. Consenso por frequência
5. Consenso por média
6. Consenso por média ponderada
7. Incertezas

## **8 Composição de mapas (20 min)**

1. Composição de mapas
2. Mapas estáticos
3. Mapas interativos

## **Bibliografia**

- Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (Tss): Assessing the accuracy of distribution models. *Journal of Applied Ecology*, 43(6), 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Araújo, M. B., R. P. Anderson, A. M. Barbosa, C. M. Beale, C. F. Dormann, R. Early, R. A. Garcia, et al. 2019. “Standards for Distribution Models in Biodiversity Assessments.” *Science Advances* 5: eaat4858.
- Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carre, J. R. Garcia Marquez, et al. 2013. “Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance.” *Ecography* 36: 27–46.
- Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, et al. 2006. “Novel Methods Improve Prediction of Species’ Distribution from Occurrence Data.” *Ecography* 29: 129–51.
- Elith, J., and J. R. Leathwick. 2009. “Species Distribution Models: Ecological Explanation and Prediction Across Space and Time.” *Annual Review of Ecology, Evolution, and Systematics* 40: 677–97.
- Elith, J., S. J. Phillips, T. Hastie, M. Dudik, Y. E. Chee, and C. J. Yates. 2011. “A Statistical Explanation of Maxent for Ecologists.” *Diversity and Distributions* 17: 43–57.
- Franklin, J. 2010. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press.

- Freeman, E. A., and G. Moisen. 2008. "PresenceAbsence: An R Package for Presence Absence Analysis." *Journal of Statistical Software* 23: 1–31.
- Guillera-Aroita, G. 2017. "Modelling of Species Distributions, Range Dynamics and Communities Under Imperfect Detection: Advances, Challenges and Opportunities." *Ecography* 40: 281–95.
- Guisan, A., and W. Thuiller. 2005. "Predicting Species Distribution: Offering More Than Simple Habitat Models." *Ecology Letters* 8: 993–1009.
- Guisan, A., W. Thuiller, and N. E. Zimmermann. 2017. *Habitat Suitability and Distribution Models with Applications in R*. Cambridge University Press.
- Guisan, A., and N. E. Zimmermann. 2000. "Predictive Habitat Distribution Models in Ecology." *Ecological Modelling* 135: 147–86.
- Hijmans, R. J. 2019. *Raster: Geographic Data Analysis and Modeling*. <https://CRAN.R-project.org/package=raster>.
- Hutchinson, G. E. 1957. "Concluding Remarks, Cold Spring Harbor Symposium." *Quantitative Biology* 22: 415–27.
- Liu, C., P. M. Berry, T. P. Dawson, and R. G. Pearson. 2005. "Selecting Thresholds of Occurrence in the Prediction of Species Distributions." *Ecography* 28: 385–93.
- Merow, C., M. J. Smith, T. C. Edwards Jr, A. Guisan, S. M. McMahon, S. Normand, W. Thuiller, R. O. Wueest, N. E. Zimmermann, and J. Elith. 2014. "What Do We Gain from Simplicity Versus Complexity in Species Distribution Models?" *Ecography* 37: 1267–81.
- Merow, C., M. J. Smith, and J. A. Silander Jr. 2013. "A Practical Guide to Maxent for Modeling Species' Distributions: What It Does, and Why Inputs and Settings Matter." *Ecography* 36: 1058–69.
- Peterson, A. T., J. Soberon, R.G. Pearson, R. P. Anderson, E. Martinez-Meyer, M. Nakamura, and M. B. Araujo. 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press.
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. "Maximum Entropy Modeling of Species Geographic Distributions." *Ecological Modelling* 190: 231–59.
- R Core Team. 2021. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>
- Roberts, D. R., V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Aroita, S. Hauenstein, et al. 2017. "Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure." *Ecography* 40: 913–29.
- Soberon, J. 2007. "Grinnellian and Eltonian Niches and Geographic Distributions of Species." *Ecology Letters* 10: 1115–23.
- Thuiller, W., M. Guéguen, J. Renaud, D. N. Karger, and N. E. Zimmermann. 2019. "Uncertainty in Ensembles of Global Biodiversity Scenarios." *Nature Communications* 10: 1446.
- Thuiller, W., B. Lafourcade, R. Engler, and M. B. Araujo. 2009. "BIOMOD - a Platform for Ensemble Forecasting of Species Distributions." *Ecography* 32: 369–73.
- Urban, M. C., G. Bocedi, A. P. Hendry, J.-B. Mihoub, G. Pe'er, A. Singer, J. R. Bridle, et al. 2016. "Improving the Forecast for Biodiversity Under Climate Change." *Science* 353: aad8466.
- Zurell, D., J. Elith, and B. Schroeder. 2012. "Predicting to New Environments: Tools for Visualizing Model Behavior and Impacts on Mapped Distributions." *Diversity and Distributions* 18: 628–34.
- Zurell, D., and J.O. Engler. 2019. "Ecological Niche Modelling." In *Effects of Climate Change on Birds*, edited by P.O. Dunn and A.P. Moller, 60–73. Oxford University Press.