



*Minicurso Como a tecnologia pode  
auxiliar a entender a distribuição das espécies?*

*Marina Morim Gomes - 2025*



## **Aula Prática Modelagem de Nicho Ecológico**

## **Materiais necessários para essa prática:**

- Notebook ou Computador
- R e RStudio Instalados no Computador que será usado
- Dados de ocorrências disponíveis em:  
[https://github.com/marinagomesdip/minicurso\\_modelagem\\_nicho/tree/main/Data/ocss\\_species](https://github.com/marinagomesdip/minicurso_modelagem_nicho/tree/main/Data/ocss_species)
- Dados climáticos disponíveis em:  
[https://github.com/marinagomesdip/minicurso\\_modelagem\\_nicho/tree/main/Data/raster](https://github.com/marinagomesdip/minicurso_modelagem_nicho/tree/main/Data/raster)

## Primeiros Passos

**ATENÇÃO:** Esse tutorial foi elaborado baseado na versão 4.5.1 do R e na versão 2023.12.1 do R Studio. Versões posteriores podem apresentar divergência na execução dos comandos.

### 1. Abrir o R Studio no seu computador

Busque o ícone na sua tela principal ou no buscador de programas

### 2. Iniciar um novo script

Vá em File no canto superior esquerdo > New File > R Script ou utilize o atalho Ctrl + Shift + N para abrir um novo script

### 3. Instale o pacote de Modelagem de Nicho que vamos utilizar (NicheToolBox – ‘ntbox’):

O tutorial completo, em idioma inglês com todos os guias pode ser encontrado [aqui](#). Neste curso, faremos da forma mais simples.

#### 3.1 Comece instalando o RTools em sua versão mais recente

O programa pode ser encontrado em <https://cran.r-project.org/bin/windows/Rtools/> . Basta instalar como qualquer outro programa em seu computador.

#### 3.2 Instale o pacote ‘devtools’ por dentro do R, com o comando indicado no nosso script.

#### 3.3 Instale o pacote ‘ntbox’ por dentro do R, com o comando indicado no nosso script.

## Abrindo o pacote através da interface shiny

A interface shiny é uma interface mais ‘amigável’ do que a versão original do R que utiliza linguagem de programação, facilitando a visualização dos processos. Neste minicurso usaremos essa interface para facilitar a prática por conta do tempo disponível.

- 1. Execute os comandos indicados no script para abrir a plataforma**  
Aguarde a execução dos pacotes e comandos pelo R. Isso pode demorar um pouco dependendo do seu processador.

Load `ntbox` in your R session and type

Hide

```
library(ntbox)  
run_ntbox()
```

- 2. Maximize a interface shiny**

- 3. Selecione a pasta de trabalho**

Observe que na aba ‘APPSettings’, a primeira configuração é selecionar um diretório do nosso trabalho. É importante observar que quando estamos executando análises/projetos em R, a organização dos diretórios (pastas) é extremamente importante para não se perder.

Vamos manter a seguinte organização dos diretórios:

MinicursoModelagem

└─ dados

└─ raster variáveis ambientais

└─ registros de ocorrência

└─ resultados

| — output\_1

| — output\_2

Na opção ‘Select workflow directory’ clique nos “...” e selecione a pasta ‘MinicursoModelagem’ na sua área de trabalho.

#### 4. Carregue as variáveis como layers dentro do programa

No subtópico “Niche layers” Vá em “Select a raster layers directory” e clique nos “...” para selecionar a pasta ‘raster\_variaveis\_ambientais’ e clique em “Load niche layers”

### Niche layers

One ne of the most important things for your data analysis in NTB is to select the folder with the raster layers that you will use as niche variables.

Please select the folder of raster layers; notice that all of them need to be in the same resolution and extension.

The supported raster formats are: .asc .bil , .tif , .nc. , .sdat , .img

Select a raster layers directory

...

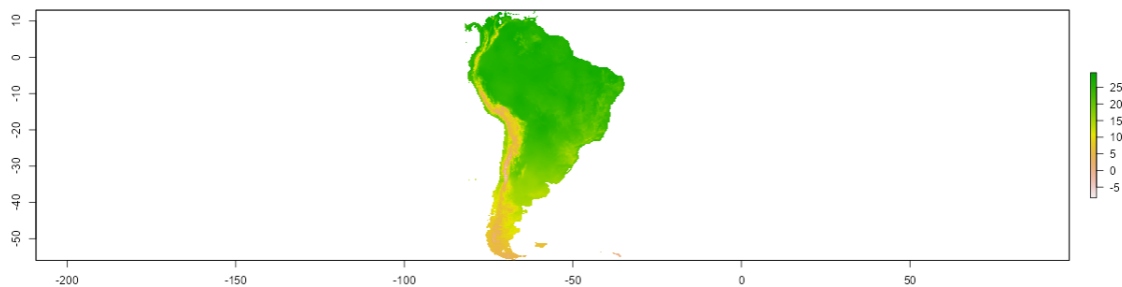
☐ Select projection layers

If you dont have environmental data ntbox provides methods from `raster` package and other sources ( [WorldClim](#), [CHELSA](#), [Bio-Oracle](#), [ENVIREM](#) ) to get environmental data.

☐ Get environmental data

⬇ Load niche layers

Após clicar nesse botão, os layers deverão aparecer na sua tela no lado esquerdo



\* Os layers utilizados nessa prática já foram recortados para a América do Sul para facilitar o tamanho dos arquivos. Para isso, utilizamos um script do R disponibilizado na pasta do curso, dentro de ‘Scripts’.

## Dados de Ocorrência da Espécie



Para essa prática usaremos uma espécie conhecida de besouros cerambycídeos, a *Cosmisoma ochraceum* (Perty, 1832) (Figura ao lado – retirada de

<http://bezbycids.com/byciddb/wdetails.asp?id=5377&w=n> –

Fotografada por Larry G. Bezark e determinado por Dilma Napp).

### 1. Importar os dados dentro da plataforma shiny para visualiza-los.

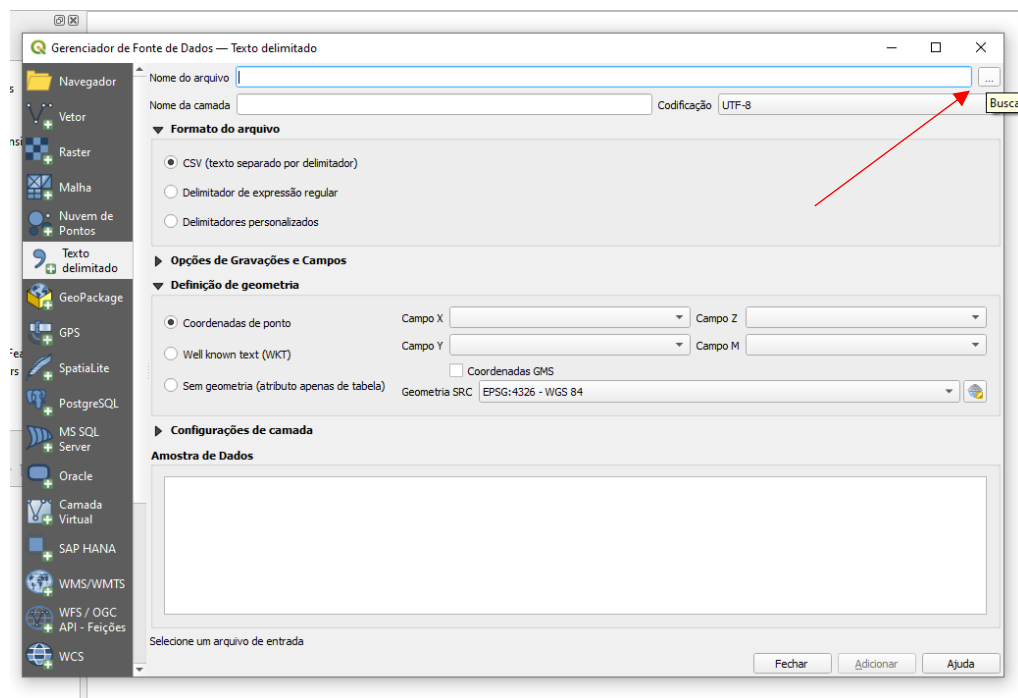
No menu superior, agora trabalharemos na aba ‘Data’, onde podemos importar dados do GBIF ou inserir nossos próprios dados. Antes de importa-los, vamos explorar a qualidade dos dados. **Lembre-se que a limpeza dos dados e a qualidade deles está diretamente relacionada com o resultado final do modelo gerado!!**

#### 1.1 GBIF

No buscador do GBIF, procure pelo nome da espécie, depois selecione o banco de dados e clique em “occurrences” para visualizar os dados tanto no formato planilha quanto no formato mapa e avalie previamente a qualidade dos dados. Eles são completos? Você os incluiria na sua análise??

#### 1.2 DADOS DE COLEÇÃO

Na pasta do github, vocês baixaram um arquivo de ocorrências da espécie, simulando um levantamento feito em uma coleção biológica. Antes de importa-lo, vamos abrir no QGIS para ver a qualidade dos dados. Abra o programa, um novo projeto e inclua os dados através do Gerenciador de Fonte de Dados (ou através do atalho Ctrl+L), clique nos três pontos e selecione o arquivo do levantamento.



Para podermos visualizar esses pontos em um mapa, vá em Complementos no menu superior > Gerenciar e Instalar Complementos. Na aba que vai abrir, em 'Tudo', busque por QuickMapService. Instale esse complemento no seu QGIS. Depois vá em 'Web' no menu superior > QuickMapServices > Settings. Mude para a aba 'More services' e clique em 'Get contributed pack'. Depois que carregar, feche essa aba, volte novamente ao menu do QuickMapServices, mas agora vá em Google > Google Satellite.

Depois de plotar, vamos verificar a coerência desses dados. Abra o Catálogo Taxônomico da Fauna Brasileira (<http://fauna.jbrj.gov.br/fauna/listaBrasil/ConsultaPublicaUC/ResultadoDaConsultaNovaConsulta.do>) e busque pelo nome da espécie. Compare os dados conhecidos de distribuição da espécie com os dados que você tem de uma coleção. Eles são coerentes? Tem dados duvidosos que poderiam ser excluídos? Exclua manualmente no csv e salve a nova planilha, sempre mantendo a original.



\* Observação: Você também pode testar a coerência das ocorrências com os rasters das variáveis ambientais. Plote o raster de elevação usando o Gerenciador de Fonte de Dados e verifique a coerência, considerando que essa espécie já foi estudada e só ocorre até o máximo de 900m de altura. De acordo com uma literatura hipotética, a espécie não é capaz de sobreviver acima dessa altitude. Para isso, use a ferramenta ‘Amostrar Valores de um raster’ dentro de Processamento > Caixa de Ferramentas, para verificar os valores associados a cada ponto. Nesse caso, a camada de entrada são seus dados de ocorrência e a camada raster é sua camada de valores. Ao executar essa função, o resultado será uma camada, que você pode acessar os valores clicando com o botão direito do mouse e indo em ‘tabelas de atributos’ para avaliar.

### 1.3 SELECIONAR DADOS DE TESTE

Idealmente, quando construímos um banco de dados com dados de espécies, utilizamos diferentes fontes independentes. Nesse caso, seríamos capazes de selecionar uma dessas fontes para testar o nosso modelo manualmente. Na prática, geralmente conseguimos poucos bancos de dados independentes, o que faz com que precisemos selecionar pontos aleatoriamente para testar o modelo. Nós iremos omitir esses pontos do programa e verificar se ele consegue construir um bom modelo que inclua esses pontos de ocorrência.

Vamos sortear 5 pontos aleatórios para teste manual posterior. Na planilha do excel, em uma nova coluna, insira a fórmula ‘aleatórioentre()’ ou ‘randbetween()’ e arraste até o final dos dados. Depois, classifique a planilha do excel a partir dessa coluna. Vamos copiar os 4 primeiros dados para uma nova planilha separada e remover desta para validar posteriormente.

## 2. Insira os dados que você já verificou na plataforma através do botão ‘browse’.

Selecione o arquivo que contém os registros de ocorrência.

**3. Remova os registros que são muito próximos.**

Para reduzir o impacto do viés amostral, é recomendado que mantenhamos apenas 1 registro por pixel. Para a escala de 10 arc-minutes, é recomendado utilizar 0.2 graus (aproximadamente 20km). Inclua esse valor, clique em ‘clean duplicates’, e, após limpar, baixe o arquivo com as ocorrências limpas. Ao utilizar a modelagem para projetos científicos, é sempre interessante que todas as etapas sejam registradas, sempre modificando os dados sem excluir os originais.

**4. Visualize os dados utilizando a aba ‘Dinamic Map’**

Nessa aba, você consegue escolher as diferentes fontes que importou os dados, além de selecionar manualmente dados para limpar. Nela também é possível incluir os dados de M (dispersão – movimento) e classificar os dados historicamente, o que não faremos no exemplo de hoje que é algo mais simples.

## Conhecendo o nicho da espécie

### 1. Vá na aba ‘Niche Space’ e em ‘Niche Data Extraction’

Nessa aba, poderemos extrair os valores do nicho relacionado com cada ponto. Nesta aba, o Niche Tool Box já separa em dados de treino e dados de teste para rodar posteriormente e podemos configurar a proporção dos dados de treino. Vamos manter essa proporção em 0.8 e clicar em ‘GO!’.

### 2. Vá na aba ‘Know niche’ e explore o nicho da espécie que você está trabalhando

No gráfico, é possível selecionar três diferentes variáveis para plotar e analisar o nicho da espécie visualmente. Entenda o formato do nicho, explore as diferentes variáveis que fazem sentido para determinar a distribuição da espécie.

### 3. Na aba ‘Niche Correlations’ é possível entender quais variáveis são relacionadas

Assim, você consegue ficar alternando entre essa aba e a anterior para ‘desenhar’ e compreender o nicho da espécie que você está estudando.

### 4. Selecione as variáveis que serão utilizadas

Para finalizar essa etapa, vamos selecionar as variáveis que iremos trabalhar nesse minicurso. Avalie quais são elas e anote para prosseguir.

## Dividindo grupos de espécies

### 1. Vá na aba ‘Niche clustering’

Nessa aba, poderemos classificar o nicho de possíveis populações, subpopulações ou espécies crípticas, através da similaridade entre o nicho no espaço ambiental.

### 2. Configure os parâmetros

Selecione o dataset inserido, as variáveis que já avaliamos serem relevantes para a espécie e o número de clusters. Após isso, clique em “GO!”.

### 3. Verifique seus resultados

Perceba como o que é próximo no espaço ambiental pode ser totalmente distinto no espaço geográfico. Lembre-se que aqui estamos calculando o nicho através dos valores das variáveis e depois projetando no espaço geográfico. A relação espaço ambiental e geográfico é extremamente importante para entendermos os resultados gerados aqui.

## Modelando uma espécie

### 1. Vá na aba ‘ENM’

Nessa aba, poderemos efetivamente modelar a espécie. Na plataforma shiny nem todos os algoritmos implementados estão disponíveis. Aqui, conseguimos explorar dois dos algoritmos mais “simples” conhecidos, o Minimum volume ellipsoid model e o BioClim. O tutorial completo com os algoritmos disponíveis na interface, você pode acessar em [https://luismurao.github.io/ntbox\\_user\\_guide.html#minimum-volume-ellipsoid-model](https://luismurao.github.io/ntbox_user_guide.html#minimum-volume-ellipsoid-model).

### 2. Minimum volume ellipsoid model - Van Aelst and Rousseeuw 2009

O modelo elipsóide proposto por Van Aelst e Rousseeuw usam a função de densidade de probabilidade multinormal para calcular o índice de adequação do nicho.

Para utiliza-lo, selecione as variáveis já selecionadas anteriormente e clique em “Go!”. Aqui, caso você tenha um polígono já elaborado para ‘M’, é possível inseri-lo e usar para projetar os dados. Após finalizar a análise, baixe seu arquivo em .asc

### 3. BIOCLIM – Booth et al. 2014

Perceba como o que é próximo no espaço ambiental pode ser totalmente distinto no espaço geográfico. Lembre-se que aqui estamos calculando o nicho através dos valores das variáveis e depois projetando no espaço geográfico. A relação espaço ambiental e geográfico é extremamente importante para entendermos os resultados gerados aqui. Após finalizar a análise, baixe seu arquivo em .asc

## Performance do Modelo

### 1. Vá na aba ‘SDM Performance’ e clique em ‘Partial ROC’

Nessa aba, podemos calcular o AUC e avaliar o quão bom ou ruim desempenhou nosso modelo. Para isso, você precisa incluir o resultado da etapa anterior, salvo em .asc e os dados de presença utilizados (a sua planilha de ocorrência). Atente-se que, neste caso, a planilha precisa ter uma ordem específica e nome específicos de coluna para que o pacote rode.

### 2. Binarization methods

Nessa aba, conseguimos visualizar os métodos que o pacote disponibiliza para calcularmos o threshold ou mesmo escolher manualmente um valor de threshold para binarizar o mapa. A partir desse mapa binarizado, poderemos testar com os dados independentes (ou aleatorizados) externamente no QGIS.

### 3. Finalização dos Mapas

Por fim, finalizaremos os mapas no QGIS, plotando nossos dados utilizados e os dados separados anteriormente.

## Literatura Utilizada

Booth, Trevor H. et al. BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. *Diversity and distributions*, v. 20, n. 1, p. 1-9, 2014.

Van Aelst, Stefan; Rousseeuw, Peter. Minimum volume ellipsoid. *Wiley Interdisciplinary Reviews: Computational Statistics*, v. 1, n. 1, p. 71-82, 2009.

Osorio-Olvera, L., Lira-Noriega, A., Soberón, J., Townsend Peterson, A., Falconi, M., Contreras-Díaz, R.G., Martínez-Meyer, E., Barve, V. and Barve, N. (2020), ntbox: an R package with graphical user interface for modeling and evaluating multidimensional ecological niches. *Methods Ecol Evol*.  
doi:10.1111/2041-210X.13452