

# Introdução à modelagem de distribuição de espécies usando a linguagem R: teoria e prática

---

Maurício Vancine  
[mauricio.vancine@gmail.com](mailto:mauricio.vancine@gmail.com)

João Giovanelli  
[jgiovanelli@gmail.com](mailto:jgiovanelli@gmail.com)



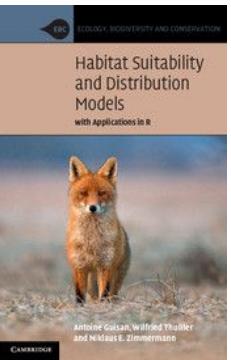
9º CONGRESSO BRASILEIRO DE  
**HERPETOLOGIA**  
CAMPINAS | 22 A 26 DE JULHO, 2019

# Apresentação do curso

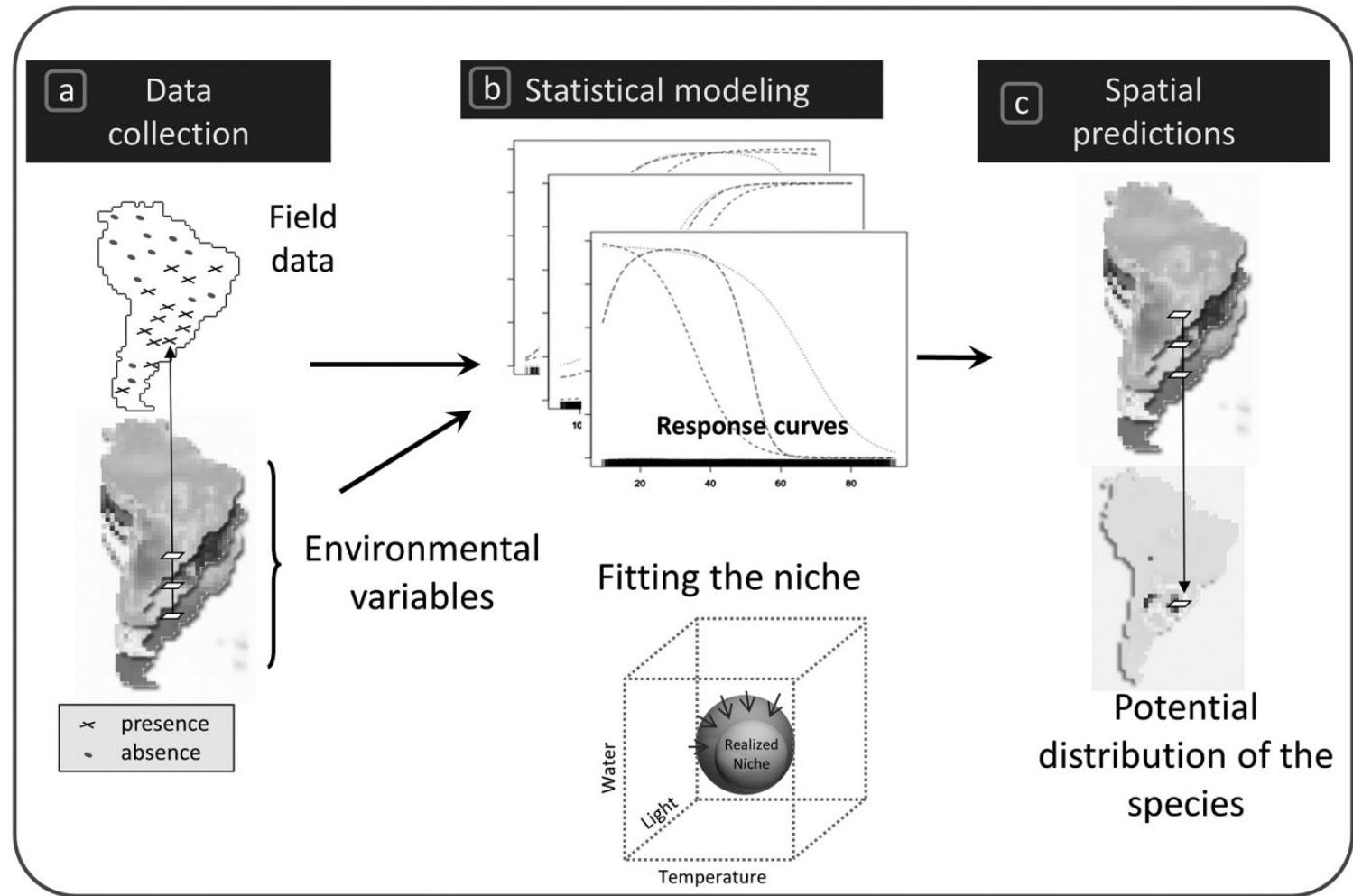
1. Introdução à Modelagem de Distribuição de Espécies (MDEs)
2. Introdução à linguagem R
3. Dados de entrada (variáveis e ocorrências)
4. Algoritmos, construção e avaliação dos MDEs
5. Automatização, consenso e mapas dos MDEs

# Avisos

- Tempo curto (7 horas)
- Conferir a instalação do R, RStudio e pacotes
- Conferir a internet
- Momentos alternados entre teoria e prática no R
- Não vamos abordar projeções (passado e futuro) ou invasão biológica
- Vamos utilizar resolução grosseira (~55 km)
- Seleção de uma espécie de interesse



# Introdução a MDEs



Guisan et al. (2017)

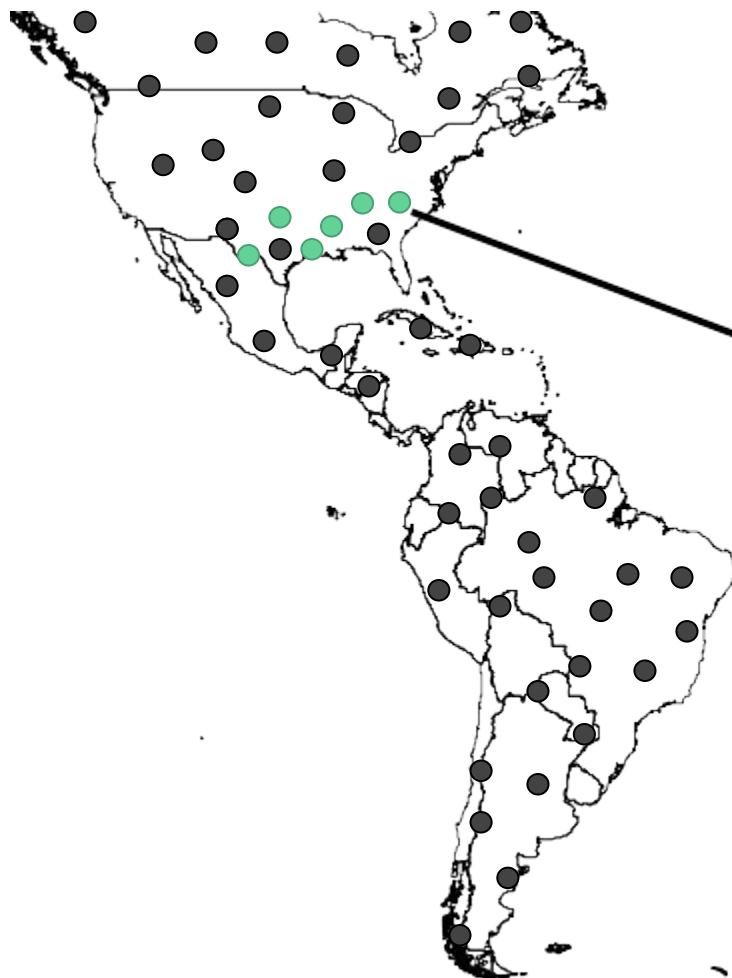
# Relação entre o espaço geográfico (G) e ambiental (E)

## Espaço geográfico (G)

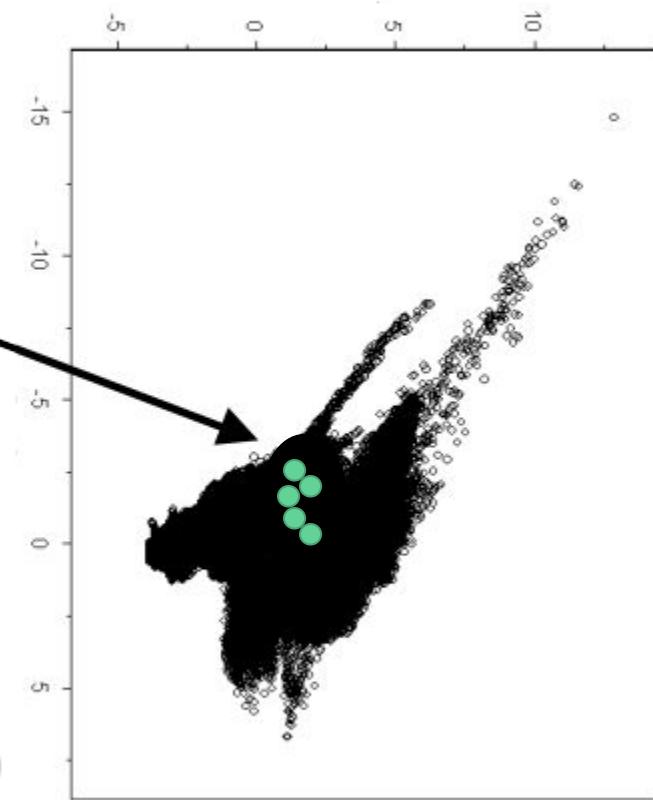


# Relação entre o espaço geográfico (G) e ambiental (E)

Espaço geográfico (G)

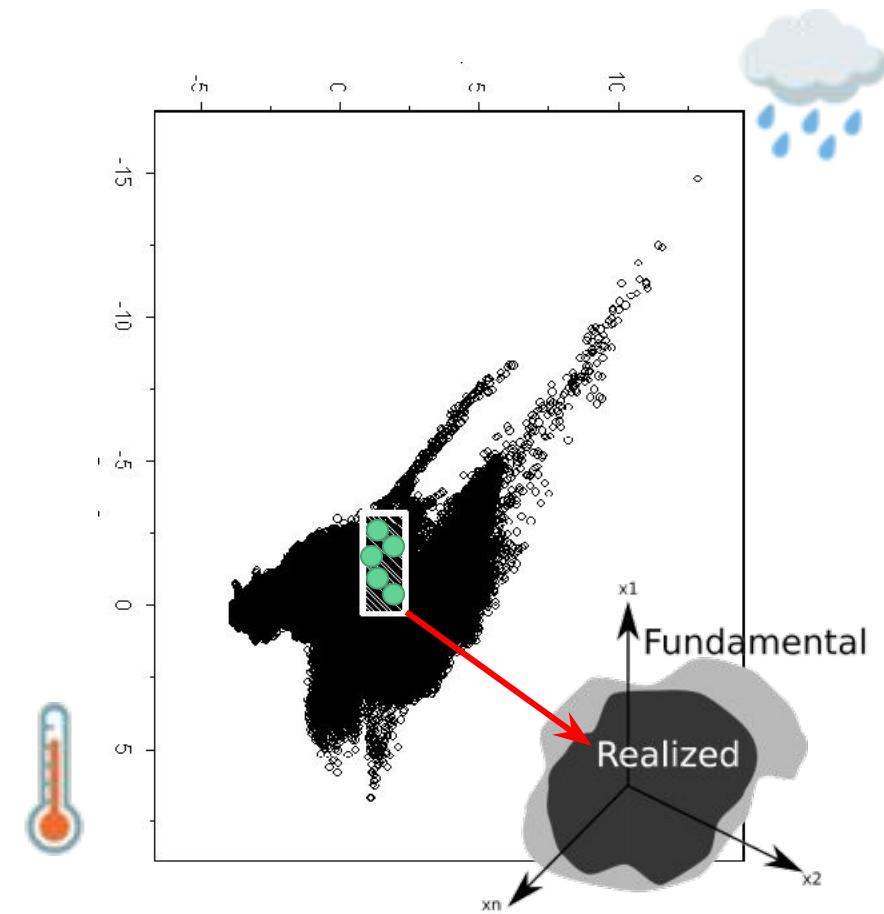


Espaço ambiental (E)



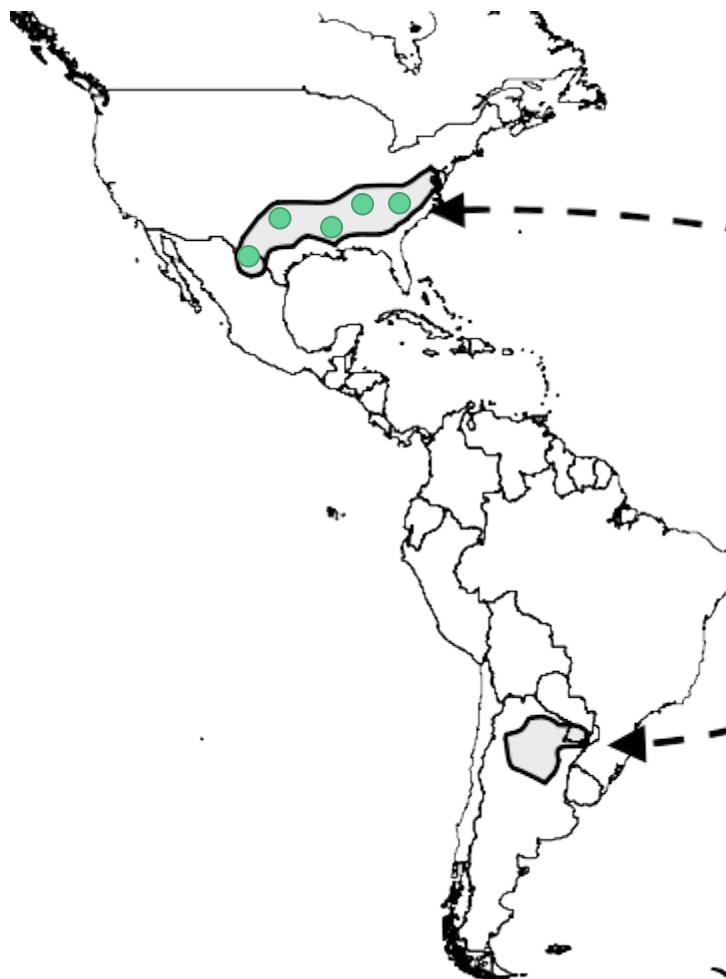
# Relação entre o espaço geográfico (G) e ambiental (E)

Espaço ambiental (E)

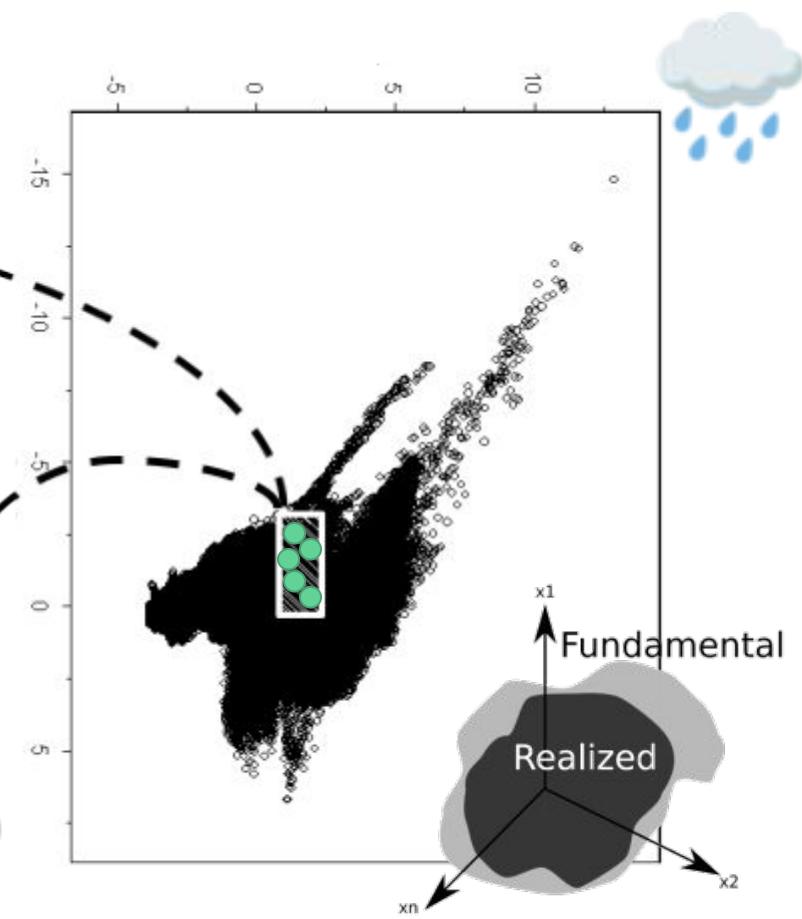


# Relação entre o espaço geográfico (G) e ambiental (E)

Espaço geográfico (G)



Espaço ambiental (E)



Como contornar  
essa extração?

# Nicho Ecológico

G



A



B



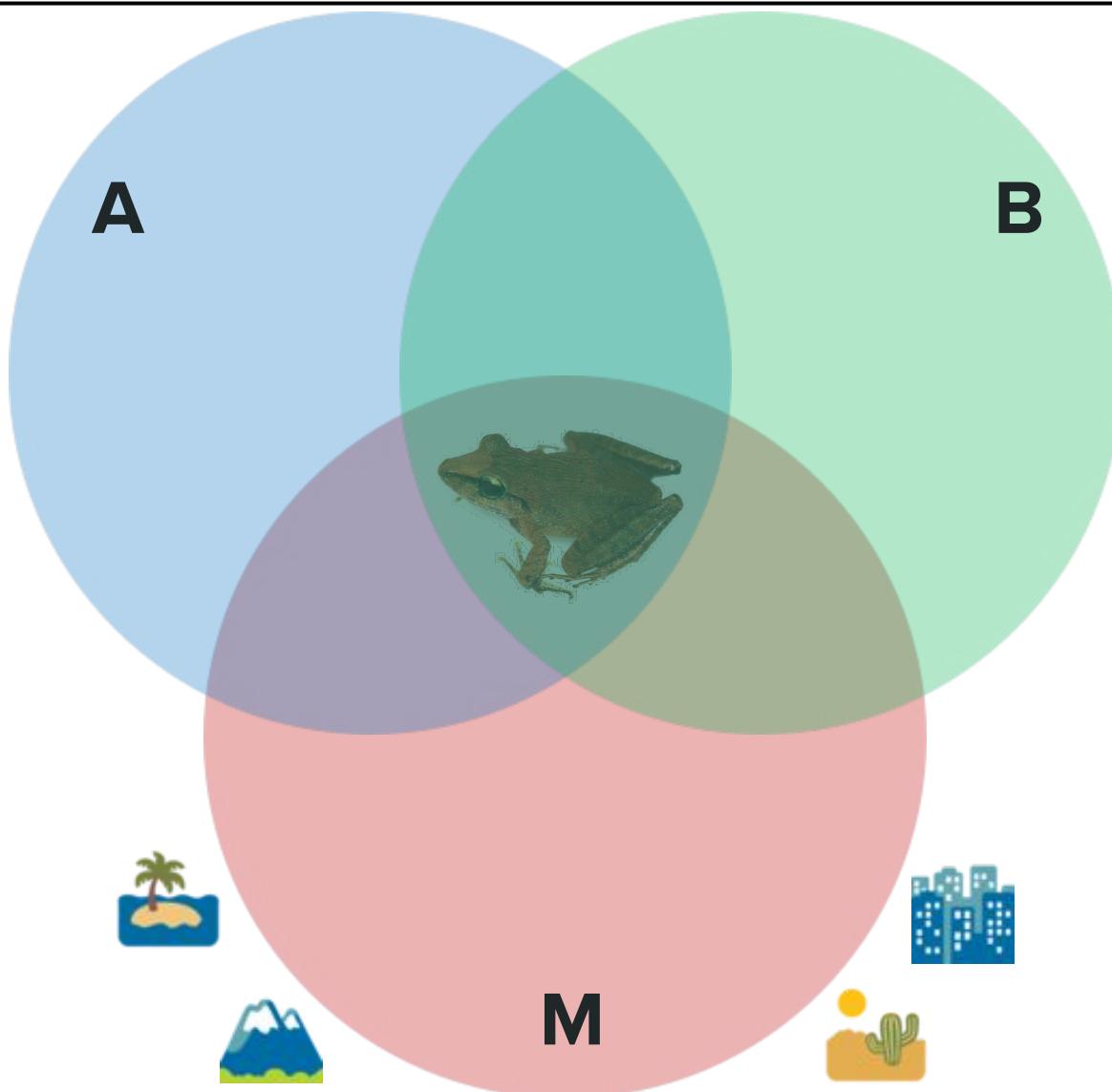
Peterson et al. (2011)

# Condições de Movimentação

G



A



M



Peterson et al. (2011)

# Condições de Movimentação

G



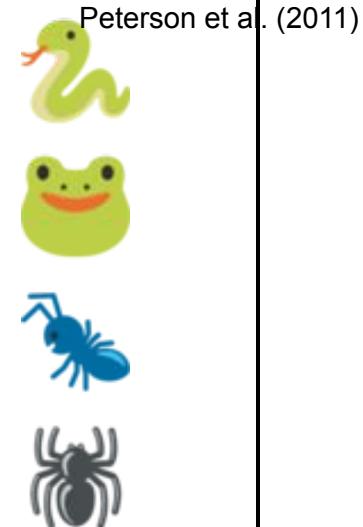
A



B



M



É possível inserir as  
interações biológicas  
nos modelos?

# Condições Modeladas

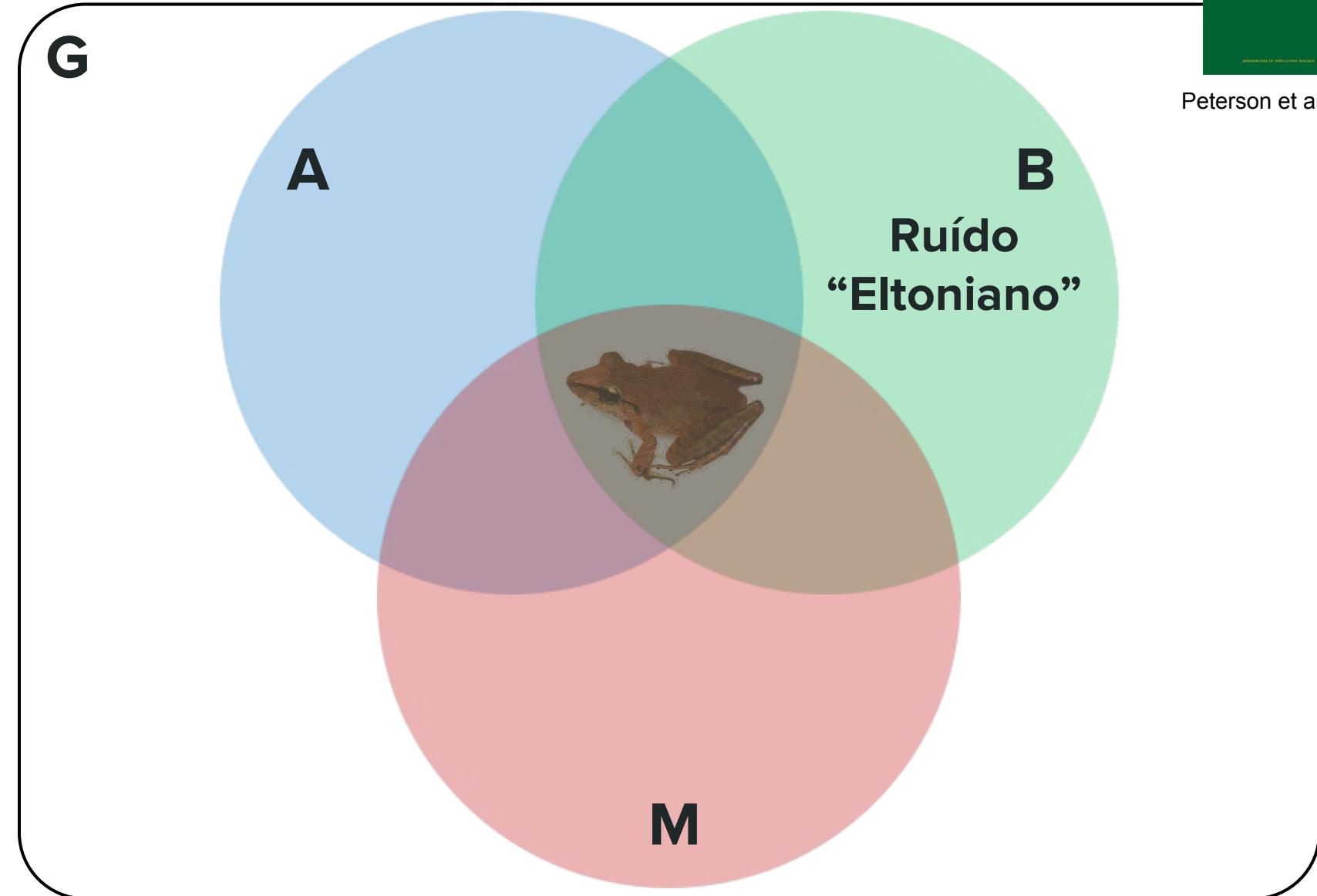
Peterson et al. (2011)

G

A

B

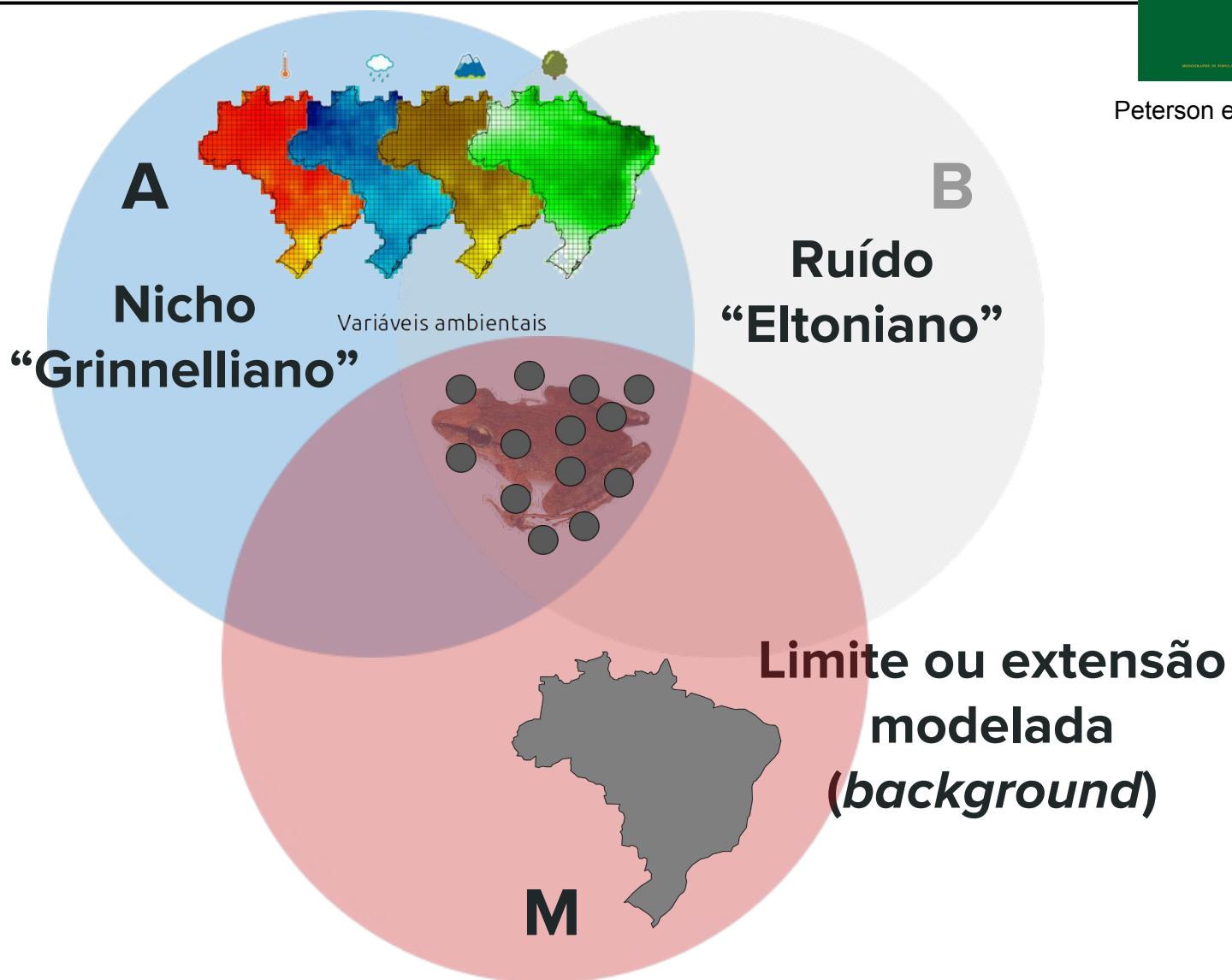
Ruído  
“Eltoniano”



M

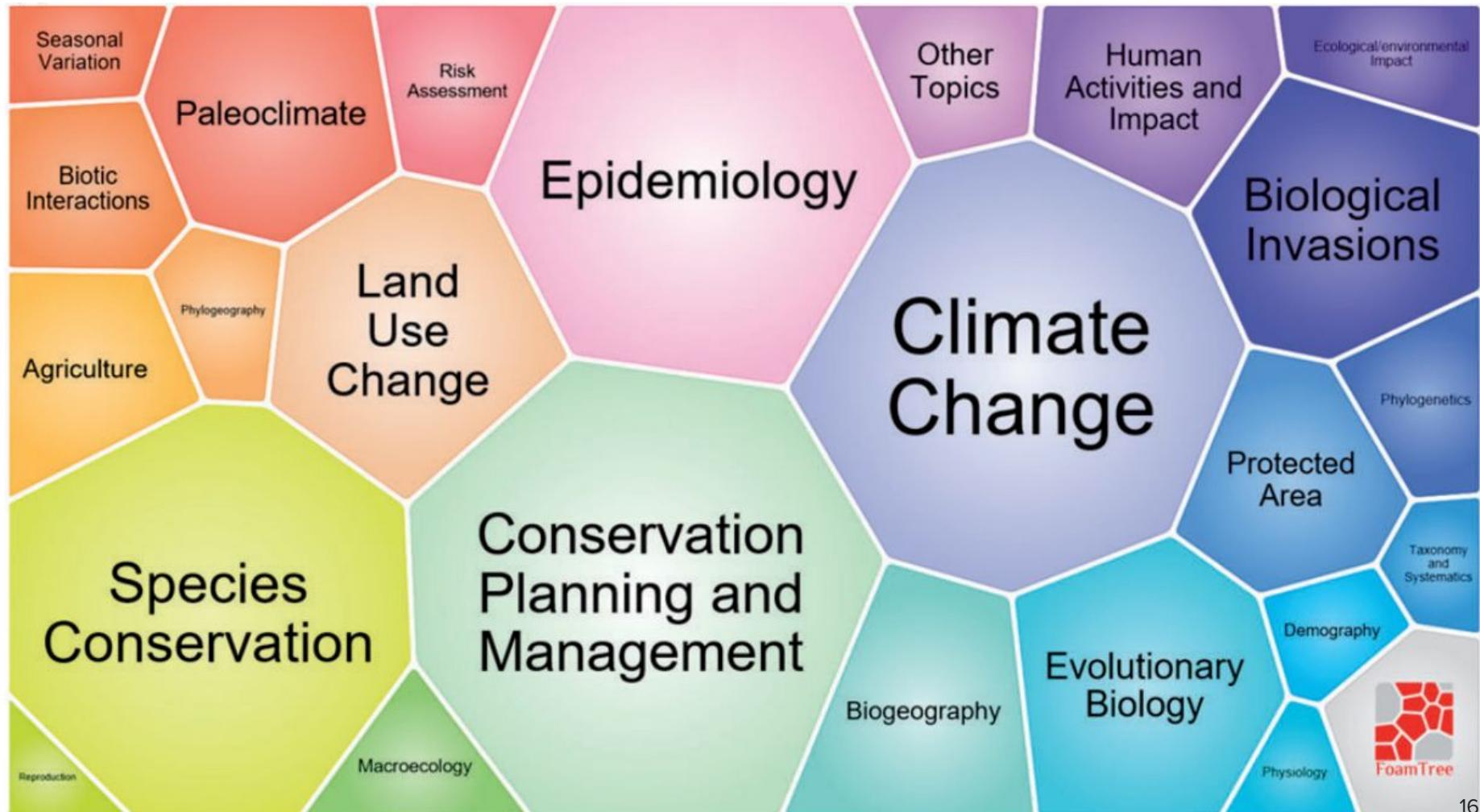
# Condições Modeladas

G

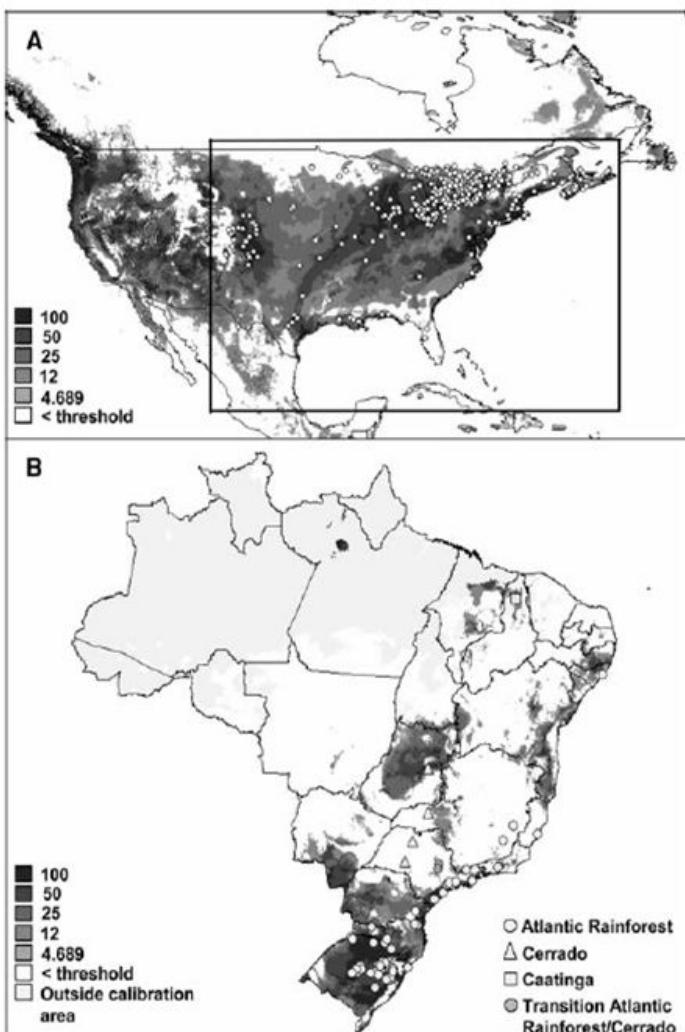


# Aplicação dos MDEs

Urbina-Cardona, N. et al. "Species Distribution Modeling in Latin America: A 25-Year Retrospective Review." *Tropical Conservation Science* 12 (2019).



# Aplicação dos MDEs



Biol Invasions  
DOI 10.1007/s10530-007-9154-5

ORIGINAL PAPER

## Predicting the potential distribution of the alien invasive American bullfrog (*Lithobates catesbeianus*) in Brazil

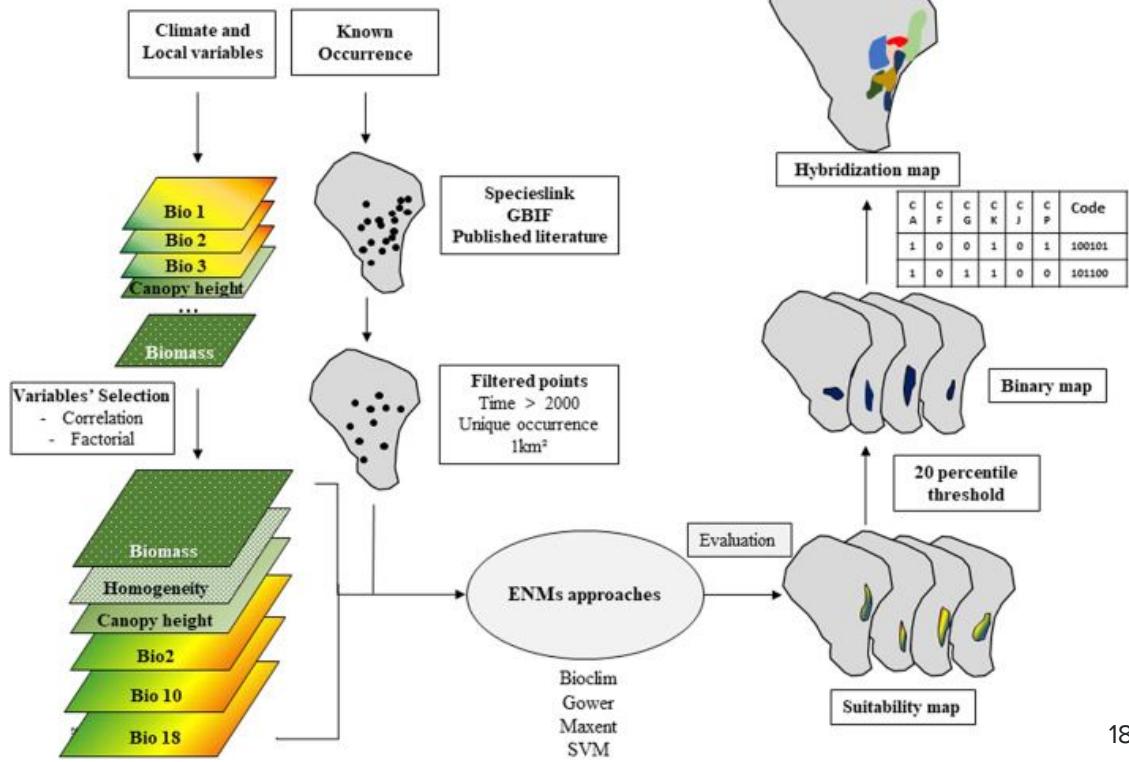
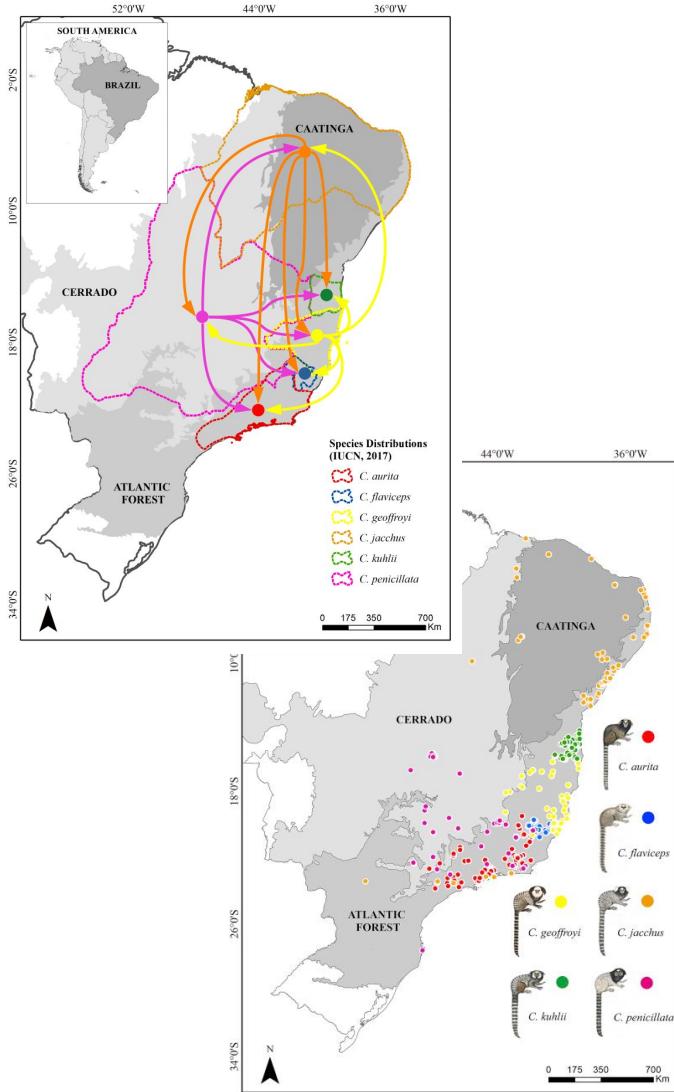
João G. R. Giovanelli · Célio F. B. Haddad ·  
João Alexandrino



Foto: Carl D. Howe

Giovanelli et al., 2008. Biological Invasions

# Aplicação dos MDEs



# Aplicação dos MDEs



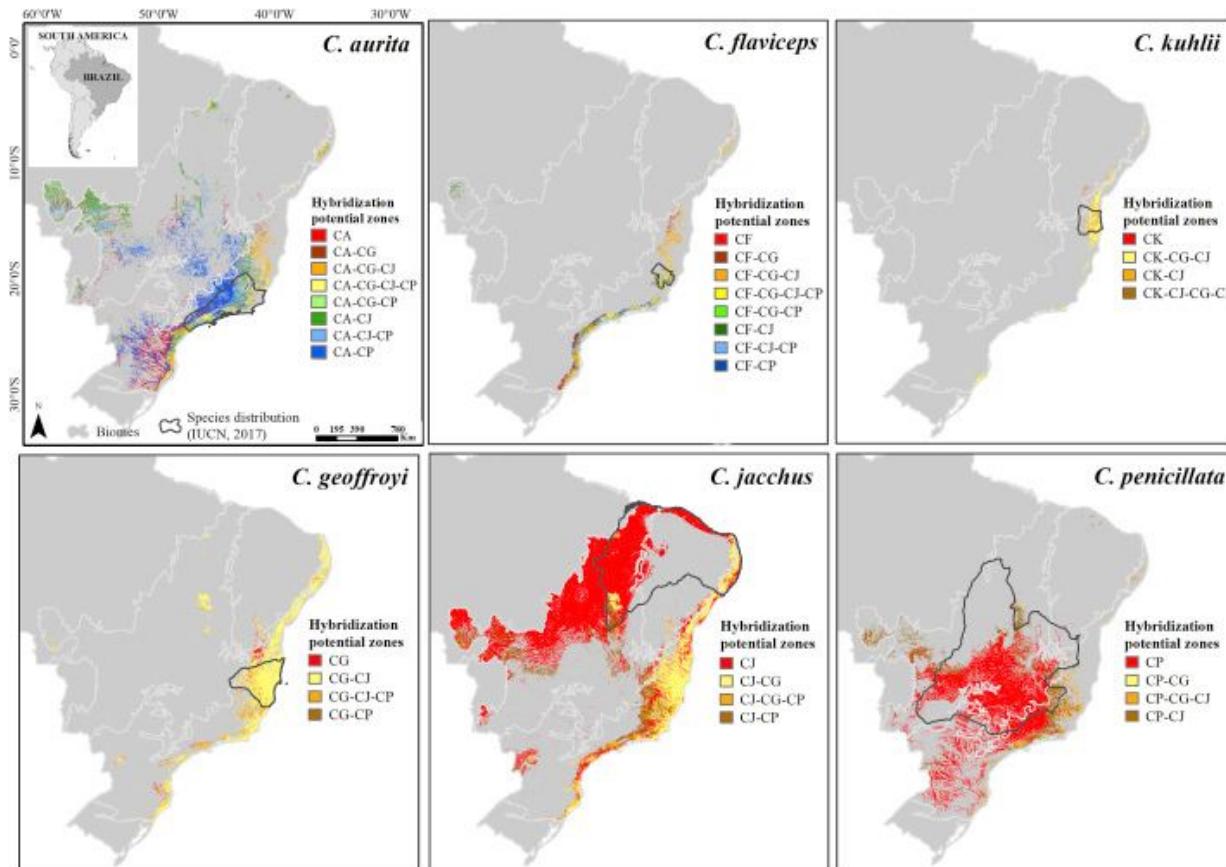
Global Ecology and Conservation  
Volume 20, October 2019, e00706



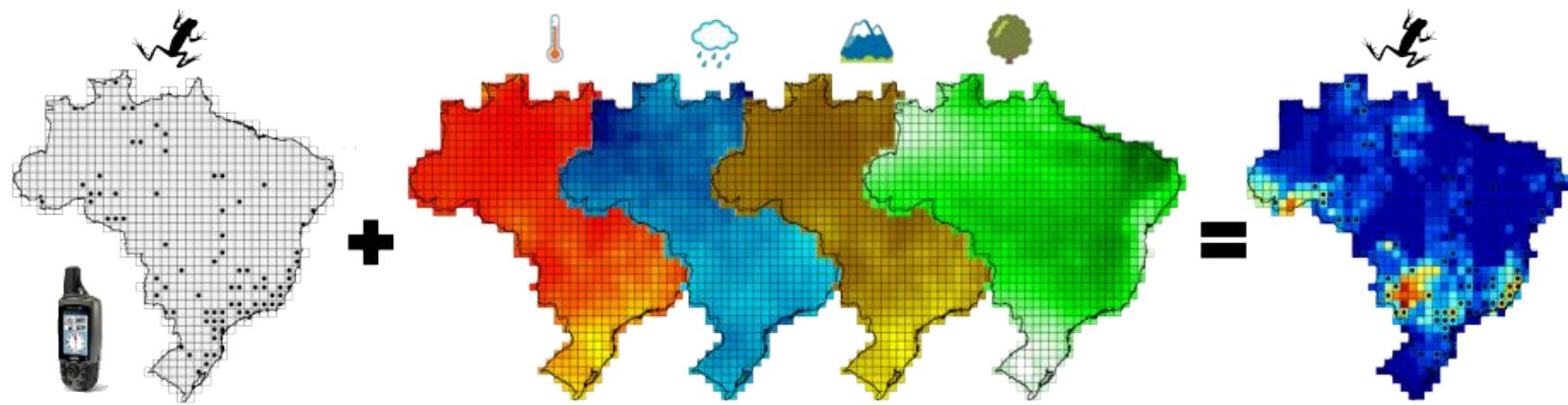
Original Research Article

## Predicting the potential hybridization zones between native and invasive marmosets within Neotropical biodiversity hotspots

Andreia Magro Moraes<sup>a</sup> , Mauricio Humberto Vancine<sup>b</sup>, Andreia Magro Moraes<sup>c</sup>, Carlos Leandro de Oliveira Cordeiro<sup>d, e</sup>, Miriam Plaza Pinto<sup>f</sup>, Adriana Almeida Lima<sup>f</sup>, Laurence Culot<sup>g</sup>, Thiago Sanna Freira Silva<sup>e</sup>, Rosane Garcia Collevatti<sup>h</sup>, Milton Cesar Ribeiro<sup>i</sup>, Thadeu Sobral-Souza<sup>j</sup>



# Operacionalização dos MDEs



"Ocorrências"

Variáveis ambientais

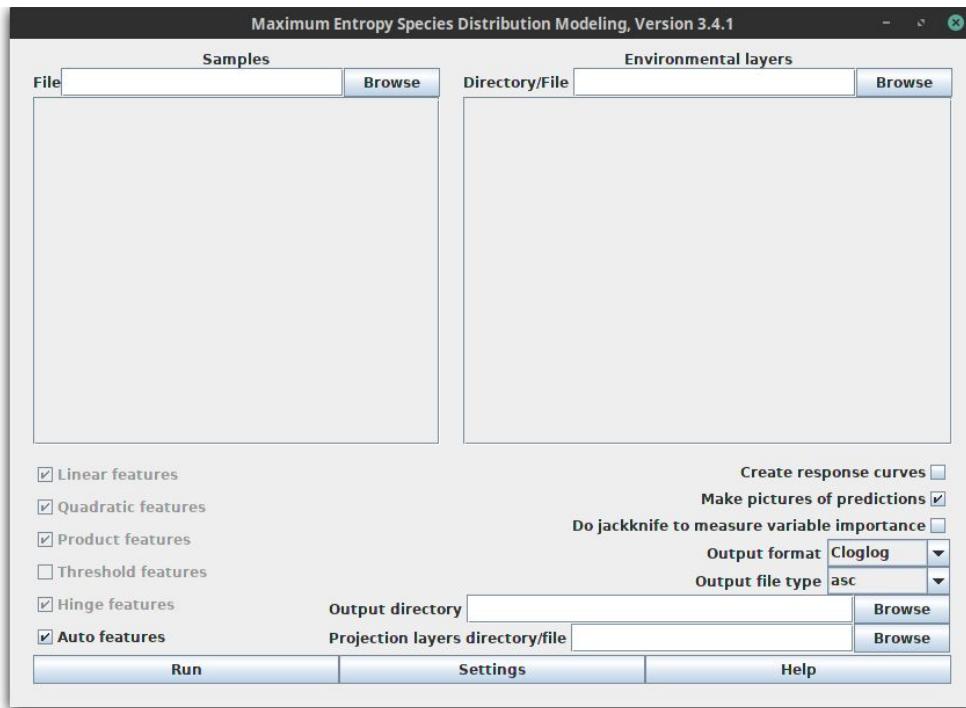
Adequabilidade

species	lon	lat
sp1	-40.2	-23.4
sp1	-38.8	-20.3
sp1	-43.3	-19.9

variaveis
temperatura
precipitação
relevo

valores
0
até
1

# Construção dos MDEs



Pacotes  
dismo  
biomod2  
sdm  
SSDM  
ENMeval  
maxlike  
maxnet  
hSDM  
iSDM  
jSDM  
zoom...

# Introdução à linguagem R

## Tópicos

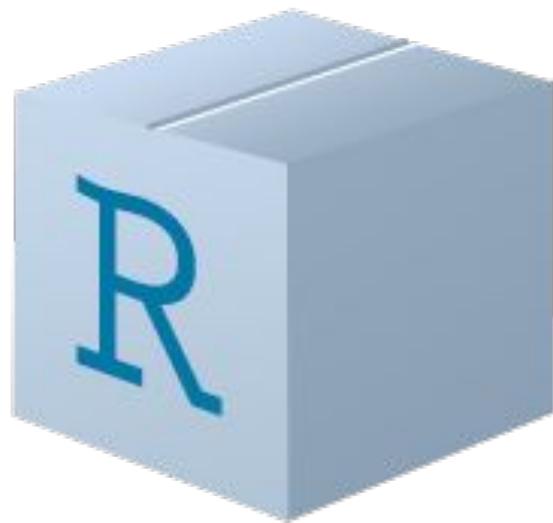
- 1.Linguagem R
- 2.Software RStudio
- 3.Script
- 4.Comentário
- 5.Atribuição
- 6.Função
- 7.Diretório
- 8.Pacotes





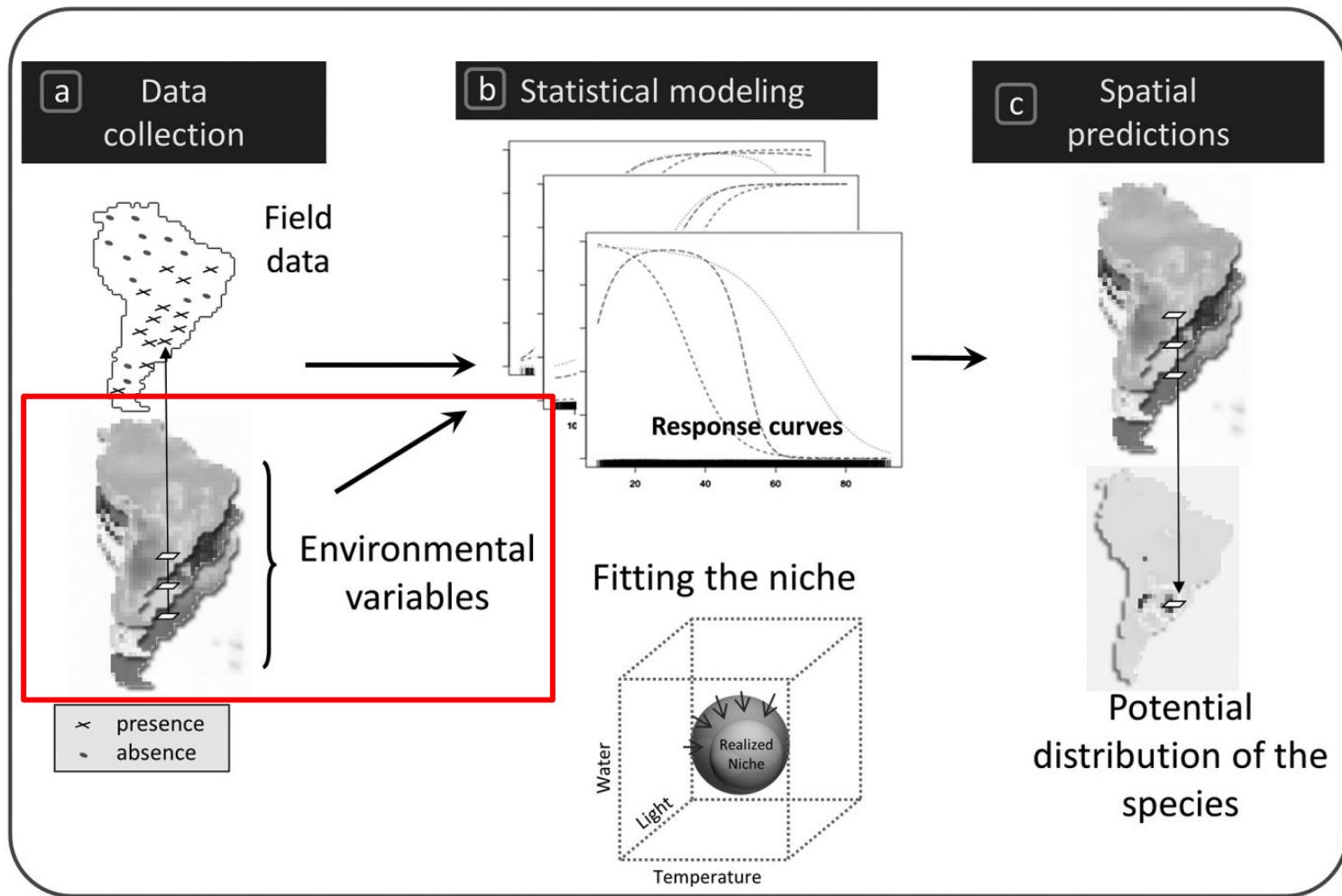
# Script - Pacotes

00\_script\_install\_packages.R

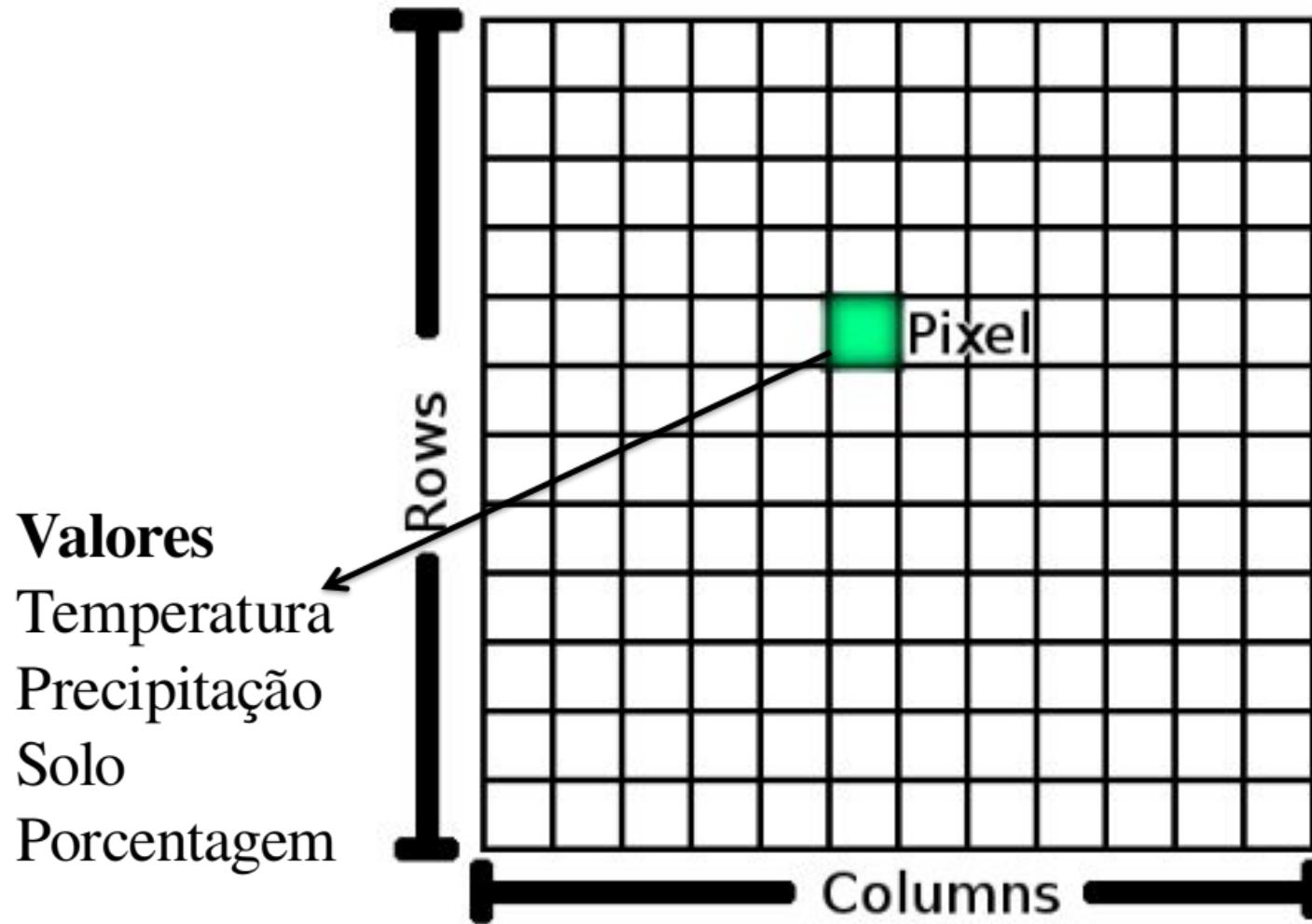


Guisan et al. (2017)

# Dados de entrada - Variáveis

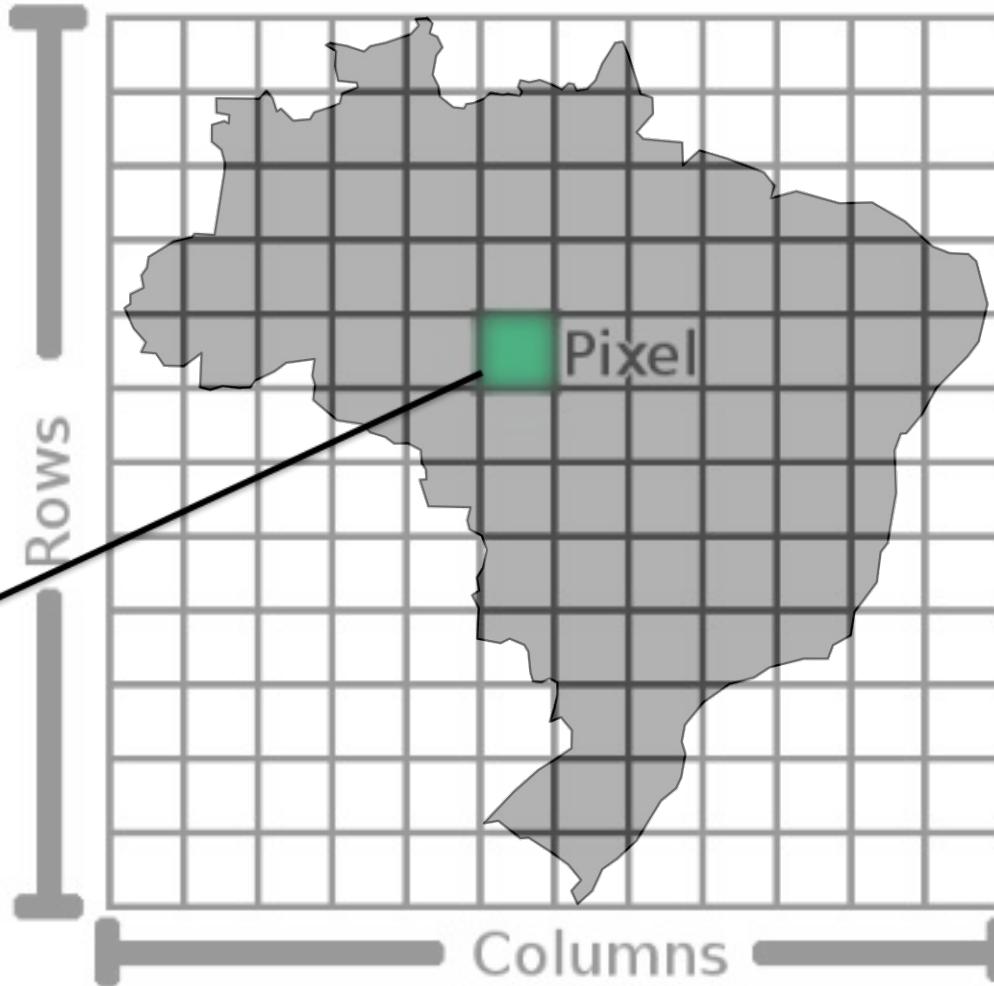


# Variáveis - Raster - Extensão e resolução

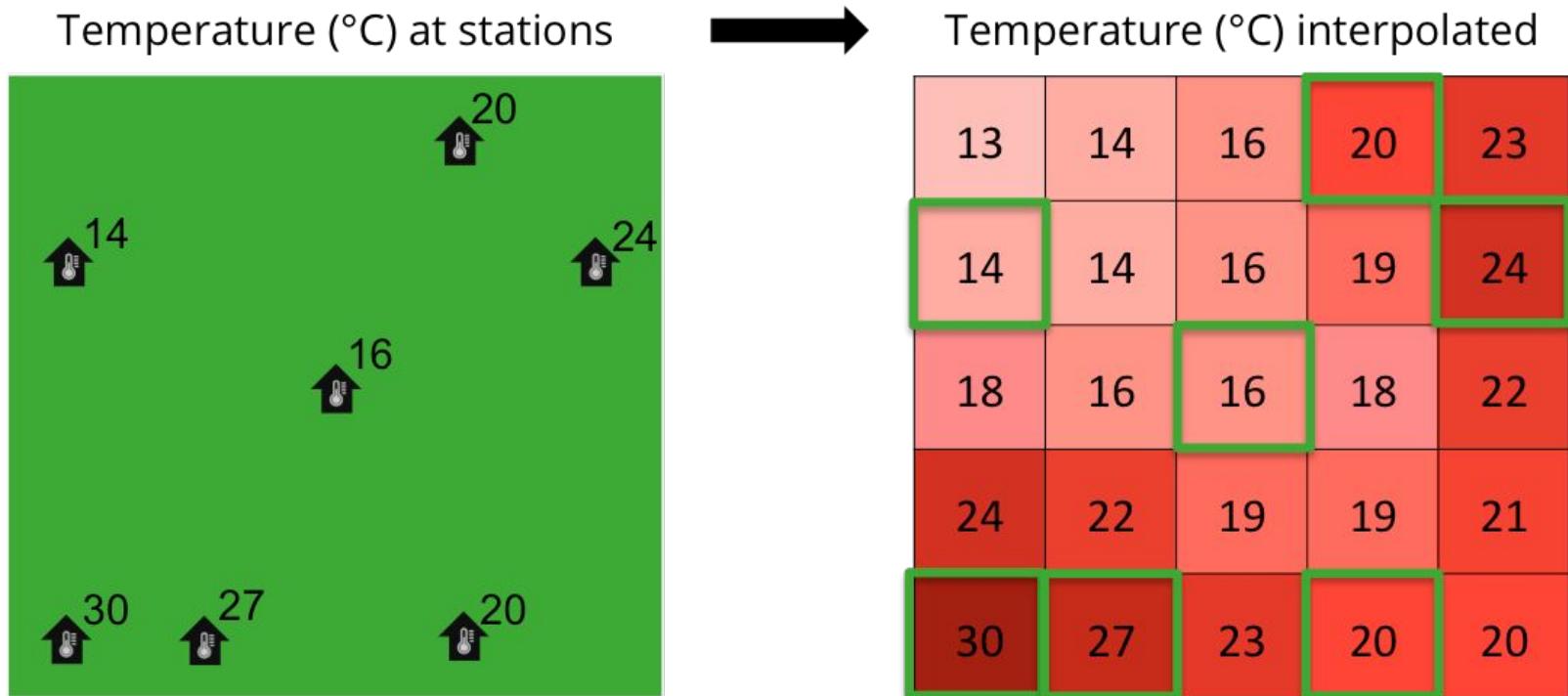


# Variáveis - Raster - Extensão e resolução

**Valores**  
Temperatura  
Precipitação  
Solo  
Porcentagem



# Dados de entrada - Variáveis



Adapted from [http://planet.botany.uwc.ac.za/nisl/GIS/spatial/chap\\_1\\_11.h](http://planet.botany.uwc.ac.za/nisl/GIS/spatial/chap_1_11.h)

# Dados de entrada - Variáveis

<http://worldclim.org/version2>

The screenshot shows the WorldClim Version 2 homepage. At the top, it says "WorldClim - Global Climate Data" and "Free climate data for ecological modeling and GIS". Below that is a "Contact" button and a "Home" link. The main title "WorldClim Version2" is centered above a paragraph of text. A table below the text lists variables and their resolutions. At the bottom, there is a red-bordered box containing another table for bioclimatic variables.

**WorldClim Version2**

WorldClim version 2 has average monthly climate data for minimum, mean, and maximum temperature and for precipitation for 1970-2000.

You can download the variables for different spatial resolutions, from 30 seconds (~1 km<sup>2</sup>) to 10 minutes (~340 km<sup>2</sup>). Each download is a "zip" file containing 12 GeoTiff (.tif) files, one for each month of the year (January is 1; December is 12).

variable	10 minutes	5 minutes	2.5 minutes	30 seconds
minimum temperature (°C)	tmin 10m	tmin 5m	tmin 2.5m	tmin 30s
maximum temperature (°C)	tmax 10m	tmax 5m	tmax 2.5m	tmax 30s
average temperature (°C)	tavg 10m	tavg 5m	tavg 2.5m	tavg 30s
precipitation (mm)	prec 10m	prec 5m	prec 2.5m	prec 30s
solar radiation (kJ m <sup>-2</sup> day <sup>-1</sup> )	srad 10m	srad 5m	srad 2.5m	srad 30s
wind speed (m s <sup>-1</sup> )	wind 10m	wind 5m	wind 2.5m	wind 30s
water vapor pressure (kPa)	vapr 10m	vapr 5m	vapr 2.5m	vapr 30s

Below you can download the standard (19) WorldClim [Bioclimatic variables](#) for WorldClim version 2. They are the average for the years 1970-2000. Each download is a "zip" file containing 19 GeoTiff (.tif) files, one for each month of the variables.

variable	10 minutes	5 minutes	2.5 minutes	30 seconds
Bioclimatic variables	bio 10m	bio 5m	bio 2.5m	bio 30s

Citation:

# Dados de entrada - Variáveis

WorldClim - Global Climate Data  
Free climate data for ecological modeling and GIS  
Contact  
Home

## Bioclimatic variables

Bioclimatic variables are derived from the monthly temperature and rainfall values in order to generate more biologically meaningful variables. These are often used in [species distribution modeling](#) and related ecological modeling techniques. The bioclimatic variables represent annual trends (e.g., mean annual temperature, annual precipitation) seasonality (e.g., annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g., temperature of the coldest and warmest month, and precipitation of the wet and dry quarters). A quarter is a period of three months (1/4 of the year).

They are coded as follows:

BIO1 = Annual Mean Temperature  
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))  
BIO3 = Isothermality ( $BIO2/BIO7$ ) (\* 100)  
BIO4 = Temperature Seasonality (standard deviation \*100)  
BIO5 = Max Temperature of Warmest Month  
BIO6 = Min Temperature of Coldest Month  
BIO7 = Temperature Annual Range ( $BIO5-BIO6$ )  
BIO8 = Mean Temperature of Wettest Quarter  
BIO9 = Mean Temperature of Driest Quarter  
BIO10 = Mean Temperature of Warmest Quarter  
BIO11 = Mean Temperature of Coldest Quarter  
BIO12 = Annual Precipitation  
BIO13 = Precipitation of Wettest Month  
BIO14 = Precipitation of Driest Month  
BIO15 = Precipitation Seasonality (Coefficient of Variation)  
BIO16 = Precipitation of Wettest Quarter  
BIO17 = Precipitation of Driest Quarter  
BIO18 = Precipitation of Warmest Quarter  
BIO19 = Precipitation of Coldest Quarter

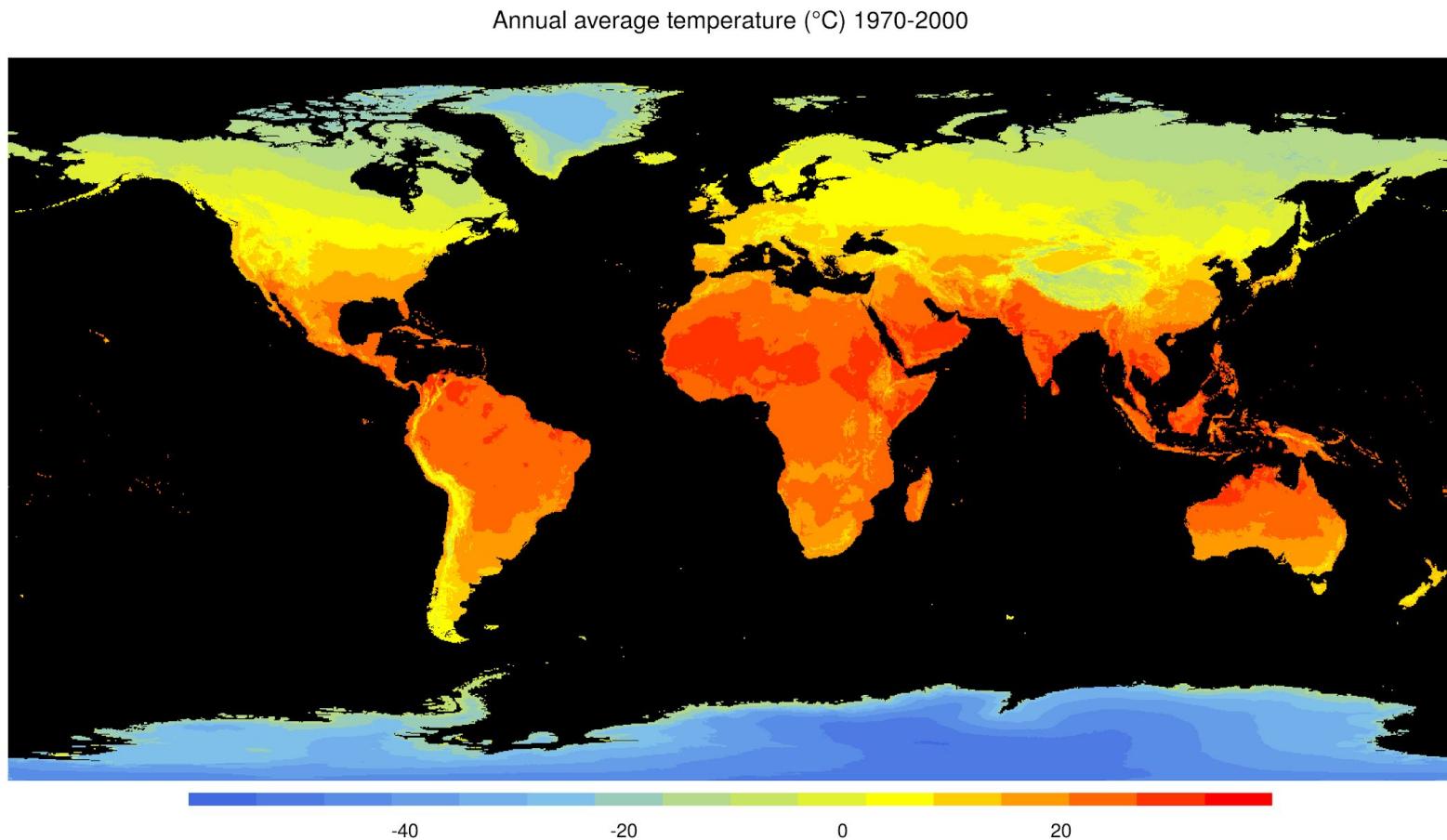
BIO01 = Temperatura média anual  
BIO02 = Variação Diurna Média de Temperatura (Média mensal (Tmax-Tmin))  
BIO03 = Isothermalidade ( $(BIO2/BIO7) (* 100)$ )  
BIO04 = Sazonalidade da Temperatura (desvio padrão \* 100)  
BIO05 = Temperatura máxima do mês mais quente  
BIO06 = Temperatura mínima do mês mais frio  
BIO07 = Amplitude térmica anual ( $BIO5-BIO6$ )  
BIO08 = Temperatura média do trimestre mais úmido  
BIO09 = Temperatura média do trimestre mais seco  
BIO10 = Temperatura média do trimestre mais quente  
BIO11 = Temperatura média do trimestre mais frio

Temperatura

BIO12 = Precipitação Anual  
BIO13 = Precipitação do mês mais chuvoso  
BIO14 = Precipitação do mês mais seco  
BIO15 = Sazonalidade da Precipitação (coeficiente de variação)  
BIO16 = Precipitação do trimestre mais chuvoso  
BIO17 = Precipitação do trimestre mais seco  
BIO18 = Precipitação do trimestre mais quente  
BIO19 = Precipitação do trimestre mais frio

Precipitação

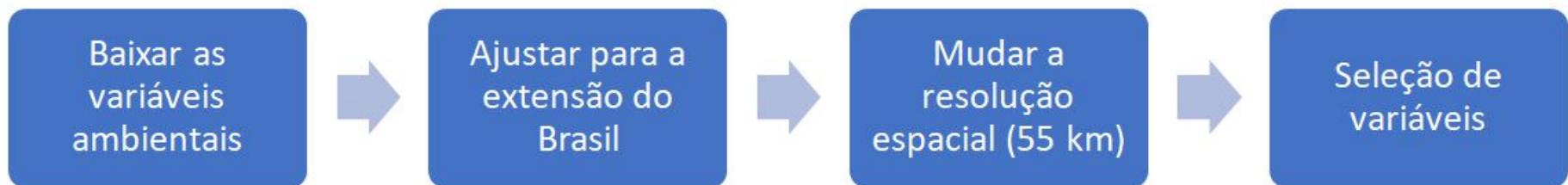
# Dados de entrada - Variáveis



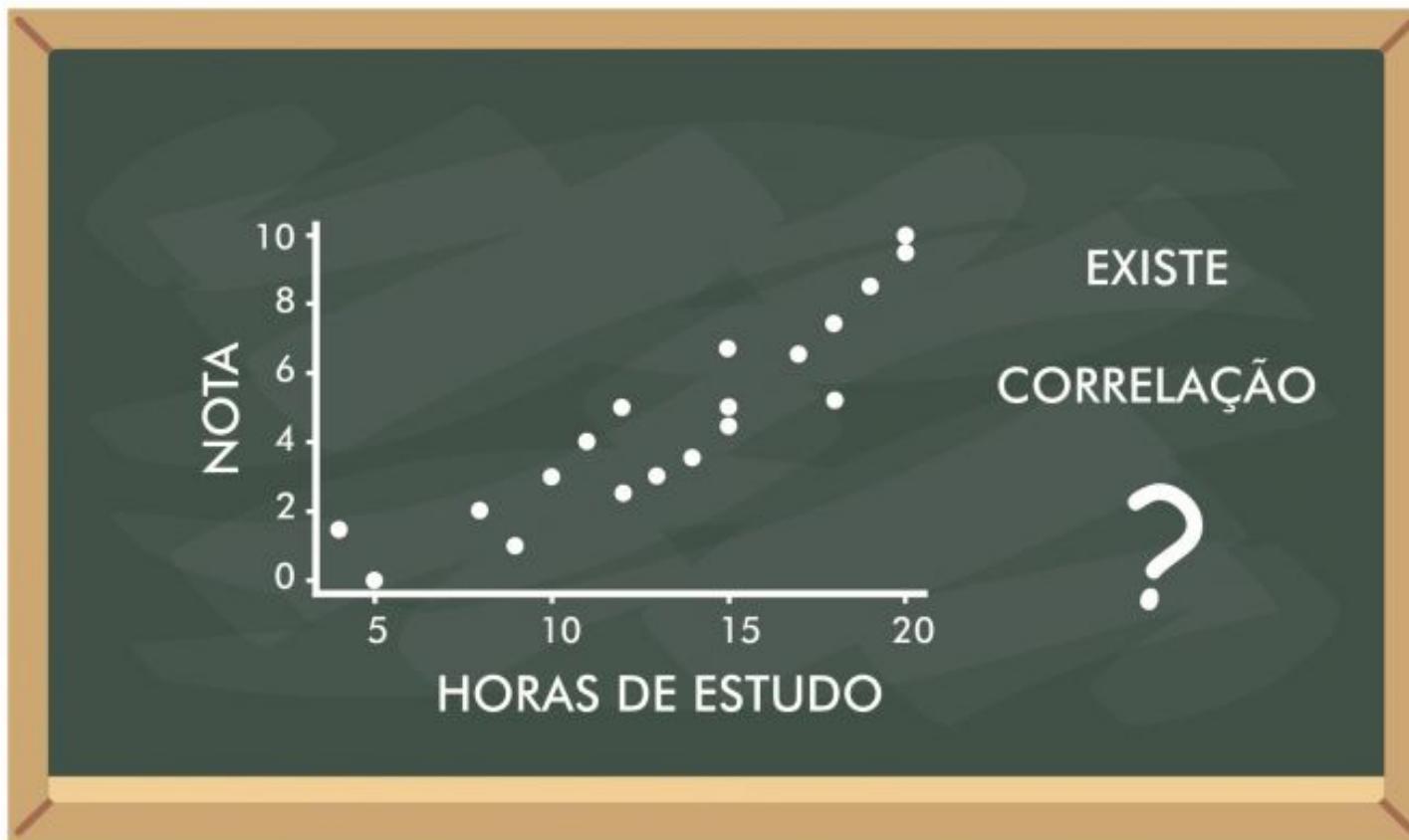


# Script - Variáveis

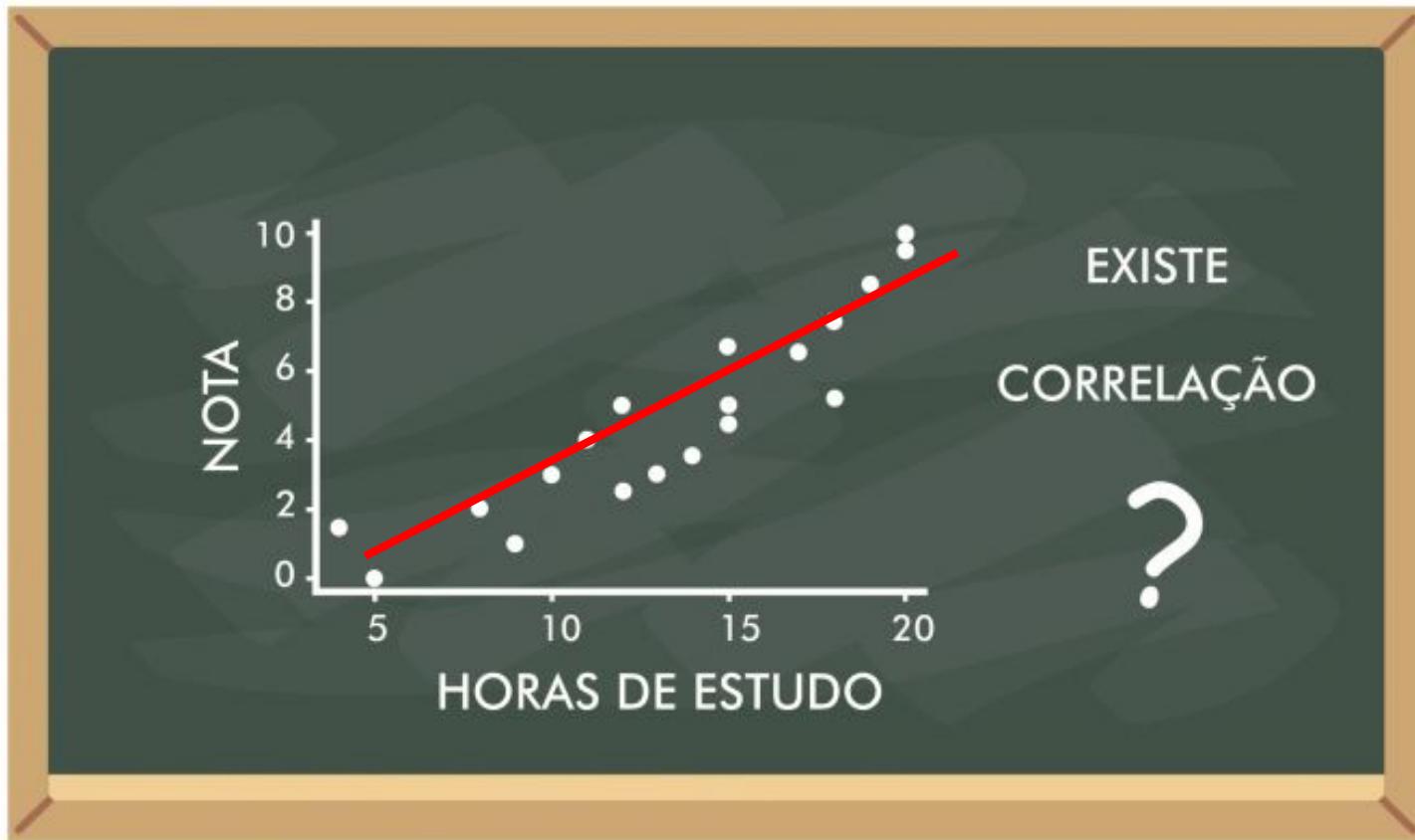
## 01\_script\_var.R



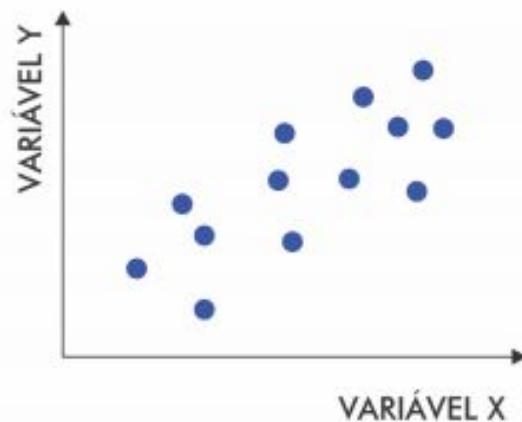
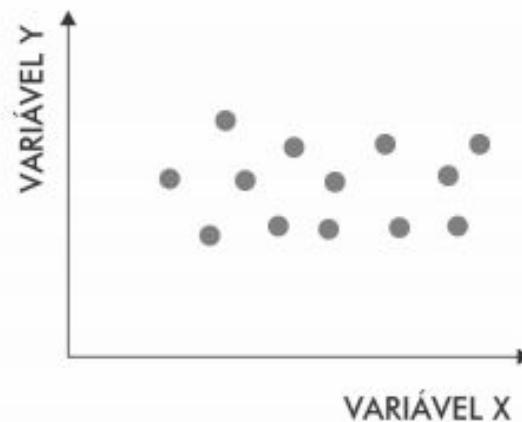
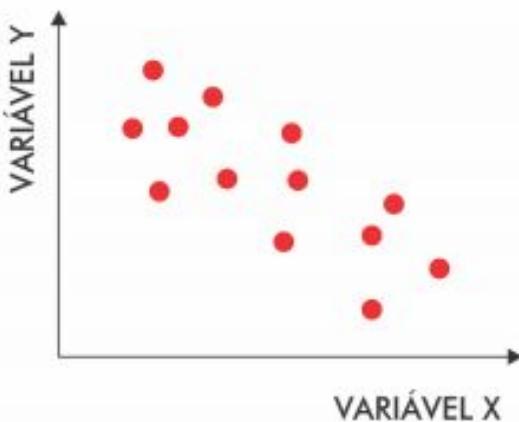
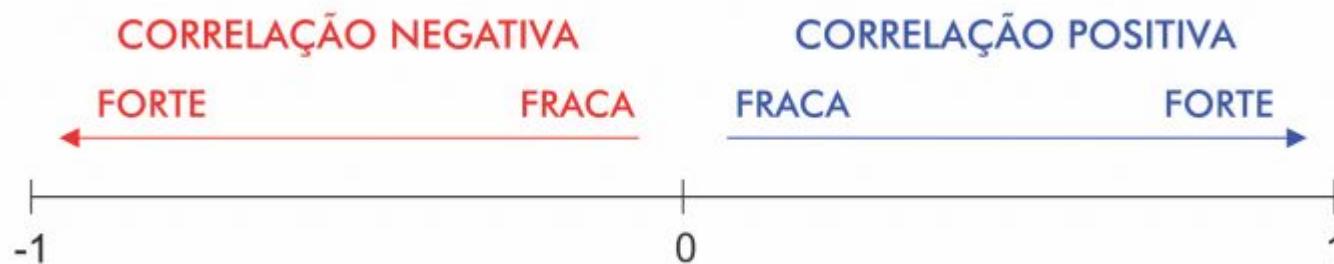
# Variáveis - Seleção - Correlação



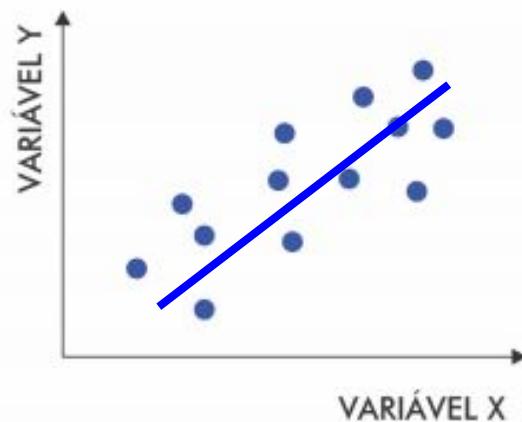
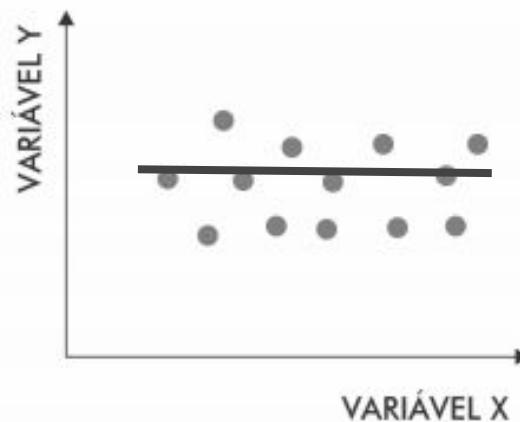
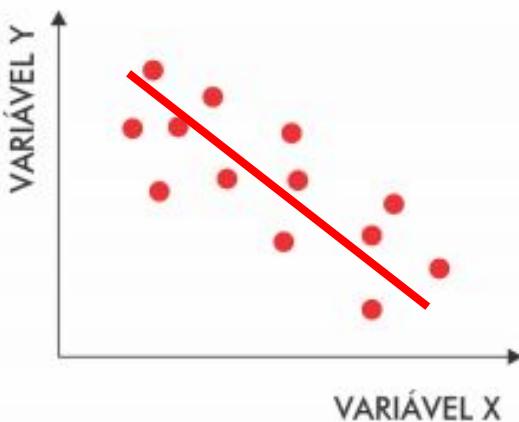
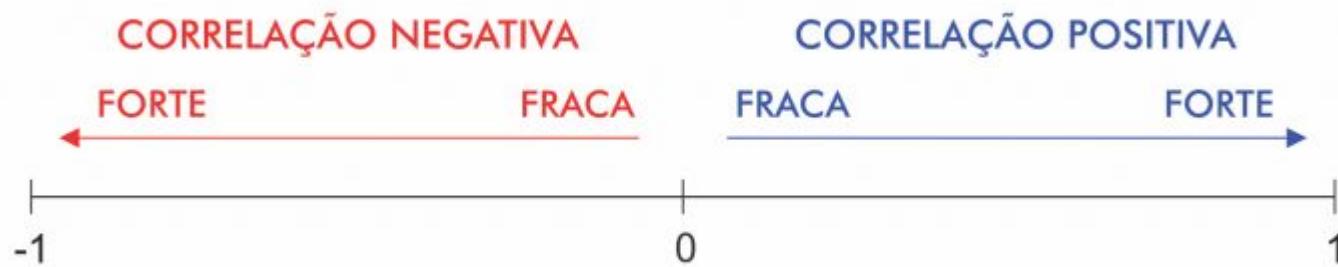
# Variáveis - Seleção - Correlação

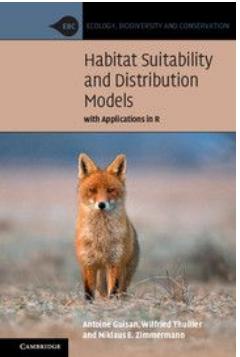


# Variáveis - Seleção - Correlação

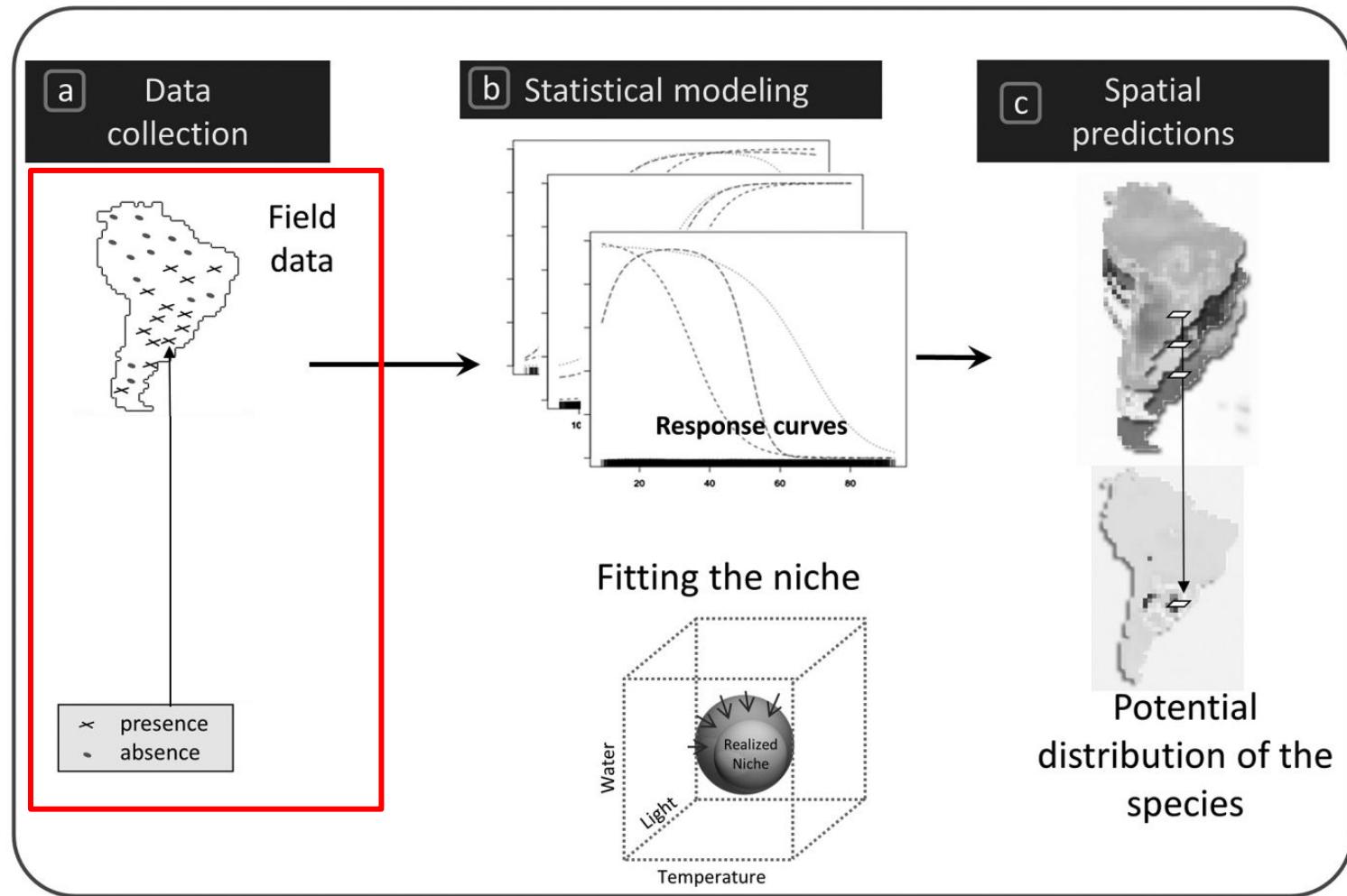


# Variáveis - Seleção - Correlação





# Dados de entrada - Ocorrências



Guisan et al. (2017)

# Ocorrências

## 1. Coleta em campo



# Ocorrências

1. Coleta em campo



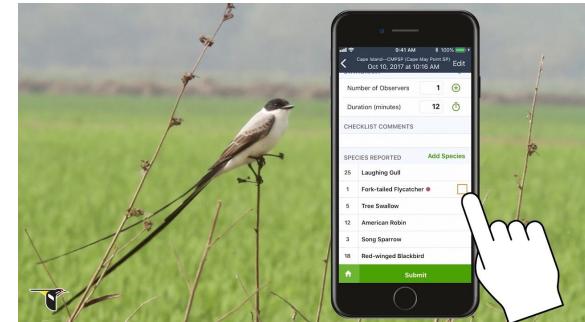
2. Literatura (artigos, data papers, ...)

# Ocorrências

1. Coleta em campo



2. Literatura (artigos, data papers, ...)



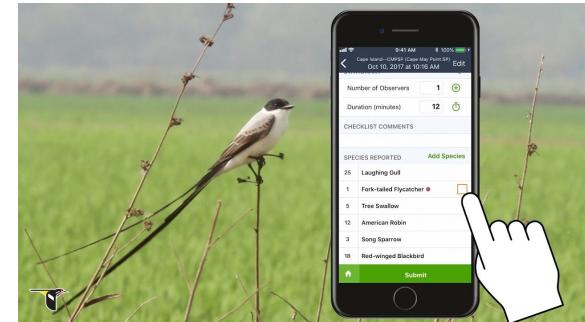
3. Naturalistas (e-Bird, iNaturalist, ...)

# Ocorrências

1. Coleta em campo



2. Literatura (artigos, data papers, ...)



3. Naturalistas (e-Bird, iNaturalist, ...)

4. Coleções científicas e museus (Museu Nacional, MZUSP, CFHB, ...)

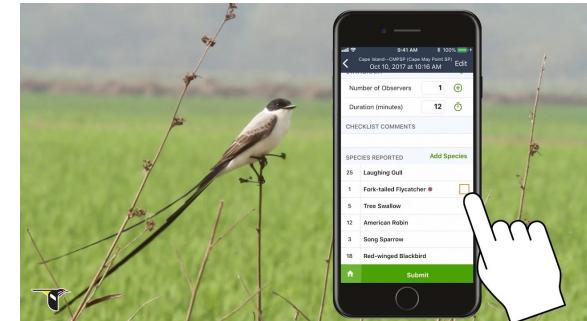


# Ocorrências

1. Coleta em campo



2. Literatura (artigos, data papers, ...)

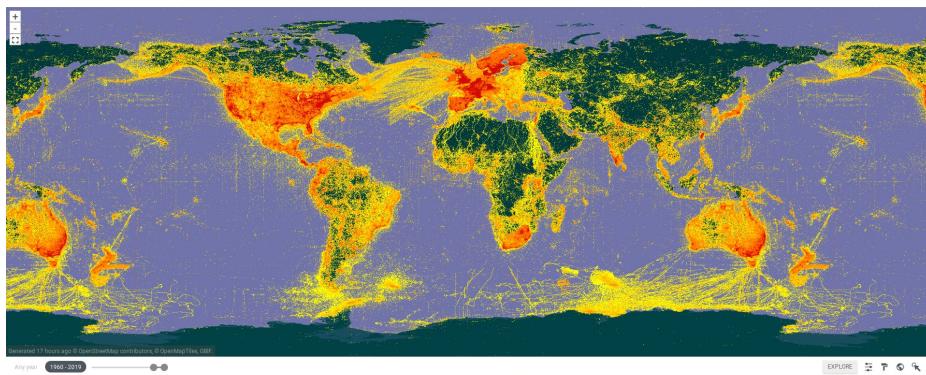


3. Naturalistas (e-Bird, iNaturalist, ...)

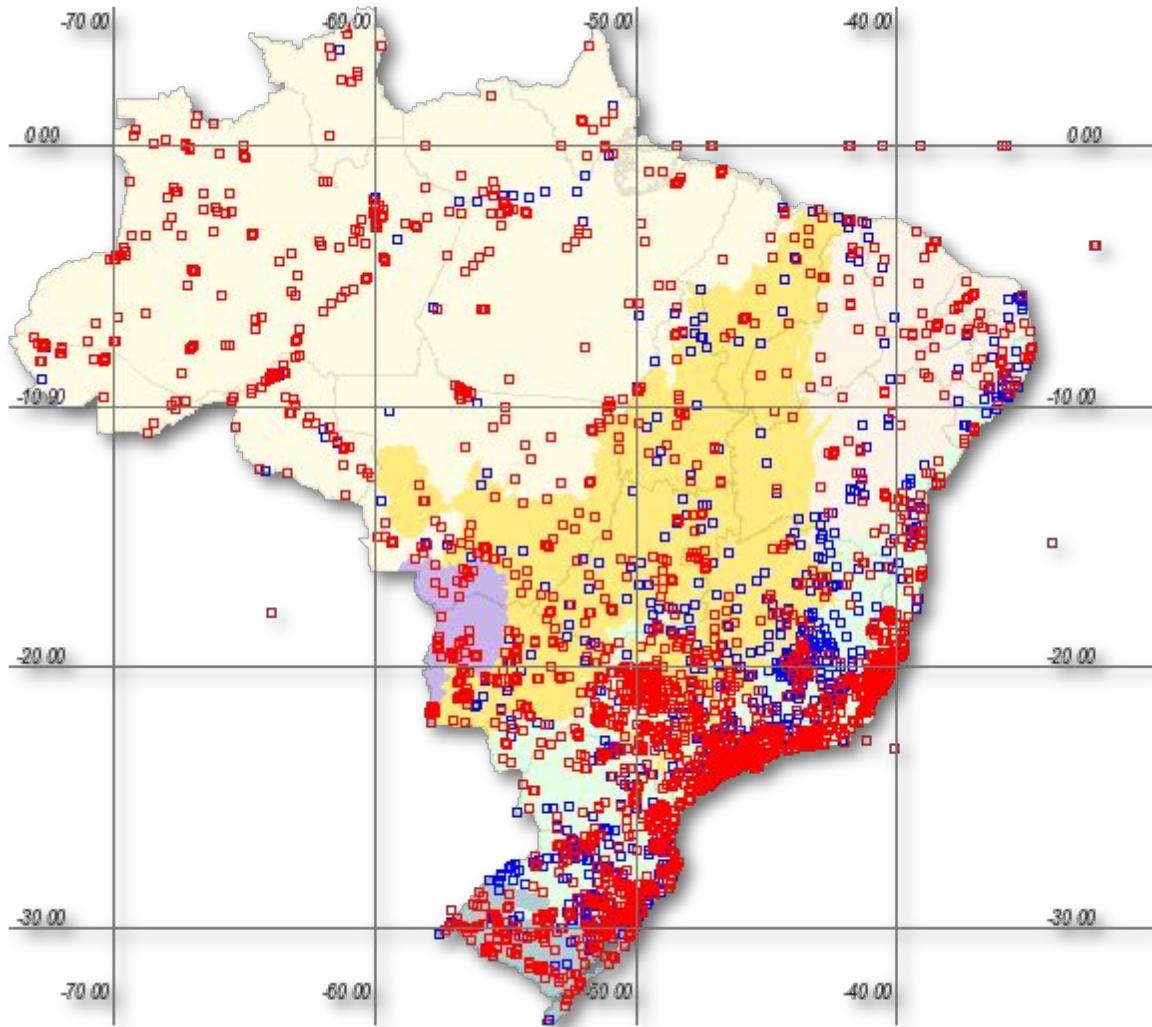
4. Coleções científicas e museus (Museu Nacional, MZUSP, CFHB, ...)



5. Banco de dados (GBIF, SpeciesLink, ...)

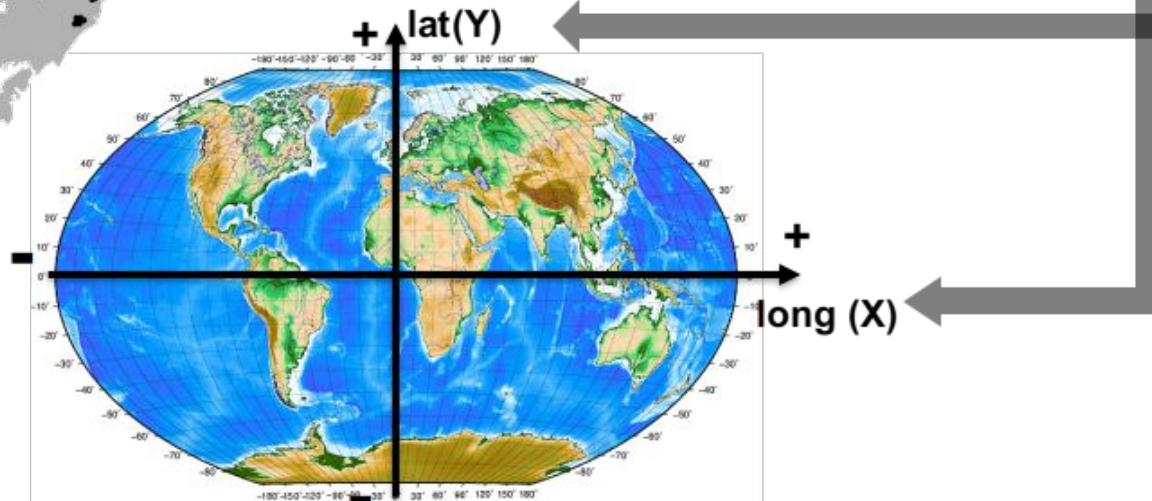
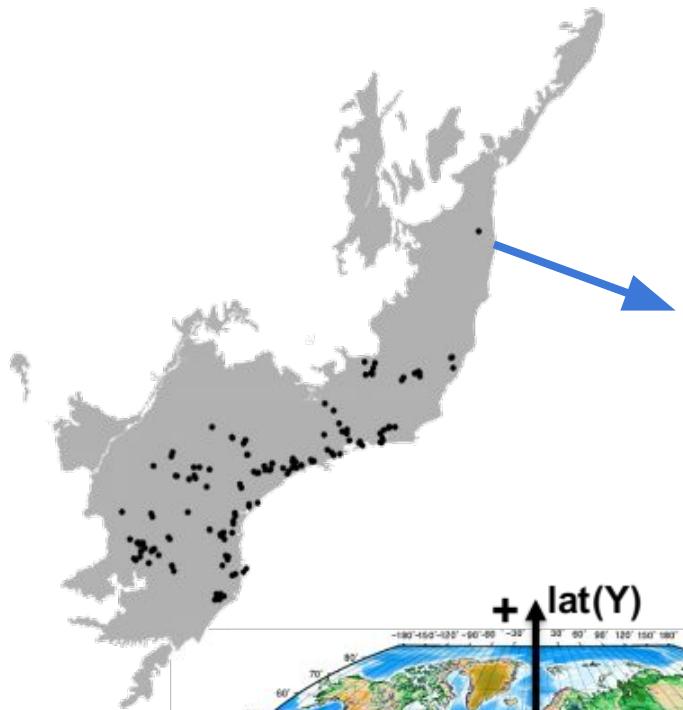


# Ocorrências



*species* link

# Ocorrências



sp	long	lat
vitreorana_uranoscopa	-52.8300	-26.4400
vitreorana_uranoscopa	-52.6836	-27.1253
vitreorana_uranoscopa	-52.5569	-26.5642
vitreorana_uranoscopa	-52.4500	-26.5667
vitreorana_uranoscopa	-52.4489	-27.0689
vitreorana_uranoscopa	-52.4147	-26.8667

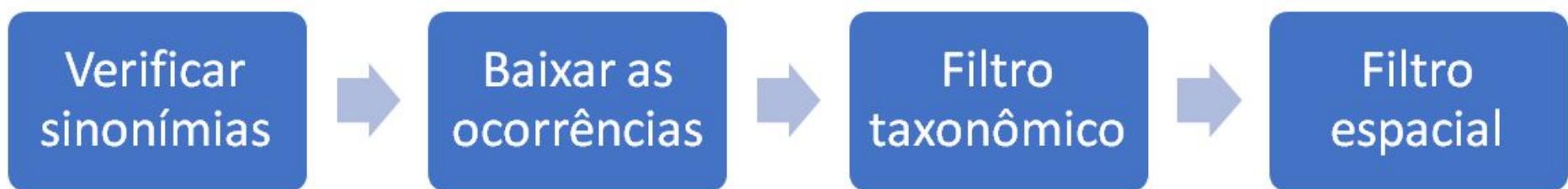
# Assumindo que a Terra é plana...





# Script - Ocorrências

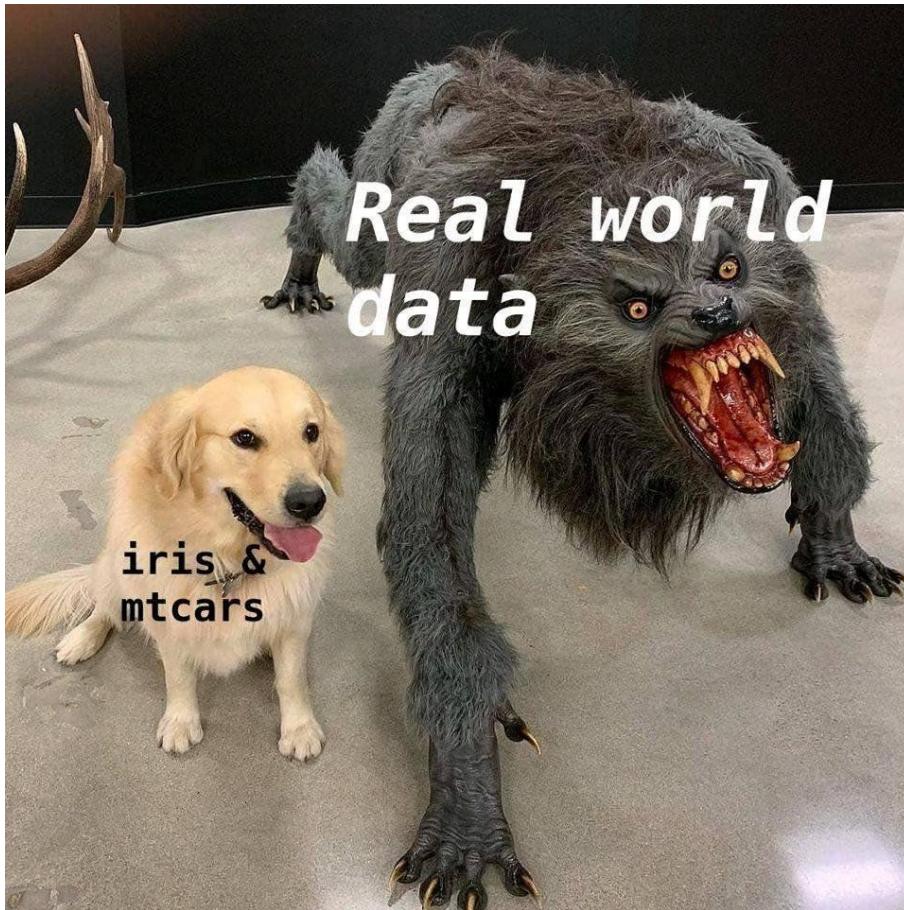
02\_script\_occ.R



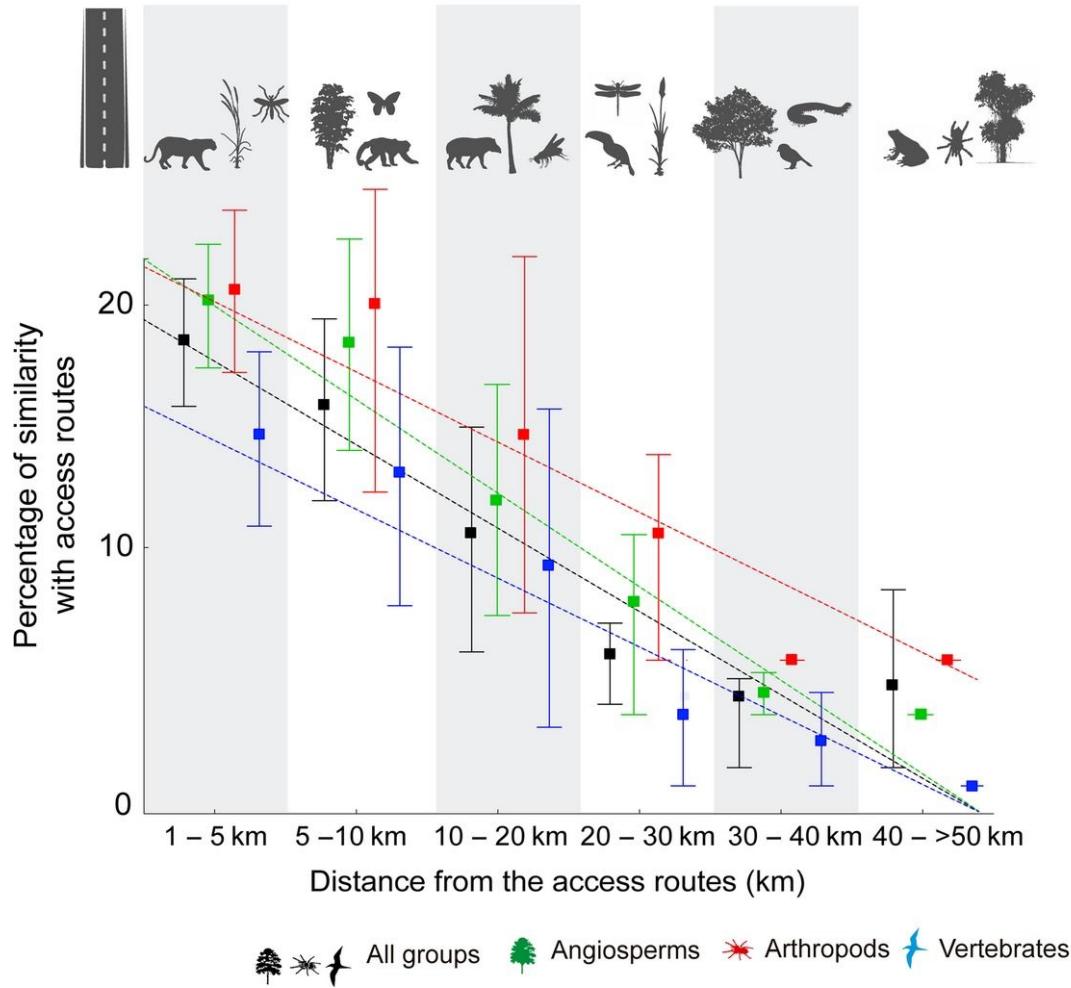


# Script - Ocorrências

02\_script\_occ.R



# Ocorrências - Filtros espaciais

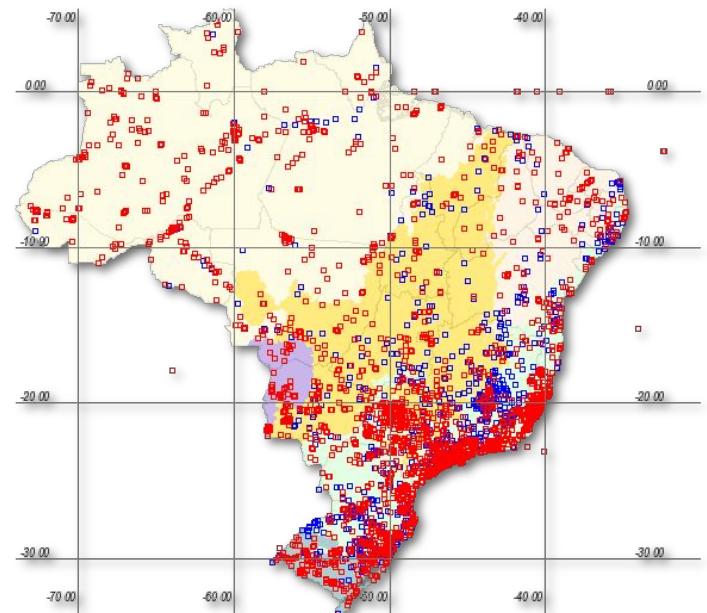


Diversity and Distributions, (Diversity Distrib.) (2016) 22, 1232–1244

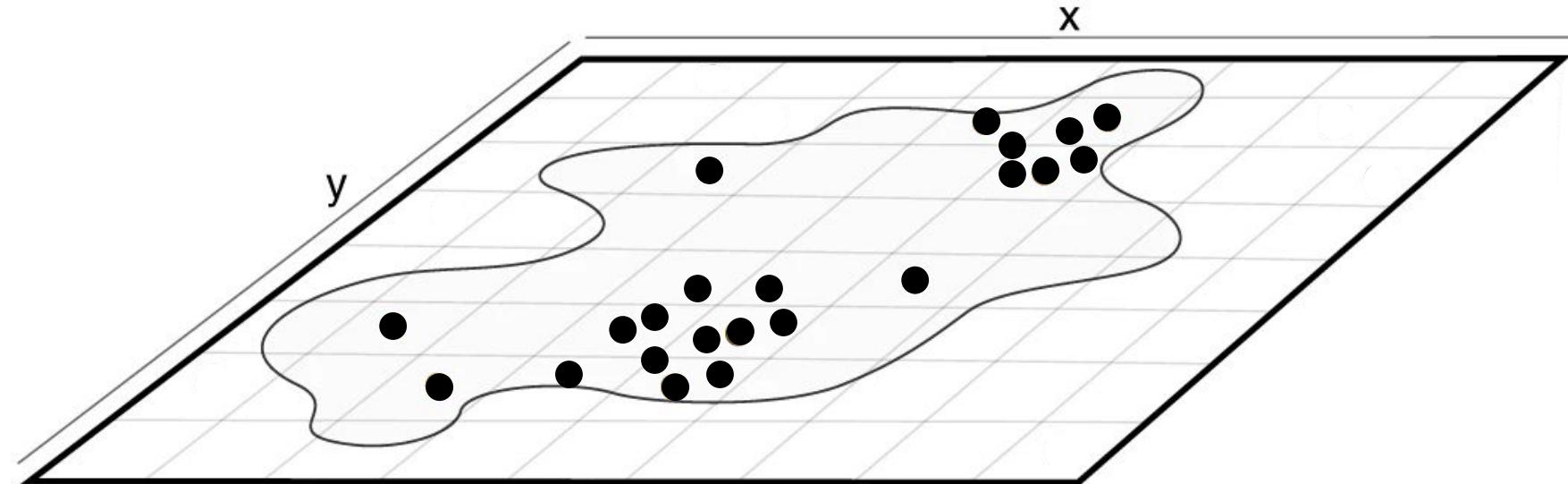


## The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity

Ubirajara Oliveira<sup>1,2\*</sup>, Adriano Pereira Paglia<sup>3</sup>, Antonio D. Brescovit<sup>4</sup>, Claudio J. B. de Carvalho<sup>5</sup>, Daniel Paiva Silva<sup>6</sup>, Daniella T. Rezende<sup>7</sup>, Felipe Sá Fortes Leite<sup>8</sup>, João Aguiar Nogueira Batista<sup>9</sup>, João Paulo Peixoto Pena Barbosa<sup>4</sup>, João Renato Stehmann<sup>9</sup>, John S. Ascher<sup>10</sup>, Marcelo Ferreira de Vasconcelos<sup>11,12</sup>, Paulo De Marco Jr<sup>13</sup>, Peter Löwenberg-Neto<sup>14</sup>, Priscila Guimarães Dias<sup>15</sup>, Viviane Gian Luppi Ferro<sup>15</sup> and Adalberto J. Santos<sup>2</sup>

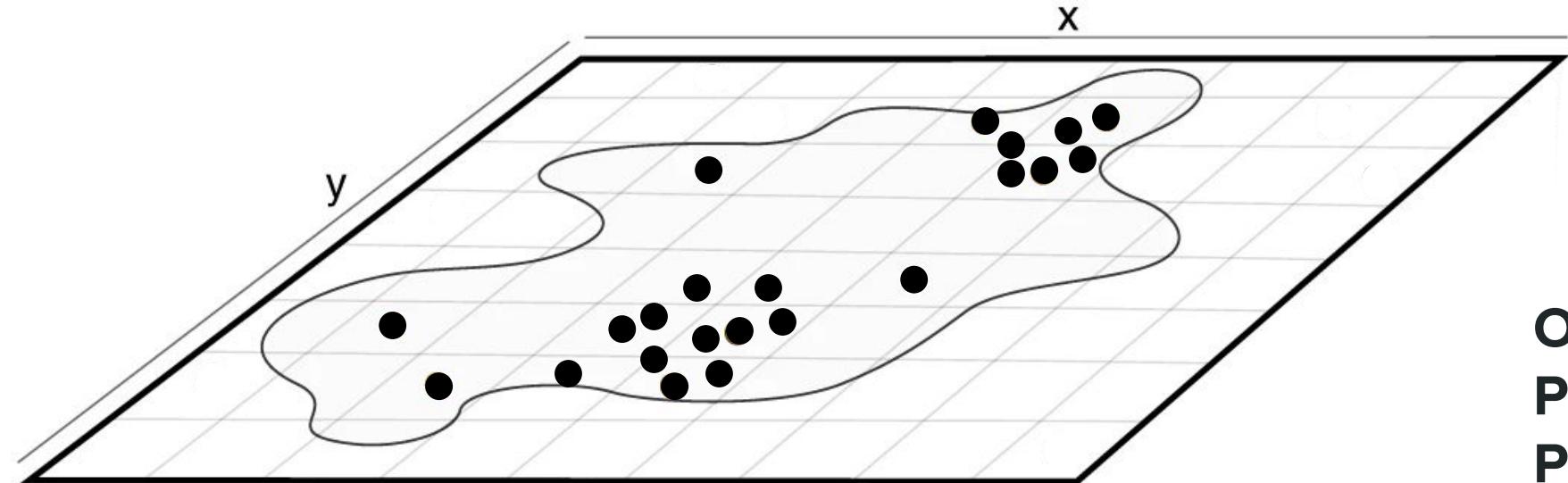


# Ocorrências - Filtros espaciais - OPPC



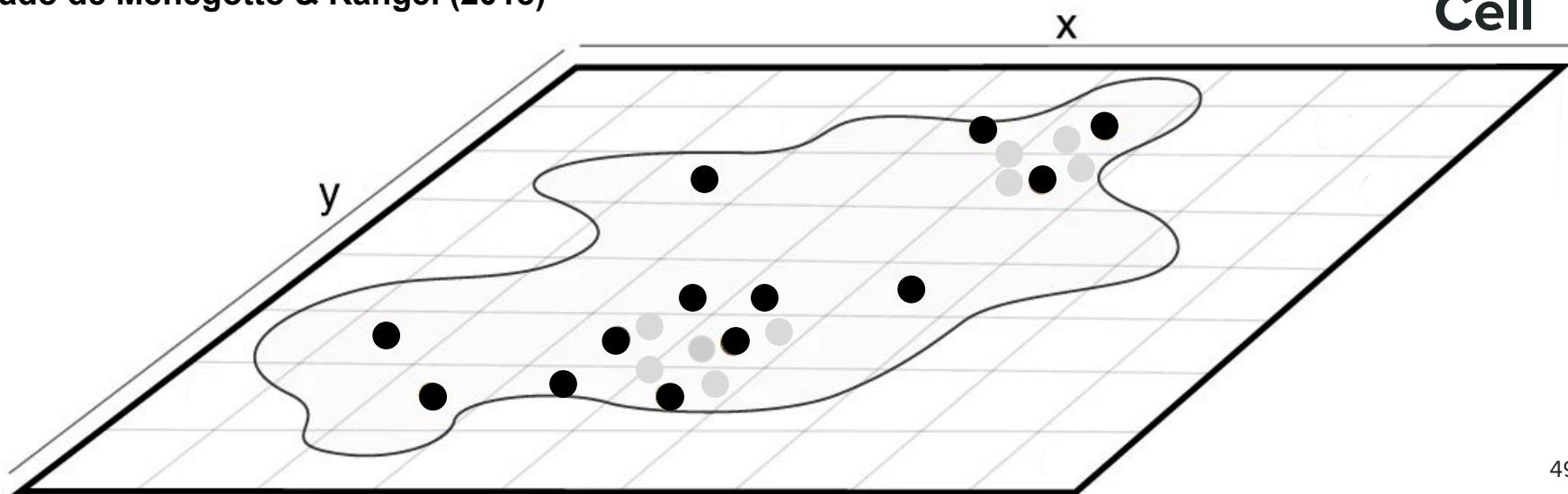
Adaptado de Menegotto & Rangel (2018)

# Ocorrências - Filtros espaciais - OPPC



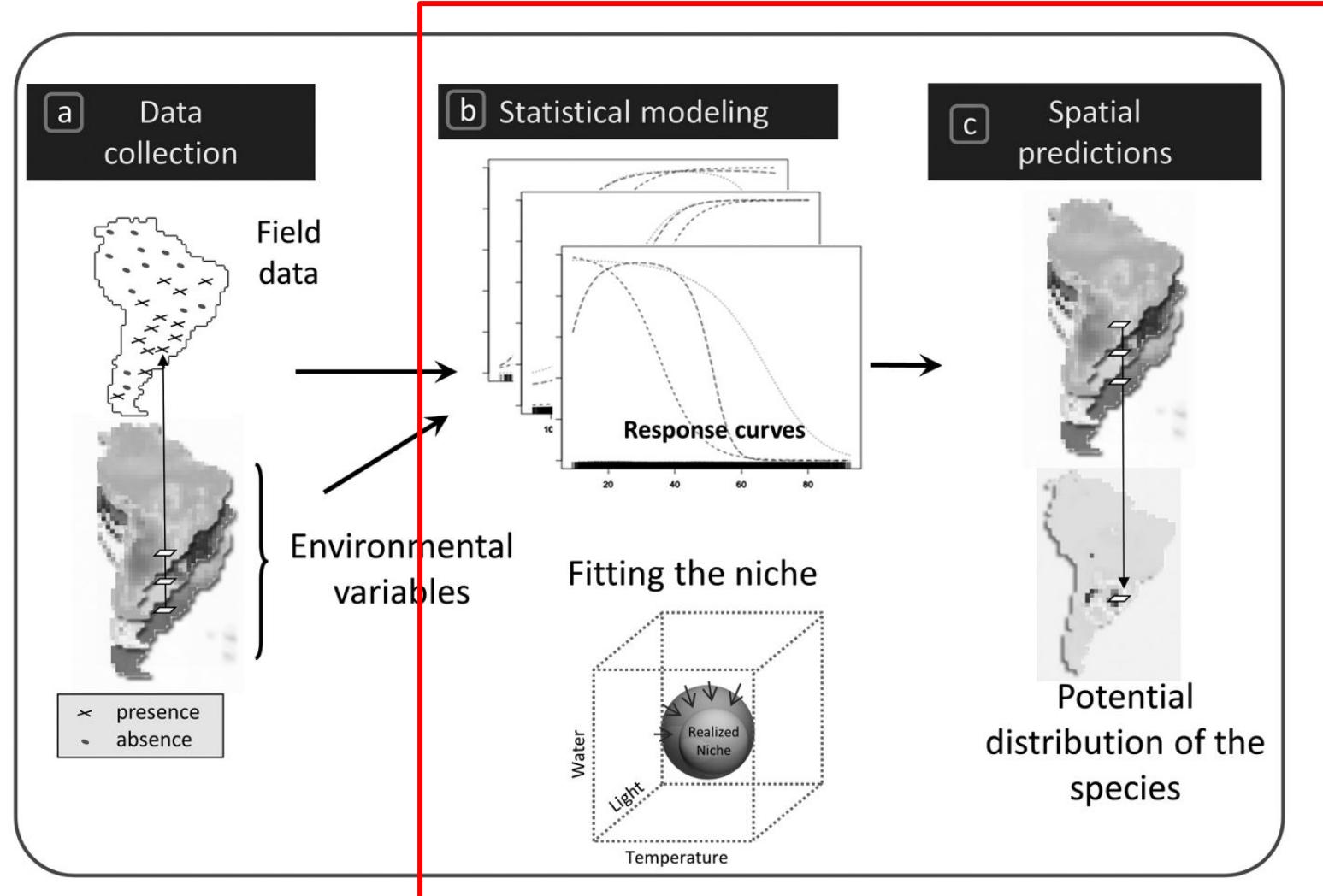
Adaptado de Menegotto & Rangel (2018)

One  
Point  
Per  
Cell



# Intervalo

# Ajuste e predição dos MDEs



Guisan et al. (2017)



# Algoritmos dos MDEs

Lima-Ribeiro &  
Diniz-Filho (2013)

Apenas presença

Bioclim  
Aquário  
Dist. Euclidiana  
Dist. Mahalanobis  
Domain (dist. Gower)  
ENFA (ecological niche factor analysis)

Presença/Background

GARP (genetic algorithm for rule-set production)  
Maxent (maximum entropy)  
SVM (support vector machine)

Aprendizado de Máquina  
(machine learning)  
“cofre”

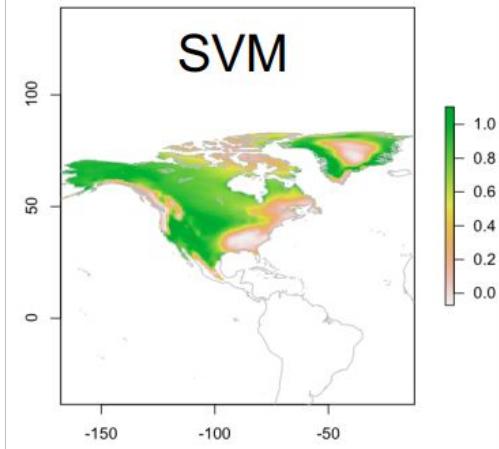
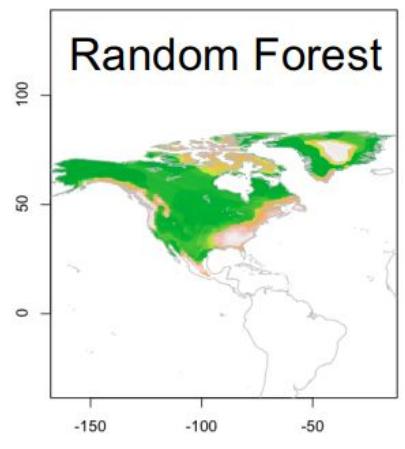
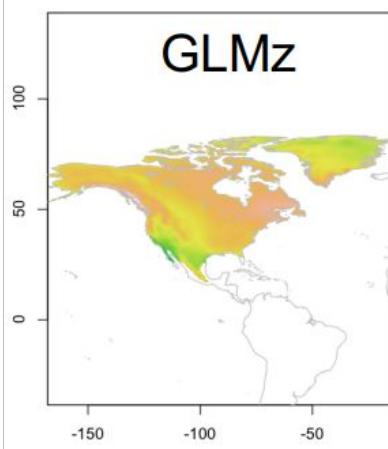
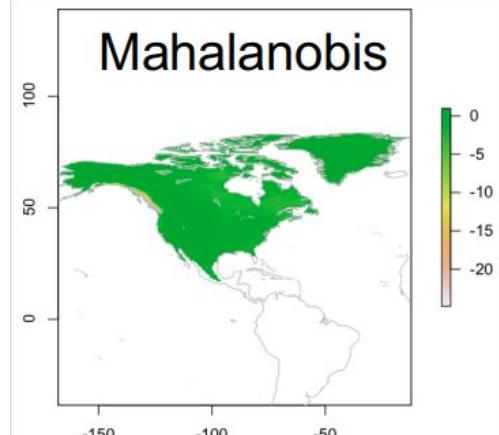
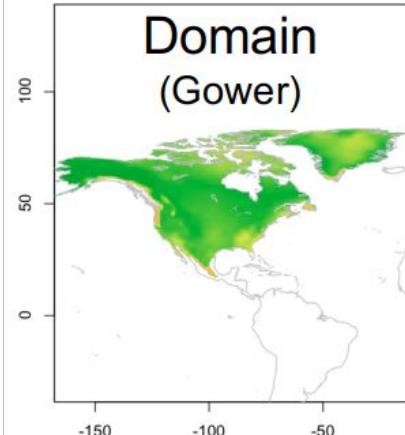
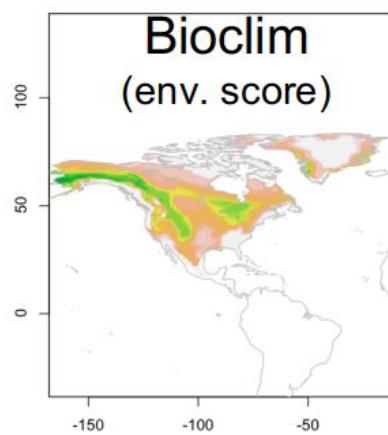
Presença/Ausência

Estatístico (“turbina”)  
GLMZ (generalized linear model)  
GAM (generalized additive model)  
FDA (flexible discriminant analysis)  
MARS (multivariate adaptive reg. splines)

BRT (boosted regression trees)  
→ GBM (gradient boosting machine)  
CART (classification and regression trees)  
RDNFOR (random forest)  
NNET (neural networks)  
→ ANN (artificial neural networks)

# Algoritmos dos MDEs

Qual usar?



# Algoritmos dos MDEs

## Consenso (*Ensemble*)



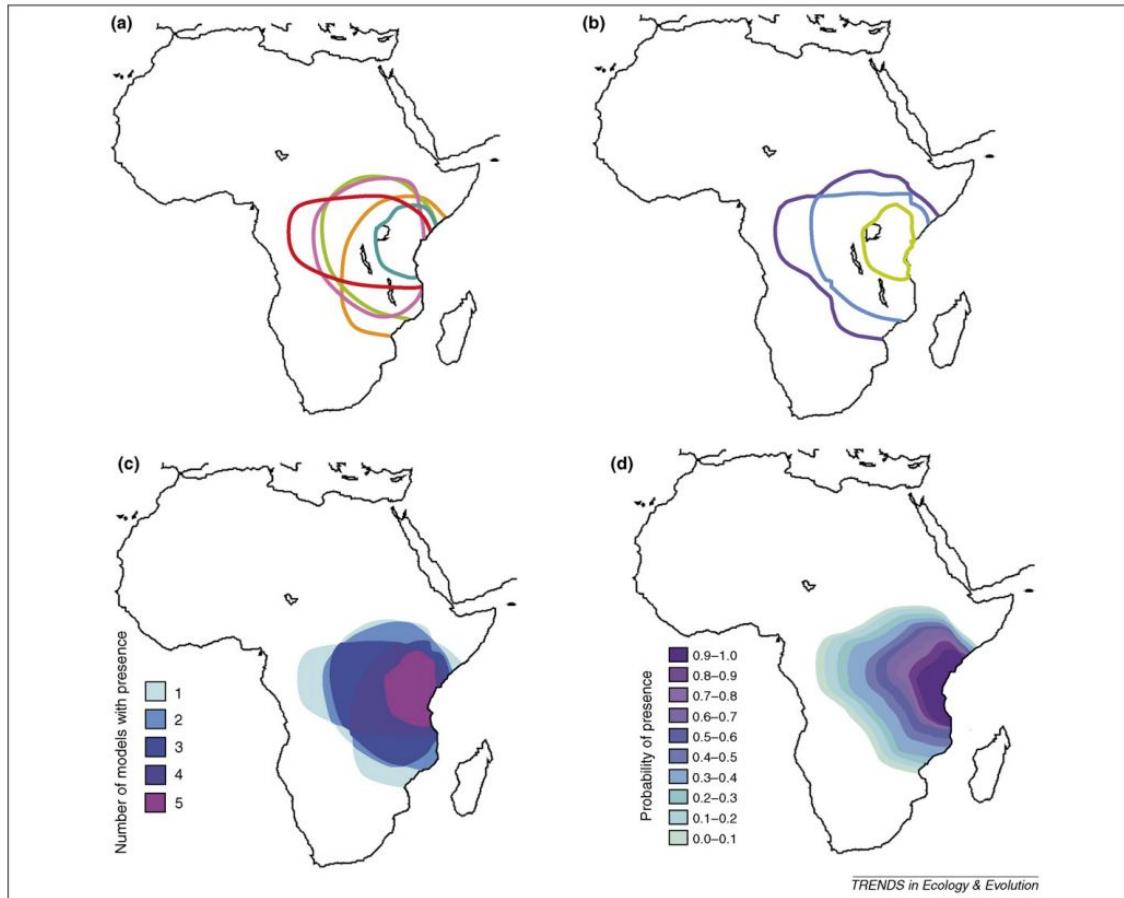
Review

TRENDS in Ecology and Evolution Vol.22 No.1

Full text provided by www.sciencedirect.com  
ScienceDirect

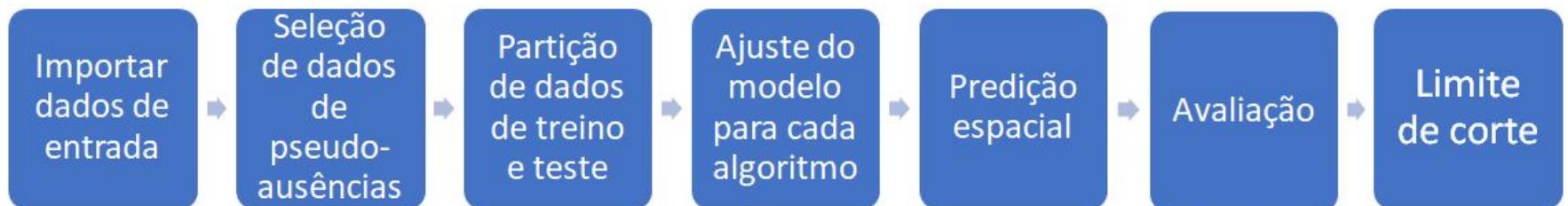
## Ensemble forecasting of species distributions

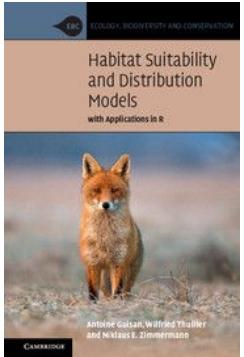
Miguel B. Araújo<sup>1</sup> and Mark New<sup>2</sup>



# Algoritmos e avaliação dos MDEs

03\_script\_sdm\_unique.R

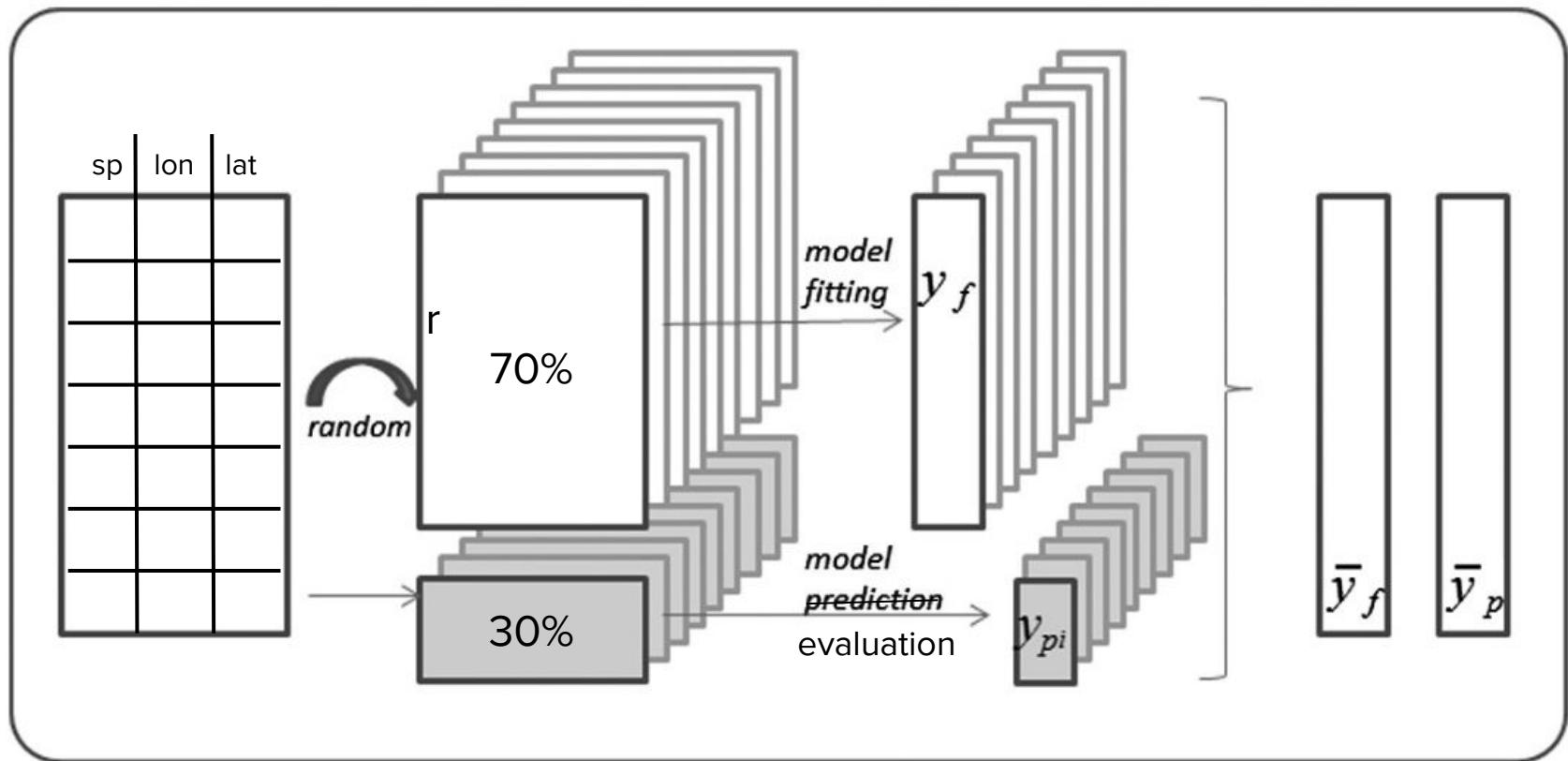




# Avaliação dos MDEs

## Ocorrências

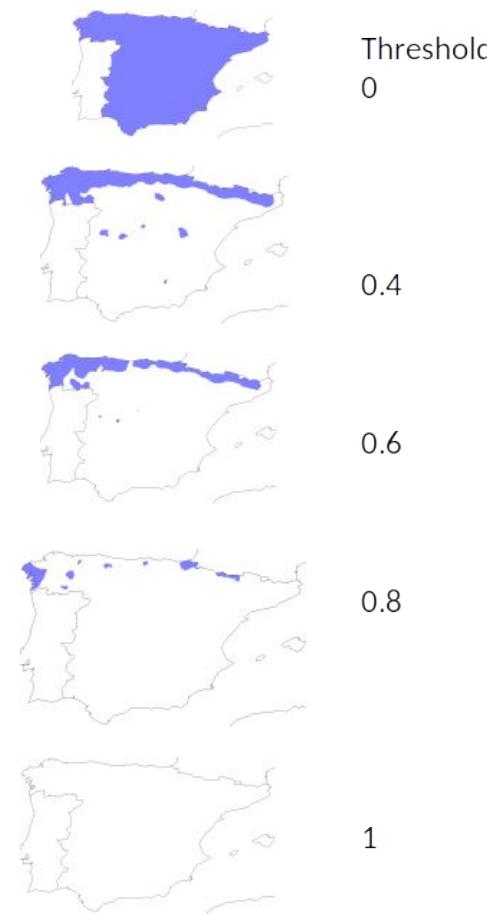
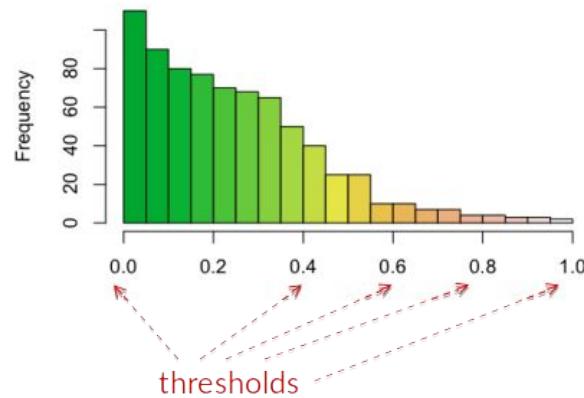
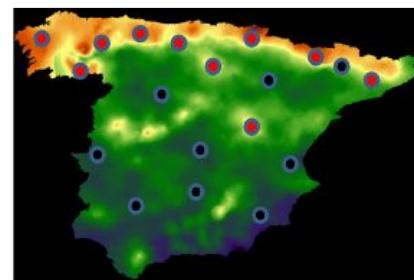
Guisan et al. (2017)





# Avaliação dos MDEs

Limite de corte (*Threshold*)

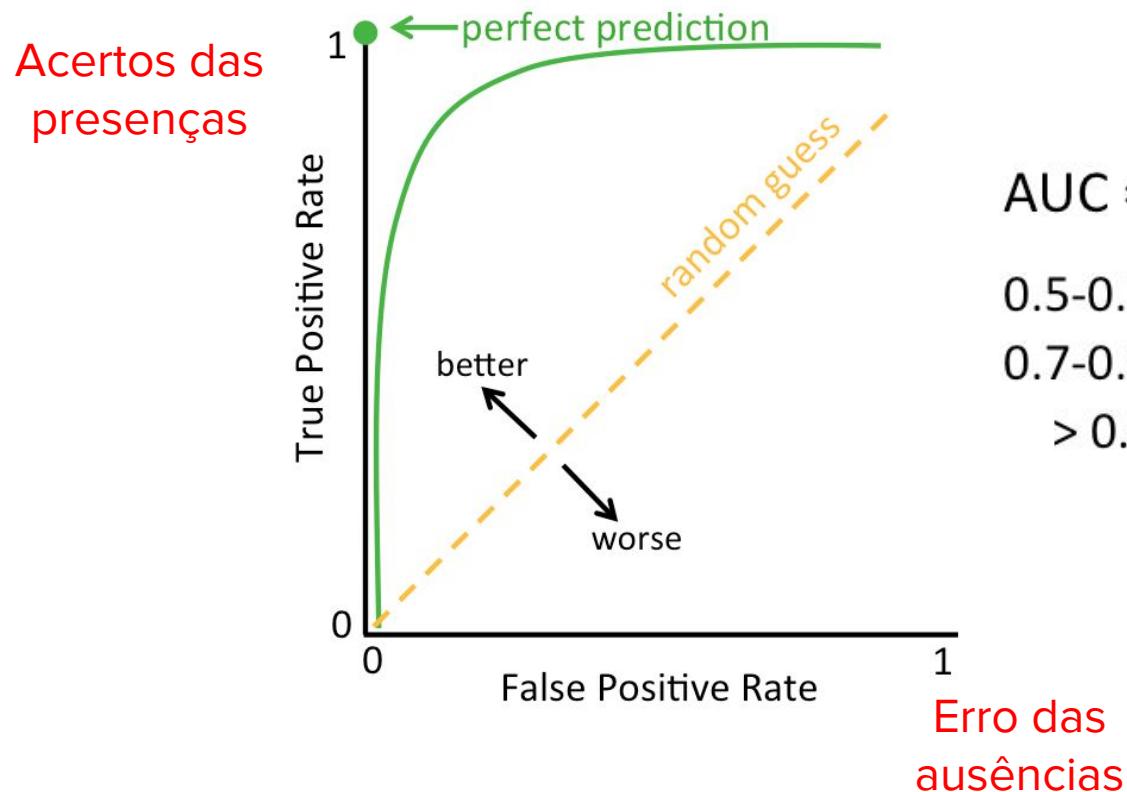


Lima-Ribeiro &  
Diniz-Filho (2013)

# Avaliação dos MDEs

## Curva ROC e AUC

Relative Operating Characteristic (ROC)



AUC = area under the curve

0.5-0.7 = poor model performance

0.7-0.9 = moderate

> 0.9 = excellent



# Avaliação dos MDEs

TSS (True skill statistic)

Lima-Ribeiro &  
Diniz-Filho (2013)

Número de sucesos menos o número de sucesos aleatórios

Valores próximos a 0 modelos não diferentes do aleatórios

Depende de um valor de corte (*threshold*)

$$\text{TSS} = \text{sensitividade} + \text{especificidade} - 1$$

# Algoritmos dos MDEs

## Consenso (*Ensemble*)



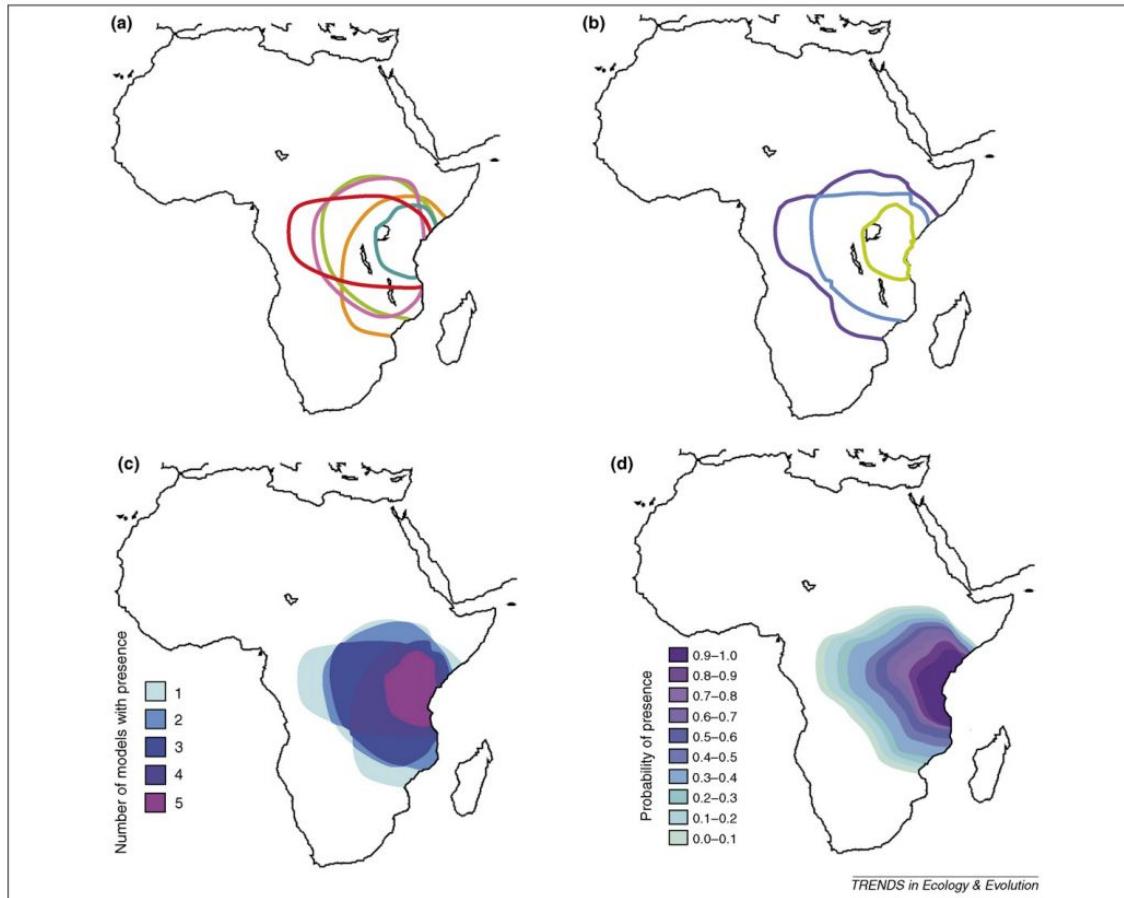
Review

TRENDS in Ecology and Evolution Vol.22 No.1

Full text provided by www.sciencedirect.com  
ScienceDirect

## Ensemble forecasting of species distributions

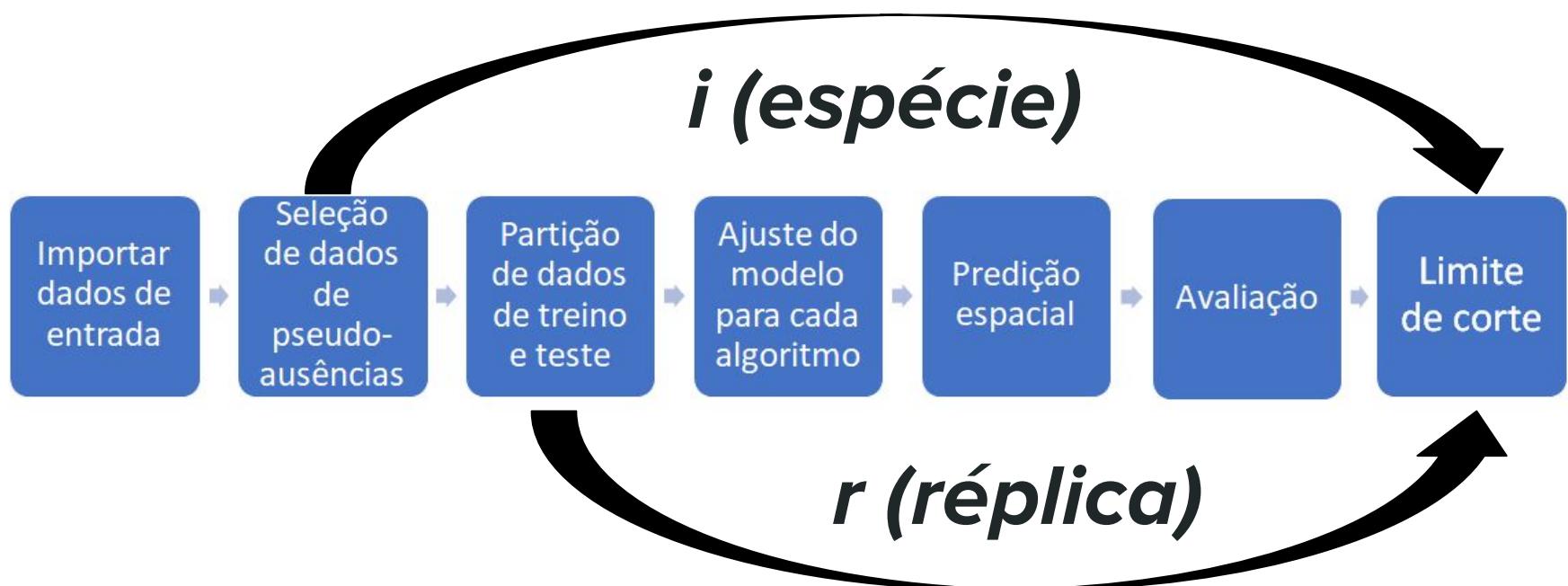
Miguel B. Araújo<sup>1</sup> and Mark New<sup>2</sup>



# Automatização

04\_script\_sdm\_multiple.R

05\_script\_summary\_eval.R

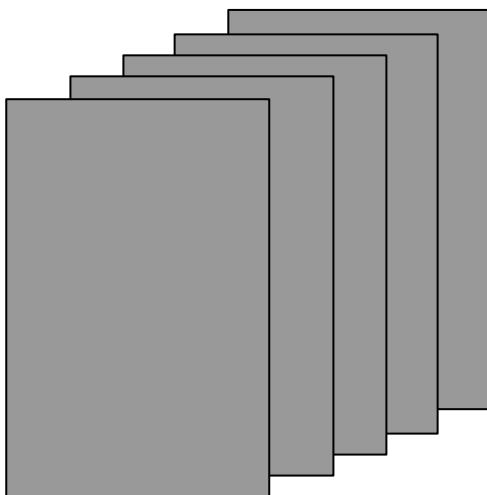


# Consenso (ensembles)

06\_script\_ens\_wei\_ave.R

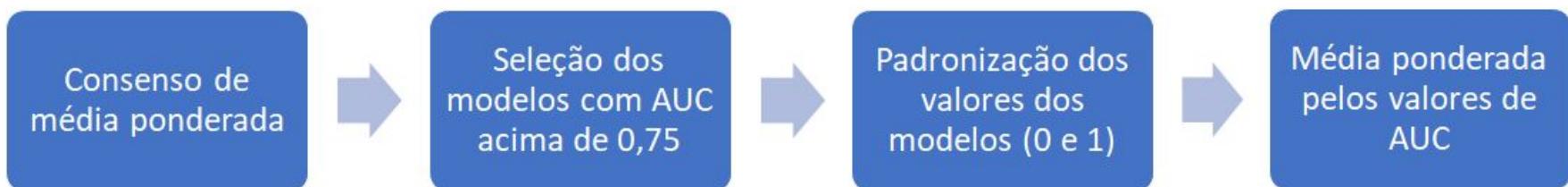


Alg x Rep = 60

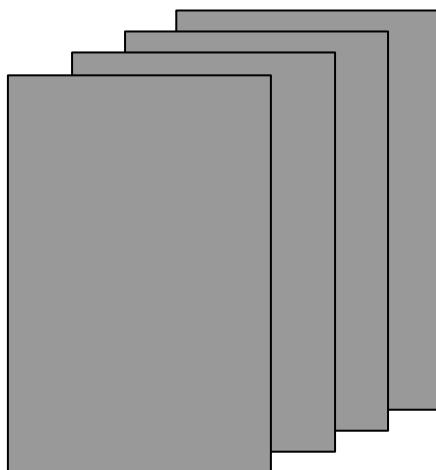


# Consenso (ensembles)

06\_script\_ens\_wei\_ave.R

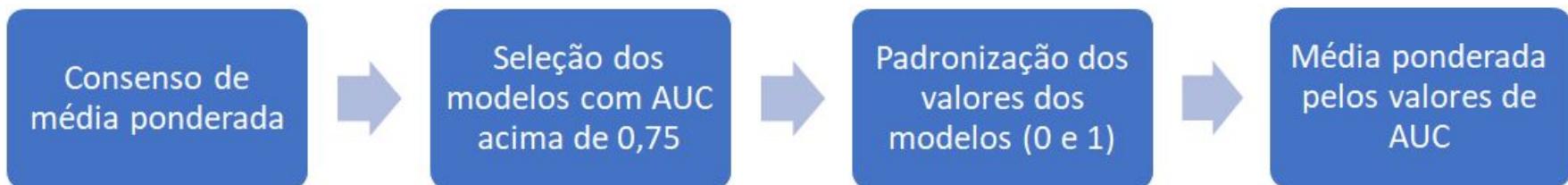


$$r = 55 \mid \text{AUC} > 0.75$$



# Consenso (ensembles)

06\_script\_ens\_wei\_ave.R



**Valores**

r1	r2	r3	....	r55
0.26	0.01	0.34	0.17	0.15
0.83	0.92	1.05	0.76	0.86
0.23	0.12	0.25	0.35	0.54
0.03	0.58	0.48	0.42	0.32
0.01	0.08	0.39	0.23	0.44
0.01	0.02	0.75	0.91	0.11

# Consenso (ensembles)

06\_script\_ens\_wei\_ave.R

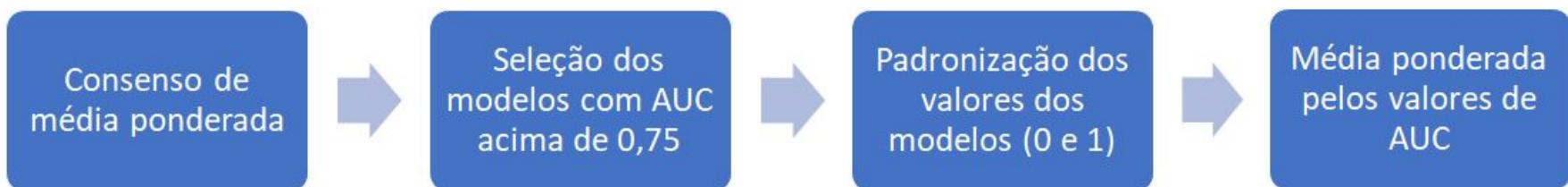


## Padronizar (0 - 1)

r1	r2	r3	....	r55
0.2	0.0	0.3	0.1	0.1
0.8	0.9	1.0	0.7	0.8
0.2	0.1	0.2	0.3	0.5
0.2	0.5	0.4	0.4	0.3
0.1	0.8	0.3	0.2	0.4
0.1	0.2	0.7	0.9	0.1

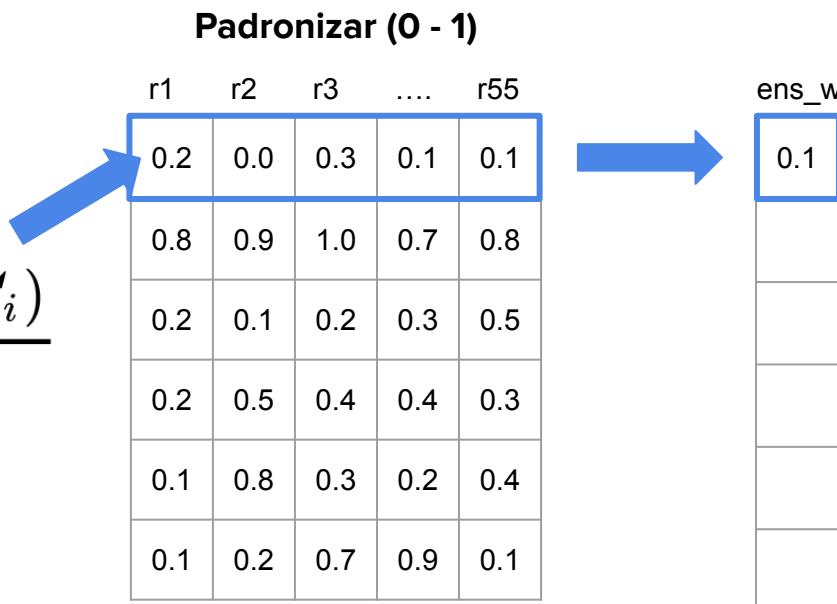
# Consenso (ensembles)

06\_script\_ens\_wei\_ave.R



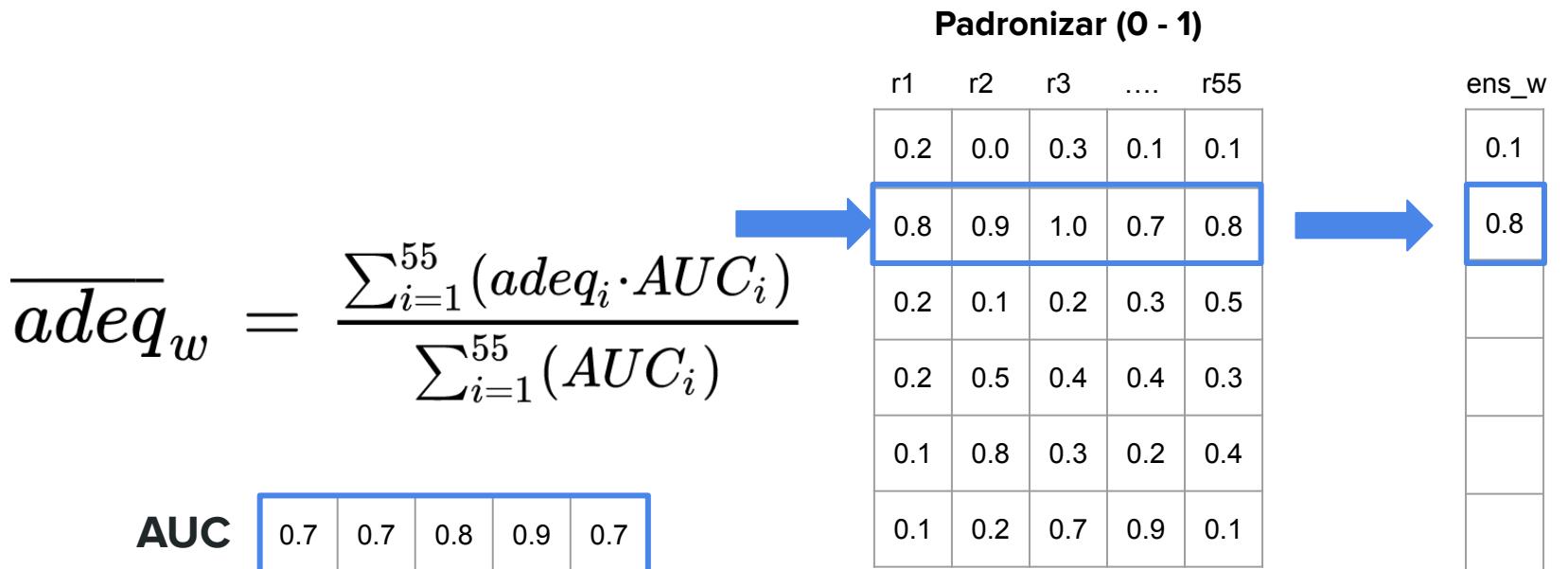
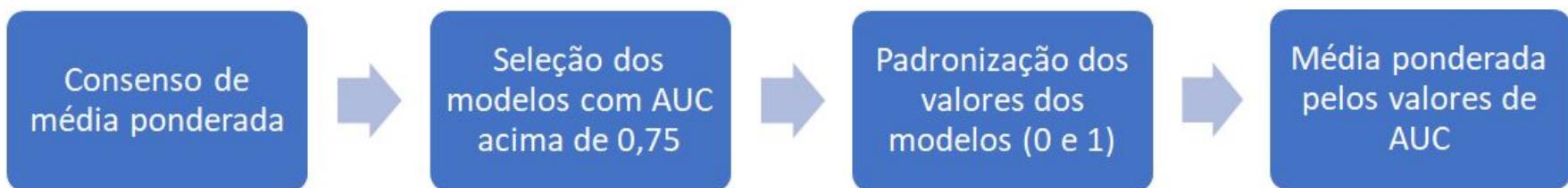
$$\overline{adeq}_w = \frac{\sum_{i=1}^{55} (adeq_i \cdot AUC_i)}{\sum_{i=1}^{55} (AUC_i)}$$

AUC	0.7	0.7	0.8	0.9	0.7
-----	-----	-----	-----	-----	-----



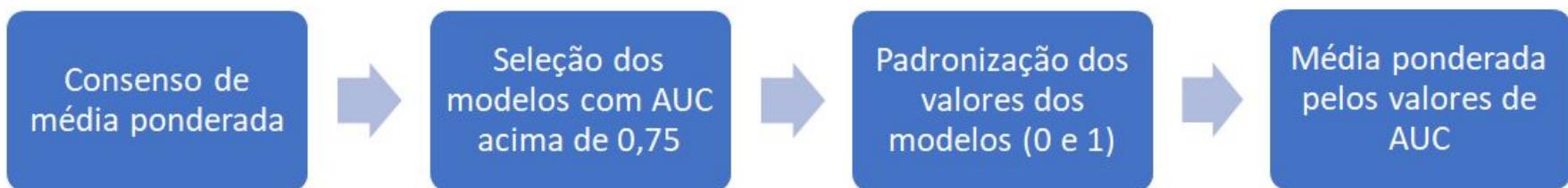
# Consenso (ensembles)

06\_script\_ens\_wei\_ave.R



# Consenso (ensembles)

06\_script\_ens\_wei\_ave.R



$$\overline{adeq}_w = \frac{\sum_{i=1}^{55} (adeq_i \cdot AUC_i)}{\sum_{i=1}^{55} (AUC_i)}$$

AUC	0.7	0.7	0.8	0.9	0.7
-----	-----	-----	-----	-----	-----

Padronizar (0 - 1)

r1	r2	r3	....	r55	ens_w
0.2	0.0	0.3	0.1	0.1	0.1
0.8	0.9	1.0	0.7	0.8	0.8
0.2	0.1	0.2	0.3	0.5	0.5
0.2	0.5	0.4	0.4	0.3	0.3
0.1	0.8	0.3	0.2	0.4	0.4
0.1	0.2	0.7	0.9	0.1	0.1

# Limite de corte (threshold)

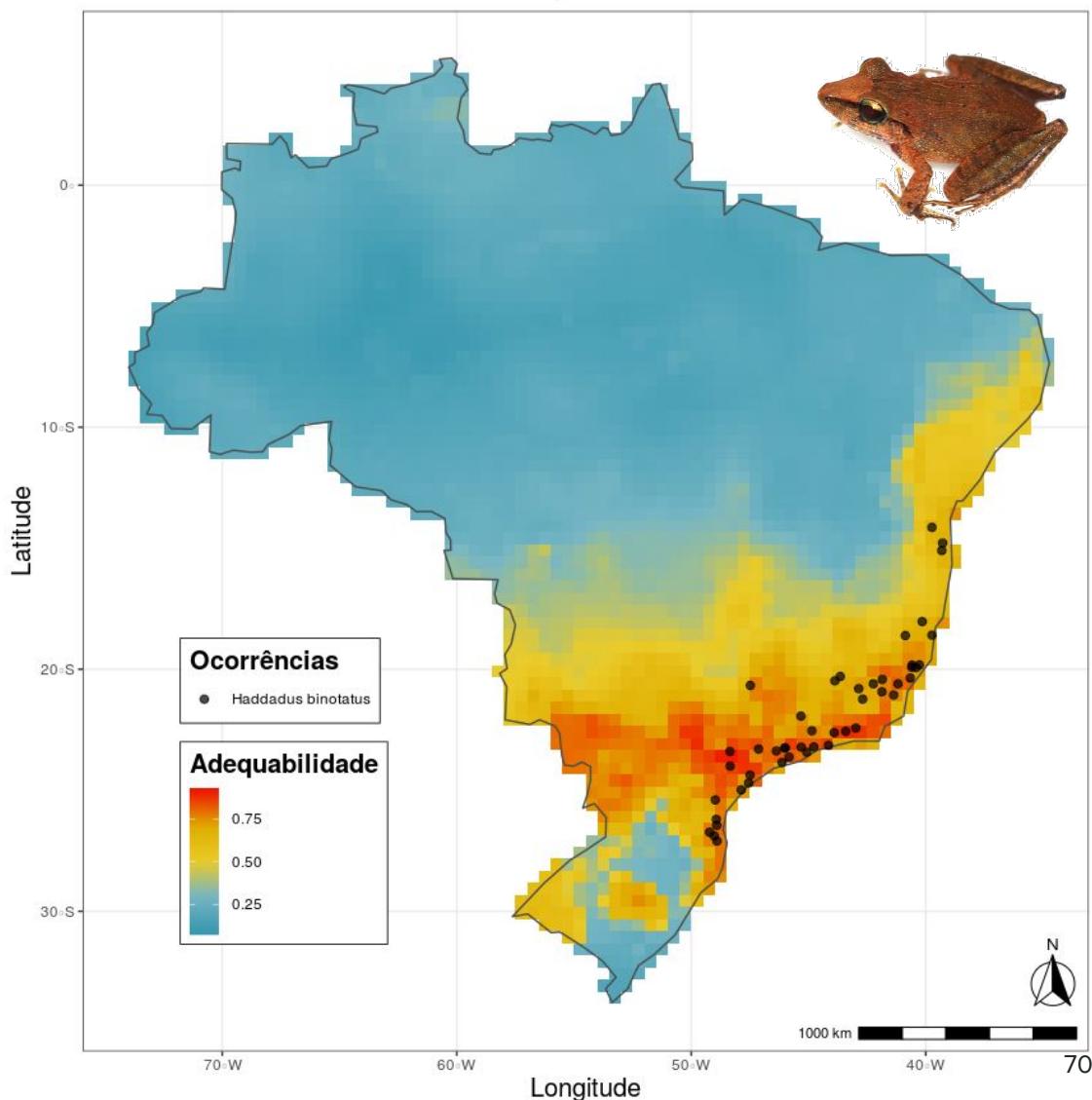
07\_script\_cut\_thr.R



# Mapas

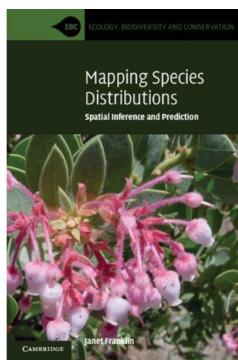
08\_script\_maps.R

*Haddadus binotatus* – Consenso média ponderada

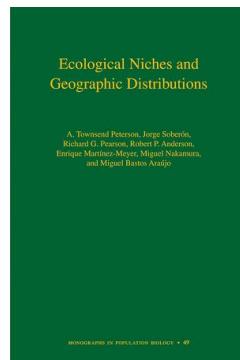


# Onde procurar mais informações?

## Livros



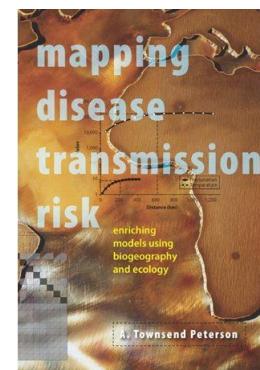
Franklin (2009)



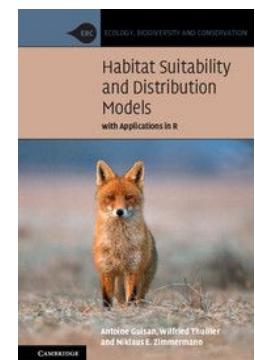
Peterson et al. (2011)



Lima-Ribeiro & Diniz-Filho (2013)



Peterson (2014)



Guisan et al. (2017)

# Fechamento

# Contatos

Maurício Vancine

[mauricio.vancine@gmail.com](mailto:mauricio.vancine@gmail.com)

João Giovanelli

[jgiovanelli@gmail.com](mailto:jgiovanelli@gmail.com)