

# modleR: a modular **workflow** for ecological niche modeling in R

Ecological Niche Modeling 2020

Andrea Sánchez-Tapia & Sara Mortara

Scientific Computation Laboratory  
Rio de Janeiro Botanical Garden

May 2020



about

about

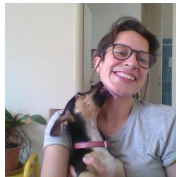


Marinez F. de Siqueira

about



Marinez F. de Siqueira

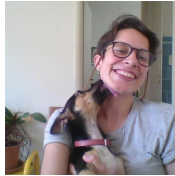


Sara Mortara

about



Marinez F. de Siqueira



Sara Mortara

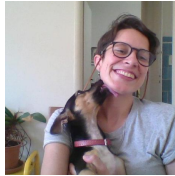


Andrea Sánchez-Tapia

about



Marinez F. de Siqueira



Sara Mortara

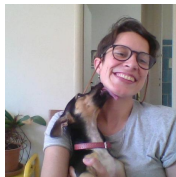


Andrea Sánchez-Tapia

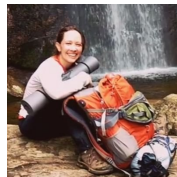
## about



Marinez F. de Siqueira



Sara Mortara



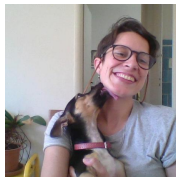
Andrea Sánchez-Tapia

- Scientific Computation Lab: Biodiversity informatics, ENM/SDM, **open science**, reproducibility

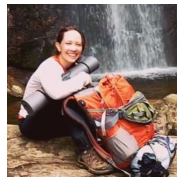
## about



Marinez F. de Siqueira



Sara Mortara



Andrea Sánchez-Tapia

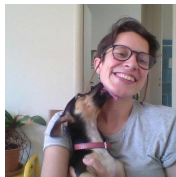
- Scientific Computation Lab: Biodiversity informatics, ENM/SDM, **open science**, reproducibility



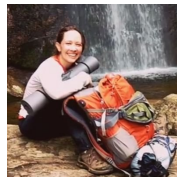
## about



Marinez F. de Siqueira



Sara Mortara



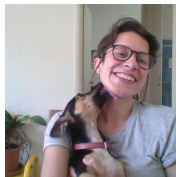
Andrea Sánchez-Tapia

- ▶ Scientific Computation Lab: Biodiversity informatics, ENM/SDM, **open science**, reproducibility
- ▶ **Scientific workflows** based in R for data downloading and cleaning, taxonomic checking

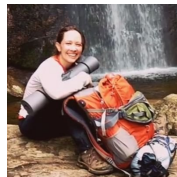
## about



Marinez F. de Siqueira



Sara Mortara



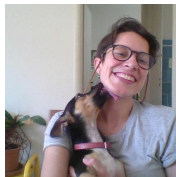
Andrea Sánchez-Tapia

- ▶ Scientific Computation Lab: Biodiversity informatics, ENM/SDM, **open science**, reproducibility
- ▶ **Scientific workflows** based in R for data downloading and cleaning, taxonomic checking

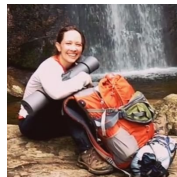
## about



Marinez F. de Siqueira



Sara Mortara



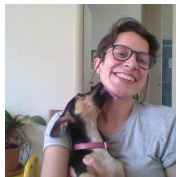
Andrea Sánchez-Tapia

- ▶ Scientific Computation Lab: Biodiversity informatics, ENM/SDM, **open science**, reproducibility
- ▶ **Scientific workflows** based in R for data downloading and cleaning, taxonomic checking
- ▶ Support for data-intensive research projects (e.g., CNCFlora - IUCN authority)

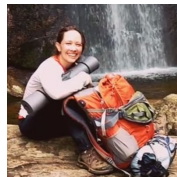
## about



Marinez F. de Siqueira



Sara Mortara



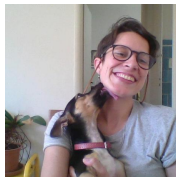
Andrea Sánchez-Tapia

- ▶ Scientific Computation Lab: Biodiversity informatics, ENM/SDM, **open science**, reproducibility
- ▶ **Scientific workflows** based in R for data downloading and cleaning, taxonomic checking
- ▶ Support for data-intensive research projects (e.g., CNCFlora - IUCN authority)

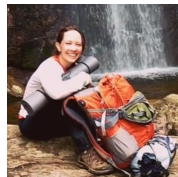
## about



Marinez F. de Siqueira



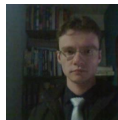
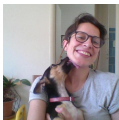
Sara Mortara



Andrea Sánchez-Tapia

- ▶ Scientific Computation Lab: Biodiversity informatics, ENM/SDM, **open science**, reproducibility
- ▶ **Scientific workflows** based in R for data downloading and cleaning, taxonomic checking
- ▶ Support for data-intensive research projects (e.g., CNCFlora - IUCN authority)
- ▶ ENM2020 Week 3: Sara Mortara - Data cleaning module (link)

# modleR developers



Andrea Sánchez-Tapia

Sara R. Mortara

Diogo S.B. Rocha

Maria Luisa Mondelli

Guilherme Gall

Felipe Barros

# overview

Scientific workflows and why building an R package

Step 1. `setup_sdmdata`  
pseudoabsence sampling

Step 2. `do_(m)any`

Step 3. `final_model`

Step 4. `ensemble_model`

Installing and using `modleR`

# **Scientific workflows and why building an R package**



why building an ENM package?

## why building an ENM package?

- ▶ We started with a **project-specific** set of scripts to execute ENM for species in the Brazilian Atlantic Forest

## why building an ENM package?

- ▶ We started with a **project-specific** set of scripts to execute ENM for species in the Brazilian Atlantic Forest
- ▶ Other projects: similar structure but flexibility needed depending on the **research question**

## why building an ENM package?

- ▶ We started with a **project-specific** set of scripts to execute ENM for species in the Brazilian Atlantic Forest
- ▶ Other projects: similar structure but flexibility needed depending on the **research question**
- ▶ Many options within R
  - GIS with **raster**, **sp**, **maps**, **rgdal**, **sf**
  - Established packages such as **dismo** (Hijmans et al 2017), **BIOMOD2** (Thuiller et al 2007)
  - Other packages **ENMeval** (Muscarella et al 2014), **sdm** (Naimi & Araújo 2016), **spThin** (Aiello-Lammens et al 2015), **zoon** (Golding et al. 2018) **wallace** (Kass et al 2018), **kuenm** (Cobos et al. 2019), **occCite** (Lowens 2020)

We needed a **workflow** rather than standalone R packages

# why building an ENM package?

folder structure and portability

# why building an ENM package?

folder structure and portability

# why building an ENM package?

folder structure and portability



# why building an ENM package?

folder structure and portability

- ▶ A single working directory per project

# why building an ENM package?

folder structure and portability

- ▶ A single working directory per project
- ▶ Different steps: different subfolders

# why building an ENM package?

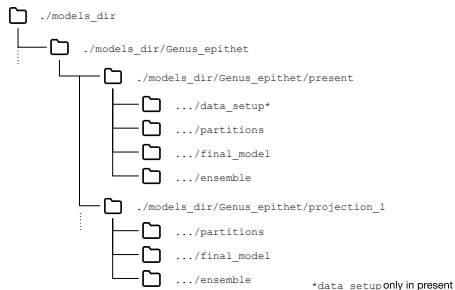
folder structure and portability

- ▶ A single working directory per project
- ▶ Different steps: different subfolders
- ▶ A consistent subfolder structure

# why building an ENM package?

## folder structure and portability

- ▶ A single working directory per project
- ▶ Different steps: different subfolders
- ▶ A consistent subfolder structure
- ▶ Relative rather than absolute paths and no `setwd()`



# why building an ENM package?

modularity

# why building an ENM package?

modularity

- ▶ Each step saves its output

# why building an ENM package?

modularity

- ▶ Each step saves its output
- ▶ The next step reads the previous output

# why building an ENM package?

modularity

- ▶ Each step saves its output
- ▶ The next step reads the previous output
- ▶ Using HD space rather than RAM



# why building an ENM package?

modularity

- ▶ Each step saves its output
- ▶ The next step reads the previous output
- ▶ Using HD space rather than RAM
- ▶ The user may enter and exit the workflow at any step

# why building an ENM package?

## modularity

- ▶ Each step saves its output
- ▶ The next step reads the previous output
- ▶ Using HD space rather than RAM
- ▶ The user may enter and exit the workflow at any step
- ▶ Parallelization and use in high performance/high throughput computational frameworks (HPC/HTC)

# why building an ENM package?

reproducibility

PERSPECTIVE

<https://doi.org/10.1038/s41559-019-0972-5>

nature  
ecology & evolution

OPEN

## A checklist for maximizing reproducibility of ecological niche models

Xiao Feng<sup>1,2,8\*</sup>, Daniel S. Park<sup>3,8</sup>, Cassandra Walker<sup>4</sup>, A. Townsend Peterson<sup>5</sup>, Cory Merow<sup>6</sup> and Monica Papes<sup>7</sup>

Reporting specific modelling methods and metadata is essential to the reproducibility of ecological studies, yet guidelines rarely exist regarding what information should be noted. Here, we address this issue for ecological niche modelling or species distribution modelling, a rapidly developing toolset in ecology used across many aspects of biodiversity science. Our quantitative review of the recent literature reveals a general lack of sufficient information to fully reproduce the work. Over two-thirds of the examined studies neglected to report the version or access date of the underlying data, and only half reported model parameters. To address this problem, we propose adopting a checklist to guide studies in reporting at least the minimum information necessary for ecological niche modelling reproducibility, offering a straightforward way to balance efficiency and accuracy. We encourage the ecological niche modelling community, as well as journal reviewers and editors, to utilize and further develop this framework to facilitate and improve the reproducibility of future work. The proposed checklist framework is generalizable to other areas of ecology, especially those utilizing biodiversity data, environmental data and statistical modelling, and could also be adopted by a broader array of disciplines.

why building an ENM package?

Thorough metadata recording:

# why building an ENM package?

Thorough metadata recording:

- ▶ parametrization options

# why building an ENM package?

Thorough metadata recording:

- ▶ parametrization options
- ▶ session information

# why building an ENM package?

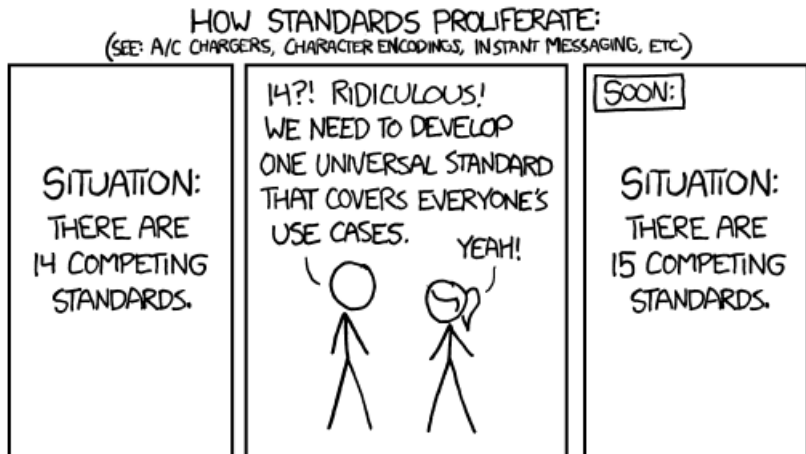
Thorough metadata recording:

- ▶ parametrization options
- ▶ session information
- ▶ packages used and their version

# why building an ENM package?

interoperability

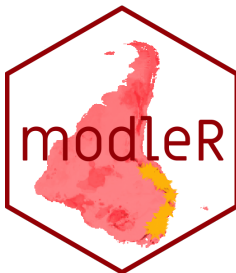
We did not create new classes or methods: communication with other packages in the R environment





# modleR

A **workflow** developed to automatize some of the common steps in ecological niche modeling



<https://model-r.github.io/modleR/>

## a **four-step** workflow

1. `setup_sdmdata()`: data setup

## a **four-step** workflow

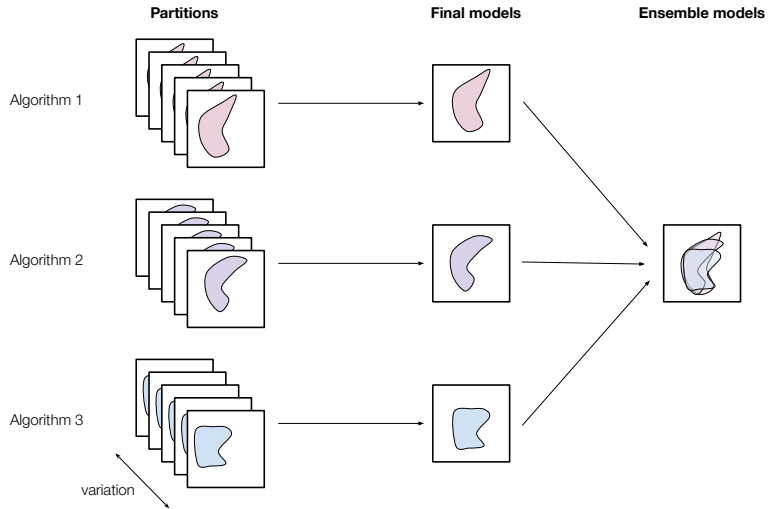
1. `setup_sdmdata()`: data setup
2. `do_any()` and `do_many()`: model fitting, projecting and evaluating

## a **four-step** workflow

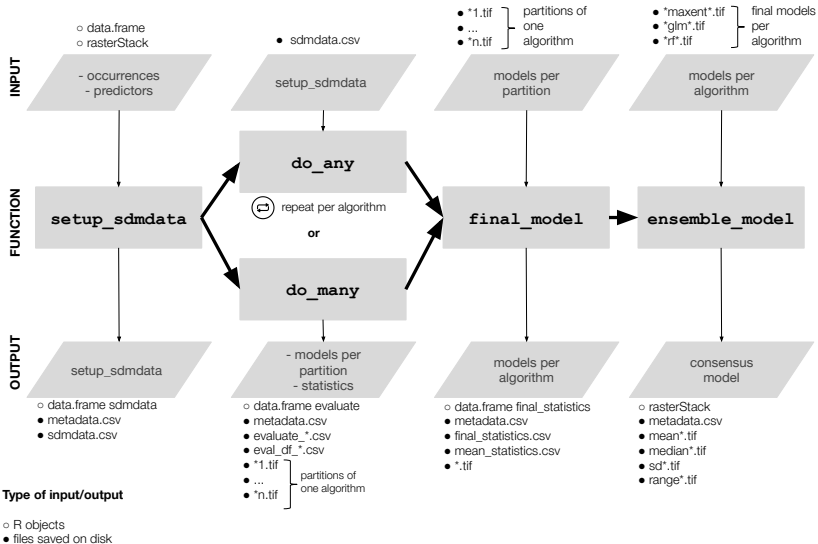
1. `setup_sdmdata()`: data setup
2. `do_any()` and `do_many()`: model fitting, projecting and evaluating
3. `final_model()`: joining partitions

## a **four-step** workflow

1. `setup_sdmdata()`: data setup
2. `do_any()` and `do_many()`: model fitting, projecting and evaluating
3. `final_model()`: joining partitions
4. `ensemble_model()`: algorithm consensus



# a four-step workflow



**Step 1.** `setup_sdmdata`



`setup_sdmdata()`: data preparation

Data preparation and cleaning should be performed previously

## setup\_sdmdata(): data preparation

Data preparation and cleaning should be performed previously

- ▶ Optional data cleaning checks: exact duplicates, NAs and one occurrence per pixel

## setup\_sdmdata(): data preparation

Data preparation and cleaning should be performed previously

- ▶ Optional data cleaning checks: exact duplicates, NAs and one occurrence per pixel
- ▶ Experimental design: bootstrap, cross-validation

## setup\_sdmdata(): data preparation

Data preparation and cleaning should be performed previously

- ▶ Optional data cleaning checks: exact duplicates, NAs and one occurrence per pixel
- ▶ Experimental design: bootstrap, cross-validation
- ▶ Pseudo-absence sampling

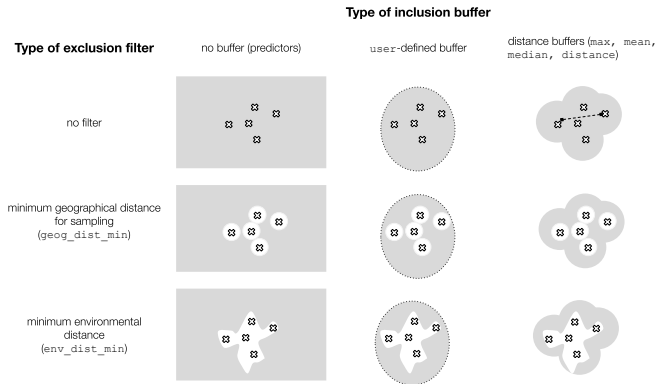
## setup\_sdmdata(): data preparation

Data preparation and cleaning should be performed previously

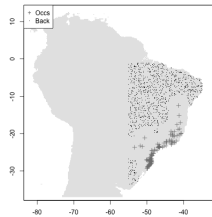
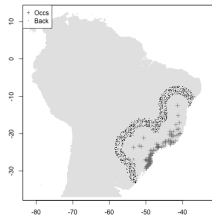
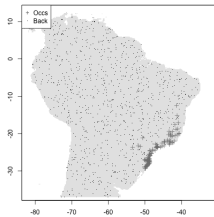
- ▶ Optional data cleaning checks: exact duplicates, NAs and one occurrence per pixel
- ▶ Experimental design: bootstrap, cross-validation
- ▶ Pseudo-absence sampling
- ▶ Control of variable correlation up to a user-defined value (e.g., 0.8)

pseudoabsence sampling

# pseudoabsence sampling options

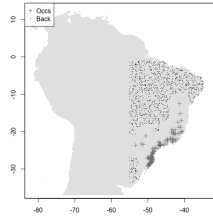
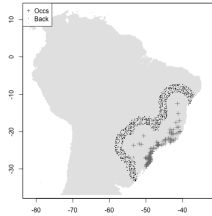
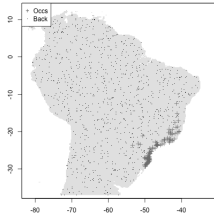


# pseudoabsence sampling



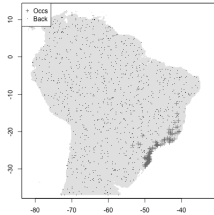


# pseudoabsence sampling

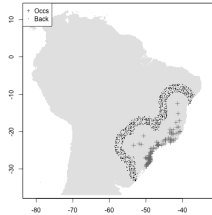


no buffer

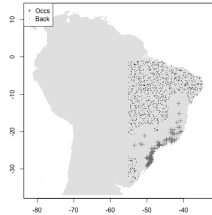
# pseudoabsence sampling



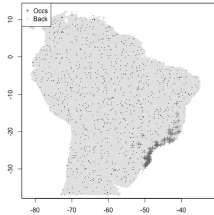
no buffer



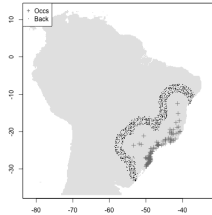
mean distance buffer  
and euclidean  
distance filter



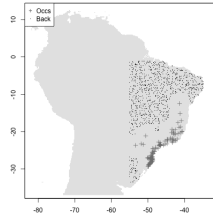
# pseudoabsence sampling



no buffer



mean distance buffer  
and euclidean  
distance filter



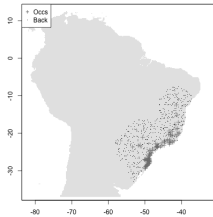
user-defined buffer  
(M) and euclidean  
distance filter

## setup\_sdmdata() output

At the end of data setup:

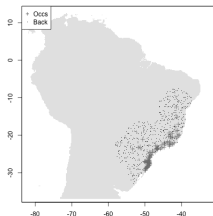
## setup\_sdmdata() output

At the end of data setup:



## setup\_sdmdata() output

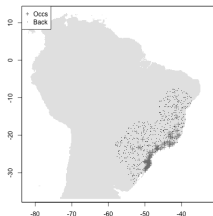
At the end of data setup:



species_name	original.n	final.n	original.n.back	final.n.back	buffer_type	dist_t	
Abarema langsdorfii	104	104	500	500	distance		
group.all	pa	lon	lat	layer.1	layer.2	layer.3	layer.4
3	1	-40.615	-19.921	-8.9611081	6.11421264	1.40630368	-7.7
3	1	-40.729	-20.016	-8.7214928	5.87764334	1.37790286	-7.7
4	1	-41.174	-20.303	-8.2562377	5.83755751	1.26637686	-8.7
4	1	-41.74	-20.493	-7.2750323	3.61264467	0.81564361	-8.7
2	1	-42.482	-20.701	-7.8781131	5.98075427	1.2590602	-9.7

## setup\_sdmdata() output

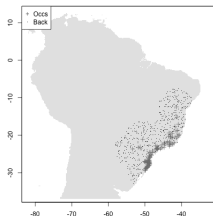
At the end of data setup:



species_name	original.n	final.n	original.n.back	final.n.back	buffer_type	dist_t	
Abarema langsdorfii	104	104	500	500	distance		
group.all	pa	lon	lat	layer.1	layer.2	layer.3	layer.4
3	1	-40.615	-19.921	-8.9611081	6.11421264	1.40630368	-7.7
3	1	-40.729	-20.016	-8.7214928	5.87764334	1.37790286	-7.7
4	1	-41.174	-20.303	-8.2562377	5.83755751	1.26637686	-8.7
4	1	-41.74	-20.493	-7.2750323	3.61264467	0.81564361	-8.7
2	1	-42.482	-20.701	-7.8781131	5.98075427	1.2590602	-9.7

# setup\_sdmdata() output

At the end of data setup:



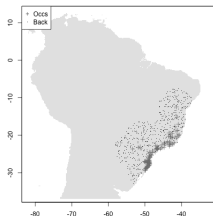
species_name	original.n	final.n	original.n.back	final.n.back	buffer_type	dist_t	
Abarema langsdorfii	104	104	500	500	distance		
group.all	pa	lon	lat	layer.1	layer.2	layer.3	layer.4
3	1	-40.615	-19.921	-8.9611081	6.11421264	1.40630368	-7.7
3	1	-40.729	-20.016	-8.7214928	5.87764334	1.37790286	-7.7
4	1	-41.174	-20.303	-8.2562377	5.83755751	1.26637686	-8.7
4	1	-41.74	-20.493	-7.2750323	3.61264467	0.81564361	-8.7
2	1	-42.482	-20.701	-7.8781131	5.98075427	1.2590602	-9.7

► metadata and session information



## setup\_sdmdata() output

At the end of data setup:



species_name	original.n	final.n	original.n.back	final.n.back	buffer_type	dist_t
Abarema langsdorfii	104	104	500	500	distance	

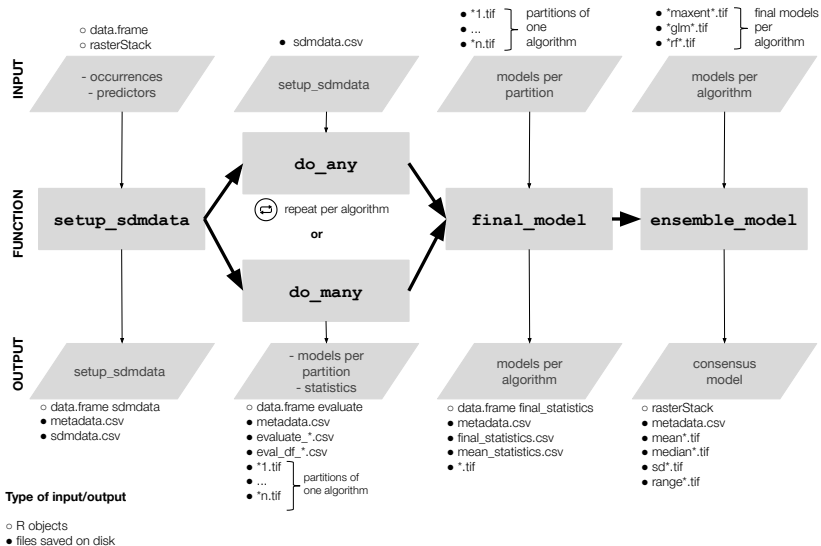
  

group.all	pa	lon	lat	layer.1	layer.2	layer.3	layer.4
3	1	-40.615	-19.921	-8.9611081	6.11421264	1.40630368	-7.
3	1	-40.729	-20.016	-8.7214928	5.87764334	1.37790286	-7.
4	1	-41.174	-20.303	-8.2562377	5.83755751	1.26637686	-8.
4	1	-41.74	-20.493	-7.2750323	3.61264467	0.81564361	-8.
2	1	-42.482	-20.701	-7.8781131	5.98075427	1.2590602	-9.

- metadata and session information
- creates a data frame `sdmdata.csv` that will be used in the next step

**Step 2.**  $\text{do}_{-}(m)$  any

# do\_(m)any(): model fitting and projection



`do_(m)any()`: model fitting and projection

## do\_(m)any(): model fitting and projection

- ▶ `do_any()` for one algorithm and partition (ex. `algo = "maxent"`)

## do\_(m)any(): model fitting and projection

- ▶ `do_any()` for one algorithm and partition (ex. `algo = "maxent"`)
- ▶ `do_many()` calls `do_any()` to fit multiple algorithms (ex: `bioclim = TRUE, maxent = TRUE`)

## do\_(m)any(): model fitting and projection

- ▶ `do_any()` for one algorithm and partition (ex. `algo = "maxent"`)
- ▶ `do_many()` calls `do_any()` to fit multiple algorithms (ex: `bioclim = TRUE, maxent = TRUE`)
- ▶ Parametrization

## do\_(m)any(): model fitting and projection

- ▶ `do_any()` for one algorithm and partition (ex. `algo = "maxent"`)
- ▶ `do_many()` calls `do_any()` to fit multiple algorithms (ex: `bioclim = TRUE, maxent = TRUE`)
- ▶ Parametrization
- ▶ The user can apply a mask



## do\_(m)any(): model fitting and projection

- ▶ `do_any()` for one algorithm and partition (ex. `algo = "maxent"`)
- ▶ `do_many()` calls `do_any()` to fit multiple algorithms (ex: `bioclim = TRUE, maxent = TRUE`)
- ▶ Parametrization
- ▶ The user can apply a mask
- ▶ Projection to different datasets (in time or space)

## do\_(m)any(): model fitting and projection

- ▶ `do_any()` for one algorithm and partition (ex. `algo = "maxent"`)
- ▶ `do_many()` calls `do_any()` to fit multiple algorithms (ex: `bioclim = TRUE, maxent = TRUE`)
- ▶ Parametrization
- ▶ The user can apply a mask
- ▶ Projection to different datasets (in time or space)
- ▶ Returns table with performance statistics → TSS, AUC, pROC, FNR, Jaccard...

do\_(m)any(): model fitting and projection

Current algorithms:

## do\_(m)any(): model fitting and projection

Current algorithms:

- ▶ bioclim, mahalanobis distance, maxent from **dismo** package

## do\_(m)any(): model fitting and projection

Current algorithms:

- ▶ bioclim, mahalanobis distance, maxent from **dismo** package
- ▶ Boosted regression trees (BRT) as implemented by `gbm.step()` function in **dismo** package

## do\_(m)any(): model fitting and projection

Current algorithms:

- ▶ bioclim, mahalanobis distance, maxent from **dismo** package
- ▶ Boosted regression trees (BRT) as implemented by `gbm.step()` function in **dismo** package
- ▶ maxent from **maxnet** package

## do\_(m)any(): model fitting and projection

Current algorithms:

- ▶ bioclim, mahalanobis distance, maxent from **dismo** package
- ▶ Boosted regression trees (BRT) as implemented by `gbm.step()` function in **dismo** package
- ▶ maxent from **maxnet** package
- ▶ GLM from base R, implemented with a stepwise selection approach

## do\_(m)any(): model fitting and projection

Current algorithms:

- ▶ bioclim, mahalanobis distance, maxent from **dismo** package
- ▶ Boosted regression trees (BRT) as implemented by `gbm.step()` function in **dismo** package
- ▶ maxent from **maxnet** package
- ▶ GLM from base R, implemented with a stepwise selection approach
- ▶ Support Vector Machines (SVM), from packages **kernlab** and **e1071**



## do\_(m)any(): model fitting and projection

Current algorithms:

- ▶ bioclim, mahalanobis distance, maxent from **dismo** package
- ▶ Boosted regression trees (BRT) as implemented by `gbm.step()` function in **dismo** package
- ▶ maxent from **maxnet** package
- ▶ GLM from base R, implemented with a stepwise selection approach
- ▶ Support Vector Machines (SVM), from packages **kernlab** and **e1071**
- ▶ Random Forests from package **randomForest**

do\_(m)any() output

At the end of the model fitting stage:

do\_(m)any() output

At the end of the model fitting stage:

- ▶ Outputs in the hard disk: .tif, .png for each partition

do\_(m)any() output

At the end of the model fitting stage:

- ▶ Outputs in the hard disk: .tif, .png for each partition
- ▶ Evaluation data frames with the performance statistics at different thresholds

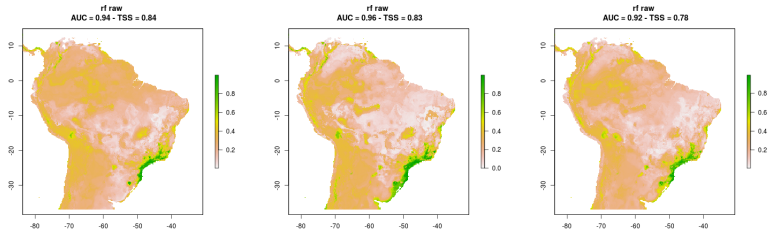
do\_(m)any() output

At the end of the model fitting stage:

- ▶ Outputs in the hard disk: .tif, .png for each partition
- ▶ Evaluation data frames with the performance statistics at different thresholds
- ▶ Metadata and session information

do\_(m)any() output

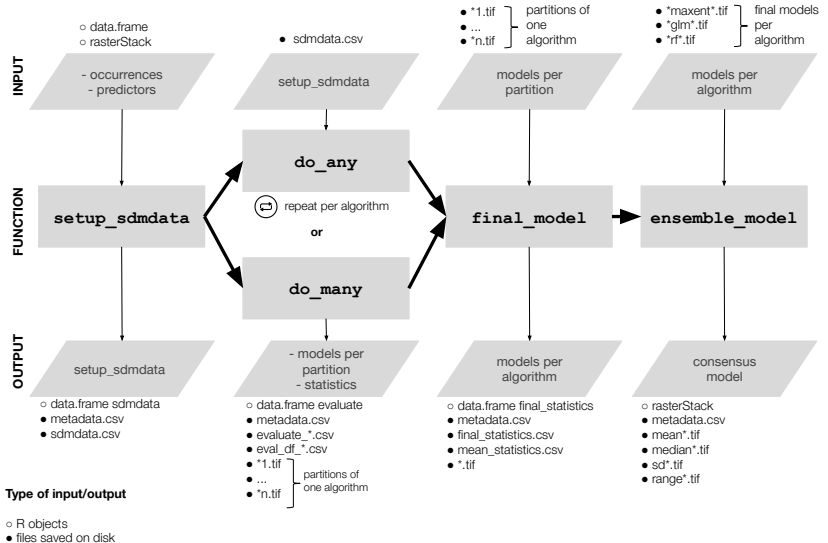
one fit model per partition



*Abarema langsoffii*, three partitions, randomForests

**Step 3.** `final_model`

# final\_model(): a model per algorithm per species





`final_model()`: a model per algorithm per species

`final_model()`: a model per algorithm per species

- ▶ The basics: a central tendency measure and uncertainty between partitions

`final_model()`: a model per algorithm per species

- ▶ The basics: a central tendency measure and uncertainty between partitions
- ▶ Which models to join? (the raw continuous model, the binary)

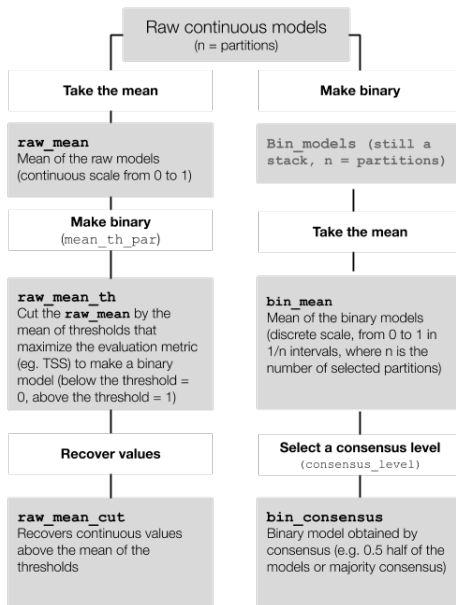
`final_model()`: a model per algorithm per species

- ▶ The basics: a central tendency measure and uncertainty between partitions
- ▶ Which models to join? (the raw continuous model, the binary)
- ▶ Some additional operations: consensus between binary models

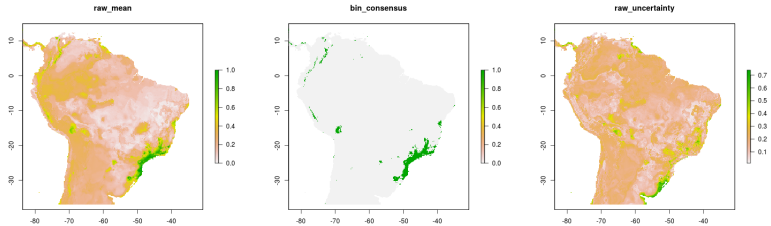
`final_model()`: a model per algorithm per species

- ▶ The basics: a central tendency measure and uncertainty between partitions
- ▶ Which models to join? (the raw continuous model, the binary)
- ▶ Some additional operations: consensus between binary models
- ▶ Uncertainty: range (max - min) between partitions

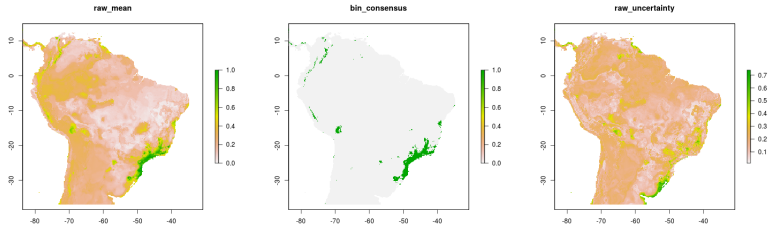
# `final_model()`: a model per algorithm per species



`final_model()`: a model per algorithm per species



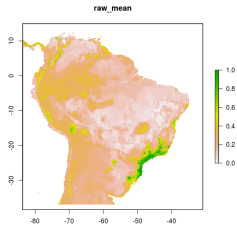
`final_model()`: a model per algorithm per species



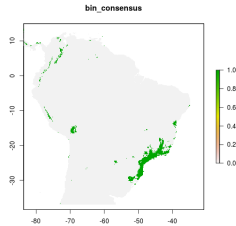
Raw mean



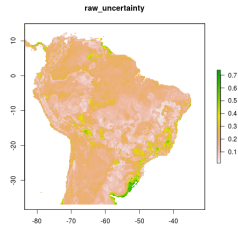
`final_model()`: a model per algorithm per species



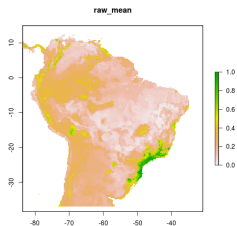
Raw mean



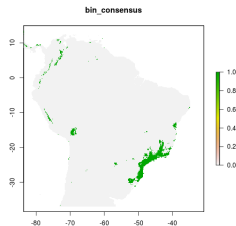
Binary consensus



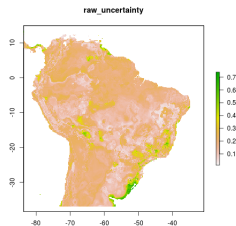
`final_model()`: a model per algorithm per species



Raw mean



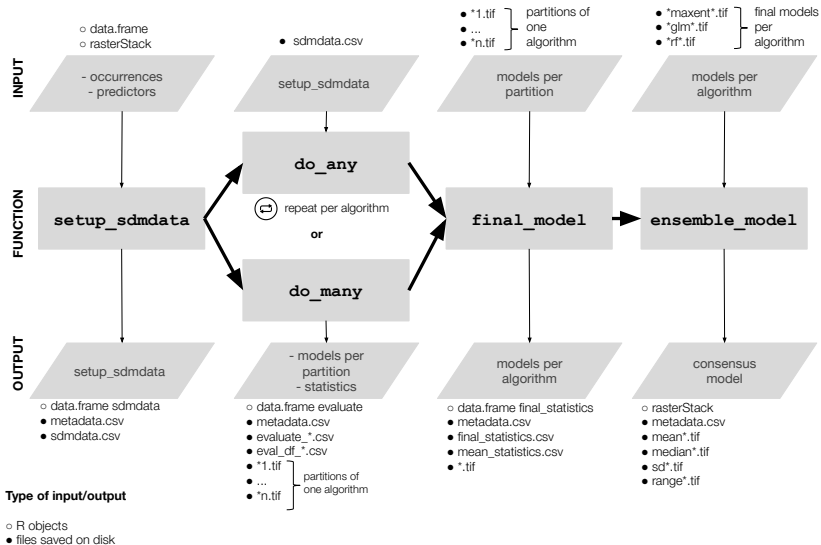
Binary consensus



Uncertainty (range)

**Step 4.** `ensemble_model`

# ensemble\_model()



`ensemble_model()`: algorithmic consensus

`ensemble_model()`: algorithmic consensus

- ▶ Mean between `final_models`

`ensemble_model()`: algorithmic consensus

- ▶ Mean between `final_models`
- ▶ Consensus

## ensemble\_model(): algorithmic consensus

- ▶ Mean between `final_models`
- ▶ Consensus
- ▶ Best-performing algorithm



## ensemble\_model(): algorithmic consensus

- ▶ Mean between `final_models`
- ▶ Consensus
- ▶ Best-performing algorithm
- ▶ PCA between algorithms

## ensemble\_model(): algorithmic consensus

- ▶ Mean between `final_models`
- ▶ Consensus
- ▶ Best-performing algorithm
- ▶ PCA between algorithms
- ▶ Range uncertainty metrics

## ensemble\_model(): algorithmic consensus

- ▶ Mean between `final_models`
- ▶ Consensus
- ▶ Best-performing algorithm
- ▶ PCA between algorithms
- ▶ Range uncertainty metrics
- ▶ Ensemble models do not necessarily perform better than individual algorithms (Zhu & Peterson 2017)

# Installing and using modleR

<https://model-r.github.io/modleR/>

# modleR: a workflow for ecological niche models

Andrea Sánchez-Tapia, Sara Mortara & Diogo S. B. Rocha

2020-04-19

Source: vignettes/modleR.Rmd

**modleR** is a workflow based on package **dismo** (Hijmans et al. 2017), designed to automatize some of the common steps when performing ecological niche models. Given the occurrence records and a set of environmental predictors, it prepares the data by cleaning for duplicates, removing occurrences with no environmental information and applying some geographic and environmental filters. It executes crossvalidation or bootstrap procedures, then it performs ecological niche models using several algorithms, some of which are already implemented in the **dismo** package, and others come from other packages in the R environment, such as **glm**, Support Vector Machines and Random Forests.



## Installing

Currently **modleR** can be installed from GitHub:

```
# Without vignette
remotes::install_github("Model-R/modleR", build = TRUE)
# With vignette
remotes::install_github("Model-R/modleR",
  build = TRUE,
  dependencies = TRUE,
  build_opts = c("--no-resave-data", "--no-manual"),
```

# other resources

## pseudoabsence

modleR **0.0.1**



Get started

Reference

Articles ▾

## Testing background point generation in modleR

Andrea Sánchez-Tapia & Sara Mortara

2020-04-18

Source: `vignettes/articles/buffer_and_randomPoints.Rmd`

This workflow tests background point generation in `modleR`. We perform tests with different types of buffer and different code options to sample pseudoabsences inside a geographic buffer. Later, we explore how different methods for sampling pseudoabsences result on different model predictions.

To run this example you will need `modleR` and the additional packages `rJava`, `raster`, and `dplyr`. To check if they are already installed and install eventually missing packages run the code below.

### Contents

The example data set

Inclusion buffers (distance-based and user-defined)

Exclusion filters in the geographic and environmental spaces

Superimposing buffers and filters

Using function `setup_sdmdata`

# other resources

## projections

modleR 0.0.1



Get started

Reference

Articles ▾

## Projecting models in modleR

Andrea Sánchez-Tapia

2020-04-19

Source: `vignettes/articles/projection.Rmd`

Here we explain the setup for projecting models in **modleR**.

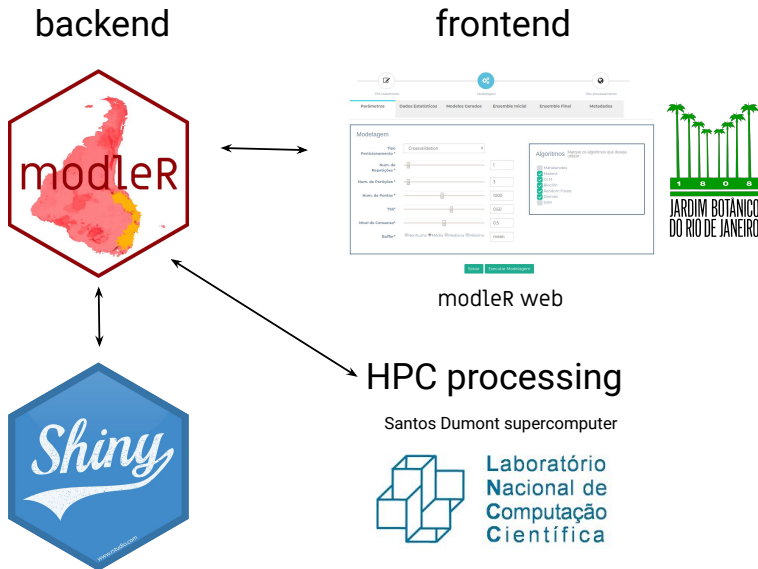
### Contents

Selecting one species as an example

Fitting and projection datasets

Final model projection

# modleR belongs to a larger framework







Pré-tratamento



Modelagem



Pós-processamento

Parâmetros

Dados Estatísticos

Modelos Gerados

Ensemble Inicial

Ensemble Final

Metadados

## Modelagem

Tipo  
Particionamento \*

Num. de  
Repetições \*

Num. de Partições \*

Num. de Pontos \*

TSS\*

Nível de Consenso\*

Buffer\* ☐ Nenhuma ☒ Média ☐ Mediana ☐ Máxima

## Algoritmos

Marque os algoritmos que deseja utilizar

- ☐ Mahalanobis
- ☒ Maxent
- ☒ GLM
- ☒ Bioclim
- ☒ Random Forest
- ☒ Domain
- ☐ SVM

Salvar

Executar Modelagem

# shiny application

modleR 3.0

Project

Species occurrence data

Environmental data

Data cleaning

Data setup

Projection setup

Modeling

Projection results?

Final models

Ensemble models

Results


Create/Open project

Select project:

Create new project

Insert project name:

Submit



A workflow to perform ecological niche modeling based on dismo

Please cite:

Sánchez-Tapia, Andrea ; de Siqueira, Marínez Ferreira ; Lima, Rafael Oliveira ; Barros, Felipe Sodré M. ; Gall, Guilherme M. ; Gadelha, Luiz M. R. ; da Silva, Luis Alexandre E. ; Osthoff, Carla . Model-R: A Framework for Scalable and Reproducible Ecological Niche Modeling. Communications in Computer and Information Science. 1ed.: Springer International Publishing, 2018, v. 796, p. 218-232.

## Model-R: A Framework for Scalable and Reproducible Ecological Niche Modeling

Andres Sánchez-Tapia<sup>1</sup>, Marínez Ferreira de Siqueira<sup>1</sup>, Rafael Oliveira Lima<sup>1</sup>, Felipe Sodré M. Barros<sup>2</sup>, Guilherme M. Gall<sup>3</sup>, Luiz M. R. Gadelha Jr.<sup>3</sup>, Luis Alexandre E. da Silva<sup>1</sup>, and Carla Osthoff<sup>3</sup>

<sup>1</sup> Botanic Garden of Rio de Janeiro, Rio de Janeiro, Brazil  
{andreasancheztapia, marinez, rafael, estevas}@brj.gov.br

<sup>2</sup> International Institute for Sustainability, Rio de Janeiro, Brazil  
f.barros@iis-rio.org

<sup>3</sup> National Laboratory for Scientific Computing, Petrópolis, Brazil  
{ggall, lgadelha, osthoff}@lncc.br

Botanical Journal of the Linnean Society, 2017, 183, 348–358. With 3 figures.

## Palaeodistribution of epiphytic bromeliads points to past connections between the Atlantic and Amazon forests

JEFFERSON RODRIGUES MACIEL<sup>1,2\*</sup>, ANDREA SÁNCHEZ-TAPIA<sup>1</sup>, MARÍNEZ FERREIRA DE SIQUEIRA<sup>1</sup> and MARCCUS ALVES<sup>1</sup>

<sup>1</sup>Jardim Botânico do Recife, Km 7,5 da BR 232, s/n, Curado 50000-230, Recife, PE, Brazil

<sup>2</sup>Instituto de Pesquisas do Jardim Botânico do Rio de Janeiro, Rua Pádua Leão 915, Jardim Botânico, Rio de Janeiro 22460-030, RJ, Brazil

<sup>3</sup>Universidade Federal de Pernambuco, Laboratório de Morfo-Taxonomia Vegetal, Av. Moraes Rego, s.n., CDU, 50670-930 Recife, PE, Brazil

Received 21 June 2016; revised 10 November 2016; accepted for publication 28 November 2016

Botanical Journal of the Linnean Society, 2015, 183, 1–10. With 9 figures.

## Environmental and geographical space partitioning between core and peripheral *Myrsine* species (Primulaceae) of the Brazilian Atlantic Forest

ANDREA SÁNCHEZ-TAPIA<sup>1</sup>, MÁRIO L. GARREIN<sup>1</sup>, MARÍNEZ F. SIQUEIRA<sup>1</sup>, KAUÍO G. GUIDONI-MARTINS<sup>2</sup>, FÁBIO B. SCARANO<sup>1,3,4\*</sup> and TATIANA T. CABELLO<sup>5\*</sup>

<sup>1</sup>Jardim Botânico do Rio de Janeiro—JBRJ, Rua Pádua Leão 915, Jardim Botânico, Rio de Janeiro, RJ, Brazil

<sup>2</sup>Universidade Vila Velha, Programa de Pós-Graduação em Ecologia de Ecossistemas, Laboratório de Ecologia Vegetal, Vila Velha, Espírito Santo, Brazil

<sup>3</sup>Universidade Federal do Espírito Santo, Centro Universitário Norte do Espírito Santo, 10 Rodovia BR 101 Norte, Km 01, Bairro Litorâneo, 29052-540, São Mateus, Espírito Santo, Brazil

<sup>4</sup>Programa de Pós-Graduação em Ecologia e Evolução, Universidade Federal do Góia, CP 131, Goiânia, GO 74001-970, Brazil

<sup>5</sup>Fundação Brasileira para o Desenvolvimento Sustentável, Rua Engenheiro Álvaro Niemeyer 76, São Conrado, 28910-180, Rio de Janeiro, RJ, Brazil

<sup>6</sup>Universidade Federal do Rio de Janeiro, Departamento de Ecologia, IB, CCS, Ilha do Fundão, 21941-970, Rio de Janeiro, RJ, Brazil

<sup>7</sup>Universidade Federal do Espírito Santo, Centro de Ciências Naturais e da Saúde, Rua Alto Universitário s.n., Laboratório de Botânica, 29050-000, Alegre, ES, Brazil

Received 20 May 2015; revised 11 December 2015; accepted for publication 28 April 2016

## ARTICLES

<https://doi.org/10.1111/1365-3113.12434>

nature  
ecology & evolution

## Strategic approaches to restoring ecosystems can triple conservation gains and halve costs

Bernardo B. N. Strassburg<sup>1,2,3\*</sup>, Hawthorne L. Beyer<sup>4</sup>, Renato Crouzeilles<sup>1,2,3</sup>, Alvaro Iribarrem<sup>1,2</sup>, Felipe Barros<sup>1</sup>, Marínez Ferreira de Siqueira<sup>1</sup>, Andrea Sánchez-Tapia<sup>1</sup>, Andrew Balmford<sup>5</sup>, Jerônimo Boelsmans Barreto Sansevero<sup>6</sup>, Pedro Henrique Santin Brancalion<sup>7</sup>, Eben North Broadbent<sup>8</sup>, Robin L. Chazdon<sup>1,2,3,4</sup>, Ary Oliveira Filho<sup>9</sup>, Toby A. Gardner<sup>10</sup>, Ascelin Gordon<sup>11</sup>, Agnieszka Latawiec<sup>1,2,3,4</sup>, Rafael Loyola<sup>12</sup>, Jean Paul Metzger<sup>13</sup>, Morena Mills<sup>14</sup>, Hugh P. Possingham<sup>15,16</sup>, Ricardo Ribeiro Rodrigues<sup>17</sup>, Carlos Alberto de Mattos Scaramuzza<sup>18</sup>, Fabio Rubio Scarano<sup>1,2</sup>, Leandro Tamboosi<sup>19</sup> and Maria Uriarte<sup>15</sup>

final remarks

## final remarks

- ▶ Reproducibility should drive any ENM workflow

## final remarks

- ▶ Reproducibility should drive any ENM workflow
- ▶ Metadata are really useful - and necessary

## final remarks

- ▶ Reproducibility should drive any ENM workflow
- ▶ Metadata are really useful - and necessary
- ▶ Any ENM workflow should be easily adapted to HPC

## final remarks

- ▶ Reproducibility should drive any ENM workflow
- ▶ Metadata are really useful - and necessary
- ▶ Any ENM workflow should be easily adapted to HPC
- ▶ Flexibility to start and leave at any step is essential to guarantee a solid evaluation of ENMs



## final remarks

- ▶ Reproducibility should drive any ENM workflow
- ▶ Metadata are really useful - and necessary
- ▶ Any ENM workflow should be easily adapted to HPC
- ▶ Flexibility to start and leave at any step is essential to guarantee a solid evaluation of ENMs
- ▶ It is not a problem to have one more package if it integrates with other packages

soon!



bioRxiv  
THE PREPRINT SERVER FOR BIOLOGY

New Results

[Comment on this paper](#)

## modleR: a modular workflow to perform ecological niche modeling in R

 Andrea Sánchez-Tapia,  Sara Ribeiro Mortara,  Diogo Souza Bezerra Rocha, Felipe Sodré Mendes Barros, Guilherme Gall,  Marínez Ferreira de Siqueira

**doi:** <https://doi.org/10.1101/2020.04.01.021105>

This article is a preprint and has not been certified by peer review [what does this mean?].

**Abstract**

Full Text

Info/History

Metrics

 Preview PDF