

# Package **jSDM** functionalities

Jeanne Clément

2021-03-19

## 1 Abstract

- Benefits of JSDMs over SDMS
- Many packages available to adjust JSDMs but which have some limitations
- package **jSDM** optimised C++ routines to reduce compilation time and new functionalities : long format data, missing data, multiple visits by site, priors specification
- Comparison of **Hmsc**, **boral** and **jSDM** librairies compilation time, accuracy and predictive performance for fitting JSDMs on Wilkinson et al. (2019) datasets
- Case-Study using a large Madagascar forest inventory, maps of biodiversity
- Expected results : compilation time reduced, equivalent or better accuracy and predictive performance
- Discussion

## 2 Introduction

### 2.1 State of the art

- Benefits of JSDMs over SDMs
- Librairies for fitting JSDMs : table to compare functionalities of **jSDM**, **HMSC**, **s-jSDM**, **boral** (JAGS), **greta**, **gjam**, **BayesComm**, **STAN** and **gllvm** (title Methods and software to fit joint species distribution models)

### 2.2 Knowledge gap

- Limitations of existing librairies
  - compilation time on large datasets
  - presence only data or presence absence data with multiple visits by site
  - datasets with missing data
  - spatially explicit models
  - specify mixed models with fixed or random species and site effects
  - (taking into account species traits and phylogeny as explicative variables of the model)

### 2.3 Response

- Implementation of **jSDM** package with optimised C++ code (Armadillo and GSL librairies)
  - Open source, new functionalities easy to implement, package évolutif
  - new functionalities : fourth corner model, long format
  - lower compilation time

- Application on Madagascar forest inventories, current (and future) climate conditions (and species traits datasets)

## 3 Methods

### 3.1 The structure of the jSDM problem

- GLMM and LVM Warton et al. (2015)
- models definition
  - link function : probit, logit, log
  - data type : presence/absence, presence/absence with multiple visits by site, abundance
  - data format : wide and long format
  - mixed models with random or fixed site and species effect
  - fourth corner models including species traits

### 3.2 Current approaches to fit the jSDM model structure

- GLMM and LVM ( $\Sigma = \Lambda\Lambda'$ ) Warton et al. (2015)
- Bayesian inference method :
  - priors specification
  - Gibbs sampling using conjugate prior formula (probit link function)
  - Metropolis-Hastings algorithm (logit and log link function)

### 3.3 Benchmarking our method against state-of-the-art jSDM implementation

- datasets Wilkinson et al. (2019)
- n.mcmc, compilation time, deviance
- compute AUC to assess the predictive performance of the models

### 3.4 Case study - Inference of biodiversity maps in Madagascar from forest inventories

#### 3.4.1 Datasets

##### 3.4.1.1 Forest inventory

##### 3.4.1.2 Current environmental variables

#### 3.4.2 Fitting joint species distribution model (JSDM)

##### 3.4.2.1 Model definition

##### 3.4.2.2 Parameters inference

##### 3.4.2.3 Evaluation of MCMC convergence

### 3.4.3 Evaluation of predictive performance

- compute AUC on inventory sites to assess the predictive performance of the model

### 3.4.4 Computation of species-species associations

### 3.4.5 Interpolation of parameters between inventory sites

- GRASS + rgrass7
- Regularized splines with tension (RST) Mitášová and Hofierka (1993)

### 3.4.6 Species richness computation

- use thresholds which maximize TSS to get presence/absence maps from occurrence probabilities maps
- sum presence/absence maps to get species richness estimation

### 3.4.7 Species turnover computation

- PCA for each pixel of the presence/absence maps (Allnutt et al. (2008))
- coordinates on the first 3 axes of the PCA reflect the composition of the species community likely to occupy the corresponding pixel
- coordinates scaled to  $[0.255]$  so that they can be represented by red colour levels for the first axis, green for the second and blue for the third
- on resulting  $\beta$  diversity map a colour difference between two pixels indicates that the species present are not the same, while pixels of identical colour host communities of similar species.

## 4 Results

### 4.1 Method validation and benchmark against state-of-the-art jSDMs

Table comparison HMSC, boral, jSDM on Wilkinson et al. (2019) datasets + birds + mites + aravo (traits) + simulated dataset n.mcmc, compilation time, deviance, predictive performance (TSS or AUC) + RMSE on simulated dataset

#### 4.1.1 Computational speed

#### 4.1.2 Accuracy of the inference about species-environment and species-species associations

comparison of TSS and RMSE obtained with different libraries

#### 4.1.3 Predicting species occurrences

comparison of AUC obtained with different libraries

## **4.2 Case study - Inference of biodiversity maps in Madagascar from forest inventories**

### **4.2.1 Predictive performance of the model**

- compute AUC on inventory sites

### **4.2.2 Estimated species-species associations**

### **4.2.3 Estimated species richness on inventory sites**

### **4.2.4 Estimated species richness at Madagascar scale**

### **4.2.5 Estimated species turnover at Madagascar scale**

## **5 Discussion**

### **5.1 Other software comparison**

Computational performance :

Inferential performance :

jSDM advantages : - lower compilation time than most other softwares - Open source - new functionalities  
easy to implement (evolutive package) - complete documentation

### **5.2 Limitation of jSDM and perspective**

- very large dataset still too long fitting, sjSDM usefull in this case
- phylogeny
- spatially explicit models
- presence only data

### **5.3 Implications and outlook for ecological data analysis**

- to predict species distributions area
- to estimate species diversity ( $\alpha$ ), species turnover ( $\beta$ ) or other diversity indices
- to predict the evolution of biodiversity under climate change
- to identify refuge zone of biodiversity
- to explain difference of vulnerability to climate change between species in relation to their functional traits

## 6 Acknowledgments

## 7 Data Accessibility

## References

- Allnutt, Thomas F., Simon Ferrier, Glenn Manion, George V. N. Powell, Taylor H. Ricketts, Brian L. Fisher, Grady J. Harper, et al. 2008. “A Method for Quantifying Biodiversity Loss and Its Application to a 50-Year Record of Deforestation Across Madagascar.” *Conservation Letters* 1 (4) : 173–81. <https://doi.org/10.1111/j.1755-263X.2008.00027.x>.
- Mitášová, Helena, and Jaroslav Hofierka. 1993. “Interpolation by Regularized Spline with Tension : II. Application to Terrain Modeling and Surface Geometry Analysis.” *Mathematical Geology* 25 (6) : 657–69. <https://doi.org/10.1007/BF00893172>.
- Warton, David I., F. Guillaume Blanchet, Robert B. O’Hara, Otso Ovaskainen, Sara Taskinen, Steven C. Walker, and Francis K.C. Hui. 2015. “So Many Variables : Joint Modeling in Community Ecology.” *Trends in Ecology & Evolution* 30 (12). Elsevier : 766–79. <https://doi.org/10.1016/j.tree.2015.09.007>.
- Wilkinson, David P., Nick Golding, Gurutzeta Guillera-Arroita, Reid Tingley, and Michael A. McCarthy. 2019. “A Comparison of Joint Species Distribution Models for Presence-Absence Data.” *Methods in Ecology and Evolution* 10 (2) : 198–211. <https://doi.org/10.1111/2041-210X.13106>.