A Julia toolkit for species distribution data

Timothée Poisot

Département de Sciences Biologiques, Université de Montréal, Montréal, Canada

timothee.poisot@umontreal.ca

Abstract   LATER

Introduction

Species Distribution Models (SDMs) are one of the most effective predictive approach to study the global distribution of biodiversity (Elith and Leathwick 2009). The training and evaluation of a SDM requires many steps, governing both its design and reporting (Zurell et al. 2020) and ultimate use and interpretation (Araújo et al. 2019). In the recent years, there has been an increase in the number of software packages and tools to assist ecologists with the development of species distribution models. As Kass et al. (2024) point out, this increase in the diversity of packages (most of them in the **R** language) is a good thing, as it can accommodate multiple workflows, and contributes to the adoption of good practices in the field.

Because the practice of species distribution modeling and analysis usually involve many different data types, tools that can provide an integrated environment are important: many existing packages have been designed independently, and therefore may suffer when it comes to interoperability. In this manuscript, we present **SpeciesDistributionToolkit** (abbreviated as **SDT**), a meta-package for the **Julia** programming language, offering an integrated environment for the retrieval, formatting, and interpretation of data relevant to the modeling of species distributions.

previous pub Dansereau and Poisot (2021)

Griffith et al. (2024) for large-scale SDM

Application description

SpeciesDistributionToolkit is released as a package for the **Julia** programming language, licensed under the open-source initiative approved MIT license. The package is registered in the **Julia** package repository and can be downloaded and installed anonymously. The full source and complete edition history is available at `https://github.com/PoisotLab/SpeciesDistributionToolkit.jl`. This page has a link to the documentation, containing a full reference for the package functions, a series of briefs how-to examples, and longer vignettes showcasing more integrative examples.

Component packages

An overview of the **SDT** package is given in Figure 1. The project is organized as a "monorepo", in which multiple packages live. This allows expanding the scope of the package by moving functionalities into new component packages, without complexifying the installation process. As **SDT** is registered in the **Julia** package repository, it can be installed with:

```
import                                    Pkg;
Pkg.add("SpeciesDistributionToolkit")
```

When loading the **SDT** package with `using SpeciesDistributionToolkit`, all component packages are automatically and transparently loaded. Therefore, users do not need to know where a specific method or function resides to use it. In the next section, we discuss how this modular design ensure that we can grow the functionality of the toolkit over time, while maintaining strict backward compatibility *and* allowing full reproducibility of an analysis.
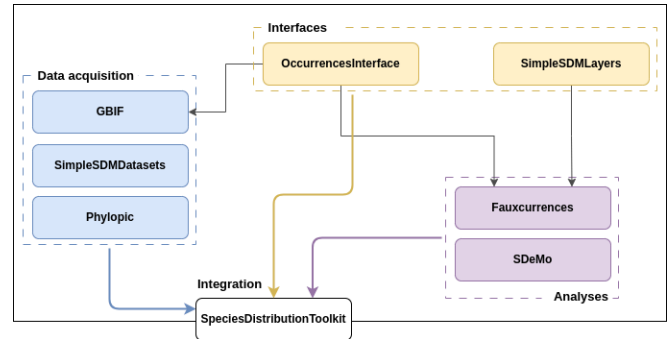

Figure 1: Overview of the packages included in **SpeciesDistributionToolkit**. The packages are color-coded by intended use, and their more specific content is presented in the main text. Note that because the package relies on *interfaces* to facilitate code interoperability, there are only three dependency relationships.

GBIF

SimpleSDMDatasets

Phylopic

OccurrencesInterface

SimpleSDMLayers

The **Fauxcurrences** packages is inspired by the work of Osborne et al. (2022), and

SDeMo

Software information

SDT uses the built-in **Julia** package manager to ensure that the version of all dependencies are kept up to date. Furthermore, we use strict semantic versioning: major versions correspond to no breaking changes in user-developed code, minor versions increase with additional functionalities, and patch releases cover minor bug fixes or documentation changes. All packages have a *CHANGELOG* file, which documents what changes are included in each release.

This strict reliance on semantic versioning solves the issues of maintaining compatibility when new functionalities are added: all releases in the *v1.x.x* branch of **SDT** depend on component packages in their respective *v1.x.x* branch, and users can benefit from now functionalities without risking to break existing code. This behavior is extensively tested, both using unit tests, and through integration testing generated as part of the online documentation.

Kellner, Doser, and Belant (2025) reported that about 20% of failures to reproduce species distribution or abundance modeling code was related to package issues. The strict reliance on semantic versioning, alongside technical choices in the **Julia** package manager and repository, means that it is possible to specify the full version of all dependencies used in a project, which addresses this important obstacle to reproducibility.

Integration with other packages

The **SDT** package benefits from close integration with other packages in the Julia universe. Notably, this includes **Makie** (and all related backends) for plotting and data visualisation, where usual plot types are overloaded for layer and occurrence data. Most data can be exported using the **Tables** interface, which allows data to be consumed by other packages like **DataFrames** and **MLJ**. Interfaces internal to Julia are also implemented whenever they make sense. Layers behave like arrays, are iterable, and broadcastable; occurrences collections are arrays and iterables.

Beyond supporting external interfaces, **SDT** defines its own internally. Access to raster data is supported by a trait-based interface for **SimpleSDMDatasets**.

Internal use of other interfaces like **StatsAPI** in **SDeMo**

one of the component packages (**OccurrencesInterface**) implements a minimalist interface to facilite the consumption of occurrence data.

Illustrative case studies

In this section, we provide a series of case studies, meant to illustrate the use of the package. The on-line documentation offers longer tutorials, as well as a series of how-to vignettes to illustrate the full scope of what the package allows. The code for each of these case studies is available as fully independent Jupyter notebooks, forming the supplementary material of this article. The example we use throughout is the distribution of *Akodon montensis* (Rodentia, family Cricetidae), and a host or orthohantaviruses (Burgos et al. 2021; Owen et al. 2010), in Paraguay. As the notebooks accompanying this article cover the full code required to run these examples, we do not present code snippets in the main text, and instead focus on explaining which component packages are used in each example.

Landcover consensus map

In this case study, we retrieve the land cover data from Tuanmu and Jetz (2014), clip them to a GeoJSON polygon describing the country of Paraguay (**SDT** can download data directly from `gadm.org`), and apply the `mosaic` operation to figure out which class is the most locally abundant. This case study uses the **SimpleSDMDatasets** package to download (and locally cache) the raster data, as well as the **SimpleSDMLayers** package to provide basic utility functions on raster data.
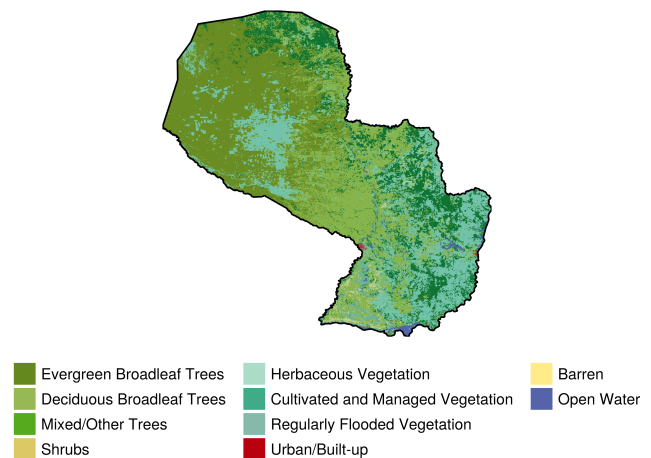

Figure 2: yeah

Using data from GBIF

(GBIF: The Global Biodiversity Information Facility 2025)
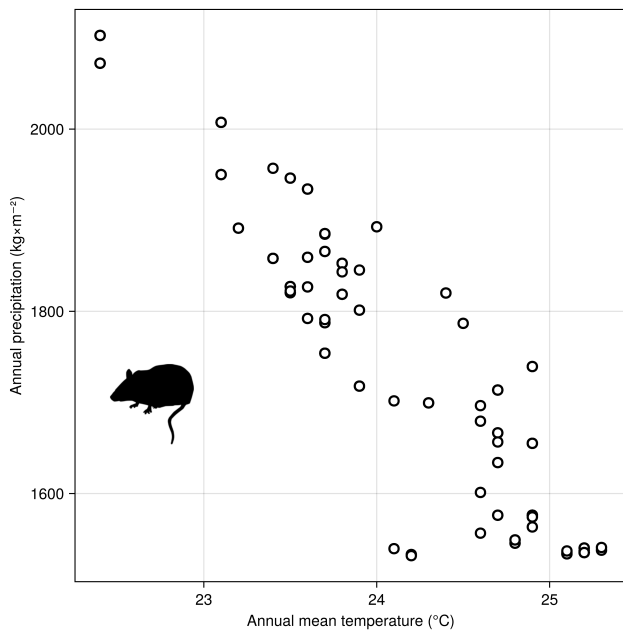
(Karger et al. 2017)

Figure 3: yeah

In practice, although the data are retrieved using the **GBIF** package, they are used internally by **SDT** through the **OccurrencesInterface** package. This package defines a small convention to handle georeferenced occurrence data, and allows to transparently integrate additional occurrence sources. By defining five methods for a custom data type, users can plug-in any occurrence data source and enjoy full compatibility with the entire **SDT** functionalities.

Training a species distribution model
(Bagnall et al. 2018)

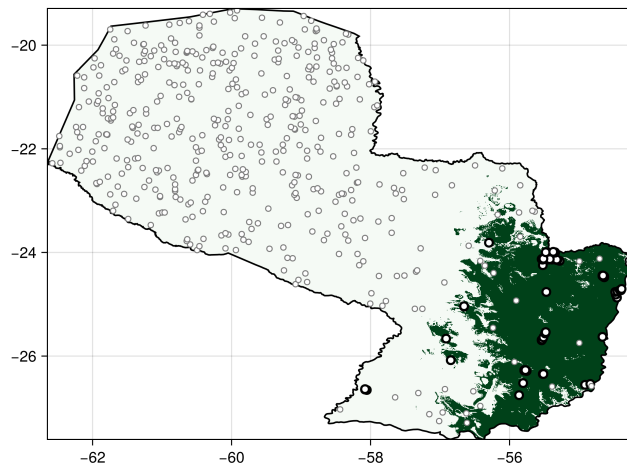PA routines from Barbet-Massin et al. (2012)

Figure 4: also yeah

Generating the distribution of a virtual species
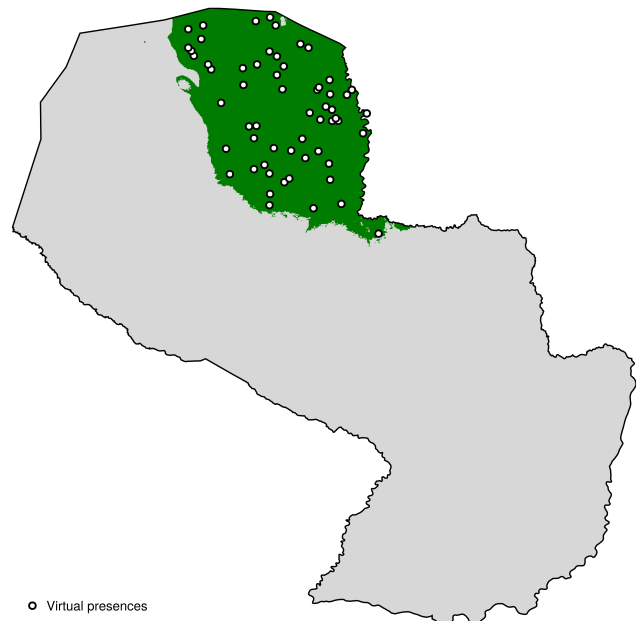(Leroy et al. 2016)

Figure 5: yeah

References

Araújo, Miguel B, Robert P Anderson, A Márcia Barbosa, Colin M Beale, Carsten F Dormann, Regan Early, Raquel A Garcia, et al. 2019. "Standards for Distribution Models in Biodiversity Assessments." *Science Advances* 5 (January): eaat4858. https://doi.org/10.1126/sciadv.aat4858.

Bagnall, A, M Flynn, J Large, J Line, A Bostrom, and G Cawley. 2018. "Is Rotation Forest the Best Classifier for Problems with Continuous Features?" *arXiv [Cs.LG]*, September.

Barbet-Massin, Morgane, Frédéric Jiguet, Cécile Hélène Albert, and Wilfried Thuiller. 2012. "Selecting Pseudo-absences for Species Distribution Models: How, Where and How Many?: How to Use Pseudo-Absences in Niche Modelling?" *Methods in Ecology and Evolution* 3 (April): 327–38. https://doi.org/10.1111/j.2041-210x.2011.00172.x.

Burgos, E F, M V Vadell, C M Bellomo, V P Martinez, O D Salomon, and I E Gómez Villafañe. 2021. "First Evidence of Akodon-Borne Orthohantavirus in Northeastern Argentina." *EcoHealth* 18 (December): 429–39. https://doi.org/10.1007/s10393-021-01564-6.

Dansereau, Gabriel, and Timothée Poisot. 2021. "SimpleSDMLayers.jl and GBIF.jl: A Framework for Species Distribution Modeling in Julia." *Journal of Open Source Software* 6 (January): 2872. https://doi.org/10.21105/joss.02872.

Elith, Jane, and John R Leathwick. 2009. "Species Distribution Models: Ecological Explanation and Prediction Across Space and Time." *Annual Review of Ecology, Evolution, and Systematics* 40 (December): 677–97. https://doi.org/10.1146/annurev.ecolsys.110308.120159.

GBIF: The Global Biodiversity Information Facility. 2025. "*What Is GBIF?*" 2025.

Griffith, Jory, Jean-Michel Lord, Michael D Catchen, Maria Isabel Arce-Plata, Manuel Fernandez Galvez Bohorquez, Matusan Chandramohan, María Camilla Diaz-Corzo, et al. 2024. "BON in a Box: An Open and Collaborative Platform for Biodiversity Monitoring, Indicator Calculation, and Reporting," October. https://doi.org/10.32942/X2M320.

Karger, Dirk Nikolaus, Olaf Conrad, Jürgen Böhner, Tobias Kawohl, Holger Kreft, Rodrigo Wilber Soria-Auza, Niklaus E Zimmermann, H Peter Linder, and Michael Kessler. 2017. "Climatologies at High Resolution for the Earth's Land Surface Areas." *Scientific Data* 4 (September): 170122. https://doi.org/10.1038/sdata.2017.122.

Kass, Jamie M, Adam B Smith, Dan L Warren, Sergio Vignali, Sylvain Schmitt, Matthew E Aiello-Lammens, Eduardo Arlé, et al. 2024. "Achieving Higher Standards in Species Distribution Modeling by Leveraging the Diversity of Available Software." *Ecography*, November. https://doi.org/10.1111/ecog.07346.

Kellner, Kenneth F, Jeffrey W Doser, and Jerrold L Belant. 2025. "Functional R Code Is Rare in Species Distribution and Abundance Papers." *Ecology* 106 (January): e4475. https://doi.org/10.1002/ecy.4475.

Leroy, Boris, Christine N Meynard, Céline Bellard, and Franck Courchamp. 2016. "Virtualspecies, an R Package to Generate Virtual Species Distributions." *Ecography* 39 (June): 599–607. https://doi.org/10.1111/ecog.01388.

Osborne, Owen G, Henry G Fell, Hannah Atkins, Jan van Tol, Daniel Phillips, Leonel Herrera-Alsina, Poppy Mynard, et al. 2022. "Fauxcurrence: Simulating Multi-species Occurrences for Null Models in Species Distribution Modelling and Biogeography." *Ecography* 2022 (July): e05880. https://doi.org/10.1111/ecog.05880.

Owen, Robert D, Douglas G Goodin, David E Koch, Yong-Kyu Chu, and Colleen B Jonsson. 2010. "Spatiotemporal Variation in Akodon Montensis (Cricetidae: Sigmodontinae) and Hantaviral Seroprevalence in a Subtropical Forest Ecosystem." *Journal of Mammalogy* 91 (April): 467–81. https://doi.org/10.1644/09-MAMM-A-152.1.

Tuanmu, Mao-Ning, and Walter Jetz. 2014. "A Global 1-km Consensus Land-cover Product for Biodiversity and Ecosystem Modelling: Consensus Land Cover." *Global Ecology and Biogeography: A Journal of Macroecology* 23 (September): 1031–45. https://doi.org/10.1111/geb.12182.

Zurell, Damaris, Janet Franklin, Christian König, Phil J Bouchet, Carsten F Dormann, Jane Elith, Guillermo Fandos, et al. 2020. "A Standard Protocol for Reporting Species Distribution Models." *Ecography* 43 (September): 1261–77. https://doi.org/10.1111/ecog.04960.