

## SOFTWARE NOTE

# `ecospat`: an R package to support spatial analyses and modeling of species niches and distributions.

Valeria Di Cola <sup>1</sup>, Olivier Broennimann <sup>1</sup>, Blaise Petitpierre <sup>1</sup>, Frank T. Breiner <sup>1,9</sup>, Manuela D'Amen <sup>1</sup>, Christophe Randin <sup>1,3,9</sup>, Robin Engler <sup>4,8</sup>, Julien Pottier <sup>6</sup>, Dorothea Pio <sup>1,7</sup>, Anne Dubuis <sup>1</sup>, Loic Pellissier <sup>5,9</sup>, Rubén G. Mateo <sup>1</sup>, Wim Hordijk <sup>1,10</sup>, Nicolas Salamin <sup>1,8</sup> & Antoine Guisan <sup>1,2</sup>

<sup>1</sup> Department of Ecology & Evolution (DEE) - University of Lausanne, Biophore Building, 1015 Lausanne, Switzerland.

<sup>2</sup> Institute of Earth Surface Dynamics (IDYST), University of Lausanne, Geopolis Building, 1015 Lausanne, Switzerland.

<sup>3</sup> Centre de Recherches sur les Ecosystèmes d'Altitude CREA. Observatoire du Mont-Blanc. Lacets du Belvédère 67 - 74400 Chamonix, France.

<sup>4</sup> Vital-IT group, SIB Swiss Institute of Bioinformatics, Génopode, Quartier Sorge, 1015 Lausanne, Switzerland.

<sup>5</sup> Landscape Ecology, Institute of Terrestrial Ecosystems, Universitätstrasse 16, ETH Zurich, 8092 Zürich, Switzerland.

<sup>6</sup> INRA, UR874-Grassland Ecosystem Research Unit, Clermont-Ferrand, France.

<sup>7</sup> Fauna & Flora International, 14 Buckingham Street, London, WC2N 6DF, UK.

<sup>8</sup> SIB Swiss Institute of Bioinformatics, Génopode, Quartier Sorge, 1015 Lausanne, Switzerland.

<sup>9</sup> Swiss Federal Research Institute WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland.

<sup>10</sup> Konrad Lorenz Institute for Evolution and Cognition Research, Martinstrasse 12, Klosterneuburg, Austria.

### Correspondence Author:

Valeria Di Cola, Olivier Broennimann

Department of Ecology & Evolution (DEE) - University of Lausanne, Biophore Building, 1015 Lausanne, Switzerland

valeria.dicola@unil.ch, olivier.broennimann@unil.ch

### Abstract

The aim of the `ecospat` package is to make available novel tools and methods to support spatial analyses and modeling of species niches and distributions in a coherent workflow. The package is written in the R language (R Development Core Team 2016) and contains several features, unique in their implementation, that are complementary to other existing R packages. Pre-modeling analyses include species niche quantifications and comparisons between distinct ranges or time periods, measures of phylogenetic diversity, and other data exploration functionalities (e.g. extrapolation detection, ExDet). Core modeling brings together the new approach of Ensemble of Small Models (ESM) and various implementations of the spatially-explicit modeling of species assemblages (SESAM) framework. Post-modeling analyses include evaluation of species predictions based on presence-only data (Boyce index) and of community predictions, phylogenetic diversity and environmentally-constrained species co-occurrences analyses. The `ecospat` package also provides some functions to supplement the `biomod2` package (e.g. data preparation, permutation tests and cross-validation of model predictive power). With this novel package, we intend to stimulate the use of comprehensive approaches in spatial modelling of species and community distributions.

## Background

Spatial ecology lies at the intersection of geography and ecology and examines the role of space in ecological processes from local to global scales. One of its overarching aims is to investigate what factors shape the distribution of species (Franklin 2009, Peterson et al. 2011) and, departing from these, to reconstruct communities and ecosystems (Levin 1992, Dungan et al. 2002, Guisan and Rahbek 2011, Hastings et al. 2011). It is now recognized that four main factors shape the distribution of species in time and space: environmental suitability, colonisation history and ability, biotic interactions, and natural and anthropogenic disturbances (Pulliam 2000, Lortie et al. 2004, Soberon 2007). In recent years, the use of niche-based, correlative species distribution models (hereafter SDMs, but other denominations are also used, such as ecological niche models ENMs, habitat suitability models HSMs, or climatic envelope models CEM; see Appendix S1 in Guisan et al. 2013) has risen spectacularly (Guisan et al. 2013), fostered by the increasing availability of numerical geodata and the tremendous progress in computer and statistical sciences (Dawson et al. 2013). As a result, many new methods have been developed (Skidmore et al. 2011) and a still increasing number of tools are becoming available to fit SDMs. Examples are *BioMapper* (Hirzel et al. 2001), *Maxent* (Phillips et al. 2006), *biomod2* (Thuiller et al. 2016), *dismo* (Hijmans et al. 2016), and *sdm* (Naimi and Araújo 2016). SDMs are increasingly used to address fundamental and applied questions in a wide range of domains, including community modeling (D'Amen et al. In Press), community phylogenetics (Ndiribe et al. 2013b), co-occurrence analyses (D'Amen et al. 2015a), biodiversity patterns (Mateo et al. 2016), anticipating and monitoring biological invasions (Guisan et al. 2014), rare species management (Breiner et al. 2015), or conservation phylogenetics (Pio et al. 2011), and the *ecospat* package offers functions to support analyses in these different fields.

Depending on where they take place in the modeling process, methods may be tentatively classified as pre-, core or post- modeling (or 'processing' in Naimi and Araújo 2016) analyses. Pre-modeling (or pre-processing) analyses are usually based on the observational data themselves, and include niche quantification (Fitzpatrick et al. 2007, Medley 2010), co-occurrence (Gotelli 2000), or phylogenetic diversity (Schweiger et al. 2008, Morlon et al. 2011) analyses. Core-modeling analyses are typically based on the quantification of species-environment relationship, using different techniques to calibrate models and use them to make spatial predictions (Guisan and Zimmermann 2000) both at the individual species and community levels. Finally, post-modeling (or post-processing) analyses make use of model predictions, for instance to incorporate a posteriori other filters of species distributions (Guisan and Rahbek 2011), such as biotic interactions (e.g. Wisz et al. 2013, D'Amen et al. 2015b), or dispersal (Dullinger et al. 2012, Engler et al. 2012). Several software and R packages exist to run core modeling analyses for individual species (see above), but fewer tools are available to support community-level modeling and the pre- and post-modeling analyses presented here (but see e.g. *dismo* Hijmans et al. 2016).

### *Aim of ecospat*

The aim of the *ecospat* package is to provide methods and utilities for spatial and/or temporal predictions of species distributions, communities' properties and related analyses (e.g. niche quantification, co-occurrence analyses). It contains a set of miscellaneous functions and methods to fit and evaluate SDMs and community models in a coherent analytical framework. The package includes novel functions (see below) and improvements of existing ones to supplement those already available in other packages. The interest of the package is that it possesses functions not only to run core modeling analyses at the individual species level, but also at the community level, and it additionally provides functions using raw observations and predictions to run pre- and post-modeling (or processing) analyses respectively.

## Novelty

The `ecospat` package includes many novel features. A major pre-modeling feature is the set of functions for niche quantification and tests of observed overlap in contrast to null distributions that were used in [Broennimann et al. \(2012\)](#), [Petitpierre et al. \(2012\)](#) and many other studies to quantify climatic niche shifts between the native and invaded ranges of alien species (see Guisan et al. 2014), to analyse niche overlap between different species (see Table 1) and to understand temporal dynamics in species niches and distributions (Nogués-Bravo 2009, Maguire et al. 2015). We are not aware of any other package (see below, section on niche analyses) – in or outside R – allowing to run the same niche analyses, including the possibility to measure niche unfilling, stability and expansion (see the COUE framework in [Guisan et al. \(2014\)](#), a shortening based on centroid shift, overlap, unfilling, and expansion of the niche), to define different environmental spaces, to correct for occurrence densities and environmental availability, and to run various tests.

In addition, two unique core modeling analyses in `ecospat` are functions to run the very recently proposed Ensemble of Small Models (ESM) approach (Breiner et al. 2015) on the one hand, and functions to implement the SESAM framework for spatial predictions of species assemblages (Guisan and Rahbek 2011, D'Amen et al. 2015a, 2015b) on the other hand. These functionalities represent new developments currently only available in `ecospat`. The package also uniquely contains functions to calculate the Boyce index (continuous or based on predefined bins Hirzel et al. 2006) to evaluate species distribution predictions when only independent presence data are available (Petitpierre et al. 2012).

A major post-modeling feature is the set of improved functions for environmentally-constrained co-occurrences analyses (based on Peres-Neto et al. 2001), where the environmental constraint is implemented by using SDM predictions. Another post-modeling functionality, nicely complementing the SESAM framework (e.g. to set phylogenetic constraints on assemblages), is the function to predict phylogenetic diversity in geographic space from the stacked predictions of species distribution models (e.g. Pio et al. 2011, 2014), which can also be used for simple pre-modeling analyses based on raw observations (Ndiribe et al. 2013a, 2013b, Pellissier et al. 2013a, 2013b).

## Technical features of the package

The `ecospat` package is written for use in the R language (R Development Core Team 2016). . The latest version 2.1.1 was recently released in CRAN (The Comprehensive R Archive Network) can be installed and used under R version 3.3 or higher (for more details see: <https://cran.r-project.org/package=ecospat> or use the link from <http://www.unil.ch/ecospat/home/menuguid/ecospat-resources/tools.html>). The `ecospat` package depends on a series of other R packages: `ade4` (Dray and Dufour 2007), `sp` (Qiao et al. 2016), `raster` (Hijmans 2015), `biomod2` (Thuiller et al. 2009), `gbm` (Zhang 2013), `MigClim` (Engler et al. 2012) and `ape` (Paradis et al. 2004). For details on the package dependencies, see Figure S1 and Table S1 in Supplementary Material.

Most functions in the `ecospat` package can be used to perform stand-alone analyses (e.g. niche or co-occurrence analyses), some functions are specifically designed to be used in supplement to other species distribution modeling packages (e.g. `biomod2`, `dismo`, `migclim`; see technical features above), but not to replace any functions in these packages. To avoid potential conflicts with other R functions, all the functions in the package start with the prefix '`ecospat.`' (see Table 1 for the main functions and their descriptions). The package `ecospat` is available from <http://cran.r-project.org/web/packages/ecospat/>, together with all example data sets referred to in the text. It is being constantly updated and we encourage interested users to look for the latest package version in CRAN or GitHub (<https://github.com/cran/ecospat>) and to report any suggestion for further improvement.

## Overview of functions and data sets

The main functions available in the `ecospat` package support either pre-, core- or post-modeling analyses, and are described in Table 1 and summarized in Figure S1 (in Supplementary Material) in relation to modeling functionalities. The `ecospat` package contains both novel features (see related section above) and more common functions already used or known elsewhere. In the latter cases, however, most functionalities were not yet made available in a wide-ranging R package, or were improved or implemented in `ecospat` in a different way than in other packages (e.g. spatial autocorrelation measure, multivariate environmental similarity surface (MESS), standard species co-occurrence analysis). It also includes basic but useful functionalities for e.g. preparing the data or projecting the models.

The package provides four test data sets to run examples: `ecospat.testData`, `ecospat.testTree`, `ecospat.testNiche.inv` and `ecospat.testNiche.nat`. The first two data sets allow illustrating the use of all community modeling functions, whereas the third and fourth ones allow illustrating niche quantification, here in the case of a hypothetical invasive species. The first data set, `ecospat.testData`, contains vegetation plots data with presence records of 50 vascular plant species (angiosperms), a set of associated environmental variables (topo-climatic) and SDM predictions for some species in an area of ~700 km<sup>2</sup> in the Western Swiss Alps (see <http://rechalpvd.unil.ch>). The second data set, `ecospat.testTree`, is a phylogenetic tree of class 'phylo' that contains data for the same 50 plant species. The third and fourth data sets, `ecospat.testNiche.inv` and `ecospat.testNiche.nat`, contain data on geographical coordinates of vegetation plots, a set of associated environmental variables (topo-climatic), occurrence sites for a hypothetical species and the prediction of its distribution in the native and invaded range.

In the next sections, we describe the key and most unique classes of methods implemented in the `ecospat` package. For the list of main functions, see Table 1 and the `ecospat` reference manual in CRAN.

### *Niche quantification, comparison and tests*

Niche quantification, comparison and tests are key components of `ecospat`. Methods for quantifying niches typically rely on either ordination techniques or on niche-based species distribution models (see Fig. 3 in Guisan et al. 2014). Compared to existing packages such as `NicheA` (Qiao et al. 2016), `spa` (Zhang 2013) and `SDMTools` (VanDerWal et al. 2014) that calculate niche overlap, the methods proposed here were initially the most comprehensive ones (see Broennimann et al. 2012). Niche functions in the `ecospat` package provide tools to quantify and compare species niches with an ordination approach. The `hypervolume` R package (Blonder et al. 2014) allows some niche quantification but without tests. Because niche functions in `ecospat` allow for direct comparisons of species–environment relationships in environmental space, no assumptions are required for the model-based approach (Broennimann et al. 2012). Likewise, the direct comparison can be made for a single variable (See example 1) or multiple variables. Most importantly, the package allows performing niche equivalency and niche similarity tests (Warren et al. 2008), which assess the statistical significance of a measured niche difference against null model niches taken randomly within a given background area. It is important to notice here that all functions in Broennimann et al. (2007), Broennimann et al. (2012), Petitpierre et al. (2012), and Broennimann et al. (2014a) are uniquely found in `ecospat` (see also Guisan et al. 2014). These functions were designed to investigate changes in the niche of invasive species, but they can also be used to compare niches between sister species (Broennimann et al. 2014b).

This approach consists of the following steps. First, an ordination technique - usually a principal component analysis (PCA) - is used to transform  $n$  correlated variables into two uncorrelated linear combinations (principal components) of the original variables. The PCA is calibrated using

environmental values from all the pixels of both the native and the invaded study areas. The axes of the PCA thus maximize the ecological variance present in the study areas. Then the PCA scores of the two species distributions, for which the niches must be compared, are projected onto a grid of cells bounded by the minimum and maximum PCA scores in the study areas. A smoothed density of occurrences for each species in each cell of the grid is then estimated using a kernel density function (in a similar way as in the package *hypervolume*). The global overlap between the niches can be calculated using metrics such as Schoener's D or Hellinger's I (see Broennimann et al. 2012 for details of the procedure and metrics). Additionally, in the case of invasive species, the niche overlap can be disentangled into three categories: unfilling, stability and expansion. This decomposition provides more information about the drivers of niche dynamic between native and invaded ranges (Petitpierre et al. 2012, Guisan et al. 2014), or outside biological invasions, about how two sister species have evolved different niches. Note that one can apply all these analyses on each variable individually (see Example 1), a very useful step to investigate which predictors are more conserved (e.g. Liu et al. 2016), and much more informative than boxplots comparisons (e.g. Fig. 1 in Mandle et al. 2010).

#### *Ensemble of Small Models (ESM)*

Another key feature of the package is the implementation of the ESM strategy to model species distributions when a limited numbers of presences or occurrences is available (e.g. for rare species, cryptic species), typically with less than 30 presence records (Wisz et al. 2008). The ESM approach was shown to reduce overfitting in the models and accordingly improve model accuracy in such cases (Lomba et al. 2010, Breiner et al. 2015). It is based on fitting numerous small models, using two (or possibly three or more) variables at each time and finally performs an ensemble of the predictions weighted by each submodel performance. It avoids overparameterization of the models, following the sample size rule-of thumb of 10:1 subjects to predictors in multiple regression (Harrell 2001). For more details on the method, see [Lomba et al. \(2010\)](#) and [Breiner et al. \(2015\)](#).

The ESM strategy can now be run and customized with four simple functions implemented in the `ecospat` package. None of the existing packages for modeling species distributions (e.g. `biomod2` (Thuiller 2003), `sdm` (Naimi and Araújo 2016), `dismo` (Hijmans et al. 2016)) currently contains the ESM approach. The ESM functions in `ecospat` depend on `biomod2` and allow the selection of the eleven different modeling techniques implemented there (and others if later added). Various evaluation indices to quantify model performance are available: AUC, TSS, Kappa, Somers D and Boyce. ESMs are therefore widely applicable and one or a set of preferred modeling techniques can be selected. The possibility to use multicore computation enables running ESMs in 'parallel' to speed up computation time. The strategy proposed with the ESM functions is as follows: (1) Calibration and evaluation of a series of small models with different techniques (i.e. GLM, Maxent, etc.); (2) Averaging them into a single ESM per technique, and finally; (3) Producing a single ensemble prediction through averaging all single-technique ESMs (i.e. running a double ensembling) (Breiner et al. 2015). Step 3 can be omitted if one does not want to run a double ensembling, because single-technique ESMs were shown to perform similarly to ESMs based on a double ensembling (i.e. fitting ESMs with different modelling techniques and ensembling them; Breiner et al. 2015). The final ESM can then be projected in a different time period or geographic space.

#### *Evaluating model predictions with presence-only data*

Additionally, the `ecospat` package provides the currently most exhaustive R implementation of the Boyce index, a presence-only and threshold-independent evaluator for SDMs' predictions (Hirzel et al. 2006). The Boyce index measures the trend in a plot showing the proportion of presences across classes of model predictions (similar to the POC-plot of Phillips and Elith 2010) and allows a conceptually correct way to evaluate predictions from presence-only models like Maxent (see Hirzel et al. 2006 for Boyce, and see Tsoar et al. 2007, Maher et al. 2014 for presence-only models). Therefore, it provides a more adequate evaluation in cases where absences do not carry reliable

information, typically in cases of non-equilibrium situations, such as with ongoing invasions by exotic species (Petitpierre et al. 2012), in case of insufficient detectability of a species (Wintle et al. 2012), or in the case where the model uses background data and not true absences (e.g. Maxent). The Boyce index measures how much model predictions differ from a random distribution of the observed presences across the prediction gradients (Boyce et al. 2002). It is the quantitative equivalent of the graphical POC plot (Phillips and Elith 2010) and often suggested to be so far the most appropriate metric in the case of presence-only models (Hirzel et al. 2006), although further research should be conducted to assess other possible options or possible refinements to evaluate presence-only predictions. The Boyce index is analogous to a Spearman correlation and varies between -1 and +1. Positive values indicate a model in which predictions are consistent with the distribution of presences in the evaluation data set, values close to zero mean that the model is not different from a random model, and negative values indicate counter predictions, i.e., predicting areas where presences are more frequent as being highly suitable for the species (Hirzel et al. 2006). Originally based on predefined (and often arbitrary) classifications of continuous suitability predictions, the Boyce index can also be threshold independent if based on a moving window, also known as continuous Boyce index (Hirzel et al. 2006). The implementation in `ecospat` proposes the standardized continuous index as default but also allows the user to parameterize predefined bins.

### *Spatial predictions of communities*

SDMs have been increasingly used over the last years to predict different community properties, such as species richness, taxonomic composition, functional and phylogenetic patterns, through stacking of individual species predictions (S-SDMs, Dubuis et al. 2011, Calabrese et al. 2014, Cord et al. 2014, D'Amen et al. In Press). A novel approach has been proposed to improve S-SDMs predictions by coupling them with richness predictions (Guisan and Rahbek 2011, D'Amen et al. 2015a, 2015b). This constraint on S-SDM predictions was proposed as one step in the larger 'spatially-explicit species assemblage modeling' (SESAM) framework. The latter seeks to reconstruct species assemblages by implementing the series of successive dispersal, habitat and biotic filters, and – if needed – the previously described macroecological constraints, as distinct modeling steps (Guisan and Rahbek 2011). SESAM, as implemented in `ecospat`, had been tested to date by applying three successive steps (D'Amen et al. 2015a, 2015b): 1) Calculating the habitat suitabilities of individual species presence for each site - these can typically be obtained by fitting an SDM for each species (i.e. S-SDM); this step represents the application of an environmental filter to the community assembly; 2) Computing richness predictions for each site - the richness prediction can be derived in different ways, for instance by summing probabilities from individual species prediction for each site, or by fitting direct macroecological richness models (MEM); this step represents the application of a macroecological constraint to the average number of species that can coexist in the considered unit; 3) Applying a biotic rule to decide which species, potentially present in the site, should be retained in the final assemblage prediction to match the predicted richness value. This last step is implemented in `ecospat` by the "probability ranking" rule (PRR), which consists of ranking the species in decreasing order of predicted probability of presence (obtained from the SDMs) and selecting species from the most probable species down the list until the sum of selected species reaches the expected richness value obtained in step 2 (D'Amen et al. 2015a, 2015b).

### *Environmentally-constrained species co-occurrence analyses*

In the context of predicting the spatial distribution of species assemblages, a key challenge is to identify significant species interactions (Wisze et al. 2013). In this regard, a novel post-modeling methodology of the `ecospat` package allows running environmentally-constrained pairwise species co-occurrence analyses based on a null model approach (as in Peres-Neto et al. 2001). This functionality uses the function `ecospat.cons_Cscore()` and allows to identify significant patterns of co-occurrence in a presence/absence species matrix, classifying each individual pair of species, as random, aggregated, or segregated. The strength of the association is calculated by the C-

score index (Stone and Roberts 1990), as  $C_{ij} = (R_i - D)(R_j - D)$ , where  $C_{ij}$  is the C score for species pair  $i$  and  $j$ ,  $R_i$  is the row total (the number of species occurrences) for species  $i$ ,  $R_j$  is the row total for species  $j$ , and  $D$  is the number of shared sites in which both species are present (Gotelli 2000). C-scores for individual species pairs and for the whole community are calculated and compared to the statistical expectation for a set of environmentally-constrained null communities. These are generated following the algorithm Ct-RA1 proposed in Peres-Neto et al. (2001), where the species presences are reassigned to sites in the matrix according to the relative probability values, maintaining fixed species frequencies (see Peres-Neto et al. 2001 for further explanations). The probability values in the sites can be obtained by fitting Species Distribution Models (SDMs) for each species. This function can also be run as a pre-modeling analysis if the environmental constraints applied to the null models are not based on the SDM predictions. This approach is expected to maximize the chance of distinguishing between mutually exclusive processes that may shape species distributions and community assembly, as environment is factored out as a possible explanation for the patterns encountered (Peres-Neto et al. 2001).

### *Evaluating community predictions*

When modeling species assemblages (Guisan and Rahbek 2011), it is known that the accumulation of small errors in individual models can produce larger errors in the final predictions (Dubuis et al. 2011, Pottier et al. 2013). Therefore, a robust evaluation of community predictions is an important step in spatial ecology analyses (Pineda and Lobo 2009, Pottier et al. 2013). The function `ecospat.CommunityEval()` allows a list of evaluation metrics (such as sensitivity, specificity, kappa, TSS, etc.) to be calculated for each site at the community level, permitting for a better understanding of the prediction accuracy and differences, for example in diverse habitat types at various elevations (e.g. Mateo et al. 2012, Pottier et al. 2013). The function can be applied to the outcomes of any community-level model, e.g. S-SDM or SESAM predictions.

### *Phylogenetic diversity measures*

The calculation of phylogenetic diversity measures on observed (pre-modeling) or predicted (post-modeling) coarse species assemblages (as in Pio et al. 2011, 2014) or local communities (Ndiribe et al. 2013a, 2013b), provides a comprehensive list of indices and optimized calculations for large data sets (which are then important for producing maps). The function `ecospat.calculate.pd()` uses a phylogenetic tree, a presence or absence (binary) matrix for each species in each location or grid cell (direct observations, S-SDM binary predictions or SESAM predictions), and different phylogenetic diversity measures (Schweiger et al. 2008). For each grid cell in the study area, the function calculates each of the phylogenetic diversity measures listed in [Schweiger et al. \(2008\)](#) based on either tree topology, branch length or minimum-spanning tree measures. Three measures are based on topology (node information only): W (standardized taxic weights), Q (basic taxic weights), and P (normalized measure of Q). Five measures are based on pairwise distances: J (intensive quadratic entropy), F (extensive quadratic entropy), AvTD (average taxonomic distinctness), TTD (total taxonomic distinctness), and Dd (pure diversity). Finally, three so-called minimum-spanning measures are based on both branch length and node information: PDroot (clade when root=FALSE; phylogenetic diversity with basal branches), PDnode (clade when root=TRUE; phylogenetic diversity), and AvPD (species; average phylogenetic diversity). These measures can be classified on the basis of whether they sum (Q, P, W, PDnode, PDroot, F, TTD, Dd) or average (AvTD, J, AvPD) the evolutionary history of all species present in an area. Specific examples using this function are found in (Pio et al. 2011, 2014). Different R packages allow similar analyses, but `ecospat` has a few novelties and differences here. For example, the `picante` R package does calculate the same phylogenetic diversity measures based on topology and minimum spanning measures, but it does not refer to them exactly with the same names (Kembel et al. 2010). The package `Phylomeasures` (Tsirogiannis and Sandel 2015) also provides functions to calculate the same measures. Unlike

*picante* and *Phylomeasures*, however, the function `ecospat.calculate.pd()` includes the additional option to calculate five measures based on pairwise distances (see Schweiger et al. 2008).

### Developed examples with *ecospat*

To illustrate the use of the *ecospat* package, we created two examples of analytical workflows that can be easily reproduced by users, involving the main set of functions in *ecospat*. In the first example, we illustrate how to compare the climatic niche of a hypothetical invasive species in its native and invaded ranges. We then predict its non-native distribution by applying an ensemble of small models (ESM) to avoid model overfitting because only few occurrences exist for the invasive species in its invaded range. We finally show how to calculate the Boyce index to evaluate the non-native prediction to assess model performance in cases like this where only presence data are available and absences may not be reliable (i.e. here the species is still expanding).

In the second example we perform a spatial analysis of a plant community, examining its phylogenetic and co-occurrence patterns and producing spatial predictions of the assemblages using novel functions in the *ecospat* package. We first compute different phylogenetic diversity measures, which can then be calculated on the predicted species assemblage as well, e.g. to estimate the difference in phylogenetic diversity under current and future climatic scenarios. We next illustrate the implementation of the SESAM framework to predict the spatial distribution of communities, applying a biotic rule to estimate the taxonomic composition of species in each site. We then calculate several evaluation indices of community-level models, in order to evaluate the community predictions. Finally, we analyze the spatial patterns of geographic overlap in the distributions of the species in the study area to illustrate the application of the environmentally-constrained null model. Complete scripts to run the two examples and to illustrate the use of the *ecospat* package are provided in the supplementary material S1 and S2.

### Example 1: Niche quantification and modeling of an invasive species

The first example illustrates the use of the functions to investigate whether the climatic niche of a one species may differ between different geographic areas, and the use of the ESM functions to model the species' distribution in the two ranges when few occurrences are available. This niche comparison framework is typically used to quantify and describe environmental niche shifts between the native and exotic ranges of invasive species or between sister species. The ESM functions can be used here to model distributions when the species is not (yet) very frequent, as typically at the start of an invasion in the exotic range. An overview of how the tools used in example 1 can be applied to different case studies is given in Table 1 (see also [Guisan et al. \(2014\)](#) and [Breiner et al. \(2015\)](#)).

In this example we first compare the niche of a species native in North America and introduced in Australia, then calculate a SDM using ESM, and finally evaluate the SDM's ability to predict the distribution with the Boyce index. The script in Supplementary Material S1 generates occurrence density, niche overlap, niche equivalency and similarity tests, and uses a brand new tool allowing the calculation and the visualization of the niche dynamic as objects in the R environment. The occurrence density generated by the function `ecospat.grid.clim.dyn()` measures the frequency of occurrences of the species for each combination of conditions (i.e. each grid cell) of the environmental space using a kernel smoother. The function `ecospat.niche.overlap()` uses the differences in occurrence densities between the two species to measure either Schönner's D or Hellinger's I metrics, both of which range from 0 (no overlap) and 1 (complete overlap). The functions `ecospat.niche.equivalency.test()` and `ecospat.niche.similarity.test()` perform tests of niche equivalency and similarity as described in [Warren et al. \(2008\)](#) but applied in environmental space, where the overlap is better assessed than in geographical space because it better takes into account climate availability and analogy between ranges (Broennimann et al. 2012, Guisan et al 2014; but note that the last version of the ENMtool by Warren et al. (2008) implemented the *ecospat* approach to test niche differences in environmental space). The niche equivalency test assesses, through random

permutations of occurrences between ranges, whether the two niches are equivalent. The niche similarity test assesses, through random shifts of the niches within available conditions in the study area, whether the species' niches are more or less similar than expected by chance. Here, we want to test for niche conservatism, so we use the alternative = "greater", i.e. the niche overlap is more similar than random expectations. Note that in the case of invasive species, we fixed the native niche as a reference and shifted only the invasive niche (`rand.type = 2`). In cases where there are no assumption about a reference niche (e.g. sister species), both niches can be simultaneously shifted (`rand.type = 1`). For details on niche equivalency and similarity tests, see [Warren et al. \(2008\)](#) and [Broennimann et al. \(2012\)](#). Figure 1A provides an example scheme of the output from the analysis of a species in the invaded and native range (Complete Script available in Supplementary Material S1). We see that the niche overlap D is 0.22, and the tests further indicate that the niches of the example species in the native and invaded ranges are not more equivalent neither not more similar than expected by chance with p-value of 1 and 0.079 respectively. Therefore, we can conclude that there is no significant climatic niche conservatism between native and invaded ranges. The niche dynamics analysis shows that this niche difference is due to the species' ability to expand into novel climates in the invaded range (niche expansion = 15%) and to the fact that the species has not (yet) colonized all the climate conditions of the native niche (niche unfilling = 28%). A visual inspection shows also that an important part of the cold climate conditions available in the native range are no longer available in the invaded range.

We next built an ESM (Ensemble of small models), by fitting with the `ecospat.ESM.Modeling()` function all possible models that contain only two predictors at a time out of all those available in the initial set of eight variables (from `ecospat.testNiche.inv`), resulting in 28 bivariate predictor combinations. Then all bivariate models were combined to an ensemble model (Lomba et al. 2010, Breiner et al. 2015) with the `ecospat.ESM.EnsembleModeling()` function. The function first evaluates each bivariate model by a cross-validated (repeated split-sample) evaluation index (in our case AUC), and builds then an average of the 28 models weighted by a cross-validation score (in our case Somers' D, which is  $AUC * 2 - 1$ ). Models with a Somers' D higher than 0 (i.e.  $AUC > 0.5$ ) were selected to build the final ensemble. For more details refer to [Breiner et al. \(2015\)](#). For details see Fig. 1B.

Finally, we evaluate the SDM's ability to predict the non-native distribution with the Boyce index, a presence only evaluator, with the function `ecospat.boyce()`. This function returns the Boyce index value: *Spearman.cor* (in our example 0.889), which is the Spearman correlation of the PE plot, a graphical plot of the predicted to expected ratio within a moving window along the suitability gradient (Fig. 1C). If the rank of the predicted to expected ratio would be completely ordered along the habitat suitability gradient, then the Boyce index would be 1. With a score of 0.889, we can assume that the SDM is useful to independently predict the species invasion in Australia (Fig. 1B). Additionally, such PE plots can be used to reclassify maps of continuous suitability values into discrete categories like potential presences or absences, which is particularly useful for conservation managers ([Hirzel et al. 2006](#)) or to recalibrate presence-only SDM's continuous suitability (Phillips and Elith 2010).

### **Example 2: Species assemblage's structure and spatial predictions**

The tools used in example 2 can be applied to provide an evaluation of the probability of occurrence of different species within a community, and to model the spatial distribution of richness and composition in a community, obtaining predictions of species assemblages more realistic than those obtained by stacking individual species predictions. For an overview of different questions that can be addressed using the tools of this example, refer to Table 1, Pio et al. (2011, 2014) and D'Amen et al. (2015a, 2015b). Along this second example, we explore different community properties of plant assemblages in the Swiss Alps. In particular, we analyze phylogenetic diversity, we predict plant species richness and composition and we use null model analysis to describe the spatial pattern of

pairwise species co-occurrence. First, we illustrate the use of the function `ecospat.calculate.pd()` by using presence-absence data with a phylogenetic tree in order to obtain different phylogenetic diversity measures, but the same function can also be applied to a predicted species assemblage. In the code in the Supplementary Material S2, we show how to load the phylogenetic tree and the data for the example. The resulting object is a list of phylogenetic diversity values for each of the grid cells in the presence/absence matrix, which can be plotted to obtain a graphical representation of the correlation of phylogenetic diversity with species richness (Fig. 2A). Second, we predict community composition with the function `ecospat.SESAM.prr()`. This function implements the final step of the SESAM assemblage, i.e. the “probability ranking rule” (PRR; D’Amen et al. 2015a). The data required for the function are two data frames of: (1) continuous probabilities from SDMs for all species (in columns) in the considered sites (in rows), and (2) richness value for each site (first column). In cases where the SDMs were fitted with presence/absence data, thus yielding true probabilities, the richness prediction can then be obtained by summing all raw probabilities (across species) at each site. Alternatively, a similar prediction can be obtained by a separate species richness model. The function returns a data frame of the same structure as the input one, containing the community prediction by the SESAM framework, i.e. binary predictions for the species in each site (see D’Amen et al. 2015a). For details see example in Fig. 2B. Third, we calculate community-predictions accuracy with different indices with the function `ecospat.CommunityEval()` that requires the observed presence-absence of the species (`eval`) and their predictions (`pred`). The returned object is a list of evaluation metrics calculated for each site (See Fig. 2C for details).

Finally, we perform a co-occurrence analysis and test for non-random patterns of species co-occurrences applying an environmentally-constrained null model (Peres-Neto et al. 2001), with the function `ecospat.cons_Cscore()`. The format required for the input data is a matrix of plots (rows) x species (columns). Input matrices should have column names (species names) and row names (sampling plots). The function returns the C-score index for the observed community (`ObsCscoreTot`), the mean of C-score for the simulated communities (`SimCscoreTot`), the p values (`PVal.less` and `PVal.greater`) to evaluate the significance of the difference between the former two indices, and finally returns the standardized effect size (SES) for the whole community (`SES.Tot`). A SES greater than 2 or smaller than -2 is statistically significant with a tail probability of less than 0.05 (Gotelli and McCabe 2002). If a community is structured by competition, we would expect the C-score to be large relative to a randomly assembled community (positive SES). Also, to test for patterns in each individual pair of species in the matrix, a table is saved in the specified working directory. This table contains the observed and expected C-score value and the standardized effect size for species pairs and it includes only those pairs with p values lower than 0.05. A further correction for false discovery rate may be needed on the basis of the dataset size (Gotelli and Ulrich 2010). This option is planned but not yet implemented in the package. Furthermore, the function automatically plots two frequency histograms: the first one represents the C-score indices from the simulated communities, where the observed C-score value is indicated by the black solid line (Fig. 2D), and the second histogram represents the distribution of the SES values calculated for all the unique combination of species pairs in the considered community (Fig. 2D). In this case, the observed C-score is significantly lower than expected by chance (in our example the p value is 1 e-4), meaning that the community is dominated by positive interactions (aggregated pattern; as can be expected in alpine plant communities under stressful conditions at high elevations; Callaway et al. 2002). The complete script to run the example is available in the Supplementary Material S2.

### **Citation of the `ecospat` package**

The latest available version of the `ecospat` package in CRAN is 2.1.1 and the license of the package is GPL (General Public License is a free, copyleft license for software and other kinds of works). It can be installed from CRAN using the usual code line as follows:

```
> install.packages("ecospat")
```

Scientists using `ecospat` functions in a published paper should cite this article or the `ecospat` package directly. Citation information can be obtained by typing:

```
> citation("ecospat")
```

on the R command prompt.

Some of the methods implemented in `ecospat` have been published (Table 1) but their scripts were only available in the related papers and so far not in open access. Users should refer to the respective publications for details of the methods and interpretation. Users of the package who apply these methods should cite as much as possible the original sources as appropriate, along with `ecospat`.

### **Acknowledgements**

We would like to thank the reviewers, Dr Nick Golding Dr. Matheus Lima-Ribeiro and an anonymous referee for their insightful comments on the paper, as these comments led us to an improvement of the work. Functions included in this package were developed in several projects mainly funded by the Swiss National Science Foundation (SNF; grants nr 31003A-125145, 31003A-152866, CR23I2\_162754), the SNF National Center of Competence in Research (NCCR) "Plant Survival", and the European Commission's sixth framework project 'EcoChange'. M.D. was supported by the FP7-PEOPLE-2012-IEF Marie Curie Action (Project nr 327987). R.G.M. was supported by the FP7-PEOPLE-2013-IEF Marie Curie Action (Project nr 622620). F.B was supported by the Federal Office for the Environment (FOEN) ('Modeling and Evaluation of Species Distributions', project number 8T20/10.0042.PJ)

## References

- Blonder, B. et al. 2014. The n-dimensional hypervolume. — *Glob. Ecol. Biogeogr.* 23: 595-609.
- Boyce, M. S. et al. 2002. Evaluating resource selection functions. — *Ecol. Model.* 157: 281-300.
- Breiner, F. T. et al. 2015. Overcoming limitations of modelling rare species by using ensembles of small models. — *Methods Ecol. Evol.* 6: 1210-1218.
- Broennimann, O. et al. 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. — *Glob. Ecol. Biogeogr.* 21: 481-497.
- Broennimann, O. et al. 2014a. Contrasting spatio-temporal climatic niche dynamics during the eastern and western invasions of spotted knapweed in North America. — *J. Biogeogr.* 41: 1126-1136.
- Broennimann, O. et al. 2007. Evidence of climatic niche shift during biological invasion. — *Ecol. Lett.* 10: 701-709.
- Broennimann, O. et al. 2014b. Influence of climate on the presence of colour polymorphism in two montane reptile species. — *Biology letters* 10: 20140638.
- Calabrese, J. M. et al. 2014. Stacking species distribution models and adjusting bias by linking them to macroecological models. — *Glob. Ecol. Biogeogr.* 23: 99-112.
- Callaway, R. M. et al. 2002. Positive interactions among alpine plants increase with stress. — *Nature* 417: 844-848.
- Cord, A. F. et al. 2014. Remote sensing data can improve predictions of species richness by stacked species distribution models: a case study for Mexican pines. — *J. Biogeogr.* 41: 736-748.
- D'Amen, M. et al. 2015a. Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models. — *J. Biogeogr.* 42: 1255-1266.
- D'Amen, M. et al. 2015b. Predicting richness and composition in mountain insect communities at high resolution: a new test of the SESAM framework. — *Glob. Ecol. Biogeogr.* 24: 1443-1453.
- D'Amen, M. et al. In Press. Spatial predictions at the community level: from current approaches to future frameworks. — *Biol. Rev.*
- Dawson, M. N. et al. 2013. An horizon scan of biogeography. — *Frontiers of Biogeography* 5:
- Dray, S. and Dufour, A. B. 2007. The ade4 package: Implementing the duality diagram for ecologists. — *J. Stat. Softw.* 22: 1-20.
- Dubuis, A. et al. 2011. Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. — *Divers. Distrib.* 17: 1122-1131.
- Dullinger, S. et al. 2012. Extinction debt of high-mountain plants under twenty-first-century climate change. — *Nature Clim. Change* 2: 619-622.
- Dungan, J. L. et al. 2002. A balanced view of scale in spatial statistical analysis. — *Ecography* 25: 626-640.
- Engler, R. et al. 2012. The MIGCLIM R package – seamless integration of dispersal constraints into projections of species distribution models. — *Ecography* 35: 872-878.
- Fitzpatrick, M. C. et al. 2007. The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? — *Glob. Ecol. Biogeogr.* 16: 24-33.
- Franklin, J. 2009. Mapping species distribution: spatial inference and prediction. — Cambridge University Press.
- Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. — *Ecology* 81: 2606-2621.
- Gotelli, N. J. and McCabe, D. J. 2002. Species co-occurrence: A meta-analysis of J. M. Diamond's assembly rules model. — *Ecology* 83: 2091-2096.
- Gotelli, N. J. and Ulrich, W. 2010. The empirical Bayes approach as a tool to identify non-random species associations. — *Oecologia* 162: 463-477.
- Guisan, A. et al. 2014. Unifying niche shift studies: insights from biological invasions. — *Trends Ecol. Evol.* 29: 260-269.

- Guisan, A. and Rahbek, C. 2011. SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. — *J. Biogeogr.* 38: 1433-1444.
- Guisan, A. et al. 2013. Predicting species distributions for conservation decisions. — *Ecol. Lett.* 16: 1424-1435.
- Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. — *Ecol. Model.* 135: 147-186.
- Harrell, F. E. 2001. *Regression Modeling Strategies: with applications to Linear Models, Logistic Regression, and Survival Analysis.* — Springer-Verlag.
- Hastings, A. et al. 2011. Spatial ecology across scales. — *Biol. Lett.* 7: 163-165.
- Hijmans, R. J. 2015. raster: Geographic Data Analysis and Modeling. R package version 2.5-2. <http://CRAN.R-project.org/package=raster>.
- Hijmans, R. J. et al. 2016. dismo: Species Distribution Modeling. R package version 1.0-15 <http://CRAN.R-project.org/package=dismo>.
- Hirzel, A. H. et al. 2001. Assessing habitat-suitability models with a virtual species. — *Ecol. Model.* 145: 111-121.
- Hirzel, A. H. et al. 2006. Evaluating the ability of habitat suitability models to predict species presences. — *Ecol. Model.* 199: 142-152.
- Kembel, S. W. et al. 2010. Picante: R tools for integrating phylogenies and ecology. — *Bioinformatics* 26: 1463-1464.
- Levin, S. A. 1992. The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. — *Ecology* 73: 1943-1967.
- Liu, X. et al. 2016. Realized climatic niches are conserved along maximum temperatures among herpetofaunal invaders. — *J. Biogeogr.*
- Lomba, A. et al. 2010. Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. — *Biol. Conserv.* 143: 2647-2657.
- Lortie, C. J. et al. 2004. Rethinking plant community theory. — *Oikos* 107: 433-438.
- Maguire, K. C. et al. 2015. Modeling species and community responses to past, present, and future episodes of climatic and ecological change. — *Annual Review of Ecology, Evolution, and Systematics* 46: 343-368.
- Maher, S. P. et al. 2014. Pattern-recognition ecological niche models fit to presence-only and presence-absence data. — *Methods Ecol. Evol.* 5: 761-770.
- Mandle, L. et al. 2010. Conclusions about Niche Expansion in Introduced *Impatiens walleriana* Populations Depend on Method of Analysis. — *Plos One* 5: e15297.
- Mateo, R. G. et al. 2016. The mossy north: an inverse latitudinal diversity gradient in European bryophytes. — *Scientific Reports* 6: 25546.
- Mateo, R. G. et al. 2012. Do Stacked Species Distribution Models Reflect Altitudinal Diversity Patterns? — *Plos One* 7: e32586.
- Medley, K. A. 2010. Niche shifts during the global invasion of the Asian tiger mosquito, *Aedes albopictus* Skuse (Culicidae), revealed by reciprocal distribution models. — *Glob. Ecol. Biogeogr.* 19: 122-133.
- Morlon, H. et al. 2011. Spatial patterns of phylogenetic diversity. — *Ecol. Lett.* 14: 141-149.
- Naimi, B. and Araújo, M. B. 2016. sdm: a reproducible and extensible R platform for species distribution modelling. — *Ecography* 39: 368-375.
- Ndiribe, C. et al. 2013a. Phylogenetic plant community structure along elevation is lineage specific. — *Ecol. Evol.* 3: 4925-4939.
- Ndiribe, C. et al. 2013b. Plant functional and phylogenetic turnover correlate with climate and land use in the Western Swiss Alps. — *J. Plant. Ecol.* rtt064.
- Nogués-Bravo, D. 2009. Predicting the past distribution of species' climatic niches. — *Global Ecology and Biogeography Letters*

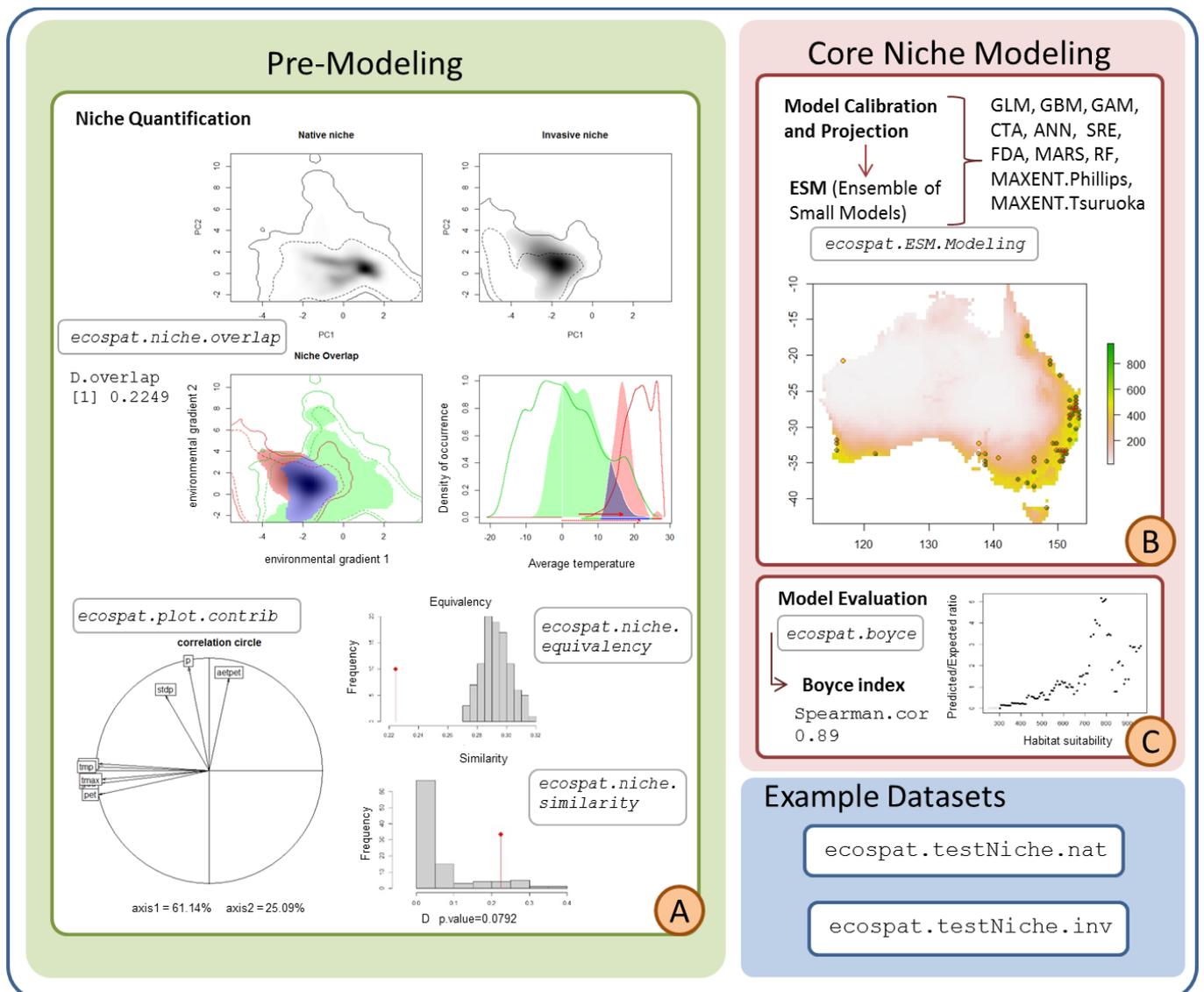
- Paradis, E. et al. 2004. APE: analyses of phylogenetics and evolution in R language. — *Bioinformatics* 20: 289-290.
- Pellissier, L. et al. 2013a. Phylogenetic alpha and beta diversities of butterfly communities correlate with climate in the western Swiss Alps. — *Ecography* 36: 541-550.
- Pellissier, L. et al. 2013b. Turnover of plant lineages shapes herbivore phylogenetic beta diversity along ecological gradients. — *Ecol. Lett.* 16: 600-608.
- Peres-Neto, P. R. et al. 2001. Environmentally constrained null models: site suitability as occupancy criterion. — *Oikos* 93: 110-120.
- Peterson, A. T. et al. 2011. *Ecological Niches and Geographic Distributions*. — Princeton University Press.
- Petitpierre, B. et al. 2012. Climatic Niche Shifts Are Rare Among Terrestrial Plant Invaders. — *Science* 335: 1344-1348.
- Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. — *Ecol. Model.* 190: 231-259.
- Phillips, S. J. and Elith, J. 2010. POC plots: calibrating species distribution models with presence-only data. — *Ecology* 91: 2476-2484.
- Pineda, E. and Lobo, J. M. 2009. Assessing the accuracy of species distribution models to predict amphibian species richness patterns. — *J. Anim. Ecol.* 78: 182-190.
- Pio, D. V. et al. 2011. Spatial Predictions of Phylogenetic Diversity in Conservation Decision Making. — *Conserv. Biol.* 25: 1229-1239.
- Pio, D. V. et al. 2014. Climate change effects on animal and plant phylogenetic diversity in southern Africa. — *Glob. Chang. Biol.* 20: 1538-1549.
- Pottier, J. et al. 2013. The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. — *Glob. Ecol. Biogeogr.* 22: 52-63.
- Pulliam, H. R. 2000. On the relationship between niche and distribution. — *Ecol. Lett.* 3: 349-361.
- Qiao, H. et al. 2016. NicheA: Creating Virtual Species and Ecological Niches in Multivariate Environmental Scenarios. — *Ecography* 39: 805-813.
- Schweiger, O. et al. 2008. A comparative test of phylogenetic diversity indices. — *Oecologia* 157: 485-495.
- Skidmore, A. K. et al. 2011. Geospatial tools address emerging issues in spatial ecology: a review and commentary on the Special Issue. — *Int. J. Geogr. Inf. Sci.* 25: 337-365.
- Soberon, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. — *Ecol. Lett.* 10: 1115-1123.
- Stone, L. and Roberts, A. 1990. The checkerboard score and species distributions. — *Oecologia* 85: 74-79.
- Thuiller, W. 2003. BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change. — *Glob. Chang. Biol.* 9: 1353-1362.
- Thuiller, W. et al. 2016. biomod2: Ensemble Platform for Species Distribution Modeling. R package version 3.3-7. <http://CRAN.Rproject.org/package=biomod2>.
- Thuiller, W. et al. 2009. BIOMOD - a platform for ensemble forecasting of species distributions. — *Ecography* 32: 369-373.
- Tsirogiannis, C. and Sandel, B. 2015. PhyloMeasures: Fast and Exact Algorithms for Computing Phylogenetic Biodiversity Measures. R package version 1.1. <http://CRAN.R-project.org/package=PhyloMeasures>.
- Tsoar, A. et al. 2007. A comparative evaluation of presence-only methods for modelling species distribution. — *Divers. Distrib.* 13: 397-405.
- VanDerWal, J. et al. 2014. SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises. R package version 1.1-20. <http://CRAN.R-project.org/package=SDMTools>.
- Warren, D. L. et al. 2008. Environmental Niche Equivalency Versus Conservatism: Quantitative Approaches to Niche Evolution. — *Evolution* 62: 2868-2883.

- Wintle, B. A. et al. 2012. Designing occupancy surveys and interpreting non-detection when observations are imperfect. — *Divers. Distrib.* 18: 417-424.
- Wisz, M. S. et al. 2008. Effects of sample size on the performance of species distribution models. — *Divers. Distrib.* 14: 763-773.
- Wisz, M. S. et al. 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. — *Biol. Rev.* 88: 15-30.
- Zhang, J. 2013. spaa: SPecies Association Analysis. R package version 0.2.1. <http://CRAN.R-project.org/package=spaa>.

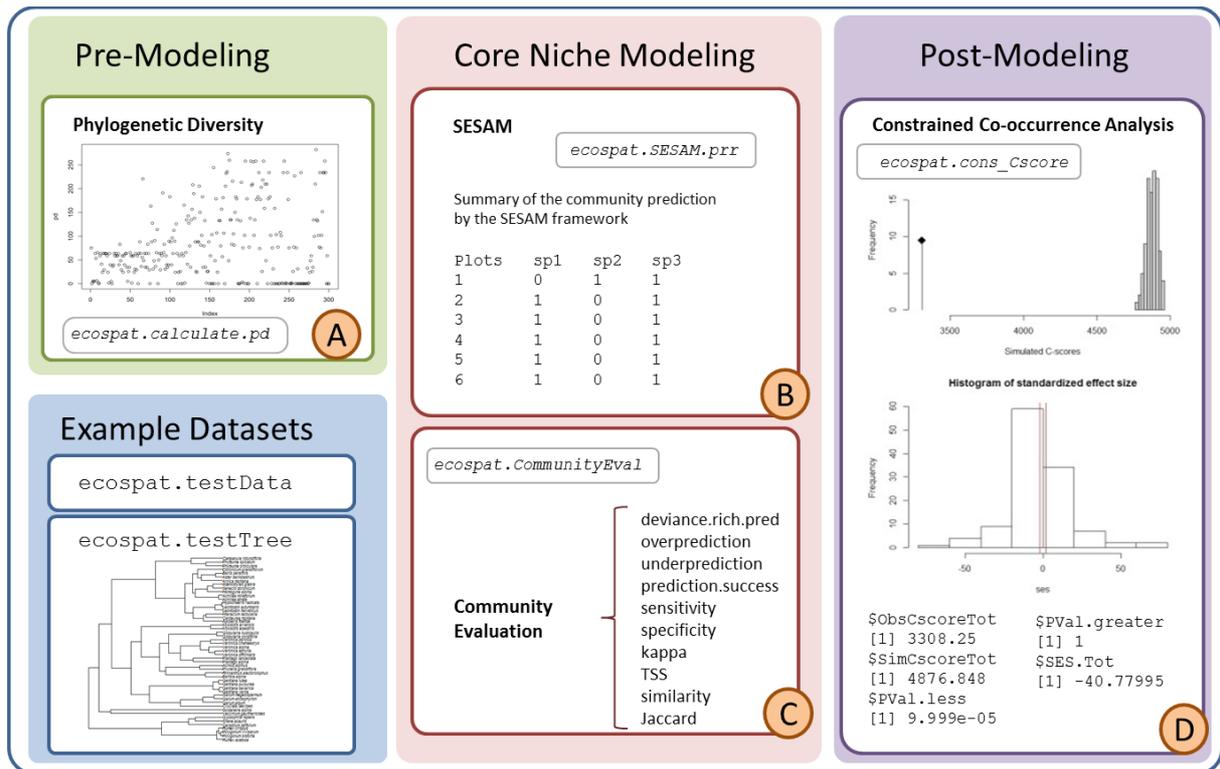
**Table 1.** Main functions, their description and published references of the *ecospat* package.

Tool	Sub-Tool	Function	Description	References where functions has been used
Pre-modelling	Spatial autocorrelation	<i>ecospat.mantel.correlogram</i>	Investigates spatial autocorrelation of environmental covariables within a set of occurrences as a function of distance.	
	Variable selection	<i>ecospat.npred</i>	Calculates the maximum number of predictors to include in the model with a desired correlation between predictors.	Petitpierre et al (2012) Science 335: 1344.
	Extrapolation Detection	<i>ecospat.exdet</i>	Assess climate analogy between a projection extent (proj) and a reference extent (cal, used in general as the background to calibrate SDMs)	
	Multivariate Environmental Similarity Surfaces	<i>ecospat.mess</i>	Calculates the MESS (i.e. extrapolation) as in Maxent (Elith et al. 2010).	Broennimann et al (2014) J. Biogeogr. 41: 1126; Mateo et al (2015) Ecography 38: 480
		<i>ecospat.plot.mess</i>	Plots the MESS extrapolation index onto the geographical space.	
	Phylogenetic diversity measures	<i>ecospat.calculate.pd</i>	Calculates phylogenetic diversity measures (Schweiger et al. 2008).	Pio et al (2011) Conserv. Biol. 25: 1229; Ndiribe et al (2013) Ecol. Evol. 3: 4925; Ndiribe et al (2013) J. Plant Ecol. Rtt064; Pio et al (2014) Glob. Change Biol. 20: 1538.
	Biotic Interactions	<i>ecospat.co-occurrence</i>	Computes an index of co-occurrences ranging from 0 (never co-occurring) to 1 (always co-occurring).	Pellissier et al (2010) Ecography 33: 1004
	Niche quantification	<i>ecospat.grid.clim.dyn</i>	Using the scores of an ordination, create a grid z of RxR pixels (or a vector of R pixels when using scores of dimension 1 or SDM predictions) with occurrence densities. Only scores of one, or two dimensions can be used.	Boulangeat et al (2012) Ecol. Lett. 15: 584; Boulangeat et al (2012) Glob. Chang. Biol. 18: 3464; Estrada-Pena et al (2012) Plos One 7: e36976; Monahan et al (2012) Plos One 7: e42097; Pellissier et al (2012) J. Evol. Biol. 25: 1658; Petitpierre et al (2012) Science 335: 1344; Poudel et al (2012) Plos One 7: e46873; Wielstra et al (2012). Plos One 7: e46671; Ahmadzadeh et al (2013) J. Biogeogr. 40: 1807; Estrada-Pena et al (2013) Exp. Appl. Acarol. 59: 351; Fourcade et al (2013) Biol. Conserv. 167: 161; Di Febbraro et al (2013) Plos One 8: e66559; Alexander (2013) Proc. R. Soc. Lond. (Biol) 280: 20131446; Ahmadzadeh et al (2013) Plos One 8: e80563; Iwamura et al (2013) Glob. Environ. Chang. 23: 1277; Koch et al (2013) Zootaxa 3745: 263; Maiorano et al (2013) Glob. Ecol. Biogeogr. 22: 302; Martinez-Cabrera et al (2013) Plos One 8: e83087; Strubbe et al (2013) Glob. Ecol. Biogeogr. 22: 962; Theodoridis et al (2013) J. Biogeogr. 40: 1278; Werner et al (2013) Ecography 36: 1127; Broennimann et al. (2014) Biol. Lett. 10: 20140638; DeChaine et al (2014) Ecol. Evol. 4: 3940; Early (2014) Glob. Ecol. Biogeogr. 23: 1356; Blonder et al (2014) Glob. Ecol. Biogeogr. 23: 595; Fourcade et al (2014) Plos One 9: e97122; Glennon et al (2014) Ecol. Lett. 17: 574; Goncalves et al (2014) Plos One 9: e111468; Grossenbacher et al (2014) Evolution 68: 1270; Grundy et al (2014) Ibis 156: 355; Le Roux et al (2014) New Phytol. 204: 230; Li et al (2014) Glob. Ecol. Biogeogr. 23: 1094; Montecino et al (2014) Aq. Inv. 9:
		<i>ecospat.niche.equivalency.test</i>	Runs a niche equivalency test (Warren et al 2008) based on two species occurrence density grids.	
		<i>ecospat.niche.similarity.test</i>	Runs a niche similarity test (Warren et al. 2008) based on two species occurrence density grids.	
		<i>ecospat.plot.niche</i>	Plots a niche z created by <i>ecospat.grid.clim.dyn()</i> .	
		<i>ecospat.plot.niche.dyn</i>	Plots niche categories and species density.	
		<i>ecospat.plot.contrib</i>	Plots the contribution of the initial variables to the analysis (i.e. correlation circle). Typically these are the eigen vectors and eigen values in ordinations.	
		<i>ecospat.niche.overlap</i>	Calculates the overlap metrics D and I based on two species occurrence density grids z1 and z2 created by <i>ecospat.grid.clim.dyn()</i> .	
<i>ecospat.plot.overlap.test</i>		Plots a histogram of observed and randomly simulated overlaps, with p-values of equivalency or similarity tests.		
<i>ecospat.niche.dyn.index</i>	Calculates niche expansion, stability and unfilling.			

				507; Pena-Gomez et al (2014) Plos One 9: e105025; Russo et al (2014) Plos One 9: e110894; Saupe et al (2014) Proc R Soc Lond (Biol) 281: 20141995; Silva et al (2014) Plos One 9: e119246; Strubbe et al (2014) Biol. Invasions 16: 759; Tingley
Core Niche Modelling	Model evaluation	<i>ecospat.cv.glm</i>	K-fold and leave-one-out cross validation for GLM.	Randin et al (2006) J. Biogeogr. 33: 1689; Pearman et al (2008) Trends Ecol. Evol. 23: 149; Engler et al (2009) Ecography 35: 872; Patsiou et al (2014) Glob. Chang. Biol. 20: 2286.
		<i>ecospat.permut.glm</i>	A permutation function to get p-values on GLM coefficients and deviance.	
		<i>ecospat.cv.gbm</i>	K-fold and leave-one-out cross validation for GBM.	
		<i>ecospat.cv.me</i>	K-fold and leave-one-out cross validation for Maxent.	
		<i>ecospat.cv.rf</i>	K-fold and leave-one-out cross validation for randomForest.	
		<i>ecospat.boyce</i>	Calculates the Boyce index (Hirzel et al. 2006).	
	<i>ecospat.CommunityEval</i>	Calculates several indices of accuracy of community predictions.	Dubuis et al (2013) J. Veg. Sci. 24:593; Pottier et al (2013) Glob. Ecol. Biogeogr. 22: 52; D'Amen et al (2015). J. Biogeogr. 42:1255.	
	Spatial predictions and projections	<i>ecospat.ESM.Modeling</i>	Calibrates simple bivariate models as in Lomba et al. 2010 and Breiner et al. 2015.	Lomba et al (2010) Biol. Conserv. 143: 2647; Breiner et al (2015) Methods Ecol. Evol. 6: 1210.
		<i>ecospat.ESM.EnsembleModeling</i>	Evaluates and averages simple bivariate models to ESMs.	
		<i>ecospat.ESM.Projection</i>	Projects simple bivariate models into new space or time.	
		<i>ecospat.ESM.EnsembleProjection</i>	Projects calibrated ESMs into new space or time.	
<i>ecospat.SESAM.prr</i>		Implement the SESAM framework to predict community composition using a 'probability ranking' rule.	D'Amen et al (2015) J. Biogeogr. 42: 1255; D'Amen et al (2015) Glob. Ecol. Biogeogr. 24: 1443	
<i>ecospat.migclim</i>		Enables the implementation of species-specific dispersal constraints into projections of species distribution models under environmental change and/or landscape fragmentation scenarios, by calling MigClim functions (Engler et al. 2012).	Bateman et al (2013) Divers. Distrib.19: 1224; Normand et al (2013) Phil. Trans. R. Soc. B 368: 20120479; Gimona et al (2015) Landsc.Ecol. 30: 771; Lurgi et al (2015) Methods Ecol. Evol. 6:247.	
Post Modelling	Variance Partition	<i>ecospat.varpart</i>	Performs variance partitioning for binomial GLM based on the deviance of two groups or predicting variables.	Randin et al (2009) J. Veg. Sci. 20: 996; Pellissier et al (2010) Ecography 33: 1004.
	Spatial predictions of species assemblages	<i>ecospat.cons_Cscore</i>	Tests for non-random patterns of species co-occurrence and calculates the C-score index for the whole community and for each species pair.	D'Amen et al (2015) J. Biogeogr. 42: 1255; D'Amen et al (2015) Glob. Ecol. Biogeogr. 24: 1443



**Figure 1.** A schematic representation of the workflow used in Example 1. The grey boxes indicate the names of the functions. (A) Plot of the occupancy of the environment by the species in the native and invaded range. Niche overlap values for native and invaded range, in terms of Schoener's D, niche similarity p-values and equivalency p-values via randomization. Niche equivalency test and falling outside the 95% confidence interval of the niche similarity test. (B) Model calibration and projection with Ensemble of Small Models (ESM). (C) Plot of the predicted to expected ratio along the suitability classes.



**Figure 2.** A schematic representation of the workflow used in Example 2. The grey boxes indicate the names of the functions. (A) Correlation of phylogenetic diversity (pd) with species richness (index). (B) Community prediction by the SESAM framework (i.e. binary predictions for all species for each site). (C) List of evaluation metrics calculated for each site. (D) Frequency histogram for simulated C-scores in the null community and frequency histogram for standardized effect size in the observed community: number of species pairs forming perfect checkerboard distributions.