

# The ecospat R package: a collection of pre-, core- and post-modeling tools to investigate species niches and distributions

Olivier Broennimann<sup>1,2</sup>, Flavien Collart<sup>1</sup> and Antoine Guisan<sup>1,2</sup>

<sup>1</sup>Department of Ecology and Evolution, University of Lausanne, Switzerland

<sup>2</sup>Institute of Earth Surface Dynamics, University of Lausanne, Switzerland

## Contents

|   |           |
|---|-----------|
| <b>1. Introduction</b>                          | <b>1</b>  |
| <b>2. Case studies</b>                          | <b>4</b>  |
| a. Modeling rare species . . . . .              | 4         |
| i. Pre-Modeling . . . . .                       | 5         |
| ii. Core-Modeling . . . . .                     | 5         |
| iii. Post-modeling . . . . .                    | 12        |
| b. Niche dynamics of invasive species . . . . . | 16        |
| i. Data preparation . . . . .                   | 16        |
| ii. Niche quantification . . . . .              | 19        |
| <b>3. Discussion and perspectives</b>           | <b>28</b> |
| <b>4. References</b>                            | <b>29</b> |

## 1. Introduction

Species distribution modeling (SDMs; Guisan et al. (2017); Peterson et al. (2011); Franklin (2010)) is a field of ecology and biogeography that has grown significantly in the last three decades (Araújo et al., 2019), with important potential conservation applications (Guisan et al., 2013). SDMs work by relating statistically (or sometime in an expert-based way, Di Febbraro et al. (2018); Fourcade et al. (2013)) species observations (simple occurrences, presence-absences, or abundances) to a set of environmental attributes (e.g. climate, topography, substrate, landscape, human influences) characterizing the sampling sites. By doing this through a set of locations in the studied region, SDMs implicitly quantify the suitability of the environment for the species (i.e. habitat or environmental suitability), and if done across the whole species' environmental range, it then quantifies the species' realized environmental niche (Araújo & Guisan, 2006; Austin et al., 1990; Pulliam, 2000; Soberón, 2007), which can be used to make geographic projections, e.g. in the future, provided some assumptions are met (e.g. equilibrium, niche stability, no niche truncation, etc; see Guisan et al. (2017); Zurell et al. (2020)). Individual SDMs can then be used for many applications,

ranging from climate change assessments (e.g. Broennimann et al. (2006); Patiño et al. (2023)), anticipating biological invasions (e.g. Broennimann et al. (2007); Barbet-Massin et al. (2018)), determining suitable areas for translocation (Ferrarini et al., 2016), or rare species management (Guisan et al., 2006; Jeliaskov et al., 2022). Individual species predictions can also be stacked across a whole taxonomic group (Dubuis et al., 2011; Kass et al., 2022) to predict community patterns, like richness, composition, or functional or phylogenetic diversity (D’Amen et al., 2015; Ferrier & Guisan, 2006; Guisan & Rahbek, 2011; Shen et al., 2023), and these multiple species predictions can be then be used to support spatial conservation planning (Moilanen et al., 2022). All these analyses and applications can now be performed thanks to numerous software packages, most notably in the R framework (Sillero et al., 2023).

One R package to perform SDMs is *ecospat* (Di Cola et al. (2017); Broennimann et al. (2023), currently version 4.0.0), a collection of R functions and data sets designed to support spatial ecology analyses, with a focus on pre-, core- and post-modeling analyses of species distributions (SDMs) (Table 1). Functions and tools within *ecospat* can be classified according to the three main group of analyses that are typically followed when building SDMs (Di Cola et al., 2017): (i) pre-modeling analyses, consisting of data preparation and exploratory analyses; (ii) core-modeling analyses, corresponding to model fitting and evaluation, and (iii) post-modeling analyses, which consist of using the models to reach study goals such as combining single species predictions to reconstruct communities (stacked SDMs; Guisan & Rahbek (2011), D’Amen, Mod, et al. (2018), Zurell et al. (2020)).

In *ecospat*, pre-modeling analyses include data extractions, designing sampling strategies, sub-sampling data, assessing spatial autocorrelation, reducing collinearity among predictors, measuring environmental similarity and analog/non-analog situations (mess), assessing niche stability within a species or niche difference among species (using a large set of functions, a key development in *ecospat*), and measuring species co-occurrences to investigate community assembly. Core-modeling analyses typically cover modeling fitting, spatial predictions and projections, and model evaluation. In *ecospat*, it includes especially fitting SDMs with the ‘Ensemble of Small Models’ approach (ESMs), that applies especially to species with small sample sizes (e.g. rare and endangered species) but also performs well for other species, and applying the different steps of the SESAM framework (Guisan & Rahbek, 2011) for modeling the spatial distribution of species assemblages and communities. It also includes evaluations of individual species models through the computation of cross-validations and evaluation metrics (such as the Boyce Index). The post-modelling analyses comprise variance partitioning, determination of occupied patches and range size using the IUCN criteria, SDM prediction map binarization, and predictions of community properties from stacking individual species (e.g. species richness; Dubuis et al. (2011)), including also functional and phylogenetic diversity (D’Amen, Mateo, et al., 2018).

In this book section, we first illustrate two study cases where the *ecospat* package can be used: (a) modeling rare species distribution and quantifying species range size and (b) anticipating biological invasions via niche comparison. We then discuss other applications of the functions in the *ecospat* package and provide some perspectives about the future development of the package.

Table 1 - List of the functions in the *ecospat* package (adapted from Di Cola et al. 2017)

| Category            | Sub category                 | Function                         | Description  |
|---------------------|------------------------------|----------------------------------|--|
| Pre-modeling        | Terrain sampling preparation | ecospat.rcls.grd                 | Reclassifies grid files to get a combined stratification from more than one grid   |
|                     |                              | ecospat.recstrat_prop            | Performs a Random Ecologically Stratified Sampling proportional  |
|                     |                              | ecospat.recstrat_regl            | Performs a Random Ecologically Stratified Sampling equal   |
|                     | Spatial autocorrelation      | ecospat.mantel.correlogram       | Investigates spatial autocorrelation of environmental covariables for occurrences  |
|                     | Variable selection           | ecospat.npred                    | Calculates the maximum number of predictors to include in the model with a desired correlation between predictors.   |
|                     |                              | ecospat.cor.plot                 | Plots correlations among predictors  |
|                     | Extrapolation detection      | ecospat.mess                     | Calculates the MESS (i.e. extrapolation) as in Maxent (Elith et al. 2010)  |
|                     |                              | ecospat.plot.mess                | Plots the MESS extrapolation index onto the geographical space.  |
|                     |                              | ecospat.climan                   | Assesses the climate analogy in a univariate and multivariate environmental space  |
|                     | Niche quantification         | ecospat.grid.clim.dyn            | Creates a grid z of Rx R pixels (or a vector of R pixels when using scores of dimension 1 or SDM predictions) with occurrence densities, using the scores of two axes from an ordination |
|                     |                              | ecospat.niche.overlap            | Calculates the overlap metrics D and I based on two species occurrence density grids z1 and z2 created by ecospat.grid.clim.dyn().   |
|                     |                              | ecospat.niche.equivalency.test   | Runs a niche equivalency test (warren et al. 2008) based on two species occurrence density grids.  |
|                     |                              | ecospat.niche.similarity.test    | Runs a niche similarity test (warren et al. 2008) based on two species occurrence density grids.   |
|                     |                              | ecospat.plot.niche               | Plots a niche z created by ecospat.grid.clim.dyn.  |
|                     |                              | ecospat.plot.niche.dyn           | Plots niche categories and species density.  |
|                     |                              | ecospat.plot.contrib             | 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.                        |
|                     |                              | ecospat.plot.overlap.test        | Plots a histogram of observed and randomly simulated overlaps, with p-values of equivalency or similarity tests.   |
|                     |                              | ecospat.niche.dyn.index          | Calculates niche expansion, stability and unfilling.   |
|                     |                              | ecospat.shift.centroids          | Draws arrows linking the centroid of the native and exotic (non-native) distribution (continuous line) and between native and invaded extent (dashed line).                              |
|                     |                              | ecospat.margin                   | Delineates the distribution's margin and its uncertainty   |
|                     |                              | ecospat.niche.dynIndexProjGeo    | Creates a raster in geography with each pixel containing a niche dynamic index   |
|                     |                              | ecospat.niche.zProjGeo           | Projects Occurrence Densities to the Geography   |
|                     |                              | ecospat.nicheNBmean              | Calculates the weighted mean niche breadth across several axes   |
|                     |                              | ecospat.nichePOSNB               | Calculates the niche breadth and niche position of taxa  |
|                     | Data preparation             | ecospat.sample.envar             | Extracts data from environmental table   |
|                     |                              | ecospat.caleval                  | Generates an evaluation and calibration dataset with a desired ratio of disaggregation.  |
|                     |                              | ecospat.CCV.createDataSplitTable | Creates a DataSplitTable for Community modeling  |
|                     |                              | ecospat.occ.desaggregation       | Removes species occurrences in a dataframe which are closer to each other than a specified distance threshold.   |
|                     |                              | ecospat.rand.pseudoabsences      | Randomly samples pseudo-absences from an environmental data frame covering   |
|                     |                              |                                  |  |
| Core niche modeling | Modeling of rare species     | ecospat.ESM.Modeling             | Calibrates simple bivariate models as in Lomba et al. 2010; Breiner et al. 2015  |
|                     |                              | ecospat.ESM.EnsembleModeling     | Evaluates and averages simple bivariate models to ESMs.  |
|                     |                              | ecospat.ESM.Projection           | Projects simple bivariate models into new space or time.   |
|                     |                              | ecospat.ESM.EnsembleProjection   | Projects calibrated ESMs into new space or time.   |
|                     |                              | ecospat.ESM.responsePlot         | Plots species response curve for Ensemble of Small Models  |
|                     |                              | ecospat.ESM.threshold            | Computes Thresholds to binarize Ensemble of Small Models   |
|                     |                              | ecospat.ESM.VarContrib           | Computes variables' contribution for Ensemble of Small Models  |
|                     |                              | ecospat.Epred                    | Calculates the weighted mean of several predictions in dataframe format (Ensemble Models)  |
|                     |                              | ecospat.maxentvarimport          | Computes variable importance for MAXENT models   |

|                  |                                 |                                      |  |
|------------------|---------------------------------|--------------------------------------|--|
|                  | Community spatial predictions   | ecospat.SESAM.prr                    | Implements the SESAM framework to predict community composition using a 'probability ranking' rule.  |
|                  |                                 | ecospat.CCV.modeling                 | Creates probabilistic prediction for all species based on SDMs or ESMs   |
| Model evaluation |                                 | ecospat.cv.glm                       | Performs a K-fold and leave-one-out cross validation for GLM.  |
|                  |                                 | ecospat.permut.glm                   | Performs permutations to get p-values on GLM coefficients and deviance.  |
|                  |                                 | ecospat.cv.gbm                       | Performs a K-fold and leave-one-out cross validation for GBM.  |
|                  |                                 | ecospat.cv.me                        | Performs a K-fold and leave-one-out cross validation for Maxent.   |
|                  |                                 | ecospat.cv.rf                        | Performs a K-fold and leave-one-out cross validation for randomForest  |
|                  |                                 | ecospat.boyce                        | Calculates the Boyce index (hizel et al 2006)  |
|                  |                                 | ecospat.CommunityEval                | Calculates several indices of accuracy of community predictions.   |
|                  |                                 | ecospat.adj.D2.glm                   | Computes the D2 metric in GLM  |
|                  |                                 | ecospat.CCV.communityEvaluation.bin  | Calculates a range of community evaluation metrics   |
|                  |                                 | ecospat.CCV.communityEvaluation.prob | Generates a number of community evaluation metrics directly based on the probability   |
|                  |                                 | ecospat.cohen.kappa                  | Calculates Kappa based on a given threshold  |
|                  |                                 | ecospat.ESM.EnsembleEvaluation       | Evaluates ensemble of small models via the pooling procedure (Collart et al. 2023)   |
|                  |                                 | ecospat.max.kappa                    | Computes the Maximum value of Kappa  |
|                  |                                 | ecospat.max.tss                      | Computes the Maximum value of the True Skill Statistic   |
|                  |                                 | ecospat.meva.table                   | Calculates evaluation metrics based on a given threshold   |
| Post modeling    | Variance partition              | ecospat.plot.kappa                   | Plots kappa value in function of the threshold value.  |
|                  |                                 | ecospat.plot.tss                     | plot TSS values in function of the threshold value   |
|                  |                                 | ecospat.poolingEvaluation            | Uses the pooling evaluation to evaluate any models   |
|                  |                                 | ecospat.varpart                      | Performs variance partitioning for binomial GLM based on the deviance of two groups or predicting variables.                                   |
|                  | Conservation                    | ecospat.rangesize                    | Quantifies the range size of a species using standard IUCN criteria  |
|                  |                                 | ecospat.occupied.patch               | Determines the occupied patch of a species using standard IUCN criteria (AOO, EOO) or predictive binary maps from Species Distribution Models. |
|                  | Other                           | ecospat.binary.model                 | Binarizes raster based on threshold  |
|                  |                                 | ecospat.mpa                          | Calculates the minimum predicted area  |
|                  | Co-occurrence analysis          | ecospat.co-occurrence                | Computes an index of co-occurrences ranging from 0 (never co-occurring) to 1 (always co-occurring).  |
|                  |                                 | ecospat.cons_Cscore                  | Tests for non-random patterns of species co-occurrence and calculates the C-score index for the whole community and for each species pair.     |
| Other functions  | Phylogenetic diversity measures | ecospat.Cscore                       | Computes and tests if C-scores higher than null model  |
|                  |                                 | ecospat.calculate.pd                 | Calculates phylogenetic diversity measures (Schweiger et al. 2008)   |

## 2. Case studies

### a. Modeling rare species

In this section, we will use the ecospat package to employ the method called “ensemble of small models (ESM)”, which was developed by Lomba et al. (2010); Breiner et al. (2015); Breiner et al. (2018) and is particularly suitable for species with small sample size, like rare and endangered species (yet also works particularly well for species with larger sample sizes but at higher computational cost than more traditional approaches; Breiner et al. (2015)). One of the major problems to model rare species is that the number of occurrences is usually scarce. Although some studies reported that species could be accurately modeled with very low sample size (e.g. 3 occurrences in van Proosdij et al. (2015)), sample size is problematic for the modeling procedure, with the risk of overfitting models when the number of occurrences is low compared to the number of predictors (Pol et al., 2016) and problem of evaluation in the case of test data with small sample sizes Collart & Guisan (2023). In general, authors are limiting the number of predictors that are put in a model using the rule of thumb that not more than one predictor term should be used per 10 occurrences. To avoid this limitation, ESMs compute bivariate models and then combine all possible bivariate models into a weighted ensemble. By averaging simple small models to an ensemble, with their respective cross-validated performance as weights, ESMs avoid overfitting without losing explanatory power through reducing the number of predictor variables, and were shown to perform significantly better than standard SDMs with

species having a low number of occurrences (Breiner et al., 2015). For this section, we will be focusing on modeling the ecological niche of the alpine plant *Veronica alpina* in the Western Swiss Alps.

## i. Pre-Modeling

The ESM functions of the *ecospat* package rely on the model fitting functions in the *biomod2* package (Thuiller et al., 2023). We thus need to first format our data by using the *BIOMOD\_FormatingData* function, where species occurrences and associated coordinates, the environmental conditions and the name of the species of interest are given. In this example we use presence-absence data for 300 plots, where *Veronica alpina* is present in 12 locations, and maps of 5 environmental predictors. The latter variables are: growing degree days above 0°C, moisture index over the growing season (average values for June to August in mm day<sup>-1</sup>), annual sum of radiation (in kJ m<sup>-2</sup> year<sup>-1</sup>), slope (in degrees), and topographic position.

```
# Load the packages
```

```
library(ecospat)
library(biomod2)
library(terra)
library(viridis)
```

```
set.seed(123)
```

```
data("ecospat.testData")
```

```
# coordinates of the plots
```

```
xy <- ecospat.testData[,2:3]
```

```
# species presences and absences
```

```
sp_occ <- ecospat.testData$Veronica_alpina
```

```
sum(sp_occ) ## Number of occurrences
```

```
[1] 12
```

```
# environmental data
```

```
env <- rast(system.file("extdata/ecospat.testEnv.tif", package="ecospat"))
```

```
# Formatting the data with the BIOMOD_FormatingData() function from the package biomod2
```

```
library(biomod2)
```

```
myBiomodData <- BIOMOD_FormatingData(resp.var = as.numeric(sp_occ),
                                     expl.var = env,
                                     resp.xy = xy,
                                     resp.name = "Veronica.Alpina",
                                     filter.raster = TRUE)
```

## ii. Core-Modeling

The function *ecospat.ESM.Modeling* is used to model the ecological niche of the species by generating bivariate models. The argument *data* is for the formatted dataset object generated by *BIOMOD\_FormatingData*. The desired algorithms can be provided in the argument *models*. Model parameters can be adapted via the argument *models.options* by giving the object from the function *BIOMOD\_ModellingOptions()* of the *biomod2* package. As in the package *biomod2*, ESM can fit 12 different algorithms: Generalized Linear Model ('GLM'), Gradient Boosted Machine ('GBM'), eXtreme Gradient Boosting Training (XGBOOST), Generalized Additive Models ('GAM'), 'CTA', Artificial Neural Network ('ANN'), 'SRE', 'FDA', 'MARS', 'RF', Maximum entropy ('MAXENT', using the java software or 'MAXNET' from the *maxnet* package).

Tuning to obtain the optimal parameters for the model can be realized with the argument *tune*. *Prevalence* can be set to build a “weighted response”. If NULL, each observation (presence or absence) will have the same weight. You can also give a specific weight to observations via the argument *Yweights*. To evaluate the models, the function performs a repeated split-sampling cross-validation using the arguments *DataSplit* and *NbRunEval*. *DataSplit* corresponds to the percentage of observations used to calibrate the models. *NbRunEval* indicates the number of times the split-sampling procedure is replicated. The function also allows user-defined cross-validations by giving a logical matrix in the argument *DataSplitTable*, where each row corresponds to an observation and each column corresponds to a run. A value TRUE means that an observation will be used for model calibration while a FALSE is for model evaluation. *weighting.score* corresponds to the evaluation metric that will be used to weight single bivariate models in the final ensemble model. The available evaluation metrics are: ‘AUC’, ‘SomersD’ (2xAUC-1), ‘Kappa’, ‘TSS’ or ‘Boyce’. *which.biva* allows to split the bivariate model procedure in several parts. For example, if *which.biva* is 1:3, only the three first variable combinations will be modeled. This allows to run different bivariate splits on different computers. However, it is better not to use this option if all models are run on a single computer. If you do so, make sure to give each of your modeling subset a unique *modeling.id*. and avoid space characters. Parallel computing can be enabled with the argument *parallel*.

```
my.ESM <-ecospat.ESM.Modeling(data = myBiomodData,
                             NbRunEval = 3,
                             DataSplit = 70,
                             DataSplitTable = NULL,
                             Prevalence = 0.5,
                             weighting.score = "SomersD",
                             models = "GLM",
                             tune = FALSE,
                             modeling.id = as.character(format(Sys.time(), "%s")),
                             models.options = NULL,
                             which.biva = NULL,
                             parallel = FALSE,
                             cleanup = FALSE,
                             Yweights = NULL)
```

The following step is to combine all the bivariate models into an ensemble. To so, we can use the function *ecospat.ESM.EnsembleModeling* which will need the object returned by *ecospat.ESM.Modeling*, the evaluation metric used to weight the bivariate models (*weighting.score*) and a *threshold* to remove poor performing models. The argument *models* allows to select one or several algorithms to realize the ensemble.

```
my.ESM.EF <- ecospat.ESM.EnsembleModeling(ESM.modeling.output = my.ESM,
                                           weighting.score = "SomersD",
                                           threshold = 0,
                                           models = NULL)
```

ESM performances resulting from the cross-validations can be observed in the object returned by *ecospat.ESM.EnsembleModeling*.

```
library(knitr)
title<-"ESM performances based on a mean or standard deviations across bivariate model performances of "
kable(t(my.ESM.EF$ESM.evaluations), caption = title, digits = 3, longtable = TRUE)
```

Table 1: ESM performances based on a mean or standard deviations across bivariate model performances of a same run

| model          | RUN1_GLM | RUN2_GLM | RUN3_GLM |
|----------------|----------|----------|----------|
| threshold      | 175      | 525      | 570      |
| sensitivity    | 1        | 1        | 1        |
| specificity    | 0.671    | 0.847    | 0.776    |
| Kappa          | 0.155    | 0.332    | 0.238    |
| AUC            | 0.771    | 0.926    | 0.859    |
| sensitivity.sd | 0        | 0        | 0        |
| specificity.sd | 0.051    | 0.039    | 0.045    |
| Kappa.sd       | 0.071    | 0.127    | 0.100    |
| AUC.sd         | 0.059    | 0.036    | 0.056    |
| TSS            | 0.671    | 0.847    | 0.776    |
| SomersD        | 0.541    | 0.853    | 0.718    |
| MPA            | 0.234    | 0.616    | 0.612    |
| Boyce          | -0.138   | 0.474    | -0.429   |
| technique      | GLM      | GLM      | GLM      |
| RUN            | RUN1     | RUN2     | RUN3     |

However, because a minimum sample size is needed to evaluate models (see Jiménez-Valverde (2020)), it is recommended to evaluate ESMs using the recently proposed pooling evaluation approach (Collart & Guisan, 2023). The function *ecospat.ESM.EnsembleEvaluation* uses this approach, which consists of pooling the suitability values predicted with the hold-out data (evaluation dataset) across replicates to measure an overall evaluation score. As the same observation (presence or absence or background point) is presumably sampled in several replicates, the suitability values for each data point are consequently averaged across replicates where they were sampled. This procedure generates a series of independent suitability values with a size approximately equal to that of the number of observations (the number of suitability values might be slightly lower than the number of original observations as some data points may not be sampled by chance in any of the  $n$  replicates), which agreement to the actual observations can then be measured using sufficient sample size. This function can compute several metrics, which can be selected with the argument *metrics*. If needed, *EachSmallModels* allows evaluating each bivariate model via the pooling evaluation.

```
my.ESM.EF.eval <- ecospat.ESM.EnsembleEvaluation(
  ESM.modeling.output = my.ESM,
  ESM.EnsembleModeling.output =
  my.ESM.EF, metrics = c("SomersD", "AUC", "MaxTSS", "Boyce"),
  EachSmallModels = FALSE)

## Evaluation dataset obtained by the pooling evaluation
pred.test <- as.data.frame(my.ESM.EF.eval$ESM.fit)

## Evaluation scores of the ESM based on the pooling evaluation
eval<-my.ESM.EF.eval$ESM.evaluations

library(knitr)
kable(eval, caption = "ESM performances based on the pooling evaluation", digits = 3)
```

Table 2: ESM performances based on the pooling evaluation

|     | AUC   | SomersD | Boyce | MaxTSS |
|-----|-------|---------|-------|--------|
| GLM | 0.847 | 0.694   | 0.73  | 0.612  |

The *ecospat* package also has numerous functions to compute model performances based on model predictions and species observation. For example, the Boyce index that only requires species presences for evaluating predictions (Hirzel et al., 2006) can be calculated with the function *ecospat.boyce*. The argument *obs* should contain the model prediction for the presences while *fit* should contain the predictions of background points. Correlation measurement can be changed via the argument *method*

```
boyce.index <- ecospat.boyce(fit = pred.test$Fit_GLM,
                             obs = pred.test$Fit_GLM[pred.test$res==1],
                             PEplot = FALSE,
                             method = "spearman")
boyce.index$cor
```

[1] 0.73

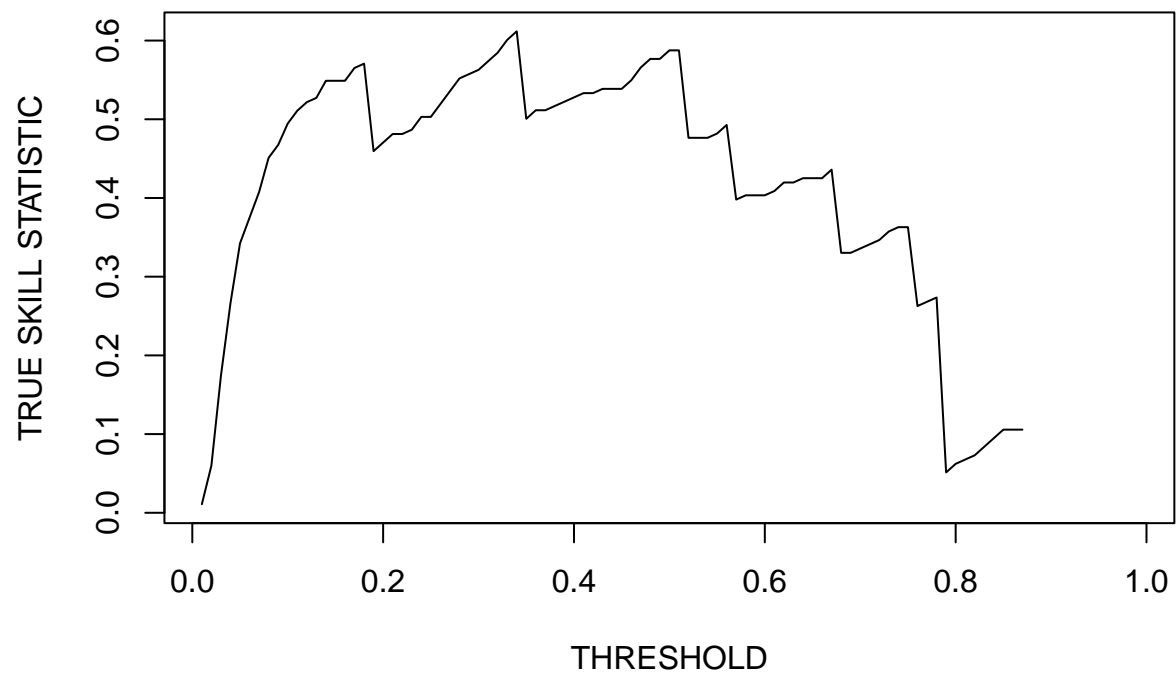
MaxTSS and MaxKappa can be calculated via the functions *ecospat.max.tss* and *ecospat.max.kappa* and the variations of TSS and Kappa metric on a threshold can be done with *ecospat.plot.tss* and *ecospat.plot.kappa*

```
MaxTSS <- ecospat.max.tss(Pred = pred.test$Fit_GLM, Sp.occ = pred.test$res)
MaxTSS$max.TSS
```

[1] 0.611715

```
ecospat.plot.tss(Pred = pred.test$Fit_GLM, Sp.occ = pred.test$res)
```

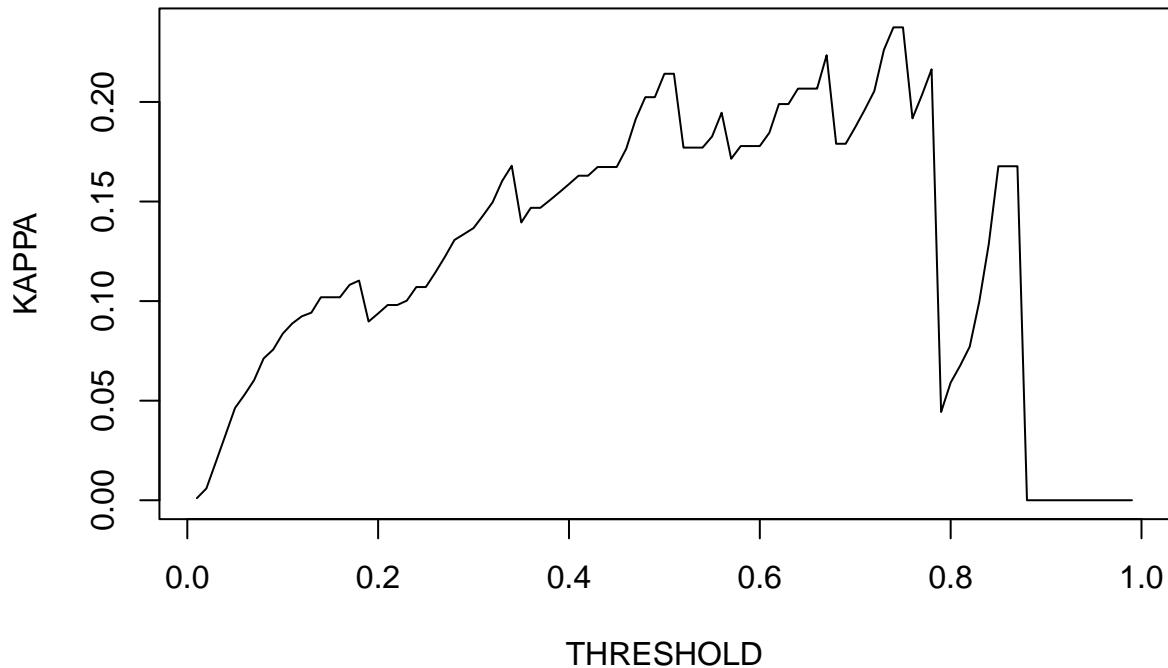




```
MaxKappa <- ecospat.max.kappa(Pred = pred.test$Fit_GLM, Sp.occ = pred.test$resp)
MaxKappa$max.Kappa
```

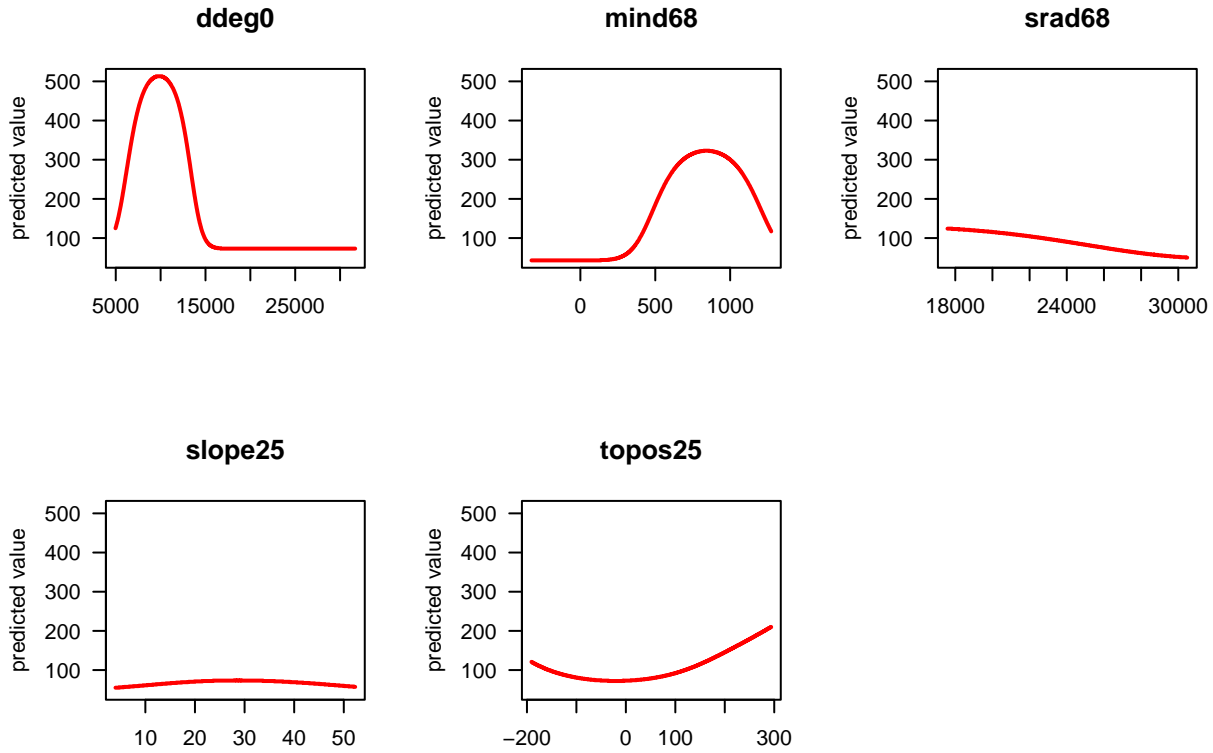
```
[1] 0.2496419
```

```
ecospat.plot.kappa(Pred = pred.test$Fit_GLM, Sp.occ = pred.test$resp)
```



Model adequacy can also be checked by observing species response curves to each environmental predictors and assess whether they make ecological sense. To do so, The function `ecospat.ESM.responsePlot` can be used. This function is an adaptation of the Evaluation Strip method proposed by Elith et al. (2005) and needs the objects returned by `ecospat.ESM.Modeling` and `ecospat.ESM.EnsembleModeling`. The statistic used generate the response curve for a single predictor while keeping all others constant can be changed via the argument `fixed.var.metric`. When looking at the curves produced with this approach, we can see that two variables seem important for our studied species. To go deeper into this, we can also compute their variable importance.

```
response.plots <- ecospat.ESM.responsePlot(
  ESM.EnsembleModeling.output = my.ESM.EF,
  ESM.modeling.output = my.ESM,
  fixed.var.metric = 'median')
```



To measure the contribution of each variable, the function `ecospat.ESM.VarContrib` can be employed. This function computes the ratio between the sum of weights of bivariate models where a focal variable was used and the sum of weights of bivariate models where the focal variable was not used. The ratio is corrected for the number of models with or without the focal variable. This ratio gives an indication on the proportional contribution of the variable in the final ensemble model (Breiner et al., 2015). A value of higher than 1 indicates that the focal variable has a higher contribution than average. For the ensemble model, a weighted mean is applied among model algorithms. In this example, we can see that only two variables seem useful to quantify the ecological niche of *Veronica alpina*.

```
var.contrib <- ecospat.ESM.VarContrib(ESM.modeling.output = my.ESM,
                                     ESM_EF.output = my.ESM.EF)

library(knitr)
kable(var.contrib, caption = "Variable contributions to ESMs", digits = 3)
```

Table 3: Variable contributions to ESMs

|         | GLM   |
|---------|-------|
| ddeg0   | 1.596 |
| mind68  | 1.326 |
| srad68  | 0.873 |
| slope25 | 0.563 |
| topos25 | 0.898 |

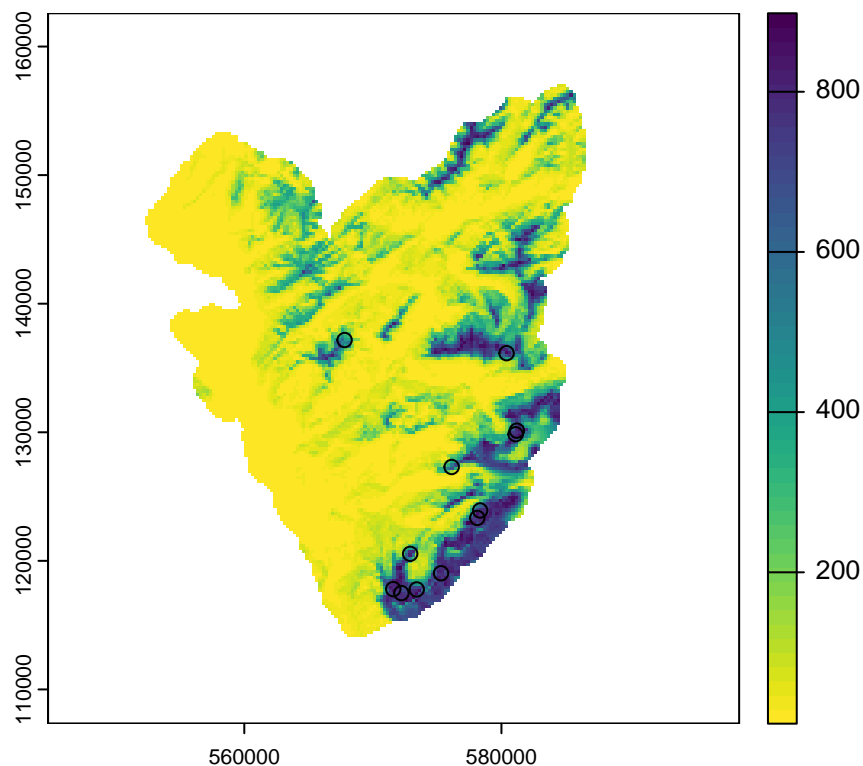
After looking at the model performances and response curves, models can be projected using two functions: `ecospat.ESM.Projection` that projects each bivariate model, and `ecospat.ESM.EnsembleProjection` that gen-

erates the ensemble of these bivariate models. The *new.env* argument allows the use of a *SpatRaster* object to project models onto a new area or time period. The argument *name.env* gives a name to the projection. Parallel computing can be enabled with the argument *parallel*. The generated maps correspond to the rounded values of habitat suitability times 1000.

```
### Projection of simple bivariate models into new space
my.ESM.proj.current <- ecospat.ESM.Projection(ESM.modeling.output = my.ESM,
                                              new.env = env,
                                              name.env = "currentTime",
                                              parallel = FALSE,
                                              cleanup = FALSE)

### Projection of calibrated ESMs into new space
my.ESM.EF.proj.current <- ecospat.ESM.EnsembleProjection(
  ESM.prediction.output = my.ESM.proj.current,
  ESM.EnsembleModeling.output = my.ESM.EF,
  chosen.models = "all")

### projected ESM of Veronica alpina at present time
plot(my.ESM.EF.proj.current, col = rev(viridis::viridis(50)))
points(xy[sp_occ==1,], col = "black")
```



### iii. Post-modeling

ESM projections can be afterwards binarized (i.e. transforming the continuous suitability index into presence-absence). To binarize these maps, diverse thresholds can be computed via the function *ecospat.ESM.threshold*.

This function also provides evaluation scores for the full model (thus, evaluating the fit of the model but not its predictive power and transferability).

```
library(knitr)

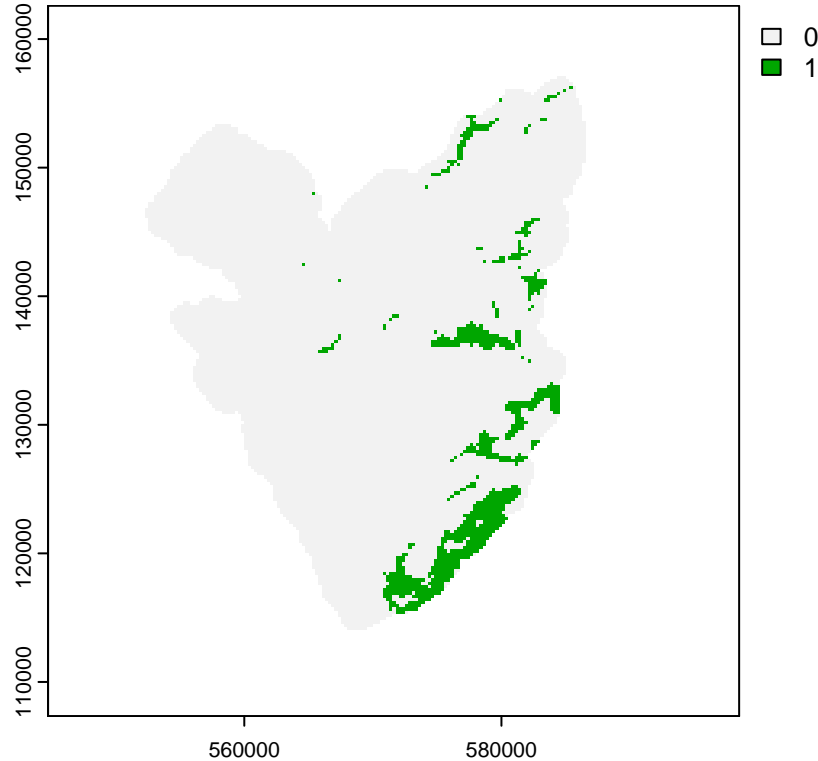
Thr <- ecospat.ESM.threshold(ESM.EnsembleModeling.output = my.ESM.EF, PEplot = FALSE)
obj <- t(Thr[,-1])
colnames(obj) = Thr[1,1]
title<-"Various threshold and fit performances of ESM"
kable(obj, caption = title, digits = 3, longtable = TRUE)
```

Table 4: Various threshold and fit performances of ESM

|                | Full_GLM_ESM |
|----------------|--------------|
| sensitivity    | 0.917        |
| specificity    | 0.845        |
| Kappa          | 0.280        |
| AUC            | 0.922        |
| sensitivity.sd | 0.083        |
| specificity.sd | 0.022        |
| Kappa.sd       | 0.070        |
| AUC.sd         | 0.023        |
| SomersD        | 0.843        |
| Boyce          | NA           |
| TSS            | 0.761        |
| TSS.th         | 0.645        |
| MPA1.0         | 0.386        |
| MPA0.95        | 0.737        |
| MPA0.90        | 0.776        |
| Boyce.th.min   | 0.340        |
| Boyce.th.max   | 0.722        |

Model projections can be afterwards binarized with the function *ecospat.binary.model* which need in the arguments *Pred*, a spatial grid and *Threshold*, the value of the threshold.

```
### Binarization of the projected ESM based on the threshold maximizing the TSS
my.ESM.EF.proj.current.bin <- ecospat.binary.model(Pred = my.ESM.EF.proj.current,
                                                    Threshold = (Thr$TSS.th)*1000)
plot(my.ESM.EF.proj.current.bin)
```



After binarizing the maps, we could quantify the species range size or the occupied patches from ESM maps and IUCN criteria. In the *ecospat* package, the function *ecospat.rangesize* and *ecospat.occupied.patch* are made for these purposes.

More precisely, *ecospat.rangesize* allows quantifying the Area of Occupancy (AOO) and the Extent of Occurrence (EOO) criteria (IUCN (2012)). Numerous parameters are available and described in the help file.

```
rangesize <- ecospat.rangesize(my.ESM.EF.proj.current.bin,
                              xy = xy[sp_occ==1,],
                              AOO.circles = TRUE,
                              lonlat = FALSE)
```

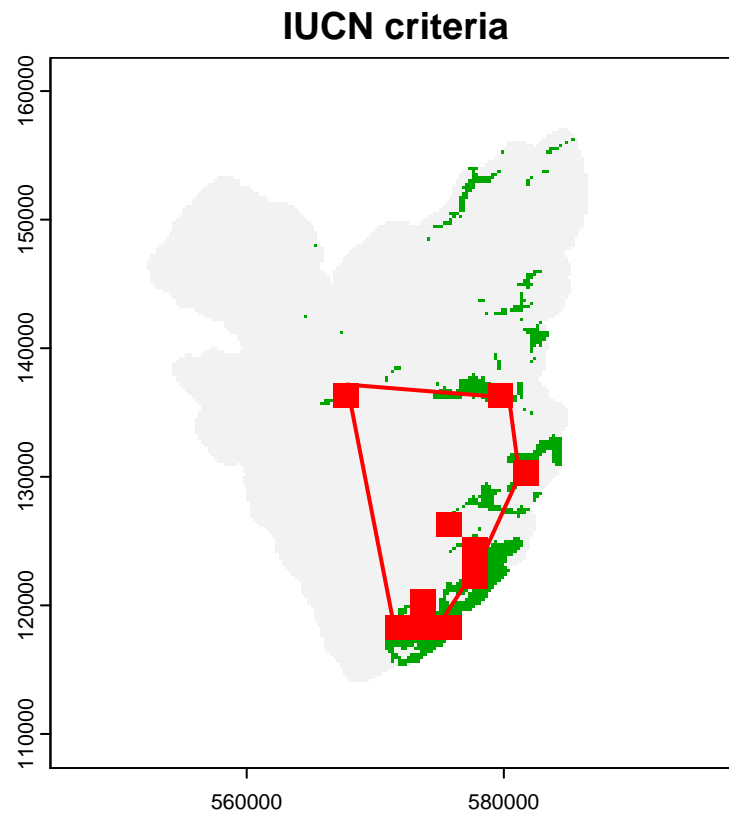
```
A00<-rangesize$RangeSize$A00
A00/1000^2 #area in km2
```

```
## [1] 40
```

```
E00<-rangesize$RangeSize$E00
E00/1000^2 #area in km2
```

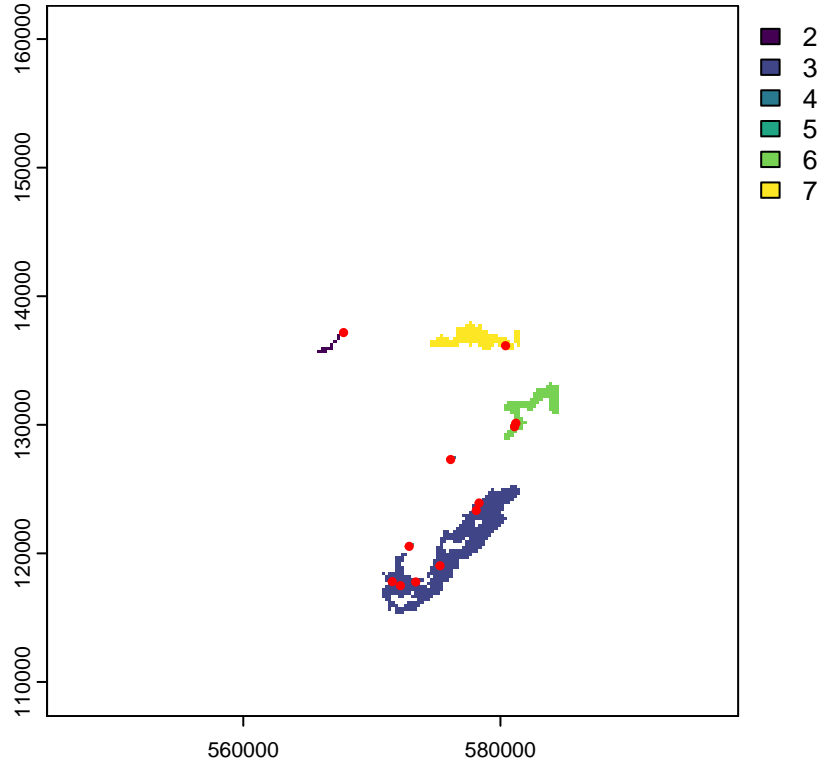
```
## [1] 175.4488
```

```
plot(my.ESM.EF.proj.current.bin,legend = FALSE, main = "IUCN criteria")
plot(rangesize$RangeObjects$A00,add = TRUE, col = "red",legend = FALSE)
plot(rangesize$RangeObjects$E00@polygons,add = TRUE, border = "red", lwd = 2)
```



*ecospat.occupied.patch* quantified the number of patches where species occupied based on species distribution predictions, species occurrences and a buffer value (in meter) around species occurrences.

```
ocp <- ecospat.occupied.patch(my.ESM.EF.proj.current.bin,
                             xy[sp_occ==1,],
                             buffer = 500)
plot(ocp,col = viridis::viridis(50)) ## occupied patches: colored areas
points(xy[sp_occ==1,],col = "red",cex = 0.5, pch = 19)
```



## b. Niche dynamics of invasive species

In this section, we use the *ecospat* package to investigate the environmental niche dynamics of an invasive species across its native and invaded ranges. We will develop the example of the spotted knapweed (*Centaurea stoebe*), an Asteraceae native to Central and Eastern Europe. In Europe, both diploid and tetraploid cytotypes can be found, but only the tetraploids have spread to North America, where it is considered an invasive species, most presumably because of its ability to produce perennial polycarpic plants (Mráz et al., 2011; Treier et al., 2009). This species served as a study case upon which most of the methodological development on niche dynamics presented in this section have been based (Broennimann et al., 2007 ; Broennimann et al., 2012 ; Broennimann et al., 2014; Broennimann & Guisan, 2008; Guisan et al., 2014; Petitpierre et al., 2012 ). Note however that, for practical reasons, the analyses presented here use occurrence and environmental data at a lower resolution than used in the original analyses, which can lead to slightly different results. The main idea of the niche dynamics analyses is to quantify realized environmental niches (sensu Hutchinson (1957), see Araújo & Guisan (2006); Broennimann et al. (2007)) by investigating the occurrence density of species in a gridded climatic space allowing for direct pixel quantification and permutation tests (Broennimann et al., 2012).

### i. Data preparation

First we need to load the R packages employed for the analyses. In addition to the package *ecospat*, the package *terra* is required for handling geo-referenced datasets, the package *geodata* to download the climatic dataset, the package *dplyr* to query, subset and summarize data, and the package *ade4* to perform the ordination analyses.



```
# Load the needed packages
library(ecospat)
library(terra)
library(tidyterra)
library(geodata)
library(dplyr)
library(ade4)
library(viridis)
```

Then we download the climatic data, here the bioclim variables of the WorldClim v2.1 dataset at 10' resolution using *worldclim\_global* from the *geodata* package. For computational efficiency, we aggregate the climatic data at 30' resolution (~50 km) using *aggregate* from *terra*. To delimit the study area for the analyses we also download (and unzip) the geo-referenced polygons of the Terrestrial Ecoregions of the World distributed by the WWF.

```
## seed set for reproducible results
set.seed(123)
## create a randomly attributed temporary local folder to download the data
tempdir<-tempdir()

## download worldclim 2.1 climate data at 10' resolution and biomes data

clim.world<-geodata::worldclim_global("bio",res=10,version="2.1",path=tempdir)
clim.world<-terra::aggregate(clim.world,3,fun=mean,na.rm=TRUE)
names(clim.world)<-sub("wc2.1_10m_", "",names(clim.world))
names(clim.world)<-sub("_", "",names(clim.world))

## download Terrestrial Ecoregions of the World dataset

teow.URL<-"https://storage.googleapis.com/teow2016/Ecoregions2017.zip"
# official website for manual download:
# https://files.worldwildlife.org/wwfcmprod/files/Publication/file
# /6kcchn7e3u_official_teow.zip
download.file(teow.URL,destfile = paste0(tempdir,"TEOW.zip"))
unzip(paste0(tempdir,"TEOW.zip"),exdir = tempdir)
biomes<-terra::vect(paste0(tempdir,"Ecoregions2017.dbf"),crs="+proj=longlat")
```

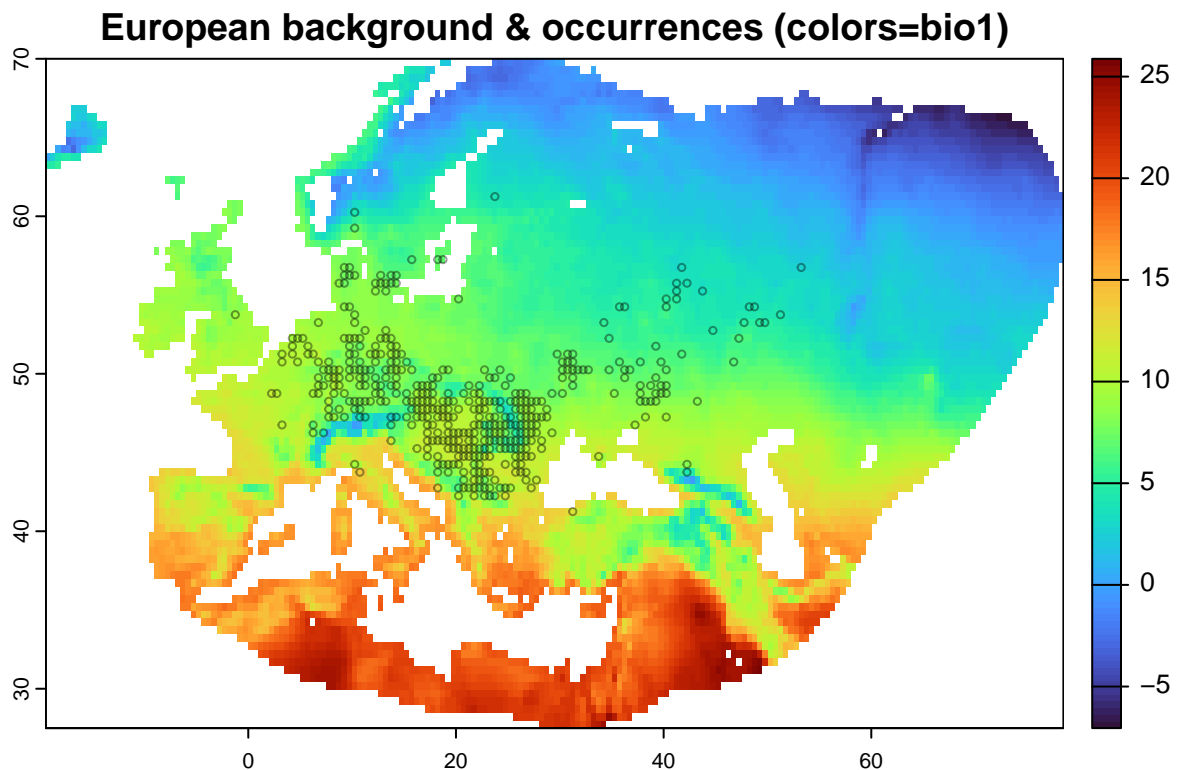
We now load and prepare the data for the two regions of analysis, namely Europe, where the species is native, and North America, where the species is invasive. The occurrence data are retrieved from the *extdata* folder included in the *ecospat* package and converted to a convenient vector format with *vect*. We thus use only occurrences of tetraploid cytotypes to ensure a fair comparison between ranges, as diploids are absent from the North American continent. We restrict the regions of analysis (background) to the biomes where the species is present in each realm using a pipe of *dyplr* functions. The biomes selected are Temperate Broadleaf Forests (4), Temperate coniferous forests (5), Boreal Forests (6), Temperate Grasslands (8) Mediterranean Forests and woodlands (12) and Deserts (13). We further restrict the background to areas distant of less than 500 km from existing occurrences (this avoids including large regions of the Palearctic realm where the species does not occur in the native range). The background vector can then be used to *crop* and *mask* the climatic data. Note that the choice of the background area has important consequences on niche quantification (Broennimann et al., 2012; Rödder & Engler, 2011), especially in the similarity tests presented below. The general advice in selecting the background is to include areas that are, or have been accessible to the species through dispersal during its evolutionary life (Barve et al., 2011). We recommend performing a sensitivity analysis on the parameters chosen for the background selection to ensure the stability and reliability of results. The plots show the background area and overlaid occurrences to visualize the final data to be analyzed.

```
## data for Europe

occ.EU.df<-read.delim(system.file("extdata/Csto4xEU.txt", package="ecospat"))
occ.EU<-vect(occ.EU.df,geom=c("x", "y"),crs="+proj=longlat")

# biomes

buf<-buffer(occ.EU,1500000) %>% aggregate() # 1500km buffer around occurrences
bkg<-biomes %>% filter(BIOME_NUM %in% c(4,5,6,8,12,13)) %>% group_by(REALM) %>%
  summarize() %>% crop(buf) # biomes in the realm where Centaurea is present
clim.EU<-crop(clim.world,bkg)
clim.EU<-mask(clim.EU,bkg)
plot(clim.EU[[1]], col=viridis(100,option="turbo"),
     main="European background & occurrences (colors=bio1)",
     #plot annual temperature (bio1)
plot(occ.EU,cex=0.5,pch=21,alpha=0.4,add=TRUE)
```



```
## data for North America

# occurrences

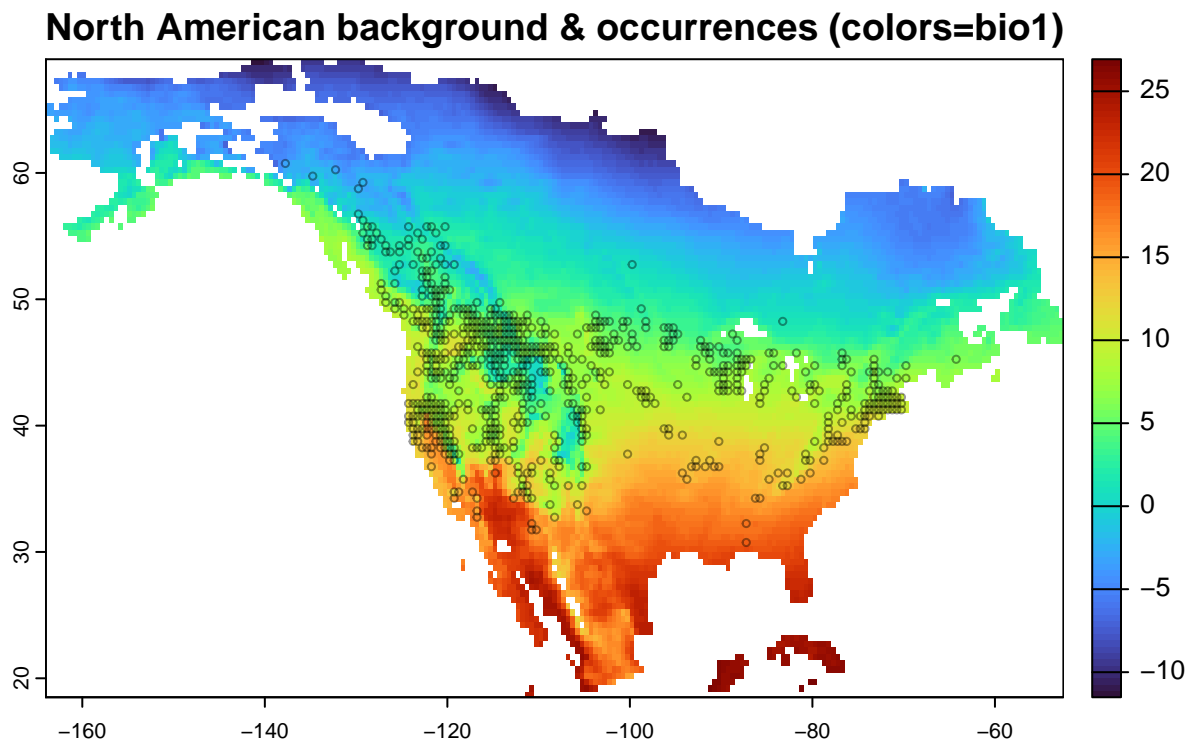
occ.NAM.df<-read.delim(system.file("extdata/Csto4xNAM.txt", package="ecospat"))
occ.NAM<-vect(occ.NAM.df,geom=c("x", "y"),crs="+proj=longlat")

# biomes
```

```

buf<-buffer(occ.NAM,150000) %>% aggregate()
bkg<-biomes %>% filter(BIOME_NUM %in% c(2,4,5,6,8,12,13)) %>% group_by(REALM) %>%
  summarize() %>% crop(buf)
clim.NAM<-crop(clim.world,bkg)
clim.NAM<-mask(clim.NAM,bkg)
plot(clim.NAM[[1]], col=viridis(100,option="turbo"),
  main="North American background & occurrences (colors=bio1)"
  #plot annual temperature (bio1)
plot(occ.NAM,cex=0.5,pch=21,alpha=0.4,add=TRUE)

```



## ii. Niche quantification

The niche quantification will be performed in two-dimensional space represented by the two first components (PC1 and PC2) of a principal component analysis (PCA). We chose to use two niche axes that maximize the variation contained in the original 19 bioclim variables from the WorldCim dataset, but the user could as well choose two specific variables relevant for the ecology of the species based on expert assessment. Niche quantification in one dimension is presented in a following section. To calibrate the PCA on the full extent of climatic values present in both background areas, we first join the European and North American datasets using *merge* from *terra*. We then extract the climatic values for each pixels of the global background (*clim.glob*), for each continental background (*clim.glob.EU*, *clim.glob.NAM*), and for each set of continental occurrences (*clim.occ.EU*, *clim.occ.NAM*) using *extract* from *terra*. The PCA is calibrated on the values of the global background using *dudi.pca* from *ade4*. The function *ecospat.plot.contrib* allows to plot the correlation circle indicating the correlation of original variables with the principal component axes with arrows. Here PC1 is mostly (but not exclusively) correlated with precipitation variables (bio12 to bio 19)

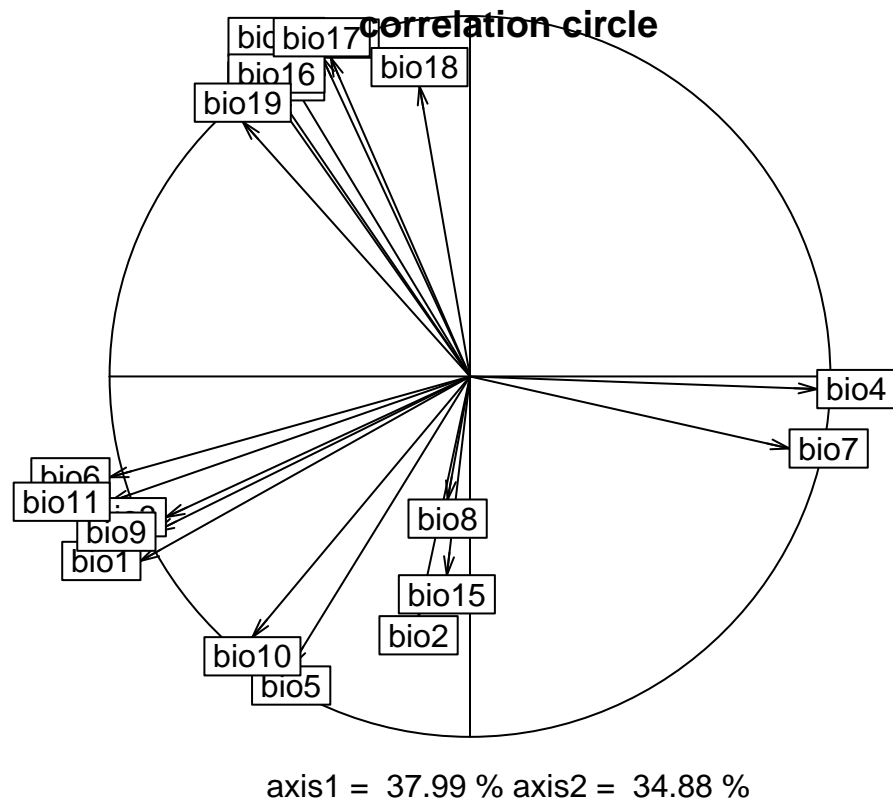
while PC2 is more correlated with temperature variables (bio1 to bio11). The inertia of each principal component (similar to the variance explained in a regression model context) is also indicated. Here the combined contribution of PC1 and PC2 explains almost 80% of the climatic variation in the original dataset. The calculation of PCA scores (*li* vectors in *dudi.pca* objects) for the pixels of the backgrounds (*scores.glob*, *scores.glob.EU*, *scores.glob.NAM*) and for the occurrences (*scores.occ.EU*, *scores.occ.NAM*) using *suprow* from *ade4* will be useful for the next step.

```
# extraction of climate values
```

```
clim<-terra::merge(clim.EU,clim.NAM)
clim.glob<-na.omit(terra::values(clim))
clim.glob.EU<-na.omit(terra::values(clim.EU))
clim.glob.NAM<-na.omit(terra::values(clim.NAM))
clim.occ.EU<-terra::extract(clim.EU,occ.EU,ID=FALSE)
clim.occ.NAM<-terra::extract(clim.NAM,occ.NAM,ID=FALSE)
```

```
## PCA calibration
```

```
pca.env<-dudi.pca(clim.glob,scannf=FALSE,nf=2)
ecospat.plot.contrib(contrib=pca.env$co,eigen=pca.env$eig)
```



```
## PCA scores
```

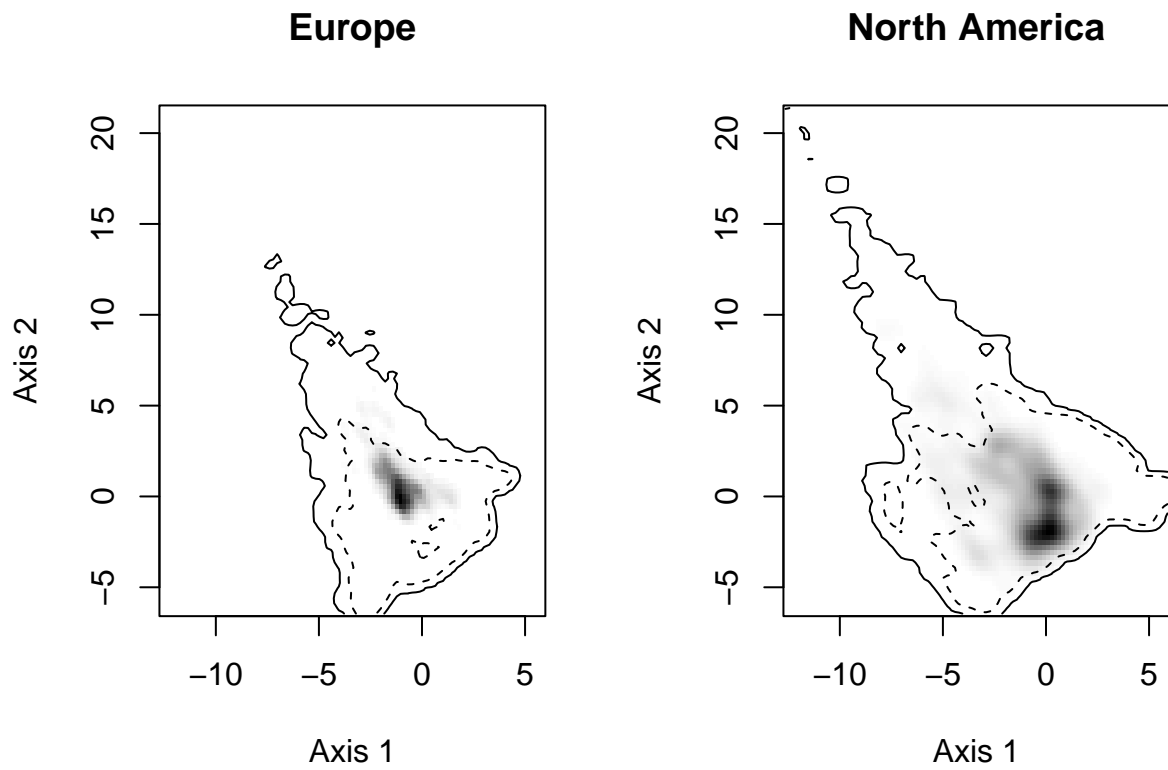
```
scores.glob<-pca.env$li
scores.glob.EU<-suprow(pca.env,clim.glob.EU)$li
scores.glob.NAM<-suprow(pca.env,clim.glob.NAM)$li
```

```
scores.occ.EU<-suprow(pca.env,clim.occ.EU)$li
scores.occ.NAM<-suprow(pca.env,clim.occ.NAM)$li
```

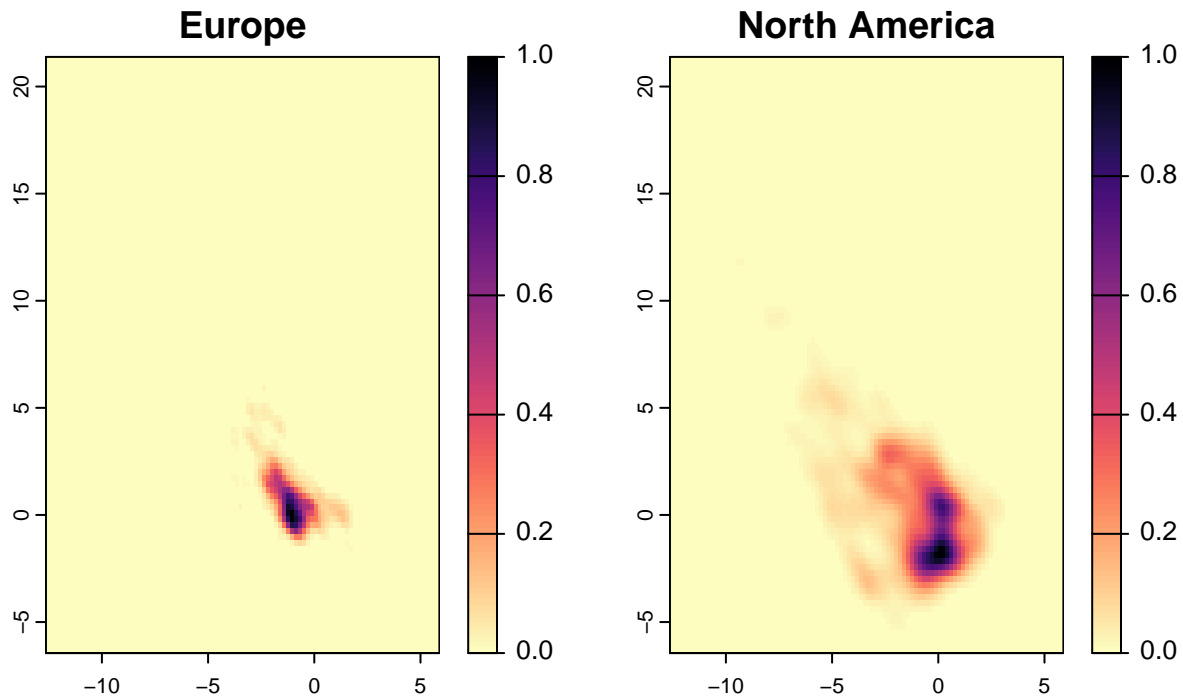
The following step is the core of the niche quantification analysis. The function *ecospat.grid.clim.dyn* first creates a gridded two-dimensional climatic space of RxR pixels bounded by the minimum and maximum values of the PCA scores of the background points (*score.glob*). It then uses kernel density estimation functions to calculate the density of occurrence in each pixel. The result is a grid of occurrence densities in a two-dimensional climatic space, with high values indicating the core of species niche, low values indicating the margin of the niche, and zero values indicating conditions outside of the niche. For comparison between ranges, occurrences densities are standardized between zero and one (object *z.uncor*). An alternative version of occurrences densities is also calculated by dividing in each pixel the occurrence density by the density of the corresponding climatic condition in the background (object *z.cor* return by *ecospat.grid.clim.dyn*; defined as occurrence occupancy in Broennimann et al. (2012)). The function *ecospat.plot.niche* allows to plot the calculated niche as a gradient of occurrence densities (shades of gray pixels) along the two first PCA axes. The contours of all climates (black solid line) and 50% most common climates (black dashed line) are also shown. Alternatively, the niches can be directly plotted with the *plot* function as *rast* objects.

```
z.EU <- ecospat.grid.clim.dyn(glob=scores.glob,
                             glob1 = scores.glob.EU,
                             sp= scores.occ.EU,
                             R=100, th.sp=0,th.env=0, kernel.method = "ks")
z.NAM <- ecospat.grid.clim.dyn(glob=scores.glob,
                              glob1 = scores.glob.NAM,
                              sp= scores.occ.NAM,
                              R=100, th.sp=0,th.env=0, kernel.method = "ks")

par(mfrow=c(1,2))
ecospat.plot.niche(z.EU,title="Europe")
ecospat.plot.niche(z.NAM,title="North America")
```



```
plot(z.EU$z.uncor,col=viridis(100,option="magma",direction = -1),main="Europe")
plot(z.NAM$z.uncor,col=viridis(100,option="magma",direction = -1),main="North America")
```



```
par(mfrow=c(1,1))
```

We can analyze the results of the niches quantification more in depth. First we want to know how much the two niches overlap. This can be achieved with the *ecospat.niche.overlap*, which calculates the Schoener's D and Hellinger I overlap indices. The parameter *cor=TRUE* allows to assess the overlap when correcting for the density of environmental conditions in the background. In our case, using *cor=FALSE*, we detect an overlap D of 0.36.

```
ecospat.niche.overlap(z1=z.EU,z2=z.NAM,cor=FALSE)
```

```
## $D
## [1] 0.3287533
##
## $I
## [1] 0.5597273
```

We can ask the question whether the overlap measured is higher than random. This can be achieved with *ecospat.niche.equivalency.test* and *ecospat.niche.similarity.test* functions, based on (Warren et al., 2008). The test of niche equivalency asks whether the overlap observed between the two niches is significantly higher than the overlap that one would get if the occurrences between the two niches would be randomly reallocated. On the other hand, the test of similarity asks whether the overlap observed between the two niches is significantly higher than the overlap one would get if the two niches would be randomly moved in the background space. Both tests have been implemented in *ecospat* for correspondence with Warren et al. (2008), though here in environmental space (while in geographical space for Warren et al. (2008); see Broennimann et al. (2012)), but the niche equivalency test almost always provides significant results even

when the overlap is very low and is thus in general not informative (Collart et al., 2021; Peterson et al., 2011). We thus only run the niche similarity test, with parameters *rand.type* = 1 to randomize both niches in the background, *rep*=100 to set the number of replications, *intersection*=NA to perform the randomization in global background without removing marginal environments, and using *overlap.alternative* = “higher” as the alternative hypothesis for the test (i.e. a p-value < 0.05 indicate that the observed overlap is significantly higher than the simulated overlap from randomized niches). Here, with a p-value of 0.25, the niche similarity test indicates that the niche of the spotted knapweed in Europe and in North America do not significantly depart from similarity.

```
sim.test<-ecospat.niche.similarity.test(z1=z.EU,z2=z.NAM,rand.type=1,rep=100,
                                         intersection=NA, overlap.alternative="higher")

sim.test$obs$D
```

```
## [1] 0.2537814
```

We can further investigate whether some parts of the native niche are not occupied in the invaded range, and conversely, if the species occupies new conditions in the invaded range. This can be done with the *expansion*, *stability* and *unfilling* indices returned by the *ecospat.niche.dyn.index* function (see COUE framework in Guisan et al. (2014)). *Stability* and *expansion* correspond to the percentages of pixels of the invasive niche which conditions present, and not present in the native range, respectively. *Unfilling* corresponds to the percentage of the native niche which is not present in the invaded range. The results indicate that the invasive niche of the spotted knapweed expands largely beyond conditions present in the native niche, while only a small fraction of the native niche is not occupied in the invaded ranges. It is possible to plot the *expansion*, *stability* and *unfilling* indices in the environmental space using the function *ecospat.plot.niche.dyn*. The plot shows the niche stability in blue, niche expansion in orange, and niche unfilling in yellow. The solid green and red contour lines indicate the extent of environmental conditions that exist in the native and invaded ranges, respectively. The dotted contour lines indicate the quantile of the most abundant environment in both ranges (here 50% most abundant, with *quant*=05). The densities of occurrences in the native range (argument *interest*=1) are displayed using gray shading. T

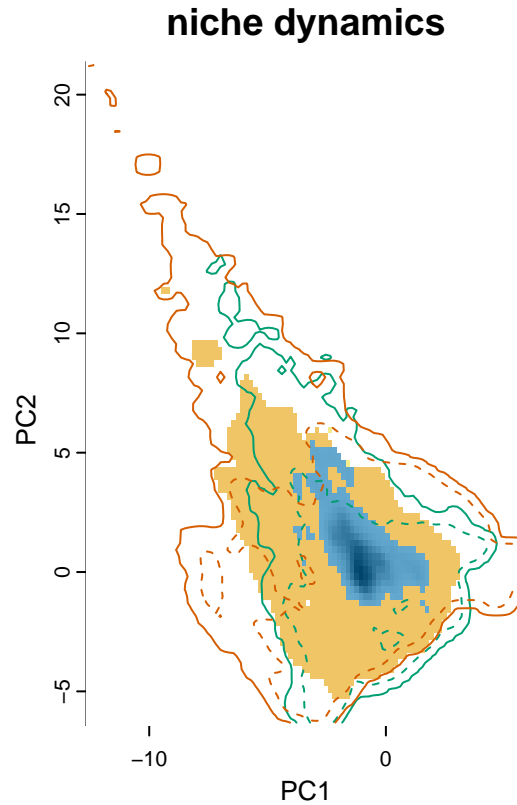
```
dyn<-ecospat.niche.dyn.index(z.EU,z.NAM,intersection=NA)
dyn$dynamic.index.w
```

```
##      expansion      stability      unfilling
## 0.5719737854 0.4280262146 0.0007058004
```

```
# colorblind friendly palette, Wong, B. (2011). Nature Methods
pal<-c("#D55E00", "#E69F00", "#F0E442", "#CC79A7", "#009E73", "#0072B2", "#56B4E9")

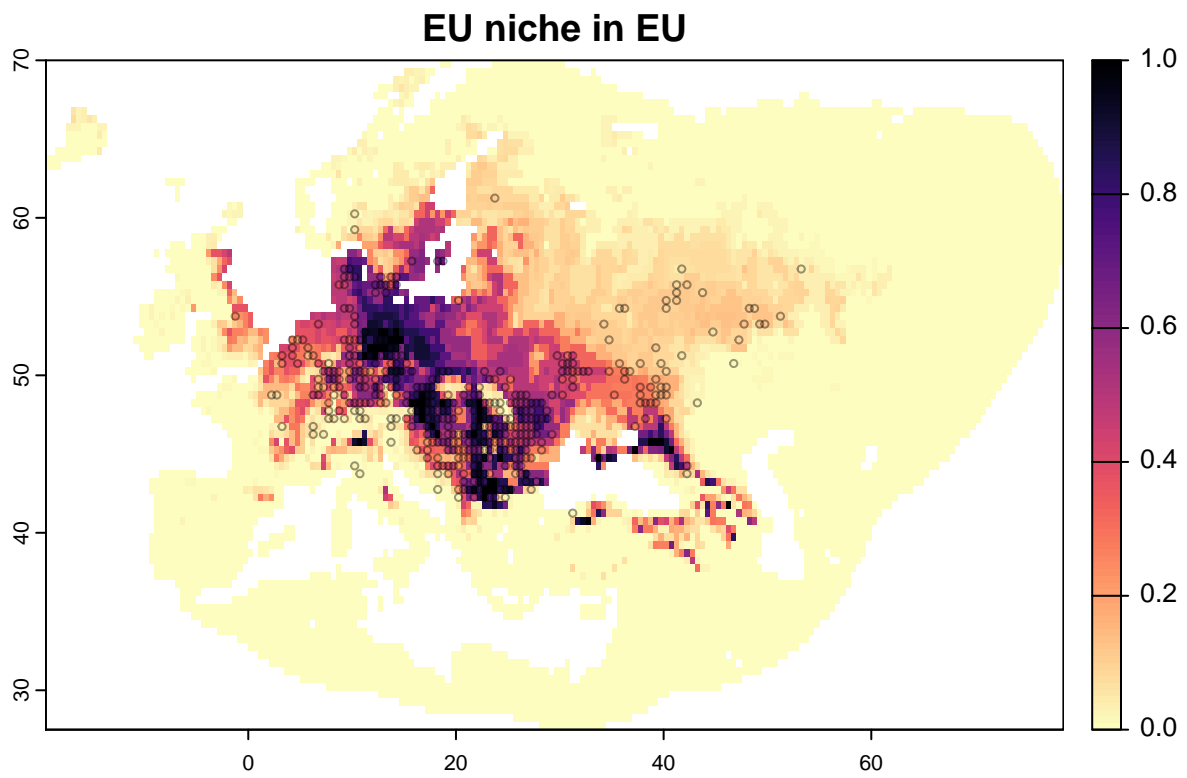
ecospat.plot.niche.dyn(z.EU,z.NAM,quant=0.5,title="niche dynamics",name.axis1 ="PC1",
                      name.axis2="PC2",interest=1,col.unf="#F0E442", col.exp="#E69F00",
                      col.stab="#0072B2", colZ1="#009E73", colZ2="#D55E00",transparency = 40)
```



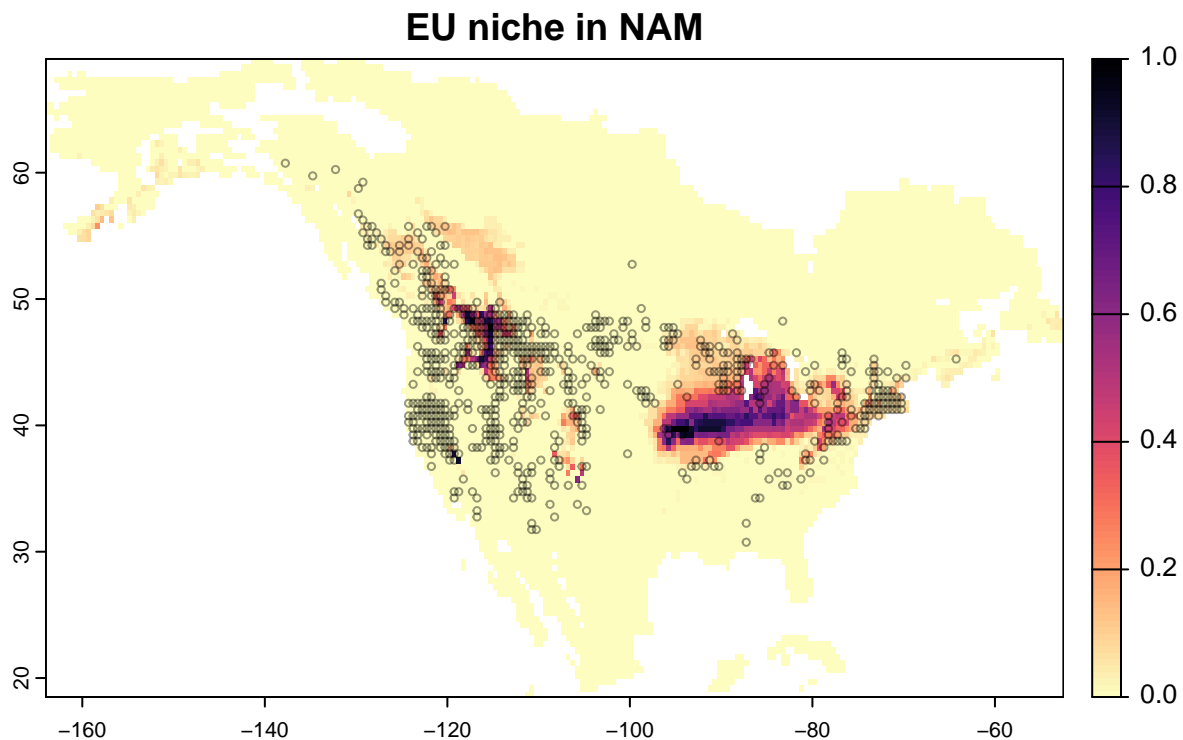


Finally, we might wonder where are the areas predicted with a high density of occurrence by the niche quantification models. Indeed, every pixel in the geography corresponds to a predicted density of occurrence in the environmental space. The function `ecospat.niche.zProjGeo` allows doing that. It is possible to project the density of occurrence of one niche (`z` object) across its own range of calibration (e.g. for Europe using `zproj=NULL` and `env=clim.EU`) but also to project the density of occurrence to another range, here for example to project the European niche of the spotted knapweed in North America, which interestingly becomes then a form of SDM.

```
geo.z.EU<-ecospat.niche.zProjGeo(z=z.EU,zproj=NULL,env=clim.EU)
plot(geo.z.EU,main="EU niche in EU",col=viridis(100,option="magma",direction=-1))
plot(occ.EU,add=TRUE,cex=0.5,pch=21,alpha=0.4)
```



```
geo.z.EUtoNAM<-ecospat.niche.zProjGeo(z=z.EU,zproj=z.NAM,env=clim.NAM)
plot(geo.z.EUtoNAM,main="EU niche in NAM",col=viridis(100,option="magma",direction =-1))
plot(occ.NAM,add=TRUE,cex=0.5,pch=21,alpha=0.4)
```

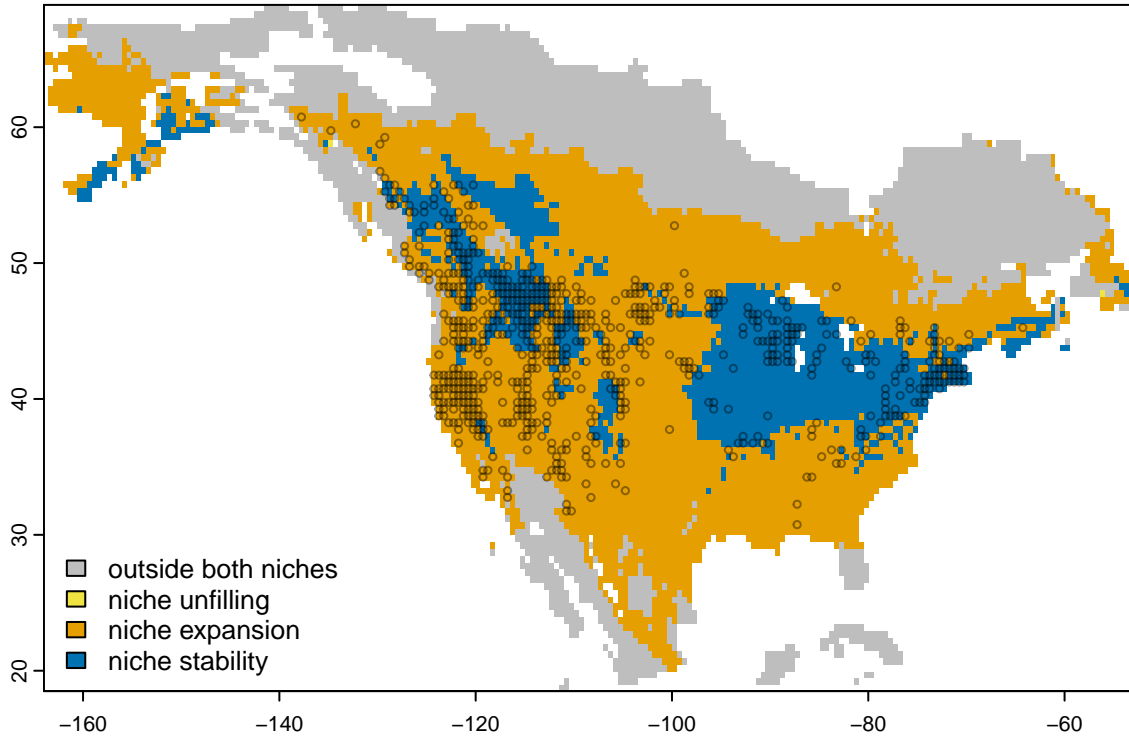


Similarly, we can also project niche dynamic indices in geography using the function `ecospat.niche.dynIndexProjGeo`. The argument `proj` controls the range of projection (i.e. `proj=2` for the North American range), and `env` sets the environmental `rast` object used for the projection. The results here show that a large part of the North East and Midwest, as well as smaller fragmented areas of the Rocky Mountains in the West have climatic conditions corresponding to the native niche of spotted knapweed, but that large areas, especially in the South correspond to an expansion of the niche.

```
geo.dyn<-ecospat.niche.dynIndexProjGeo(z.EU,z.NAM,proj=2,env=clim.NAM)

plot(geo.dyn,col = c("grey", "#F0E442", "#E69F00", "#0072B2"),
     main="niche dynamics in NAM",
     plg=list(legend=c("outside both niches", "niche unfilling",
                      "niche expansion", "niche stability"), x="bottomleft"))
plot(occ.NAM,add=TRUE,cex=0.5,pch=21,alpha=0.4)
```

## niche dynamics in NAM



### 3. Discussion and perspectives

We have illustrated the use of the *ecospat* R package to model a rare alpine plant species with ensemble of small models (ESMs; Lomba et al. (2010); Breiner et al. (2015)) and to quantify and compare niches in the native and invaded ranges of an invasive plant species with the COUE framework (Broennimann et al., 2012; Guisan et al., 2014). These two examples only used part of the functions available in *ecospat*. Other examples could have been developed for instance along the steps of the spatially-explicit species assemblage modelling (SESAM) frameworks (Guisan & Rahbek, 2011), including pre-modelling functions to analyse co-occurrences, biotic interactions and assembly of species to form assemblages (D’Amen, Mod, et al., 2018; Scherrer, Mod, et al., 2019), post-modelling of assemblages and communities from individual species predictions (D’Amen et al., 2015), or the pre- and post- analyses and modelling of spatial patterns of phylogenetic diversity, at the level of individual species (D’Amen, Mateo, et al., 2018; Pio et al., 2014) or communities (Ndiribe et al., 2013). For an example of community modelling using these functions, we send interested readers to Di Cola et al. (2017) (example 2) and to the related published references (some cited above or in D’Amen et al. (2015)). By containing some unique features (COUE niche framework, ESMs, SESAM framework), the *ecospat* package is complementary to other R packages for SDMs and spatial analyses in ecology, biogeography and conservation (see e.g. Joo et al. (2019); Meynard et al. (2019); Sillero et al. (2023), Kass et al. in review).

Furthermore, *ecospat* is complemented by three other packages initially developed in our *ecospat* group: *migclim* (Engler et al., 2012), *covsel* (Adde, Rey, Fopp, et al. (2023)) and *n-sdm* (Adde, Rey, Brun, et al. (2023)). The *migclim* package allows integration of dispersal constraints into projections of species distribution models (see e.g. Engler et al. (2009)), while *covsel* allows for efficient selection of predictors in SDMs out of large sets of candidate covariates. The *n-sdm* package is very complementary to *ecospat* in the sense that it allows modelling large number of species on high-power computing platforms (HPCs; i.e. clusters) using

three wishable properties: (i) a nested hierarchical approach (the ‘n’ in n.sdm) to avoid niche-truncation (when the study area is too small to capture all environmental requirements of a species; Chevalier et al. (2021); Chevalier et al. (2022)), (ii) the covsel approach for an efficient selection of predictors (Adde, Rey, Fopp, et al., 2023), allowing to consider also predictors that describe the environmental characteristics in the neighborhood of observations (i.e. focal windows analyses; as in Scherrer, Christe, et al. (2019)), not only in the pixel contain the observation; and (iii) the ESM approach described above to also include rare and infrequent species when modelling very large numbers of species, e.g. to assess how biodiversity might evolve under various environmental change scenarios at a national scale (Adde et al. in prep.).

Finally, ecospat is continuously evolving by increasing and adapting the number of functions to help researchers develop robust and efficient spatial analyses and models to predict the distribution of species and communities. For instance, the ESM functions are currently being optimized to run more efficiently and with less dependencies (Collart et al. in prep.), and the niche functions were expanded to include a ‘niche margin index’ (NMI; Broennimann et al. (2021); Pearman, Broennimann et al. in press) allowing to measure not only niche innerness (i.e. what SDM predictions do) but also niche outerness. Last, as the migclim package currently only runs under ancient versions of R, we foresee to integrate it into a next version of ecospat.

## 4. References

- Adde, A., Rey, P.-L., Brun, P., Külling, N., Fopp, F., Altermatt, F., Broennimann, O., Lehmann, A., Petitpierre, B., Zimmermann, N. E., Pellissier, L., & Guisan, A. (2023). *N-SDM*: a high-performance computing pipeline for Nested Species Distribution Modelling. *Ecography*, 2023(6). <https://doi.org/10.1111/ecog.06540>
- Adde, A., Rey, P.-L., Fopp, F., Petitpierre, B., Schweiger, A. K., Broennimann, O., Lehmann, A., Zimmermann, N. E., Altermatt, F., Pellissier, L., & Guisan, A. (2023). Too many candidates: Embedded covariate selection procedure for species distribution modelling with the covsel R package. *Ecological Informatics*, 75, 102080. <https://doi.org/10.1016/j.ecoinf.2023.102080>
- Araújo, M. B., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O’Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, 5(1). <https://doi.org/10.1126/sciadv.aat4858>
- Araújo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33(10), 1677–1688. <https://doi.org/10.1111/j.1365-2699.2006.01584.x>
- Austin, M. P., Nicholls, A. O., & Margules, C. R. (1990). Measurement of the Realized Qualitative Niche: Environmental Niches of Five Eucalyptus Species. *Ecological Monographs*, 60(2), 161–177. <https://doi.org/10.2307/1943043>
- Barbet-Massin, M., Rome, Q., Villemant, C., & Courchamp, F. (2018). Can species distribution models really predict the expansion of invasive species? *PLOS ONE*, 13(3), e0193085. <https://doi.org/10.1371/journal.pone.0193085>
- Barve, N., Barve, V., Jimenez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., Soberon, J., & Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, 222(11), 1810–1819. <https://doi.org/10.1016/j.ecolmodel.2011.02.011>
- Breiner, F. T., Guisan, A., Bergamini, A., & Nobis, M. P. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, 6(10), 1210–1218.
- Breiner, F. T., Nobis, M. P., Bergamini, A., & Guisan, A. (2018). Optimizing ensembles of small models for predicting the distribution of species with few occurrences. *Methods in Ecology and Evolution*, 9(4), 802–808. <https://doi.org/10.1111/2041-210x.12957>
- Broennimann, O., Di Cola, V., & Guisan, A. (2023). *Ecospat: Spatial ecology miscellaneous methods*. <https://cran.r-project.org/web/packages/ecospat/ecospat.pdf>
- Broennimann, O., Fitzpatrick, M. C., Pearman, P. B., Petitpierre, B., Pellissier, L., Yoccoz, N. G., Thuiller, W., Fortin, M. J., Randin, C., Zimmermann, N. E., Graham, C. H., & Guisan, A. (2012). Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, 21(4), 481–497. <https://doi.org/10.1111/j.1466-8238.2011.00698.x>

- Broennimann, O., & Guisan, A. (2008). Predicting current and future biological invasions: both native and invaded ranges matter. *Biology Letters*, 4(5), 585–589. <https://doi.org/10.1098/rsbl.2008.0254>
- Broennimann, O., Mráz, P., Petitpierre, B., Guisan, A., & Müller-Schärer, H. (2014). Contrasting spatio-temporal climatic niche dynamics during the eastern and western invasions of spotted knapweed in North America. *Journal of Biogeography*, 41(6), 1126–1136. <https://doi.org/10.1111/jbi.12274>
- Broennimann, O., Petitpierre, B., Chevalier, M., González-Suárez, M., Jeschke, J. M., Rolland, J., Gray, S. M., Bacher, S., & Guisan, A. (2021). Distance to native climatic niche margins explains establishment success of alien mammals. *Nature Communications*, 12(1). <https://doi.org/10.1038/s41467-021-22693-0>
- Broennimann, O., Thuiller, W., Hughes, G., Midgley, G. F., Alkemade, J. M. R., & Guisan, A. (2006). Do geographic distribution, niche property and life form explain plants' vulnerability to global change? *Global Change Biology*, 12(6), 1079–1093. <https://doi.org/10.1111/j.1365-2486.2006.01157.x>
- Broennimann, O., Treier, U. A., Müller-Schärer, H., Thuiller, W., Peterson, A. T., & Guisan, A. (2007). Evidence of climatic niche shift during biological invasion. *Ecology Letters*, 10(8), 701–709. <https://doi.org/10.1111/j.1461-0248.2007.01060.x>
- Chevalier, M., Broennimann, O., Cornuault, J., & Guisan, A. (2021). Data integration methods to account for spatial niche truncation effects in regional projections of species distribution. *Ecological Applications*, 31(7). <https://doi.org/10.1002/eap.2427>
- Chevalier, M., Zarzo-Arias, A., Guélat, J., Mateo, R. G., & Guisan, A. (2022). Accounting for niche truncation to improve spatial and temporal predictions of species distributions. *Frontiers in Ecology and Evolution*, 10. <https://doi.org/10.3389/fevo.2022.944116>
- Collart, F., & Guisan, A. (2023). Small to train, small to test: Dealing with low sample size in model evaluation. *Ecological Informatics*, 75, 102106. <https://doi.org/10.1016/j.ecoinf.2023.102106>
- Collart, F., Hedenäs, L., Broennimann, O., Guisan, A., & Vanderpoorten, A. (2021). Intraspecific differentiation: Implications for niche and distribution modelling. *Journal of Biogeography*, 48(2), 415–426. <https://doi.org/10.1111/jbi.14009>
- D'Amen, M., Mateo, R. G., Pottier, J., Thuiller, W., Maiorano, L., Pellissier, L., Ndiribe, C., Salamin, N., & Guisan, A. (2018). Improving spatial predictions of taxonomic, functional and phylogenetic diversity. *Journal of Ecology*, 106(1), 76–86. <https://doi.org/10.1111/1365-2745.12801>
- D'Amen, M., Mod, H. K., Gotelli, N. J., & Guisan, A. (2018). Disentangling biotic interactions, environmental filters, and dispersal limitation as drivers of species co-occurrence. *Ecography*, 41(8), 1233–1244. <https://doi.org/10.1111/ecog.03148>
- D'Amen, M., Rahbek, C., Zimmermann, N. E., & Guisan, A. (2015). Spatial predictions at the community level: from current approaches to future frameworks. *Biological Reviews*, 92(1), 169–187. <https://doi.org/10.1111/brv.12222>
- Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., D'Amen, M., Randin, C., Engler, R., Pottier, J., Pio, D., Dubuis, A., Pellissier, L., Mateo, R. G., Hordijk, W., Salamin, N., & Guisan, A. (2017). ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, 40(6), 774–787. <https://doi.org/10.1111/ecog.02671>
- Di Febbraro, M., Sallustio, L., Vizzarri, M., De Rosa, D., De Lisio, L., Loy, A., Eichelberger, B. A., & Marchetti, M. (2018). Expert-based and correlative models to map habitat quality: Which gives better support to conservation planning? *Global Ecology and Conservation*, 16, e00513. <https://doi.org/10.1016/j.gecco.2018.e00513>
- Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P., & Guisan, A. (2011). Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, 17(6), 1122–1131. <https://doi.org/10.1111/j.1472-4642.2011.00792.x>
- Elith, J., Ferrier, S., Huettmann, F., & Leathwick, J. (2005). The evaluation strip: new and robust method for plotting predicted responses from species distribution models. *Ecological Modelling*, 186(3), 280–289. <https://doi.org/10.1016/j.ecolmodel.2004.12.007>
- Engler, R., Hordijk, W., & Guisan, A. (2012). The MIGCLIM R package – seamless integration of dispersal constraints into projections of species distribution models. *Ecography*, 35(10), 872–878. <https://doi.org/10.1111/j.1600-0587.2012.07608.x>
- Engler, R., Randin, C. F., Vittoz, P., Czáká, T., Beniston, M., Zimmermann, N. E., & Guisan, A. (2009). Predicting future distributions of mountain plants under climate change: does dispersal capacity matter?

- Ecography*, 32(1), 34–45. <https://doi.org/10.1111/j.1600-0587.2009.05789.x>
- Ferrarini, A., Selvaggi, A., Abeli, T., Alatalo, J. M., Orsenigo, S., Gentili, R., & Rossi, G. (2016). Planning for assisted colonization of plants in a warming world. *Scientific Reports*, 6(1). <https://doi.org/10.1038/srep28542>
- Ferrier, S., & Guisan, A. (2006). Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, 43(3), 393–404. <https://doi.org/10.1111/j.1365-2664.2006.01149.x>
- Fourcade, Y., Engler, J. O., Besnard, A. G., Rödder, D., & Secondi, J. (2013). Confronting expert-based and modelled distributions for species with uncertain conservation status: A case study from the corncrake (*Crex crex*). *Biological Conservation*, 167, 161–171. <https://doi.org/10.1016/j.biocon.2013.08.009>
- Franklin, J. (2010). Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions*, 16(3), 321–330. <https://doi.org/10.1111/j.1472-4642.2010.00641.x>
- Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N. G., Lehmann, A., & Zimmermann, N. E. (2006). Using Niche-Based Models to Improve the Sampling of Rare Species. *Conservation Biology*, 20(2), 501–511. <https://doi.org/10.1111/j.1523-1739.2006.00354.x>
- Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., & Kueffer, C. (2014). Unifying niche shift studies: Insights from biological invasions. *Trends in Ecology & Evolution*, 29(5), 260–269. <https://doi.org/10.1016/j.tree.2014.02.009>
- Guisan, A., & Rahbek, C. (2011). SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, 38(8), 1433–1444. <https://doi.org/10.1111/j.1365-2699.2011.02550.x>
- Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models, with applications in r*.
- Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., Regan, T. J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, M., J. R., Setterfield, S. A., Elith, J., Schwartz, M. W., Wintle, B. A., Broennimann, O., Austin, M., Ferrier, S., ... Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, 16(12), 1424–1435. <https://doi.org/10.1111/ele.12189>
- Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199(2), 142–152. <https://doi.org/10.1016/j.ecolmodel.2006.05.017>
- Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22(0), 415–427. <https://doi.org/10.1101/sqb.1957.022.01.039>
- IUCN. (2012). *IUCN red list categories and criteria: Version 3.1, 2nd edition*. IUCN Species Survival Commission & The World Conservation Union.
- Jeliazkov, A., Gavish, Y., Marsh, C. J., Geschke, J., Brummitt, N., Rocchini, D., Haase, P., Kunin, W. E., & Henle, K. (2022). Sampling and modelling rare species: Conceptual guidelines for the neglected majority. *Global Change Biology*, 28(12), 3754–3777. <https://doi.org/10.1111/gcb.16114>
- Jiménez-Valverde, A. (2020). Sample size for the evaluation of presence-absence models. *Ecological Indicators*, 114, 106289. <https://doi.org/10.1016/j.ecolind.2020.106289>
- Joo, R., Boone, M. E., Clay, T. A., Patrick, S. C., Clusella-Trullas, S., & Basille, M. (2019). Navigating through the r packages for movement. *Journal of Animal Ecology*, 89(1), 248–267. <https://doi.org/10.1111/1365-2656.13116>
- Kass, J. M., Guénard, B., Dudley, K. L., Jenkins, C. N., Azuma, F., Fisher, B. L., Parr, C. L., Gibb, H., Longino, J. T., Ward, P. S., Chao, A., Lubertazzi, D., Weiser, M., Jetz, W., Guralnick, R., Blatrix, R., Lauriers, J. D., Donoso, D. A., Georgiadis, C., ... Economo, E. P. (2022). The global distribution of known and undiscovered ant biodiversity. *Science Advances*, 8(31). <https://doi.org/10.1126/sciadv.abp9908>
- Lomba, A., Pellissier, L., Randin, C. R., Vicente, J., Moreira, F., Honrado, J., & Guisan, A. (2010). Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*, 143(11), 2647–2657. <https://doi.org/10.1016/j.biocon.2010.07.007>
- Meynard, C. N., Leroy, B., & Kaplan, D. M. (2019). Testing methods in species distribution modelling using virtual species: what have we learnt and what are we missing? *Ecography*, 42(12), 2021–2036. <https://doi.org/10.1111/ecog.04385>

- Moilanen, A., Lehtinen, P., Kohonen, I., Jalkanen, J., Virtanen, E. A., & Kujala, H. (2022). Novel methods for spatial prioritization with applications in conservation, land use planning and ecological impact avoidance. *Methods in Ecology and Evolution*, 13(5), 1062–1072. <https://doi.org/10.1111/2041-210x.13819>
- Mráz, P., Bouchier, R. S., Treier, U. A., Schaffner, U., & Müller-Schärer, H. (2011). Polyploidy in Phenotypic Space and Invasion Context: A Morphometric Study of *Centaurea stoebe*s.l. *International Journal of Plant Sciences*, 172(3), 386–402. <https://doi.org/10.1086/658151>
- Ndiribe, C., Pellissier, L., Antonelli, S., Dubuis, A., Pottier, J., Vittoz, P., Guisan, A., & Salamin, N. (2013). Phylogenetic plant community structure along elevation is lineage specific. *Ecology and Evolution*, 3(15), 4925–4939. <https://doi.org/10.1002/ece3.868>
- Patiño, J., Collart, F., Vanderpoorten, A., Martin-Esquivel, J. L., Naranjo-Cigala, A., Mirolo, S., & Karger, D. N. (2023). Spatial resolution impacts projected plant responses to climate change on topographically complex islands. *Diversity and Distributions*, 29(10), 1245–1262. <https://doi.org/10.1111/ddi.13757>
- Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R., Martínez-Meyer, E., Nakamura, M., & Araújo, M. P. (2011). *Ecological niches and geographic distributions*. Princeton University Press.
- Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C., & Guisan, A. (2012). Climatic niche shifts are rare among terrestrial plant invaders. *Science*, 335(6074), 1344–1348. <https://doi.org/10.1126/science.1215933>
- Pio, D. V., Engler, R., Linder, H. P., Monadjem, A., Cotterill, F. P. D., Taylor, P. J., Schoeman, M. C., Price, B. W., Villet, M. H., Eick, G., Salamin, N., & Guisan, A. (2014). Climate change effects on animal and plant phylogenetic diversity in southern Africa. *Global Change Biology*, 20(5), 1538–1549. <https://doi.org/10.1111/gcb.12524>
- Pol, M. van de, Bailey, L. D., McLean, N., Rijdsdijk, L., Lawson, C. R., & Brouwer, L. (2016). Identifying the best climatic predictors in ecology and evolution. *Methods in Ecology and Evolution*, 7(10), 1246–1257. <https://doi.org/10.1111/2041-210x.12590>
- Pulliam, H. R. (2000). On the relationship between niche and distribution. *Ecology Letters*, 3(4), 349–361. <https://doi.org/10.1046/j.1461-0248.2000.00143.x>
- Rödder, D., & Engler, J. O. (2011). Quantitative metrics of overlaps in Grinnellian niches: advances and possible drawbacks. *Global Ecology and Biogeography*, 20(6), 915–927. <https://doi.org/10.1111/j.1466-8238.2011.00659.x>
- Scherrer, D., Christe, P., & Guisan, A. (2019). Modelling bat distributions and diversity in a mountain landscape using focal predictors in ensemble of small models. *Diversity and Distributions*, 25(5), 770–782. <https://doi.org/10.1111/ddi.12893>
- Scherrer, D., Mod, H. K., Pottier, J., Litsios-Dubuis, A., Pellissier, L., Vittoz, P., Götzenberger, L., Zobel, M., & Guisan, A. (2019). Disentangling the processes driving plant assemblages in mountain grasslands across spatial scales and environmental gradients. *Journal of Ecology*, 107(1), 265–278. <https://doi.org/10.1111/1365-2745.13037>
- Shen, T., Song, L., Corlett, R. T., Guisan, A., Wang, J., Ma, W.-Z., Mouton, L., Vanderpoorten, A., & Collart, F. (2023). Disentangling the roles of chance, abiotic factors and biotic interactions among epiphytic bryophyte communities in a tropical rainforest (Yunnan, China). *Plant Biology*, 25(6), 880–891. <https://doi.org/10.1111/plb.13570>
- Sillero, N., Campos, J. C., Arenas-Castro, S., & Barbosa, A. M. (2023). A curated list of R packages for ecological niche modelling. *Ecological Modelling*, 476, 110242. <https://doi.org/10.1016/j.ecolmodel.2022.110242>
- Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, 10(12), 1115–1123. <https://doi.org/10.1111/j.1461-0248.2007.01107.x>
- Thuiller, W., Georges, D., Gueguen, M., Engler, R., Breiner, F., Lafourcade, B., & Patin, R. (2023). *biomod2: Ensemble platform for species distribution modeling*. <https://CRAN.R-project.org/package=biomod2>
- Treier, U. A., Broennimann, O., Normand, S., Guisan, A., Schaffner, U., Steinger, T., & Müller-Schärer, H. (2009). Shift in cytotype frequency and niche space in the invasive plant *Centaurea maculosa*. *Ecology*, 90(5), 1366–1377. <https://doi.org/10.1890/08-0420.1>
- van Proosdij, A. S. J., Sosef, M. S. M., Wieringa, J. J., & Raes, N. (2015). Minimum required number of specimen records to develop accurate species distribution models. *Ecography*, 39(6), 542–552. <https://doi.org/10.1111/ecog.01509>
- Warren, D. L., Glor, R. E., & Turelli, M. (2008). Environmental niche equivalency versus conservatism:



Quantitative approaches to niche evolution. *Evolution*, 62(11), 2868–2883. ://000260348100012  
Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillerá-Arroita, G., Guisan, A., Lahoz-Monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T., Rapacciuolo, G., Schmatz, D. R., Schröder, S.-D., B., Thuiller, W., Yates, K. L., ... Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography*, 43(9), 1261–1277. <https://doi.org/10.1111/ecog.04960>