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

Olivier Broennimann[1,2], Flavien Collart[1] and Antoine Guisan[1,2]

[1]Department of Ecology and Evolution, University of Lausanne, Switzerland
[2]Institute of Earth Surface Dynamics, University of Lausanne, Switzerland

November 17, 2023

1. Introduction

Species distribution modeling (SDMs; W. Guisan A. (2017); Peterson (2011); Franklin (2010)) is a field of ecology and biogeography that has grown significantly in the last three decades (A. Araújo M. B. 2019), with important potential conservation applications (T. Guisan A. 2013). SDMs work by statistically relating (or sometime in an expert-based way) 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 range, it then quantifies the species' realized environmental niche (Austin 1990; Pulliam 2000; Soberón 2007; G. Araújo M. B. 2006) and can be used to make projections, e.g. in the future, provided some assumptions are met (e.g. equilibrium, niche stability, no niche truncation, etc; see W. Guisan A. (2017); Zurell (2020)). Individual SDMs can then be used for many applications, from climate change assessments (e.g. T. Broennimann O. (2006), anticipating biological invasions (e.g. U. A. Broennimann O. Treier (2007)), or managing rare species (B. Guisan A. 2006), and can be stacked (Dubuis 2011) to predict community patterns, like richness, composition, or functional or phylogenetic diversity (Ferrier 2006; R. Guisan A. 2011; D'Amen 2015; Shen 2023).

Three main steps can be followed to build SDMs:g (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 (Di Cola 2017). All these applications can now be performed thanks to numerous software packages, most notably in the R framework (Sillero (2023),Kass et al. in prep.). One of them is 'ecospat' (Di Cola (2017), Broennimann et al. 2023, currently vers. 4.0.0), a collection of R functions and data sets for the support of spatial ecology analyses with a focus on pre-, core- and post-modelling analyses of species distributions (SDMs) (Table 1).

In the ecospat package, pre-modeling analyses include data extractions, designing sampling strategies, sub-sampling data, assessing spatial autocorrelation, reducing collinearity among predictors, measuring environmental similarity (mess), assessing niche stability within a species or niche difference among species (using a large set of functions, key development in 'ecospat'), and measuring species co-occurrences to investigate community assembly. Core-modeling analyses include . . .

In this book section, we first illustrate two aimed at presenting 2 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.

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 (2010); A. Breiner F. T. (2015); M. P. Breiner F. T. (2018), which is particularly suitable for rare species. 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 (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. 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 an ensemble. By averaging simple small models to an ensemble, 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 A. Breiner F. T. (2015). For this section, we will be focusing on modeling the ecological niche of *Veronica alpina* in the Western Swiss Alps.

i. Pre-Modeling

The ESM functions of the ecospat package relies on biomod2. We thus need to first format our data by using the function 'BIOMOD_FormatingData', 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. We also have 5 environmental predictors in our study area. These variables are the growing degree days (with a 0°C threshold), moisture index over the growing season (average values for June to August in mm day-1), the annual sum of radiation (in kJ m-2 year-1), Slope (in degrees), and the topographic position .

```r
# Load the packages
library(ecospat)
library(biomod2)
library(terra)
library(viridis)
```

```r
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
```

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

```r
# Formatting the data with the BIOMOD_FormatingData() function from the package biomod2
myBiomodData <- biomod2::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 corrends 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*

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*.

```
t(my.ESM.EF$ESM.evaluations)
```

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 pooling evaluation (Collart 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. 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). This function can compute several metrics, which can be selected with the argument *metrics*. If needed, *EachSmallModels* allows to evaluate each bivariate models 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","MaxKappa",
                                      "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
my.ESM.EF.eval$ESM.evaluations
```

Table 2: ESM performances based on the pooling evaluation

	AUC	SomersD	Boyce	MaxKappa
GLM	0.847	0.694	0.73	0.25

Ecospat package also has numerous functions to compute model performances on your own. By providing model predictions and species observation.

For example, the Boyce index which only requires presences 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$resp==1],
                             PEplot = FALSE,
                             method = "spearman")
boyce.index$cor
```
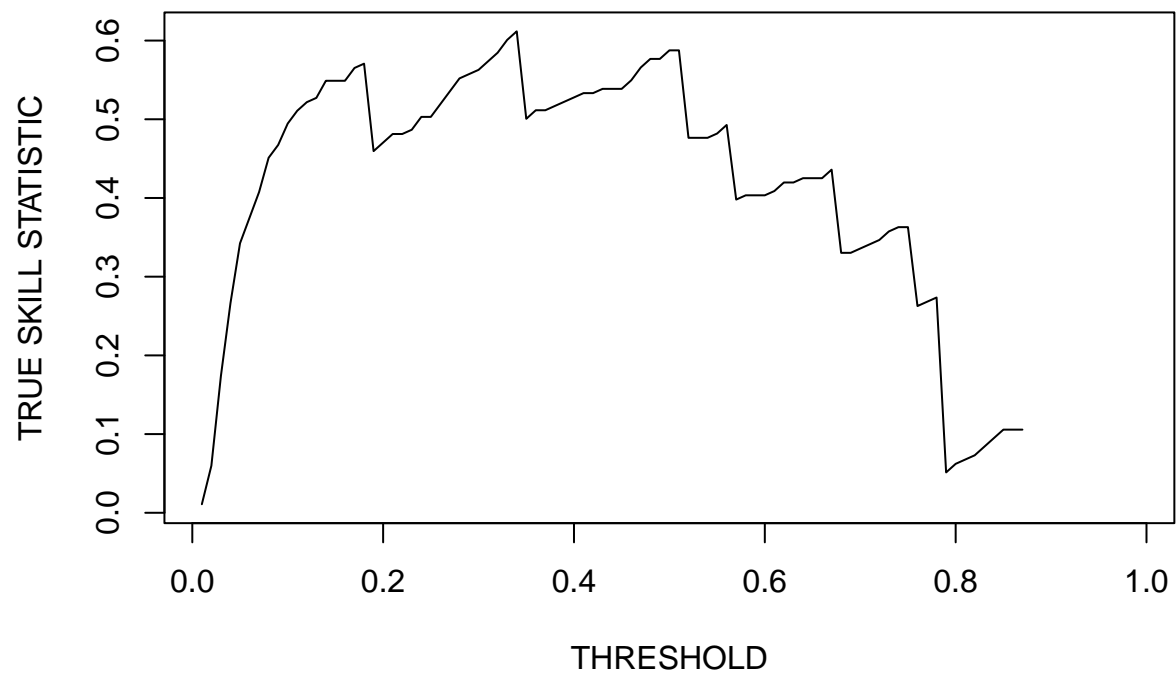
```
## [1] 0.73
```

MaxTSS and MaxKappa can be estimated 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$Fit_GLM,
                          Sp.occ = pred.test$resp)
MaxTSS$max.TSS
```

```
## [1] 0.611715
```
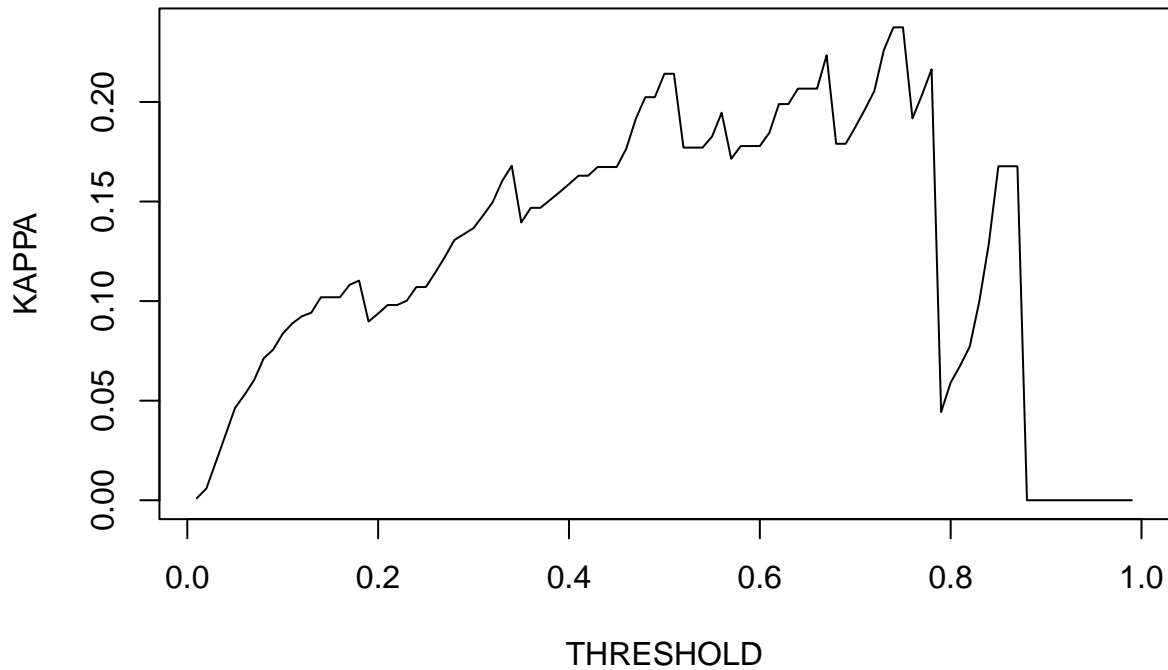
```
ecospat.plot.tss(Pred = pred.test$Fit_GLM,
                 Sp.occ = pred.test$resp)
```

```
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 performances can also be checked by observing species response curves to each environmental predictors. To do so, The function *ecospat.ESM.responsePlot* can be used. This function is an adaptation of the Evaluation Strip method proposed by F. Elith J. (2005) and needs the objects returned by *ecospat.ESM.Modeling* and *ecospat.ESM.EnsembleModeling*. The statistic used to keep constant the other predictor while generated the response curve for a predictor can be changed via the argument *fixed.var.metric*

```
response.plots <- ecospat.ESM.responsePlot(ESM.EnsembleModeling.output =
                                            my.ESM.EF,
                                    ESM.modeling.output = my.ESM,
                                    fixed.var.metric = 'median')
```
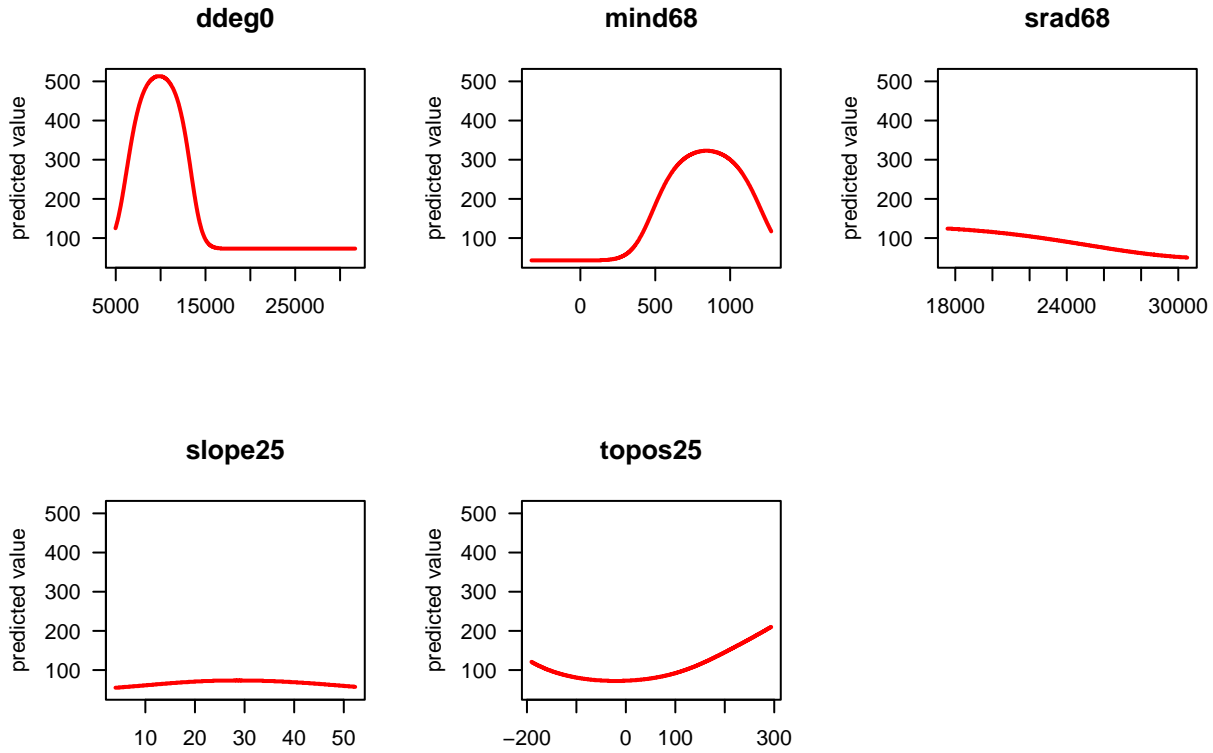
To measure the contribution of each variable, you can use the function *ecospat.ESM.VarContrib.* This function computes the ratio between sum of weights of bivariate models where a focal variable was used and sum of weights of bivariate models where the focal variable was not used. The ratio is corrected 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. 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.

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

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* which projects each bivariate model and *ecospat.ESM.EnsembleProjection* which generates the ensemble of these bivariate models. *new.env* argument allows to project models onto a new

data.frame, *SpatRaster* of ecological values while *name.env* allows to give a name to the projection. Parallel computing can be enabled with the argument *parallel*

```r
### 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")
```

```r
### 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")
```
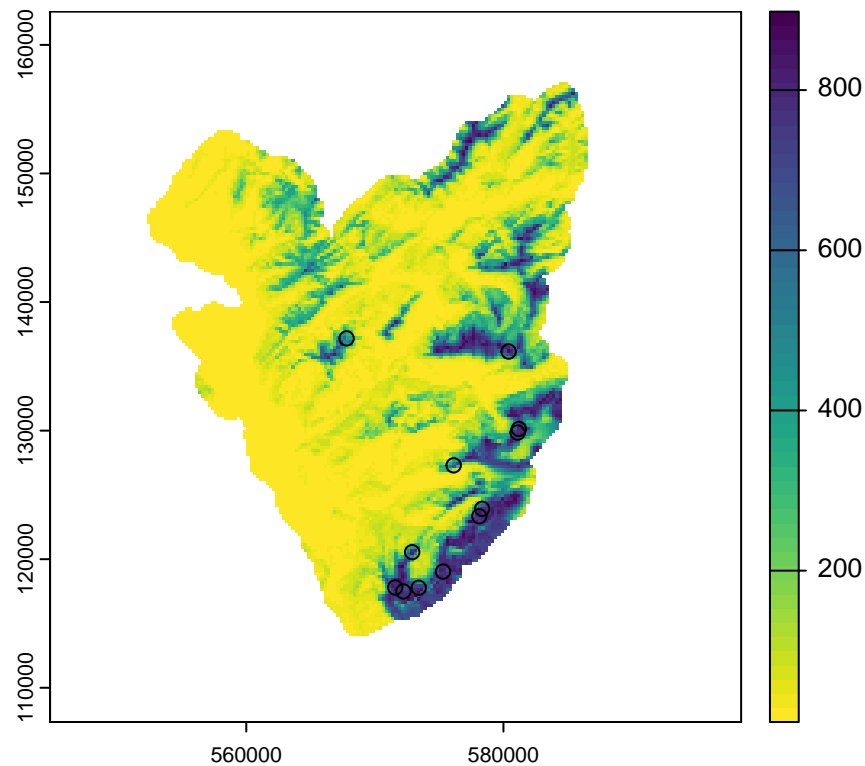
iii. Post-modeling

ESM projections can be afterwards binarized. 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 transferability).

```
Thr <- ecospat.ESM.threshold(ESM.EnsembleModeling.output = my.ESM.EF,
                             PEplot = FALSE)
t(Thr)
```

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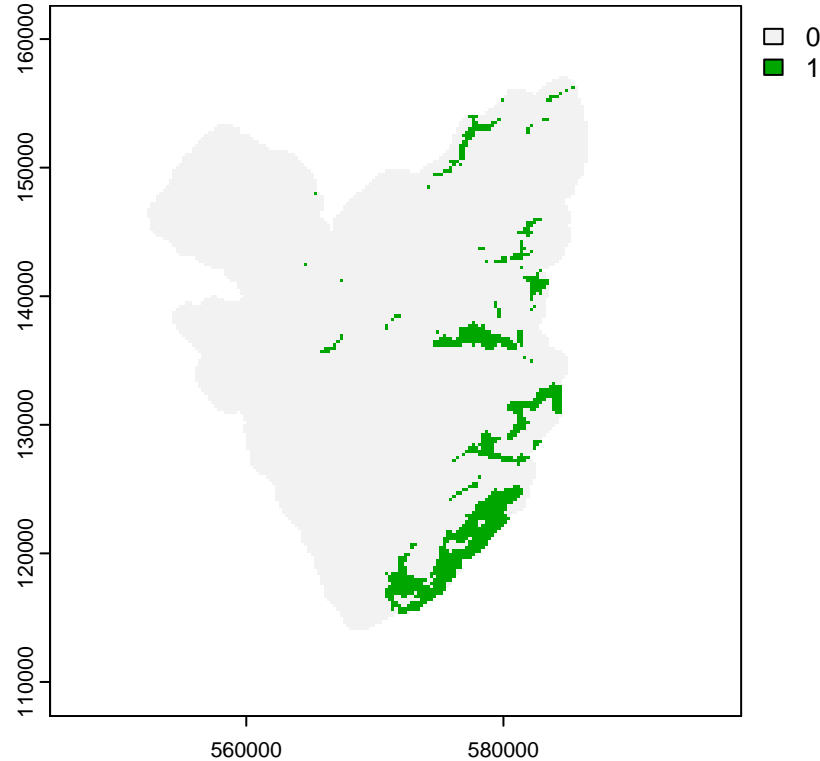
After binarizing maps, one could quantify the species range size or the occupied patches from ESM maps and IUCN criteria. In the ecospat package, the function *ecospat.rangesize* and *ecospat.occupied.patch* are made for these purposes.

More precisely, *ecospat.rangesize* allows quantifying Area of Occupancy AOO and the Extent of Occurrence EOO. 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)

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

11

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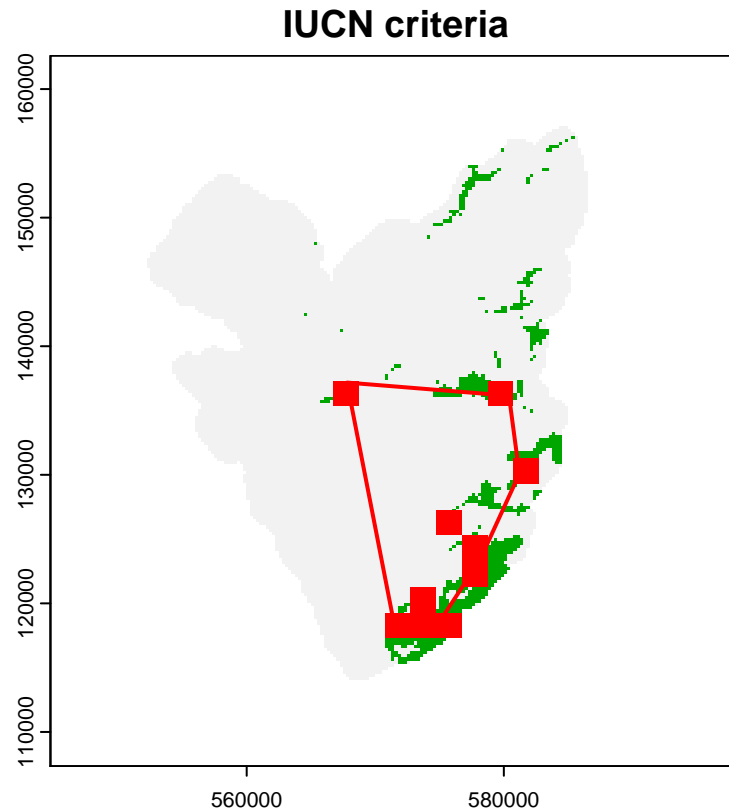
ecospat.occupied.patch quantified the number of patches where species occupied based on species distribution predictions, species occurrences and a buffer value (in meter) around species occurrences.

```
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)
```
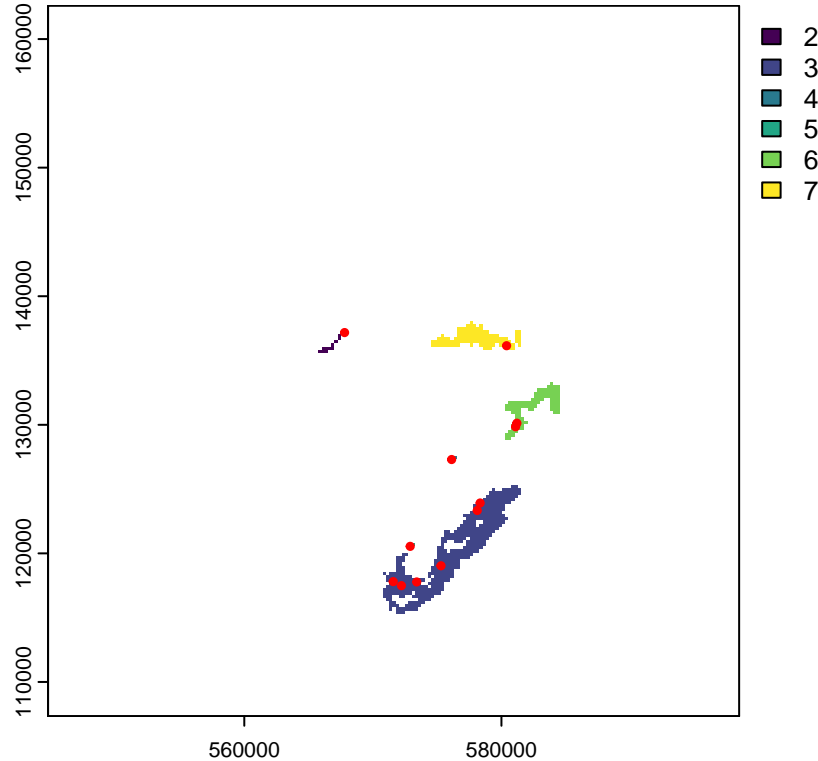
b. Niche dynamics of invasive species

In this section, we will use the ecospat package to investigate the niche dynamics of an invasive species across its native and invasive 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 (Treier 2009 ; Mráz 2011). This species served as a study case upon which most of the methodological development on niche dynamics presented in this section have been based (U. A. Broennimann O. Treier 2007 ; F. Broennimann O. 2012 ; Petitpierre 2012 ; P. Guisan A. 2014). Note however 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 environmental niches by investigating the occurrence density of species in a gridded climatic space allowing for direct pixel quantification and permutation tests.

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 worlclim 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/wwfcmsprod/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 Paleoarctic
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, 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)
```


European background & occurrences (colors=bio1)
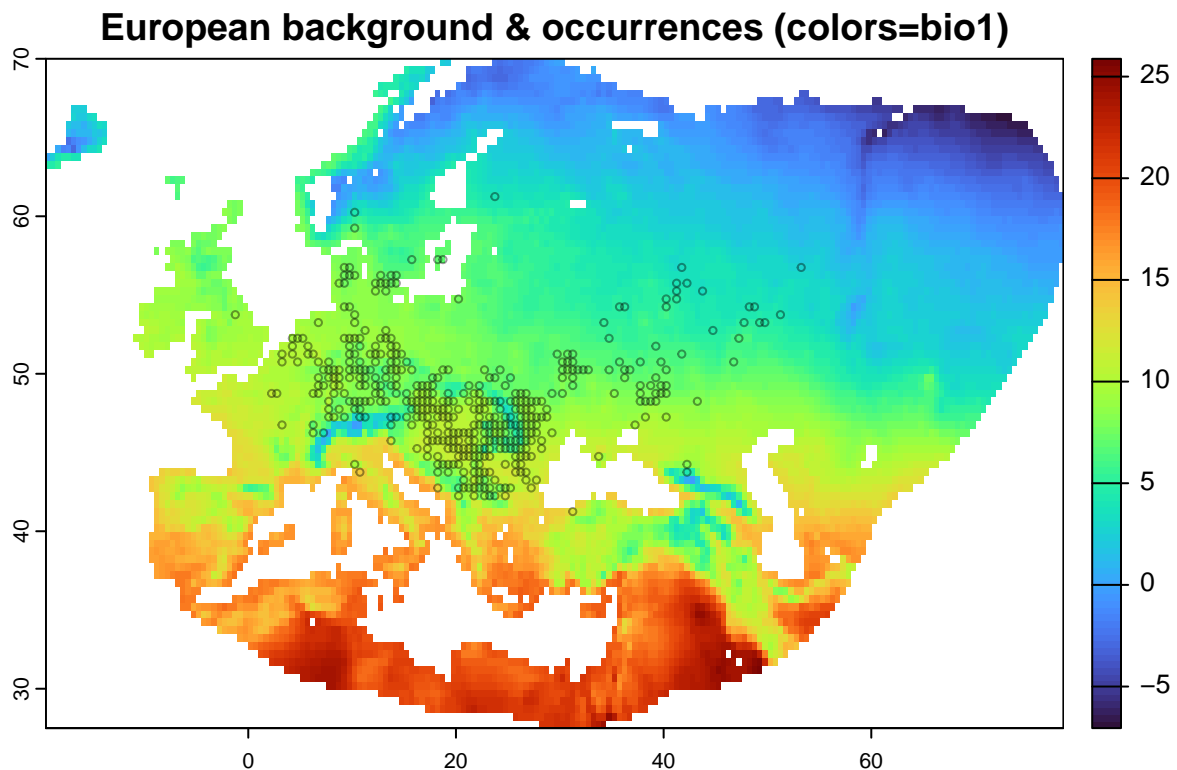
```
## data for North America

# 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,1500000) %>% 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)
```


North American background & occurrences (colors=bio1)
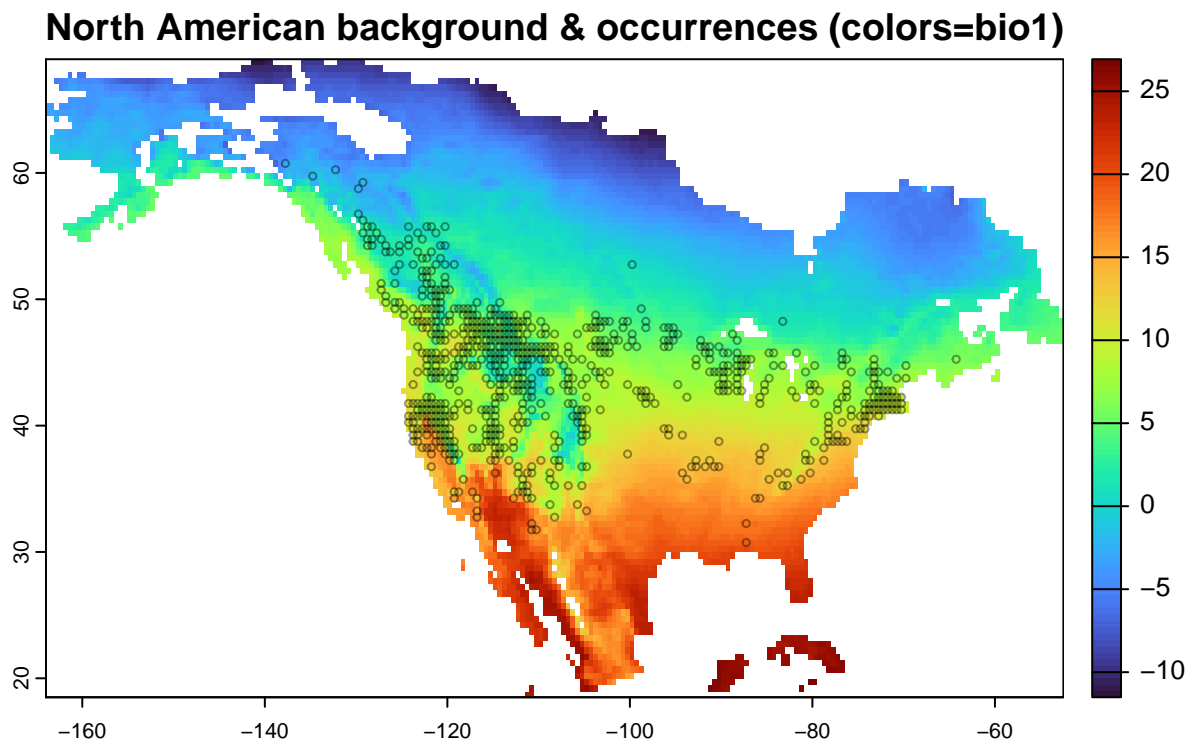
ii. Niche quantification

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.

```r
# 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)
```


axis1 = 37.99 % axis2 = 34.88 %
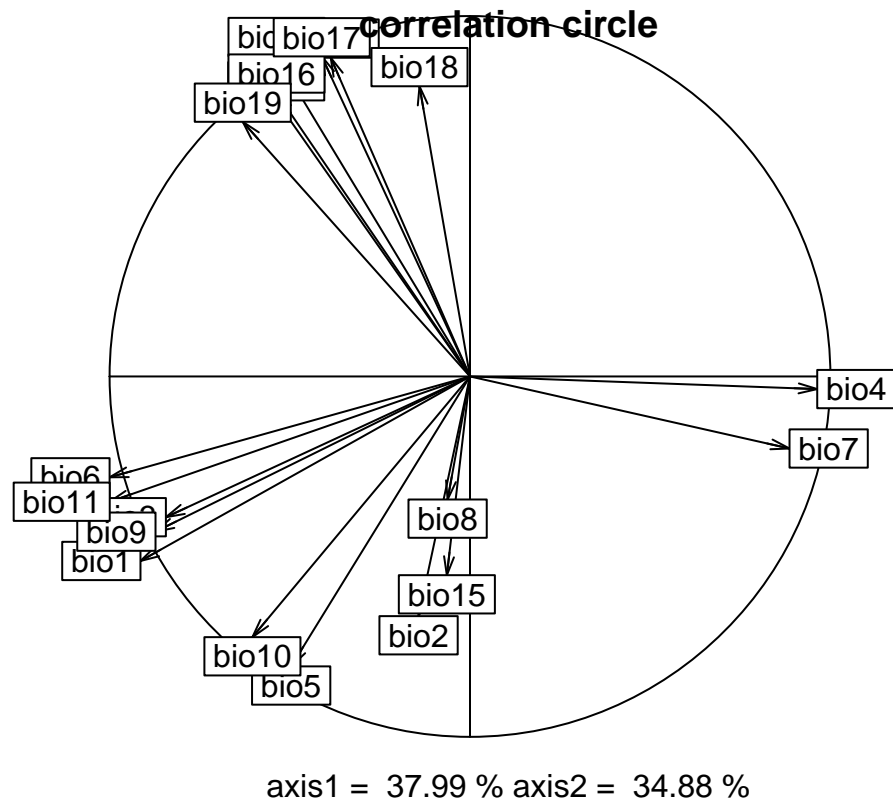
```r
## PCA scores

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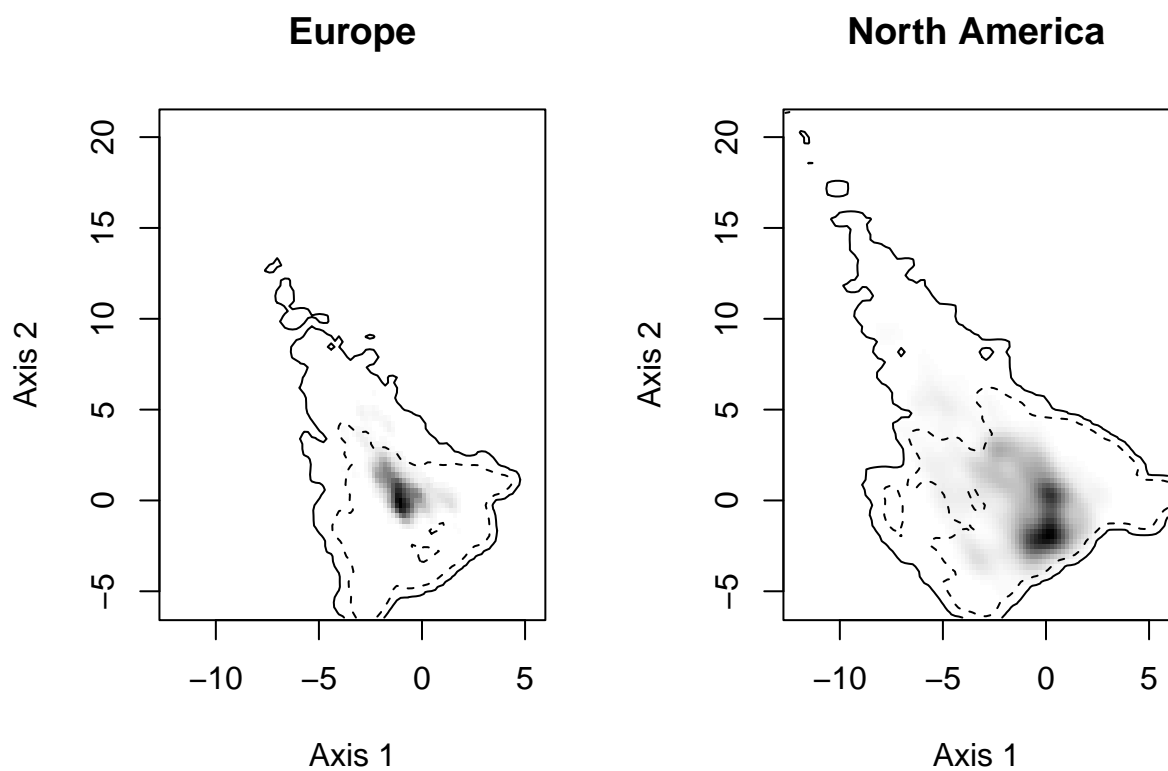
```r
plot(z.EU$z.uncor,col=viridis(100,option="magma",direction = -1),main="Europe")
plot(z.NAM$z.uncor,col=viridis(100,option="magma",direction = -1),main="North America")
```
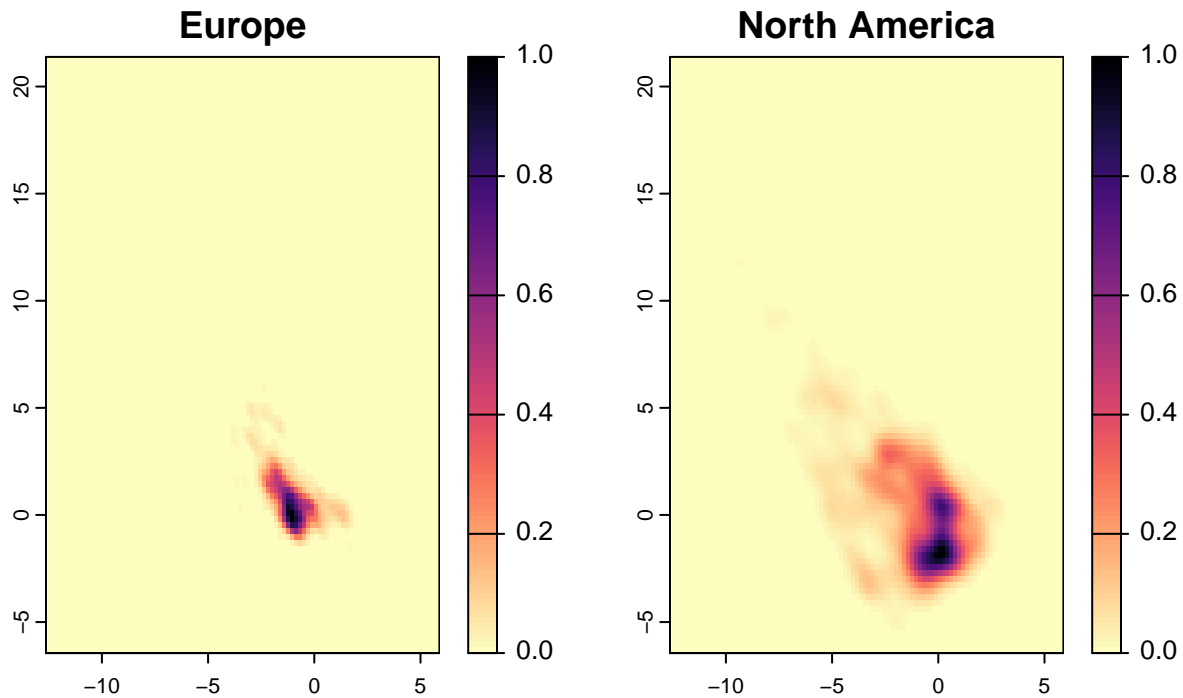
```
par(mfrow=c(1,1))
```

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
```

Now 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 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 2008), but the niche equivalency test almost always provides significant results even when the overlap is very low and is thus in general not informative (**collart2020?**; Peterson 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 invasive range. This can be done with the e*xpansion* , *stability* and *unfilling* indices returned by the *ecospat.niche.dyn.index* function. *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 e*xpansion*, *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)
```
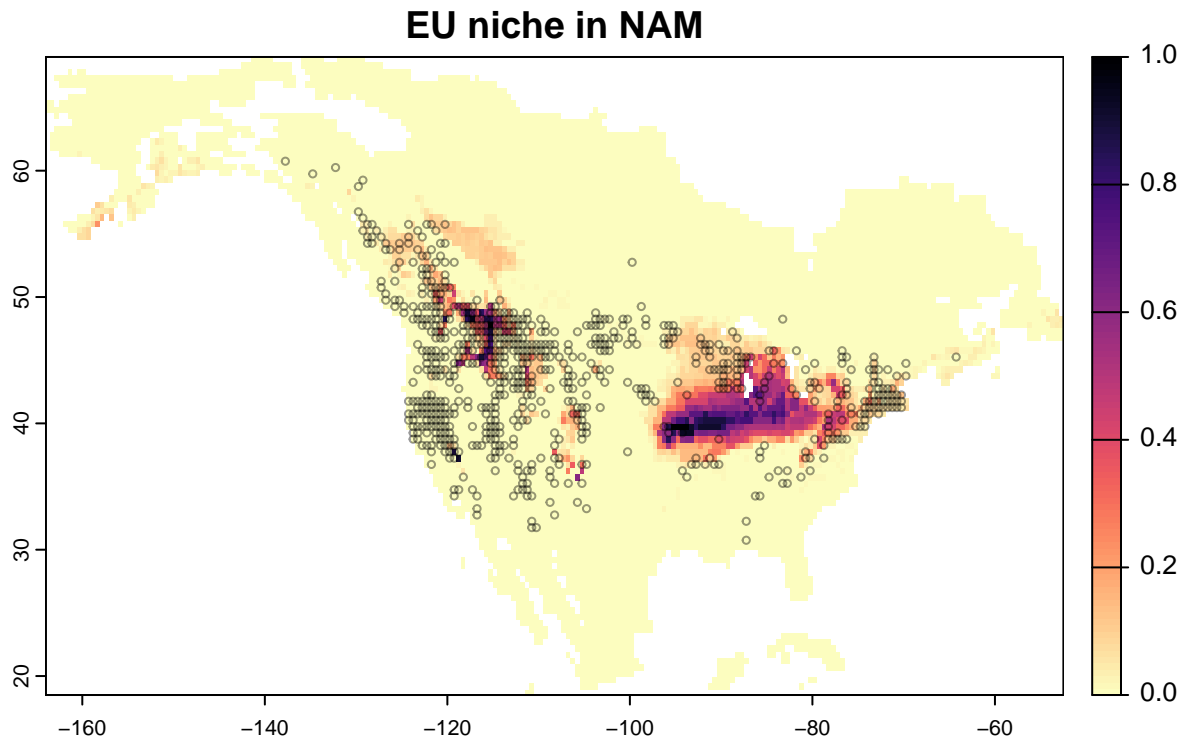
niche dynamics
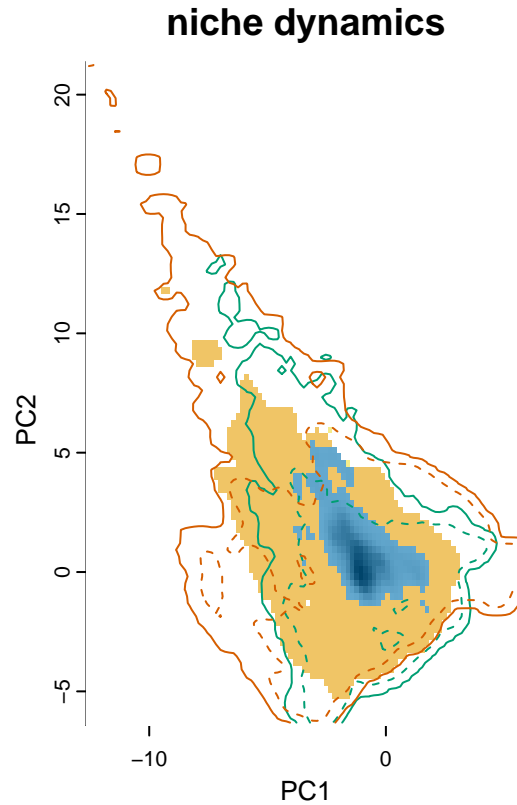
Finally, we might wonder where are the areas predicted with a high density of occurrence by the niche quantification models. Indeed, every pixel in the geography corresponds to a predicted density of occurrence in the environmental space. The function *ecospat.niche.zProjGeo* allows to do that. It is 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.

```
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)
```
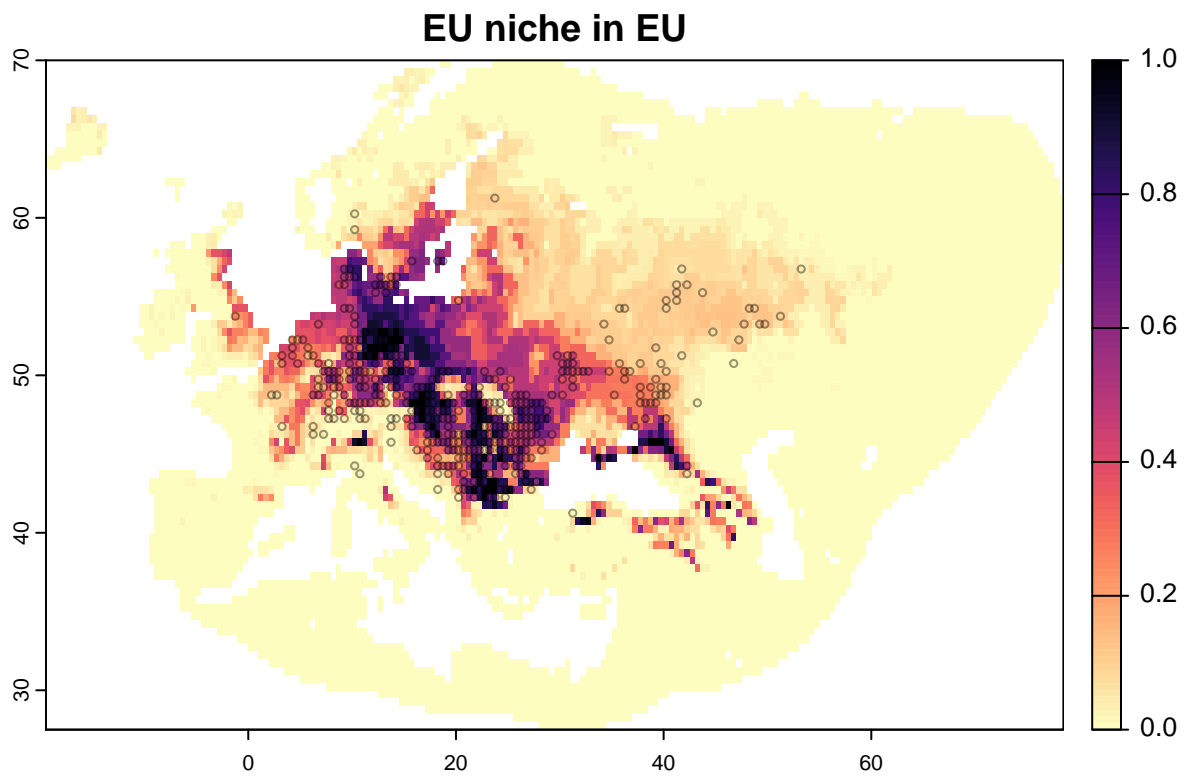
EU niche in EU


```
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)
```

EU niche in NAM

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 conrresponding 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

- outside both niches
- niche unfilling
- niche expansion
- niche stability
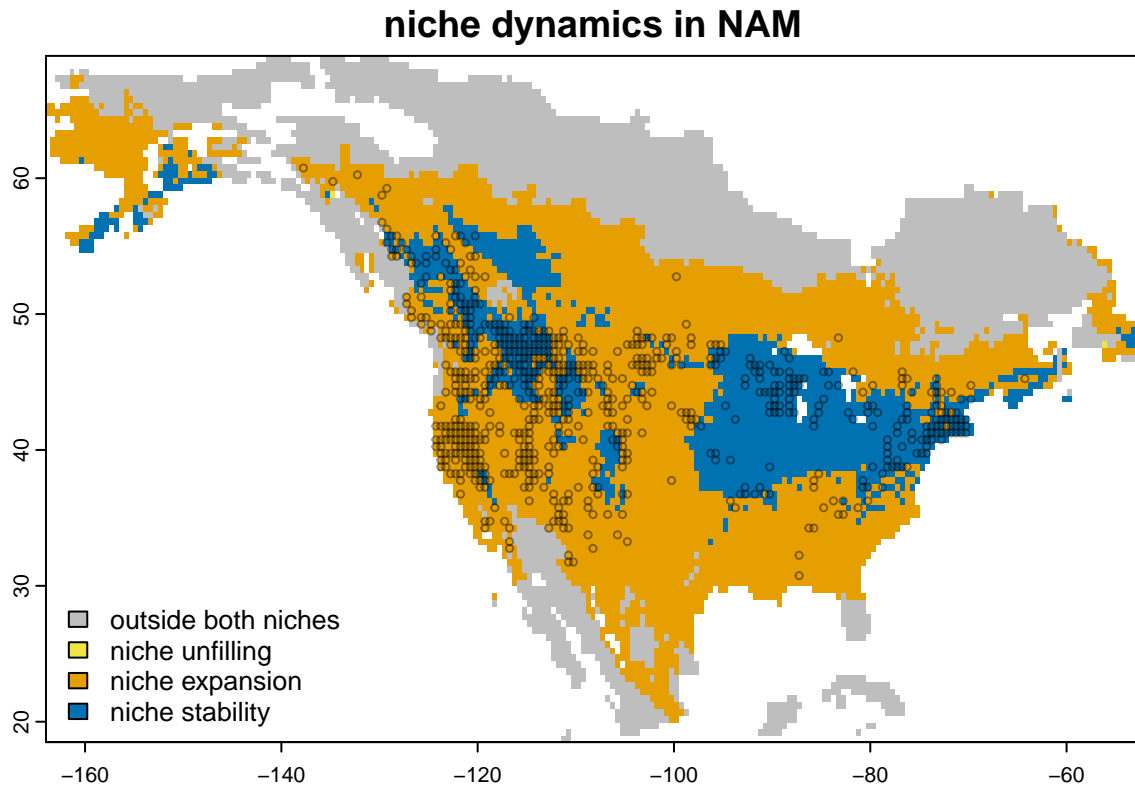
3. Conclusion and perspectives

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

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.cor	Investigates spatial autocorrelation of environmental covariables within a set of occurrences as a function of distance.
	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 (K. Elith J. 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 R × 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 an equivalency test (Warren 2008)based on two species occurrence density grids.
		ecospat.niche.similarity.test	Runs the similarity test (Warren 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.env	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	Spatial predictions, projections of rare species	ecospat.ESM.Modeling	Calibrates simple bivariate models as in Lomba (2010) and A. Breiner F. T. (2015)
		ecospat.ESM.EnsembleModeling	Evaluate 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, projections	ecospat.SESAM.prr	Implements the SESAM framework to predict community composition using a 'probability ranking' rule.

	Model evaluation	ecospat.CCV.modeling	Creates probabilistic prediction for all species based on SDMs or ESMs
		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 (Hirzel 2006)
		ecospat.CommunityEval	Evaluates 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 [@collart2023]
		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
		ecospat.plot.kappa	Plots kappa value in function of the threshold value.
		ecospat.plot.tss	plot TSS value in function of the threshold value.
		ecospat.poolingEvaluation	Uses the pooling evaluation to evaluate any models
Post modeling	Variance partition	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.occupiedpatch	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
Other functions	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.
		ecospat.Cscore	Computes and tests if C-scores > null model
	Phylogenetic diversity measures	ecospat.calculate.pd	Calculates phylogenetic diversity measures (Schweiger 2008)

4. References

Araújo, Anderson, M. B. 2019. "Standards for Distribution Models in Biodiversity Assessments." *Science Advances* 5 (1). https://doi.org/10.1126/sciadv.aat4858.

Araújo, Guisan, M. B. 2006. "Five (or so) Challenges for Species Distribution Modelling." *Journal of Biogeography* 33 (10): 1677–88. https://doi.org/10.1111/j.1365-2699.2006.01584.x.

Austin, Nicholls, M. P. 1990. "Measurement of the Realized Qualitative Niche: Environmental Niches of Five Eucalyptus Species." *Ecological Monographs* 60 (2): 161–77. https://doi.org/10.2307/1943043.

Barve, N., V. Barve, A. Jimenez-Valverde, A. Lira-Noriega, S. P. Maher, A. T. Peterson, J. Soberon, and F. Villalobos. 2011. "The Crucial Role of the Accessible Area in Ecological Niche Modeling and Species Distribution Modeling." *Ecological Modelling* 222 (11): 1810–19. https://doi.org/10.1016/j.ecolmodel.2011.02.011.

Breiner, A., F. T. 2015. "Overcoming Limitations of Modelling Rare Species by Using Ensembles of Small Models." *Methods in Ecology and Evolution* 6 (10): 1210–18.

Breiner, M. P., F. T. 2018. "Optimizing Ensembles of Small Models for Predicting the Distribution of Species with Few Occurrences." Edited by Nick Isaac. *Methods in Ecology and Evolution* 9 (4): 802–8. https://doi.org/10.1111/2041-210x.12957.

Broennimann, Fitzpatrick, O. 2012. "Measuring Ecological Niche Overlap from Occurrence and Spatial Environmental Data." *Global Ecology and Biogeography* 21 (4): 481–97. https://doi.org/10.1111/j.1466-8238.2011.00698.x.

Broennimann, Thuiller, O. 2006. "Do Geographic Distribution, Niche Property and Life Form Explain Plants' Vulnerability to Global Change?" *Global Change Biology* 12 (6): 1079–93. https://doi.org/10.1111/j.1365-2486.2006.01157.x.

Broennimann, U. A., O. Treier. 2007. "Evidence of Climatic Niche Shift During Biological Invasion." *Ecology Letters* 10 (8): 701–9. https://doi.org/10.1111/j.1461-0248.2007.01060.x.

Collart, Guisan, F. 2023. "Small to Train, Small to Test: Dealing with Low Sample Size in Model Evaluation." *Ecological Informatics* 75 (July): 102106. https://doi.org/10.1016/j.ecoinf.2023.102106.

D'Amen, Rahbek, M. 2015. "Spatial Predictions at the Community Level: From Current Approaches to Future Frameworks." *Biological Reviews* 92 (1): 169–87. https://doi.org/10.1111/brv.12222.

Di Cola, Broennimann, V. 2017. "Ecospat: An R Package to Support Spatial Analyses and Modeling of Species Niches and Distributions." *Ecography* 40 (6): 774–87. https://doi.org/10.1111/ecog.02671.

Dubuis, Pottier, A. 2011. "Predicting Spatial Patterns of Plant Species Richness: A Comparison of Direct Macroecological and Species Stacking Modelling Approaches." *Diversity and Distributions* 17 (6): 1122–31. https://doi.org/10.1111/j.1472-4642.2011.00792.x.

Elith, Ferrier, J. 2005. "The Evaluation Strip: New and Robust Method for Plotting Predicted Responses from Species Distribution Models." *Ecological Modelling* 186 (3): 280–89. https://doi.org/10.1016/j.ecolmodel.2004.12.007.

Elith, Kearney, J. 2010. "The Art of Modelling Range-Shifting Species." *Methods in Ecology and Evolution* 1 (4): 330–42. https://doi.org/10.1111/j.2041-210x.2010.00036.x.

Ferrier, Guisan, S. 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.

Franklin, J. 2010. "Moving Beyond Static Species Distribution Models in Support of Conservation Biogeography." *Diversity and Distributions* 16 (3): 321–30. https://doi.org/10.1111/j.1472-4642.2010.00641.x.

Guisan, Broennimann, A. 2006. "Using Niche-Based Models to Improve the Sampling of Rare Species." *Conservation Biology* 20 (2): 501–11. https://doi.org/10.1111/j.1523-1739.2006.00354.x.

Guisan, Petitpierre, A. 2014. "Unifying Niche Shift Studies: Insights from Biological Invasions." *Trends in Ecology & Evolution* 29 (5): 260–69. https://doi.org/10.1016/j.tree.2014.02.009.

Guisan, Rahbek, A. 2011. "SESAM - a New Framework Integrating Macroecological and Species Distribution Models for Predicting Spatio-Temporal Patterns of Species Assemblages." *Journal of Biogeography* 38 (8): 1433–44. https://doi.org/10.1111/j.1365-2699.2011.02550.x.

Guisan, Tingley, A. 2013. "Predicting Species Distributions for Conservation Decisions." Edited by Hector Arita. *Ecology Letters* 16 (12): 1424–35. https://doi.org/10.1111/ele.12189.

Guisan, W., A. 2017. *Habitat Suitability and Distribution Models, with Applications in r.*

Hirzel, Le Lay, A. H. 2006. "Evaluating the Ability of Habitat Suitability Models to Predict Species Presences." *Ecological Modelling* 199 (2): 142–52. https://doi.org/10.1016/j.ecolmodel.2006.05.017.

Jiménez-Valverde, A. 2020. "Sample Size for the Evaluation of Presence-Absence Models." *Ecological Indicators* 114 (July): 106289. https://doi.org/10.1016/j.ecolind.2020.106289.

Lomba, L., A. 2010. "Overcoming the Rare Species Modelling Paradox: A Novel Hierarchical Framework

Applied to an Iberian Endemic Plant." *Biological Conservation* 143 (11): 2647–57. https://doi.org/10.1016/j.biocon.2010.07.007.

Mráz, Bourchier, P. 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.

Peterson, Soberón, A. T. 2011. *Ecological Niches and Geographic Distributions.* Princeton: Princeton University Press.

Petitpierre, Kueffer, B. 2012. "Climatic Niche Shifts Are Rare Among Terrestrial Plant Invaders." *Science* 335 (6074): 1344–48. https://doi.org/10.1126/science.1215933.

Pulliam, H. R. 2000. "On the Relationship Between Niche and Distribution." *Ecology Letters* 3 (4): 349–61. https://doi.org/10.1046/j.1461-0248.2000.00143.x.

Schweiger, Klotz, O. 2008. "A Comparative Test of Phylogenetic Diversity Indices." *Oecologia* 157 (3): 485–95. https://doi.org/10.1007/s00442-008-1082-2.

Shen, Song, T. 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): 880–91. https://doi.org/10.1111/plb.13570.

Sillero, Campos, N. 2023. "A Curated List of R Packages for Ecological Niche Modelling." *Ecological Modelling* 476 (February): 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): 1115–23. https://doi.org/10.1111/j.1461-0248.2007.01107.x.

Treier, Broennimann, U. A. 2009. "Shift in Cytotype Frequency and Niche Space in the Invasive Plant *Centaurea Maculosa*." *Ecology* 90 (5): 1366–77. https://doi.org/10.1890/08-0420.1.

van Proosdij, M. S. M., A. S. J. 2015. "Minimum Required Number of Specimen Records to Develop Accurate Species Distribution Models." *Ecography* 39 (6): 542–52. https://doi.org/10.1111/ecog.01509.

Warren, Glor, D. L. 2008. "Environmental Niche Equivalency Versus Conservatism: Quantitative Approaches to Niche Evolution." *Evolution* 62 (11): 2868–83. ://000260348100012.

Zurell, Franklin, D. 2020. "A Standard Protocol for Reporting Species Distribution Models." *Ecography* 43 (9): 1261–77. https://doi.org/10.1111/ecog.04960.