

Tutorial for SDMPlay: 5/ Model extrapolation

2021-02-26

Species distribution modelling (SDM) has been developed for several years to address conservation issues, to assess the direct impact of human activities on ecosystems and to predict the potential distribution shifts of invasive species (see Elith et al. 2006, Pearson 2007, Elith and Leathwick 2009). SDM relates species occurrences with environmental information and can predict species distribution on their entire occupied space. **This approach has been increasingly applied to Southern Ocean case studies, but requires corrections in such a context, due to the broad scale area, the limited number of presence records available and the spatial and temporal aggregations of these datasets.**

SDMPlay is a pedagogic package that will allow you to compute SDMs, to understand the overall method, and to produce model outputs. The package, along with its associated vignettes, highlights the different steps of model calibration and describes how to choose the best method to generate accurate and relevant outputs. SDMPlay proposes codes to apply a popular machine learning approach, BRT (Boosted Regression Trees) and introduces MaxEnt (Maximum Entropy). It contains occurrences of marine species and environmental descriptor datasets as examples associated with several vignette tutorials.

Objectives of tutorial #5/ Spatial extrapolation

Considering the reference dataset of environmental conditions for which species presence-only records are modelled, extrapolation corresponds to the part of the projection area for which one environmental value at least falls outside of the reference dataset. Because occurrence sampling is limited in the Southern Ocean and that environmental conditions are contrasting in this broad scale area, it was observed that model predictions are actually extrapolation in proportions up to 75% in some cases ! (Guillaumot et al. 2020). It is therefore important to measure extrapolation and to provide extrapolation uncertainty maps along with model predictions. Here, we used the Multivariate Environmental Similarity Surface (MESS) index to quantify model uncertainty associated to extrapolation (Elith et al. 2010), but other approaches exist (Owens et al. 2013).

See also...

- **Tutorial #1/ Compute Species Distribution Models**
Focusses on data structure, data preparation and general model computing.
- **Tutorial #2/ SDM outputs**
Presents the main outputs you can generate with your SDM.
- **Tutorial #3/ Importance of model calibration**
Highlights the procedure to accurately calibrate your model and proposes some methods to limit the influence of several biases.
- **Tutorial #4/ Spatial cross-validation**
Cross-validation is a method to validate your model. When working with presence data spatially aggregated, the cross-validation procedure should be adapted. This tutorial provides some elements to apply this method (referring to Guillaumot et al. 2019).

Let's begin ! Evaluate extrapolation of your case study!

We will work with the *Ctenocidaris nutrix* case study. Load the occurrence records and environmental layers.

```
library(SDMPplay)
library(raster)
data("ctenocidaris.nutrix")
data(predictors2005_2012)
ctenocidaris.nutrix.occ <- ctenocidaris.nutrix[,c(7,8)]

# we will just keep 3 descriptors for the example
predictors2005_2012 <- raster::subset(predictors2005_2012,c(1:3))
```

```
library(RColorBrewer)
data("worldmap")
my.palette.oranges <- brewer.pal(n = 9, name = "Oranges")
my.palette.blue <- rev(brewer.pal(n = 9, name = "Blues"))
```

```
library(dismo)
# Measure the extrapolation score (MESS: Multivariate Environmental Similarity Surface)

# Extract environmental values at presence-only records location
envi_scores <- raster::extract(predictors2005_2012, ctenocidaris.nutrix.occ)
#head(envi_scores)

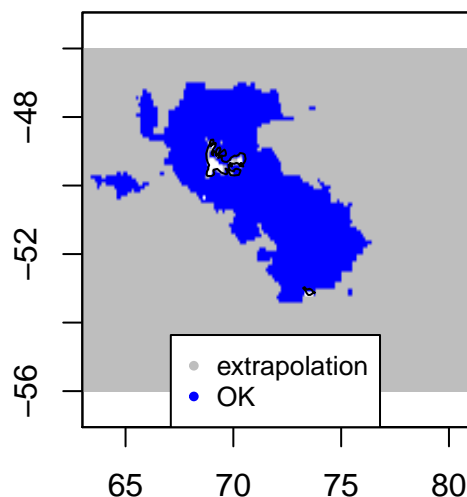
# Calculate MESS score
x <- dismo::mess(predictors2005_2012,envi_scores)

y <- x; values(y)<- values(x)>0
y <- reclassify(y,cbind(FALSE,0)) # MODEL EXTRAPOLATES (MESS <0)
y <- reclassify(y,cbind(TRUE,1)) # MODEL DOES NOT EXTRAPOLATE (MESS >0)

MESS_layer <- mask(y,subset(predictors2005_2012,1)) # remove land pixel (layer #1 = depth)
```

MESS_layer is a RasterLayer with values contained between 0 and 1. The model extrapolates for 0 and does not extrapolate for 1 value.

```
# Plot the result
plot (MESS_layer, col=c("grey","blue"), legend=F)
points(worldmap, type="l")
legend("bottom", legend=c("extrapolation","OK"),
      col=c("grey","blue"),
      pch=20, bg = "white", cex=0.8)
```



You can then overlap this layer to SDM predictions when interpreting your results or preparing your maps for your publications! See examples in Guillaumot et al. (2020).

```
# Calculate the proportion of the area where extrapolation occurs
MESS<- reclassify(MESS_layer,cbind(1,NA))

# compare the number of pixels = 0 to the number of total pixels of the area
length(which(!is.na(values(MESS))))*100 / length(which(!is.na(values(subset(predictors2005_2012,1)))))

## [1] 82.89967
```

Assess which environmental descriptors are responsible for extrapolation at each pixel

```
# create an empty raster to initiate a Rasterstack
stack_amelio_MESS <- subset(predictors2005_2012,1); values(stack_amelio_MESS) <- NA

# Loop to calculate the value of dissimilarity of each environmental descriptor
# For each pixel, it will be determined if extrapolation occurs for each environmental descriptor

for (k in 1:nlayers(predictors2005_2012)){
  presvals <- raster::extract(subset(predictors2005_2012, k),
                              ctenocidaris.nutrix.occ)
  x_amelio <- dismo::mess(subset(predictors2005_2012, k),presvals)
  stack_amelio_MESS <- stack(stack_amelio_MESS,x_amelio)
}
```

```

# Delete the first layer of the stack that was empty (initialization)
stack_amelio_MESS <- dropLayer(stack_amelio_MESS,1)
names(stack_amelio_MESS) <- names(predictors2005_2012)

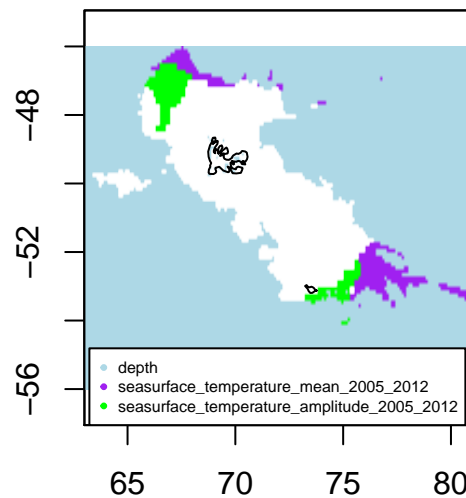
# Search for the environmental layer that is responsible for the lower MESS score
# (i.e. responsible for the extrapolation)

MESS_amelio <- which.min(stack_amelio_MESS)
MESS_amelio <- mask(MESS_amelio, MESS) # keep only areas where extrapolation occurs

# Plot the result

plot (MESS_amelio, col=c("lightblue","purple", "green"), legend=F)
points(worldmap, type="l")
legend("bottom", legend=names(predictors2005_2012),
      col=c("lightblue","purple", "green"),
      pch=20, bg = "white", cex=0.5)

```



```

# Calculate the contribution of each environmental descriptor in extrapolation
table_mess_amelio <- matrix(data=NA, nrow = 1, ncol= nlayers(predictors2005_2012))
colnames(table_mess_amelio) <-c("depth", "mean_temp", "amplitude_temp")

for (k2 in 1:nlayers(predictors2005_2012)){
  table_mess_amelio[1,k2] <- length(which(values(MESS_amelio)==k2))*100/
    length(which(!is.na(values(subset(predictors2005_2012,1)))))
}

table_mess_amelio

```

```
##           depth mean_temp amplitude_temp
## [1,] 78.68054  2.344581      1.874545
```

REFERENCES

- Elith, J., Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J Hijmans, R., Huettmann, F., ... & A Loiselle, B. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129-151.
- Elith, J. & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677-697.
- Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1(4), 330-342.
- Guillaumot, C., Artois, J., Saucède, T., Demoustier, L., Moreau, C., Eléaume, M., ... & Danis, B. (2019). Broad-scale species distribution models applied to data-poor areas. *Progress in Oceanography*, 175, 198-207.
- Guillaumot, C., Moreau, C., Danis, B. & Saucède, T. (2020). Extrapolation in species distribution modelling. Application to Southern Ocean marine species. *Progress in Oceanography*, 188, 102438.
- Owens, H. L., Campbell, L. P., Dornak, L. L., Saupe, E. E., Barve, N., Soberón, J., ... & Peterson, A. T. (2013). Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, 263, 10-18.
- Pearson, R.G. (2007). Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*, 50.