

Ecological Niche in R-GRASS (ENiRG)

F. Canovas, C. Magliozzi, F. Mestre,
J.A. Palazon & M. Gonzalez-Wanguemert

April 30, 2015

1 ENFA analysis using the command line

1.1 Data preparation

Reading the package.

```
> library(ENiRG)
```

Importing species occurrence data from a data R file. Usually you will have a file with point locality data representing the known distribution of the studied species. Below is an example of the *Apis mellifera* records stored in the list of class *enirg* (Canovas *et al.*, 2002, 2008, 2014).

```
> data(apis.enirg)
> lina <- apis.enirg$presences
> head(lina)
```

	Long	Lat	Lina
SXL	-8.08555	41.90560	1
DJ	-3.31659	38.71414	2
BUR	-8.55524	42.18557	2
AYO	-1.05745	39.06090	2
MOn	-0.61467	39.33986	2
XAB	0.10622	38.76471	3

The ecogeographical variables should be stored in raster format. Almost all relevant formats can be used (check [GDAL library](#) for compatibility). Qualitative and quantitative predictors should represent variables of interest for the species. Below an example of importing a raster map. If not specified,

the output name of the uploaded file is set to the original one as default. Maps are automatically uploaded to a GRASS session, which should be initiated.

Users can check metadata regarding the predictors and summarize the current settings of GRASS environment. The user can select the preferred format between the tabular of GRASS or the more terse R output.

```
> map.info(map.name="tann", format="R")
```

A list of available maps can be imported, also searching for a particular pattern.

```
> list.maps(type="rast")
```

```
$rast
 [1] "African"           "African_hsm"       "African_hsm_class"
 [4] "African_pred"     "li_Mar"            "li_Spec1"
 [7] "li_Spec2"         "mntcm"             "mxtwm"
[10] "pond_std_mntcm"   "pond_std_mxtwm"    "pond_std_rfann"
[13] "pond_std_rfdm"    "pond_std_rfwm"     "pond_std_tann"
[16] "rfann"            "rfdm"              "rfwm"
[19] "std_mntcm"        "std_mxtwm"         "std_rfann"
[22] "std_rfdm"         "std_rfwm"          "std_tann"
[25] "tann"             "temp"              "test"
[28] "test.hsm.class"   "test_hsm"          "test_pred"
```

When making predictions, standardization of *egv* can be advantageous, finding the maximum likelihood estimates. If not specified, the output name of the standardized file will contain the prefix "std_" plus the original name. More than one map can be standardized at the time.

1.2 Configuring and running ENFA analysis

The names for the ecogeographical variables should be parsed throughout the main function *enirg* in vectorial format. In the following example the standardized map are used to perform the analysis.¹

```
> # loading the environmental information in batch
> predictor.names <- c("tann", "mxtwm", "mntcm",
>                    "rfann", "rfdm", "rfwm")
> predictor.maps <- paste("std_", predictor.names, sep="")
```

¹NOTE: The names of the raster type files must correspond to the names of the ecogeographical variables already loaded into GRASS.

```

> file.names <- paste(system.file(package = "ENiRG"),
>                     "/ext/", predictor.names,
>                     ".asc", sep="")

> import.egvs(file.names, predictor.names)

> # standardization
> stdz.maps(predictor.names, predictor.maps)

> # performing the Ecological Niche Factor Analysis (ENFA)
> enirg(presences.table = lina, qtegv.maps = predictor.maps,
>       species.name = "African", nf = 1, scannf = FALSE,
>       method = "normal") -> apis.enirg

```

After setting predictor variables and occurrence points, the analysis can be run. You can use `scannf=TRUE` directly on R console to interactively choose the number of factors. NOTE: set `load.map=FALSE` and `method="large"` to perform the calculations over large data sets.

Results are returned into a list object.

```

> str(apis.enirg)

```

List of 17

```

$ call           : language enirg(presences.table = lina, ...
$ nf            : num 1
$ cw            : num [1:6] 1 1 1 1 1 1
$ species       : chr "African"
$ egvs          : chr [1:6] "std_tann" ...
$ qt.egvs       : chr [1:6] "std_tann" ...
$ ql.egvs       : NULL
$ presences     : 'data.frame': 135 obs. of  3 variables:
..$ Long: num [1:135] -8.086 -3.317 -8.555 -1.057 -0.615 ...
..$ Lat : num [1:135] 41.9 38.7 42.2 39.1 39.3 ...
..$ Lina: int [1:135] 1 2 2 2 2 3 3 3 3 3 ...
$ total.marginality : num 0.51
$ marginalities   : num [1:6, 1] 0.603 0.413 0.422 ...
..- attr(*, "dimnames")=List of 2
.. ..$ : chr [1:6] "std_tann" ...
.. ..$ : chr "Marginality"
$ total.specialization: num 0.483
$ specializations  : num [1:6, 1] 0.476 -0.056 -0.13 ...
..- attr(*, "dimnames")=List of 2

```

```

.. ..$ : chr [1:6] "std_tann" ...
.. ..$ : chr "Spec1"
$ co          : num [1:6, 1:2] 0.603 0.413 ...
..- attr(*, "dimnames")=List of 2
.. ..$ : chr [1:6] "std_tann" ...
.. ..$ : chr [1:2] "Mar" "Spec1"
$ mar        : Named num [1:6] 0.833 0.57 0.582 ...
..- attr(*, "names")= chr [1:6] "std_tann" ...
$ m          : num 1.9
$ s          : num [1:5] 3.863 2.123 1.076 0.762 0.562
$ obs.li     : 'data.frame': 135 obs. of 4 variables:
..$ cat      : int [1:135] 1 2 3 4 5 6 7 8 9 10 ...
..$ presences: int [1:135] 1 2 2 2 2 3 3 3 3 3 ...
..$ li_Mar   : num [1:135] -1.267 1.205 -0.84 0.768 1.634 ...
..$ li_Spec1 : num [1:135] 0.2999 -0.0134 0.0892 ...
- attr(*, "class")= chr "enirg"

```

The user is able to plot the results of the factor analysis throughout the function *enirg.plot*. (Figure 1). NOTE: for large data set, method should be set to "simplified" (Figure 1).

```
> enirg.plot(apis.enirg, method="extended")
```

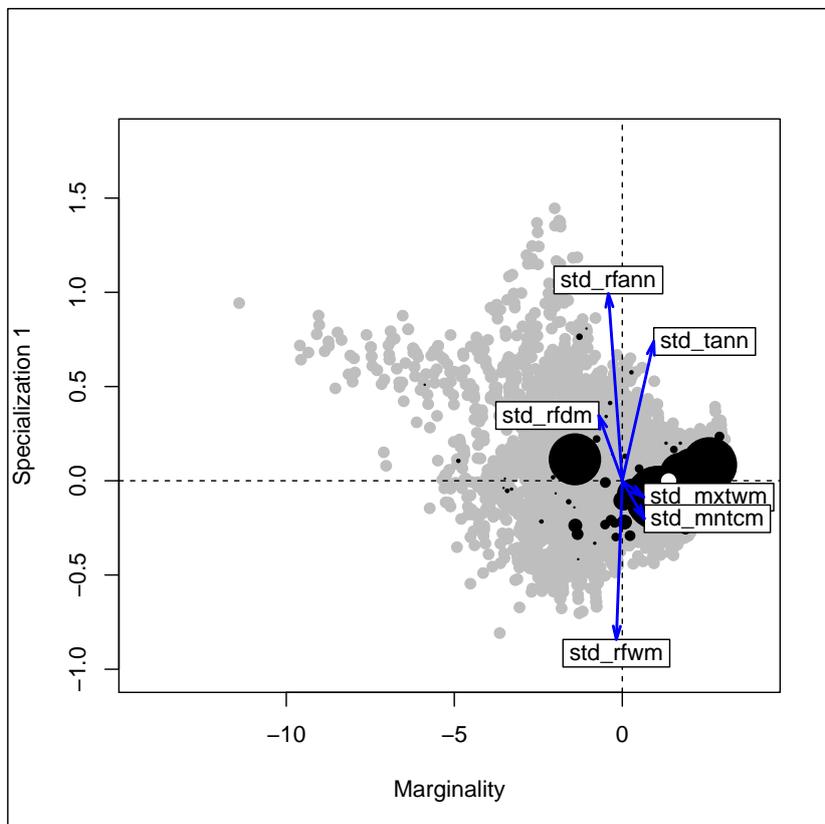


Figure 1: ENFA ordination as obtained by setting method="extended".

"li" maps for marginality and specialization/s can be plotted into R using *raster* package (Figure 2). NOTE: not recommended for large data sets. Maps can be then viewed and managed through GRASS and QGIS.

```
> require(raster)
> plot(raster(readRAST6("li_Mar")))
```

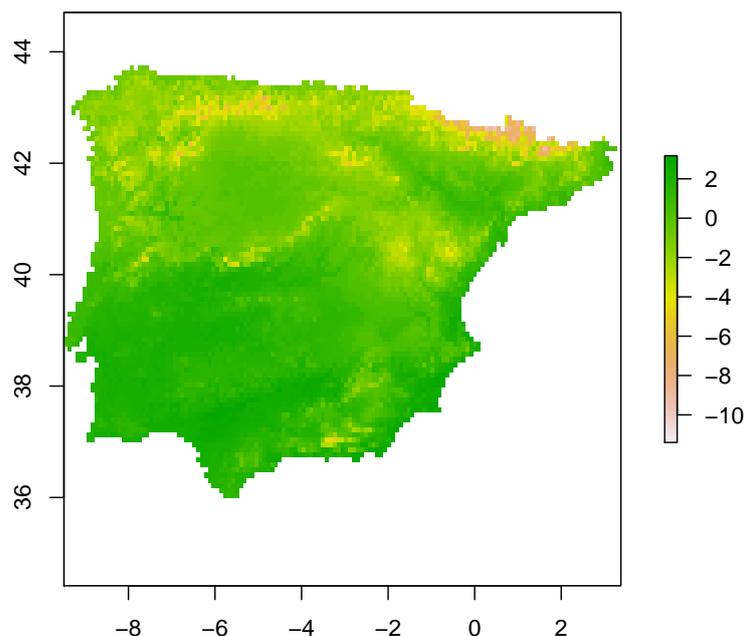


Figure 2: Marginality factor.

Habitat suitability map can be calculated using *enirg.predict*. If the user does not provide a new set of maps, the calculations will be done over the same ecogeographical variables, which were used to run *enirg* function.

```
> enirg.predict(apis.enirg, load.map = TRUE,
>               method = "normal") -> apis.hsm
```

To perform Boyce continuous classification index, two vector objects are needed. First, a vector containing the predictions for our observations. It can be extracted from the list generated using *predict.enirg()*

```
> data.pred.obs <- apis.hsm$predictions$predicted
```

Second, a vector containing all the prediction values for the entire area of study. It can be generated as follow:

```
> data.pred.map <- apis.hsm$African_hsm@data@values
```

, without non data values:

```
> data.pred.map <- data.pred.map[which(data.pred.map > 0)]
```

Finally, running the manual classification. NOTE: Try with the following intervals for the best classification fit:

- unsuitable = 0.25
- marginal = 0.69
- suitable = 0.81

```
> boyce(data.pred.obs, data.pred.map)
```

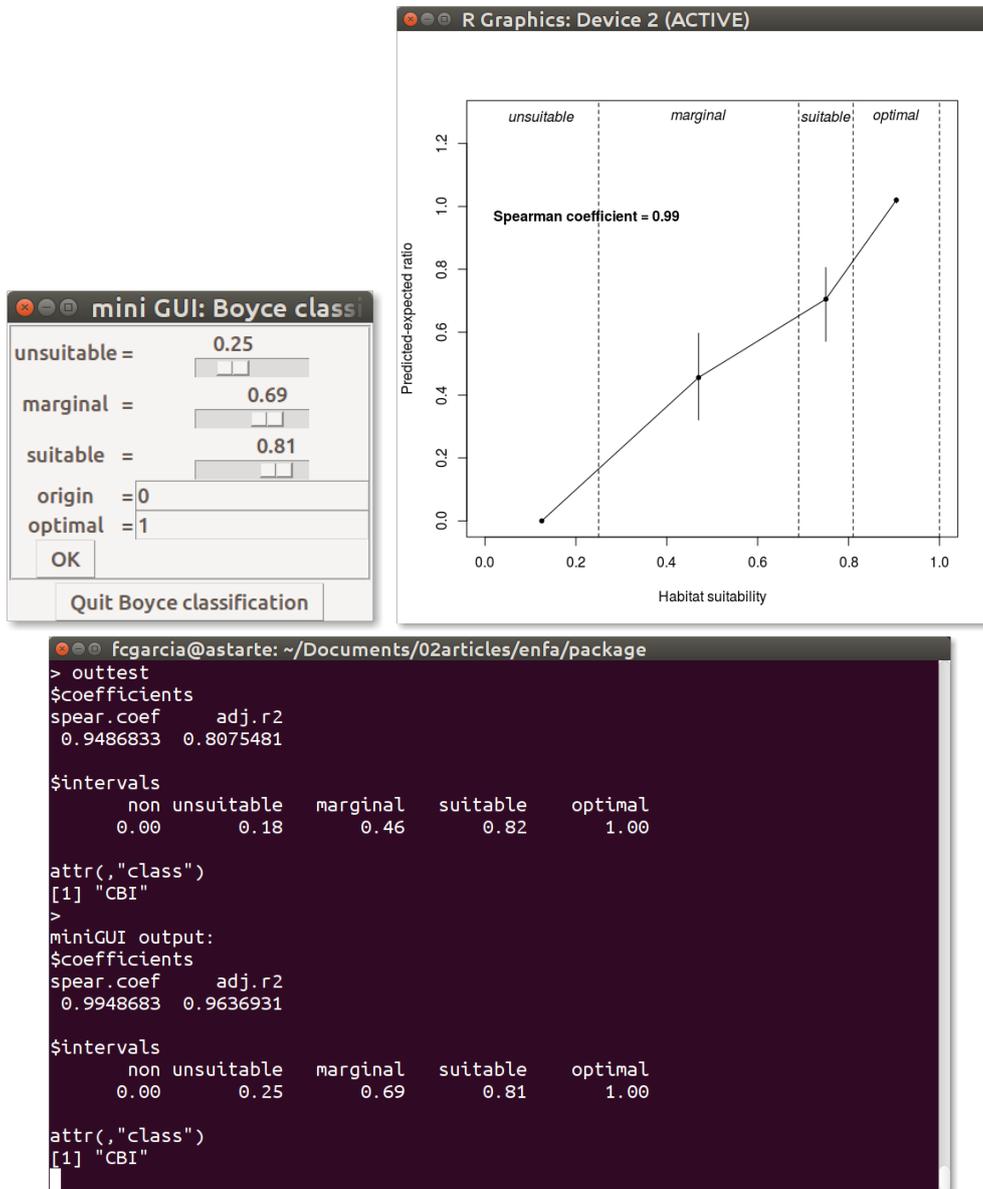


Figure 3: Classification process.

Now, the user is able to reclassify the HSM map into the four suitability classes according to (Boyce *et al.*, 2002) (unsuitable, marginal, suitable, optimal).

```

> data(cbi.results)
> classify.map(apis.enirg, cbi.results,

```

```

+           output.name="African_hsm_class",
+           load.map = TRUE) -> apis.hsm.class

unsuitable  marginal  suitable
      0.25      0.69      0.81
0.750000 rounded up to 0
0.750000 rounded up to 0
1.190000 rounded up to 1
1.190000 rounded up to 1
1.310000 rounded up to 1
1.310000 rounded up to 1
African_hsm_class raster map was sucessfully classify!

```

The HSM can be plotted into R using *raster* package (Figure 4). NOTE: maps can be also viewed and managed through GRASS and QGIS.

```

> plot(apis.hsm.class[[1]], legend = FALSE,
+      col = c("red", "yellow", "blue", "green"))
> contour(apis.hsm.class[[1]], add=T)
> legend("bottomright",
+       legend = c("unsuitable", "marginal", "suitable", "optimal"),
+       fill = c("red", "yellow", "blue", "green"))

```

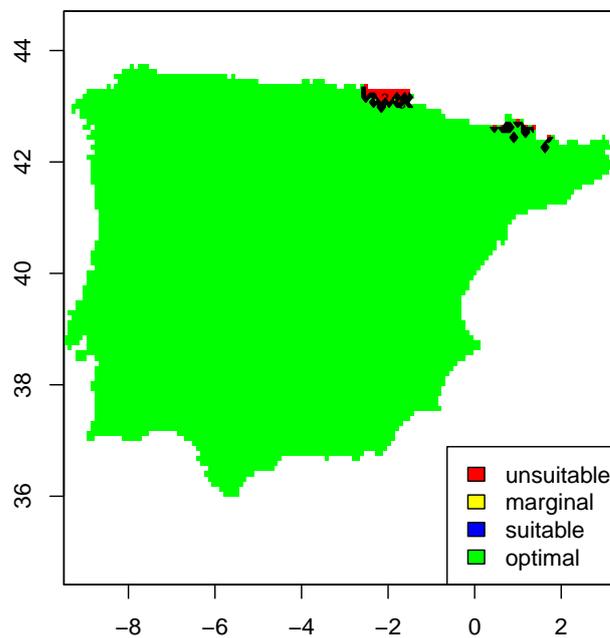


Figure 4: Habitat suitability map.

2 Important note

For Unix-based systems, *method="large"* option is also available, when using ENiRG for very large data sets. To use that functionality, the user should start GRASS in a appropriate mapset and then start R. Methods are the same as the ones used in this vignette.

3 References

References

- Boyce, M.S., Vernier, P.R., Nielsen, S.E., Schmiegelow, F.K.A. (2002). Evaluating resource selection functions. *Ecological Modelling* 157, 281-300.
- Canovas, F., De la Rúa, P. Serrano, J. and Galian, J. (2002). Mitochondrial DNA variability in *Apis mellifera iberica* populations from Galicia. *Archivos de Zootecnia*, 51: 441-448
- Canovas, F., De la Rúa, P. Serrano, J. and Galian, J. (2008). Geographical patterns of mitochondrial DNA variation in *Apis mellifera iberiensis* (Hymenoptera: Apidae). *Journal of Zoological Systematics and Evolutionary Research*, 46(1): 23-30
- Canovas, F., De la Rúa, P. Serrano, J. and Galian, J. (2014). Analysing and describing a contact area of two distinct evolutionary units. An ecological perspective. *Journal of Insect Conservation*. DOI: 10.1007/s10841-014-9701-1.