

# SPEDInstabR: An algorithm based on a fluctuation index for selecting predictors in species distribution modeling



Cástor Guisande <sup>a,\*</sup>, Emilio García-Roselló <sup>b</sup>, Jürgen Heine <sup>b</sup>, Jacinto González-Dacosta <sup>b</sup>, Luis González Vilas <sup>a</sup>, Baltasar J. García Pérez <sup>c</sup>, Jorge M. Lobo <sup>d</sup>

<sup>a</sup> Facultad de Ciencias, Universidad de Vigo, 36310 Vigo, Spain

<sup>b</sup> Departamento de Informática, Edificio Fundación, Universidad de Vigo, 36310 Vigo, Spain

<sup>c</sup> Escuela Superior de Ingeniería Informática, Edificio Politécnico s/n, Campus As Lagoas, Universidad de Vigo, 32004 Orense, Spain

<sup>d</sup> Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales (CSIC), c/ José Gutiérrez Abascal 2, 28006 Madrid, Spain

## ARTICLE INFO

### Article history:

Received 1 September 2016

Received in revised form 17 November 2016

Accepted 17 November 2016

Available online 28 November 2016

### Keywords:

Biogeography

Environmental data

Species distribution models

Use-availability

Variable selection

## ABSTRACT

Here, we present SPEDInstabR, available as an R package on CRAN and as an RWizard application on <http://www.ipez.es/RWizard>, which provides tools for the identification of the environmental factors that better discriminate between the conditions prevailing in the area of a species and those existing in the geographical background over which the study is carried out. This could include the world, countries, regions, river basins, etc. or the extent of occurrence of the species estimated by using convex hull,  $\alpha$ -shape or Kernel density distributions. The procedure consists of dividing each factor into a number of intervals or bins decided by the user, calculating the number of records in each bin, separately considering the cells where the species occur and those of the selected geographical background. A peak of instability is observed when there are important differences in the factor comparing the bins of presence with the corresponding ones of extent. We consider that those factors with a higher percentage contribution to the Instability index better discriminate between the cells of presence and the extent. We tested the algorithm using virtual species, comparing the generated selections with those produced by MaxEnt.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

### 1.1. Background

In the absence of exhaustive information, predictions for estimating the geographical distribution of organisms or the niche of taxonomical entities have become one of the most successful research lines in Macroecology and Biogeography (Hortal et al., 2012). Depending on the research purpose, these procedures are frequently denominated as Species Distribution Models (SDMs) or Environmental Niche Models (ENMs) (Peterson et al., 2011). SDM/ENM models generate their estimations by relating a dataset of geographic occurrences, almost exclusively in the form of binary presence-absence data, with the environmental characteristics of the localities in which these occurrences are observed, which act as predictors. A general characteristic of these modeling practices is that they very often generate output values without using the information provided by reliable absence information (Lobo et al., 2010). As it is very difficult to certify the absence of a species in a locality (Isaac and Pocock, 2015), these modeling exercises generally use as a response variable the available information about the occurrence of species (presences) and a undetermined but

high number of “absences” selected at random from the considered territory (the so-called background absences). This is the classic procedure that has been well-known for a long time used in the Resource Selection Functions (RSF; Johnson, 1980), which compare the environmental conditions in the occurrence localities (use) against those existing in the whole territory (availability). It has been shown that RSFs do not estimate the probability of occurrence (Hastie and Fithian, 2013), and that their use only allows a measure of the intensity of the own data used in the model training to be obtained (Aarts et al., 2012). Indeed, the apparent accuracy of the results provided by the models using presences and background absences is a consequence of selecting inappropriate geographical extents (Hijmans, 2012), in which the rates of “absence” success are magnified as a consequence of including zero values beyond the environmental and geographical ranges of the species (“naughty noughts”; Austin and Meyers, 1996; Lobo et al., 2010).

The comparison of the environmental characteristics existing in the presence localities against those available in the territory of interest (geographical background or GB) can make it, however, easier to recognize the predictors with higher probabilities of being influential to explain the environmental preference of the species at the considered extent (Johnson, 1980; Beyer et al., 2010; Acevedo et al., 2012). The number and identity of the selected predictors in an SDM or ENM model are two of the most important sources of uncertainty in the obtained results (Beaumont et al., 2005). To minimize their effect in

\* Corresponding author.

E-mail address: [castor@uvigo.es](mailto:castor@uvigo.es) (C. Guisande).

model predictions, it is worth selecting the minimum number of uncorrelated explanatory variables whose relevance is well founded on biological reasons (Synes and Osborne, 2011). Although the existence of biases in the observed survey effort within the selected GB may seriously hinder such selection of predictors (Kadmon et al., 2004), a recommendable procedure would be first removing collinear variables: identifying them using, for example, variance inflation factors iteratively (VIF; Dormann et al., 2013). Subsequently, the application of a procedure similar to that used by RSFs, comparing the environmental conditions in the presences with those available in the GB, may further aid to select the most relevant variables able to explain the preferences of the species.

Use-availability and related procedures are not without their drawbacks (Garshelis, 2000), being classically carried out using chi-square tests, some indices such as the Manly-Chesson or the Johnson indices, discriminant analysis, logistic regressions or binomial generalized linear models. These indices require the use of discrete environmental unities (e.g. habitats); when it comes to the use of continuous environmental values (e.g. mean temperature values between presence and background values) it is convenient to normalize the data (Manly et al., 2002). For example, Ecological Niche Factor Analyses (ENFA; Hirzel et al., 2002) examine both marginality (viz. the ecological distance between the species optimum derived from occurrences and the mean values within the reference area) and specialization (i.e. the ratio of the ecological variance between species observations and the reference area). These calculations require that the selected environmental variables be normalized.

## 1.2. Aims

In this study, an Instability index is presented with the purpose of discriminating the variables with a higher likelihood of being relevant to explain the occurrence of species when true absence data are lacking. To do this, continuous variables were binned and the conditions in the considered territory were compared against those in the observed presence localities, thus allowing the use of any continuous variable independently of their distribution as though they were discrete. We first explain the rationale and use of this index to examine their performance subsequently, comparing them with the variable importance measures provided by MaxEnt (Phillips et al., 2006), the software for modeling species distribution that is most widely used (Ahmed et al., 2015).

## 2. Material and methods

### 2.1. Installation and platform availability

SPEDInstabR is an application that can be installed as a plug-in of RWizard, an open-source and easy-to-use graphical user interface for the R environment (see <http://www.ipez.es/RWizard/>). However, SPEDInstabR may also be installed and started as a regular R add-on package, because it is available on CRAN (R Development Core Team, 2016). Although SPEDInstabR can be used without the need for RWizard, it is only when it is used in conjunction with RWizard that all its capabilities are truly revealed. To use SPEDInstabR in RWizard, it is necessary to install the file SPEDInstabR.v1.1.RWA, available on the website <http://www.ipez.es/RWizard>, from the main menu of RWizard “RWizard applications → Install new application”. In this way, SPEDInstabR will be added as another RWizard Application (Guisande et al., 2015; Guisande, 2016a; Guisande, 2016b, 2016c). Before the installation of SPEDInstabR, it is first necessary to install RWizard. RWizard is an open-source interface under GNU General Public License that has been developed in C# on the Net platform. It is designed as an interface to facilitate the interaction with R (see video demonstration at <http://www.ipez.es/RWizard>). The only requirement for the installation of RWizard is that R and Net Framework 4.0 must be already installed, and then the file Setup.RWizard.V2.3.EXE must be installed.

Geographic data about countries, regions, and provinces, etc. available in RWizard were obtained from <http://www.openstreetmap.org> (OpenStreetMap, 2016) and freshwater habitats from González-Vilas et al. (2016).

### 2.2. Instability algorithm

The data required by SPEDInstabR (i.e. environmental data of presences and background data for the desired predictors and the selected GB) may be easily obtained from the software ModestR (see <http://www.ipez.es/modestr/>; García-Roselló et al., 2013, 2014, 2015). Alternatively, the user may directly apply the procedure implemented in SPEDInstabR to derive potential distribution maps by the so-called “niche of occurrence” application of ModestR. The Appendix gives a thorough guide to the process of obtaining the files required by SPEDInstabR.

Before running the algorithm, the user may select the most appropriate GB to compare the environmental conditions in the presence localities against those in the selected geographical background. This geographical extent can be the entire Earth, an arbitrary area selected by the user (e.g. countries, regions, river basins), or the extent of occurrence (EOO) of the species. We consider that the most appropriate option to select GB should be the one where the species may have equal access to the available environmental conditions (e.g. basins with presence observations). The use of SPEDInstabR within ModestR allows: i) different EOOs from occurrence points such as convex hulls,  $\alpha$ -shapes or Kernel density distributions to be generated; ii) the grain size or resolution most appropriate to the circumstances (e.g. cells of  $5' \times 5'$ ) to be selected; and iii) the desired environmental variables to be included and selected, such as ASCII files. Once selected, each of the considered continuous environmental variables is divided into quantiles or bins decided by the user (the default is 30), which divide the values of each variable into groups with the same number of observations. Both the number of observations (cells) in the selected GB belonging to each bin for the considered variable and the number of presence observations in each bin is calculated. As more bins are selected by the user, the number of observations is reduced for each one, thus diminishing the power to discriminate or detect the variables characterizing presence localities. We suggest applying the very robust Freedman-Diaconis rule (Birgé and Rozenholc, 2006) to calculate both the number and the width of bins ( $h$ ) as:

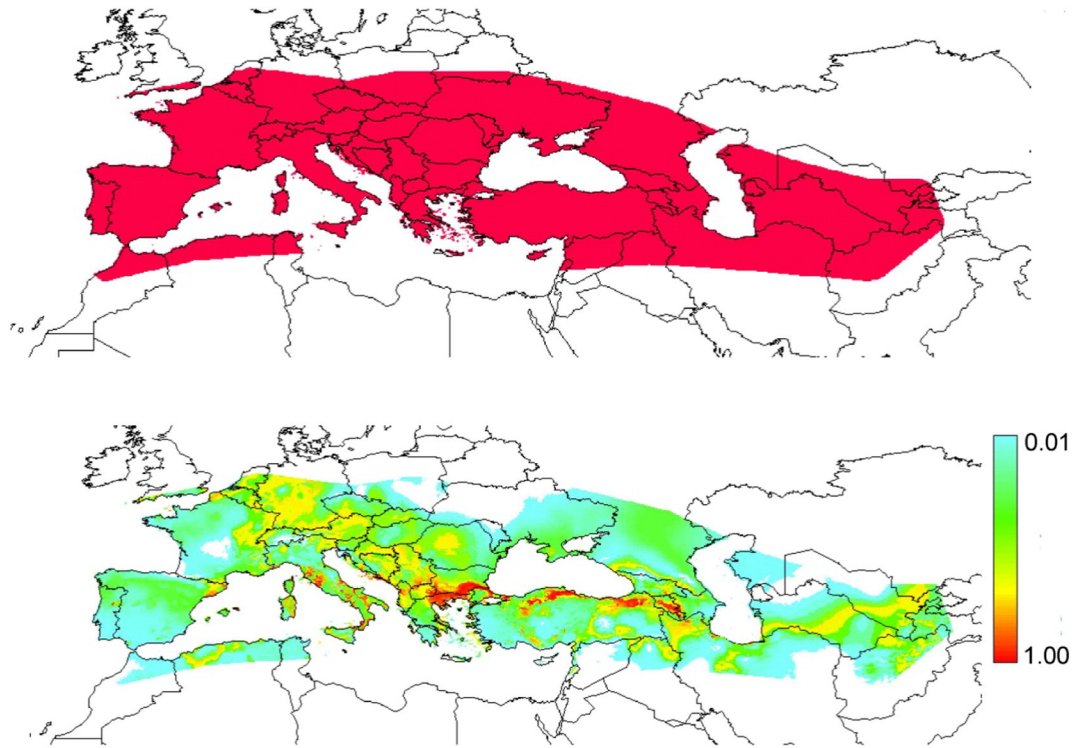
$$h = 2 \times \frac{IQR}{n^{1/3}}$$

where IQR is the interquartile range ( $Q3-Q1$ ) of the variable values in the presence localities, and  $n$  is the number of available presence observations. For each of the considered bins, the relative frequency of the environmental variable data as well as that of presence observation are then used to calculate an index of Instability ( $I$ ).  $I$  is a modification of the fluctuation index of Dubois (1973) modified by Guisande et al. (2011) as:

$$I = \sum_{z=1}^n \sum_{j=1}^i p_{zj} \log_2 \frac{p_{zj}}{\bar{p}_{zj}}$$

$$p_{zj} = \frac{|I_{zje} - I_{zjp}|}{\sum |I_{zje} - I_{zjp}|}$$

where  $n$  is the number of environmental variables,  $i$  is the number of intervals or bins,  $p_{zj}$  is the relative proportion, considering all variables and intervals, of the absolute difference between the interval  $j$  of the variable  $z$  obtained in the cells where the species is present ( $I_{zjp}$ ) and the cells of the whole GB ( $I_{zje}$ ), and  $\bar{p}_{zj}$  is the reference state that is calculated as the mean of all the  $p_{zj}$  values. If  $I_{zjp}$  and  $I_{zje}$  are the same, as zero



**Fig. 1.** Geographical area selected to build the virtual species (above) and example of one virtual species (below) with continuous values in the 0–1 range reflecting a Gaussian symmetrical unimodal response to the bioclimatic variables BIO3 and BIO18 added together, but with different weights (0.7 and 0.3, respectively).

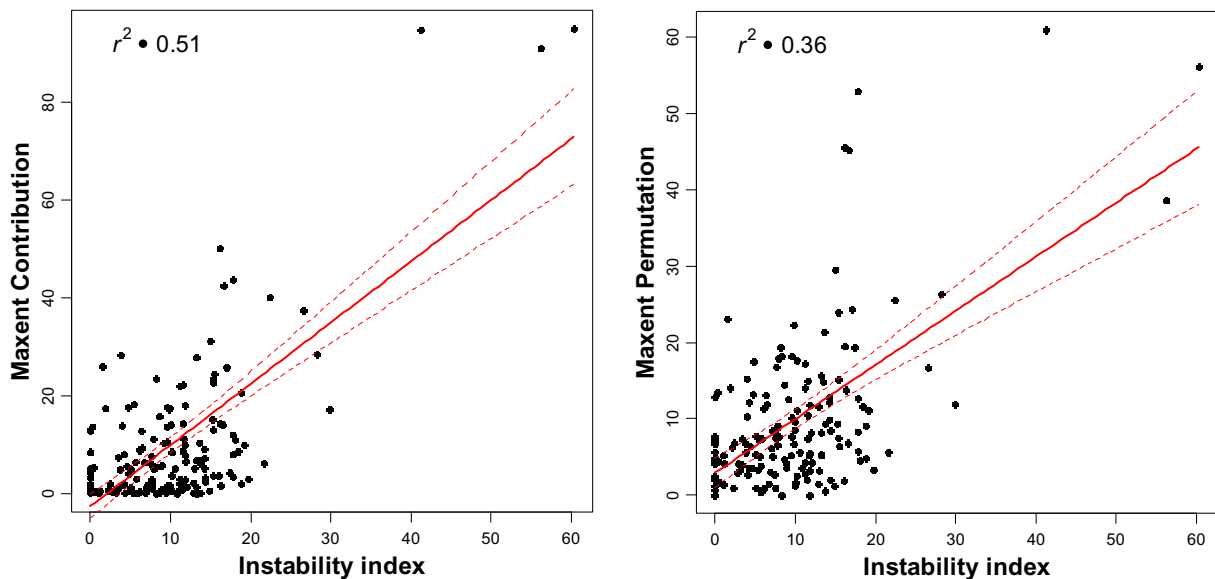
values cannot be included in the algorithm, the minimum instability value is assigned that is obtained for this species.

For each environmental variable, a peak of instability is observed for a bin when there are important differences in the relative frequency of the cells with presence data compared with those of the GB, thus suggesting that some values of this variable seem to be preferred by the species. Once the Instability index is calculated for each species, the values are standardized to a range between 0–1, being thus possible to estimate the percentage of contribution of each environmental variable to the complete index value. We consider that those environmental variables with a higher percentage of contribution to the Instability

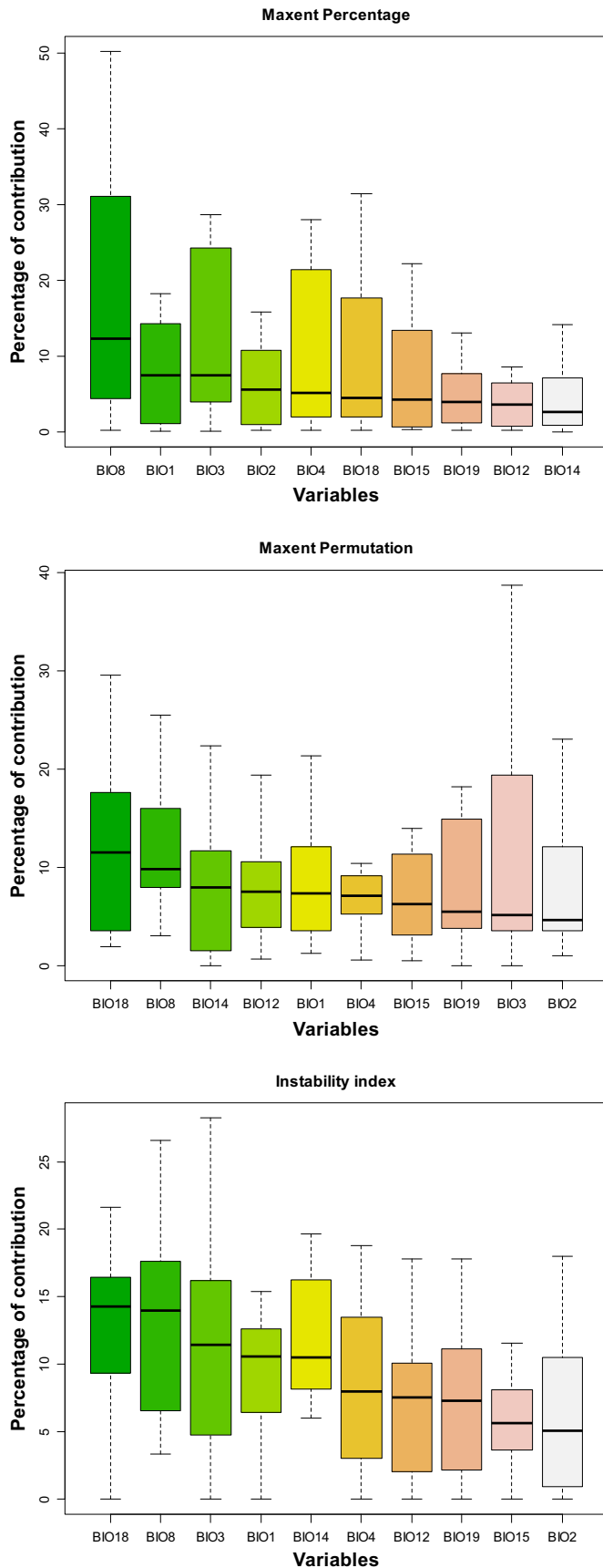
index are those with the highest capacity of discrimination between areas of presence and the geographical background.

### 3. Results

From the 19 bioclimatic variables of the WorldClim dataset (Hijmans et al., 2005), we only used the 10 with a variance inflation factor (VIF) lower than 15: annual mean temperature (BIO1), mean diurnal range (BIO2), isothermality (BIO3), temperature seasonality (BIO4), mean temperature of wettest quarter (BIO8), annual precipitation (BIO12), precipitation of the driest month (BIO14), precipitation seasonality



**Fig. 2.** Relationship between the percentage contribution of each variable for each species obtained with the Instability index and both the percentage of contribution obtained with MaxEnt contribution (upper panel) and Maxent permutation (lower panel).



(BIO15), precipitation of the warmest quarter (BIO18) and precipitation of the coldest quarter (BIO19). Therefore, collinear variables were removed. The resolution of the variables was  $5' \times 5'$ .

The capacity of the Instability index to recover the relevant explanatory variables was assessed by building virtual species whose occurrence is completely determined by known climatic variables. First, an area within the western Palearctic region was selected as GB (Fig. 1; 132,929 cells of  $5' \times 5'$ ). In this area, a third part of the complete environmental range around mean values [mean  $\pm$  (maximum – minimum values / 3)] is selected for the variables BIO3 ( $n = 88,113$  cells), BIO8 ( $n = 84,347$ ) and BIO18 ( $n = 77,813$ ), and the obtained climatic values are normalized to 0–1 range ( $N$ -values) so that they have a Gaussian symmetrical unimodal response. With the data of these three climatic variables, eight virtual species were built: i) three using only the  $N$ -values of each one of the variables; ii) one using the  $N$ -values of BIO3 and BIO8 added together ( $N$ -values BIO3 +  $N$ -values BIO8) with the same weights (0.5 and 0.5); iii) one using BIO3 and BIO18 added together but with different weights (0.7 and 0.3, respectively); iv) one using BIO8 and BIO18 added together but with different weights (0.2 and 0.8); v) one using BIO3, BIO8 and BIO18 added together with different weights (0.5, 0.2 and 0.3, respectively); and finally vi) a virtual species using BIO3, BIO8 and BIO18 with different weights (0.5, 0.2 and 0.3, respectively) but multiplied together ( $N$ -values BIO3 \*  $N$ -values BIO8 \*  $N$ -values BIO18). All the continuous values obtained in this way that were higher than zero were transformed to binary (1/0), and 0.1% and 1% of the presence cells were selected at random for each virtual species. Thus, the capacity of recovering the true influential variables (BIO3, BIO8 and BIO18) from the initial set of 10 climatic variables was measured on 16 occasions under different situations differing in complexity and sample completeness. These 16 virtual species are available in the supplementary material in the file Virtual\_ModestRDB.

The results of the Instability index were compared with those provided by the MaxEnt software (Percent Contribution and Permutation Importance; further details in the tutorial available at <https://www.cs.princeton.edu/~schapire/maxent/>).

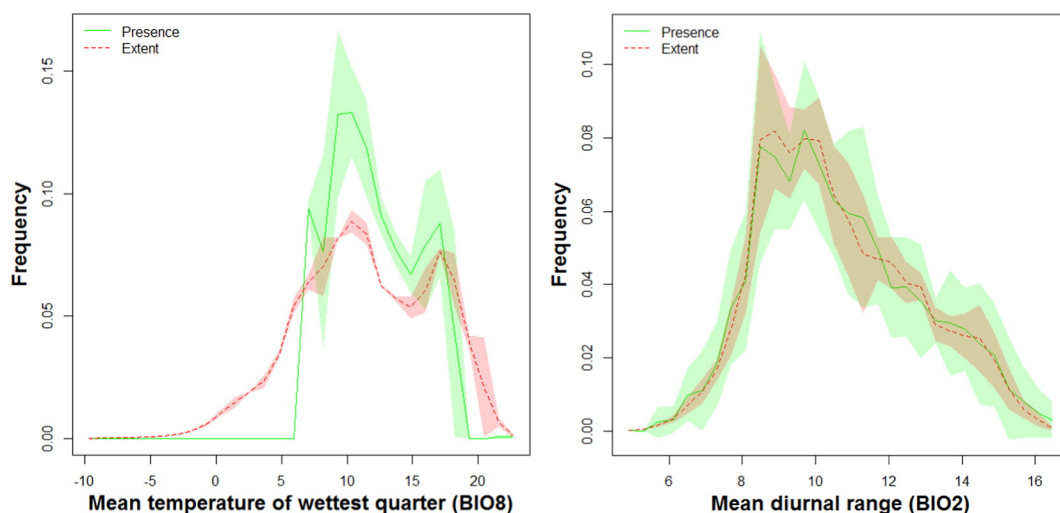
There was a significant relationship between the percentage of contribution obtained using the Instability index with both MaxEnt contribution and MaxEnt permutation (Fig. 2), although a better fit was obtained with the former ( $r^2 = 0.51$ ) than the latter ( $r^2 = 0.36$ ). However, only the Instability index showed that the variables with a higher percentage of contribution to explain the differences between the cells of presence and those of the GB were those selected for modeling the virtual species: BIO3, BIO8 and BIO18 (Fig. 3).

With SPEDInstabR, it is also possible to show for each factor the frequency of the records observed in each bin, for just one or a group of selected species (see the Appendix for details). For instance, Fig. 4 shows for BIO8 that the frequency of records in the cells of presence is higher than in those of the GB at intermediate values (Fig. 4; upper panel). However, for the BIO2 considering all the species, the frequency of intervals was similar in both the cells of presence and in those of the GB (Fig. 4; lower panel).

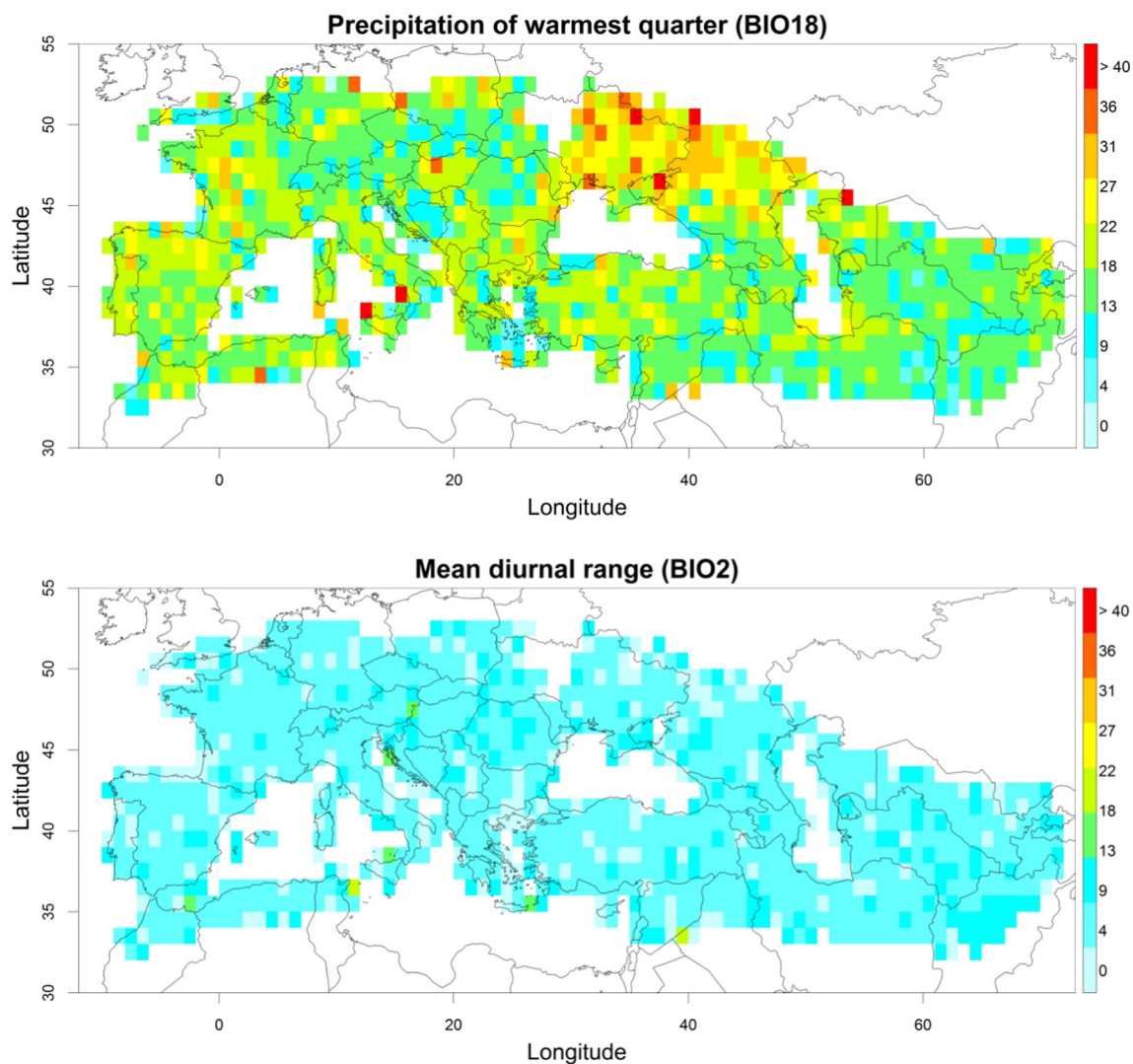
With ModestR or SPEDInstabR, it is also possible to obtain the mean contribution of any variable considering all or a selected group of species present in the cell, in the scale used for the calculations (see the Appendix for details). Fig. 5 shows the mean contribution of one of the most important variables (BIO18) and of a variable with a low mean contribution (BIO2). Therefore, it is possible to analyze potential changes in the differential contribution of variables across a geographical area.

**Fig. 3.** Boxplots with the mean contribution of all variables considering all species obtained using Maxent contribution (upper panel), MaxEnt permutation (middle panel) and the Instability index (lower panel). Outliers are not shown. Annual mean temperature (BIO1), mean diurnal range (BIO2), isothermality (BIO3), temperature seasonality (BIO4), mean temperature of the wettest quarter (BIO8), annual precipitation (BIO12), precipitation of the driest month (BIO14), precipitation seasonality (BIO15), precipitation of the warmest quarter (BIO18) and precipitation of the coldest quarter (BIO19).





**Fig. 4.** Mean  $\pm$  SD frequencies of each standardized interval for the variable BIO8 (upper panel) and for the variable BIO2 (lower panel) considering all species. The upper limit of the interval for each variable is shown on the x-axis.



**Fig. 5.** Map with the mean percentage contribution considering all species of precipitation of the warmest quarter (BIO18; upper panel) and mean diurnal range (BIO2; lower panel) in cells of 5'  $\times$  5'.

## 4. Conclusions

Identification of the main factors controlling the distribution of species is one of the most important issues in ecology, particularly in studies about SDMs and ENMs, where there are many modeling methods with different approaches. For all of them, predictor selection is an important issue affecting the models' predictive ability (Elith et al., 2006; Austin and Meyers, 1996).

Our algorithm allows continuous variables to be managed without the need of normalizing them. Using virtual species, we demonstrated that is able to identify the main factors affecting the species' distribution. This is particularly true under the complex situations that are common in nature, with different factors affecting it to varying degrees; under such circumstances, MaxEnt, the most widely used method for modeling species' distributions, fails.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.ecoinf.2016.11.004>.

## References

- Aarts, G., Fieberg, J., Matthiopoulos, J., 2012. Comparative interpretation of count, presence-absence and point methods for species distribution models. *Methods Ecol. Evol.* 3, 177–187.
- Acevedo, P., Jiménez-Valverde, A., Lobo, J.M., Real, R., 2012. Delimiting the geographical background in species distribution modelling. *J. Biogeogr.* 39, 1383–1390.
- Ahmed, S.E., McInerney, G., O'Hara, K., Harper, R., Salido, L., Emmott, S., Joppa, L.N., 2015. Scientists and software - surveying the species distribution modelling community. *Divers. Distrib.* 21, 258–267.
- Austin, M., Meyers, J., 1996. Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. *For. Ecol. Manag.* 85, 95–106.
- Beaumont, L.J., Hughes, L., Poulsen, M., 2005. Predicting species distributions: use of climatic parameters in BIOCLIM and its impact on predictions of species' current and future distributions. *Ecol. Model.* 186, 250–269.
- Beyer, H.L., Haydon, D.T., Morales, J.M., Frair, J.L., Hebblewhite, M., Mitchell, M., Matthiopoulos, J., 2010. The interpretation of habitat preference metrics under use-availability designs. *Philos. Trans. R. Soc. B* 365, 2245–2254.
- Birgé, L., Rozenholc, Y., 2006. How many bins should be put in a regular histogram. *ESAIM Probab. Stat.* 10, 24–25.
- Development Core Team, R., 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46.
- Dubois, D.M., 1973. An index of fluctuations, Do, connected with diversity and stability of ecosystems: applications in the Lotka-Volterra model and in an experimental distribution of species. Rapport de sythèse III, Programme National Sur l'environnement Physique et Biologique Mer. Commission Interministérielle de La Politique Scientifique. Liège, Project.
- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distribution from occurrence data. *Ecography* 29, 129–151.
- García-Roselló, E., Guisande, C., González-Dacosta, J., Heine, J., Pelayo-Villamil, P., Manjarrés-Hernández, A., Vaamonde, A., Granado-Lorencio, C., 2013. ModestR: a software tool for managing and analyzing species distribution map databases. *Ecography* 36, 1202–1207.
- García-Roselló, E., Guisande, C., Heine, J., Pelayo-Villamil, P., Manjarrés-Hernández, A., González-Vilas, L., González-Dacosta, J., Vaamonde, A., Granado-Lorencio, C., 2014. Using ModestR to download, import and clean species distribution records. *Methods Ecol. Evol.* 5, 703–713.
- García-Roselló, E., Guisande, C., Manjarrés-Hernández, A., González-Dacosta, J., Heine, J., Pelayo-Villamil, P., Vari, R.P., Vaamonde, A., Granado-Lorencio, C., Lobo, J.M., 2015. Can we derive macroecological patterns from primary GBIF data? *Glob. Ecol. Biogeogr.* 24, 335–347.
- Garshelis, D.L., 2000. Delusions in habitat evaluation: measuring use, selection, and importance. In: Boitani, L., Fuller, T.K. (Eds.), *Research Techniques in Animal Ecology: Controversies and Consequences*. Columbia University Press, New York, pp. 111–164.
- González-Vilas, L., Guisande, C., Vari, R.P., Pelayo-Villamil, P., Manjarrés-Hernández, A., García-Roselló, E., González-Dacosta, J., Heine, J., Pérez-Costas, E., Granado-Lorencio, C., Palau-Ibars, A., Lobo, J.M., 2016. Geospatial data of freshwater habitats for macroecological studies: an example with freshwater fishes. *Int. J. Geogr. Inf. Sci.* 30, 126–141.
- Guisande, C., 2016a. Niche estimation. R package version. 1:3 Available: <http://CRAN.R-project.org/package=EnvNicheR> (Accessed 20 January 2016).
- Guisande, C., 2016b. An algorithm for morphometric characters selection and statistical validation in morphological taxonomy. R package version. 1:1 Available: <http://CRAN.R-project.org/package=VARSEDIG> (Accessed 2 February 2016).
- Guisande, C., 2016c. Estimation of the relative importance of factors affecting species distribution based on stability concept. R package version. 1:1 Available: <http://CRAN.R-project.org/package=SPEDInstabR> (Accessed 5 June 2016).
- Guisande, C., Barreiro, A., Vaamonde, A., 2011. Tratamiento de Datos Con R, Statistica Y SPSS. Díaz de Santos, Madrid.
- Guisande, C., Heine, J., García-Roselló, E., González-Dacosta, J., Pérez-Costa, E., García-Pérez-Schofield, B., González-Vila, L., Vaamonde, A., Lobo, J.M., 2015. FactorsR: an RWizard application for identifying the most likely causal factors in controlling species richness. *Diversity* 7, 385–396.
- Hastie, T., Fithian, W., 2013. Inference from presence-only data; the ongoing controversy. *Ecography* 36, 864–867.
- Hijmans, R.J., 2012. Cross-validation of species distribution models: removing spatial sorting and calibration with a null model bias. *Ecology* 93, 679–688.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978.
- Hirzel, A.H., Hausser, J., Chessel, D., Perrin, N., 2002. Ecological-niche factor analysis: how to compute habitat suitability maps without absence data? *Ecology* 83, 2027–2036.
- Hortal, J., Lobo, J.M., Jiménez-Valverde, A., 2012. Basic questions in biogeography and the (lack of) simplicity of species distributions: putting species distribution models in the right place. *Nat. Conservacao* 10, 108–118.
- Isaac, N.J.B., Pocock, M.J.O., 2015. Bias and information in biological records. *Biol. J. Linn. Soc.* 115, 522–531.
- Johnson, D.H., 1980. The comparison of usage and availability measurements for evaluating resource preference. *Ecology* 61, 65–71.
- Kadmon, R., Farber, O., Danin, A., 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecol. Appl.* 14, 401–413.
- Lobo, J.M., Jiménez-Valverde, A., Hortal, J., 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 33, 103–114.
- Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L., Erickson, W.P., 2002. *Resource Selection by Animals*. Kluwer Academic Publishers, Dordrecht.
- OpenStreetMap, 2016. <http://www.openstreetmap.org/copyright> (Accessed 23 February 2016).
- Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R., Martínez-Meyer, E., Nakamura, M., Araújo, M., 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton.
- Phillips, S.J., Anderson, R.P., Schapire, S.E., 2006. Maximum entropy modelling of species geographic distributions. *Ecol. Model.* 190, 231–259.
- Syns, N.W., Osborne, P.E., 2011. Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. *Glob. Ecol. Biogeogr.* 20, 904–914.