OXFORD

Article

A procedure to assess the spatial variability in the importance of abiotic factors affecting distributions: the case of world freshwater fishes

Ana M. Manjarrés-Hernández[a], Cástor Guisande[b,*], Emilio García-Roselló[c], Patricia Pelayo-Villamil[d], Jacinto González-Dacosta[c], Jürgen Heine[c], Luis González Vilas[b], Carlos Granado-Lorencio[e], Santiago R. Duque[a], and Jorge M. Lobo[f]

[a]Instituto Amazónico de Investigaciones-IMANI, Universidad Nacional de Colombia, Leticia, A.A. 215, Colombia 910001, [b]Facultad de Ciencias del Mar, Universidad de Vigo, Vigo 36310, Spain, [c]Departamento de Informática, Edificio Fundición, Universidad de Vigo, Vigo 36310, Spain, [d]Grupo de Ictiología, Universidad de Antioquia, Medellín, Colombia 050010, [e]Departamento de Biología Vegetal y Ecología, Facultad de Biología, Universidad de Sevilla, Sevilla 41080, Spain, and [f]Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales (CSIC), Calle de José Gutiérrez Abascal 2, Madrid 28006, Spain

*Address correspondence to Cástor Guisande. E-mail: castor@uvigo.es.

Abstract

Understanding the factors shaping species' distributions is a key longstanding topic in ecology with unresolved issues. The aims were to test whether the relative contribution of abiotic factors that set the geographical range of freshwater fish species may vary spatially and/or may depend on the geographical extent that is being considered. The relative contribution of factors, to discriminate between the conditions prevailing in the area where the species is present and those existing in the considered extent, was estimated with the instability index included in the R package SPEDInstabR. We used 3 different extent sizes: 1) each river basin where the species is present (local); 2) all river basins where the species is present (regional); and 3) the whole Earth (global). We used a data set of 16,543 freshwater fish species with a total of 845,764 geographical records, together with bioclimatic and topographic variables. Factors associated with temperature and altitude show the highest relative contribution to explain the distribution of freshwater fishes at the smaller considered extent. Altitude and a mix of factors associated with temperature and precipitation were more important when using the regional extent. Factors associated with precipitation show the highest contribution when using the global extent. There was also spatial variability in the importance of factors, both between species and within species and from region to region. Factors associated with precipitation show a clear latitudinal trend of decreasing in importance toward the equator.

Key words: anisotropic predictors, environmental data selection, geographical background, non-stationary predictors

The geographical area where a species is distributed is a complex expression of its ecology and evolutionary history (Brown 1995), being determined by several factors that interact dynamically and with different strengths at different scales (Gaston 2003; Pearson and Dawson 2003). Four main classes of factors determine areas in which a species is found: abiotic and biotic factors, the regions that are accessible to dispersal by the species from some source area, and the evolutionary capacity of populations of the species to adapt to new conditions (Soberón and Peterson 2005). In this context, the estimation of the factors correlated with the distribution of organisms is the necessary first step to infer the possible mechanisms able to explain why the location, shape, extent, and boundaries of species' ranges are as they are (Van de Pol et al. 2016), as well as to project the probable distribution of species in absence of exhaustive distribution information (Peterson et al. 2011). The explanatory variables for these purposes are generally established by examining the relationships between their values and the abundance, density, cover or presence/absence data in the case of individual species. In these studies both the information of the predictors and of the response variable come from different localities, which are managed in a "global" way independently of their spatial location. Following Ramón Margalef (1968) we could say that species "do not dance on the head of a pin" and those simple stationary and isotropic global models cannot represent appropriately the complex nature of the relations between environmental conditions and species occurrences, which may differ spatially.

The spatial variation in the weights of a predictor has been estimated, for example by examining the change in the explanatory capacity of a variable across distance classes or resolutions (Steffan-Dewenter et al. 2002; Chust et al. 2004), a particularly important issue in the case of freshwater fishes (Radinger et al. 2015), due to the linear dependence of local characteristics. However, it is the statistical technique geographically weighted regression (GWR) that has been used in ecology to estimate the local influence of a predictor variable when its relevance varies spatially (Osborne and Suarez-Seoane 2002; Foody 2004; Bickford and Laffan 2006; Holloway and Miller 2015). GWR can calculate local regression coefficients of each explanatory variable considering subsamples of neighbors, which are weighted differently by their distance to the focal point (see Fotheringham et al. 2002). Although having many advantages, GWR is computer-intensive, difficult to apply in the case of binary response variables (presence–absence), and subjected to the same requirements of ordinary least squares regressions.

The selection of the most relevant environmental variables is key in the so-called species distribution models (SDMs) or ecological niche models (ENMs), in which environmental characteristics of the localities in which a species is observed are used as predictors to account for species occurrences (Peterson et al. 2011). As reliable absence information about species is generally lacking and difficult to obtain, the selection of the variables with a higher probability of being influential in predicting the distribution of species can be carried out by comparing the environmental values in the occurrence localities (use or presences) against those existing in the selected territory (availability or background absences). This is the classic procedure used in the resource selection functions (Johnson 1980) to recognize the environmental predictors that contribute most to explain the preferences of the species. These use/availability or presence/background absence comparisons are rarely carried out considering the non-stationary nature of the species occurrence–environment relationships; that is, the lack of constancy of these relationships across the geography. Mcnew et al. (2013) is the first

and unique study that we know using a logistic GWR in the context of a use-availability scheme.

An instability index that does not require normalized data has been proposed recently to discriminate the variables with a higher likelihood of being relevant for presence/background absence data (Guisande 2016; Guisande et al. 2016). Here, this instability index is used to examine and visualize the spatial variation in the comparative importance of a high number of continuous predictors, thus showing the anisotropic and non-stationary character of predictors. Worldwide freshwater fish species data are used to exemplify the capacity of the proposed procedure to describe the comparative relevance of temperature and precipitation variables in explaining species distributions at different scales as well as showing their contrasting non-stationary character.

Material and Methods

Estimation of the relative contribution of factors to species' distribution

The relative contribution of the selected explanatory variables on the distribution of each one of the species was estimated with the recently proposed instability index implemented in the R package SPEDInstabR (Guisande 2016; Guisande et al. 2016; see help manual in Guisande 2016, https://cran.r-project.org/web/packages/SPEDInstabR/SPEDInstabR.pdf), which is based on the fluctuation index of Dubois (1973) modified by Guisande et al. (2006, 2011). This instability index has been designed to improve the correct identification of the variables controlling species distributions, when these are continuous and difficult to normalize. To do that, the SPEDInstabR application of RWizard (Guisande et al. 2014) allows the reliable identification of the environmental factors that better discriminate between the conditions prevailing in presence locations, against those in the geographical background or area over which a study was carried out. Once selected, these variables can be used to estimate the probable and potential occurrence of the species from fragmentary data (Pelayo-Villamil et al. 2012; Lobo 2016) but also for other purposes as estimating the variables with a higher probability of being relevant to explain the distribution of species.

The variables selected are 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 geographical background belonging to each bin for the considered variable and the number of presence observations in each bin is calculated. 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 (Guisande et al. 2006) as:

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

$$p_{zj} = \frac{\left| I_{zje} - I_{zjp} \right|}{\Sigma \left| I_{zje} - I_{zjp} \right|},$$

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 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 geographical background, 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 and 1, being thus possible to estimate the percentage of contribution of each environmental variable to the complete index value. 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.

Geographical extent

Due to the importance of the selected geographical extent both in the results and in the discrimination capacity of SDMs (Barve et al. 2011; Acevedo et al. 2012; Niamir et al. 2016), 3 progressively widening extents were used to determine the comparative importance of environmental predictors. The more restricted used extent (E1) is delimited as each one of the river basins of level 2 (González-Vilas et al. 2016), where there are observations of each species. Thus, if a species has presence data in 3 different basins of level 2, the contribution of predictors is individually estimated for each basin. The rationale to use this geographical extent is that basins would encompass the set of accessible localities for each species, and that the comparative relevance of each environmental predictor will be better determined when the effect of dispersal limitations are minimized (Peterson et al. 2011; Acevedo et al. 2012). The geographical extent that follows includes all those river basins of level 2 where the species is present (E2); that is, assuming that all the presence localities are accessible and connected, even if they belong to different basins. Finally, the whole Earth (E3) was also used as a geographical extent. The number of intervals in which each factor is divided to compare the conditions in presence localities versus the selected geographical extent was the default option in SPEDInstabR.

Origin of freshwater fish species data

The data set of geographical records for freshwater fishes developed by Pelayo-Villamil et al. (2015) was updated to reflect the taxonomic changes and new species described until the end of April 2016. Online Appendix 1 shows a detailed description of all sources obtained from the Global Biodiversity Information Facility (GBIF) used in this data set, and other sources used such as web pages, museums, etc. are described in Pelayo-Villamil et al. (2015). Records were downloaded and filtered using the data cleaning facilities available in the ModestR software (García-Roselló et al. 2013, 2015): 1) records with the same latitude and longitude were not included; 2) records with the latitude and longitude 0° were not included; 3) duplicated records were not included; and 4) habitat data cleaning (see García-Roselló et al. [2014] for details). At the end of April 2016, 16,543 species of freshwater fishes were recognized as valid by systematists and are available in IPez (http://www.ipez.es, Guisande et al. 2010), so this taxonomic list of species was used in our study. Of these, 16,479 species (99.6% of the total) have associated geographical information for a total of 845,764 geographical records (without duplication). From species with

distribution information (with available geographical records), in the extent E1 only those species with more than 30 records were considered (1,124 species with a total of 732,604 geographical records). In the case of the extents E2 and E3, we used in our study only those species with more than 100 records (846 freshwater fish species with a total of 723,874 geographical records; see Online Appendix 2 for detailed description of the species included in the analysis).

Used environmental variables

From the 19 bioclimatic variables of the WorldClim data set for all the Earth's terrestrial area (Hijmans et al. [2005], see the acronyms in the web site http://www.worldclim.org/bioclim), we firstly select those with a variance inflation factor (VIF) lower than 20. VIF quantify the multicollinearity of predictors (Dormann et al. 2013) and in our case this value was selected in order to eliminate the variables showing the most severe multicollinearity. We used this unusual high VIF threshold to only eliminate those highly correlated environmental variables in order to maximize the differences in the selected predictors when the considered extent varies.

As consequence, the following 10 bioclimatic variables were finally selected: annual mean temperature (BIO1), mean diurnal range (BIO2), isothermality (BIO3), temperature seasonality (BIO4), mean temperature of wettest quarter (BIO8), annual precipitation (BIO12), precipitation of driest month (BIO14), precipitation seasonality (BIO15), precipitation of warmest quarter (BIO18), and precipitation of coldest quarter (BIO19). We also used other variables (see Pelayo-Villamil et al. 2015), such as human population density (number of people per km^2 in year 2000), slope (topographic slope in degrees), slope-aspect (which is defined as the compass direction to which a slope faces measured in degrees), altitude (meters), vegetation index (VI), terrestrial primary production (TPP, $g \ C \ m^{-2} \ d^{-1}$), and TH24 (topographic heterogeneity calculated for the 24 surroundings cells, see Pelayo-Villamil et al. 2015). We therefore used both direct and indirect variables (Austin 2007) probably related with unconsidered ones, such as solar radiation, river flow, etc.

Human population density comes from a globally consistent, spatially explicit map based on the Gridded Population of the World dataset, Version 3 (GPWv3). To develop the global data set, national population data are transformed from their native spatial units, which are usually administrative (such as state or county-level) and of varying resolutions to a global grid of quadrilateral, latitude–longitude cells at a resolution of 2.5 arc min, and then downscaled to 6 arc min. A proportional allocation gridding algorithm, utilizing more than 300,000 national and sub-national administrative units, is used to assign population values to the 1-degree grid cells. Population densities show the number of humans per square kilometer, based on census data available in 2000 and with estimates when necessary to fill in missing or incomplete data.

The source for slope and aspect comes from combining data from NASA's Shuttle Radar Topography Mission covering the land surface from 60° south to 60° north. The data for the rest of the Northern Hemisphere (60–90° north) come from digital elevation models (digital versions of paper-based topographic maps) produced by the US Geological Survey. The data for the remainder of the Southern Hemisphere (60–90° south) come from the "RAMP II" project of the Radarsat Antarctic Mapping Project Digital Elevation Model, Version 2.

VI and TPP (in $g \ C \ m^{-2} \ d^{-1}$) come from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument aboard
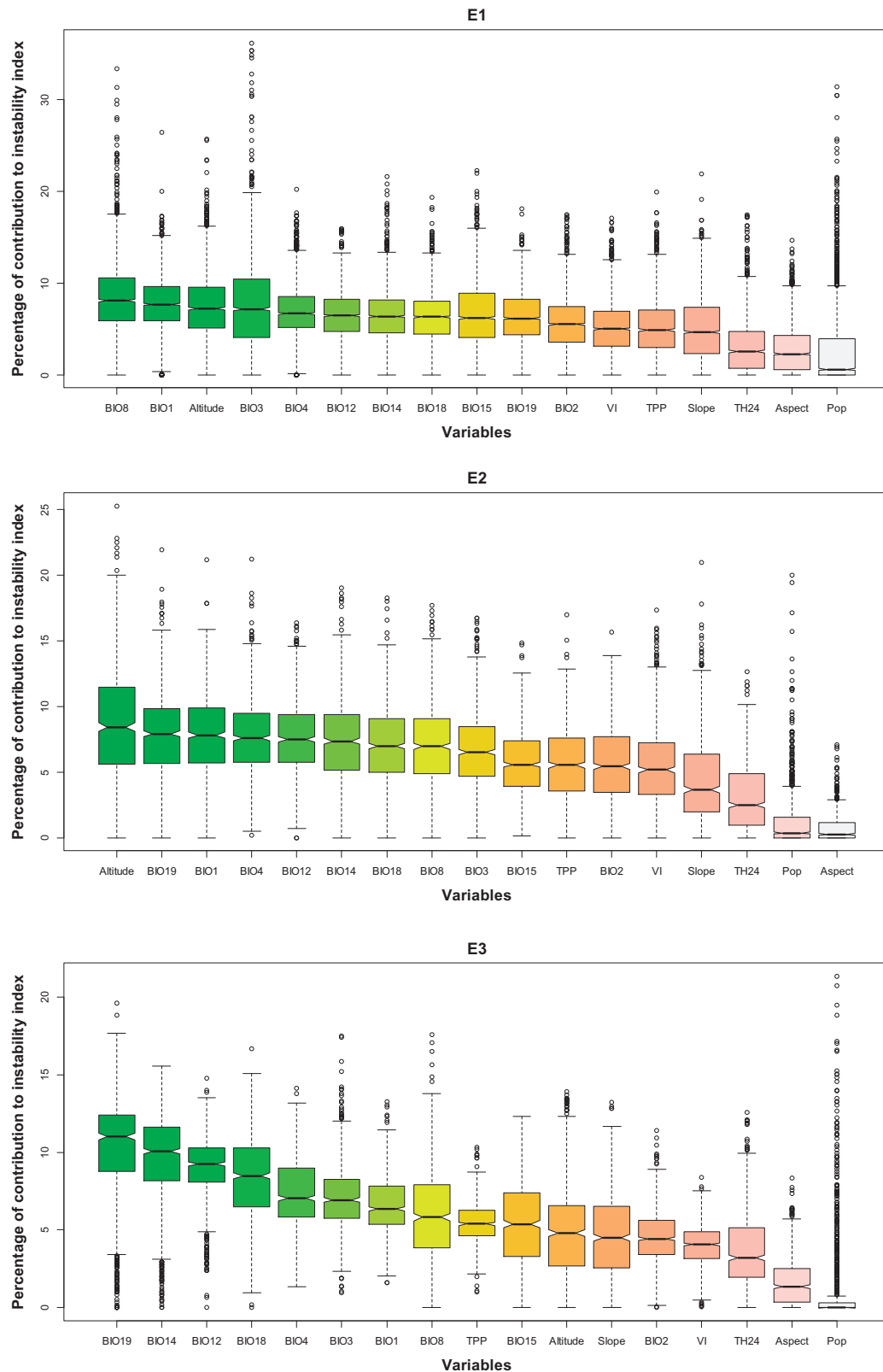
Figure 1. Boxplots representing the median contribution of each one of the considered variables comparing the conditions of each freshwater fish species in their presence cells of 5′ × 5′ against the conditions prevailing in the cells i) included in each river basin of level 2 where there are observations of each species (E1); ii) all river basins where the species is present (E2); and iii) the complete world as extent (E3). The limit of the bars indicates the minimum and maximum, the limit of the box indicates the first and third quartile, and the points are the outliers. If the notches of 2 groups do not overlap it seems to be an evidence that the 2 medians differ. BIO1, annual mean temperature; BIO2, mean diurnal range; BIO3, isothermality; BIO4, temperature seasonality; BIO8, mean temperature of wettest quarter; BIO12, annual precipitation; BIO14, precipitation of driest month; BIO15, precipitation seasonality; BIO18, precipitation of warmest quarter; BIO19, precipitation of coldest quarter; number of people per km^2, Pop, population density; topographic slope in degrees, slope; slope-aspect (which is defined as the compass direction to which a slope faces measured in degrees); altitude (meters); VI, vegetation index; TPP, terrestrial primary production in g C m^{-2} d^{-1}; and TH24, topographic heterogeneity calculated for the 24 surroundings cells.
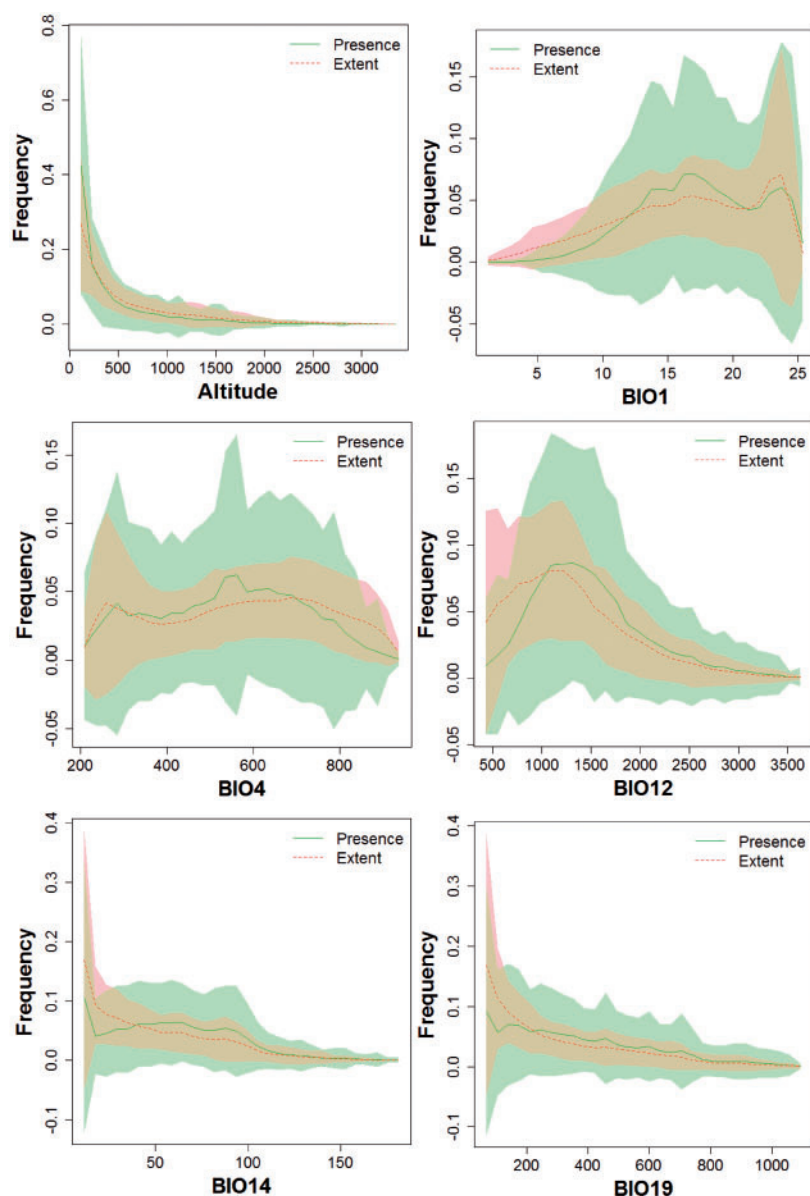
Figure 2. Mean ± SD frequencies (lines ± shaded areas) for each variable interval for the cells with presence records and the cells of the E2 geographical extent (all river basins of level 2 where the species is present). BIO1, annual mean temperature; BIO4, temperature seasonality; BIO12, annual precipitation; BIO14, precipitation of driest month; and BIO19, precipitation of the coldest quarter.

NASA's Terra satellite. Specifically, monthly data of terrestrial net primary productivity and VI from 2001 to 2010 were obtained by averaging available information for each pixel of selected variables using the statistical software RWizard (Guisande et al. 2014). The net primary productivity indicates how much carbon dioxide is taken up by vegetation during photosynthesis minus how much carbon dioxide is released when plants respire. The values indicate how fast carbon was taken in, or released, for every square meter of land over the indicated time span. Values range from -1.0 g C m^{-2} d^{-1} to 6.5 g m^{-2} d^{-1}. A negative value means decomposition or respiration exceeded carbon absorption; in other words, more carbon was released into the atmosphere than was absorbed by the plants. We also include the VI as a productivity variable. This variable represents a measure of the greenness of Earth's landscapes.

All these variables were included in the analyses at a resolution of $5' \times 5'$ (≈ 100 km^2). The use of the same resolution for the different extents may be influencing the provided results but conversely using different resolutions may complicate the estimation of the effect that we try to measure (the extent) as showed by other authors (Gillingham et al. 2012).

Results

At the more restricted extent (E1), variables associated to temperature (BIO8, BIO1, BIO3, and BIO4) and the altitude were the factors with the highest relative contribution to explain the distributions of each individual freshwater fish species (Figure 1 upper panel). Altitude and a mix of factors both associated with temperature (BIO1 and BIO4) and precipitation (BIO19, BIO12, and BIO14) become important factors at E2 (regional extent; see medium panel in Figure 1). Lastly, some variables associated with
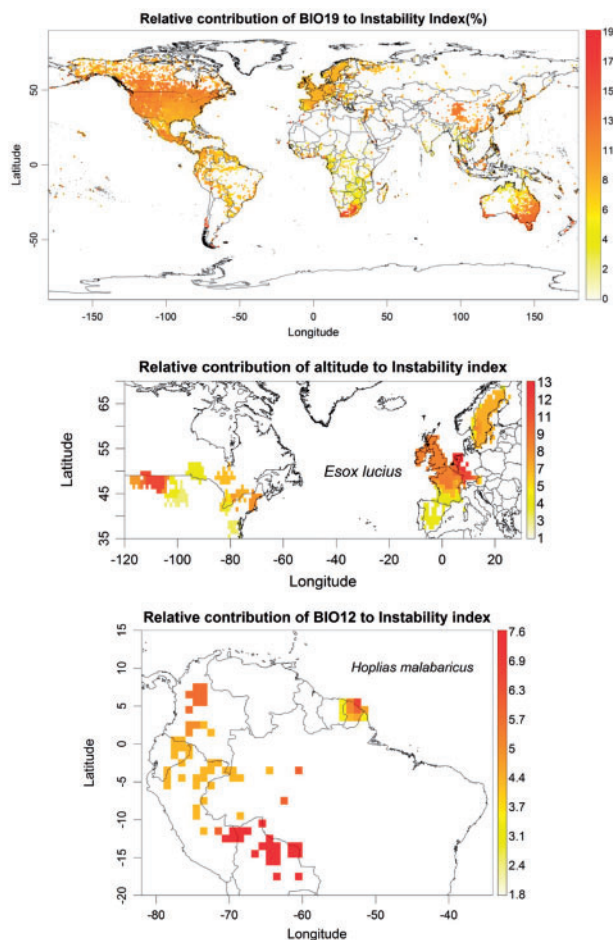
Figure 3. Map representing the geographical variation in the mean percentage of contribution of precipitation of the coldest quarter (BIO19, upper panel) for all freshwater fish species in the cells of 1°, as measured by the instability index. These contribution scores were calculated using the SPEDInstabR algorithm comparing the values in the presence cells against those present in all the cells of the E2 geographical extent (all river basins of level 2 where the species is present). Percentage of contribution of altitude on the distribution *Esox lucius* (medium panel) using the E1 geographical extent, and contribution of annual precipitation (BIO12) for the species *Hoplias malabaricus* (lower panel) at the same extent.

precipitation (BIO19, BIO14, BIO12, and BIO18) were those that had the highest contributions when using the world as extent (E3; lower panel in Figure 1). Therefore, the contribution of the factors associated with precipitation seems to increase with the size of the considered extent, while the contribution of the factors associated with temperature variables increased as the size of the extent decreases.

The differences in the frequency of presence cells along each variable gradient versus the frequency of the cells at the selected extent can be geographically represented (Figure 2). The results for the E2 situation were described, although the patterns were similar for the E1 and E3 extents (shown in Online Appendix 3, Figure A3.5 for E1 and Figure A3.6 for E3). Thus, presence cells seem to be comparatively more frequent at lower altitudes (<250 m) and less frequent, at altitudes from 300 m to 2,000 m, than the frequencies of the cells of the E2 geographical extent (Figure 2). At higher altitudes (>2,000 m) both frequencies seem to be similar. In the case of annual mean temperatures (BIO1), presence cells were less frequent

until 13 °C, approximately, but more frequent at temperatures from 13°C to 22 °C (Figure 2). The frequencies of presence cells were higher at intermediate values of temperature seasonality (BIO4), but freshwater species seem to avoid areas with high temperature seasonality (Figure 2). In the case of BIO12 (annual precipitation), the frequencies of presence cells were lower than those of the geographical extent in places with <1,200 mm, approximately, but higher at greater precipitation values. Both the precipitation of the driest month (BIO14) and the precipitation of the coldest quarter (BIO19) showed a similar pattern (Figure 2), with the frequencies of the presence cells being lower at lower values and higher at intermediate values than the frequencies of cells at the considered extent.

The spatial variability in the relative contribution of each factor, both for all species (mean contribution) and single species, show that the same variable may exercise a different influence depending on the geographical location and the considered extent (Figure 3 upper panel). This spatial variability can be observed in all the considered environmental variables (not shown). In fact, a clear latitudinal gradient can be observed in the variation of the percentage of contribution of some variables, which may vary depending on the considered extent (i.e., BIO14 and BIO19 see Figure 4). The contribution of an indirect variable such as altitude seems to be highly dependent on the considered extent, although is generally lower at the higher latitudes of the southern hemisphere. The contribution of annual mean temperature was rather homogeneous and erratic (Figure 4), while the variation in the contribution of precipitation variables (i.e., BIO19 and BIO14 in Figure 4) show a clear pattern. Tropical and subtropical basins were significantly less influenced by precipitation variables than northernmost and southernmost high-latitude basins. Interestingly, in the northern hemisphere the contribution of precipitation variables seems to be higher when the considered extent increases.

Discussion

In this study, we were able to demonstrate that the identification of the factors accounting for the distribution range of the species clearly depend on the extent to which these relationships are examined, a topic which has long been hypothesized (MacArthur 1972; Wiens 2015) and that only recently has been considered fundamental to correctly estimate the predictor functions relating species occurrences with environmental variables (Barve et al. 2011; Acevedo et al. 2012). Our study also shows that the influence of explanatory variables changes spatially in a non-stationary way, as demonstrated in other studies (Osborne and Suarez-Seoane 2002; Foody 2004; Bickford and Laffan 2006; Hortal et al. 2011; Mcnew et al. 2013; Holloway and Miller 2015). In our study, altitude is an important factor that influences the smaller considered extent distribution patterns, but not at global extents. Factors associated to temperature are also important at local extent, whereas factors associated with precipitation seem to be more important at global extents. Moreover, the relative contribution of factors to explain the distribution of freshwater fish species varies from region to region, due partially to the different species present in each area, but mainly due to the spatial variation in the importance of the factors accounting for freshwater species' distributions. This spatial variability in the importance of environmental factors has been formerly documented in the case of fishes (Windle et al. 2010; Radinger et al. 2015). Here, a latitudinal and geographical gradient in the importance of abiotic factors has been shown.
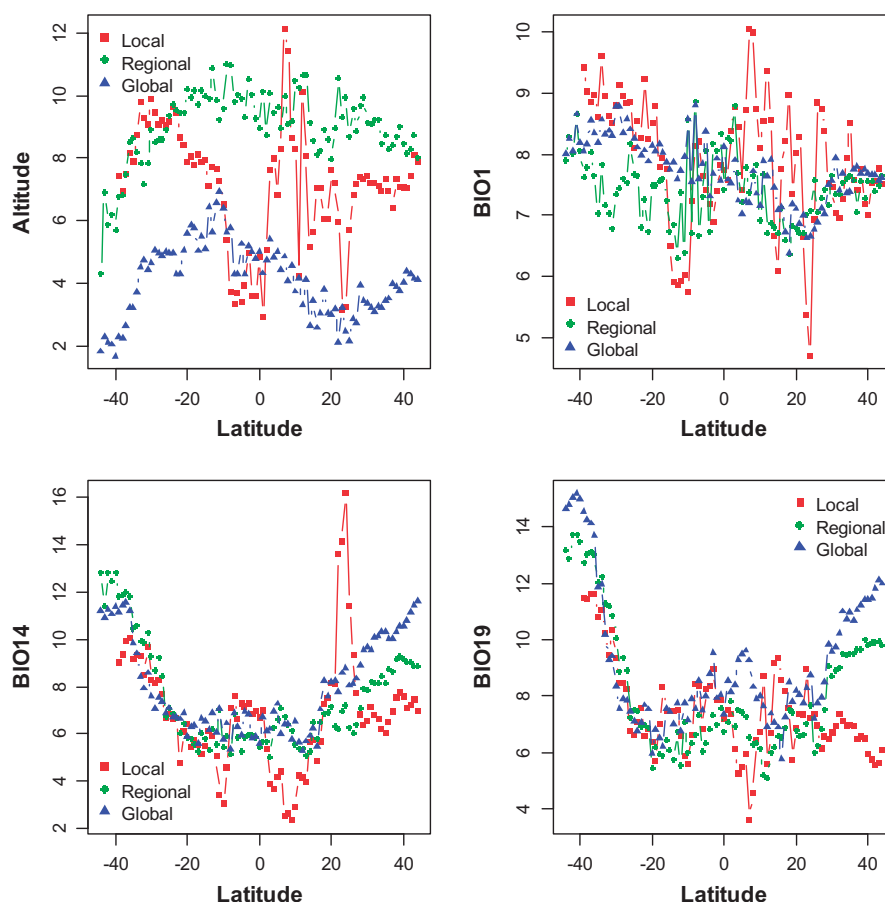
Figure 4. Mean latitudinal contribution to the values of the instability index (%) of altitude, annual mean temperature (BIO1), precipitation of driest month (BIO14), and precipitation of coldest quarter (BIO19), depending on the considered geographical extent. Higher latitudes (90–45° both north and south) were not included in the plots due to the low number of species in those regions.

The frequency distribution of presence data compared with those existing in the occupied basins (Figure 2) is in agreement with classic environmental gradients (Griffiths et al. 2014): the frequency of occurrence of freshwater fishes diminishes at temperatures below 13 °C and precipitations lower than 1,200 mm, approximately. However, freshwater fishes in tropical and subtropical basins are comparatively less influenced in their distributions by precipitation variables than those located in northernmost and southernmost high-latitude basins. Interestingly, in the northern hemisphere the contribution of precipitation variables seems to be higher when the considered extent increases, so that precipitation variables increase in importance at higher latitudes, both south and north, although this latitudinal trend is more or less pronounced depending on the extent considered. Thus, when the variability in the climatic conditions is minimized by considering only a geographical background limited to the basins in which a species occurs, the comparative relevance of precipitation variables in explaining occurrence and distribution seems to be higher than that of temperature under temperate conditions. We may hypothesize that the higher levels of annual precipitation in tropical and subtropical basins suggest that precipitation is not a limiting factor in these areas, so its effect on fish distribution is not as important as in higher latitudes, where hydrological variations related to precipitation may be a more limiting factor when compared with temperature.

Of course, the procedure proposed to identify the most important factors in determining the occurrence of individual freshwater fishes should be managed with caution. The provided results are not only limited by potential errors in the number and characteristics of the used environmental variables (Van Neil et al. 2004), the existence of false presence records (Tyre et al. 2003) or survey biases (Pelayo-Villamil et al. 2015), but also by the lack of consideration of the historical biogeography of the species (Wiens and Donoghue 2004) or species interactions (González-Salazara et al. 2013). However, in spite of this, our findings clearly suggest that different predictors can be detected as relevant at different scales. This fact has important implications for studies about SDMs and ENMs, for which predictor selection is an important issue affecting models' predictive ability (Austin 2007). We suggest when modeling species distribution, that careful attention should be paid to the selection of the extent used to estimate the SDMs and ENMs as previously suggested (Barve et al. 2011; Acevedo et al. 2012), and that a single set of predictors for each species should not be used, but instead different predictors and predictor functions for each region where the species is present (Mcnew et al. 2013). These are not new issues and we are aware of how its consideration complicates the building of models capable of providing reliable estimations of species distributions. How to select the most appropriate extent for each species, and how to manage the variation in the parameters of the environmental predictors obtained at different extents to generate reliable regional predictions are key issues to address (see, e.g., Suárez-Seoane et al. 2014). Be that as it may, modelers should justify how they have managed these 2 questions or why it has not done so.

Supplementary Material

Supplementary material can be found at https://academic.oup.com/cz.

References

Acevedo P, Jimenez-Valverde A, Lobo JM, Real R, 2012. Delimiting the geographical background in species distribution modelling. *J Biogeogr* **39**: 1383–1390.

Austin M, 2007. Species distribution models and ecological theory: a critical assessment and some possible approaches. *Ecol Model* **200**:1–19.

Barve N, Barve V, Jiménez-Valverde A, Lira-Noriega A, Maher SP et al., 2011. Crucial role of the accessibility area in ecological niche modeling and species distribution modeling. *Ecol Model* **222**:1810–1819.

Bickford S, Laffan S, 2006. Multi-extent analysis of the relationship between pteridophyte species richness and climate. *Glob Ecol Biogeogr* **15**:588–601.

Brown JH, 1995. *Macroecology*. Chicago (IL): University of Chicago Press.

Chust G, Petrus JL, Ducrot D, Ventura D, 2004. Scale dependence of insect assemblages in response to landscape pattern. *Landsc Ecol* **18**:41–57.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G et al., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**:27–46.

Dubois DM, 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, Project Mer. Commision Interministérielle de la Politique Scientifique. Liège.

Foody GM, 2004. Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Glob Ecol Biogeogr* **13**:315–320.

Fotheringham AS, Brundson C, Charlton M, 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. West Sussex, UK: John Wiley and Sons, Ltd.

García-Roselló E, Guisande C, González-Dacosta J, Heine J, Pelayo-Villamil P et al., 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 et al., 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 et al., 2015. Can we derive macroecological patterns from primary GBIF data? *Glob Ecol Biogeogr* **24**:335–347.

Gillingham PK, Huntley B, Kunin WE, Thomas CD, 2012. The effect of spatial resolution on projected responses to climate warming. *Divers Distrib* **18**: 990–1000.

Gaston K, 2003. *The Structure and Dynamics of Geographic Ranges*. Oxford: Oxford University Press.

González-Salazara C, Stephensb CR, Marquet PA, 2013. Comparing the relative contributions of biotic and abiotic factors as mediators of species' distributions. *Ecol Model* **248**:57–70.

González-Vilas L, Guisande C, Vari RP, Pelayo-Villamil P, Manjarrés-Hernández A et al., 2016. Geospatial data of freshwater habitats for macroecological studies: an example with freshwater fishes. *Int J Geogr Inf Sci* **30**: 126–141.

Griffiths D, McGonigle C, Quinn R, 2014. Climate and species richness patterns of freshwater fish in North America and Europe. *J Biogeogr* **41**: 452–463.

Guisande C, 2016. Estimation of the relative importance of factors affecting species distribution based on stability concept. R package version 1.3. Available from: http://CRAN.R-project.org/package=SPEDInstabR (accessed 6 November 2016).

Guisande C, Heine J, Gomzález-DaCosta J, García-Roselló E, 2014. Vigo, Spain: RWizard Software. University of Vigo.

Guisande C, Barreiro A, Maneiro I, Riveiro I, Vergara-Castaño AR et al., 2006. *Tratamiento de Datos*. Madrid: Ediciones Díaz de Santos.

Guisande C, Barreiro A, Vaamonde A, 2011. *Tratamiento de Datos con R, Statistica y SPSS*. Madrid: Ediciones Díaz de Santos.

Guisande C, García-Roselló E, Heine J, González-Dacosta J, Gozález Vilas L et al., 2016. SPEDInstabR: an algorithm based on a fluctuation index for selecting predictors in species distribution modeling. *Ecol Inform* **37**: 18–23.

Guisande C, Manjarrés-Hernández A, Pelayo-Villamil P, Granado-Lorencio C, Riveiro I et al., 2010. IPez: an expert system for the taxonomic identification of fishes based on machine learning techniques. *Fish Res* **102**:240–247.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A, 2005. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* **25**: 1965–1978.

Holloway P, Miller JA, 2015. Exploring spatial scale, autocorrelation and nonstationarity of bird species richness patterns. *ISPRS Int J Geo-Inform* **4**: 783–798.

Hortal J, Diniz-Filho JAF, Bini LM, Rodriguez MA, Baselga A et al., 2011. Ice age climate, evolutionary constraints and diversity patterns of European dung beetles. *Ecol Lett* **14**:741–748.

Johnson DH, 1980. The comparison of usage and availability measurements for evaluating resource preference. *Ecology* **61**:65–71.

Lobo JM, 2016. The use of occurrence data to predict the effects of climate change on insects. *Curr Opin Insect Sci* **17**:62–68.

MacArthur RH, 1972. *Geographic Ecology*. New York (NY): Harper and Row.

Margalef R, 1968. *Perspectives in Ecological Theory*. Chicago (IL): The University of Chicago Press.

Mcnew LB, Gregory AJ, Sandercock BK, 2013. Spatial heterogeneity in habitat selection: nest site selection by greater prairie-chickens. *J Wildl Manage* **77**:791–801.

Niamir A, Skidmore AK, Toxopeus AG, Real R, 2016. Use of taxonomy to delineate spatial extent of atlas data for species distribution models. *Glob Ecol Biogeogr* **25**:227–237.

Osborne PE, Suarez-Seoane S, 2002. Should data be partitioned spatially before building large-scale distribution models? *Ecol Model* **157**:249–259.

Pearson RG, Dawson TP, 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Glob Ecol Biogeogr* **12**:361–371.

Pelayo-Villamil P, Guisande C, González-Vilas L, Carvajal-Quintero JD, Jiménez-Segura LF et al., 2012. ModestR: Una herramienta informática para el estudio de los ecosistemas acuáticos de Colombia. *Actual Biol* **34**: 225–239.

Pelayo-Villamil P, Guisande C, Vari RP, Manjarrés-Hernández A, García-Roselló E et al., 2015. Global diversity patterns of freshwater fishes: potential victims of their own success. *Divers Distrib* **21**:345–356.

Peterson AT, Soberón J, Pearson RG, Anderson R, Martínez-Meyer E et al., 2011. *Ecological Niches and Geographic Distributions*. Princeton (NJ): Princeton University Press.

Radinger J, Wolter C, Kail J, 2015. Spatial scaling of environmental variables improves species–habitat models of fishes in a small, sand-bed lowland river. *PLoS One* **10**:e0142813.

Soberón J, Peterson AT, 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodivers Inform* **2**:1–10.

Steffan-Dewenter I, Münzenberg U, Burger C, Thies C, Tscharntke T, 2002. Scale-dependent effects of landscape context on three pollinator guilds. *Ecology* **83**:1421–1432.

Suárez-Seoane S, Virgós E, Terroba O, Pardavila X, Barea-Azcón JM, 2014. Scaling of species distribution models across spatial resolutions and extents along a biogeographical gradient: the case of the Iberian mole *Talpa occidentalis*. *Ecography* **37**:279–292.

Tyre AJ, Tenhumberg B, Field SA, Niejalke D, Parris K et al., 2003. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecol Appl* **13**:1790–1801.

Van de Pol M, Bailey LD, McLean N, Rijsdijk L, Lawson CR et al., 2016. Identifying the best climatic predictors in ecology and evolution. *Methods Ecol Evol* **7**:1246–1257.

Van Neil KP, Laffan SW, Lees BG, 2004. Effect of error in the DEM on environmental variables for predictive vegetation modelling. *J Veg Sci* **15**: 747–756.

Wiens JJ, 2015. The niche, biogeography and species interactions. *Philos Trans R Soc B* **366**:2336–2350.

Wiens JJ, Donoghue MJ, 2004. Historical biogeography, ecology and species richness. *Trends Ecol Evol* **19**:639–644.

Windle MJS, Rose GA, Devillers R, Fortin MJ, 2010. Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic. *ICES J Mar Sci* **67**:145–154.