International Journal of Geographical Information Science, 2015
http://dx.doi.org/10.1080/13658816.2015.1072629

Taylor & Francis
Taylor & Francis Group

Geospatial data of freshwater habitats for macroecological studies: an example with freshwater fishes

Luis González Vilas[a], Cástor Guisande[a]*, Richard P. Vari[b], Patricia Pelayo-Villamil[c], Ana Manjarrés-Hernández[d], Emilio García-Roselló[e], Jacinto González-Dacosta[e], Jürgen Heine[e], Elisa Pérez-Costas[a], Carlos Granado-Lorencio[f], Antoni Palau-Ibars[g] and Jorge M. Lobo[h]

[a]*Facultad de Ciencias, Universidad de Vigo, Vigo, Spain;* [b]*Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA;* [c]*Grupo de Ictiología, Universidad de Antioquia, Medellín, Colombia;* [d]*Instituto Amazónico de Investigaciones-IMANI, Universidad Nacional de Colombia, Leticia, Colombia;* [e]*Departamento de Informática, Edificio Fundición, Universidad de Vigo, Vigo, Spain;* [f]*Departamento de Biología Vegetal y Ecología, Facultad de Biología, Universidad de Sevilla, Sevilla, Spain;* [g]*Biodiversidad, I+D+i y Recursos Hídricos, Dirección de Medio Ambiente y Cambio Climático España y Portugal, ENDESA, Lleida, Spain;* [h]*Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales (CSIC), Madrid, Spain*

(*Received 19 March 2015; accepted 9 July 2015*)

Global data sets are essential in macroecological studies. File formats of the few available data sets of freshwater ecosystems, however, are either incompatible with most macroecological software packages, incomplete, or of coarse spatial resolutions. We integrated more than 460 million geographical coordinates for freshwater habitats in the FRWater data set, partitioned into seven different habitats (lentic, wetlands, reservoirs, small rivers, large rivers, small ditches, large ditches, small channels, large channels, small drains and large drains) in ModestR (http://www.ipez.es/ModestR). A comprehensive collection of geospatial rasters was assembled, one for each of the seven freshwater habitats, with the area in km^2 occupied by each habitat presented in cells of 5 arc-minute resolution. The utility of FRWater was evaluated using hierarchical partitioning via the identification of the contribution of the seven different freshwater habitats to both species richness and rarity. To this end, we used a data set of 836,123 geographical records of the 16,216 species of freshwater fishes recognized as valid by systematists at the end of 2014. Areas in North America and Europe are the most detailed in the FRWater data set, evidencing the higher quality of data sources in those regions. The number of geographical coordinates is much lower for Africa, Asia, Australia, and South America where many water bodies remain unmapped. In light of the variation in information quality at continental levels, we performed and present comparative analyses for Europe versus South America at local (5′ × 5′ grid cells) and regional (5° × 5° grid cells) scales. The relative contribution of small rivers to both species richness and rarity was highest under almost all analyses, followed by lentic habitats and large rivers. The areas of different habitats moreover explained a relatively high proportion of the observed variance in geographic rarity. Our findings corroborate previous findings that the greater contribution of small rivers to species richness is probably due to these habitats promoting geographical rarity. Hence, species richness is favored by the isolation resultant from, and the refuges associated with, small river basins and via the diversification processes promoted by such isolation.

Keywords: freshwater habitats; freshwater fishes; species richness; rarity

*Corresponding author. Email: castor@uvigo.es

Introduction

Macroecology searches for generalized patterns over broad spatial and temporal scales (Gaston and Blackburn 2000). As such it is highly dependent on the availability of reliable global scientific data sets to detect patterns and identify the probable causal factors most likely controlling species abundance and distribution and, hence, species richness.

The last two decades have seen the development of multiple global databases applicable to environmental and ecological questions as a consequence of the growing availability of global satellite data, vast improvements in data processing and storage capabilities, and enhanced data access through online servers and web portals (Hastings *et al*. 1991, Graham *et al*. 2004, UNEP 2012). The wide use of global data sets facilitated the growth of macroecological studies, particularly the development of species distribution model applications (Alroy 2003, Robinson *et al*. 2011), notwithstanding limitations such as incomplete and fragmented data and geographic bias (Bedia *et al*. 2013, Thorson *et al*. 2014). Available information in global data sets, nonetheless, remains primarily restricted to terrestrial and marine rather than freshwater ecosystems.

The area represented by different freshwater habitats may be an important factor affecting not only freshwater aquatic species but also terrestrial ones; however, detailed global geographic information on freshwater habitats is very scarce. HYDRO1k is a global geographic database including streams and drainage basins data derived from the USGS 30 arc-second digital elevation model of the world (US Geological Survey 2000). Another global product based on digital elevation data is HydroSHEDS, which provides different geo-referenced data sets at various scales, such as river networks, watershed boundaries, drainage directions, and flow accumulations (Lehner *et al*. 2008, Lehner and Grill 2013). Lehner and Döll (2004) proposed the development of a global database of lakes, reservoirs, and wetlands by developing a Geographic Information System (GIS) combining the best currently available data sources. The Global Water System Project Digital Water Atlas, in turn, compiles data from different sources to create a set of annotated maps, such as the global distribution of large water reservoirs and dams (Lehner *et al*. 2011). Another global database of freshwater habitats is Natural Earth (www.naturalearthdata.com), a collection of vector and raster maps that includes river, lake, and reservoir data at three different spatial scales. Arbault *et al*. (2014), in turn, collected data from different sources to develop a global database of rivers and streams with a focus on energy-level evaluation for those systems. Other freshwater resource data sets are available at national, regional, or local scales from government and/or nonpublic institutions, such as the Australian Hydrological Geospatial Fabric (Geofabric, http://www.bom.gov.au/water/geofabric/index.shtml).

Unfortunately, several impediments hinder the application of all these aforementioned data sets in macroecological studies. First, the conversion of the data sets to file formats compatible with modeling software is rarely straightforward. Second, data assembly is usually challenging and time-consuming since the databases use different structures at varying spatial resolutions. Finally, data sets are commonly incomplete and of coarse spatial resolution.

The aim of this study is to present a new global data set of freshwater habitats (FRWater), which is integrated into ModestR, a powerful software aimed at working

with species distribution maps and taxonomic data (García-Roselló *et al*. 2013). FRWater is available at a spatial resolution of 5 arc-minutes in the ASCII raster format used by many species distribution modeling packages, such as Garp (http://www.nhm.ku.edu/desktopgarp) and MaxEnt (https://www.cs.princeton.edu/~schapire/maxent).

The main advantage of FRWater compared to other freshwater global geographic data sets is that users can utilize the tools and facilities available in ModestR to manage and analyze information about species distributions within FRWater without the necessity of data assembly and/or pre-processing. Other strengths of FRWater are its global coverage, refined spatial resolution (1 arc-second), and its discrimination among seven different types of freshwater aquatic habitats.

Large-scale variables associated with the variation of species richness have been widely studied in the case of freshwater fish species (e.g., Pelayo-Villamil *et al*. 2015). However, the relative contribution of different aquatic habitats to species richness is usually not incorporated in global assessments. Herein we show the utility of FRWater to determine the contribution of the different freshwater habitats to explain the global variation in the species richness and rarity of freshwater fish species.

Methods

FRWater data set

Geographic data for the creation of FRWater were taken from OpenStreetMap, a free collaborative mapping project (http://www.openstreetmap.org). OpenStreetMap was selected as the data source for three primary reasons. First, data are available under the Open Database License, and users are thus allowed to freely share, modify, and use the database. Second, it provides global coverage, integrating geospatial data from different sources with data continuously improved and updated through user contributors. Finally, it uses a topological data structure with four core elements and a tagging system that simplifies the identification of water features.

The FRWater data set provides a collection of lines and polygons showing the geographic borders of freshwater habitats. Seven different types of freshwater habitats are included in the data set, with a division of some habitats into small and large categories (Table 1). Small water features such as small rivers, channels, ditches, and drains are depicted as lines, while all other freshwater habitat types are represented by polygons (Table 1). Subsequently during data rasterization, small water features are rasterized to the minimum resolution of 1 arc-second. Note that the ground area covered by 1 arc-second is latitude-dependent. Thus, polygons are water bodies with an area larger than approximately 955 m^2 at the equator, but only slightly larger than 100 m^2 near the poles. This discrepancy is subsequently corrected by the use of ModestR (see below). In addition to the freshwater habitats *per se*, the FRWater data set also includes islands within the large habitats (polygons). Therefore, island area is accounted for when determining the total area of the water body in question.

Integration of OpenStreetMap data into Modest R was a complicated process due to the different structure of the geographic information and the large volume of data (more than 460 million geographical coordinates). The process involved four steps:

Table 1. Types of freshwater habitats included in the data set.

Habitat	Type	Description
Lentic habitats	Polygons	Lakes, ponds, bogs
Reservoirs	Polygons	Artificial and natural reservoirs
Wetlands	Polygons	Marshlands, swamps
Large rivers	Polygons	Main rivers, deltas, estuaries
Small rivers	Lines	Streams, creeks, rivulets, brooks
Large channels	Polygons	Man-made waterways for transportation; large waterways for
Small channels	Lines	irrigation
Large ditches	Polygons	Narrow nonlined artificial waterways for draining land or
Small ditches	Lines	removing storm water
Large drains	Polygons	Minor lined artificial waterways for carrying storm water or
Small drains	Lines	industrial discharge; channelized streams

(1) The identification based on the OpenStreetMap tagging system (Table 2) and download of ways and multipolygon relations associated with freshwater habitats. A way is an ordered list of geographical coordinates describing open lines or closed polygons, while a multipolygon relation is a combination of ways that forms one or more closed polygons (see more information at http://www.openstreetmap.org). In the case of multipolygon relations, the ways defined as *outer* were associated with their corresponding habitat (Table 2), while *inner* ways were identified as islands.
(2) The download of the geographical coordinates associated with the previously identified elements and the generation of a preliminary DAT file for each way or relation, assigning to each a unique ID and the corresponding type of habitat (Table 2).
(3) The importation of the DAT files into ModestR.
(4) Rasterization to a spatial resolution of 1 arc-second.

FRWater data set was integrated into ModestR, which consists of four applications: MapMaker, a GIS environment to build and analyze species distribution maps; DataManager, which allows the management of species and taxonomic databases;

Table 2. Habitat assigned to blocks in FRWater data set according to the tags (key-value) in OpenStreetMap.

OpenStreetMap tag (key-value)	FRWater habitat
Natural – water	Lentic habitats
Natural – wetlands	Wetlands
Waterway – river	Large rivers (polygons)/small rivers (lines)
Waterway – riverbank	Large rivers (polygons)/small rivers (lines)
Waterway – stream	Large rivers (polygons)/small rivers (lines)
Waterway – canal	Large channels (polygons)/small channels (lines)
Waterway – ditch	Large ditches (polygons)/small ditches (lines)
Waterway – drain	Large drains (polygons)/small drains (lines)
Water – reservoir	Reservoirs
Landuse – reservoir	Reservoirs

MRFinder, a tool to find the species present in specific areas and MRMapping for building maps (http://www.ipez.es/ModestR, García-Roselló *et al*. 2013). ModestR with the integrated FRWater data set requires Microsoft Windows XP or a later version, a 64-bit CPU, a minimum of 4 GB of RAM memory, and 20 GB of hard-drive space. ModestR is available via mirrors in several countries to facilitate access to this large software package. Additional useful features available in the different applications of ModestR are summarized in Table 3. More details about error-checking, quality control, and quality assurance procedures are included in the supplemental data.

An example with freshwater fishes

Species distributions

The data set of geographical records for freshwater fishes developed by Pelayo-Villamil *et al*. (2015) was updated to reflect taxonomic changes and new species described to the end of 2014 using the menu facility of ModestR (Pelayo-Villamil *et al*. 2012, García-Roselló *et al*. 2013). The data was filtered using the data cleaning facilities available in ModestR (García-Roselló *et al*. 2014). These include data cleaning when downloading records from GBIF and habitat data cleaning to remove duplicates (records of the same specimen sent by different collections). Additional data cleaning involved the removal of common errors in GBIF data, e.g., records with geographic coordinates of 0° longitude and 0° latitude, and records for which the longitude and latitude values are identical and likely represent erroneous repetitive data entry. Finally, erroneous synonyms were corrected and habitat filtering was applied under which records in marine settings were considered invalid.

We did not make any distinction between native and alien species, so these were handled in comparable modes. At the end of 2014, 16,216 species of freshwater fishes were recognized as valid by systematists and are available in IPez (http://www.ipez.es, Guisande *et al*. 2010). Of these, 13,379 species (82.5% of the total) have associated geographical information with a total of 836,065 records.

For the analyses, we used range maps, instead of point-to-grid maps, with the range maps generated by estimating the extent of occurrence (EOO) for each species with ModestR (García-Roselló *et al*. 2015). To do this, we selected an α-shape procedure with an α value of 6, because global scale studies of macroecological patterns demonstrate that this simple method is a more parsimonious option for extrapolating species distributions from primary data (García-Roselló *et al*. 2015).

Extent of occurrence area

Once range maps have been drawn for each species, we use the following equation to calculate the EOO area of each species (probable distribution in km^2) with ModestR so as to avoid the biases generated by the use of geographical coordinates:

$$1.852 \times \frac{12,756.2\pi}{21,600} \cos\left(\text{latitude} \times \frac{\pi}{180}\right)$$

The above equation is used for the calculation of the ground area (in km^2) corresponding with an individual pixel of $1' \times 1'$. The value 1.852 is a nautical mile (distance of 1 minute

Table 3. Main features of ModestR software.

Application	Main features
MapMaker	Expert-drawn maps
	Occurrence-based maps, with online access to GBIF data
	Importation of shapefiles, KML, CSV, ASC rasters
	Exportation of species distribution and habitat maps to high-resolution image files, shapefiles, KML, rasters
	Cleaning facilities by discriminating between habitats using environmental-based rules and dispersal capacity
	Estimation of AOO, extent of occurrence (by convex hull, alpha shape, or kernel density), environmental occurrence, kernel density maps, or niche of occurrence
	Environmental layers (polar coordinates systems using several environmental variables)
	Integration of environmental data rasters
	Raster clipping using species presence areas or any arbitrary template
MRFinder	Searching for species present in specified areas, discriminating between habitats
	Pre–post filtering species searching by taxonomy
	Rare species and custom categories filters
	Integrated database of predefined areas (World administrative areas and World river basins
	Importation/exportations of shapefiles, KML
	Summary of environmental conditions and of the area covered (in km^2), discriminating by different freshwater habitats
	Summary of species found by specified area
	Richness measures, latitudinal gradients, rarity index, AOO index, EOO index, patch index, and latitudinal range index in selected areas
DataManager	Creation of species distribution databases structured by taxonomy
	Taxonomy importation from CSV, phyloXML or ITIS database, and exportation to CSV
	Species distribution importation from CSV, shapefiles, Darwin core archives, probability rasters (e.g., Maxent)
	Batch downloading of distribution data for any taxonomic rank from GBIF
	Estimation of area covered (in km^2), discriminating by different freshwater habitats
	Estimation of the richness in selected freshwater habitats
	Bulk species distribution data exportation to CSV, raster, Darwin core template, or high resolution image files
	Estimation of AOO, extent of occurrence (by convex hull, alpha shape, or kernel density), and data cleaning features
	Bulk raster clipping feature using species AOO or EOO
	Estimation of environmental variables contribution to species presence
	Richness measures, latitudinal gradients, rarity index, AOO index, EOO index, patch index, and latitudinal range index for any taxonomic rank
MRMapping	Creation of maps with multiple species distribution data
	Drag & drop feature to add data from any ModestR database or map file
	Distribution data grouping feature by any taxonomic rank (classes, orders, families, etc.)
	Importation of shapefiles and KML
	Exportation of distribution and habitat maps to high resolution image files, shapefiles, KML, rasters
	Merging and intersection calculation between any set of species or taxonomic ranks
	Estimation of AOO for any taxonomic rank
	Estimation of percentage of spatial overlap between any set of species or taxonomic ranks (orders, families, species, etc.)

of arc measured along any meridian) in km, 12,756.2 is twice the radius of the earth in km and finally 21,600 is the number of arc-minutes in a full circumference. This equation was applied to all the pixels occupied by the species using the central location of each pixel as the latitude value, and all the obtained values were finally summed to calculate the distribution area in km^2.

Subsequently, we estimate the average EOO area values for all the species present in each cell (i.e., average EOO). Rare species may have a naturally restricted range within a widespread habitat, may be specialists in narrowly distributed habitats, or may have had the size of their original ranges reduced by human activity. Although concepts for considering a species as rare and methods to estimate rarity may differ (Laffan and Crisp 2003), the common factor is the use of small range sizes (Flather and Sieg 2007) and can be considered interchangeable (Kunin and Gaston 1997). In our case, a low value of the EOO index indicates the occurrence of a higher number of species with small geographic range sizes or geographical rarity (Gaston and Fuller 2009).

Statistical analyses

The contribution of each of the seven freshwater habitats to the explanation of the variation in species richness and rarity values was estimated by using a hierarchical partitioning procedure directed to alleviate the frequent multicollinarity of predictors (Chevan and Sutherland 1991), which was estimated with the *hier.part* function in the R software (Walsh and Mac Nally 2014). The advantage of hierarchical partitioning over stepwise-selection techniques is that hierarchical partitioning is focused on drawing inferences about the likely causality of variables, whereas stepwise-selection techniques focus on finding a single best predictive function (Mac Nally 2000). The presence of multicollinearity is, in turn, considered calculating the variance inflation factor which provides an index that measures how much the variance of an estimated regression coefficient is increased because of collinearity (Fox and Weisberg 2011). This analysis was performed with the package *usdm* (Naimi 2014). This technique was used to exclude, among correlated habitats, those with lower causal contributions to the dependent variable.

All scripts used for the statistical analyses are based on previous ones (Guisande *et al*. 2006, 2011), and the most appropriate graph for each statistical analysis was determined following Guisande and Vaamonde (2012). These scripts are integrated into the software RWizard (Guisande *et al*. 2014).

Integration of geographic records of the species and FRWater

ModestR previously allowed the organization of a set of different environmental variables (García-Roselló *et al*. 2013) and now additionally includes global geographic data for freshwater habitats (the FRWater data set). One of the purposes of this feature is to integrate the exportation of geographic records of species distributions with those of the environmental variables and/or area occupied by any of the above described freshwater habitats into grid cells of different sizes. We used the DataManager application of ModestR (García-Roselló *et al*. 2013) to export the species richness jointly with the data of the area occupied by each one of the considered freshwater habitats within $5' \times 5'$ (local scale) and $5° \times 5°$ (regional

scale) grid cells. In the case of rarity, the data were exported to a single grid cell size of 60′ × 60′.

Results

Table 4 summarizes the FRWater data set by habitat utilizing two parameters: area of freshwater habitat (Figure 1(a)) and number of geographical coordinates (Figure 1(b)). Seasonal water bodies or waterways without a permanent flow (geographical elements tagged as intermittent in OpenStreetMap) were also included in the analysis. Areas were computed from the rasterized map with a 1 arc-second resolution. The data set, which includes 14,685,790 blocks (lines and polygons), is composed of more than 460 million geographical coordinates. The total area of freshwater habitat is approximately 2.86 million km^2 (around 1.9% of total Earth land area). In terms of numbers of geographical coordinates, the small river habitat is the most abundant category of all the considered freshwater habitats (45.7%). Nonetheless, lentic habitats cover the largest area in total, encompassing more than one-half of the area (57.7%). Despite having a lower number of

Table 4. Number of geographical coordinates ($\times 10^5$) and area ($km^2 \times 10^3$) computed for each habitat in the FRWater data set.

Habitat	Europe	America	Africa	Asia	Oceania	Total	%
Number of geographic coordinates ($\times 10^5$)							
Lentic habitats	331.0	988.4	12.3	121.0	23.8	1476.5	31.98
Reservoirs	10.7	23.3	1.8	11.2	0.5	47.5	1.03
Wetlands	87.0	319.3	5.9	12.2	4.5	429.0	9.29
Large rivers	128.2	162.2	12.3	65.2	6.9	374.8	8.12
Small rivers	378.1	1351.2	81.9	354.2	24.7	2190.0	47.45
Large channels	3.6	0.4	0.2	0.5	0.1	4.7	0.11
Small channels	9.0	10.0	1.4	8.1	0.3	28.7	0.62
Large ditches	0.1	0.3	0.1	0.0	0.0	0.5	0.01
Small ditches	18.7	15.7	1.9	2.7	0.1	39.1	0.85
Large drains	0.2	0.1	0.0	0.0	0.0	0.3	0.01
Small drains	15.9	3.9	1.2	3.0	0.4	24.4	0.53
Total	982.4	2874.8	118.9	578.2	61.3	4615.5	
%	21.3	62.3	2.6	12.5	1.3		
Area ($km^2 \times 10^3$)							
Lentic habitats	209.1	801.6	232.7	338.8	65.2	1647.4	57.66
Reservoirs	19.7	86.1	16.5	39.5	0.9	162.8	5.7
Wetlands	67.3	196.9	100.3	76.7	5.2	446.4	15.62
Large rivers	27.1	149.2	39.2	157.2	4.9	377.6	13.21
Small rivers	33.5	99.1	16.8	50.6	5.0	204.8	7.17
Large channels	0.4	0.1	0.04	0.1	0.1	0.7	0.024
Small channels	1.6	1.7	0.56	4.1	0.1	8.1	0.29
Large ditches	0.01	0.1	0.001	0.001	0	0.1	0.004
Small ditches	2.5	2.3	0.28	0.5	0.007	5.6	0.19
Large drains	0.01	0.004	0.0003	0.01	0.0003	0.03	0.002
Small drains	2.2	0.4	0.2	0.8	0.1	3.6	0.13
Total	363.2	1337.5	406.6	668.4	81.4	2857.2	
%	12.7	46.8	14.2	23.4	2.9		

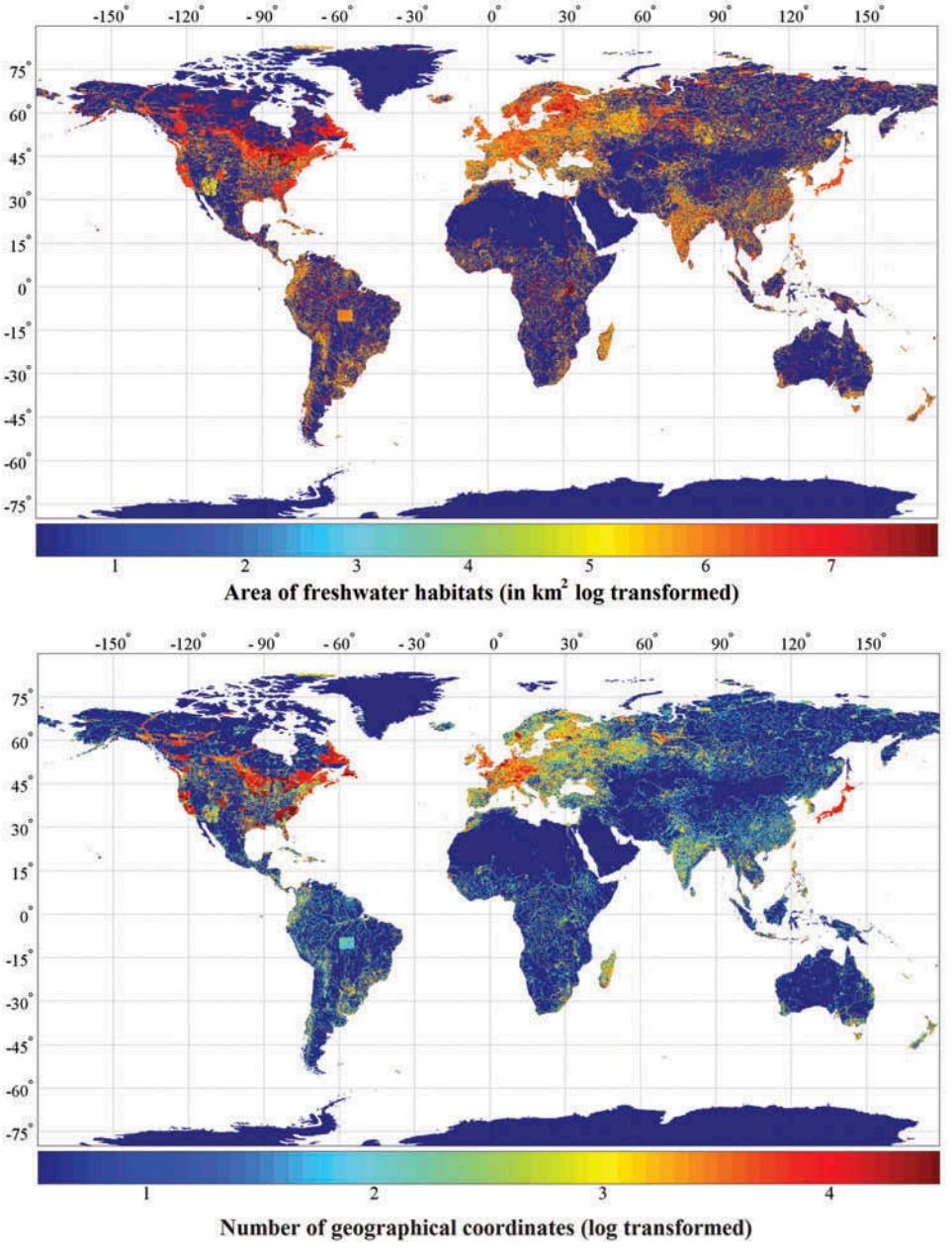

Figure 1. World map showing in 5′ grid cells: (a) the area of freshwater habitat in km^2 log transformed and (b) the number of geographic coordinates obtained from the FRWater data set.

coordinates, the lentic, wetland, and large river habitats each also have greater individual total areas than does the small river habitat which covers only 7.2% of the total area. Remarkably, reservoirs cover 5.7% of the total area, despite involving less than 1% of the coordinates. It is critical to note that in so far as the small

habitats (channels, ditches, drains, and rivers) are lines, it is impossible to estimate their areas to the same accuracy as the polygons (large habitats). Therefore, the reported summary areas in km^2 for small habitats should be taken with caution.

It is also important to emphasize that the available FRWater data set must be considered provisional because many water bodies in some regions of the World remain unmapped. Thus, information quality varies among areas. Areas in North America and Europe (Figure 2(a) and (b)) are overall the most detailed, evidencing a higher data source quality, whereas in Africa (Figure 2(c)), Asia (Figure 2(d)), Australia, and South America, the numbers of geographical coordinates are much lower. Moreover, some regions have patchy data quality with areas with a greater volume of data proximate to areas with lesser data. Most such patches are in Canada, where some areas lacking information are surrounded by areas with high numbers of geographical coordinates. Conversely, there is also a large patch with high numbers of coordinates in central South America (see Figure 1) that is surrounded by a large area with more limited data. Presumably additional data gathering will allow many areas in South America to look like this patch in some future version of FRWater.

Considering the unequal quality and detail of the available maps for freshwater habitats, we examine the predictive capacity of these habitats on species richness for two contrasting continental areas: Europe and South America. The freshwater habitat with the highest contribution to species richness, in South America at both local (5′ × 5′) and regional (5° × 5°) scales and in Europe at local scale (5′ × 5′), was small

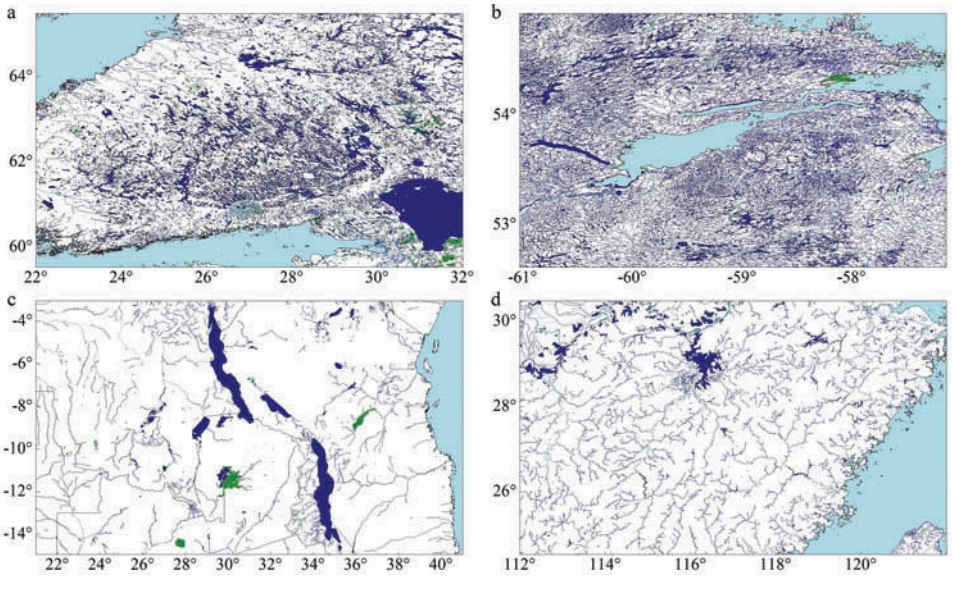

Figure 2. Examples of maps showing the freshwater habitats available in the FRWater data set for four geographical areas: (a) southern Finland, (b) Newfoundland region of Canada, (c) central East Africa, and (d) Southwestern China. Maps were obtained by exporting JGP files with MapMaker (García-Roselló *et al*. 2013). The ModestR menu offers the user the option to specify a specific color for each of the different freshwater habitats and to show all or just a selection of habitats.

rivers followed by lentic and/or large river habitats (Figure 3). The contributions of large and small rivers were similar in Europe and at the regional scale (5° × 5°). A stepwise multiple regression showed that the percentage of variance explained by freshwater habitats was relatively low, particularly in South America, at the local scale but larger at the regional scale.

Due to the lack of important differences in the relative contribution of the habitats to species richness at differing scales, in the analysis of rarity, we studied only a grid size cell of 60′ × 60′. Small rivers were also the habitat with the highest contribution to rarity in both continents (Figure 4). It is important, however, to highlight that freshwater habitats explained a relatively high proportion of the variance observed in the EOO index, 53% in Europe and 58% in South America.

Discussion

Factors affecting geographical rarity in freshwater fishes are unknown because rarity is poorly predicted by climatic and/or productivity variables (Pelayo-Villamil *et al*. 2015). This supports the hypothesis that nonclimatic or nonenergetic factors, such as geographical isolation, could have influenced the distribution of freshwater fish rarity (Hubert and Renno 2006, Leprieur *et al*. 2011, Dias *et al*. 2013, Griffiths *et al*. 2014). Our study demonstrated that one of these factors is the type of freshwater habitat, given that freshwater habitats explained a relatively high proportion of the observed variance in rarity.

Leaving aside the misleading apparent differences resulting from non-natural factors (e.g., different intensities of research focused on freshwater systems in various geographical areas, Pelayo-Villamil *et al*. 2015), it is well known that a limited number of climatic and productivity variables account for species richness among freshwater fishes (Tedesco *et al*. 2005, 2012, Oberdorff *et al*. 2011, Griffiths *et al*. 2014, Pelayo-Villamil *et al*. 2015). It has been hypothesized that another important factor contributing to species richness in freshwater fishes is isolation resultant from river basin boundaries and the refuge and diversification processes promoted by such isolation (Schleuter *et al*. 2012, Dias *et al*. 2013, Toussaint *et al*. 2014, Pelayo-Villamil *et al*. 2015). Our results corroborate this hypothesis, because the small river habitat is the primary contributor explaining the geographical variation of both rarity and species richness.

At broad spatial scales, biodiversity in freshwater habitats is expected to increase with area (Ricklefs 2004); a hypothesis demonstrated in freshwater fishes (Oberdorff *et al*. 1995, Lévêque *et al*. 2008). According to the Modifiable Areal Unit Problem (Fotheringham and Wong 1991), results of spatial analysis are strongly influenced by the definition of areal units, so that the effect of freshwater habitat on rarity and/or species richness might vary with scale. In studies comparing the biodiversity patterns for macrophytes and macroinvertebrates among habitats and at different scales (Williams *et al*. 2004, Davies *et al*. 2008), rivers were identified as the most species-rich habitat at the individual site level (alpha diversity), whereas ponds were found to contribute most to biodiversity at the regional level (gamma diversity). We observed, however, that regardless of spatial scale, the small river habitat was almost always that with the highest contribution to both species richness and rarity for freshwater fishes at both local and regional scales.

Available information in FRWater is presently incomplete due to the lack of hydrological data in various regions of the World. Therefore, it is not advisable to use the data set

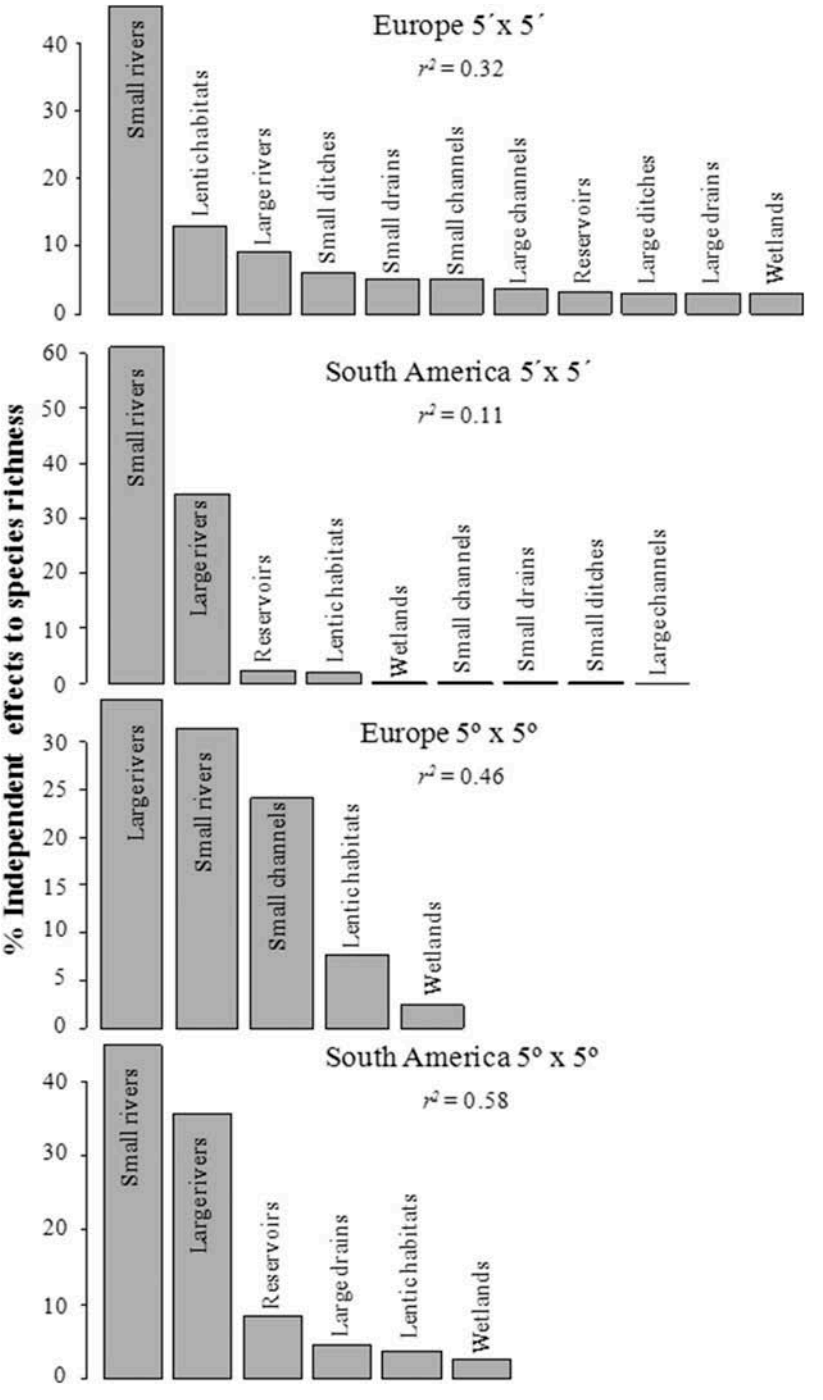

Figure 3. Relative contribution estimated by using hierarchical partitioning, of the different freshwater habitats significantly related with species richness at local (5′ × 5′ grid cells) and regional (5° × 5° grid cells) scales in Europe (−25° to 40° longitude and 34° to 75° latitude) and South America (−82° to −34° longitude and −57° to 13° latitude). The regression coefficient (r^2) of the stepwise multiple regression performed to species richness as the dependent variable, and the freshwater habitats as independent variables is also shown. Dependent and independent variables are all log-transformed.

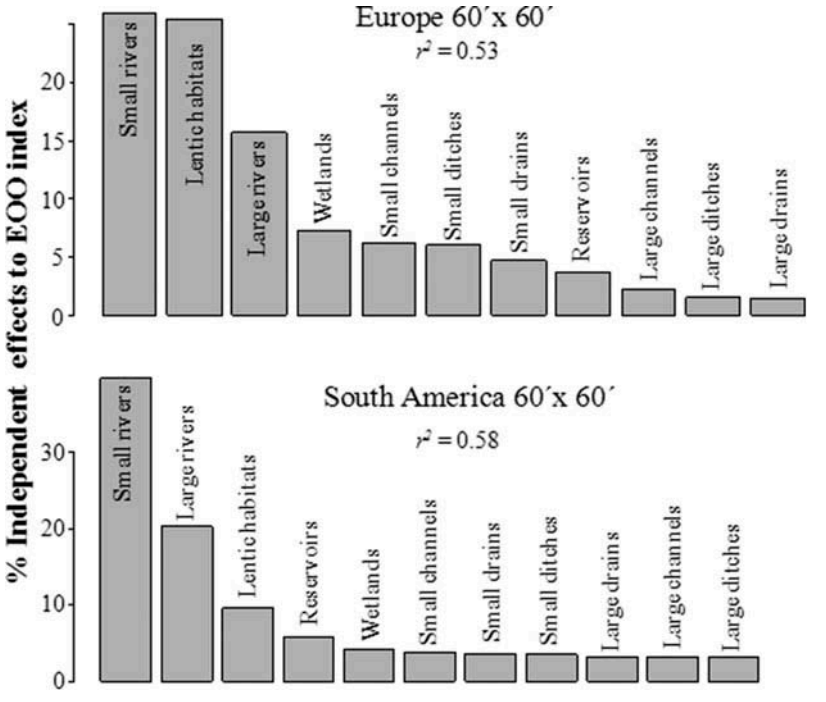

Figure 4. Relative contribution estimated by using hierarchical partitioning of the different freshwater habitats significantly related with the EOO index estimated by using hierarchical partitioning in grid cells of 60′ × 60′ in Europe (−25° to 4° longitude; 34° to 75° latitude) and South America (−82° to −34° longitude; −57° to 13° latitude). Also shown is the regression coefficient (r^2) of the stepwise multiple regression performed to the EOO index as the dependent variable and the freshwater habitats as independent variables. Dependent and independent variables are all log-transformed.

to compare areas that are well mapped with areas poorly mapped, i.e., Europe versus South America, or North America versus South America, etc. Instead, we recommend that each area be evaluated separately. Nonetheless, the available data can be productively used in macroecological studies, particularly in the data-rich regions of Europe and many portions of North America. Furthermore, in our example, we obtained similar results comparing areas that are well (Europe) and poorly mapped (South America) in terms of water bodies and demonstrated an important potential functionality of the FRWater data set. We consider that the joint use of this source of data together with the different tools provided within ModestR will facilitate the visualization and analysis of large data sets for managers, macroecologists, and biogeographers.

Acknowledgments

We thank ENDESA for technical support.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the ENDESA [110/02168].

Supplemental data

Supplemental data for this article can be accessed here.

References

Alroy, J., 2003. Global databases will yield reliable measures of global biodiversity. *Paleobiology*, 29, 26–29. doi:10.1666/0094-8373(2003)029<0026:GDWYRM>2.0.CO;2

Arbault, D., *et al*., 2014. A first global and spatially explicit emergy database of rivers and streams based on high-resolution GIS-maps. *Ecological Modelling*, 281, 52–64. doi:10.1016/j.ecolmodel.2014.03.004

Bedia, J., Herrera, S., and Gutiérrez, J.M., 2013. Dangers of using global bioclimatic datasets for ecological niche modeling. Limitations for future climate projections. *Global and Planetary Change*, 107, 1–12. doi:10.1016/j.gloplacha.2013.04.005

Chevan, A. and Sutherland, M., 1991. Hierarchical partitioning. *American Statistician*, 45, 90–96.

Davies, B., *et al*., 2008. Comparative biodiversity of aquatic habitats in the European agricultural landscape. *Agriculture, Ecosystems & Environment*, 125, 1–8. doi:10.1016/j.agee.2007.10.006

Dias, M.S., *et al*., 2013. Natural fragmentation in river networks as a driver of speciation for freshwater fishes. *Ecography*, 36, 683–689. doi:10.1111/j.1600-0587.2012.07724.x

Flather, C.H. and Sieg, C.H., 2007. Species rarity: definition, causes, and classification. *In*: M.G. Raphael and R. Molina, eds. *Conservation of rare or little-known species: biological, social, and economic considerations*. Washington, DC: Island Press, 40–66.

Fotheringham, A.S. and Wong, D.W.S., 1991. The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A*, 23, 1025–1044. doi:10.1068/a231025

Fox, J. and Weisberg, S., 2011. *An R companion to applied regression*. 2nd ed. Thousand Oaks, CA: Sage.

García-Roselló, E., *et al*., 2013. ModestR: a software tool for managing and analyzing species distribution map databases. *Ecography*, 36, 1202–1207. doi:10.1111/j.1600-0587.2013.00374.x

García-Roselló, E., *et al*., 2014. Using ModestR to download, import and clean species distribution records. *Methods in Ecology and Evolution*, 5, 708–713. doi:10.1111/mee3.2014.5.issue-7

García-Roselló, E., *et al*., 2015. Can we derive macroecological patterns from primary Global Biodiversity Information Facility data? *Global Ecology and Biogeography*, 24, 335–347. doi:10.1111/geb.2015.24.issue-3

Gaston, K.J. and Blackburn, T.M., 2000. *Pattern and process in macroecology*. Oxford: Blackwell Science.

Gaston, K.J. and Fuller, R.A., 2009. The sizes of species' geographic ranges. *Journal of Applied Ecology*, 46, 1–9. doi:10.1111/jpe.2009.46.issue-1

Graham, C.H., *et al*., 2004. New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution*, 19, 497–503. doi:10.1016/j.tree.2004.07.006

Griffiths, D., McGonigle, C., and Quinn, R., 2014. Climate and species richness patterns of freshwater fish in North America and Europe. *Journal of Biogeography*, 41, 452–463. doi:10.1111/jbi.12216

Guisande, C., *et al*., 2014. *RWizard software* [online]. Spain, University of Vigo. Available from: http://www.ipez.es/RWizard [Accessed 27 February 2015].

Guisande, C., *et al*., 2006. *Tratamiento de datos*. Madrid: Díaz de Santos.

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

Guisande, C., *et al*., 2010. IPez: an expert system for the taxonomic identification of fishes based on machine learning techniques. *Fisheries Research*, 102, 240–247. doi:10.1016/j.fishres.2009.12.003

Guisande, C. and Vaamonde, A., 2012. *Gráficos estadísticos y mapas con R*. Madrid: Díaz de Santos.

Hastings, D.A., Kinemena, J.J., and Clark, D.M., 1991. Development and application of global databases: considerable progress, but more collaboration needed. *International Journal of Geographical Information Systems*, 5, 137–146. doi:10.1080/02693799108927837

Hubert, N. and Renno, J.-F., 2006. Historical biogeography of South American freshwater fishes. *Journal of Biogeography*, 33, 1414–1436. doi:10.1111/jbi.2006.33.issue-8

Kunin, W.E. and Gaston, K.J., 1997. *The biology of rarity: causes and consequences of rare – common differences*. London: Chapman and Hall.

Laffan, S.W. and Crisp, M.D., 2003. Assessing endemism at multiple spatial scales, with an example from the Australian vascular flora. *Journal of Biogeography*, 30, 511–520. doi:10.1046/j.1365-2699.2003.00875.x

Lehner, B. and Döll, P., 2004. Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, 296, 1–22. doi:10.1016/j.jhydrol.2004.03.028

Lehner, B. and Grill, G., 2013. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 27, 2171–2186. doi:10.1002/hyp.9740

Lehner, B., *et al*., 2011. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, 9, 494–502. doi:10.1890/100125

Lehner, B., Verdin, K., and Jarvis, A., 2008. New global hydrography derived from space-borne elevation data. *EOS Transactions, American Geophysical Union*, 89, 93–94. doi:10.1029/2008EO100001

Leprieur, F., *et al*., 2011. Partitioning global patterns of freshwater fish beta diversity reveals contrasting signatures of past climate changes. *Ecology Letters*, 14, 325–334. doi:10.1111/ele.2011.14.issue-4

Lévêque, C., *et al*., 2008. Global diversity of fish (Pisces) in freshwater. *Hydrobiologia*, 595, 545–567. doi:10.1007/s10750-007-9034-0

Mac Nally, R., 2000. Regression and model-building in conservation biology, biogeography and ecology: the distinction between – and reconciliation of – 'predictive' and 'explanatory' models. *Biodiversity and Conservation*, 9, 655–671. doi:10.1023/A:1008985925162

Naimi, B., 2014. *Uncertainty analysis for species distribution models. R package, ver. 3.5-0* [online]. Available from: CRAN.R-project.org/package=usdm [Accessed 30 June 2014].

Oberdorff, T., Guegan, J.-F., and Hugueny, B., 1995. Global scale patterns of fish species richness in rivers. *Ecography*, 18, 345–352. doi:10.1111/eco.1995.18.issue-4

Oberdorff, T., *et al*., 2011. Global and regional patterns in riverine fish species richness: a review. *International Journal of Ecology*, 2011, 12. Article ID 967631. doi:10.1155/2011/967631.

Pelayo-Villamil, P., *et al*., 2012. ModestR: Una herramienta informática para el estudio de los ecosistemas acuáticos de Colombia. *Actualidades Biológicas*, 34, 225–239.

Pelayo-Villamil, P., *et al*., 2015. Global diversity patterns of freshwater fishes – potential victims of their own success. *Diversity and Distributions*, 22, 1703–1714.

Ricklefs, R.E., 2004. A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, 7, 1–15. doi:10.1046/j.1461-0248.2003.00554.x

Robinson, L.M., *et al*., 2011. Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Global Ecology and Biogeography*, 20, 789–802. doi:10.1111/j.1466-8238.2010.00636.x

Schleuter, D., *et al*., 2012. Geographic isolation and climate govern the functional diversity of native fish communities in European drainage basins. *Global Ecology and Biogeography*, 21, 1083–1095. doi:10.1111/geb.2012.21.issue-11

Tedesco, P.A., *et al*., 2012. Patterns and processes of global riverine fish endemism. *Global Ecology and Biogeography*, 21, 977–987. doi:10.1111/geb.2012.21.issue-10

Tedesco, P.A., *et al*., 2005. Evidence of history in explaining diversity patterns in tropical riverine fish. *Journal of Biogeography*, 32, 1899–1907. doi:10.1111/jbi.2005.32.issue-11

Thorson, J.T., Cope, J.M., and Patrick, W.S., 2014. Assessing the quality of life history information in publicly available databases. *Ecological Applications*, 24, 217–226. doi:10.1890/12-1855.1

Toussaint, A., *et al*., 2014. Historical assemblage distinctiveness and the introduction of widespread non-native species explain worldwide changes in freshwater fish taxonomic

dissimilarity. *Global Ecology and Biogeography*, 23, 574–584. doi:10.1111/geb.2014.23.issue-5

UNEP (United Nations Environment Programme), 2012. *Global Environment Outlook 5 (GEO5) – Environment for the future we want*. Nairobi: UNEP.

US Geological Survey, 2000. *HYDRO1k Documentation* [online]. Available from: http://webgis.wr.usgs.gov/globalgis/metadata_qr/metadata/hydro1k.htm [Accessed 30 June 2014].

Walsh, C. and Mac Nally, R., 2014. Hierarchical partitioning. *R package version 1.0-4* [online] Available from: http://CRAN.R-project.org/package=hier.part [Accessed 30 June 2014].

Williams, P., *et al*., 2004. Comparative biodiversity of rivers, streams, ditches and ponds in an agricultural landscape in southern England. *Biological Conservation*, 115, 329–341. doi:10.1016/S0006-3207(03)00153-8