

ModestR: a software tool for managing and analyzing species distribution map databases

Emilio García-Roselló, Cástor Guisande, Jacinto González-Dacosta, Juergen Heine, Patricia Pelayo-Villamil, Ana Manjarrés-Hernández, Antonio Vaamonde and Carlos Granado-Lorencio

E. García-Roselló (erosello@uvigo.es), J. González-Dacosta and J. Heine, Computer Science Dept, Univ. of Vigo, Campus Lagoas-Marcosende s/n, ES-36310 Vigo, Spain. – C. Guisande, Faculty of Sea Sciences, Univ. of Vigo, Campus Lagoas-Marcosende s/n, ES-36310 Vigo, Spain. – P. Pelayo-Villamil, Ichthyology group, Univ. of Antioquia, A.A. 1226, Medellín, Colombia. – A. Manjarrés-Hernández, Amazonia Research Inst., Univ. National of Colombia, Km 2 vía Tarapacá, Leticia, Colombia. – A. Vaamonde, Faculty of Economy, Univ. of Vigo, Torrecedeira 105, ES-36208 Vigo, Spain. – C. Granado-Lorencio, Faculty of Biology, Univ. of Sevilla, Avenida Reina Mercedes s/n, ES-41012 Sevilla, Spain.

The ModestR package consists of three applications: MapMaker, DataManager and MRFinder. MapMaker facilitates making range maps by drawing the areas, by importing existing data or using the Global Biodiversity Information Facility portal. It can discriminate between different habitats, thereby making data cleaning tasks easier. DataManager allows the management of taxonomically structured databases for range maps. MRFinder supports querying ModestR databases to find the species present in specific areas. Possible applications include the compilation and management of species distribution databases, cleaning data and computing aggregated data to perform subsequent analyses in other packages thanks to emphasized interoperability.

ModestR package has been developed with the primary aim of providing the scientific community with an easy-to-use but powerful tool for managing species distribution data. It is designed to be simple and intuitive even for users not familiar with general-purpose Geographical Information Systems tools (GIS) that are broadly used for species distribution mapping. ModestR supports databases structured on hierarchical taxonomy. It provides features to easily clean, manage, analyse and summarise large species distribution datasets. It offers high interoperability with the other software tools widely used for pre-post processing data related to distributions, such as spreadsheets, GIS software, and SDM software. Particular attention, however, has been paid to facilitate data exchange with R statistical environment (R Development Core Team) and subsequently with packages widely used in distribution analysis, such as SDMtools (VanDerWal et al. 2012) or dismo (Hijmans et al. 2012). Moreover, a new interface is being developed, which is specifically designed to link the output of ModestR with R environment. Data exchange, however, is also possible with other stand-alone applications such as SAM spatial analysis software (Rangel et al. 2010) and SDM software such as Maxent (Phillips et al. 2006).

ModestR software design

The ModestR package consists of three applications: MapMaker, DataManager and MRFinder. MapMaker

allows users to easily draw range maps by means of an intuitive interface. A map made with MapMaker will be stored in a ModestR database linked to taxonomic data. This database can be created and managed with DataManager, which also allows maps to be processed, and information to be aggregated and exported for subsequent analysis using other packages. MRFinder allows querying a ModestR database to retrieve the species that are present in a specific area, and subsequently calculate and export aggregated data from those species.

To be useful to the broadest range of users, it was decided to prioritise the ease of use and the innovative and useful features in ModestR, together with the data interoperability with existing tools that already provide many other capabilities that different users may need. The following sections provide further details about each of the applications in ModestR.

MapMaker description

The MapMaker application provides a simple but powerful user interface for building precise individual species distribution maps. ModestR incorporates habitats, a concept that is specific for species distribution maps. Five habitats are currently supported in MapMaker: sea, land, small rivers, large rivers and lentic habitats.

There are several ways of using MapMaker to create a distribution map. First, by importing a CSV file with

occurrence coordinates, which can be useful when analysing field samples.

Second, by selecting a species (synonyms of the species can also be included) and importing data from the Global Biodiversity Information Facility (GBIF 2010) online portal. The data is cleaned as all duplicates are removed and those locations that are not in the habitats selected by the user are considered invalid, with a current error of ± 0.1 minutes. As an example, the upper map of Fig. 1 shows the geographical

records for the white shark *Carcharodon carcharias* in MapMaker, obtained from the GBIF database. MapMaker automatically classifies valid (red) and invalid samples (black) in Fig. 1 depending on whether coordinates are within or outside the habitat of the species (this shark is a marine species). Data cleaning can be reinforced by applying user-defined rules involving environmental variables. This requires providing MapMaker environmental data files in CSV or ESRI ASC format. The user can always manually reclassify samples.

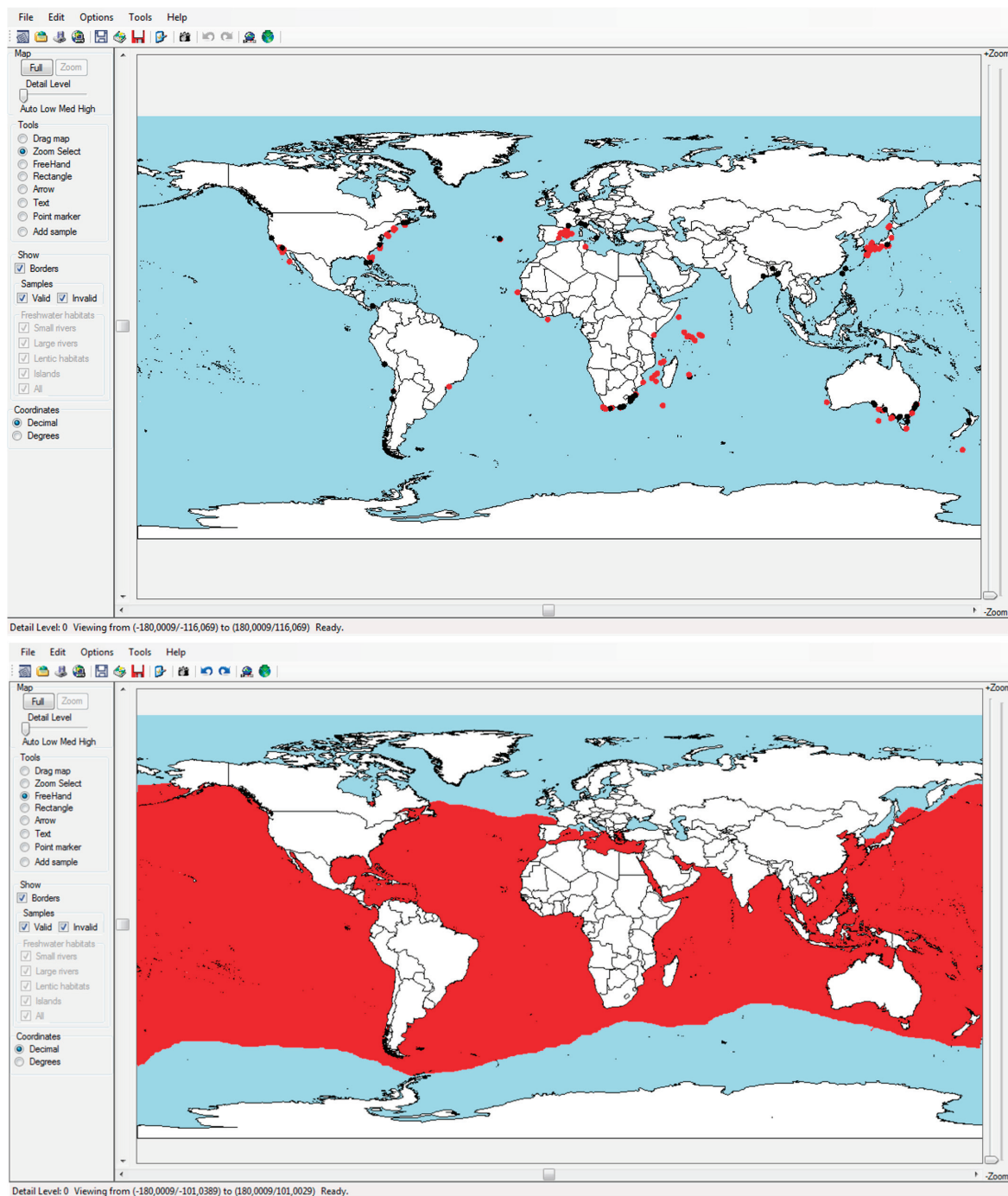


Figure 1. Range maps for the white shark *Carcharodon carcharias* made with MapMaker. The upper map was made with data obtained from the online GBIF database; red and black points correspond to samples classified as valid and invalid by MapMaker, respectively. The lower map was made using the freehand tool, with information about the hypothetical distribution of this species (Compagno et al. 2005).

Third, MapMaker also supports expert-drawn maps that usually consist of several close areas that specify species presence. The user can select an area with the freehand tool and choose the type of habitats within this area occupied by the species. MapMaker will then fill in the corresponding portion of the map. For example, the lower map of Fig. 1 shows the distribution of *Carcharodon carcharias* made using the freehand tool with information about the potential distribution of this species, according to Compagno et al. (2005). This feature makes drawing distribution maps very intuitive, as the user can immediately see the real area selected.

Four, MapMaker can also import range maps from files in KML or shapefile formats commonly used by GIS tools.

Finally, MapMaker can create maps by importing ESRI ASCII grid probability files generated by distribution modelling software such as Maxent (Phillips et al. 2006) and dismo (Hijmans et al. 2012). In this case, a probability threshold has to be entered, and MapMaker will build a range map with all the areas where the probability of species presence is higher than this threshold.

Maps drawn with MapMaker can be saved in standalone files or in a ModestR database by assigning the map to a particular species. They can also be exported to several formats, such as high-resolution BMP or JPEG, CSV, KML, or shapefile (a free Geospatial Data Abstraction Library package is used for the latter option).

DataManager description

DataManager provides integrated management of taxonomic information and distribution maps in ModestR databases. Although there is an option for manually introducing

the taxa, a ModestR database will usually be populated by importing existing taxonomic data from files in CSV or in phyloXML formats. Once the taxonomic data is introduced, distribution maps for any species can be stored in the database, either by saving the map directly to the database from MapMaker, or by importing data maps from DataManager. Map import/export features in DataManager are the same as in MapMaker, with the difference being that bulk tasks can be done straightforwardly for all species from any branch of the taxonomy. For example, a single operation in DataManager can download data for all species of an order from GBIF, or import samples for several species from a CSV file, or export all species of a taxonomic group to JPEG files, or import several ESRI probability files generated by an SDM software and building and storing the corresponding maps in the database. This feature allows users to easily complete tasks that involve the import, processing and export of large amounts of data (i.e. Fig. 2).

The database contents are presented in DataManager as a hierarchical taxonomy tree with classes at the highest level, and species at the lowest level. An icon signals species that have an associated map. Several filtering and searching options are available to make it easier to find taxa and maps (Fig. 3, upper panel). Moreover, any modifications in the taxonomy can be directly made onto the tree.

This taxonomic tree facilitates the understanding and management of the data structure, as well as the taxonomic relationships among the species. It also allows users to explore the species distribution patterns at higher taxonomic ranks than species, which is a valuable feature in species distribution analysis (Gruenwald and Zhang 2008).

DataManager also performs the rasterization of distribution maps, currently to a 1×1 minute cell, which can be

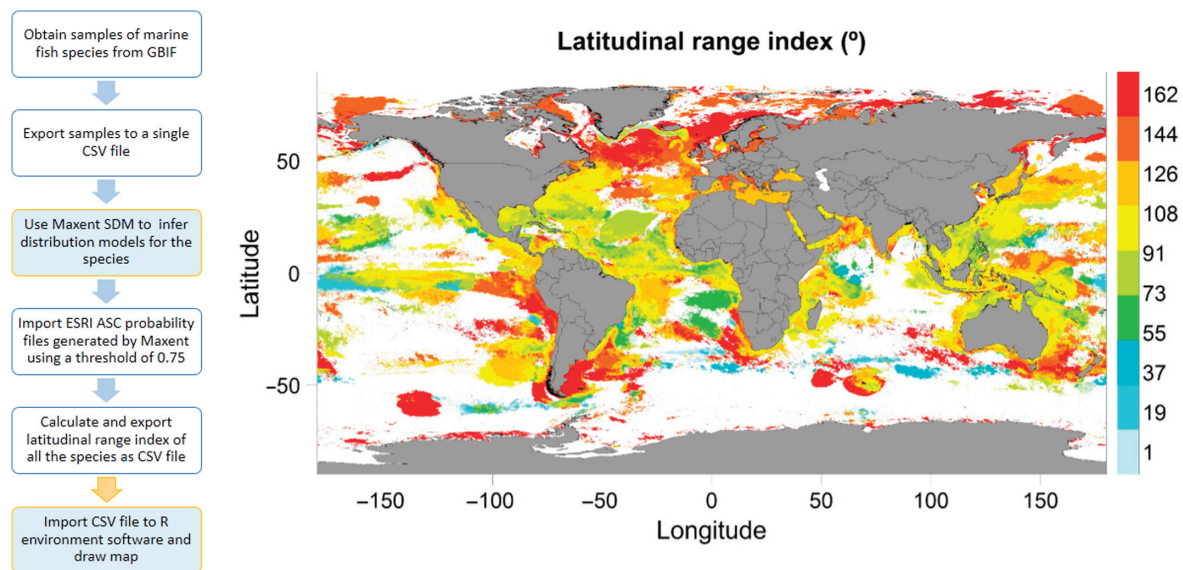


Figure 2. Steps used to obtain a latitudinal range index in cells of $5' \times 5'$ of fishes belong to orders with only marine species, using maps created by importing the predictions of the distribution models for all the species inferred with Maxent ver. 3.3.3, based on occurrence records previously obtained from the GBIF by DataManager, and using the Bio-ORACLE global dataset as environmental factors (Tyberghein et al. 2012). A record was considered valid when the probability of the distribution models obtained was higher than 0.75 (this setting can be modified by the user). Results are drawn with R environment software. Each step described on the left flowchart required only one single operation in DataManager (provided that required ModestR taxonomy database previously existed). Steps with shaded background indicate operations performed with other applications and importing/exporting data to/from DataManager.

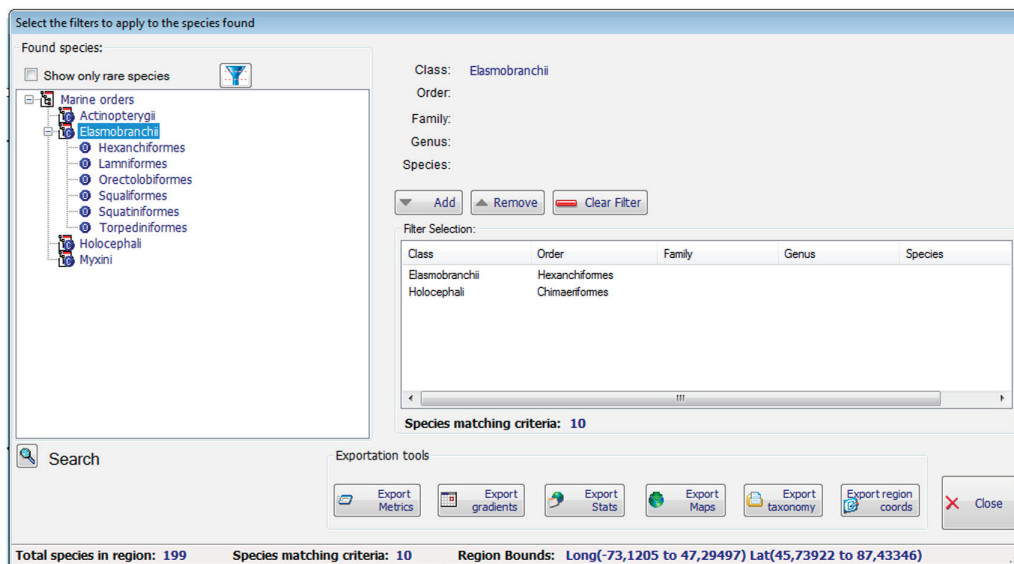
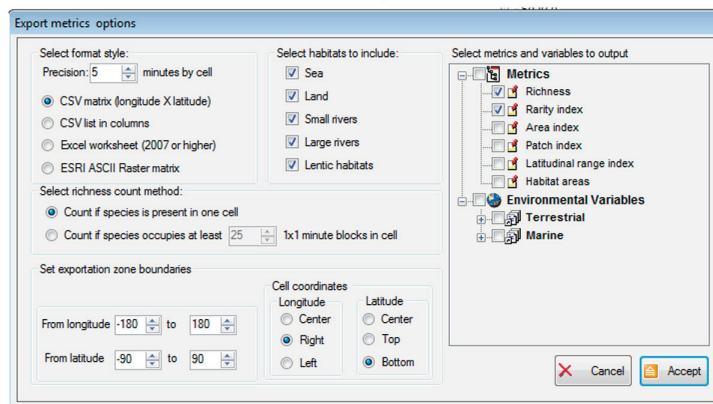
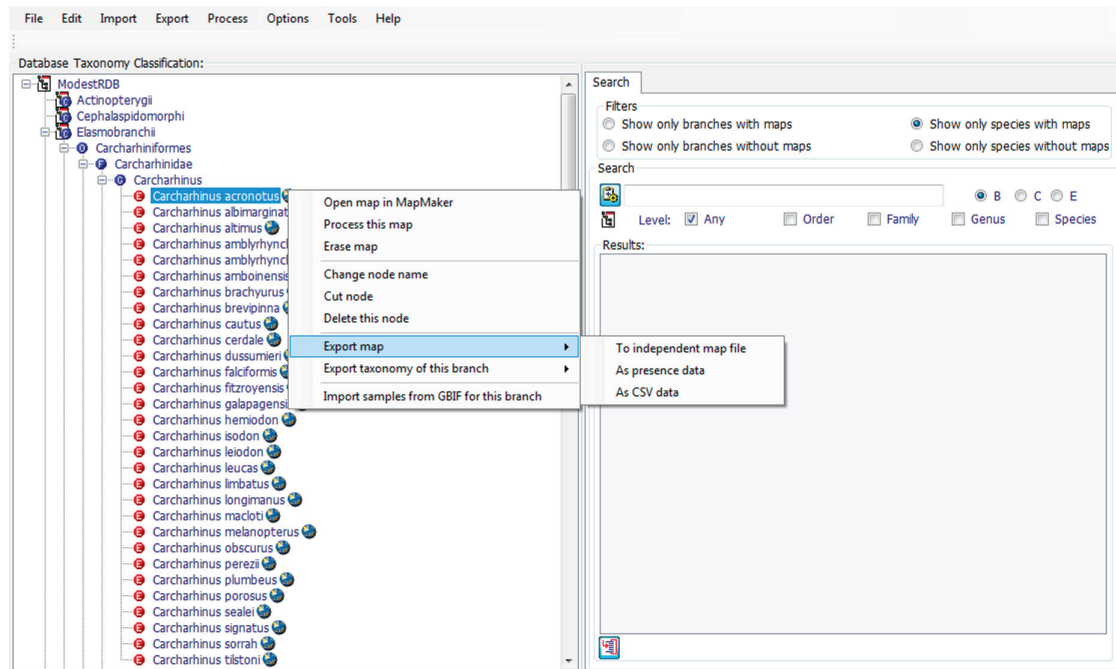


Figure 3. Top: the menu of DataManager showing the taxonomy of IPez (Guisande et al. 2010). The left tree displays taxonomic data. A world icon beside the species indicates that there is a map assigned to the species. Several filtering and searching options are available in the right panel. Middle: the DataManager dialog box used to configure exportation settings; the left tree allows selecting which metrics will be calculated as well as environmental variables. Bottom: the MRFinder dialog box where species found in an area are shown. Filters can be added to select specific taxonomic groups from the species found and several export tools are available.

considered an appropriate resolution for a broad range of usages. The resulting matrix is stored in the database to be used to perform subsequent calculations. As rasterization is a laborious task that can take a significant amount of time, particularly for expert-drawn maps (far less for sample-based maps), the user can decide when this task can be carried out, thus avoiding hindering his/her work.

Once maps have been rasterized, all the data exporting options of DataManager are fully available. It is possible to export aggregated data from any branch of the taxonomy, from the class to the species level, which is a feature that is not usually available in software not specifically designed for species distribution mapping purposes (Gruenwald and Zhang 2008).

In order to generate metrics from one species or from any branch of the taxonomy, output cell resolution can be selected from 1×1 to 60×60 minutes. A tree shows the available metrics that DataManager can calculate, which include several biogeographical indices and the area of each type of habitat for each considered cell (Fig. 3, middle panel). DataManager also allows configuring environmental variables to be included in the output. This data can be easily imported from files in CSV or ESRI ASC format as for example a Bi-Oracle dataset (Tyberghein et al. 2012). The purpose of this feature is to integrate exportation of aggregated data for species distributions and environmental variables. Output data can be exported to CSV format, arranged as a matrix or as a list of values, to a raster in ESRI ASCII format, or to a Excel-xlm file (Fig. 3, middle panel). These files can also be easily imported to spreadsheets, the R environment software and GIS tools.

Latitudinal gradients for the supported metrics can also be calculated for bands of any height from 1° to 10° of latitude. These are particularly useful to analyse distribution patterns.

Samples from any taxonomic group can be exported to a single CSV file, making it easy to use them in other packages, including spreadsheets, statistical or SDM software. When existing maps are expert-drawn range maps, pseudo-samples may also be generated. Finally, samples from maps can also be exported to a CSV format designed to be easily combined with available GBIF tools to generate Darwin Core archives.

MRFinder description

MRFinder is the final application in ModestR package. Its user interface is very similar to that of MapMaker (Fig. 1), but MRFinder allows the user to select an arbitrary zone in the world map, whether by drawing it freehand for an irregular zone, or by entering precise coordinates for a rectangular zone. MRFinder performs a search for species in the selected area, showing the results of the query in a taxonomic tree where filters can be applied to select specific taxonomic groups (Fig. 3, lower panel). It is also possible to select only rare species from those found in an area. The criteria for considering a species as rare may vary among studies, but they are usually associated to small range size, and this was the criteria used in ModestR. A species is considered to be rare when it is only found in

a particular region and nowhere else, and its geographical distribution is restricted to an area smaller than a size defined by the user.

As well as the same data aggregation and exportation features already available in DataManager (exporting metrics, latitudinal gradients, taxonomy lists, samples, etc.), statistical data such as the area of occupancy, maximum, mean and minimum latitude and longitude, and the reference area/occupied area ratio, can also be generated. These results are calculated and summarized for each taxonomic level of the species, from species up to classes, using the selected area boundaries as reference (even in the case of an irregular area). In this way MRFinder allows different analyses to DataManager: whereas in DataManager the analyses can only be performed by aggregating the species of a taxonomic group, in MRFinder they can include species from different groups that are present in the same area, allowing for more complex analyses.

Package installation

The ModestR software package is free for academic non-commercial use. It can be downloaded from ModestR website at <www.ipez.es/ModestR> or from persistent URL <www.purl.org/ModestR>. It requires Microsoft Windows XP or later and the .Net framework 4.0. Additional materials, including a tutorial, an R package with several functions to make maps with data exported from ModestR, and sample databases, are available at the ModestR website. Geographic data about country borders were obtained from the web page <www.diva-gis.org> (Hijmans et al. 2001), while geographic data for the freshwater habitats were obtained from <www.openstreetmap.org/>.

To cite ModestR or acknowledge its use, cite this Software note as follows, substituting the version of the application that you used for 'version 0':

García-Roselló, E., Guisande, C., González-Dacosta, J., Heine, J., Pelayo-Villamil, P., Manjarrés-Hernández, A., Vaamonde, A. and Granado-Lorencio, C. 2013. ModestR: a software tool for managing and analyzing species distribution map databases. – *Ecography* 36: 1202–1207 (ver. 0).

References

- Compagno, L. et al. 2005. *Sharks of the world*. – Princeton Univ. Press.
- GBIF 2010. GBIF data portal. – <<http://data.gbif.org>>.
- Gruenwald, L. and Zhang, J. 2008. Embedding and extending GIS for exploratory analysis of large-scale species distribution data. – In: *Proceedings of the 16th ACM SIGSPATIAL conference on advances in geographic information systems*. New York, USA, article no. 28.
- Guisande, C. et al. 2010. IPEZ: an expert system for the taxonomic identification of fishes based on machine learning techniques. – *Fish. Res.* 102: 240–247.
- Hijmans, R. et al. 2001. Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. – *Plant Genet. Res. Newslett.* 27: 15–19.

- Hijmans, R. J. et al. 2012. dismo: species distribution modeling. – R package ver. 0.7-23, <CRAN.R-project.org/package=dismo>.
- Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. – *Ecol. Model.* 190: 231–259.
- Rangel, T. F. L. V. B. et al. 2010. SAM: a comprehensive application for spatial analysis in macroecology. – *Ecography* 33: 46–50.
- Tyberghein, L. et al. 2012. Bio-ORACLE: a global environmental dataset for marine species distribution modelling. – *Global Ecol. Biogeogr.* 21: 272–281.
- VanDerWal, J. et al. 2012. SDMTools: tools for processing data associated with species distribution modelling exercises. – R package, ver. 1.1-13, <<http://CRAN.R-project.org/package=SDMTools>>.