# ModestR Software

# User's Manual

**v.5.5**






WebSite: http://www.ipez.es/ModestR

If you use ModestR software and you publish any results or images obtained by means of ModestR software you can cite it in your work as:

> *E.García-Rosello et al. (2013) ModestR: a software tool for managing and analyzing species distribution map databases. Ecography 36(11):1202–1207, online in http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0587.2013.00374.x/abstract . D.O.I: 10.1111/j.1600-0587.2013.00374.x*

# Contents

**LIMITATION OF LIABILITY**
In no event shall This document be liable for any damages (including, without limitation, lost profits, business interruption, or lost information) rising out of 'Authorized Users' use of or inability to use This document, even if advised of the possibility of such damages. In no event will This document be liable for inaccurate results, software errors, loss of data or for indirect, special, incidental, consequential (including lost profit), or other damages based in contract, tort or otherwise. This document shall have no liability with respect to the content of this document or any part thereof, including but not limited to errors or omissions contained therein, libel, infringements of rights of publicity, privacy, trademark rights, business interruption, personal injury, loss of privacy, moral rights or the disclosure of confidential information.

# 1 Introduction

ModestR software is a software package to help academic community managing species distribution information. It is currently composed of three applications:

- MapMaker: aimed to build distribution maps in a simple way, either by selected the zones where the species is found (range maps), or by adding occurrences or occurrences in the map from CSV files o GBIF database. It also provides several tools to analyse and transform a map (such as data cleaning, hulling, etc.)
- DataManager allows creating taxonomic databases where maps should be stored associated to a species. It also performs calculations and exports data like presence or richness matrices, that can be processed with other software tools. Most of the features available in MapMaker to analyse and transform maps are also available in DataManager, with the crucial difference that while in MapMaker they will be applied to one single map, in DataManager they could be applied to a whole collection of maps launching a sort of batch task.
- MRFinder allows to find in the distribution maps stored in a ModestR database which species are present in an area of the world, simply by selecting this area in the map. Then it can generate several statistical and summary data output.
- MRMapping is a tool to easily make maps grouping several species data by any rank, and presenting several distributions in the same map. This makes easy to represent de distribution of any group of species, of genus, a family, etc… or to show where two or more species (or groups of species) overlap. MRMapping can also calculate overlapping areas.

In the next sections we introduce those tools. We also describe the different importation and exportation options and formats supported by ModestR applications in the sections 6, 7 and 8 respectively.

## 1.1 ModestR assumptions

Before all, it is interesting to point out some assumptions and facts about ModestR that can be important to correctly understand how it works, its limitations, and to properly interpret results obtained with ModestR:

- **ModestR map projection and datum:** ModestR maps use a simple equirectangular projection also called the equidistant cylindrical projection or geographic projection. The projection maps meridians to vertical straight lines of constant spacing (for meridional intervals of constant spacing), and circles of latitude to horizontal straight lines of constant spacing (for constant intervals of parallels). The projection is neither equal area nor conformal, but is probably the easier to visualize and understand, particularly for users not experienced in GIS (more info: https://desktop.arcgis.com/en/arcmap/10.3/guide-books/map-projections/equirectangular.htm and https://en.m.wikipedia.org/wiki/Equirectangular_projection).
  The used datum is the WGS 84 (World Geodetic System 1984), using simple latitude and longitude coordinates in degrees, minutes and seconds.
- **ModestR spatial accuracy:** ModestR stores geographical data, such as occurrence coordinates or shape vertexes, using longitude and latitude coordinates with at least 5

decimals. This is equivalent to a precision of less than ten meter, which is enough for most usages. This should not be confused with the precision of some outputs generated by ModestR, which can be significantly lower. But spatial data such as species distributions are stored with high precision, and can be retrieved with such this high precision if needed.

- **ModestR type of maps**: ModestR supports three types of presence maps:
  - Occurrence maps or maps based on occurrence records. Each occurrence is basically a pair of longitude and latitude coordinates that identify a location where a species is present.
  - Range or area maps: range or area maps are maps where species presence is demarcated by polygons or areas. Being more complex, range maps require more computing time to be processed by ModestR.
  - Mixt maps: ModestR supports maps that mix occurrences and areas, even if we don't encourage this option, as this makes more complex to process maps and more difficult to correctly interpreting results. Anyway, ModestR will generally handle mixt maps the same way than area maps.

- **Areas, metrics and presence resolution**: ModestR calculations including areas, presence detection, metrics (such as richness), etc, are done with a resolution of ±1' (1' stands for 1 arcminute). If you use environmental variables into ModestR, usually as raster/grid data (for example in ESRI ASC format), maximal supported resolution is of 1', but coarser resolutions are supported by ModestR and recommended, particularly when working with worldwide rasters, as they allow considerably faster calculations (most of those data come in resolutions of 5'x5').
  Occurrences habitat checking is currently done with a resolution of aprox. 1" (1 arcsecond).
  For occurrence (occurrence) maps, the area corresponding to an occurrence cannot be accurately evaluated. Therefore, in terms of species presence area, ModestR makes a simple assumption, considering that the area of occurrence equals the area of the cell of 1'x1' (1' stands for 1 arcminute) where it is located. But that this is just a convenient assumption and not an accurate value. Other areas calculated from occurrence maps, such as Extent of Occurrence area using convex hull or alpha shape, will be accurately calculated, as they are polygons but not just points.
  In the case of range maps, areas can be accurately calculated, after dividing them in small cells with a precision of 1'x1'.
  As stated before, this precision of 1'x1' for many ModestR outputs does not affect the precision of the input spatial data (such as occurrences or shapes), which are stored with a very high precision.

- **Species habitats**: ModestR supports assigning several different habitats to a species. But when processing a map it assumes that a species cannot have different excluding habitats in different locations. That is, if a species has a validated presence area on a marine habitat and another on land habitat, for example, ModestR will assume that both habitats are valid for this species. This assumption is done in several calculations and map transformations, like EOO calculation, niche estimations or hull transformations, for example.

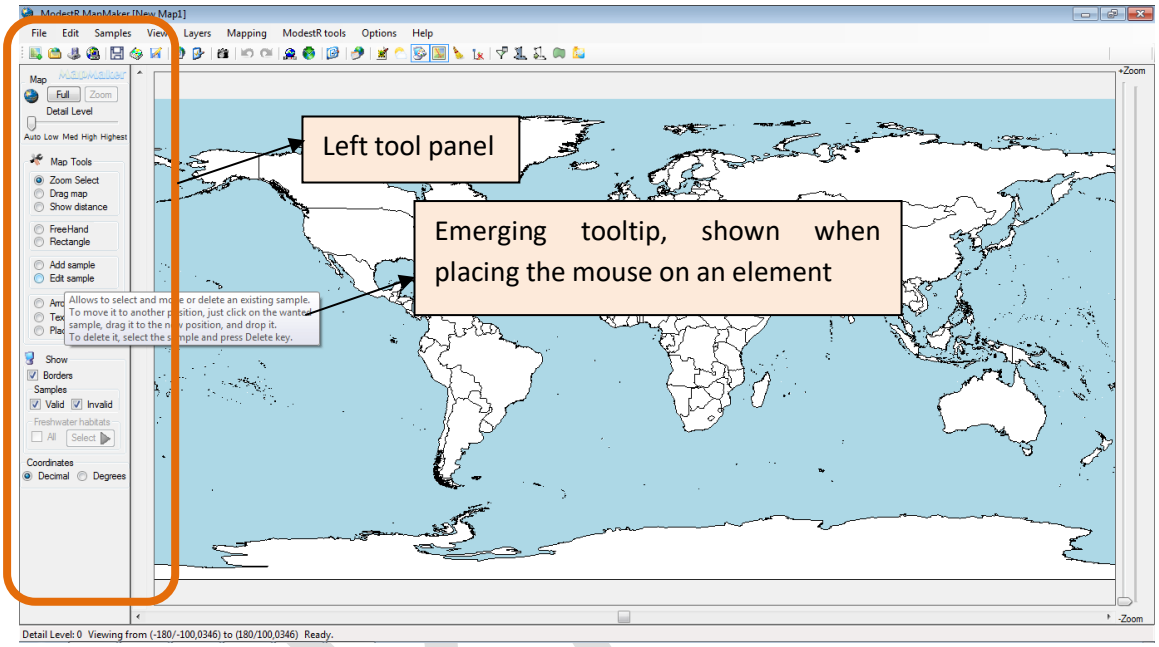
- **Taxonomy restrictions**: ModestR supports taxonomy databases that have to include from class to species levels. It assumes that there cannot have two taxons of the same level with the same name. For example you can't have an order named X in a class A, and another order also named X in another class B. In this case ModestR will only keep the first order inserted with the name X.

- **Default settings:** ModestR provides many features to work with distribution data. Some of them can be somewhat complex, and there are usually several parameters to adjust some data processing features. Most of the times, ModestR will provide default settings that can be valid for the common cases. In this way, ModestR tries to be simple to use, particularly for new users. But once familiar with those settings, ModestR allows you to modify and parametrize many features.

- **Task cancelation and backups**: to offer a better interactive user experience, ModestR has been designed to allow users canceling running tasks whenever it is possible. This feature is particularly useful in applications like DataManager, that is aimed to process large dataset. But take into account that even if ModestR allows you canceling a task at any moment, the modifications made on data until this moment will usually not be undone! That is why ModestR offers *Undo* features in most interactive applications like MapMaker or MRMapping, and data backup warnings in applications like DataManager. Particularly in DataManager it is probably recommendable to use integrated backup feature (or just make your own copy of the database) before launching tasks that will modify data. This can prevent from losing original data in case of failure or if you decide to cancel the task before its completion, but when some data have been already modified.

# 2 First steps with MapMaker

Mapmaker is an easy-to-use software application to build distribution maps. The MapMaker user interface is quite simple, as the most usual tools are always visible at the left tool panel. You also have contextual menus that will appear when you click with the right button of the mouse on the map. In addition, quite all menus, buttons and other elements have a tooltip that briefly explains their usage, and that will be shown when you place the mouse over an element.




## 2.1 Opening MapMaker and moving across the world map

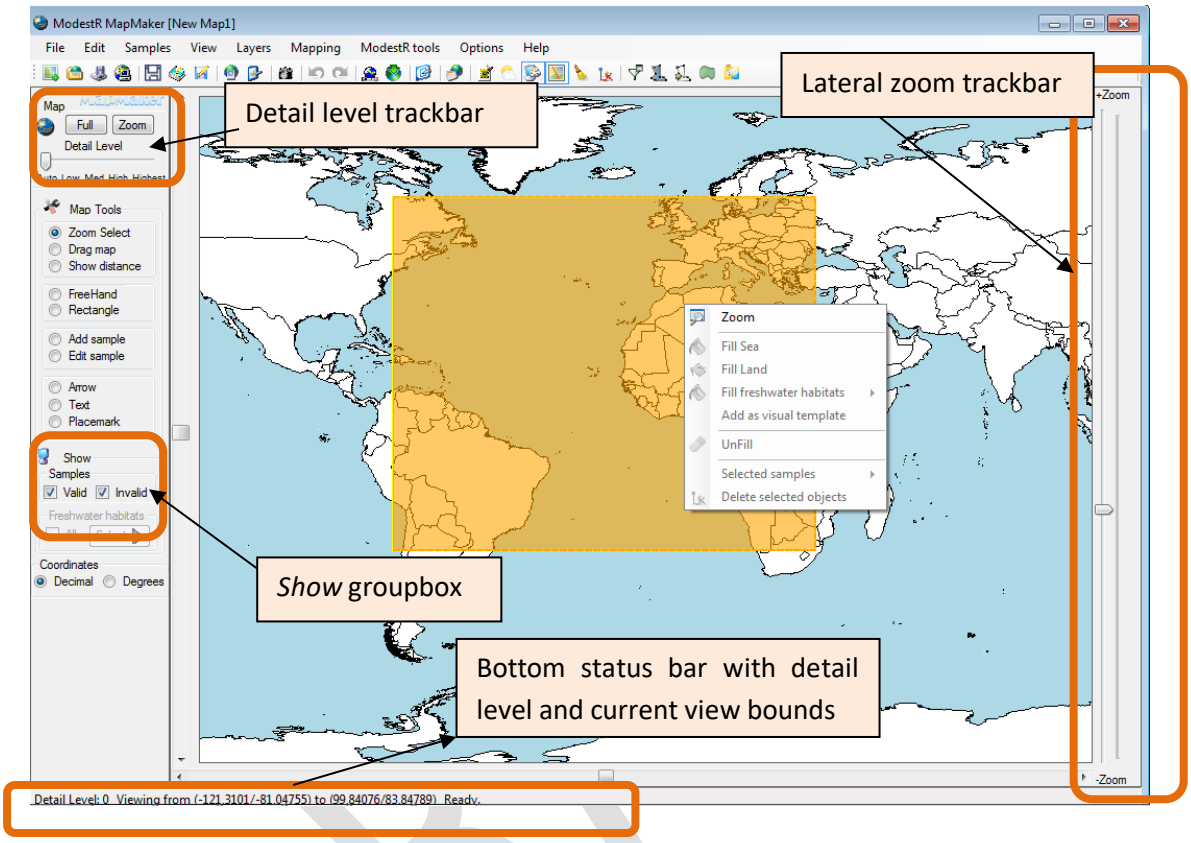
When you open MapMaker, it will automatically show the world map in a full view with a lower detail level[1]. You can zoom in any part of the map selecting the *Zoom Select* tool in the left tool panel, and then selecting the zone in the world map. When the wanted zone is selected, press the *Zoom* button or select the *Zoom* option on the contextual menu. You can also use the mouse wheel to zoom in and out (click first on the map to set the focus on it), but this option can perform very slowly, depending on your computer capabilities. Finally, you can use the lateral Zoom trackbar to modify the zoom level.

MapMaker have several detail levels, from 0 (lower) to 4 (highest). It will automatically changes to a higher or a lower detail level regarding the deep of the zoom. Anyway, you can force a detail level using the *Detail Level* trackbar at the top of the left tools panel. Take into account that the higher the level, the slower will be displaying the map and moving across it. MapMaker will display the current level of detail in the bottom status bar. You can select the elements that will be shown on the map in the *Show* groupbox left tool panel, such as borders or freshwater habitats. Freshwater habitats are only visible in the higher level[2], so the checkboxes to make

[1] World map data comes from from OpenStreetMap with ©OpenStreetMap contributors, http://www.openstreetmap.org/copyright

[2] From ModestR v3.0, freshwater data is provided separately from ModestR package, which only contains a world map with land data. If you need freshwater data, you can download it from Modestr website Once

them visible will only be enabled when the map is displayed using this detail level. You can directly come back to the main world view using the *Full* button on the top of left tool panel.

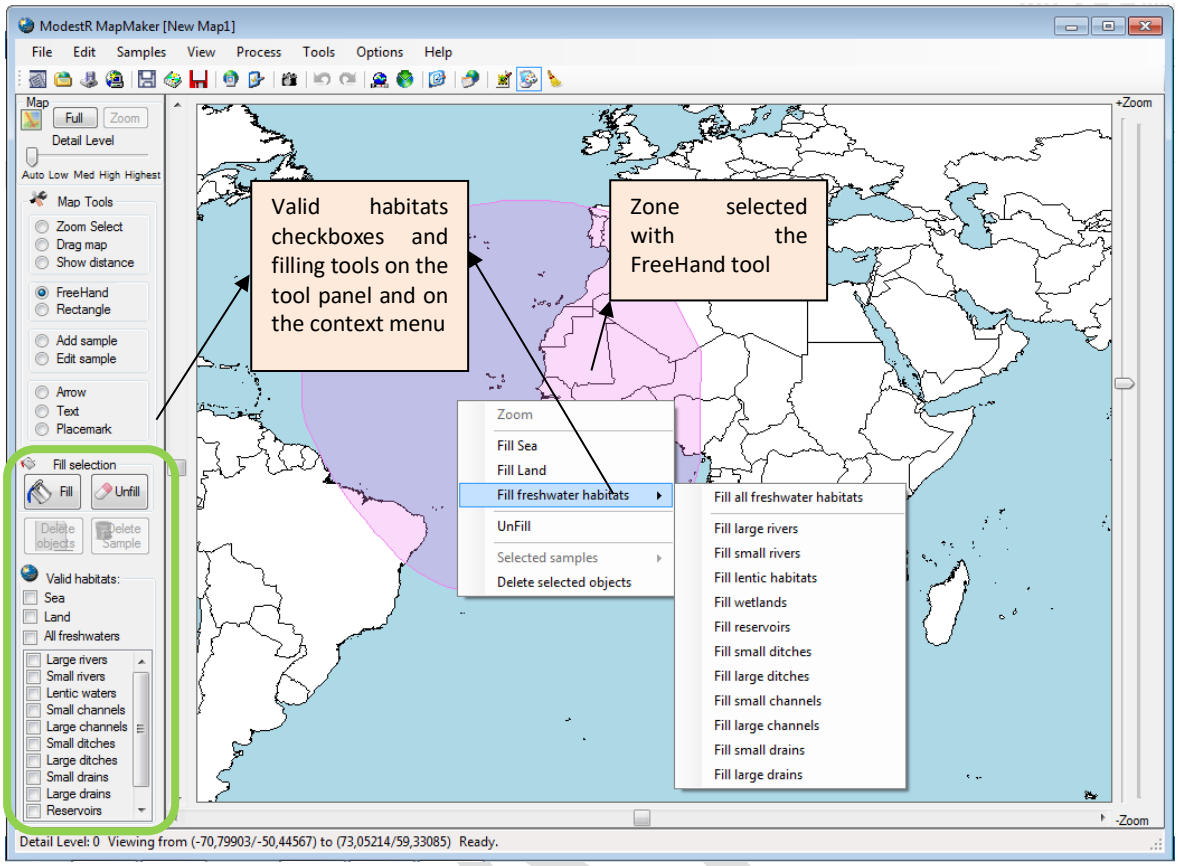



To drag the map, you can use the *Drag map* tool. When dragging, an arrow will appear, indicating where the location situated at the beginning of the arrow will appear when you end the dragging operation.

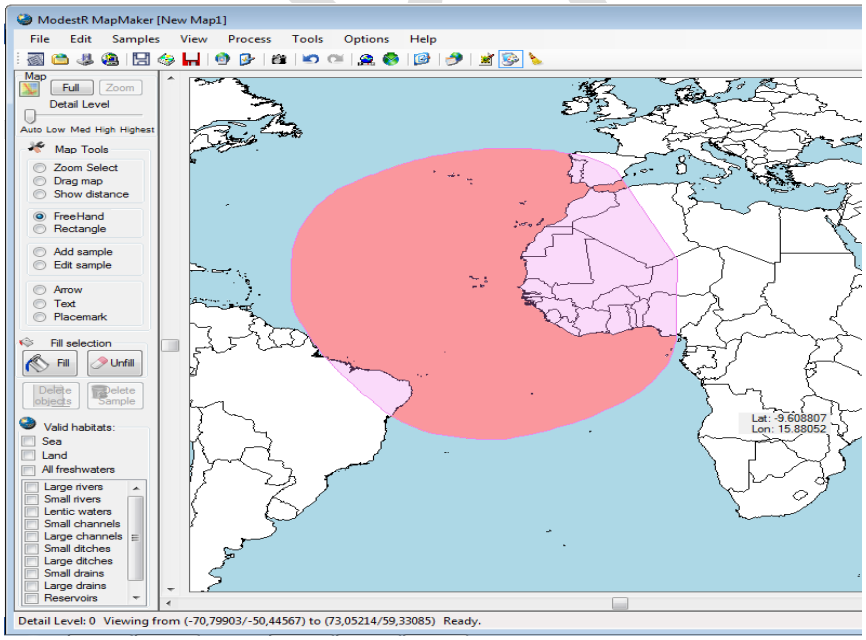
## 2.2 Making a range distribution map in MapMaker

You can make two main kinds of distribution maps in MapMaker: range maps, and occurrence (also called occurrence) maps. To make a range map, you simply have to draw on the map a zone or area occupied by a species, typically using the *Freehand* tool. Once a zone selected on the map, the *Fill Selection* and the *Valid habitats* groupboxes will be displayed at the left tool panel. Then select which habitat of this zone is occupied by a species, using the checkboxes of the *Valid habitats* groupbox. Finally, use the Fill button on the Fill *Selection* groupbox, or alternatively use the wanted option of the contextual menu that you can display clicking with the right button on the map.

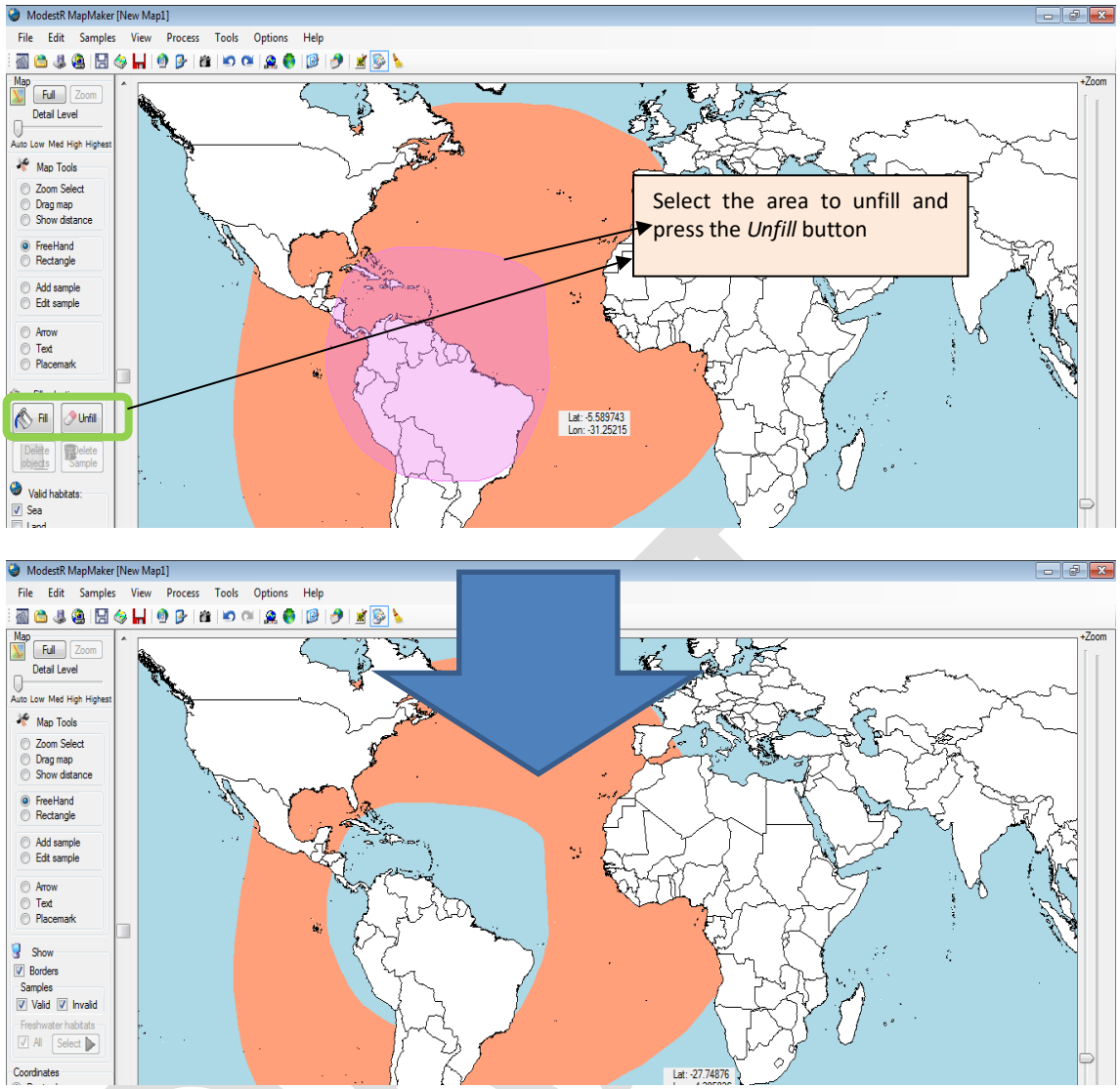
installed, you will be able to show those data in ModestR. Moreover, ModestR will be able to discrimate between freshwater habitats when creating maps, cleaning data or calculating areas, for example.




You can see that only the corresponding parts of this zone will be filled in, to indicate the presence of the species. Filling colors as well as map element colors can be modified in the *Options/Preferences* menu.




You can unfill part of a selection just selecting a zone and using *Unfill* button.



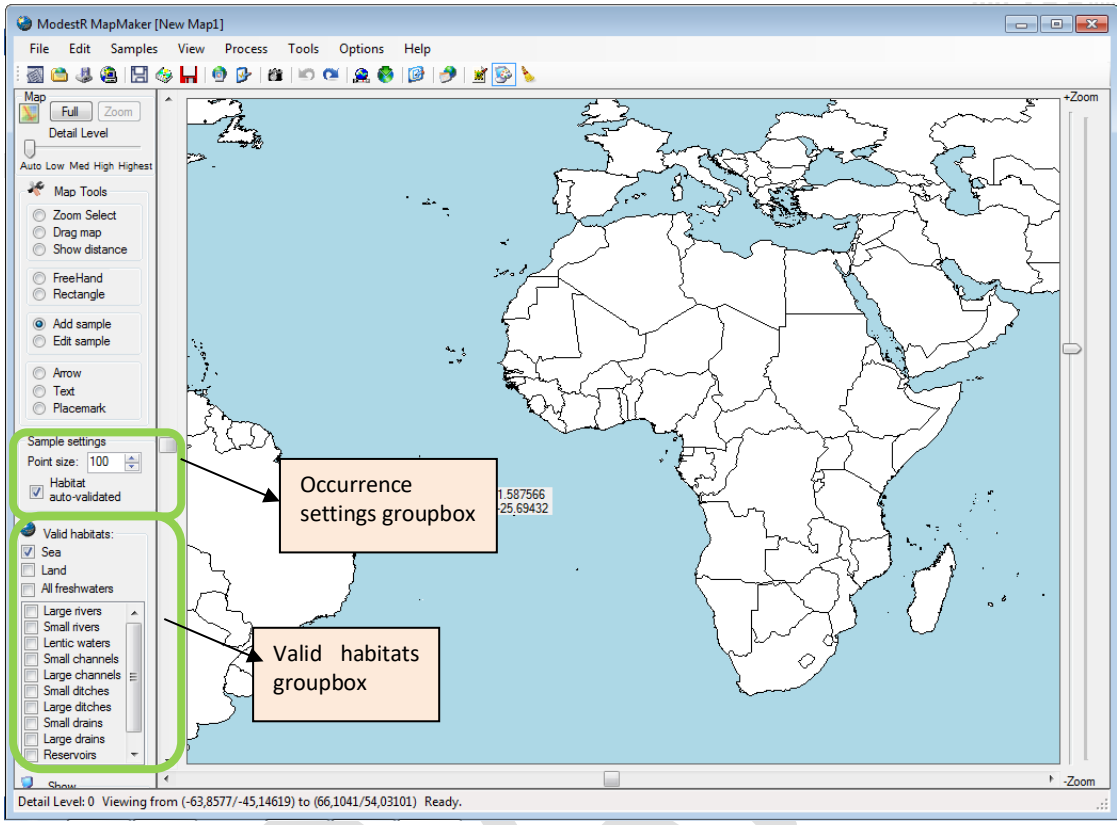

If you want to fully undo an operation, you would rather use the *Undo* button of the toolbar, or the *Edit/Undo* option in the main menu, or the Ctrl+Z key combination. Operations that can be undone in MapMaker are those that affect the map distribution data, such as adding or removing occurrences, presence areas, etc. Operations that only visualize data such as displaying environmental data or shapes (explained later in this chapter) cannot be undone, because they do not really imply a modification of the map, just a change in display options.

## 2.3 Making an occurrence-based distribution map in MapMaker

To make a distribution map basing on occurrences or occurrences of a species, you should simply have to add the occurrences on the map, using one of three options: manually adding occurrences, importing them from a CSV file, or downloading them from online GBIF portal. Each one of those options is explained below:

1. Manually: selecting the *Add occurrence* tool in the left tool panel. Then select the type of occurrence in the *Occurrence settings* groupbox (this is explained later in the section), and the valid habitats in the *Valid Habitats* groupbox that will be displayed in the left tool panel. And finally click on the position in the map where the occurrence has to be added. You can drag the occurrence to the exact target location while you have the left

mouse button pressed. Once you release the button, the occurrence will be placed on the map. Obviously this option is only useful when having to add few occurrences.



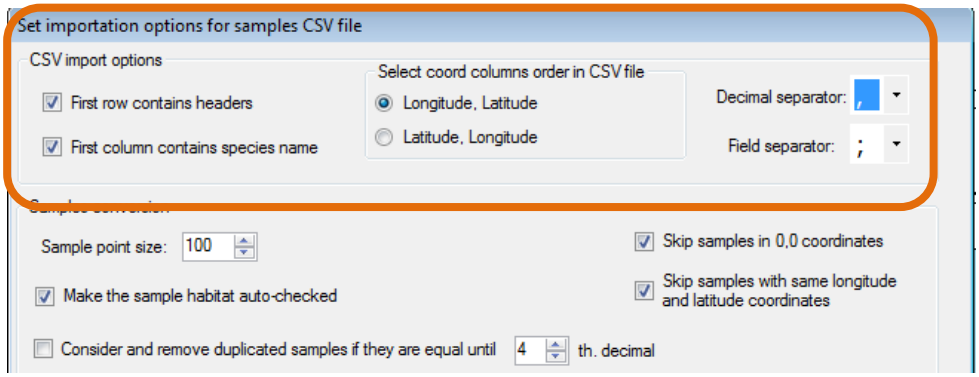

2. Another option is to import occurrences from a CSV file. On the main menu, select *File/Import/Occurrence data from CSV*. You will have to select the file where the CSV data is located. Expected format is a list of occurrences in the form:

   [Species]; Longitude ;Latitude (where *[Species] field is optional)*

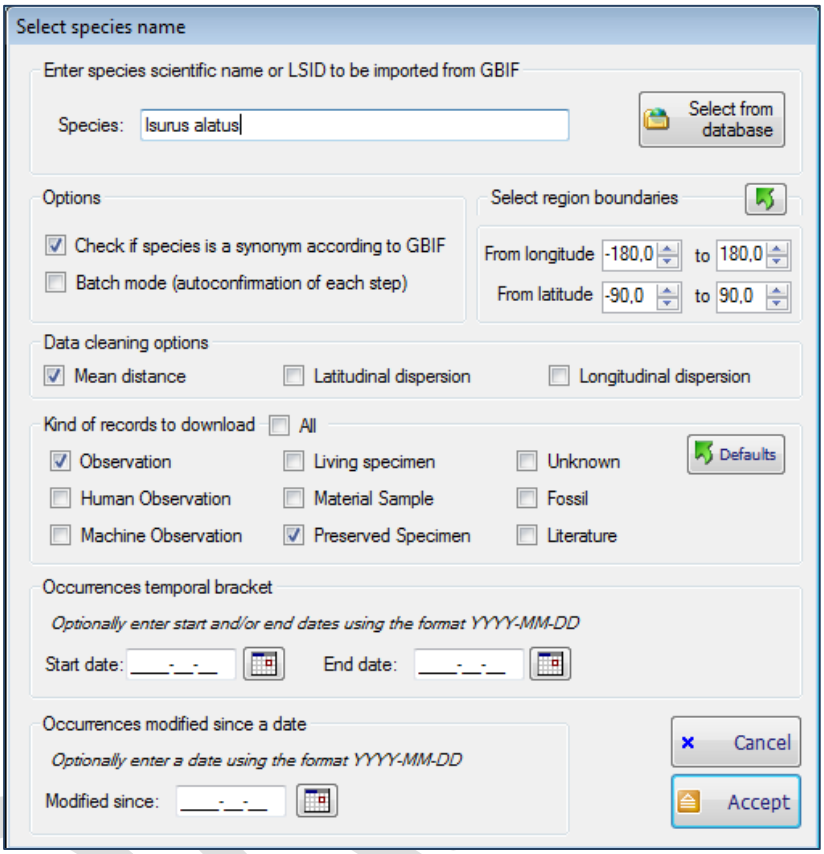
   or

   [Species] ;Latitude; Longitude (where *[Species] field is optional)*

   Data can optionally contain a headers row. Details of the expected format are explained in 7.1. Anyway, you can select the right options regarding the format of the source file in the importation dialog box.

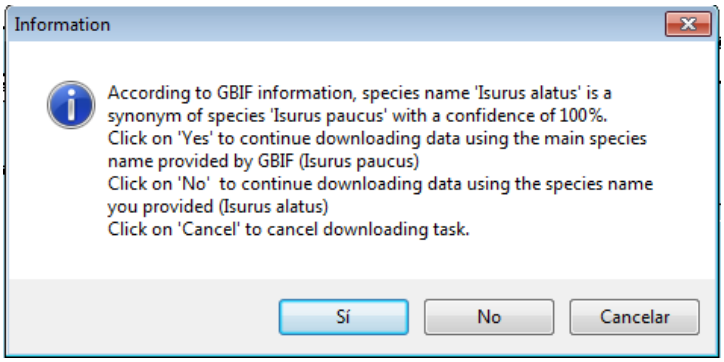


3. Importing occurrences from the GBIF online database. On the main menu, select *File/Import/Occurrences from online GBIF database* or the equivalent button of the toolbar. You will have to write the desired species name. Alternatively you can select it from a ModestR database, if you already have one populated with taxonomic data; this will be explained later on this tutorial. You can also enter an LSID that identifies a species, then MapMaker will try to resolve it.

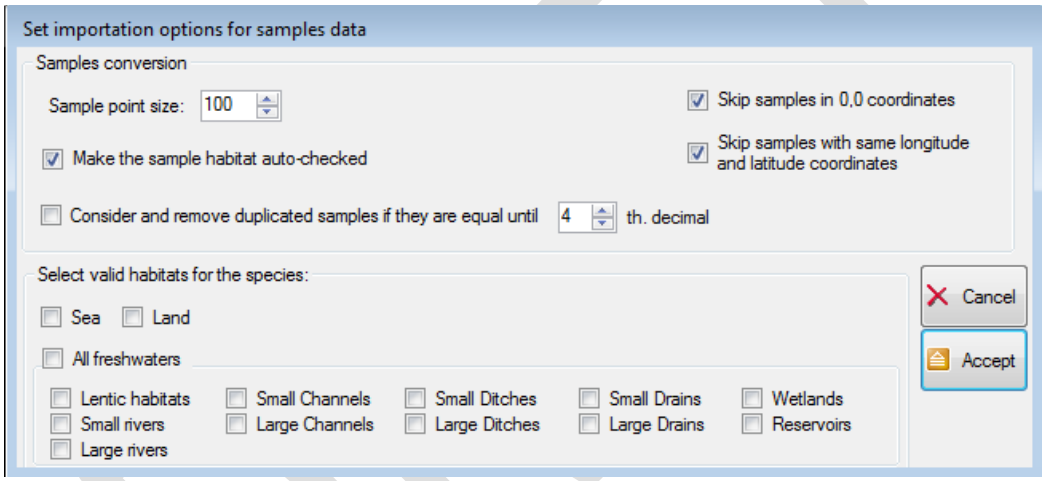



Then you can choose if you want MapMaker to look for synonyms in GBIF database. You can also select the type of records you want to download from GBIF, the rectangular area to be used as boundaries (by default it is the whole world), the start and end date of the occurrences to retrieve, and the occurrences added or modified into GBIF database since a particular date. Remember that to obtain a brief explanation of what each option and field is, just place the mouse on it to see a tooltip. However, a more complete description can be found on GBIF documentation. You can also select to apply a data cleaning once data downloaded, using some dispersal capacity variables. By default, mean distance variable is selected to perform data cleaning. Before using other variables, it is recommended to see section 2.6 for more details about dispersal capacity cleaning.

If you select to look for synonyms, MapMaker will first check if the species you entered is considered as a synonym of another one according to GBIF, and ask you if you want to use it to search occurrences in GBIF.




When importing occurrences either from a CSV file or from GBIF database, or when manually adding a occurrence, you will have to select how MapMaker will treat those occurrences. Several options will be displayed in a dialog box when importing from GBIF or CSV files (when adding manually an occurrence only some options will be available and they will be displayed in the left tool panel):




- Occurrence point size: select the size of the visual point that will indicate an occurrence in the map. This size has no importance when processing map. It is only a visual parameter.
- Skip occurrences in 0,0 coordinates: don't process occurrences whose coordinates are 0º longitude and 0º latitude, as those occurrences are usually erroneous data.
- Skip occurrences whose longitude and latitude coordinates have the same value, as those occurrences are usually erroneous data.
- Consider and remove duplicated occurrences if they are equal until the Xth decimal: this allows setting when two occurrences will be considered as duplicated, thus avoiding adding them twice. If, for example, we configure this option to take into account until the 2th decimal, occurrences like (122.4567;9.4568) and (122.4598;9.4560) will be considered duplicated. If you don't check this option, all occurrences will be imported and added to the map, even if they are duplicated.
- Make the occurrence habitat auto-checked: using this option an occurrence will be automatically checked by MapMaker to determine if it is situated in a valid or invalid habitat, and therefore it will be considered valid or invalid with a precision of aprox. ±1 seconds[3].

[3] This requires installing the last software updates. If you want to distinguish freshwaters, you also have to install freshwater data in ModestR (available in ModestR website). Anyway, as freshwater habitats are

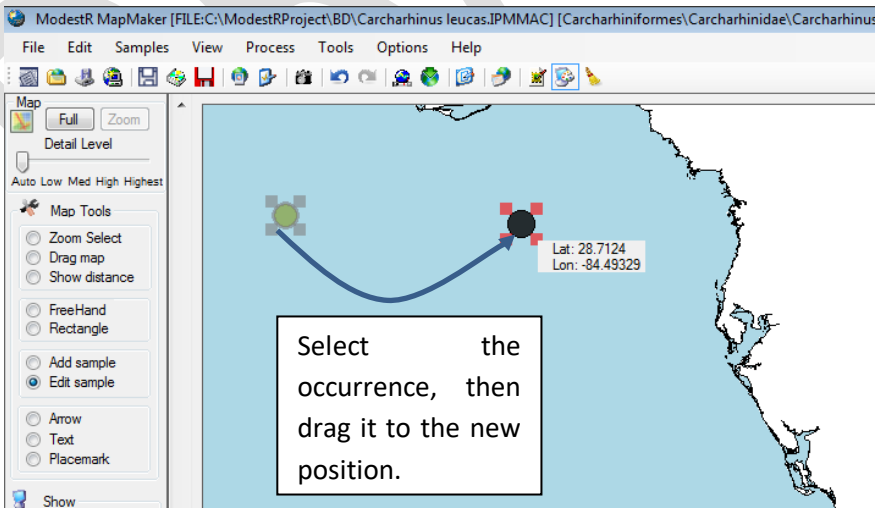
You can set different colors for valid an invalid occurrences to easily distinguish them in the Options/Preferences option on the main menu of MapMaker. Moreover, only valid occurrences will be used when processing and extracting data from a map (like presence or richness data). If on the contrary you don't set an occurrence as dynamic, it will be considered as manually validated/invalidated. By default, it will be valid, and you have to manually set it as invalid if needed. Therefore, we recommend you to try the dynamic validation.

- Select valid habitats for the occurrences: you have to select which habitats are valid for the occurrences you are importing (this option appears on the left tool panel when manually adding occurrences). You can select one or more occurrences, but it is mandatory to select at least one habitat. If you selected to make the occurrences as habitat auto-checked, those habitats will be used to validate/invalidate a occurrence. If you don't selected to make the occurrences as dynamic, this information will NOT be used, as occurrences have to be manually validated/invalidated. But it will be stored and used in case you later modify an occurrence to make it as habitat auto-checked.

## 2.4 Modifying occurrences in a map

Once occurrences added to a MapMaker map, you may want to modify them. MapMaker allows the following possibilities:

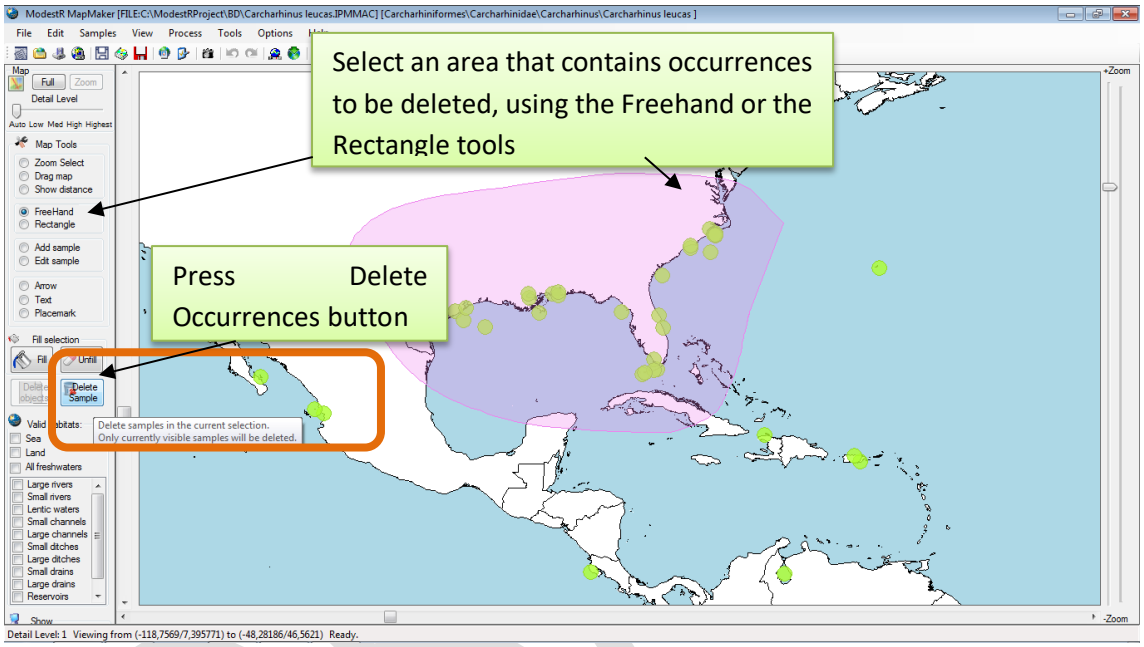
- Move a occurrence: sometimes you can want to relocate an occurrence to a near location. For example, to relocate an occurrence incorrectly placed in the sea for a few meters, for a terrestrial species. To do that, you can use the *Edit occurrence* tool from the left tool panel. With this tool selected, go to the map, click on the wanted occurrence and drag it to the new position, then drop it.




- Delete one single occurrence: To do that, you can use the *Edit occurrence* tool from the left tool panel. With this tool selected, go to the map, click on the wanted occurrence and press the *Delete* key. You can also click on the occurrence with the right mouse button and select the *Delete occurrence* option.

still being reviewed and updated, in some areas checking feature may only distinguishes between land and sea habitats.

- Delete a group of occurrences: to delete several simply select an area that contains the occurrences you want to delete using the *Freehand* or the *Rectangle* tools from the left tool panel, and press the *Delete occurrences* button on the *Fill Selection* groupbox. Or use the context menu item *Selected occurrences/Delete occurrences* (pressing the right mouse button to show it). Take into account that only occurrences that are currently visible in the selected area will be deleted. If for example you set display selections to only show invalid occurrences, only those occurrences will be affected by a delete operation (occurrences can be shown/hidden using *View/Map Show Options/Show Occurrences* menu or in *Show* options in the left tool panel).




- Modify occurrences marker size: to modify the size of the visual point that will indicate an occurrence in the map, you can select an area that contains the occurrences you want to modify using the *Freehand* or the *Rectangle* tools from the left tool panel, and then use the context menu item *Selected occurrences/Set size* (pressing the right mouse button to show it). To modify the size of all occurrences of the map, use the *Occurrences/Modify occurrences marker size* in the main menu. Remember that this size has no importance when processing map. It is only a visual parameter.
- Validate/unvalidate occurrences: if you want to manually validate or invalidate occurrences, either dynamic occurrences or not, you can select an area that contains the occurrences you want to modify using the *Freehand* or the *Rectangle* tools from the left tool panel, and then use the context menu item *Selected occurrences* (pressing the right mouse button to show it). You will find options to validate/unvalidate all selected occurrences, or only those that are valid/invalid. To validate/unvalidate all occurrences of the map, use the options in the *Occurrences* menu of the main menu. Remember that if you manually validate or invalidate occurrences, this has two side effects: (a) occurrences become not habitat auto-checked. So they will not be checked anymore, but keep the valid/invalid status you manually set. And (b) manually validated occurrences will be took into account when calculating data (presence or richness for

example), while invalidated one will be stored with the map, but not took into account for any calculation.

- Set occurrences as habitat auto-checked: if you added or modified occurrences setting them as manually validated/unvalidated, you may want to reset them as habitat auto-checked, this way MapMaker will automatically check their validity regarding the valid habitats they have associated. To do that you can select an area that contains the occurrences you want to modify using the *Freehand* or the *Rectangle* tools from the left tool panel, and then use the context menu item *Selected occurrences/Set as dynamic occurrences* (pressing the right mouse button to show it). To make all occurrences of the map to be dynamic, use the option in the *Occurrences* menu of the main menu.
- Modify occurrences valid habitats: you may have downloaded or imported occurrence data and added them to the map indicating some habitats as valid, but now you want to modify this selection, for example, to add more freshwater habitats as valid for a species. To do that, just go to the *Occurrences/Modify occurrences valid habitats* in the main menu.

## 2.5 Importing data to a MapMaker map

Besides importing occurrences from GBIF or CSV files, as explained below, MapMaker can also import existing maps or data from other sources, such as distribution models generated with Maxent[4], ESRI shapefiles, or KML files. Each one of those importation features is explained in more detail in chapter 7.

## 2.6 Data cleaning using automatic range-based rules

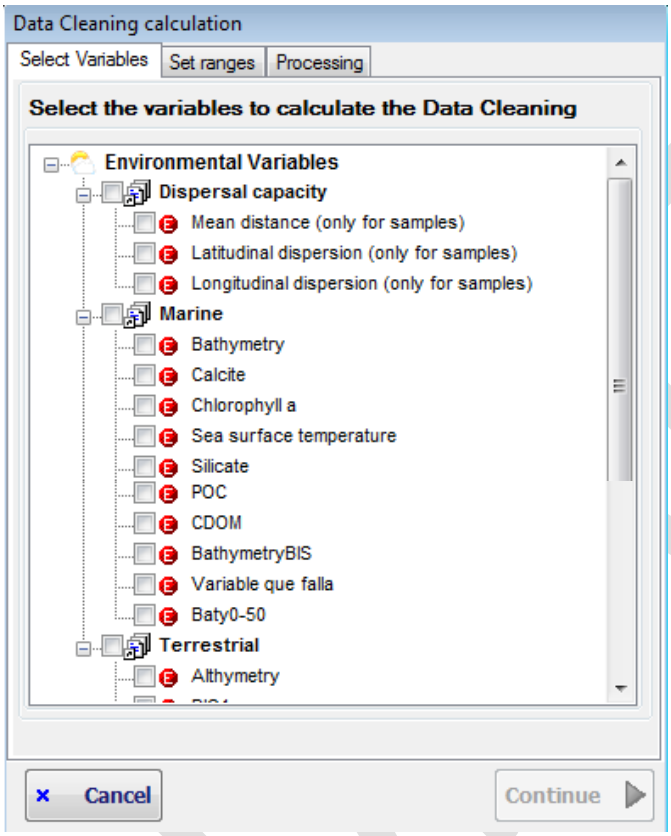
MapMaker incorporates the possibility of using automatic validation rules basing on outliers detection to determine the validity of occurrences as well as areas of presence according to environmental variables. This allows for example to invalidate occurrences located in places that have some environmental conditions such as temperature or altitude values outside a certain range. Or to delete from an area of presence the parts where some environmental conditions are not respected. To do that, ModestR will firstly determine the range of values took by those variables on the current presence data, and allow the user to select how to determine outliers. To determine current ranges for variables in occurrence maps, one single value for each variable is took in the coordinates of the occurrence. For range maps (with areas of presence in turn of occurrences) a value for each variable is took for each occupied cell of 1'x1' (so for example for a presence area of 10'x4', 40 values will be took for each variable). Data cleaning will be done for each individual occurrence, and, in case of areas, for each 1'x1' cell of the presence areas. So in case of areas the minimal resolution supported when cleaning is of 1'x1'.

Environmental data needed to use this features are not included in ModestR, so they have to be provided by the user as data files in ESRI ASCII or CSV format. There exist free datasets available in Internet, such as Bi-Oracle (Tyberghein et al. 2012) for example that are fully compatible with ModestR.

[4] Futher information about Maxent software can be found in: Steven J. Phillips, Miroslav Dudík (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31(2):161-175. DOI: 10.1111/j.0906-7590.2008.5203.x

Therefore, the first step will be to integrate those data in ModestR, to allow it to use them when needed. To do that, you can use the menu *Layers/Manage environmental variables*, either in MapMaker or in DataManager. Details about how to integrate environmental data in ModestR are explained in section 6.26.

Once environmental data has been made available to ModestR, automatic data cleaning can be started using the menu *Mapping/Data Cleaning/Automatic environmental based cleaning* of MapMaker. A dialog box like the shown below will appear:




In this dialog box you can select on the tree the variables you want to use in the data cleaning. Besides the environmental variables you have integrated in ModestR, you could see three specifically calculated variables related to dispersal capacity which are also available: *mean distance*, *latitudinal dispersion* and *longitudinal dispersion*. Mean distance for each occurrence record is calculated as the mean distance from this record to all the other records. Latitudinal dispersion and longitudinal dispersion merely are the latitude or longitude[5] value of each occurrence record. These variables allow validating records basing on their relative geographical dispersion, so records excessively distant from the others will be cleaned. It must be pointed out that those variables can only be used on maps that contain occurrences, but not on range maps.

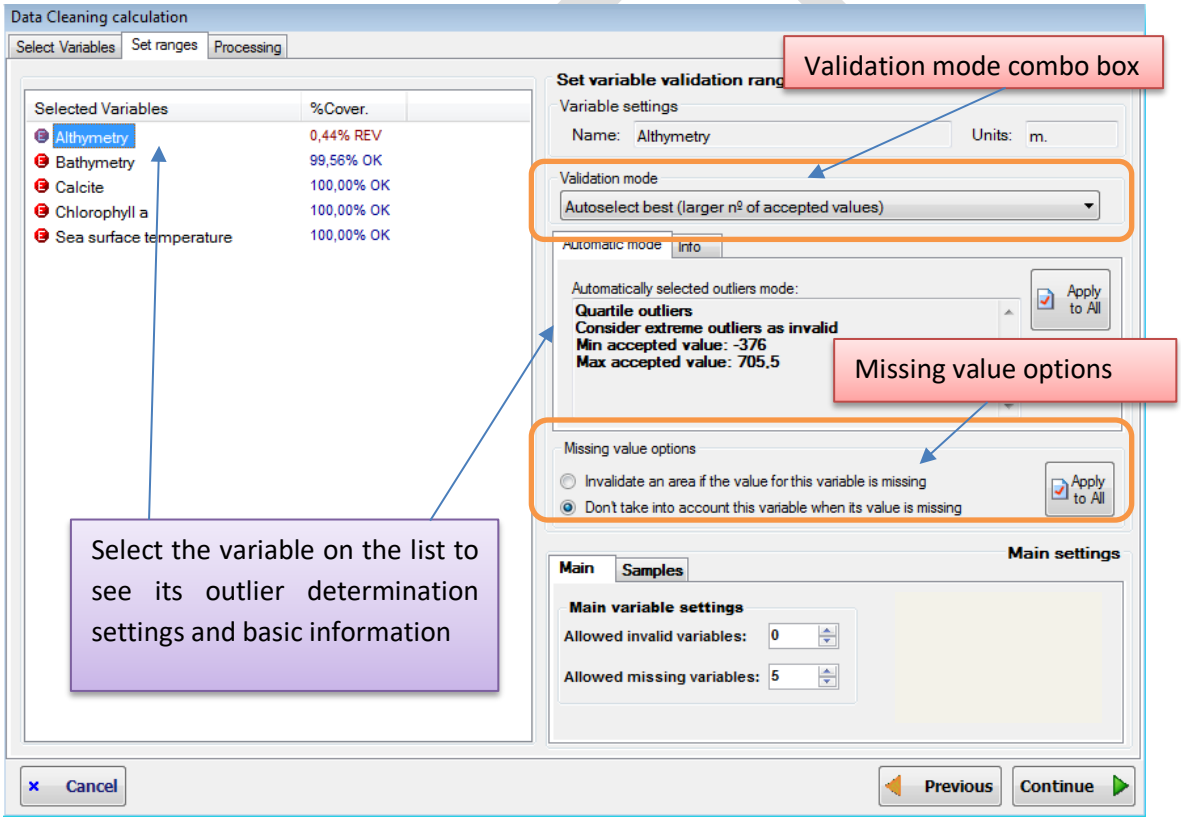
Once all whished variables selected, click on the *Continue* button to go to the next step. ModestR will then determine the range of values took by the selected variables on the current presence

[5] Take into account that using longitudinal dispersion (which is merely the longitude value of each occurrence) to perform data cleaning can lead to biased results, because longitude values of occurrences near the extremes (i.e. -180º or 180º) are very different (so potential outliers), while occurrences can actually be quite close.

data, whether areas or occurrences. It will also determine the coverage percent of each variable; that is, the percentage of presence areas/occurrences where there is a not-null value for a variable. This can helps you to determine if a variable is really useful to perform data cleaning.

Those data will be shown in the next step, as shown in the figure below. You can see on the left the list of selected variables, with a new column where the coverage percent is presented. As you can see, in this example the *Althymetry* variable has a very low coverage, smaller than 1%. ModestR will even show you a warning message in those cases, and show the values in another color. In this example this low coverage is because presence data correspond to a marine species, so there are no althymetry data in those areas. You most likely remove this variable from the list, by going to the previous step (using the *Previous* button*)* and deleting it from the list of selected variables. Then you will return to this step using the *Continue* button.

Before continuing, it is interesting to point out that many of the options and settings displayed on this dialog box are autoexplained: just by placing the mouse cursor on one element, usually a short emerging explanation will be shown.




Then you can select how outliers will be detected for each variable. To do that you must select a variable on the left listbox. The corresponding settings for this variables will be shown on the right-up panel. The available settings are:

- **Validation mode:** this is the way the values for this variable on the presence zones (whether occurrences or areas) will be evaluated to determine if some occurrence/area has to be cleaning or not. Any validation mode will determine the range (minimum and maximum) of valid values for the variable, so any value that fall outside this range will

be considered an outlier[6]. The difference between different validation modes will be in the way they use to determine this range. By default, the validation mode will be set to *"Autoselect best"*. But you can manually select between the following modes:

- ***Autoselect best**:* this mode will evaluate which one of the other available validation modes returns the less outliers, and uses this mode to clean data. That is, with this mode you will always obtain the minimal number of outliers. The specifically selected mode will be described in the box below, under the "*Automatic mode"* tab, and it will be one of the available modes that are described below: *Quartile outliers* or *JackKnife outliers*.
  *Autoselect best* mode will not be available for variables with a coverage of 0%; that is, without not-null values for the current presence areas.
- ***Quartile outliers***: this option uses a validation based on the values of the Q1 and Q3 quartiles of the values took by a variable in the current presence areas/occurrences. This option will consider any value of this variable as invalid if it is outside the range [Q1 - X(Q3 - Q1) , Q3 + X(Q3 - Q1)] where X will be:
  - X=1.5 if you select to consider both mild and extreme outliers as invalid.
  - X=3 if you select to consider only extreme outliers as invalid.
  - You can also set a custom value of X in the options shown for this validation mode.
- ***JackKnife outliers**:* this validation mode to detect outliers is described in Chapman's *Principles and Methods of Data Cleaning* (2005) and it is considered quite reliable. ModestR supports two variants of this method: the one proposed by CRIA and the one proposed in DIVA-GIS (both are described in Chapman's book).
- ***Custom***: in this case it is the user that will manually set the minimum and maximum values of the range of valid values for a variable. Any value that fall outside this range will be considered an outlier.

- **Info tab:** under the validation mode combo box, a tabbed box is shown where information and options for the currently selected validation mode are displayed. There is also a *Info* tab where you can consult basic statistics for the currently selected variable, such as maximum and minimum values that this variable has on the current presence areas/occurrences, mean value, quartiles, and jackknife ranges.
- **Missing value options:** under the tabbed box that show the information for the variable, two radiobuttons allow you to select what to do when there is no value for the variable in a particular area/occurrence. You can select between two possibilities:
  - Invalidate an area/occurrence if the value for the variable is missing (that is, if the variable have a null value in this location): in this case, an area/occurrence will be cleaned (considered invalid) as soon as there is no value for the variable for this area/occurrence, no matter the values for the other variables.
  - Don't take into account this variable when its value is missing: in this case, if there is no value for the variable for this area/occurrence, this variable will be

[6] It must be pointed out that for *mean distance* variable (included in the *Dispersal capacity* group), the minimum value of the acceptance range is always set to zero, to avoid eventual elimination of occurrences with small mean distances to the other ones.

just not considered to make a decision. ModestR will check the values took by the other variables to decide about the validity of the area/occurrence.

Take into account that those options are *per variable*. That is, you can set different options for each variable. A missing value for a variable can lead to immediately invalidate an area/occurrence, while for another variable it will not be took into account. You should decide in a per-variable basis the most adequate option.

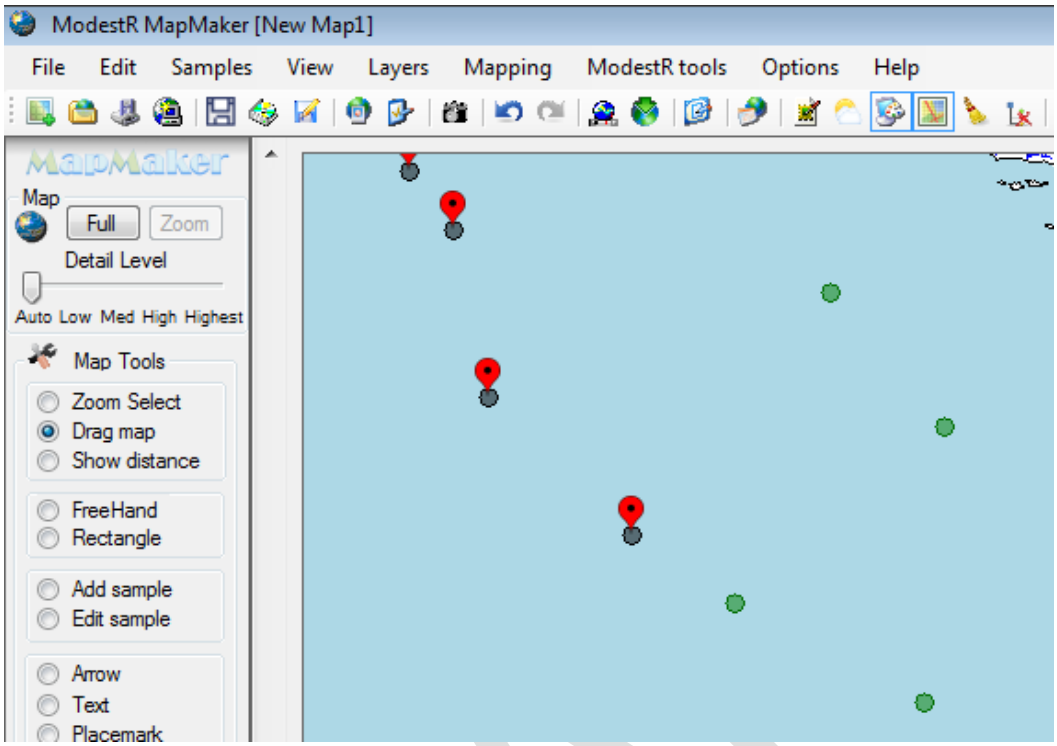
- **Apply to all buttons:** either in the tabbed box that displays the options for the currently selected validation mode or in the missing value options, you can see a button *Apply to All*. This button can be used to replicate the corresponding settings for the current variable to all variables in an easy way.

Finally, on the right-bottom side of the dialog box, you can see a tabbed box with main settings. While settings described above are for each variable, those main settings have global effects. They are:

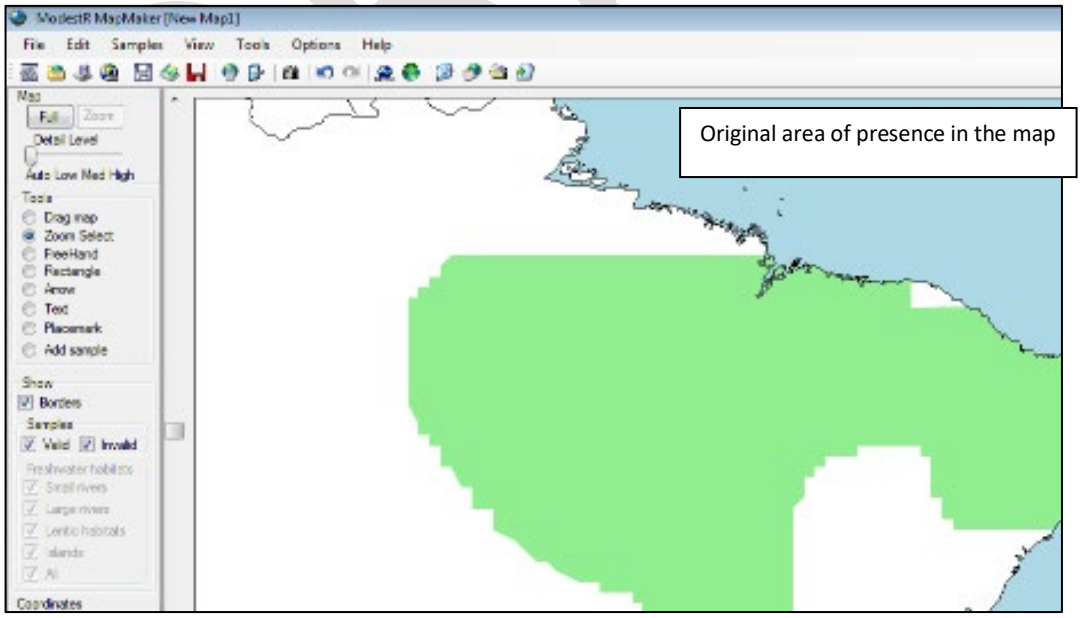
- **Allowed invalid variables:** Indicates how many variables can take invalid values, that is, outliers (according to their validation rule) before considering an area or occurrence as invalid for the species. The default value is 0, and the maximum is N-1 where N is the number of selected variables. Usually this field will be 0, because typically you will want to clean a occurrence/areas as soon as one of the selected variables takes an outlier value. But this setting provides a way to be more flexible. For example you can select 10 variables and allow that up to two variables of these take outlier values before considering a occurrence/area as invalid.
- **Allowed missing variables:** Indicates how many variables can be missing (have a null value) before considering an area or occurrence as invalid for the species. By default it will be set to N where N is the number of selected variables. That is, if all variables have null values for a occurrence/area, the area will be considered as valid. Take into account that this parameter will be applied only if the variables with missing values are not individually configured as directly invalidating an area when their value is missing (as explained above). That is, this rule will be applied only if after applying the rules for each individual variable the occurrence/area has not already been considered as invalid.
- **Occurrences tab:** in this tab page you have to select which occurrences apply to (all, only already valid ones, only already invalid ones, or none). You can also select to add placemarks to validated/invalidated occurrences to make easier localizing them on the map. This tab will only be shown if the map contains occurrences.
- **Areas tab:** in this tab page you can select to validate or not validate areas of presence. To visualize the zones that don't comply with the rule, you can add a shape, usually using a semitransparent showy color (see section 2.16 for more information about shapes). The shape doesn't modify map presence data, as it's just a visual helping element. This tab will only be shown if the map contains areas.

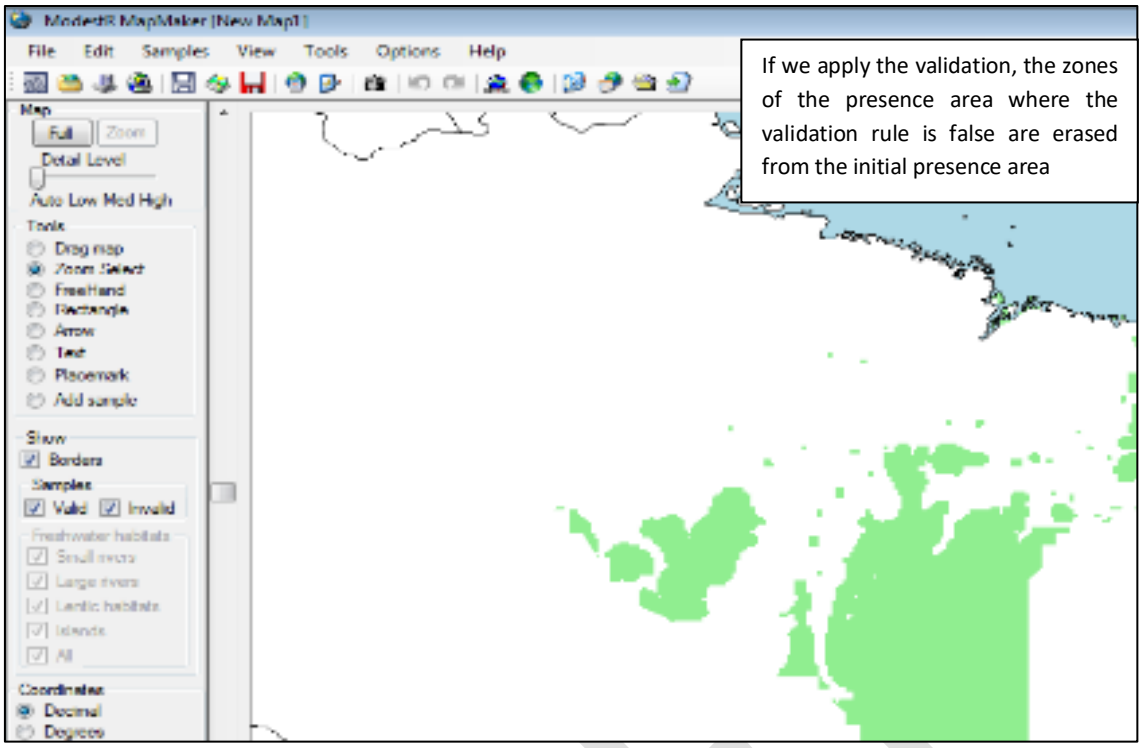
Once all those settings configured, press the *Continue* button to perform data cleaning. ModestR will apply validation rule for each variable on each occurrence and on each 1'x1' cell of each presence area.

For example, in this map, after the validation, a red placemark indicates the occurrences located in points that don't comply with the validation rule. Green ones indicate occurrences located in points that comply with the validation rule (colors are eligible by the user).




If you have selected to apply validation to areas, the zones of the area where the validation rule is not true will be deleted. Areas validation can require a certain amount of time, depending on the complexity of the area.



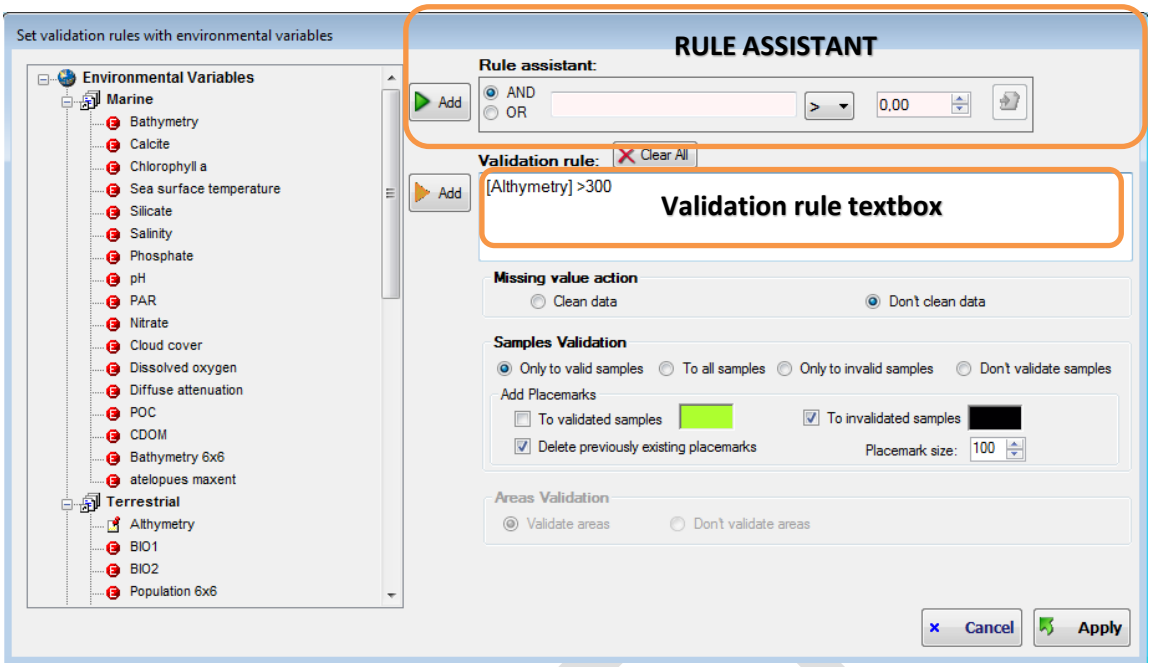



Once a data cleaning task ended, and only if this cleaning involved occurrences, MapMaker will allow you to save a report of the cleaning operations done in CSV format containing a list of the cleaned occurrences and details about the variable ranges they didn't comply with. As said before, this report will only be generated when cleaning occurrences, and will not include details regarding areas cleaning. The reason for this is that even a relatively small region will easily contains thousands of 1'x1' cells, which results in a very long and rarely useful report.

## 2.7 Data cleaning using custom validation rules

Besides automatic cleaning, MapMaker incorporates the possibility of using custom validation rules to determine the validity of occurrences as well as areas of presence according to environmental variables. As mentioned before in section 2.6, environmental data needed to use this features are not included in ModestR, so they have to be provided by the user as data files in ESRI ASCII or CSV format and integrated in ModestR as explained in section 6.26.

Once environmental data available to ModestR, validation rules can be added using the menu *Mapping /Data Cleaning/Custom environmental based cleaning* of MapMaker. A dialog box like the shown below will appear.




In this dialog box you can select any environmental variable and add it to the expression that will be used as validation rule. You can either use the Rule assistant that helps you to construct the rule, or just write it into the validation rule box. To use the Rule assistant, follow those steps:

- Select the environmental variable in the tree
- Click on the *Add* button aside the Rule assistant
- Select the adequate relation operation (>, <, =, …) and enter a value to compare to.
- Select it the rule will be added to the validation rule as an AND or an OR rule (only relevant when there are more than one rule).
- Click on the button to *Add this rule to Validation*. The rule part will be added to the validation rule textbox.

Alternatively you can directly write or modify a validation rule in the Validation rule textbox, using the lower *Add* button to add variables from the tree to the expression.

Validation expression have to be logical or Boolean (that is, evaluated as true or false). Variables used in the expression have to be written inside brackets (this will be automatically done if you use the *Add* buttons to add them. AND and OR operators are allowed. Using the right mouse button on the validation rule box a contextual menu will appear showing more options and operators.

Once a valid rule introduced, you have to select which occurrences apply to (all, only already valid ones, only already invalid ones, or none). You can also select to add placemarks to validated/invalidated occurrences to make easier localizing them on the map. You can also select to validate areas of presence using the validation rule.

Use the *Apply* button to clean presence data according to the validation rule. In this case, if you selected to apply validation to occurrences, occurrences located in a point where the rule is true will be set as valid. Otherwise they will be set as invalid. Nevertheless, if an occurrence is

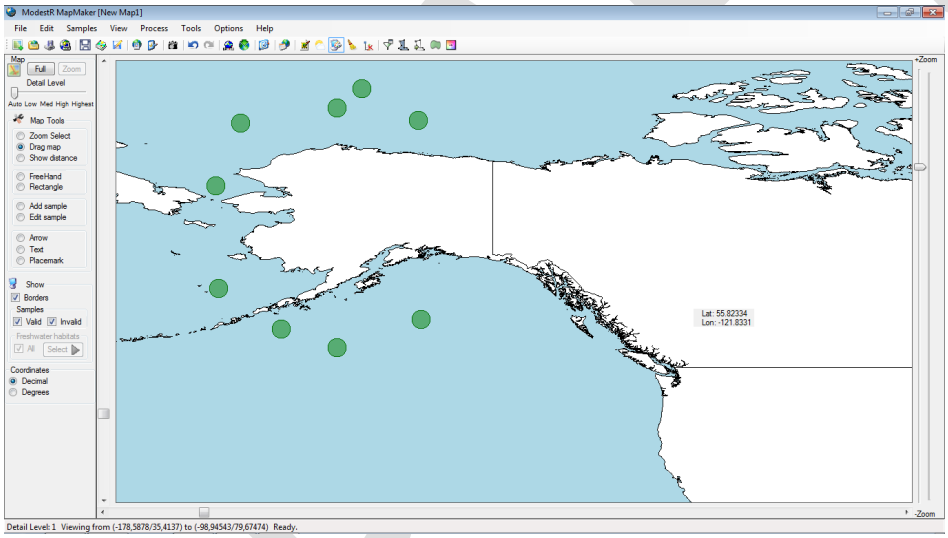
dynamic (as explained in previous section) and it is invalid because of its habitat, it will always remain invalid.

An example of how results can look like is shown in section 2.6 above.

## 2.8 Using hull transformations tools

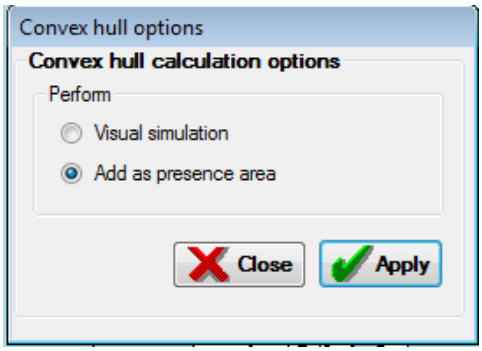
Hull transformations tools allows you to see/add a hull as an area of presence to a map. This feature can be used to estimate the Extent Of Occurrence[7] (EOO) of a species. ModestR provides three different hull transformations: convex hull, concave hull and density-based hull. Concave hull uses the alpha shape algorithm. Density-based hull uses a normal kernel smoothing algorithm. Those features are mainly aimed to see/add a hull calculated from occurrence data (occurrences). Therefore, they commonly do not take into account presence areas other than occurrences.

Convex hull transformation calculates a single minimal convex polygon that contains all the occurrences present on a map. Areas cannot be took into account. At least three occurrences are required to calculate convex hull. As an example, we'll assume we have a map with the occurrences shown below.

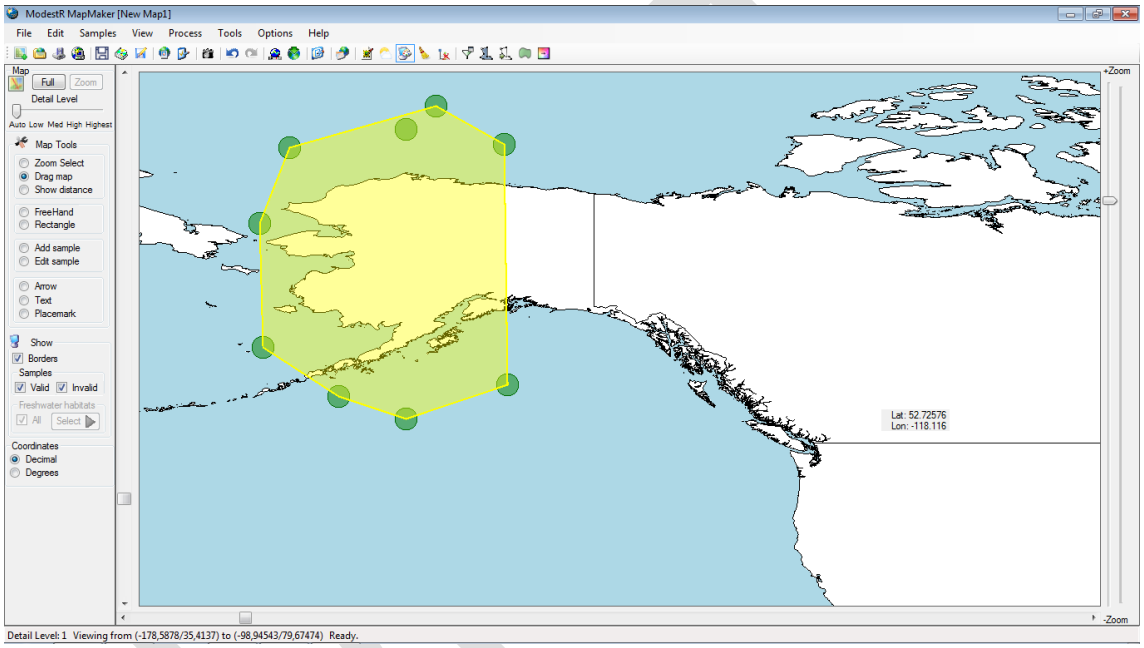



To calculate convex hull area for those occurrences, just go to *Mapping /Hull transformation/Convex hull* menu item. A dialog box will be shown where you can select the wanted feature:

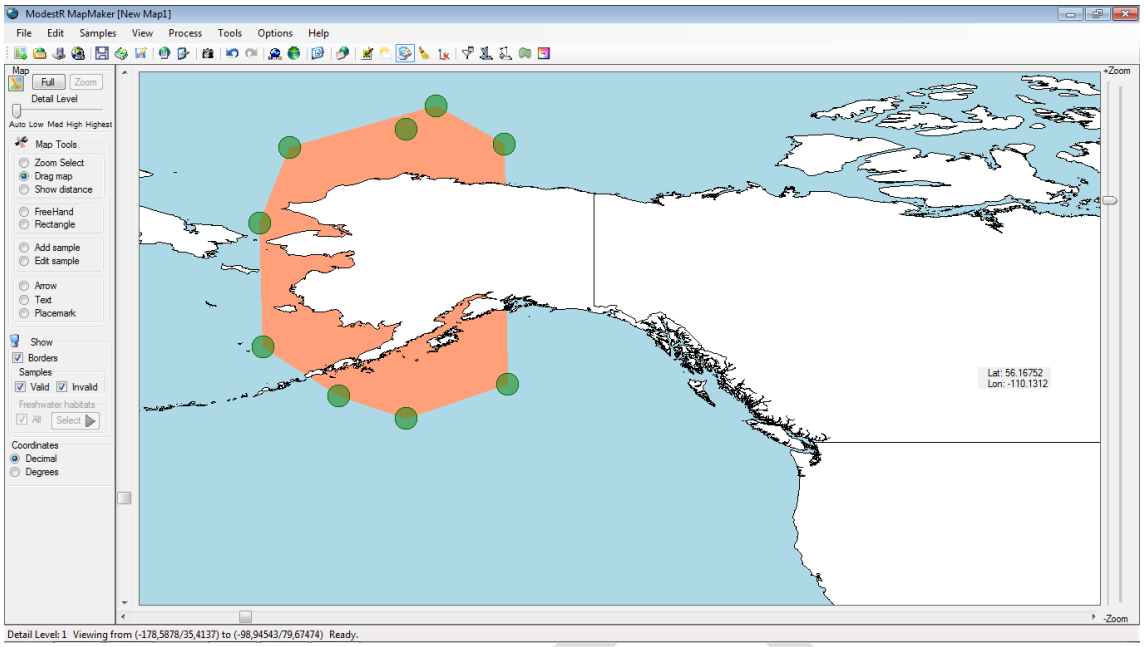
[7] Extent of occurrence (EOO) is defined by IUCN as the area contained within the shortest continuous imaginary boundary which can be drawn to encompass all the known, inferred or projected sites of present occurrence of a taxon (IUCN 2013)




- Visual simulation: this option will just add a shape to the map that shows the convex hull. No modification is made on the map. Shapes can be deleted using the *Edit/Delete all shapes* option or the corresponding button of the toolbar. However shapes are not saved with the map. For the example map above, this will be the result:




- Add as presence area: this option will add a presence area to the map corresponding to the convex hull. Existing data will not be deleted, but this adds a new presence area to the map. The habitats assigned as valid to this new area will be inferred from the already existing presence data (occurrences and areas). You can undo changes using *Edit/undo* or modify this new area as any other area, using *Unfill* feature (see section 2.2). For the example map above, this will be the result:




### 2.8.1 Using convex hull with areas

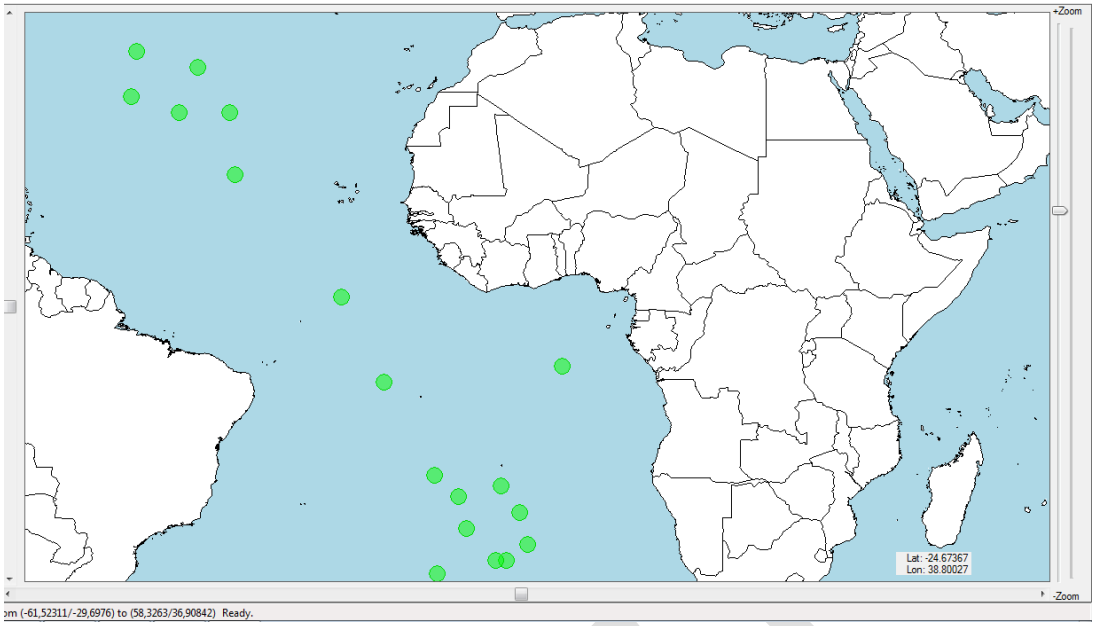
Convex hull is mainly intended to be used with occurrence data. But it can also be used with area data, as ModestR can use the points of the vertexes of the areas to calculate convex hull. However, when you add a presence area to MapMaker, you have to select a valid habitat, and MapMaker will visually show the target area according to this habitat. But this is actually just a visual presentation. MapMaker doesn't have the real data of the resulting area (this requires a posterior rasterization process that will be made by DataManager). That is why calculating the convex hull of an area in MapMaker is not allowed. When processing maps in DataManager the maps are rasterized and only real presence area is took into account. Therefore in DataManager convex hull including areas can be accurately calculated (see section 3.8).

### 2.8.2 Using alpha shape hull with occurrences

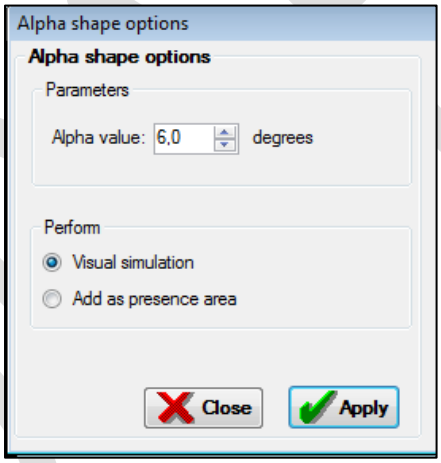
The alpha shape algorithm calculates one or more polygons that contain some or all the occurrences on a map, depending on a parameter (usually called the alpha value). Intuitively, the α-shape algorithm can be seen as rolling a circular object of radius α making it to lean against the standing out nails on a board, that represent the points of a set. If we connect the nails that constitute the edges of the set for this object of radius α, we obtain one or more closed areas that constitute the α-shape (see Edelsbrunner & Mücke (1994) work for more details). As the α-value is increased, the α-shape tends to the convex hull, while it tends to the initial set of points when it is decreased.

This feature cannot be used with areas, so it is mainly useful in occurrence maps. At least four occurrences are required to calculate alpha shape (but depending on their relative distances it can be impossible to calculate a not-null alpha shape).

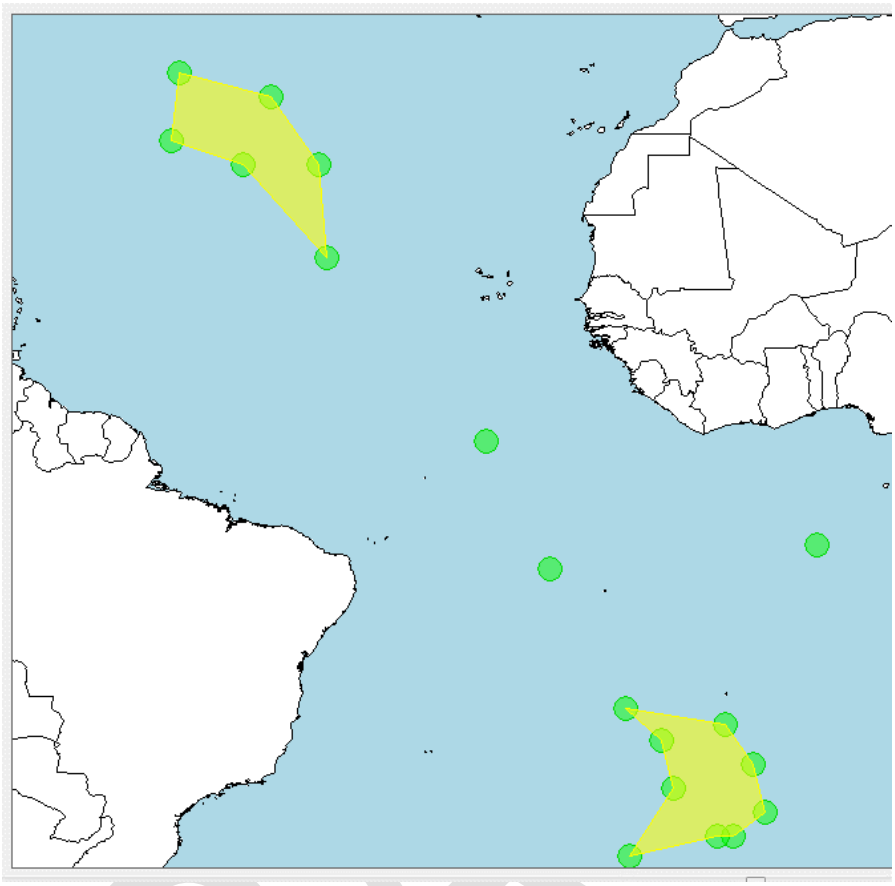
As an example, we'll assume we have a map with the occurrences shown below.




To calculate concave hull area using the alpha shape for those occurrences, just go to *Mapping /Hull transformation/Alpha shape* menu option. A dialog box will be shown where you can select the wanted feature:




- Alpha value: this value is the main parameter we have to provide to the alpha shape algorithm.
- Visual simulation: this option will just add a shape to the map that shows the alpha shape hull. No modification is made on the map. Shapes can be deleted using the *Edit/Delete all shapes* option or the corresponding button of the toolbar. However shapes are not saved with the map. For the example map above and an alpha of 6 degrees, this will be the result:



As you can see, alpha shape can return several separated polygons, and it may not contains all the occurrences, depending on the alpha value and the distances between the occurrences. You can easily test alpha shape with different alpha values to see the differences and select the value you consider more accurate.

- Add as presence area: this option will add a presence area to the map corresponding to the alpha shape hull. Existing data will not be deleted, but this adds a new presence area to the map. The habitats assigned as valid to this new area will be inferred from the already existing presence data (occurrences and areas). You can undo changes using *Edit/Undo*, or modify this new area using *Unfill* feature (see section 2.2).

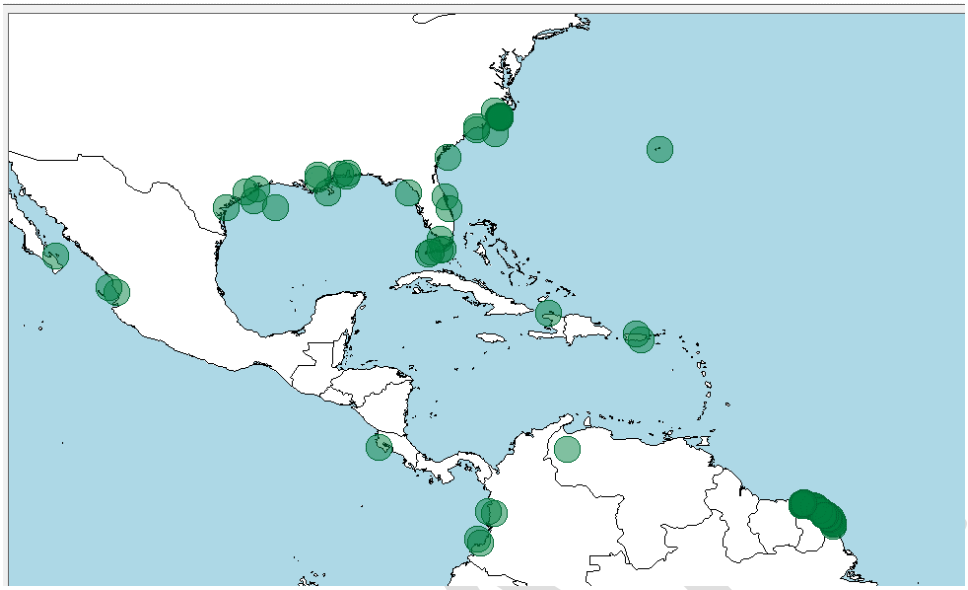
### 2.8.3 Using density-based hull with occurrences

The density estimation algorithm calculates one or more polygons that contain all the occurrences on a map, depending on a parameter (usually called the bandwidth). The density is estimated using a kernel smoothing normal distribution algorithm[8]. As the bandwidth is increased, the estimated hull is broader, while it tends to the initial set of points when it is decreased.

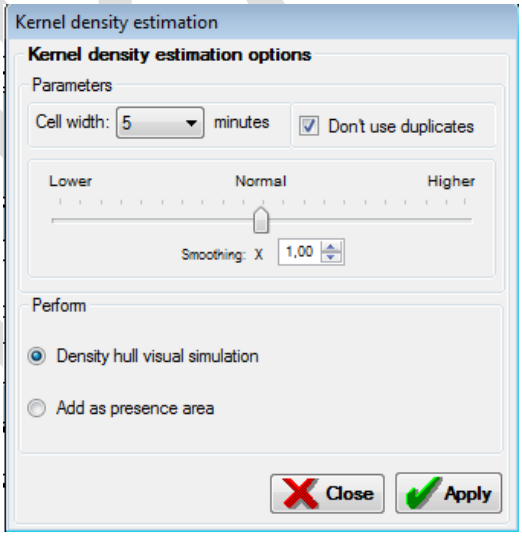
[8] For more information on kernel smoothing you can see Wand, M. P. & Jones, M. C. (1995) Kernel Smoothing. Chapman and Hall, London.

Density hull in ModestR is calculated using as cut-off threshold the minimal density value for the occurrences present on the map. This way, all occurrences will always be included in the calculated hull. At least three occurrences are required to calculate density.

This feature cannot be used with areas, so it is mainly useful in occurrence maps. As an example, we'll assume we have a map with the occurrences shown below.

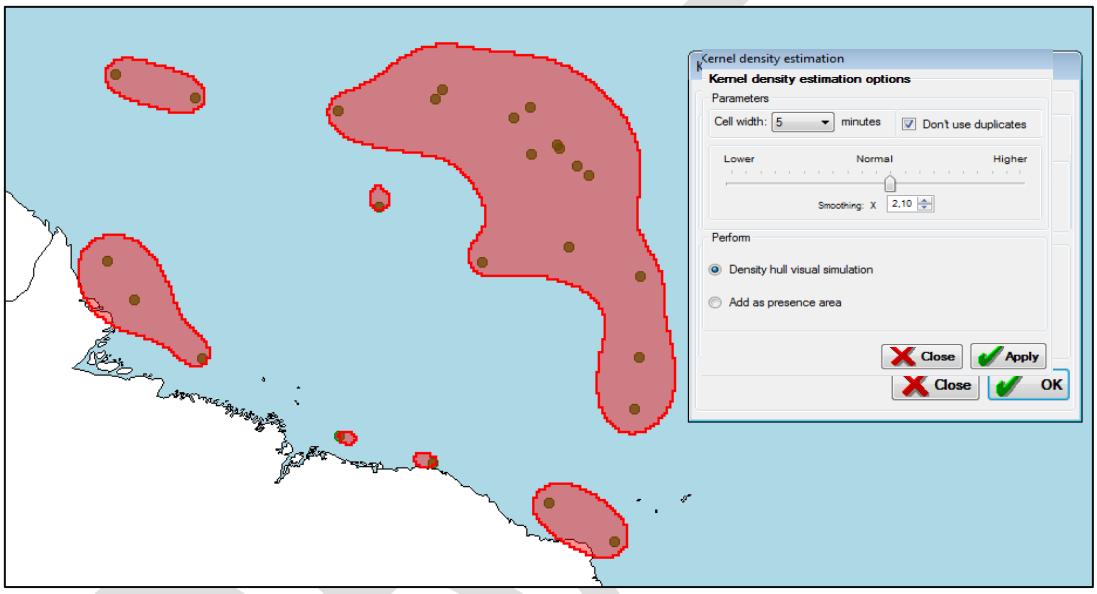


To calculate density estimation hull, just go to *Mapping /Hull transformation/Kernel Density/Distribution Map* menu option. A dialog box will be shown where you can set the parameters and select the wanted feature:




- Cell width in minutes: allows you to set the cell size to be used to calculate density. Density will be calculated for each cell and assumed to be the same in all the cell. So the smaller the cell size, the more precise is the density calculation. But take into account that the time to perform the calculations grows geometrically as cell size diminishes.

- Don't use duplicates: if enabled, only distinct occurrences (usually considered until the 5th decimal) will be used to calculate kernel density. This avoids biasing density because there are many duplicates of some occurrences.
- Smoothing: smoothing factor determines how smooth the density matrix will be. The greater it is, the more the density will spread out far around presence areas. Default factor (x1) has been adjusted to return a density area with a standard deviation of the distances between presence areas near the original standard deviation of the distances between occurrences.
- Density hull visual simulation: this option shows a shape of the hull calculated using the density estimation with the set parameters. For the example below, this will be the result:




As you can see, density hull usually return several separated polygons, but it will contains all the occurrences (even if the polygons can be very small for some occurrences). You can easily test density hull with different parameters to see the differences and select the value you consider more accurate. Anyway, remember that default smoothing value (x1) has been adjusted to return a density area with a standard deviation of the distances between presence areas near the original standard deviation of the distances between occurrences.

- Add as presence area: this option will add a presence area to the map corresponding to the density hull. Existing data will not be deleted, but this adds a new presence area to the map. The habitats assigned as valid to this new area will be inferred from the already existing presence data (occurrences and areas). However, you can undo changes using *Edit/Undo* or modify this new area as any other area, using *Unfill* feature (see section 2.2).

If you want to display a density map, just for presentation or analysis purposes, you can also use the option to add a kernel density map to the data, as explained in section 2.9.5.

## 2.9 Working with rasters

To make easier visually analyzing relationships between species distributions and environmental variables, or just to have appealing images to use in some documents, presentations, etc., you can add raster data to MapMaker maps. This is only for visual purposes and it doesn't affect distribution maps. Raster data will appear when exporting a MapMaker map to an image file (see section 7.6 to export images from MapMaker).

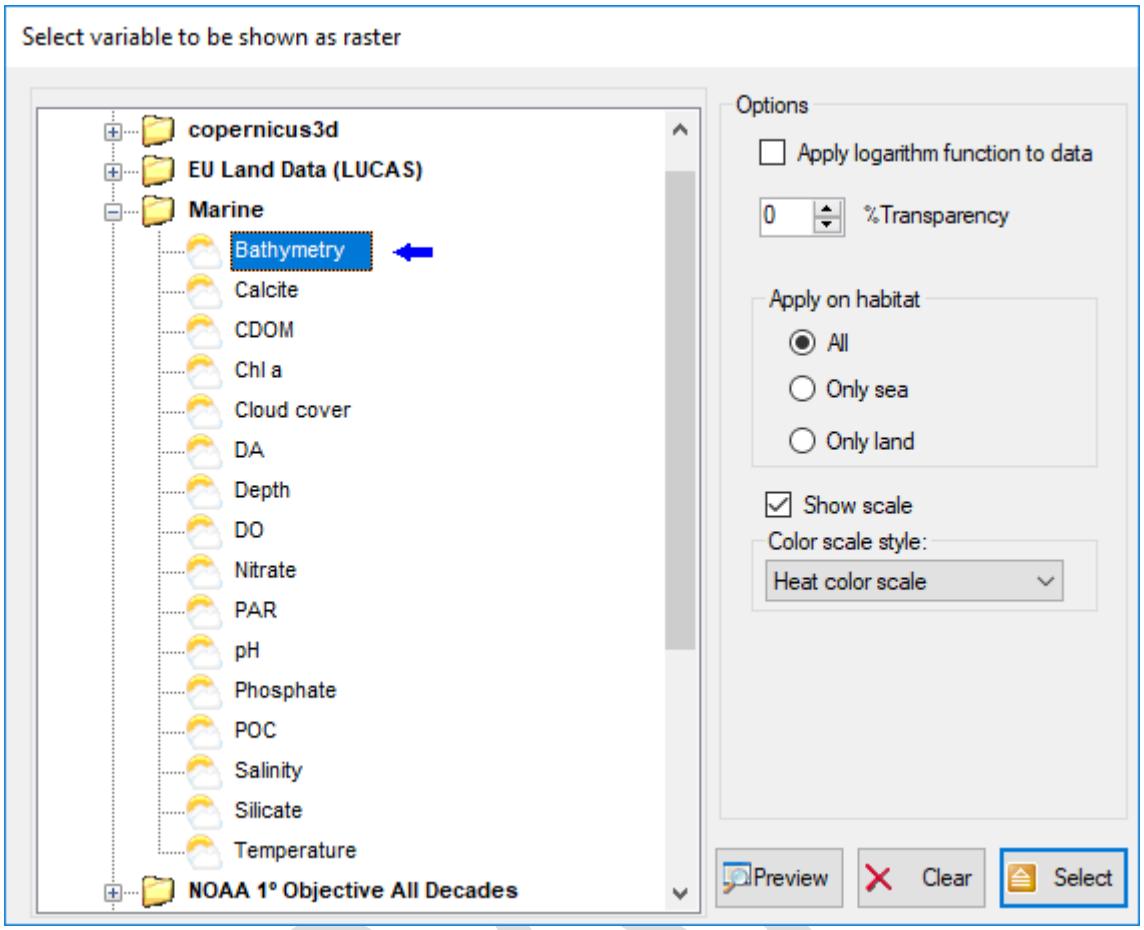
There are two main ways to add raster data: using environmental data that you already have integrated in ModestR, o directly loading raster data from an ESRI ASC file. We explain both of them in the next subsections. It must be pointed out that currently ModestR only supports raster data with a maximal resolution of 1'. Moreover, as using a raster of a resolution 1' for all the world is a heavy duty, we recommend to ordinarily use rasters of a resolution of 5', which may be enough precise for mere visualizing purposes.

### 2.9.1 Adding environmental data as raster

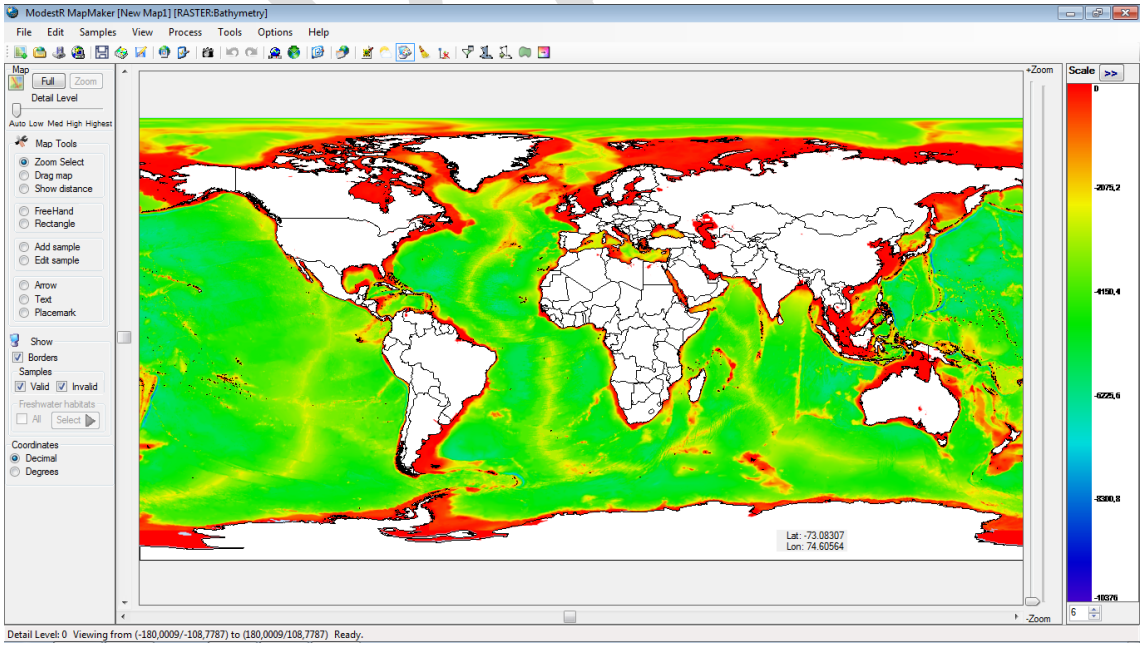
Environmental data needed to use this features are not included in ModestR, so they have to be provided by the user as data files in ESRI ASCII, GeoTiff, netCDF or CSV format. There exist free datasets available in Internet, such as Bi-Oracle or Copernicus (https://www.copernicus.eu ) for example that are fully compatible with ModestR.

Therefore, the first step will be to integrate those data in ModestR, to allow it to use them when needed. To do that, you can use the menu *Layers/Manage environmental variables*, either in MapMaker or in DataManager. Details about how to integrate environmental data in ModestR are explained in section 6.26.

Then, you can easily add one environmental variable to the map using the MapMaker menu item *Layers/Rasters/Add environmental data raster*, or the corresponding button of the toolbar. A dialog box will show to you the available variables. You can select only one variable (it should be visually confusing to show more than one). You can also select if the raster has to be shown only on sea or o land habitat, or on both.




You can preview the raster first. Once selected, the raster will be shown on the map like this:




A color scale will be shown on the right of the map, but it can be hidden using the >> button

You can change the raster to be shown just by doing the same steps and selecting another variable. To temporally hide/show this raster, you can use the menu item *View/Show selected*

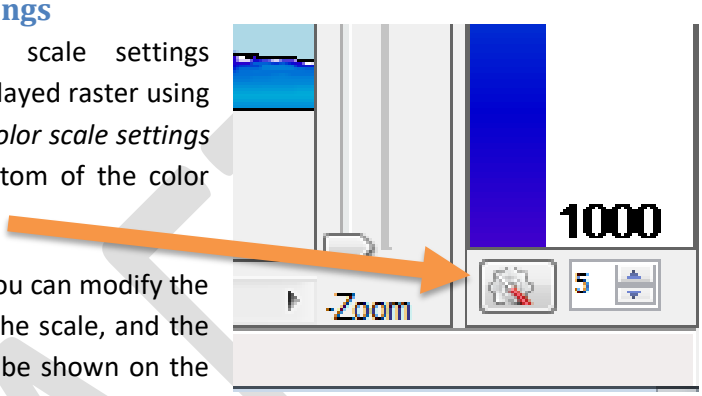
*rasters*, or the corresponding button of the toolbar. To definitely clear the raster from the map, you can use the menu item *Layers/Rasters/ Clear raster*, or the corresponding button of the toolbar.

### 2.9.2 Loading raster files

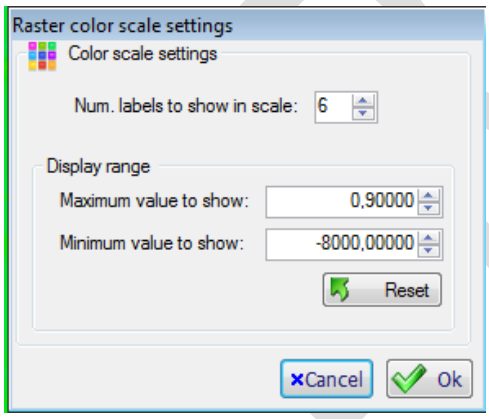
You can also directly load raster data form an ESRI ASC or a GeoTiff file using the menu item *Layers /Rasters/Load raster from ESRI ASC/GeoTiff file.* Everything else works the same way as explained in previous section for rasters using environmental data integrated in ModestR. You can also directly drag&drop a raster file with those formats onto MapMaker in order to show it in the map.

### 2.9.3 Raster color scale settings

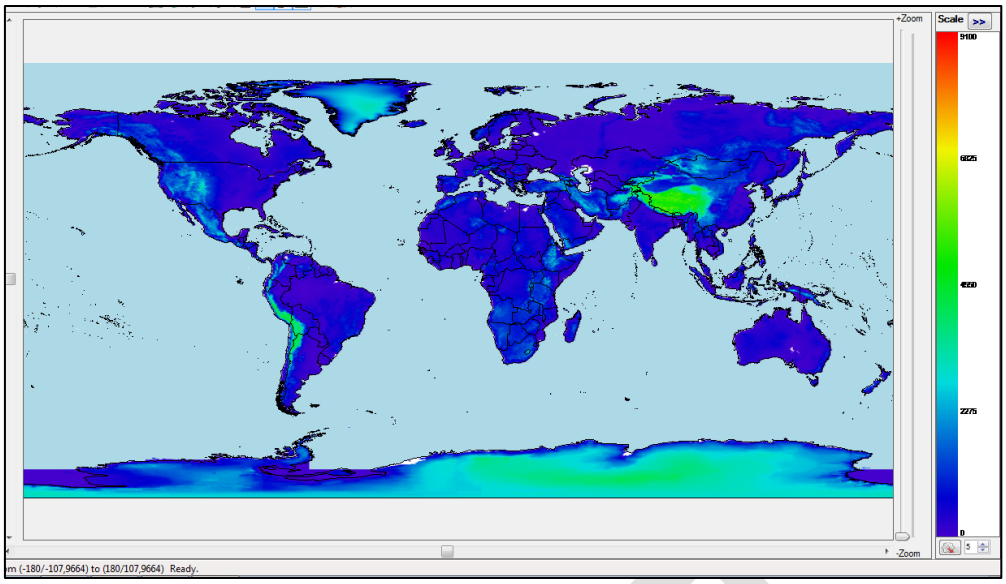
You can change lateral color scale settings corresponding to the currently displayed raster using the menu option *Layers /Rasters/Color scale settings* or the button that appears at bottom of the color scale.



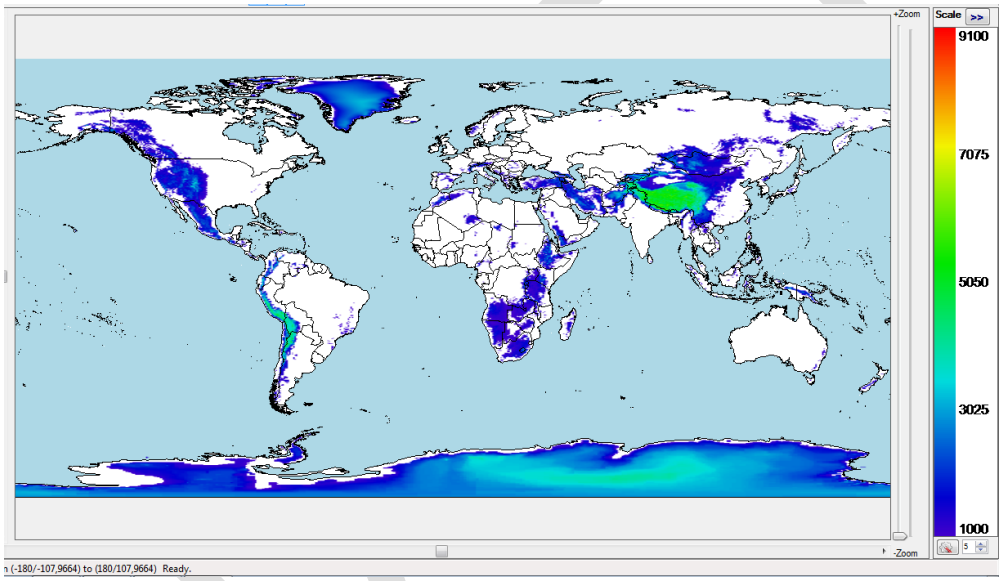

A dialog box will be shown where you can modify the number of labels to be shown on the scale, and the maximum and minimum values to be shown on the map. If you change this range, areas of the raster with a value outside the range will not be shown. This can be useful to show only areas where an environmental variable is within a specific range of values.




For example, in the following figure altimetry data are shown using default options:



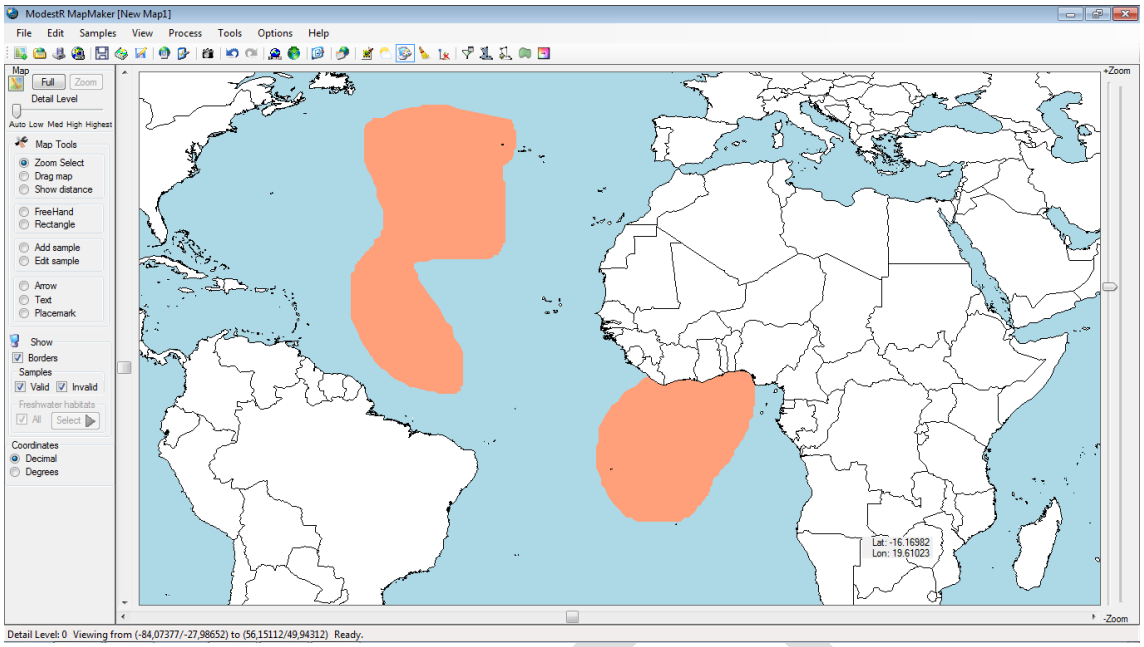

And this will be the result after setting minimum value to show to 1000:




### 2.9.4 Raster data clipping

MapMaker can cut off or clipped a raster to match a shape. This shape can be the species area of presence, or another arbitrary shape (see 2.16 for shape features). The raster to be clipped can be a raster file previously loaded (see 2.9.2) or one already integrated in ModestR. Details about how to integrate environmental data in ModestR are explained in section 6.26.

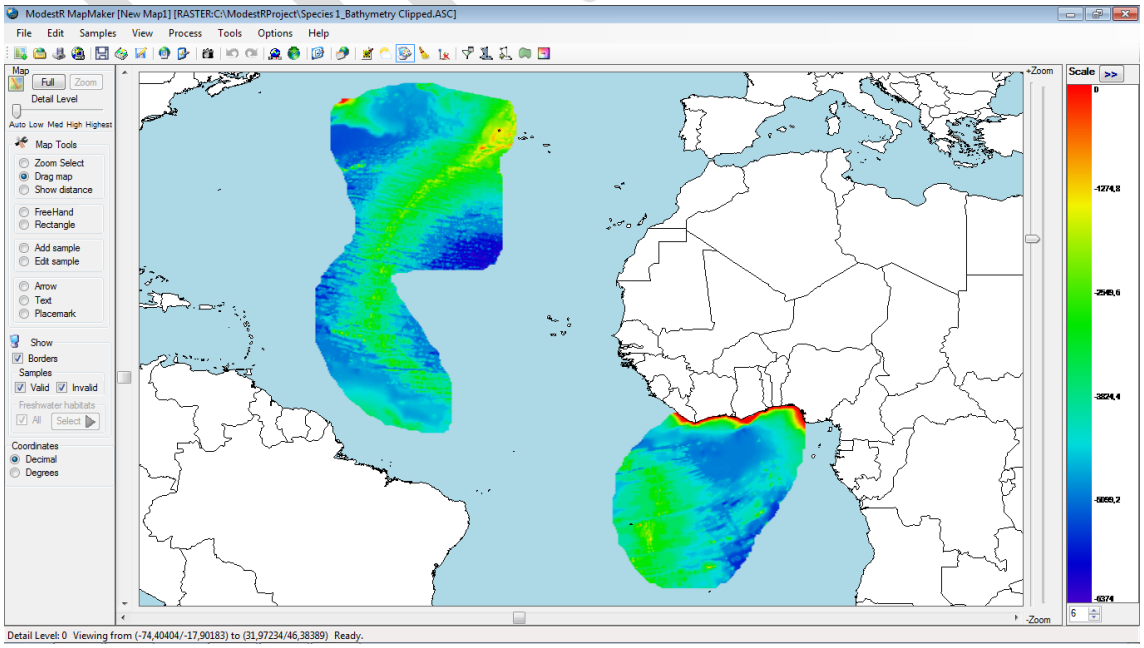
For example, let's suppose we have this map with two presence areas in the sea:




Now we want to export Bathymetry data from the corresponding raster just for those areas. We'll go to *File/Export/Environmental data/For presence areas.* A dialog box will show to you the available variables (those previously integrated in ModestR). We'll select the wanted one (it could be more than one, to generate several clipped files). Next, we'll have to selected the target folder where the file will be saved, in ESRI ASC format. Be careful because any existing file with the same name will be overwritten!

A report in CSV format is also generated with data about clipped areas in $km^2$ and simple statistical measures (max, min, mean, std and quartiles) for the environmental data in the clipped areas (see 7.9).

Here we can see the resulting raster, which only contains data for the selected areas:



The same process can be done to cut off and export raster data for shapes. We'll just have to add the wanted shapes to the map (see 2.16) and use the *File/Export/Environmental data/For each shape*.

If you previously have loaded a raster from a file (see 2.9.2) and you want to cut off it, you can go to *Layers/Rasters/Clip current raster…* and select whether to clip it to the current species distribution or to the shapes currently added to the map.

To cut off a raster following an arbitrary shape, you can also use MRFinder, as explained in section 4.12.3.

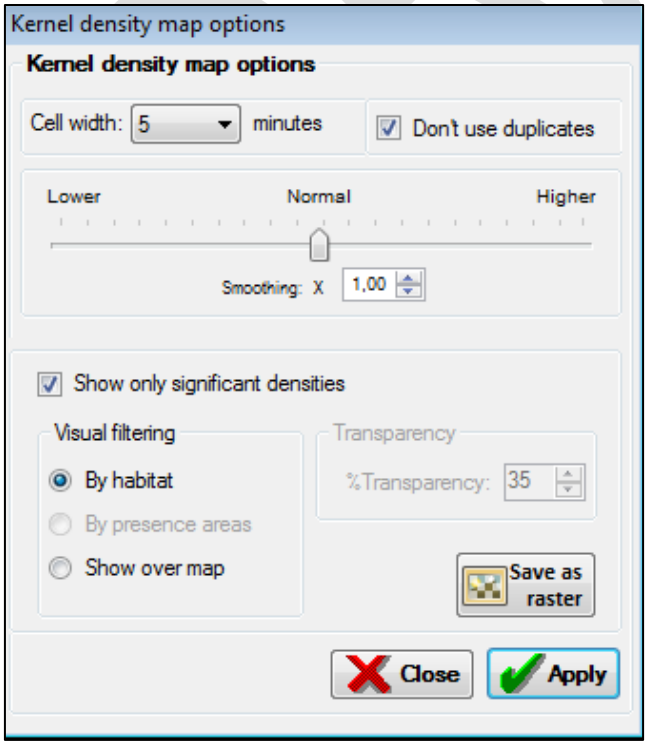
To easily cut off rasters using as shapes the corresponding EOO of a set of species, you can use DataManager, as explained in section 6.19.

### 2.9.5 Adding a kernel density map as raster

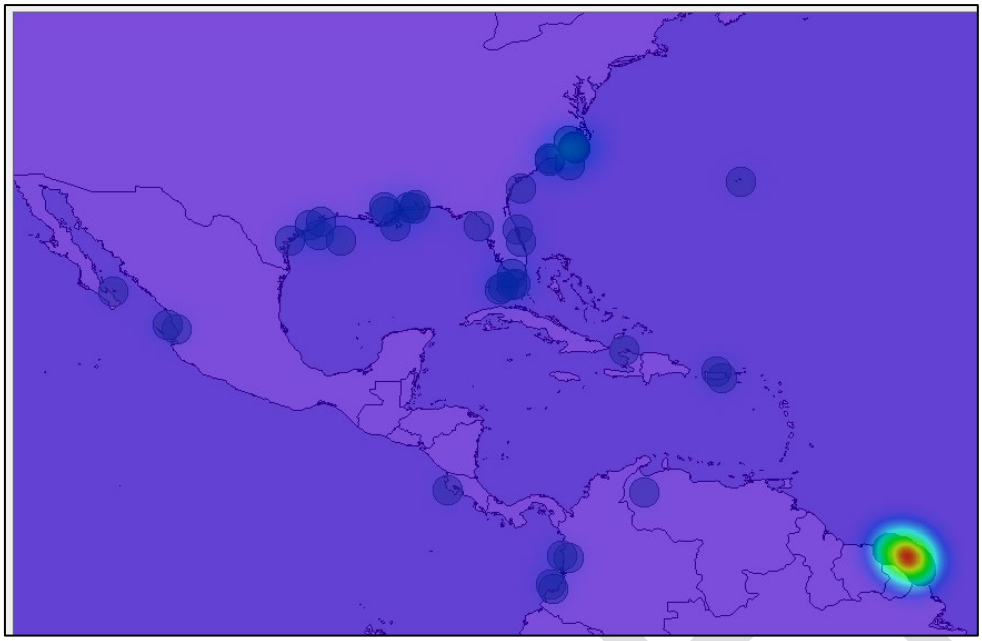
In section 2.8.3 we described how to calculate a hull using a kernel density algorithm. But you can also show a kernel density map just for informative and presentation purposes.

Even if MapMaker will allow you to use it in maps that contain areas and occurrences, this option only takes into account occurrences. It calculates density gradients from the occurrences and shows them as a density map. The shown density map is just a shape that doesn't modify the map. As it is calculated and displayed as a raster, it will clear any other raster you have previously activated.

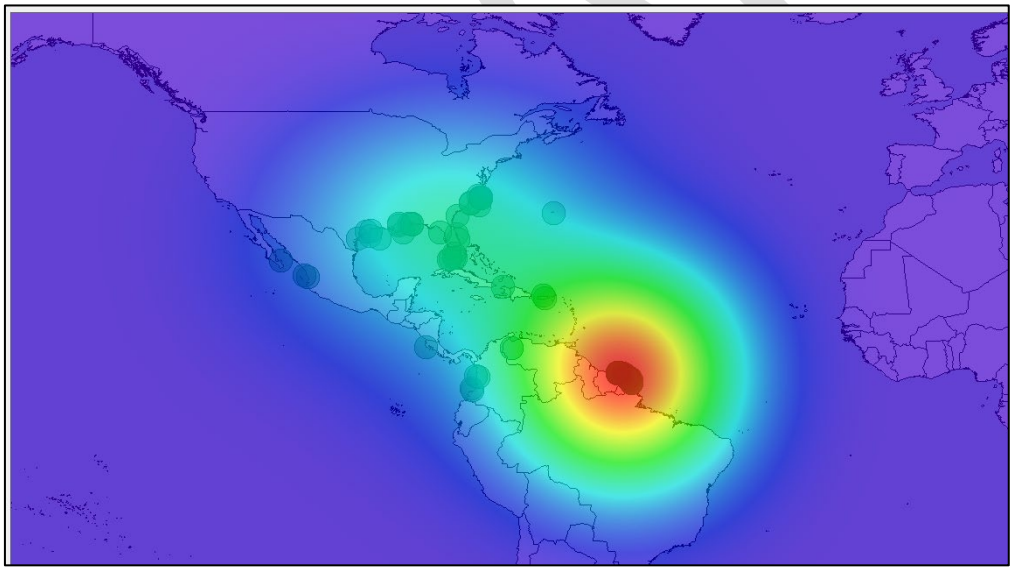
To calculate and show this map, go to *Mapping /Hull transformation/Kernel Density/Density Map*. In the dialog box, you can select the resolution (5' is a good balance between precision and calculation speed). You can also select to show only significant densities, and to filter density map by habitat, by presence area, or to show in over the map.



For example, if you use the *Show over map* option, the result can be something like this:

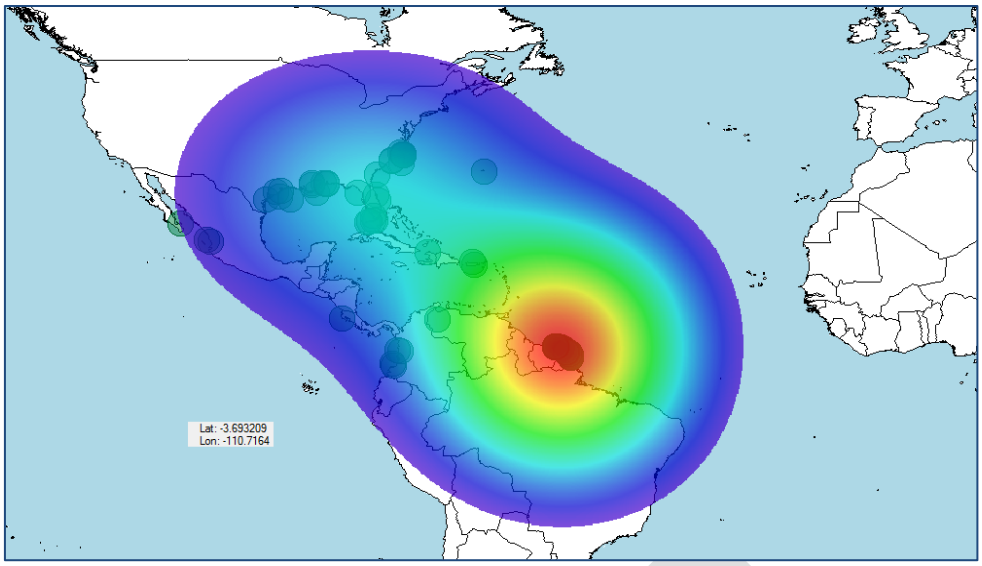


To obtain more visually appealing maps (but with more blurred density areas), you can increase smoothing. For example, using a x11 factor, the result is:

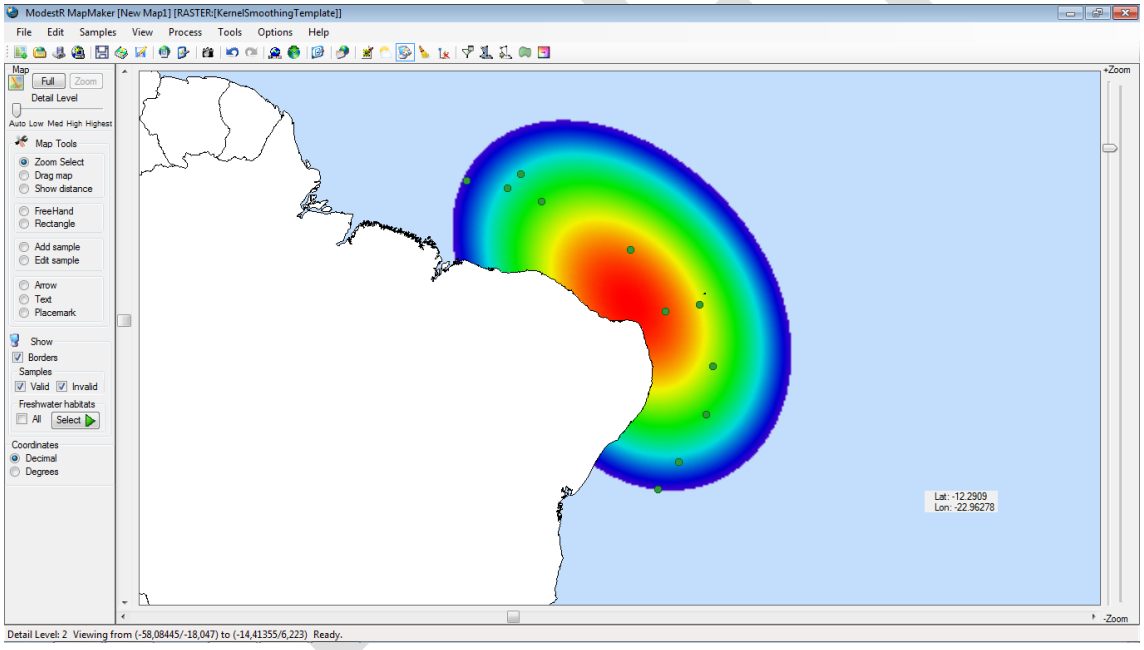


You can also modify transparency to increase/decrease visibility of the density map regarding the underlying map.

You can select to show only areas where the density is equal or greater than minimal density at an occurrence record. Areas with lower densities will not be shown. For the example map shown below and using the default parameters, the result will be:

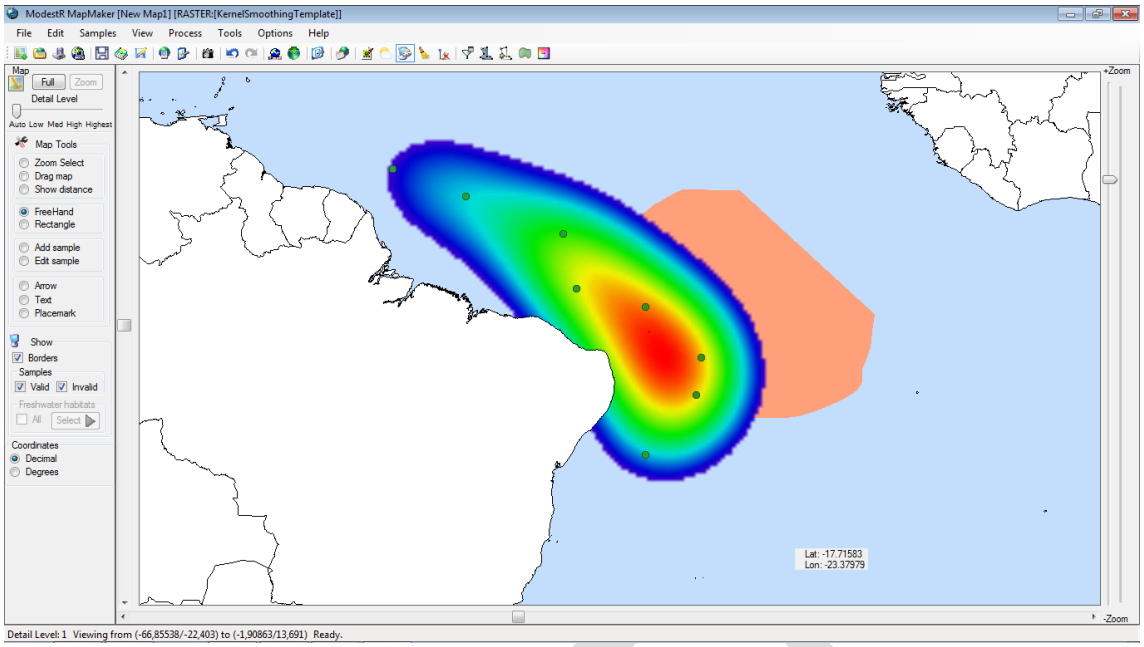



If you select the visual filtering by habitat, MapMaker will only show density map on the habitats allowed for the species. For example, for a marine species, this could be the result (a very high smoothing value has been used for visual purposes):

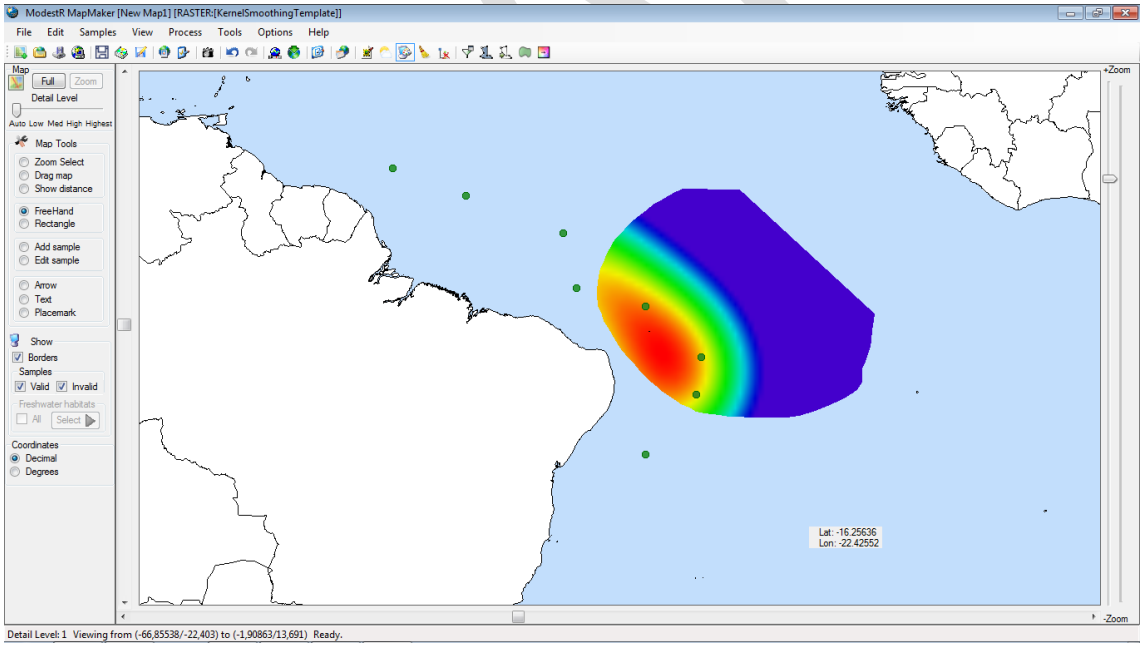



The visual filtering by habitat is well fitted for occurrence maps. But it can be less adequate for maps that also contain areas, as density map will be always shown OVER the presence areas, hiding them.

For example, if in the folowing map we have occurrences and also a presence area. Therefore the density map will be calculated using only occurrences, and shown over presence areas, hiding them totally or partially, like this:




The option of visual filtering by presence areas can be very interesting to compare density estimated from occurrences and presence areas added using hulling or Environmental Occurrence options. This option will use the presence areas as stencils to show density map *into* those areas. This makes easy to compare how well those areas fit into the estimated density from occurrences. This should be the result in the previous map:




When you want to make density maps like the explained above but for many species, in turn of doing it one by one in MapMaker, it could be more useful to do it in DataManager, that has a feature for that (see section 6.25).

## 2.10 Calculating Environmental Occurrence

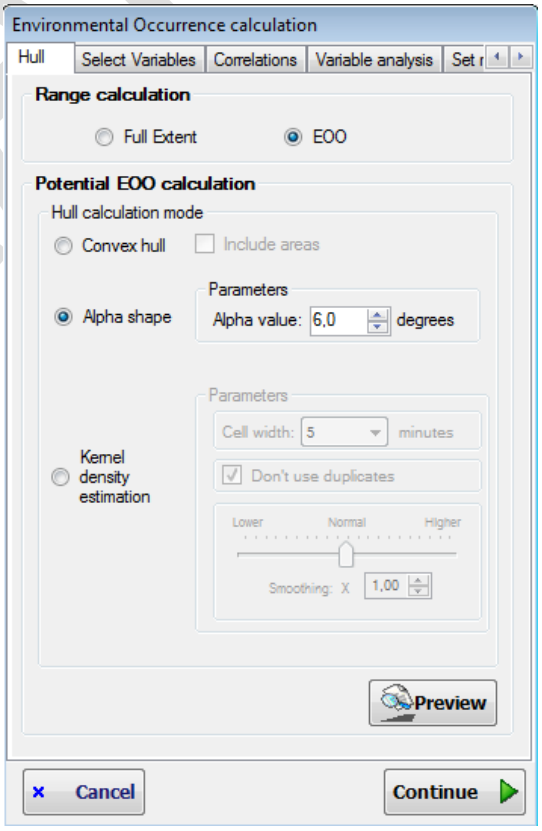
MapMaker can estimate the Environmental Occurrence[9] of a species, that can be defined as all the available areas for the species within a specific range (typically the EOO, but it can also be all the world), but excluding the habitats not occupied by the species and considering only those areas within the range of environmental conditions suitable for the presence of the species. This option is intended to be applied on occurrence-based maps, as range maps already include presence areas, but it can be also applied to those type of maps, with some restrictions. Environmental data needed to use this features are not included in ModestR, so they have to be provided by the user as data files in ESRI ASCII or CSV format. There exist free datasets available in Internet, such as Bi-Oracle (Tyberghein et al. 2012) for example that are fully compatible with ModestR.

Therefore, the first step will be to integrate those data in ModestR, to allow it to use them when needed. To do that, you can use the menu *Layers/Manage environmental variables*, either in MapMaker or in DataManager. Details about how to integrate environmental data in ModestR are explained in section 6.26.

To apply Environmental Occurrence to a distribution map, you should go to the *Mapping/ Environmental Occurrence* menu option. ModestR will guide you across several steps to perform this task.

The first step is selecting the range calculation for the environmental occurrence and the EOO mode. There are two options for the range:

- Full extent: the environmental occurrence will be calculated for all available environmental data. For example, if you use environmental variables that contain data for all the world, the environmental occurrence will be calculated for all the world. If you use environmental variables that contain data for a single country, the environmental occurrence will be calculated for this country.
- EOO (Extent of occurrence): environmental occurrence will be calculated only for the area contained in the EOO of the species. This EOO will be calculated according to the settings you have to enter in this same step.



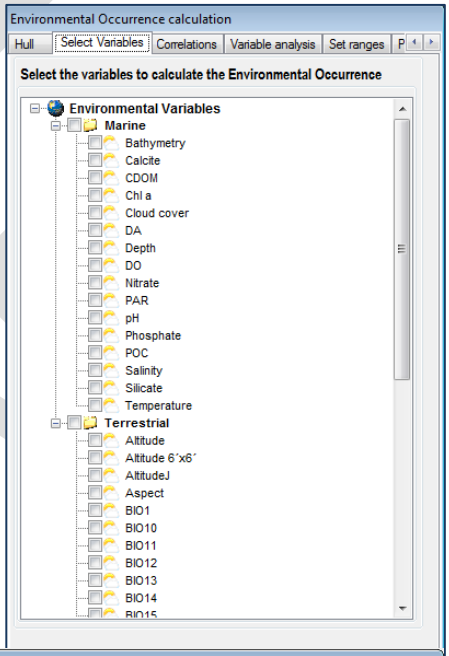

[9] Environmental Occurrence was formerly called « Niche of Occurrence » in previous versions of ModestR. But from version 2.1, "Niche of Occurrence " names a different feature explained later in this manual.

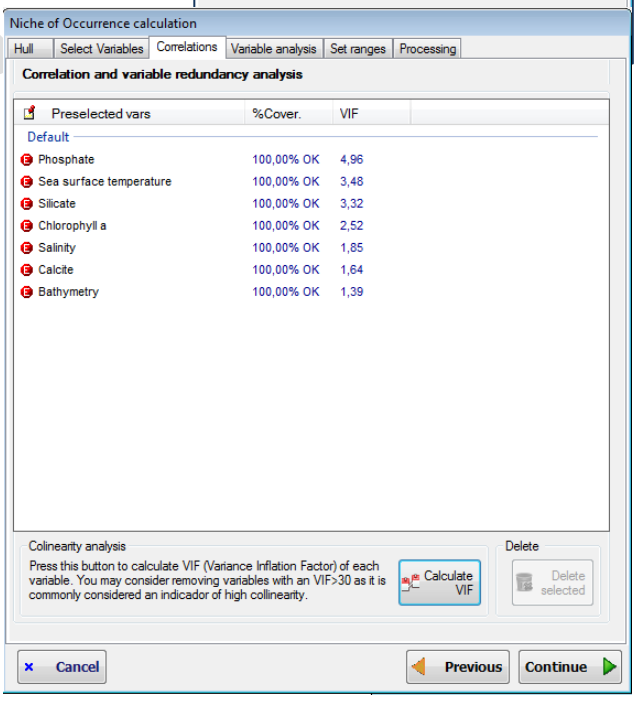
Regarding the way the EOO will be estimated, you can select any of the methods already available to calculate a hull on a distribution map (see section 2.8 before). If EOO hull cannot be calculated using the selected option, you cannot continue to the next step. This can happens if for example there are too few occurrences in a map (see section 2.8 before for more details about each hull method).

Even if you select "Full extent" range, you have to select a EOO mode, because those settings will also be used for variable contribution analysis.

Next step is selecting the environmental variables that will be used to calculate Environmental Occurrence. Of course, in the ideal case, you should select the variables that affect the most the species distribution. Anyway, you can always select most variables, and ModestR will help you to choose the best ones in the next steps.

Next step allows you to optionally calculate Variance Inflation Factor (VIF). A VIF value higher than 30 is commonly considered an indicator of collinearity among the variables. Thus, you have the change choice of removing those variables significantly related and, therefore, that are redundant. But this is an optional step. If you just want to include all variables, you can simply continue. Take into account that if two variables are very correlated, the way that one of them if a linear function of the other one, VIF cannot be calculated. Involved variables will be marked to allow you to remove one of them if you want.



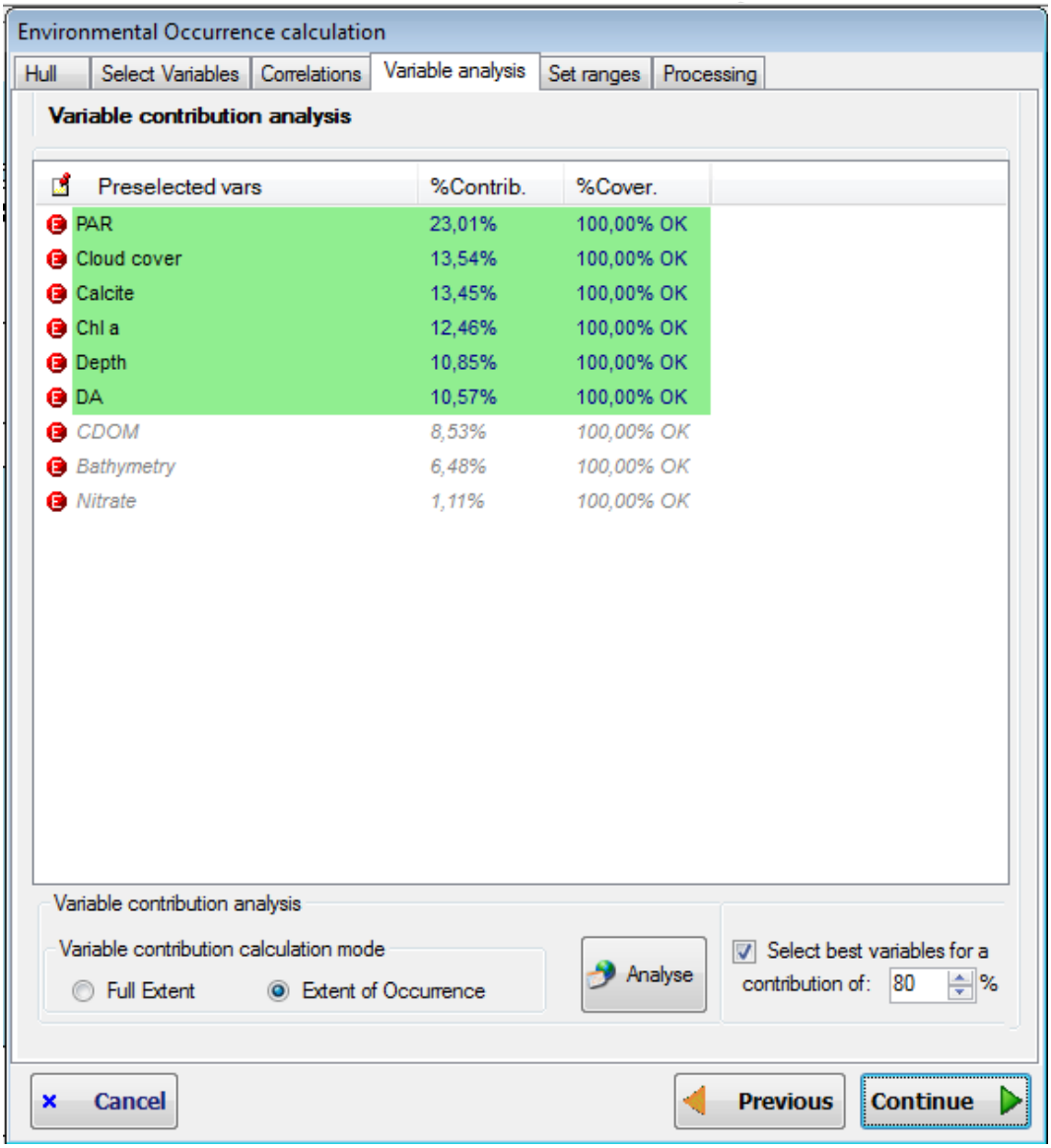




As you could see, when calculating VIF, you can also see the coverage percent of each environmental variable into the EOO. That is, for how much area of the estimated EOO there are values for each environmental variable. This can also allow you to remove inadequate variables, that have a very low coverage in the species EOO. For example, a variable such as Bathymetry will have a quite null coverage for a terrestrial species.

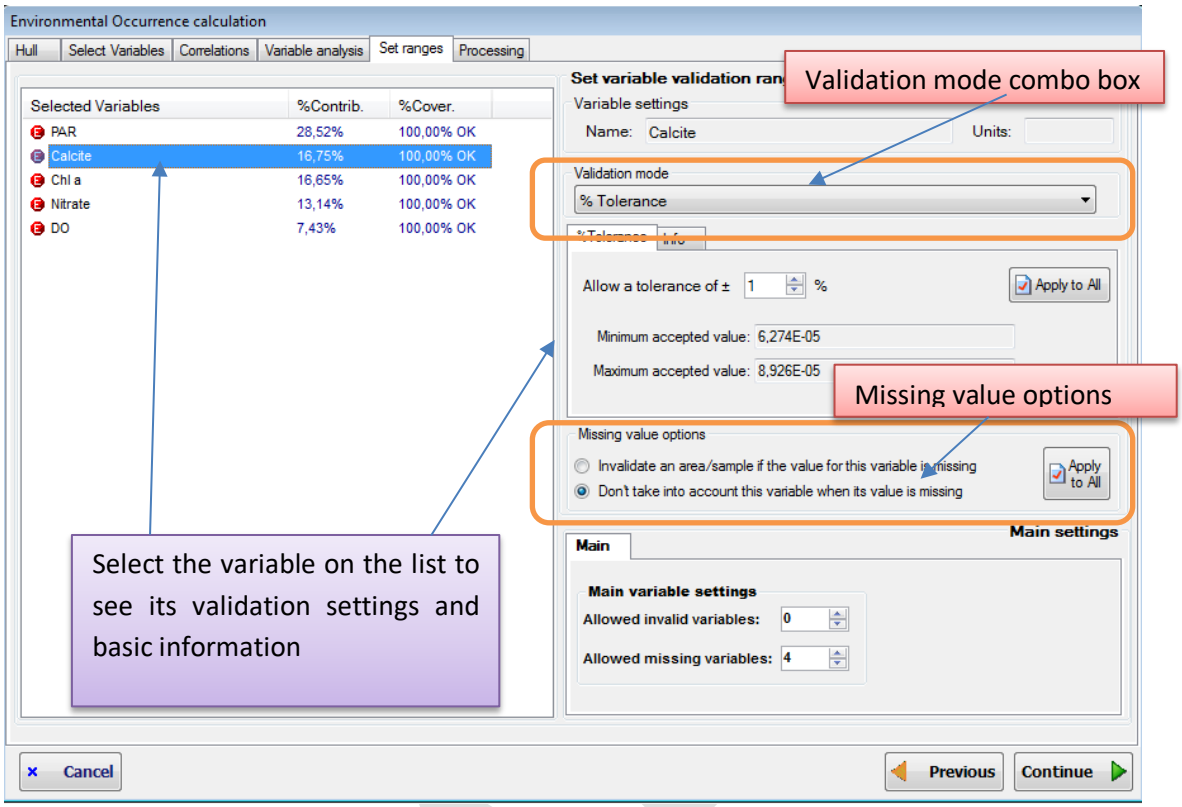
In the next step you can perform a statistical analysis for selecting those variables that are more affecting the most the distribution of the species. The environmental variables selected can be checked either for the full extent where environmental data are available or just within the EOO of the species (the EOO

is calculated using the settings you entered in an earlier step). Then an index of instability is calculated using the fluctuation index of Dubois (1973) modified by Guisande et al. (2006). Once the instability index is calculated is possible to estimate the percentage of contribution of each variable to the index. Those environmental variables with a higher percentage contribution to the instability index can be considered as those more affecting the most the distribution of the species. In order to include only the variables with higher contribution, in ModestR is possible to select a threshold percentage with the accumulative percentages of the variables. But this is an optional step. If you just want to include all variables, you can simply continue.

Same as when calculating VIF, you can also see the coverage percent of each environmental variable into the EOO.




In the next step, ModestR will first determine the range of values took by each one of the selected variables in the estimated EOO. Before continuing, it is interesting to point out that many of the options and settings that are displayed on this dialog box are autoexplained: just by placing the mouse cursor on one element, usually a short emerging explanation will be shown.




Then you can select how validation will be done for each variable. To do that you must select a variable on the left listbox. The corresponding settings for this variables will be shown on the right-up panel. The available settings are:

- **Validation mode:** this is the way the values for this variable on the presence zones (whether occurrences or areas) will be evaluated to determine if some occurrence/area has to be removed from estimated EOO or not. There are two validation modes:
  - ***% Tolerance*:** the default validation mode is *% Tolerance*. It consists of using the range (minimum and maximum) of valid values for the variable, increased by a tolerance percent (by default 1%, but you can modify it), so any value that fall outside this range will be considered invalid and the corresponding area removed from the EOO.
  - ***Custom***: in this case it is the user that will manually set the minimum and maximum values of the range of valid values for a variable. Any value that fall outside this range will be considered invalid.
- **Info tab:** under the validation mode combo box, a tabbed b ere information and options for the currently selected validation mode ere is also a *Info* tab where you can consult basic statistics for the curren ble, such as maximum and minimum values that this variable has on the current presence areas/occurrences, mean value, quartiles, and jackknife ranges.
- **Missing value options:** under the tabbed box that show the information for the variable, two radiobuttons allow you to select what to do when there is no value for the variable in a particular area/occurrence. You can select between two possibilities:
  - Invalidate an area/occurrence if the value for the variable is missing (that is, if the variable have a null value in this location): in this case, an area/occurrence

will be cleaned (considered invalid) as soon as there is no value for the variable for this area/occurrence, no matter the values for the other variables.

- Don't take into account this variable when its value is missing: in this case, if there is no value for the variable for this area/occurrence, this variable will be just not considered to make a decision. ModestR will check the values took by the other variables to decide about the validity of the area/occurrence.

Take into account that those options are *per variable*. That is, you can set different options for each variable. A missing value for a variable can lead to immediately invalidate an area/occurrence, while for another variable it will not be took into account. You should decide in a per-variable basis the most adequate option.

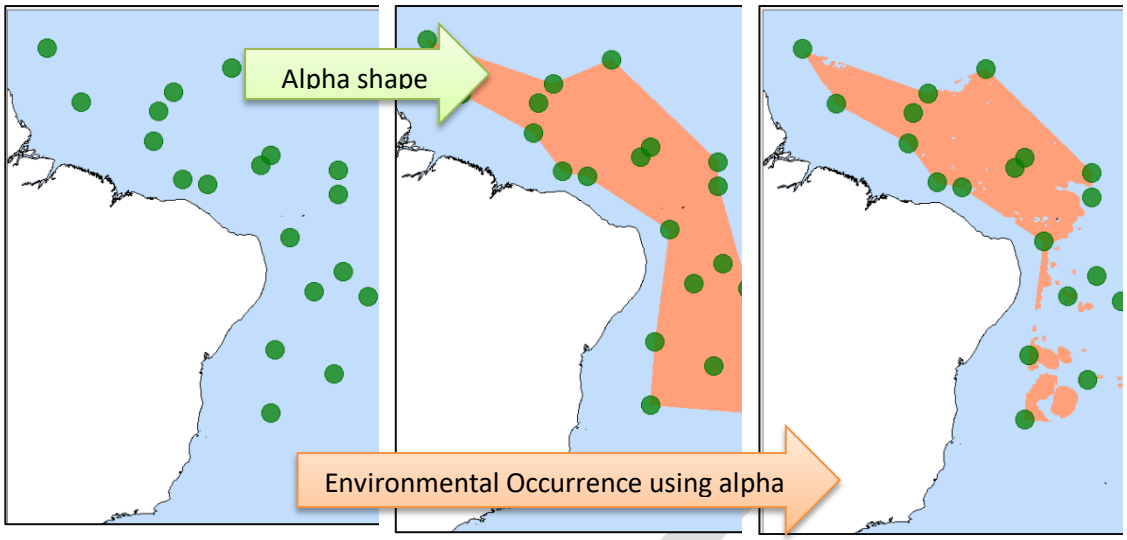
- **Apply to all buttons:** either in the tabbed box that displays the options for the currently selected validation mode or in the missing value options, you can see a button *Apply to All*. This button can be used to replicate the corresponding settings for the current variable to all variables in an easy way.

Finally, on the right-bottom side of the dialog box, you can see a tabbed box with main settings. While settings described above are for each variable, those main settings have global effects. They are:

- **Allowed invalid variables:** Indicates how many variables can take invalid values, that is, outliers (according to their validation rule) before considering an area or occurrence as invalid for the species. The default value is 0, and the maximum is N-1 where N is the number of selected variables. Usually this field will be 0, because typically you will want to clean a occurrence/areas as soon as one of the selected variables takes an outlier value. But this setting provides a way to be more flexible. For example you can select 10 variables and allow that up to two variables of these take outlier values before considering a occurrence/area as invalid.
- **Allowed missing variables:** Indicates how many variables can be missing (have a null value) before considering an area or occurrence as invalid for the species. By default it will be set to N where N is the number of selected variables. That is, if all variables have null values for a occurrence/area, the area will be considered as valid. Take into account that this parameter will be applied only if the variables with missing values are not individually configured as directly invalidating an area when their value is missing (as explained above). That is, this rule will be applied only if after applying the rules for each individual variable the occurrence/area has not already been considered as invalid.

Once all those settings configured, press the *Continue* button to perform Environmental Occurrence calculation. Once ended, the resulting estimated presence areas will be added to the distribution map.

In the figures below you can see an example of applying Environmental Occurrence to an occurrence map, compared to just the EOO using the alpha shape, for example. As you can see, here Environmental Occurrence used the EOO estimated by the alpha shape but removed several areas that didn’t complied with the environmental conditions estimated for the species from its initial presence data.




## 2.11 Creating 2D compounded environmental layers (2DCEL's)

A two-dimensional compounded environmental layer (from now 2DCEL) in ModestR is a data layer somewhat similar to a raster. But it should not be confused with raster environmental data or environmental variables, which can be imported from raster files (see section 2.9.). While a raster usually contains information about a single variable (e.g. altitude, temperature…), ModestR 2DCEL's are a compound of several variables, using an approach basing on a polar coordinates system. In a large sense, this can be compared to a sort of multidimensional scaling approach, where multiple variables (that can be somewhat considered as "dimensions") are "scaled" in a two-dimensional polar coordinates system. 2DCEL's can have two applications in ModestR:

- To visualize the different environments that are present, according to a set of environmental variables, in a single graph.
- To calculate the niche of occurrence of a species. In this process, the areas where a species can potentially be present are determined basing on their proximity to the environmental conditions where the species is present in the polar coordinates system.

You can create CEL's for the entire world or for a selected area. Environmental data needed to use this features are not included in ModestR, so they have to be provided by the user as raster files in ESRI ASCII, GeoTiff, netCDF or CSV format. There exist free datasets available in Internet, such as Bi-Oracle (Tyberghein et al. 2012) for example that are fully compatible with ModestR.

Therefore, the first step will be to integrate those data in ModestR, to allow it to use them when needed. To do that, you can use the menu *Layers/Manage environmental variables*, either in MapMaker or in DataManager. Details about how to integrate environmental data in ModestR are explained in section 6.26.

To create a new 2DCEL, go to the menu *Mapping/Niche of occurrence/Create new compounded environmental layer (CEL)/2DCEL.* The following dialog box will be shown, where you have to select at least two environmental variables to be used to generate the 2DCEL (details about how to integrate environmental data in ModestR are explained in section 6.26.).

A relevant consideration is that is highly recommended to use variables with the same cell size (5 minutes, for example) when creating a 2DCEL. Environmental variables with different cell sizes are supported, but this can lead to less accurate results. Anyway, the 2DCEL will always be calculated using the lowest cell size among the used variables, and all variables will be previously downscaled to this cell size.

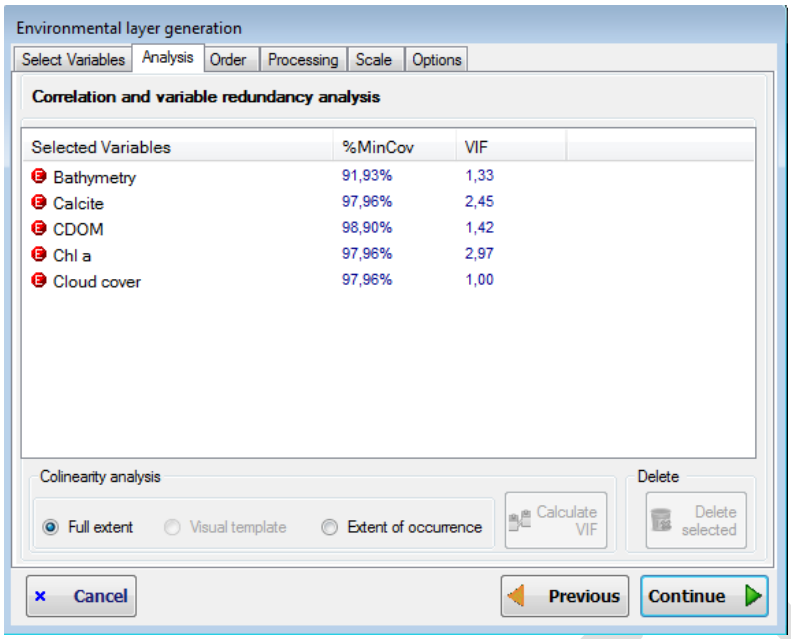
In this dialog box you can also select the *Geographical extent*, which can be:

- Full extent: the CEL will be calculated for the area for which selected environmental data is available. For example, if you selected environmental variables that cover the entire world, the 2DCEL to be created will also cover the entire world.
- Shape: this option will only be available if you previously added a shape to the map, before starting the 2DCEL creation process. Briefly, a shape is an area that is added to the map just for visual purposes (see section 2.16 for more details about shapes). For example, if you added a shape that demarcates a specific country, and you select the "Shape" layer extent mode in this dialog box, the 2DCEL to be created will only covers this specific country (supposing that environmental data are also available for this area, of course).
- Extent of Occurrence: this option will only be available if there exists occurrence data in the map, before starting the 2DCEL creation process. In this case, you can select the "Extent of Occurrence" layer extent mode in this dialog box, to delimitate the 2DCEL to be created to the Extent of Occurrence (EOO) that can be calculated from the current occurrence data in the map. If you select this option, a panel will appear on the right side for you to select how EOO should be calculated (see section 2.8 for more details about the different ways of calculating EOO).

Once those options set, click on *Continue* button to go to next step.

In the next step, selected variables are shown in a list. You can optionally calculate Variance Inflation Factor, by clicking on the *Calculate VIF* button. But this is just optional. You can skip this step and click on *Continue* button to go to next step.

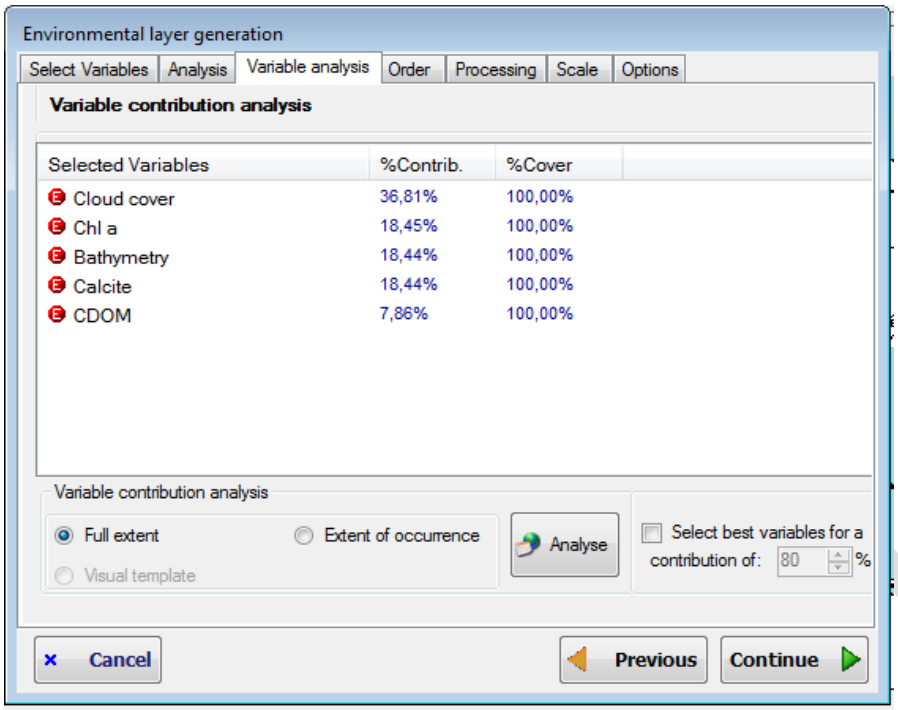
In the case you click on the *Calculate VIF* button, the VIF of each variable will be shown in the list. VIF can be useful to detect high correlated variables, and then decide to remove some variables. To remove a variable, just select in on the list and use the *Delete selected* button.




Similarly to the *Layer extent* option described in the previous step, VIF can be calculated for the full extent, for a shape or for the EOO (given that a shape or occurrence data are present on the map, as explained before). Take into account that extent used to calculate VIF can be different from the extent selected for the 2DCEL. ModestR doesn't establish any limitation in this sense, so the convenience of using one or another option is left to the user. Click on *Continue* button to go to next step.

If occurrence data are present on the map when you started to create the 2DCEL, the next step will allow you to perform a contribution analysis. In other case, this step will be automatically skipped. If occurrence data are present on the map, you can optionally calculate the contribution of each variable to the presence of the species, by clicking on the *Analyse* button. This is just optional. You can skip this step and click on *Continue* button to go to next step.

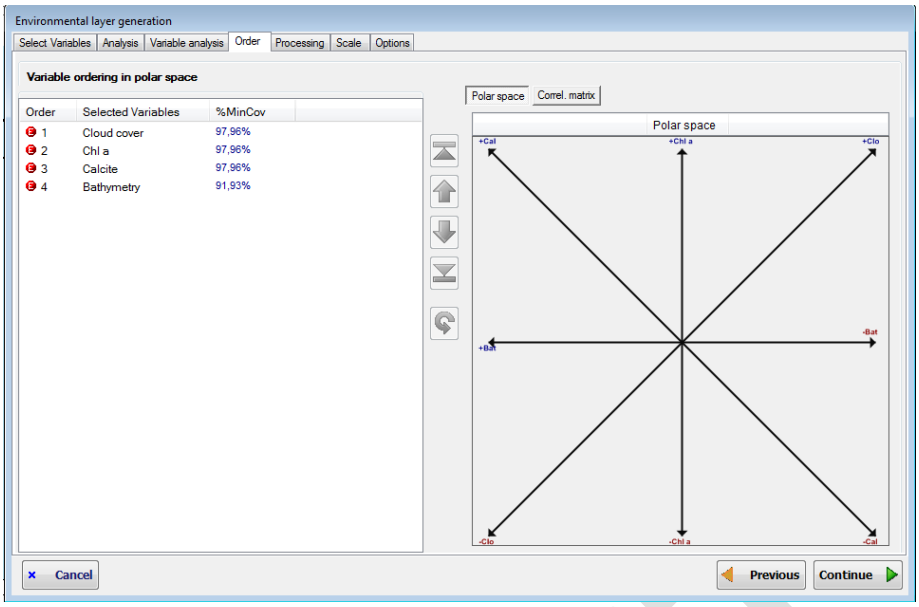
Even if the resulting 2DCEL can be used as any other layer, the aim of this option is to create a 2DCEL using the variables more relevant for a specific species, possibly to posteriorly use it to calculate Niche of Occurrence of this particular species.



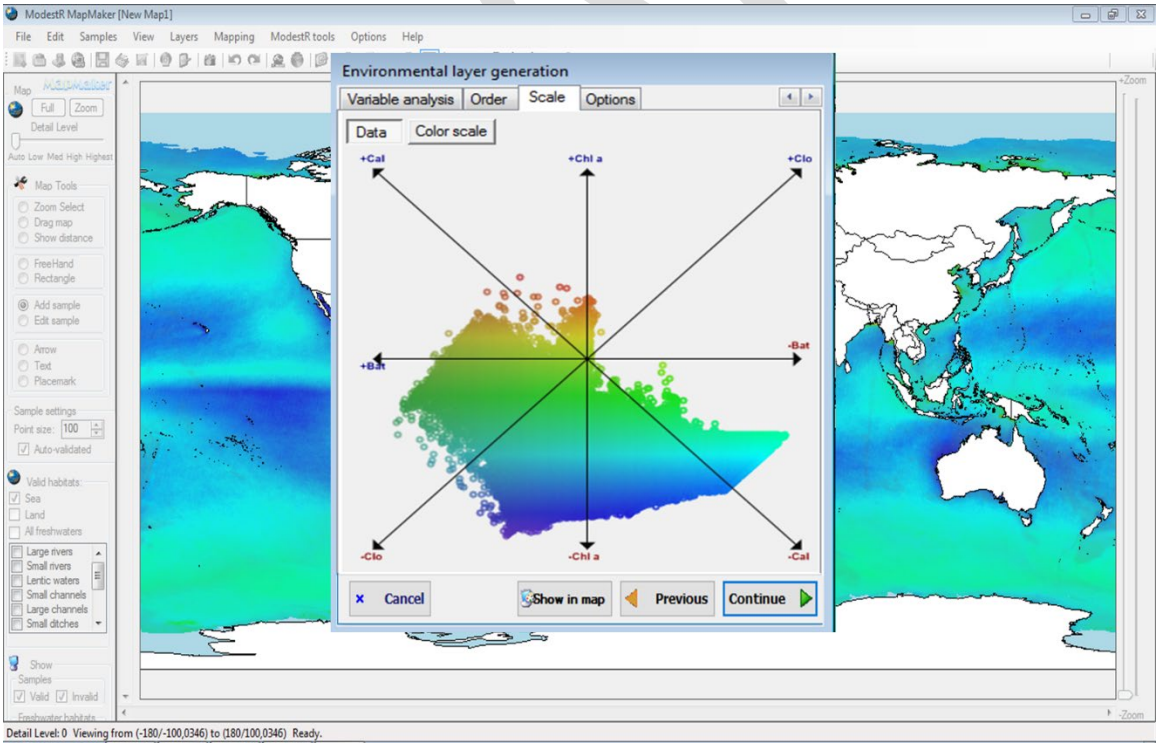

In the case you click on the *Analyse* button, the contribution of each variable will be shown in the list. Contributions can be useful to determine which variables seems to be more influential on the species presence/absence (see section 2.10 for more details). Once contributions calculated, you can check the option *Select best variables for a contribution of XX%*, for MapMaker to select automatically the variables with the higher contribution to reach the indicated % (that you can modify). Selected variables will be highlighted in the list.

Similarly to the *Layer extent* option described in the previous step, variable contribution can be calculated for the full extent, for a shape or for the EOO (given that a shape or occurrence data are present on the map, as explained before). Take into account that extent used to calculate variable contribution could be different from the extent selected for the 2DCEL. ModestR does not establish any limitation in this sense, so the convenience of using one or another option is left to the user.

Click on *Continue* button to go to next step. In this step, definitely selected variables are shown on the list, and on the right panel a representation of how they will be ordered to calculate polar coordinated is displayed. Variables are automatically ordered according to their relative correlations: the next variable to another is the one which as the higher correlation with it. You can eventually manually change order by selecting a variable in the list and using the buttons located on the right of the list.

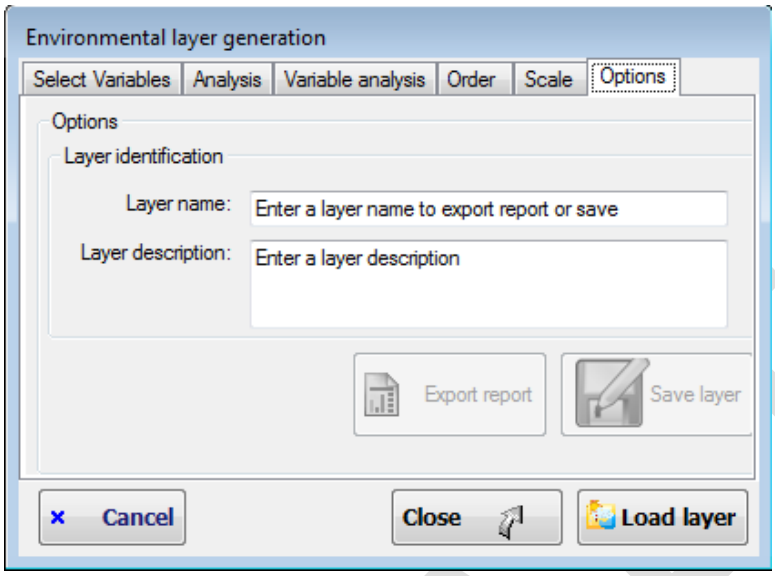



Click on *Continue* button to go to next step. In this step, 2DCEL will be calculated according to entered settings. This can take some time. The distribution of the environmental data in the polar coordinates system will be shown in the dialog box. The resulting layer can be also shown on the map using the *Show in map* button, as in the figure here below. But remember that is a 2DCEL, it is independent from any species data; it'll be in a subsequent process of NOO that this layer can be used to estimate a species distribution.




Click on *Continue* button to go to next step. The final step allows you to save and/or to export a report of the generated layer. In both cases, you should enter a layer name before. Then you can export a report to a folder, and/or save the layer to be able to use it later.

Exported report will contain several files such as an ASC raster for the X and for the Y components, and a graph of the polar coordinates with the data distribution (see 7.11 for details).




Finally, you can close this dialog box, or load the layer in the current work session (even if you not saved it), to be able to use it just during the current work session.

As pointed out before, it is highly recommended to use environmental variables with the same cell size (5 minutes, for example) when creating a 2DCEL to obtain more accurate results. Another important issue is that 2DCEL's do not store environmental data that was used to generate it, they just store a link to those data. So it is important that you don't delete neither modify environmental variables used to generate a layer, because this will make environmental layer unusable.

## 2.12 Calculating 2D niche of occurrence (NOO2D)

MapMaker can estimate the Niche Of Occurrence[10] of a species (NOO2D), that can be defined as the set of areas which are located "near" the areas where the species is already present (e.g. where there are occurrences) in the polar coordinates system generated for a specific 2DCEL. Here the 2DCEL refers to a polar coordinates system generated from the combination of values from several environmental variables, as explained in section 2.11 above. And "near" should be understood assuming a probability density-based criterion: a kernel density estimation is calculated for the polar coordinates system, using the presence of the species. The lowest density where the species is already present (e.g. where there is a occurrences) is used as cutoff. Therefore, all the areas with a density above this cutoff value will be considered as being part of the Niche Of Occurrence of the species. In other words, the Niche Of Occurrence of the species is composed of all the areas whose environmental conditions are "near" to those where the

[10] Environmental Occurrence was formerly called « Niche of Occurrence » in previous versions of ModestR. But from version 2.1, "Niche of Occurrence" names a different feature based on environmental layers described in section 2.11.

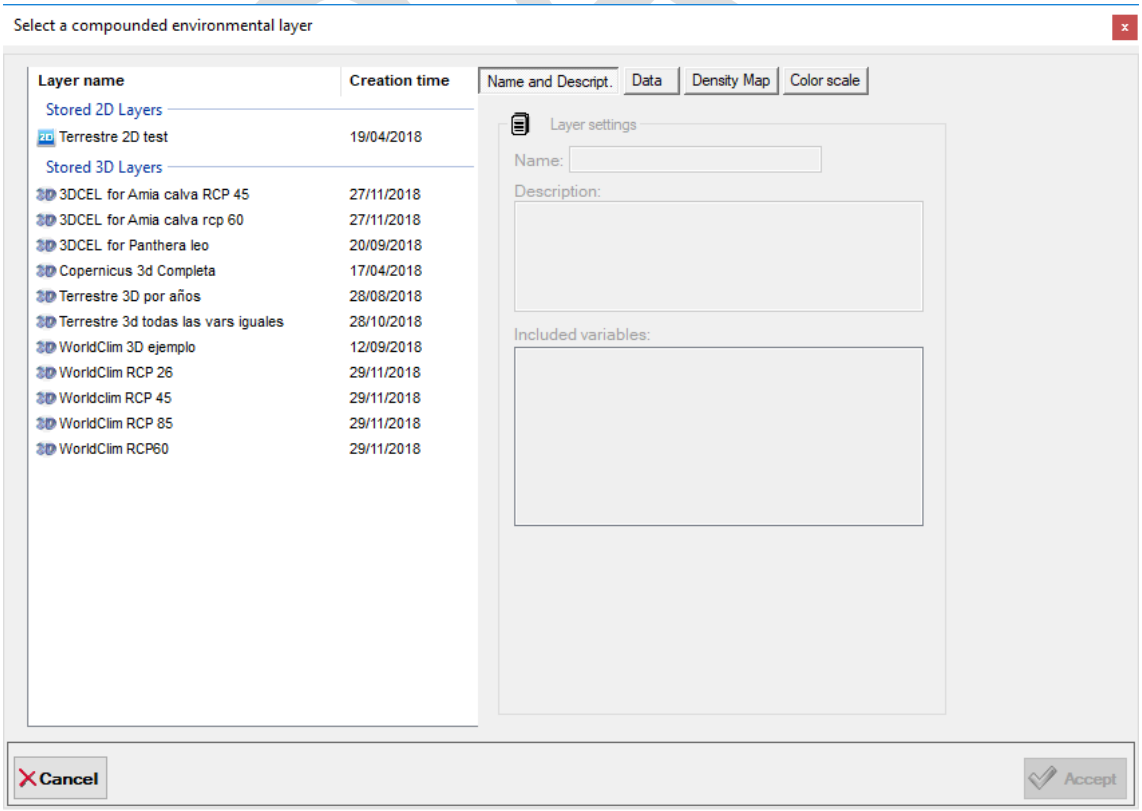
species is already present in a polar coordinates system (2DCEL) defined using a set of environmental variables.

This option is intended to be applied on occurrence-based maps, as range maps already include presence areas. It also requires that you have a 2DCEL already created, as explained in section 2.11. You can use a previously stored 2DCEL, for example. It is very important to point out that, as seen before, a 2DCEL is built upon several environmental variables. And to perform niche of occurrence, those original environmental variables are needed, besides the 2DCEL itself. Consequently, to obtain accurate results, those original environmental variables must still exist, with the same names and data than when the 2DCEL was created. Errors or inaccurate results can be obtained if some variable has been modified since the 2DCEL creation. To avoid those problems as much as possible, from ModestR v.3.1 when you delete or modify an environmental variable, any 2DCEL that used this variable will be deleted.

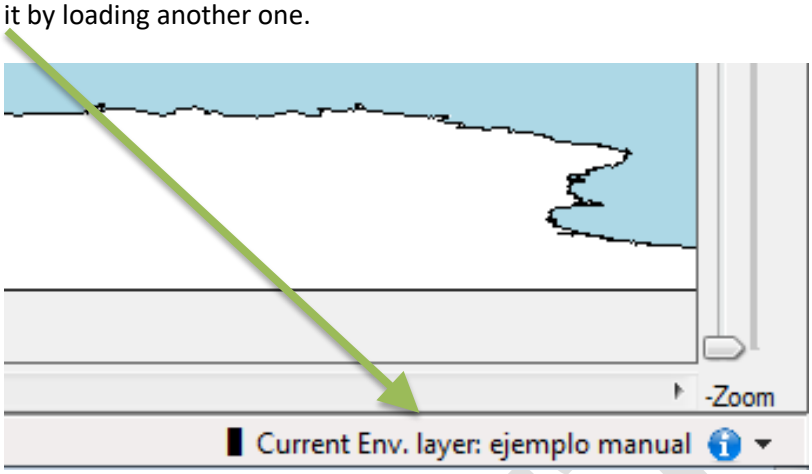
The first step to calculate the NOO2D of a species in MapMaker is loading the wanted species data, for example downloading data from GBIF, or from an existing ModestR database.

Next, you have to select the 2DCEL that will be used to calculate the NOO2D. This determines the polar coordinates system and the environmental data to be used. To select a 2DCEL you have several possibilities:

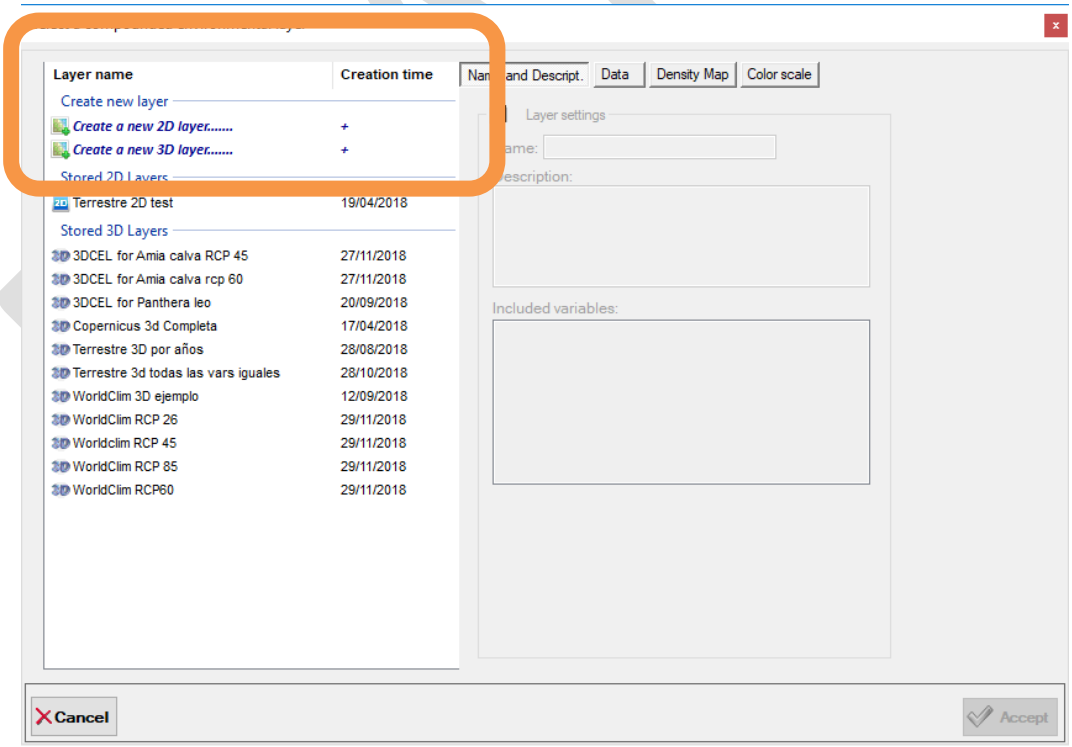
- Go to menu option *Mapping/Niche of occurrence/Load compounded environmental layer*. Select the wanted CEL in the dialog box that will appear, and click on *Accept* button.



Once loaded, a CEL can be used during all the current work session. The currently loaded CEL is displayed in the right-bottom corner of the MapMaker window. You can change it by loading another one.




- Directly go to *Mapping/Niche of occurrence* and select between *Density Map* or *Distribution map*, according to the map you want to obtain. The dialog box to choose the NOO2D will be shown, allowing you to select a stored CEL or to create a new one. In this last case, the steps will be those explained in section 2.11.

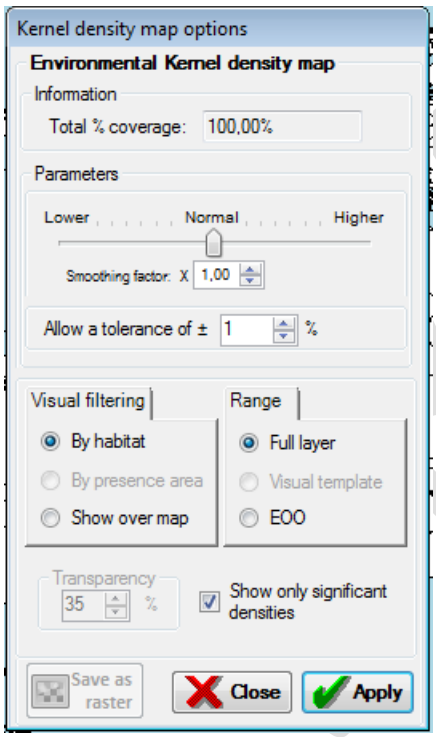


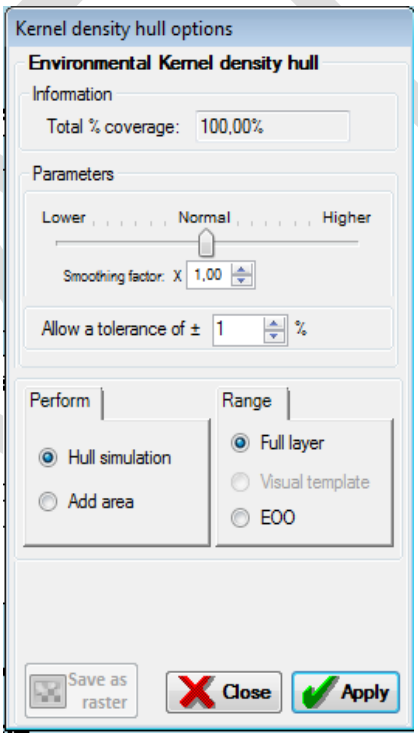

Once selected or created the 2DCEL to be used to calculate NOO2D, you can go to menu *Mapping/Niche of occurrence*[11] and select between:

[11] Excepted if you already started the process by going to *Mapping/Niche of occurrence/Density Map* or *Distribution map*, of course.

- *Density map*: this option calculates and show the density map corresponding to environmental conditions for the species. In this map you can see which areas are "near" (that have high density) to the environmental conditions where the species is already present (using the selected 2DCEL as reference), and those which are "far" (that have low density). This option does not modify the map, it's just for visualization purposes.
- *Distribution map*: this option calculates the NOO2D for the species using the selected 2DCEL as reference. In terms of calculations, it's very similar to the density map. The difference is that in turn of showing a density map, the areas that have a density above the cutoff value (described before in this section) are added to the map as presence areas. So the species map is modified, as NOO2D is added.

In both cases, the displayed dialogs are quite the same. As you can see, the main difference is just in the display options. For density map, you can set how density map will be show ('*Visual filtering*' options), while for distribution map you can select to show a simulation, or to add NOO2D to the map as presence areas ('*Perform*' options).



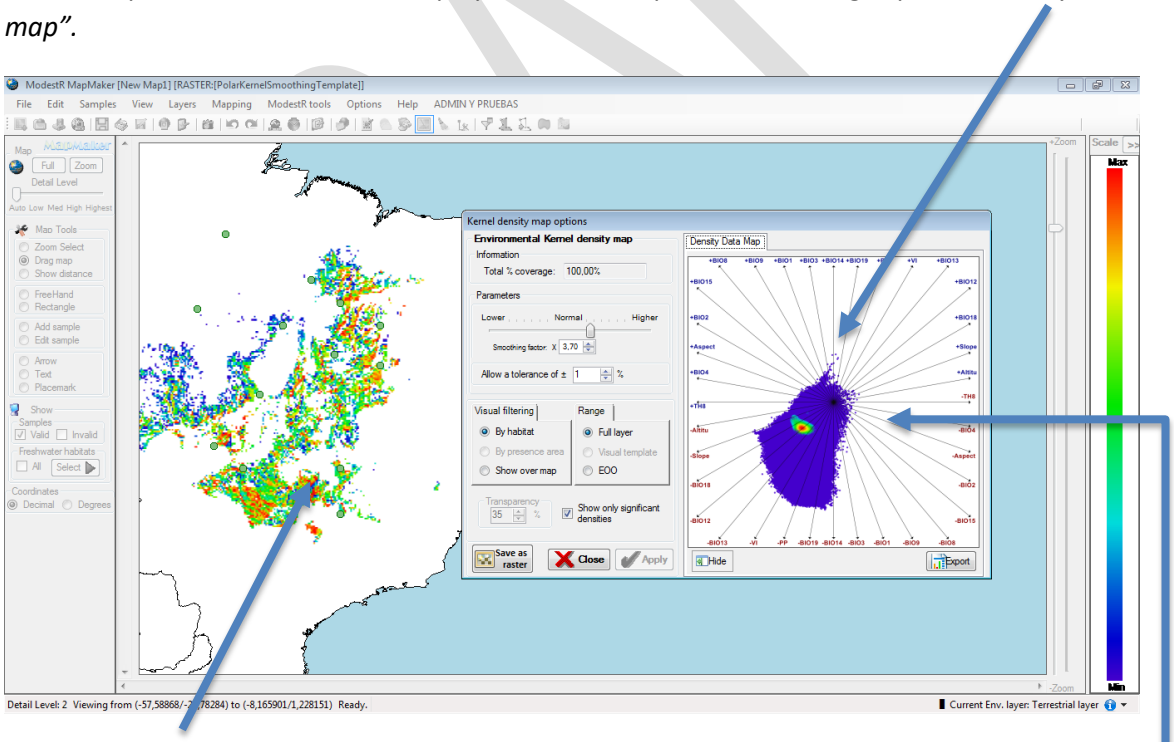




Besides those settings, in both displayed dialogs we can see and eventually modify the following settings:

- Total % coverage: This value indicates the % of presence data that are included in currently selected 2DCEL and that could be used to calculate kernel density. A value too low indicates that few presence data can be used, so results could be poorly significant. A low value can eventually appear if you are using a 2DCEL that does not cover all the world, and/or current presence data are out of the area covered by the 2DCEL. Or for example if you are using a 2DCEL generated with environmental variables for the sea areas, and presence data are on land areas.

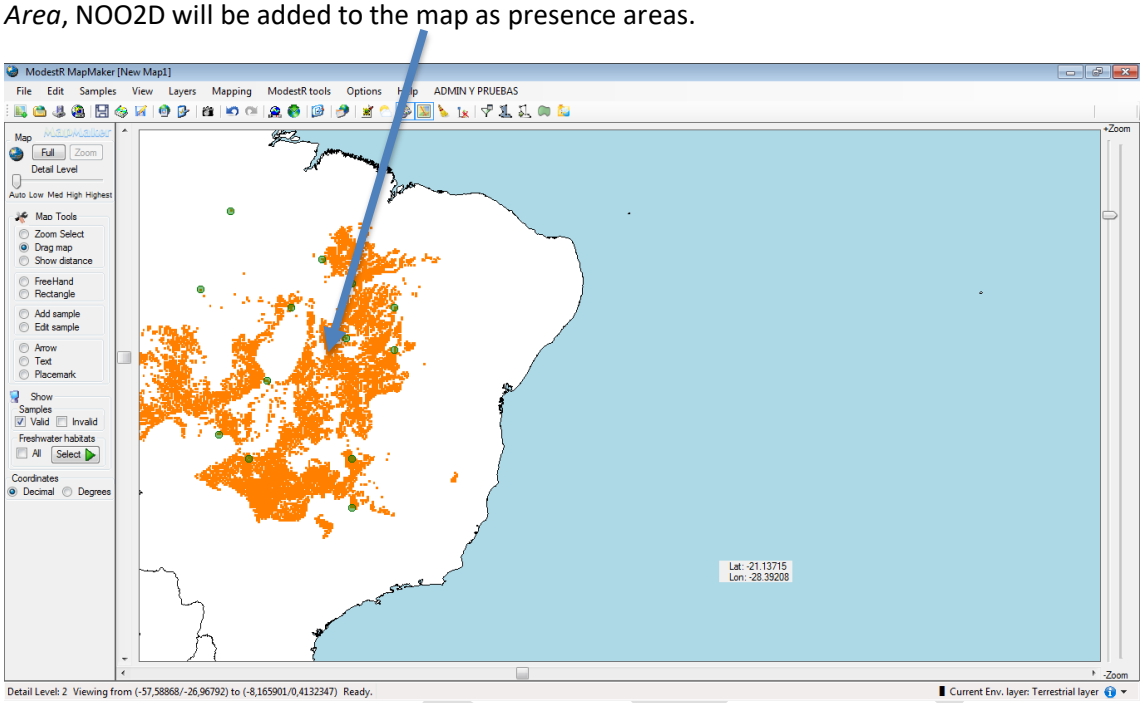
- Smoothing factor: determines how smooth the density matrix will be. The greater it is, the more the density will spread out far around presence areas. Default factor is x1 .
- Allow a tolerance of ± X: Apply a % of tolerance on the potential areas of presence to consider them as valid, with regard to the range between min and max values for each environmental variable involved in the currently selected 2DCEL in the presence data of the species.
- Range: determines the range or maximal area for which NOO2D will be calculated. Options are:
    - Full layer: to calculate NOO2D for all the area covered by the currently selected 2DCEL.
    - Shape: if you have a shape currently added on the map, you can calculate NOO2D only within the bounds of this shape. For example you may want to limit the NOO2D to the river basins where the species is present (see 2.16.2). Note that shapes have to be previously added to the map to be able to use this feature.
    - EOO: you can calculate NOO2D only within the bounds of the EOO (Extent of Occurrence) of the current species. If you select this option, EOO options will be shown on the right panel, where you can set how EOO will be calculated.

Once settings established, you just have to click on *Apply button* to start NOO2D calculation. Once completed, results will be displayed on the map, and on the right panel "*Density data map*".




For *Density map* option, the data shown in the map correspond to the areas that would be part of the NOO2D of the species, with the color corresponding to the density of each area. Areas with higher density are those near from the environmental conditions of the species in the 2DCEL, as shown in the *Density data map* shown on the right panel.

For *Distribution map,* if you selected *Hull simulation,* the areas highlighted in the map correspond to the areas that would be part of the NOO2D of the species. If you selected *Add Area*, NOO2D will be added to the map as presence areas.



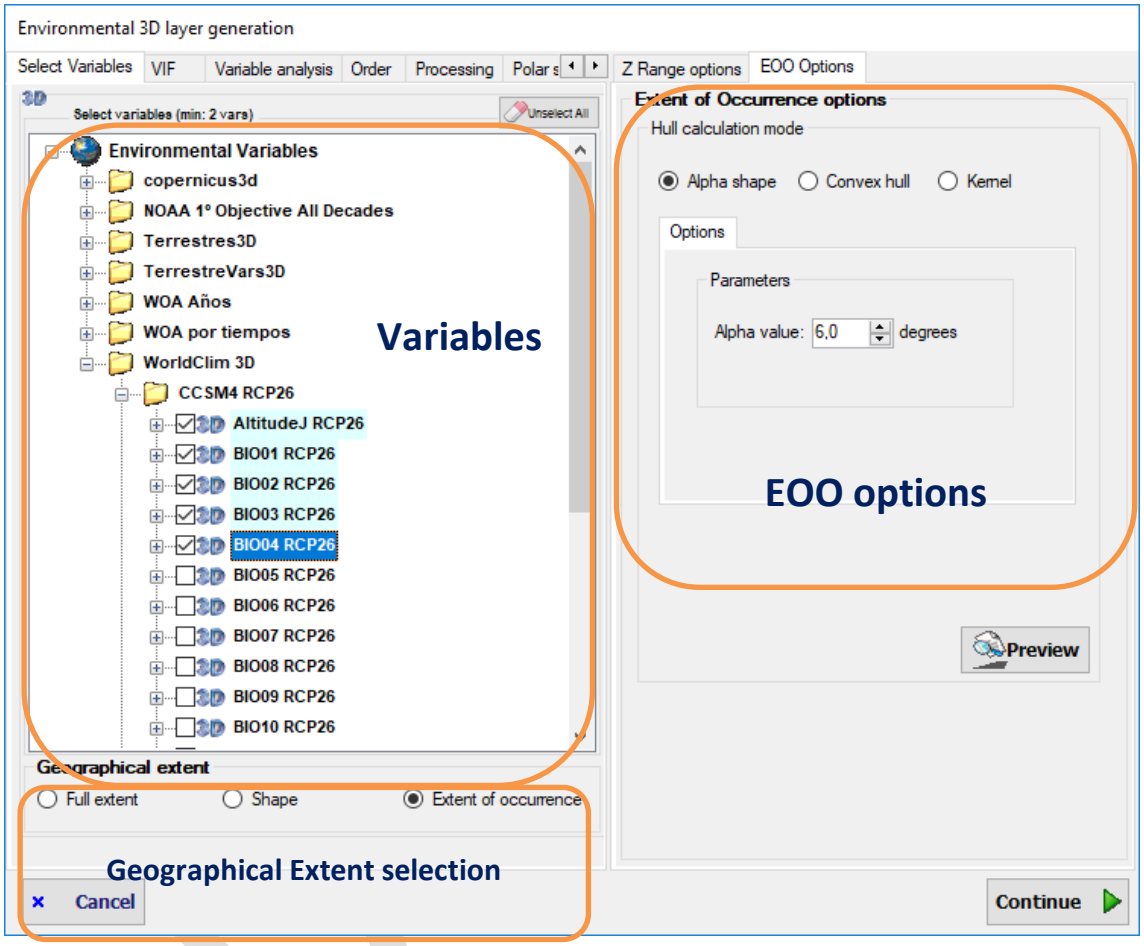

## 2.13 Creating 3D compounded environmental layers (3DCEL's)

We have seen previously 2D CEL's which are a compound of several 2D variables (rasters). But ModestR also allows for creating 3D CEL's, which are a compound of several 3D variables. A 3D variable is a set of 2D variables (rasters) for a given Z dimension, such as a set of temperature rasters for several years. The details about integrating 3D data and defining 3D variables in ModestR are explained in section 6.27.

To create a new 3DCEL, go to the menu *Mapping/Niche of occurrence/Create new compounded environmental layer (CEL)/3DCEL.* The first step starts with the delimitation of the Geographical Extent (GE) to be used as working area. Several methods are available, such as the Extent of Occurrence (EOO) that can be calculated as an alpha shape, a convex hull, or a spatial kernel density; or any shape, as for example river basins with presence observations; or considering the whole world. In MapMaker the option of using a shape as GE is only available if the shape has been added to the map before starting the NOO3D creation. For example, we can add river basins with presence observations using the option *Layers/Shapes/Add river basins with occurrences* once the species occurrences are loaded, and before starting the NOO3D calculation.

The following step consists in selecting the potential predictor environmental variables to assess the suitability of the species. When building a 3DCEL only 3D datasets are shown as variables. At least two variables have to be selected, but usually more variables are selected. Not any combination of variables is allowed. The restrictions to be respected are:

- ✓ All the selected 3D datasets should have the same precision and range. That is, all the rasters should have the same cell size and the same coordinates.
- ✓ All the selected 3D datasets should have the same Z meaning (such as year, depth…)
- ✓ All the selected 3D datasets should have the same layers; that is, the same Z values. For example, if we select as a variable a 3D dataset of temperatures for each year between 2000 and 2010, we cannot also select another 3D dataset of precipitations that does not contains data for the same years.




In this first step the Z range should also be specified. That is, the range of Z values that will be included in the 3DCEL to be created, an that, obviously will determine the available range of Z values to subsequently calculate the NOO3D. If we are using variables that are 3D datasets basing on time such as WorldClim climate simulations, where each layer (i.e. Z value) is a year, the Z range will specify which range of years will be included in the 3DCEL. Conversely, if we are using WOA data where each layer corresponds to a different depth, the Z range will specify which depth range will be included in the 3DCEL.

The available options are:

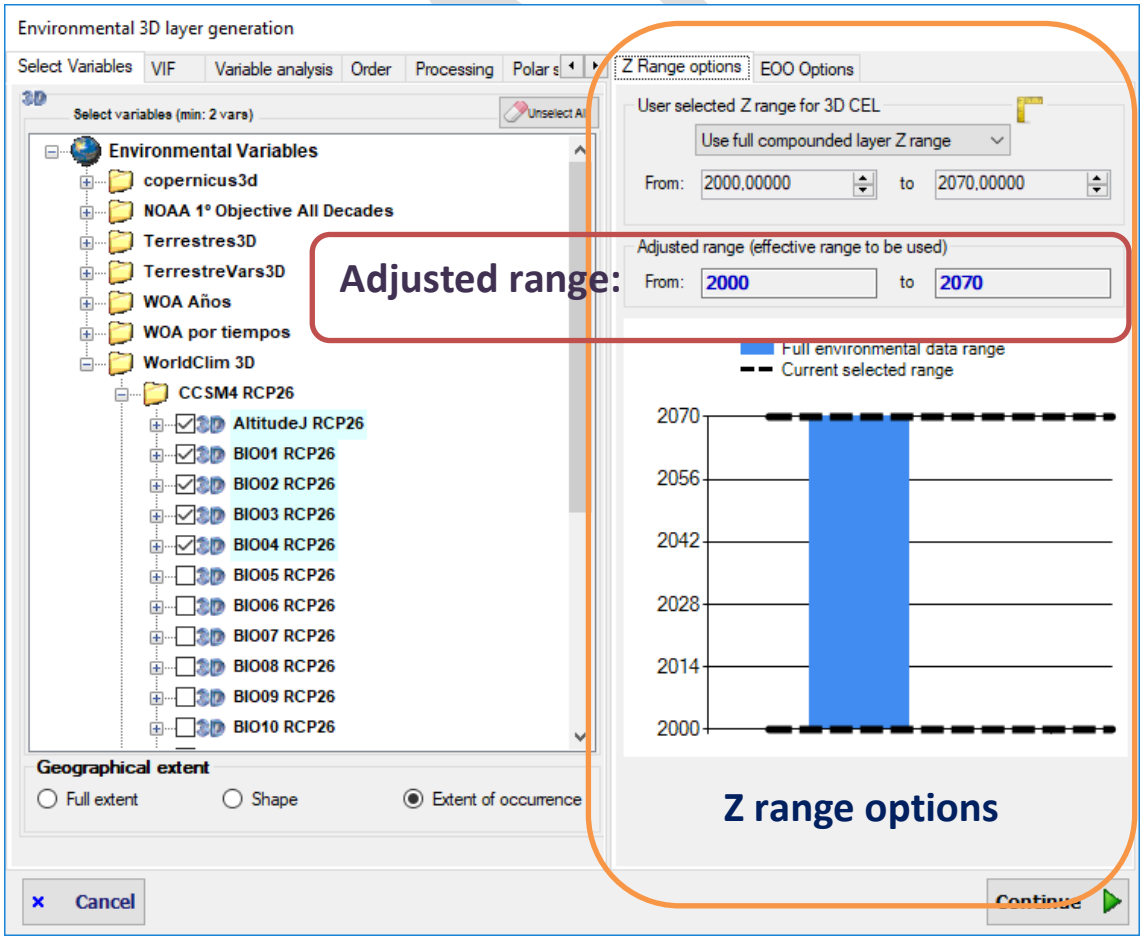
a) **Use full compounded layer Z range**: the whole range of the selected variables will be used. For example, if we select as variables several 3D datasets of environmental data (temperature, precipitations, etc..) that include layers (rasters) for each year between 2000 and 2010, the whole range (2000-2010) will be used,

b) **Adjust to occurrence Z values range** (when a species is currently loaded): the range is adjusted to the species occurrences range, but always limited by the variables range. For example, if we select as variables several 3D datasets of environmental data (temperature, precipitations, etc..) that include layers (rasters) for each year between 2000 and 2010, and the occurrences of the species currently loaded include year data between 1990 and 2005, the effective Z range will be 2000-2005 (as there are no data before 2000 in the selected variables).

c) **Use custom values**: a custom range entered by the user will be used, but always limited by the Z range of the selected variables.

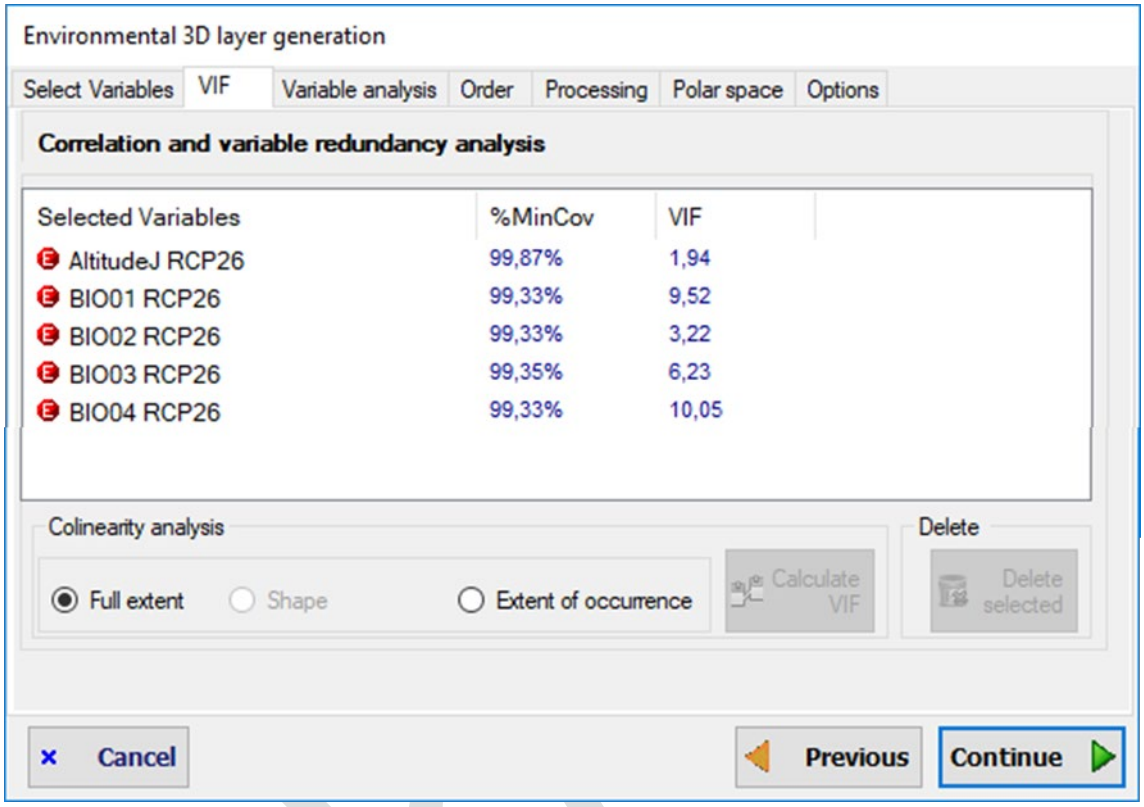
When selecting options b) or c), the selected range is adjusted to the nearest Z values available in the selected variables that include the selected range. For example, if we select as variables several 3D datasets that include layers for the years 2000, 2005, 2010 and 2015; then we establish a custom Z range between 2006 and 2012, the adjusted Z range will be 2005-2015 This adjusted Z range is the one that will be effectively used to build the 3DCEL.

The adjusted range is shown in the options panel, and a graph is displayed to help visually understand the adjusted Z range. In this graph the Z range of the selected variables (i.e. the maximal available range ) is showed as a column, whereas the effectively selected range is shown as two black stroke lines.




In the next step the redundancy among the selected variables can be optionally measured using the variance inflation factor (VIF). A VIF value higher than 30 is usually considered an indicator

of collinearity among variables, and users are provided with the opportunity of removing redundant variables. The VIF can be calculated selecting any GE (geographical extent) allowed, such as the full world, an arbitrary shape, or the EOO. The selected GE doesn't have to be the same than the selected in the previous step as the range of the 3DCEL, even if the most usual option is using the same GE.




In the next step the contribution of each selected variable to the distribution of the species can be optionally analyzed using an Instability Index proposed recently (Guisande, 2016; Guisande et al., 2017). The explanatory variables with a higher percentage of contribution to the Instability Index are assumed those that most affect the distribution of the species in the accessible area and we can whish those to be used in the subsequent steps, discarding the other ones. Several parameters can be set by the user in order to customize this analysis:

- **Z type**: allows selecting the type of Z dimension to be used. That is, the meaning you want to give to Z. It can be depth, year, date&time, or any other (custom). Date&Time refers to the full date and time information (year, month, day etc.) whereas year only takes the year of the temporal information available. The default value is set according to the Z meaning associated to the selected variables. The settings relating to the Z range will depend on the selected Z type.
- **Occurrence selection**: you can choose between using only occurrences that have a Z value (ej. year, depth…) to perform calculations; or using all occurrences, whether they have a Z value or not. Of course, the Z value that will be took into account is the corresponding to the Z type previously selected.
- **Z value assignation**: sets how a Z value will be assigned to each occurrence.

- First of all, we can select between using occurrence Z value if it is present, or not. In the first case, the occurrences with Z value will always be took into account, and their associated Z value used to perform analysis.
  If we choose not using occurrence associated Z value, even if it is present, we can be presented with two situations, depending on the options previously chosen in the occurrence selection setting: (1) if we selected using only occurrences that have a Z value, in any case only those occurrences will be used; but not their associated Z value. Of course, this is not very usual, as habitually we will prefer to use the occurrence associated Z value if it exists. (2) If we selected using all occurrences, whether they have a Z value or not, all occurrences will be used; but not their associated Z value in case they have one.
  An example can be when estimating the distribution of a species along several years, for example using data from climatic simulations. In this case we can have occurrences of a species that can eventually have an associated date, together with variables for different years, probably a recent year and several future years (simulations). It this case we'll probably want to assign all occurrences to the recent year in order to perform the analysis.
- If we selected to use all occurrences (so some of them may not have a Z value), or only the occurrences with a Z value but not their associated Z value, we have to choose which substitute Z value will be assigned to those occurrences. There are three options:
  - Use the mean of the Z value, calculated using the occurrences with a Z value. For example we may use this option when working with Z=depth; we may have some occurrences from a species with a depth value, and some others without it. In this case we may want to assign to those that don't have a value the mean depth value of those that do have a depth value.
  - Use a custom Z value entered by the user; in this case the value should be in the Z range to be used (see next settings). It can be selected from a list that shows the existing Z values in the selected variables, or manually entered.
  - The third option consists of computing a mean environmental layer for each variable and for the selected Z range, and directly assigning environmental values from those layers to the occurrences without Z value (or to all selected occurrences, depending on the previous settings).

- **Z range for contribution calculation**: this parameter sets the Z range to be used to perform calculations. This also determines which occurrences will be used. Only occurrences with a Z value within the selected range will be used. Options are:
  - Use full compounded layer range: the whole Z range of the variables will be used. For example, if we are using Z=year and the selected variables are from 2000 to 2100, the Z range to be used will be this one.
  - Use range of occurrences Z: only species Z range (according to occurrences Z values) that are within the compounded layer range will be used. For example, if we are using Z=depth and the selected variables are from 0m to 5000m, but the occurrences of the species have depth values ranging from 0 to 300 the Z range to

be used will be this last one, excluding the layers of the variables outside this range. This option is only available if there are occurrences of the species with Z value.

- Use custom values: a custom range entered by the user will be used. Of course, this range cannot be larger than the compounded layer range, that is, the Z range of the selected variables.

- Range adjustment mode: when selecting as Z range that of occurrences, or a custom one, this range should be adjusted to the existing Z values available in the selected variables. This can be done in two ways:
  - Strict: range is used without adjustment. Thus, only Z values of the variables included in the range will be used.
  - Nearest: initial range minimum and maximum are adjusted to the nearest available Z values of the variables to obtain a new range that includes the original one. For example if Z=depth and Z values exists in the variables for depths 5, 10 and 15, and user selected range is (6,8), the adjusted range will be (5,10), because there are the nearest existing values.

- **Layer selection**: this parameter determines which layers of the selected variables and within the selected Z range will be used. The options are:
  - Use only layers with occurrences: Only layers in the selected Z range where some occurrences exists will be used to calculate variable contributions. For example, if the selected Z range is [0 to 100], and there are layers for Z=0, Z=10, Z=20, etc... then only those layers where there are occurrences for the corresponding Z value will be used to calculate contributions. This is the default option and the most usual one, as usually we'll only consider environmental conditions most corresponding to the occurrences to calculate contribution.
  - Use all layers: All existing layers in the selected Z range will be used to calculate variable contributions, whether there are occurrences or not for each layer. For example, if the selected Z range is [0 to 100], and there are layers for Z=0, Z=10, Z=20, etc... then all those layers will be used to calculate contributions even if there are no occurrences for some of those Z values.

- **Filter variables by habitats**: apply a filter to remove areas of each variable that are outside the valid habitats for the species. For example, if a variable contains data for marine and terrestrial areas, but the species is only marine, this filter will remove terrestrial areas from the variable before performing calculations. This options requires a lot of extra time, so disable it if you know it is unnecessary to filter variables.

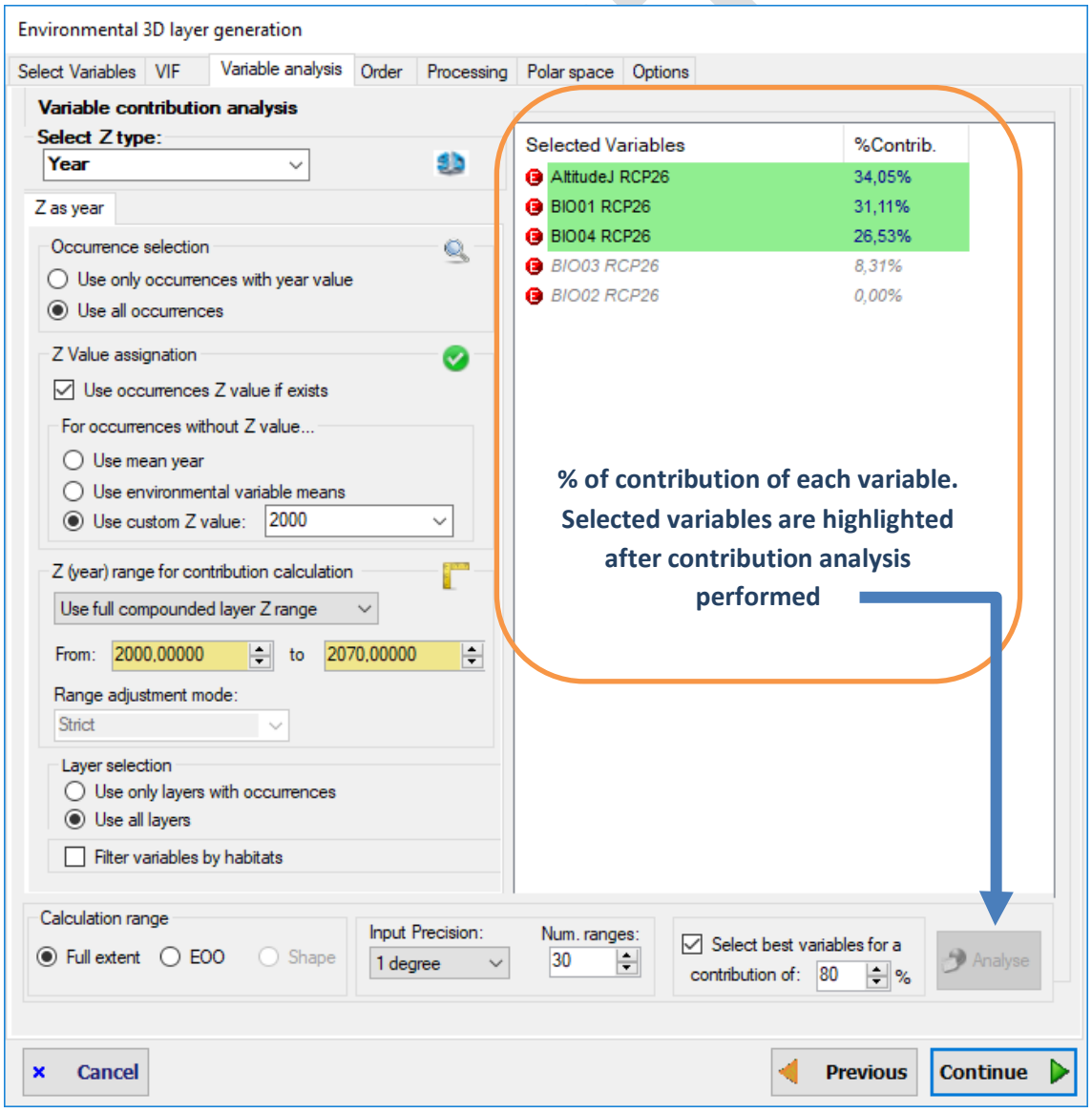
- **Calculation range**: selects the GE or spatial area to be used to calculate contributions. Therefore, only values of the selected variables for this are will be used to calculate contributions. Available options are the same than in previous steps: the full world, an arbitrary shape, or the EOO. The selected GE doesn't have to be the same than the selected in the previous step as the range of the 3DCEL, even if the most usual option is using the same GE.

- **Input precision**: this option allows setting the grid precision that will be used to calculate contributions. This can be set taking into account selected variables precision and

performance issues: selecting a higher precision (i.e. lower cell size) than that of the selected variables does not bring improvement; and the higher the precision, the slower the calculations.

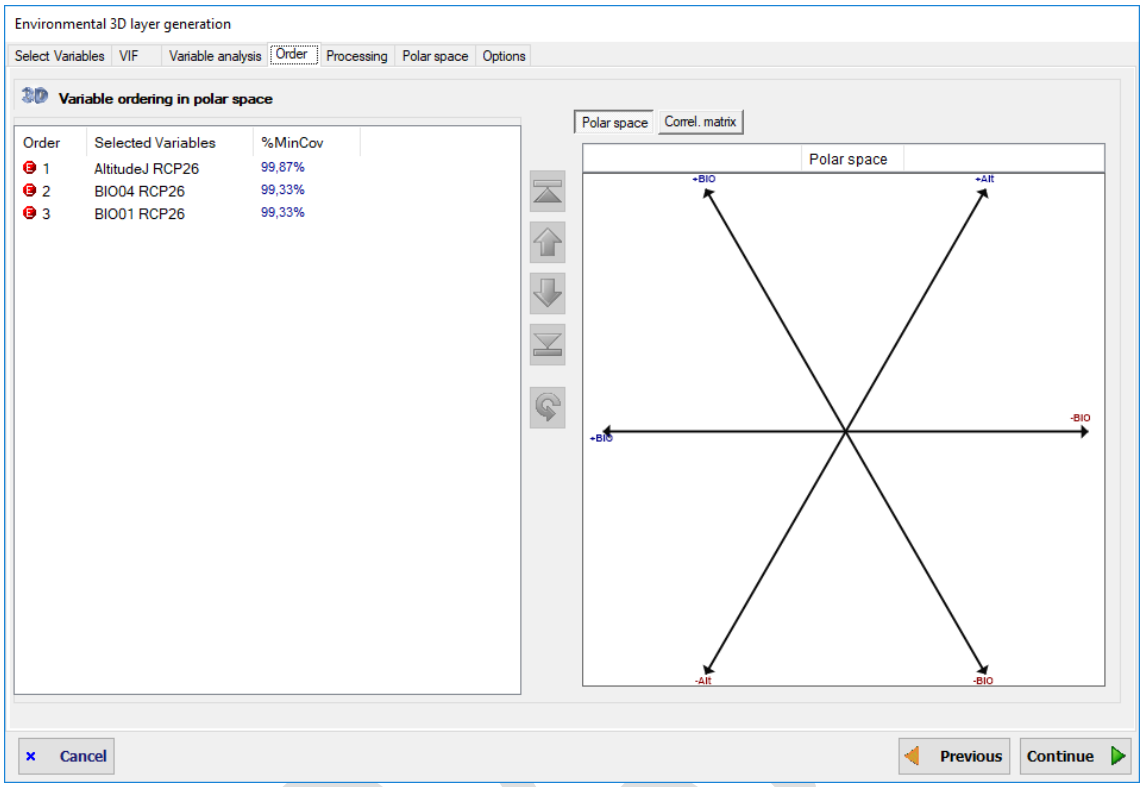
- **Num. ranges**: specifies the number of ranges to be used to calculate contributions. Modifying this parameters needs knowing contribution algorithm in detail, so see Guisande et al. (2017).
- **Select best variables for a contribution of XX%**: once contributions calculated, it selects automatically the variables with the higher contribution to reach the indicated % (that you can modify). Selected variables will be highlighted in the list.

Once those parameters set, the contributions are calculated clicking on the "Analyse" button (this may be done several times, changing the parameters and recalculating contribution). Anaway, this step is optional and we can just go to the next step without analyzing contribution neither discard variables regarding its contribution value.




In this step, definitely selected variables are shown on the list, and on the right panel a representation of how they will be ordered to calculate polar coordinated is displayed. Variables

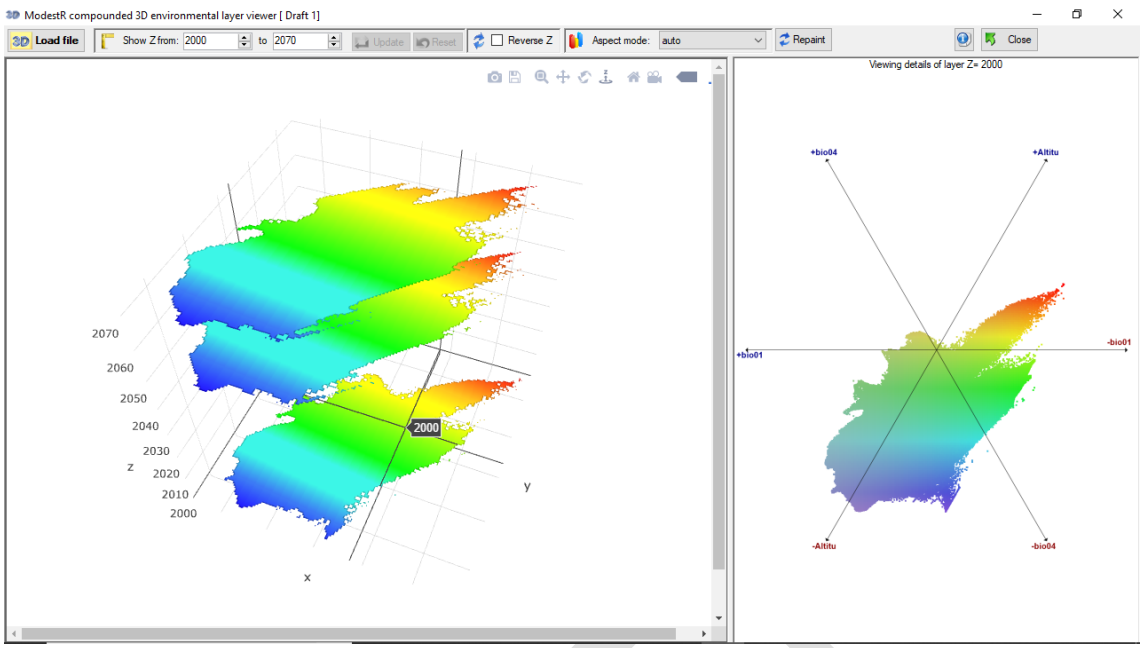
are automatically ordered according to their relative correlations: the next variable to another is the one which as the higher correlation with it. You can eventually manually change order by selecting a variable in the list and using the buttons located on the right of the list.



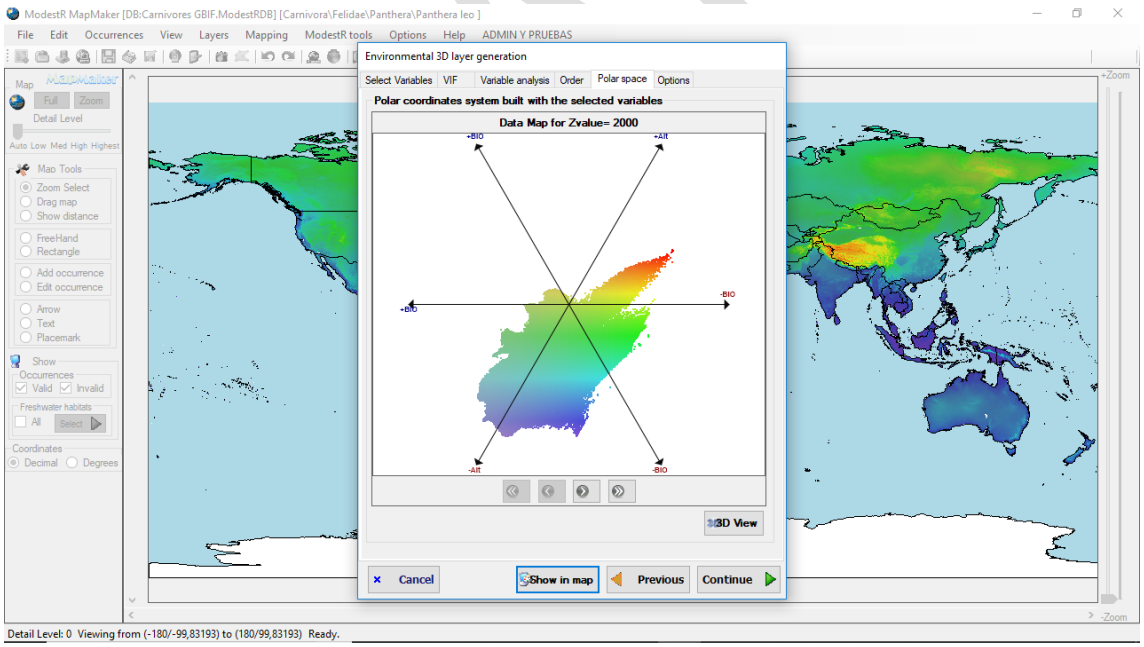

In this step, the 3DCEL will be calculated according to entered settings. This can take some time. To build the polar coordinate system the range of each variable is calculated taking all layers of this variable. For example, if Z is depth, every available layer of temperature for each depth will be used to determine the range of this variable. This results in a reference 2D polar coordinate system that represents the whole environmental space to take into account, including all Z layers (ej: every available depths; or every time when Z is time). Once this reference space defined, it is straightforward to obtain the corresponding 2D subspace for each Z layer. Therefore the resulting 3DCEL is a collection of 2D polar coordinate systems along the Z dimension.

The distribution of the data in the polar coordinates system will be shown in a graph. As this is a 3DCEL with several layers, scroll buttons allow to display next/previous layers, and an indicator on top of the graph shows the Z value of layer currently displayed.

A button also allows opening a 3D viewer where an interactive 3D graph allows to explore in a 3D manner the 3DCEL. When moving the mouse cursor over the layers of the 3D graph, the corresponding 2D layer is shown on the right panel. This viewer has other useful features such as allowing to show only a subrange of layers or reversing the Z index, that may be useful when working with depths, for example.

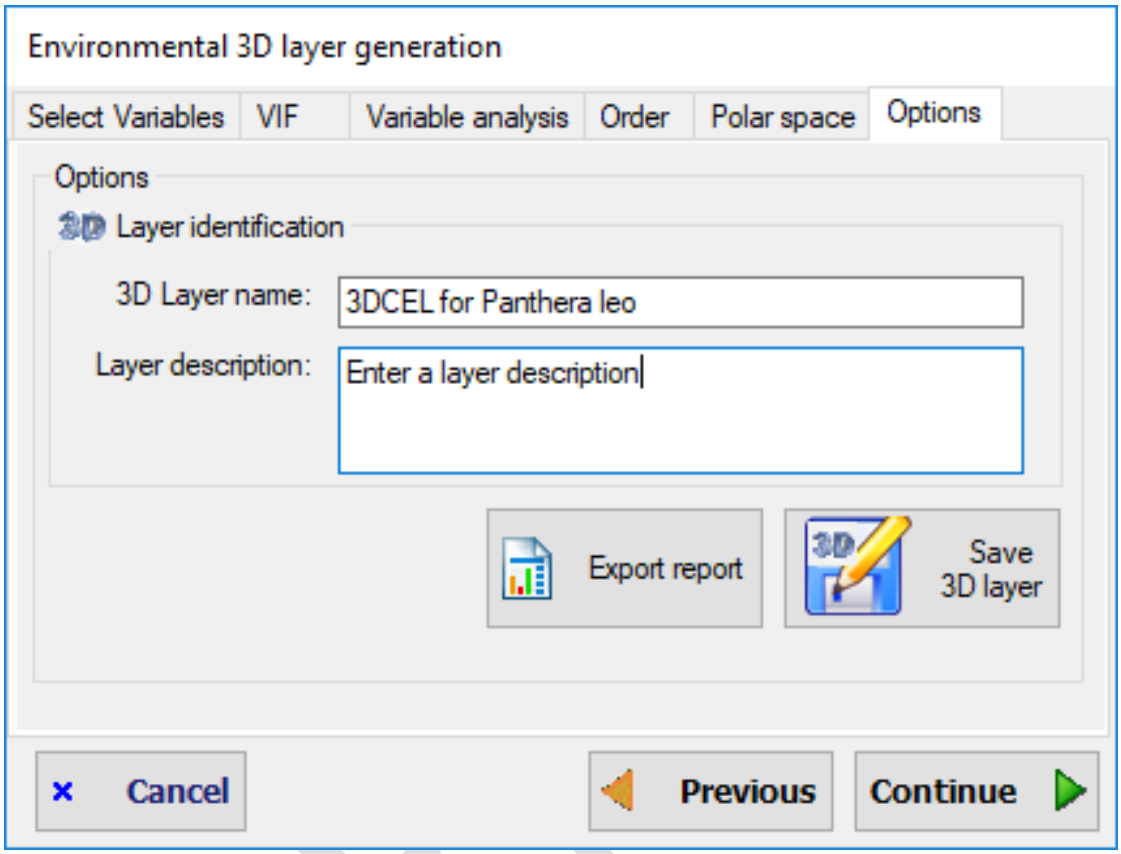



The resulting 3DCEL can be also shown layer by layer on the map using the *Show in map* button, as in the figure here below. But remember that is a CEL, it is independent from any species data; it'll be in a subsequent process of NOO that this layer can be used to estimate a species distribution.




The final step allows to save and/or to export a report of the generated layer. In both cases, you should enter a layer name before. Then you can export a report to a folder, and/or save the layer to be able to use it later.

Exported report will contain several files such as the correlation matrix in a CSV file and the 3D view of the 3DCEL in a file visualizable using the MR3DCELViewer included in ModestR (go to *Tools* menu in any of the ModestR applications to run this viewer).




Storing the generated 3DCEL is not mandatory. We can just continue and use this on-the-fly 3DCEL in this worksession, but it'll be lost when we'll close MapMaker. Therefore, if a 3DCEL is susceptible of being used more times, it's better to store it.

## 2.14 Calculating 3D niche of occurrence (NOO3D)

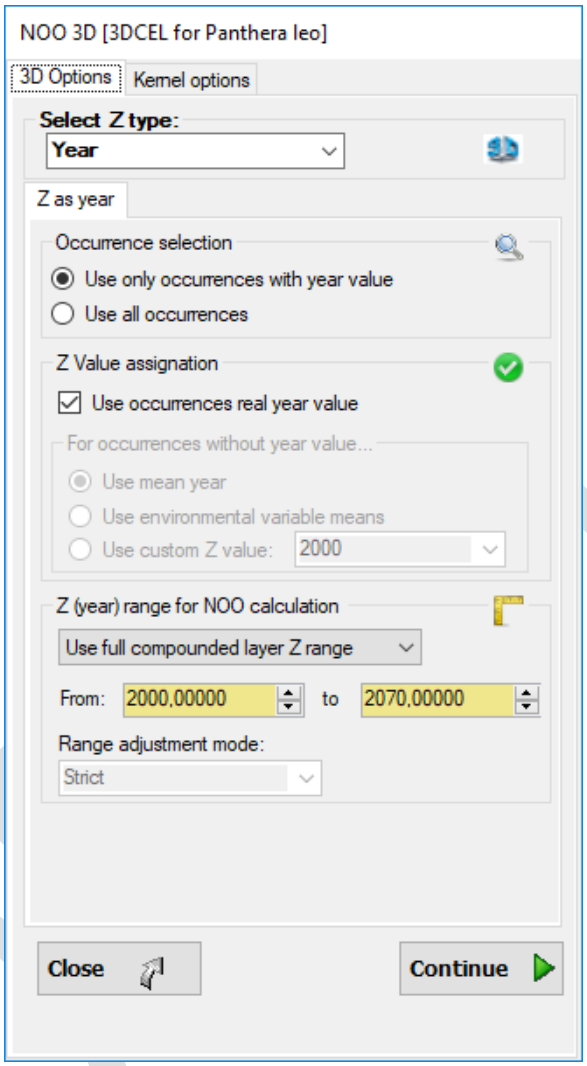
NOO3D is a generalization of NOO2D (see 2.12) for three dimensions[12], therefore used 3DCEL's (see previous section). NOO3D calculation can start using *Mapping/Niche of occurrence/Density map* menu option, or *Mapping/Niche of occurrence/Distribution map* menu option. The difference between both options is the output. In the first case we'll obtain a density map visually showing the areas where is more/less likely for the species to be present, but no map modification will be done. In the *Distribution map* option a presence/absence calculation is done, and new presence areas are added.

As for NOO2D, the first step will be select or create a 3DCEL to be used to calculate the NOO3D. Once a 3DCEL has been selected, the NOO3D itself can be calculated for a species.

The first step is setting the parameters regarding the occurrences and the Z range to be used. The available options are a subset of those described previously for the variable contribution

[12] Being strict, NOO3D is not a full 3D model, but more a 2.5D.That is, a collection of 2D models performed using a single reference space, along a 3th dimension (ej: time or depth).

calculation, but in this case they will affect how NOO3D is calculated: Z type, Occurrence selection, and Z value assignation have been already explained previously for the variable contribution calculation; and the Z range for NOO calculation is analogous to the explained Z range for contribution calculation.

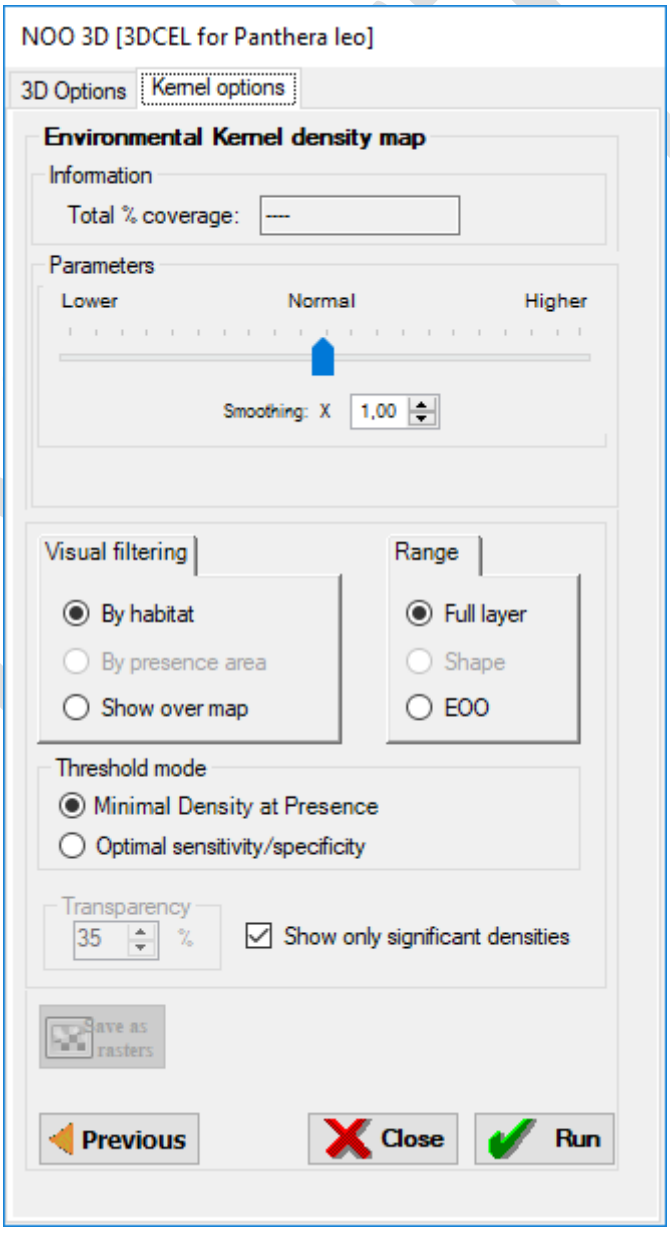



In the next step the available options allow to customize the kernel density estimation:

- **Kernel options**. The suitability of the species in the 3DCEL environmental space is calculated using a kernel density function. Therefore the areas more suitable to the species will be those nearest those where more occurrences exists. Users are allowed to tune two important parameters of this function: smoothing factor, which determines the kernel spread around presence areas in the density matrix (default is x1); and tolerance (default is 1%), which limits potential areas of presence to those within an environmental range defined from known occurrences according to this value. It must also be signaled that for those species with low number of records, for which was not possible to calculate the kernel density, only the areas where there are occurrences will be considered suitable areas for the species.
- **Range**: the spatial boundaries where the NOO3D will be calculated. The available options are the same than those for the 3DCEL (alpha shape, convex hull, spatial kernel

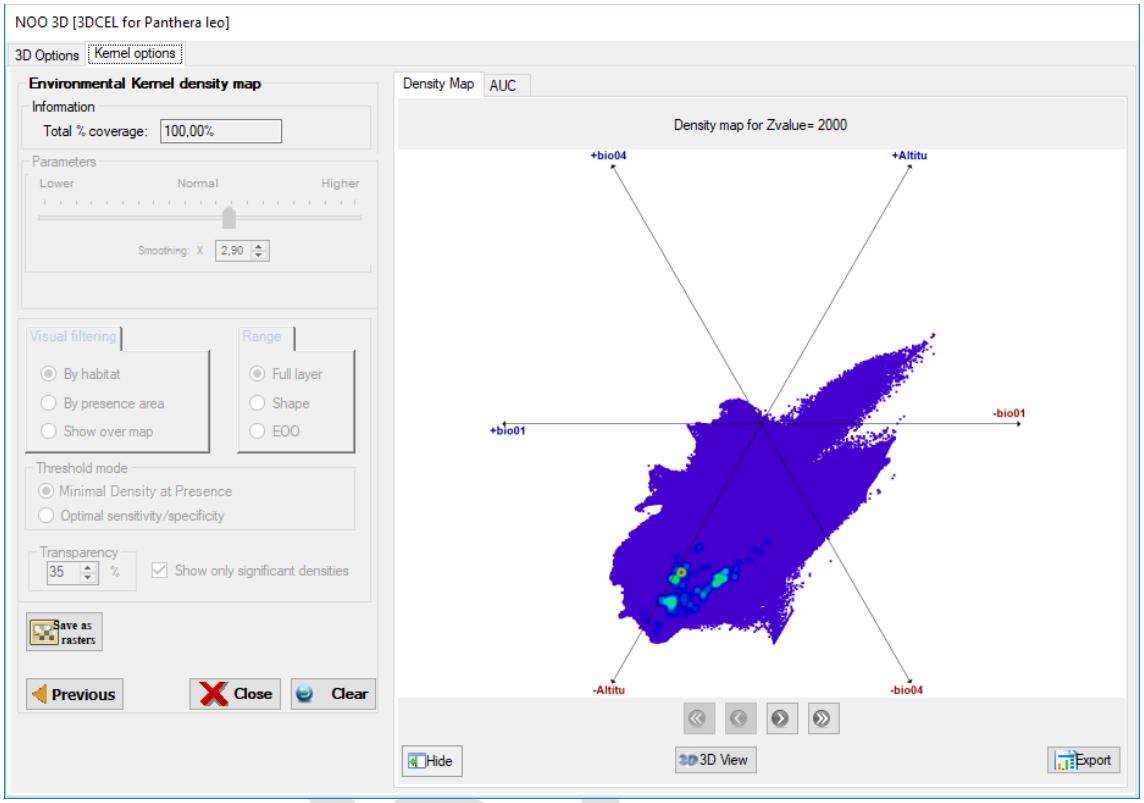
density, river basins with presence observations, or considering the whole world). But both options should not be confused. The 3DCEL boundaries directly settle the maximum geographical extent to be considered, and therefore also restricts the environmental space to be considered. The NOO3D only can establish the maximum spatial area to be considered for the species, which may be the same than for the 3DCEL, or a more restricted area. Obviously it has no sense to try to establish larger boundaries for the NOO3D than those of the 3DCEL used to calculate it. Nevertheless, apart from this limitation, the decision of the suitability of any 3DCEL/NOO3D boundaries combination is left to the user.

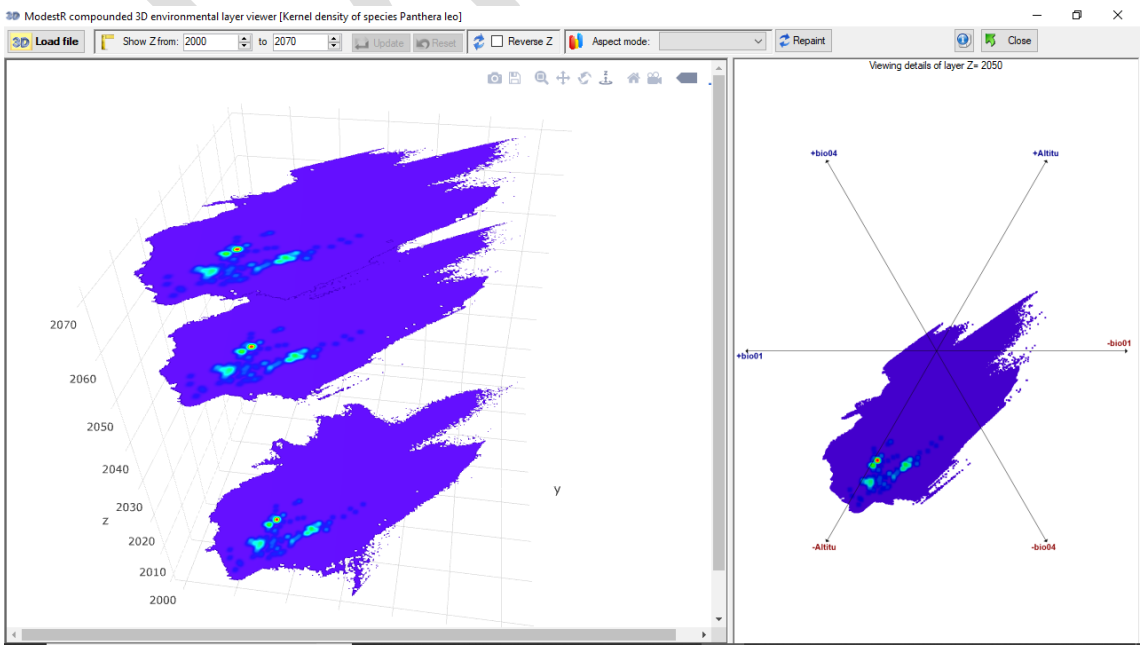
- The **threshold mode** to determine the species presence/absence. It can be the minimal density value at a known presence or the optimal balance between sensibility and specificity.
- Other visualization options are also available, such as filtering the density map by habitat, setting transparency, etc.



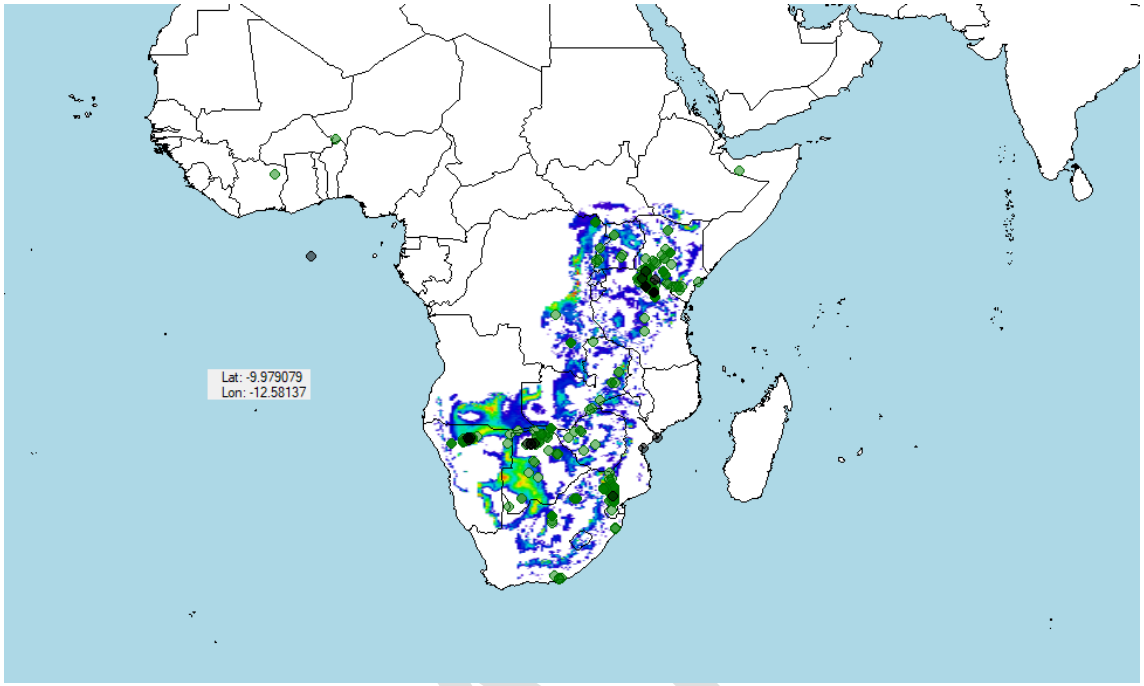
Once the parameters set, the density map is calculated clicking on the *Run* button. Once computed, the output results from an NOO3D are:

- A polar density map that shows the density (suitability) of the species in the environmental space of the 3DCEL. There is a map for each Z level, and a 3D view is also available.

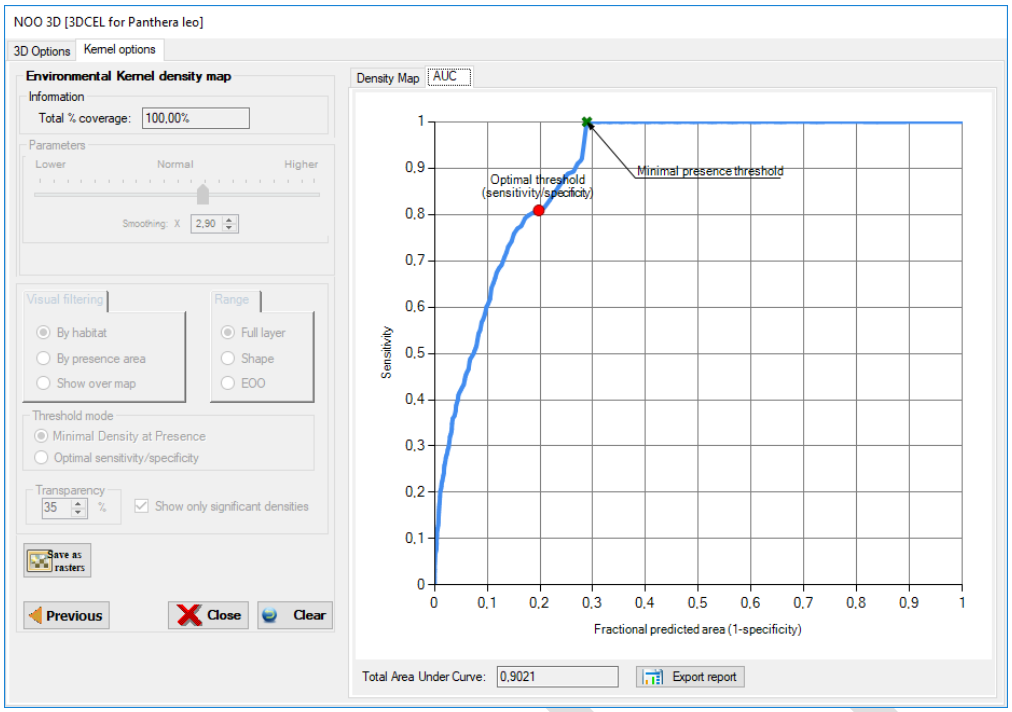


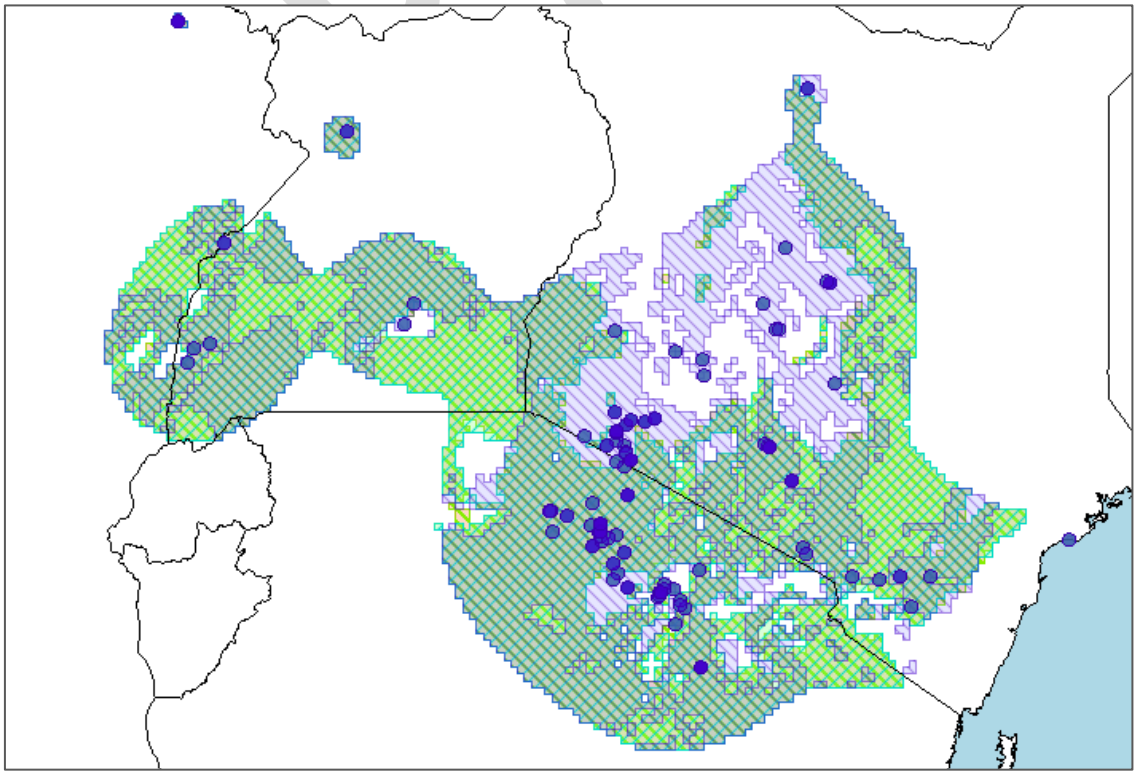
- A spatial density map for each Z level, representing the suitability level of the species in the different areas.




- The ROC curve and the AUC value computed from verified presences and background pseudo-absences. NOO seeks to provide distribution estimations when reliable and unbiased absences are not available. In these situations, measuring the discrimination capacity of a model results by the classic ROC (receiver characteristics curve) and AUC (area under the curve) values cannot be calculated due to the lack of information about the false positive rate (1-specificity) (see Lobo et al., 2008; Jiménez-Valverde, 2012). Phillips (2017) propose to replace the false positive rate by the fraction of the total study area predicted present. The consequence of this strategy to circumvent the problem of the lack of absence data is that model repre-sentations correctly predict as many as possible presences in the smallest possible area; that is, a predicted area inevitably very similar in shape and extent to a simple density function of the available presence data used in the training process. Even if it is a wrong procedure of validation, ModestR allows calculating these pseudo-ROC and pseudo-AUC values.




- A multiple-layer distribution map generated from the density map and using the selected threshold value to build a binary (presence-absence) species distribution map, where presence areas are considered the niche of occurrence of the species. This map contains a presence area for each Z value in the selected range that can be visualized in MRMapping, which allows showing several overlapping distribution maps.

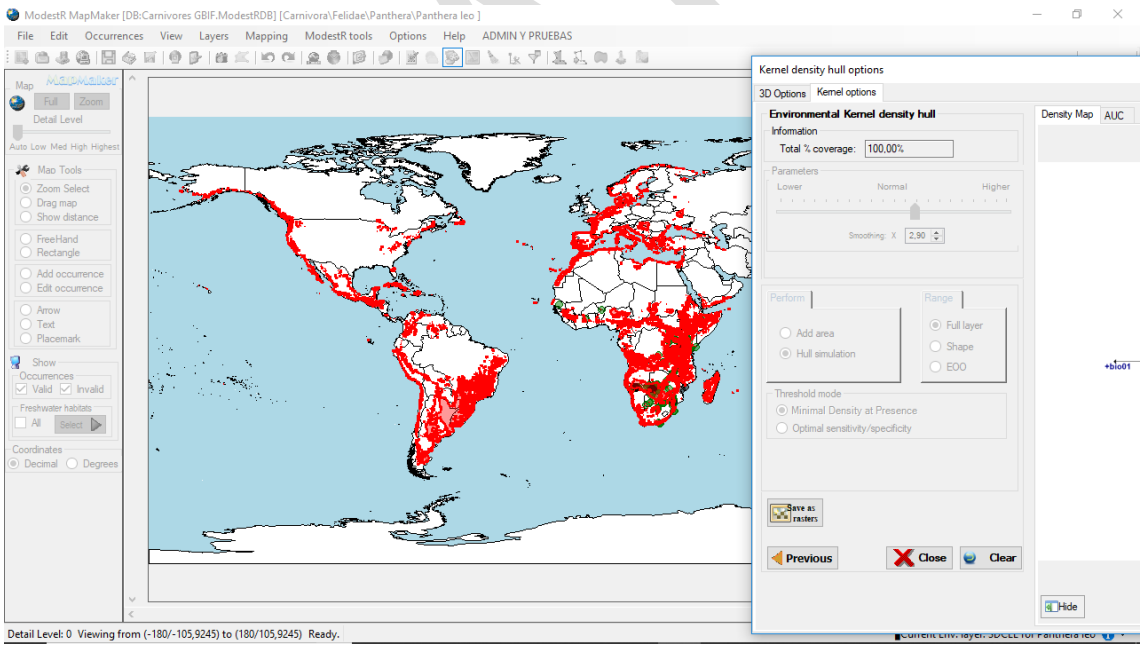


*Detail of a small area where the several distribution layers of the species for each Z layer can be seen in different colors, in MRMapping.*

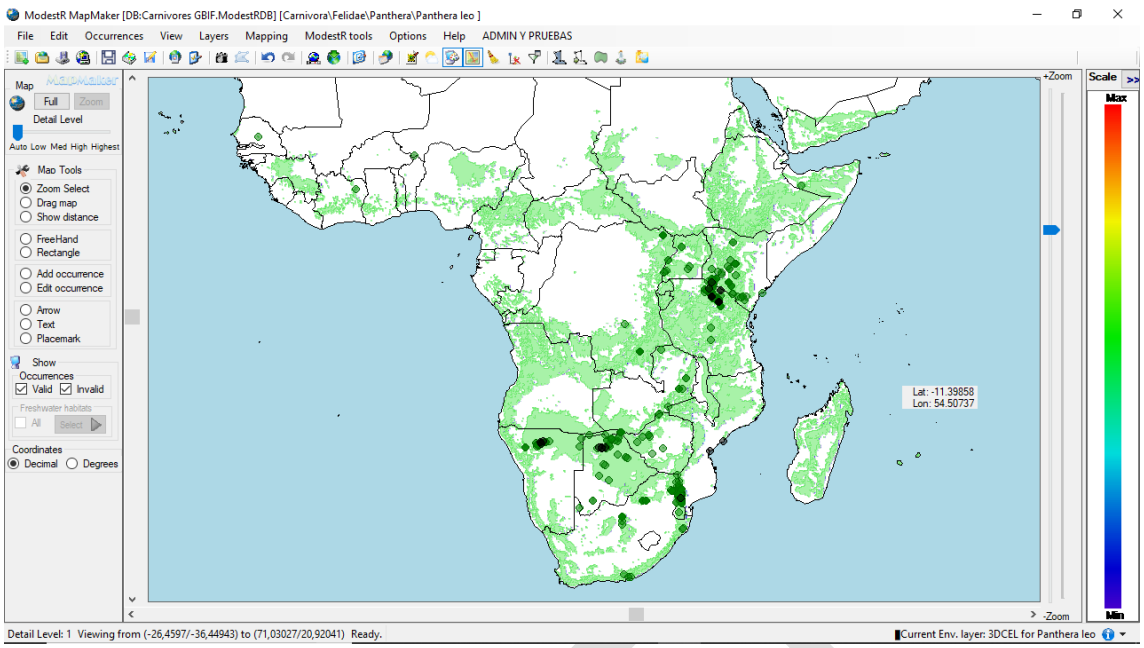
This density map can be calculated several times changing the settings, just by clicking on the *Clear* button that appears after the first time in the dialog box, changing the parameters and clicking again on the *Run* button.

This is the process done using *Mapping/Niche of occurrence/Density map* menu option. As said before this feature provides a density map visually showing the areas where is more/less likely for the species to be present, but no map modification will be done. Using the *Mapping/Niche of occurrence/Distribution map* menu option a presence/absence calculation is done, and new presence areas are added. In this last case the available options are the same, excepted those related to visualization, as the output will be a set of presence areas in turn of a density map.

In the case of the *Distribution map* feature we can choose between performing a simulation, which just adds the likely presence areas as visual elements to the map (so not took into account as part of the species distribution). Or adding areas, which definitely adds presence areas to the species distribution map; this option also exports a multiple-layer distribution map visualizable in MRMapping as when performing a density map.




*Result of performing Distribution map in simulation mode. The likely presence areas are shown as visual elements in the map, but are not added as presence areas.*




*Result of performing Distribution map in add areas mode. The likely presence areas (in green) are definitely added to the species distribution map.*

It is very important to remark that when performing a *Distribution map* the presence areas obtained for each Z layer are added to the species distribution map. As ModestR doesn't save associated Z value information for each one of those areas, this may not be always the wanted result.

For example, if we are calculating the distribution of a marine species for a range of depths, obtaining a distribution map where the presence areas for the whole range of depths has been added may be useful, even if the detail about the depth corresponding to each presence area will not be stored. But if we are calculating how the distribution of a species may change along a range of years, for example using climate simulations, a distribution map where the presence areas for the whole range of years has been added and are not distinguishable doesn't have sense.

That is why both when calculating density maps or distribution maps in a NOO3D context a multiple-layer distribution map for the species is exported in the format used by MRMapping application. This map contains a distinct presence area for each Z layer. This allows showing several overlapping distribution maps, analyzing results layer by layer, exporting distributions for individual layers, etc.

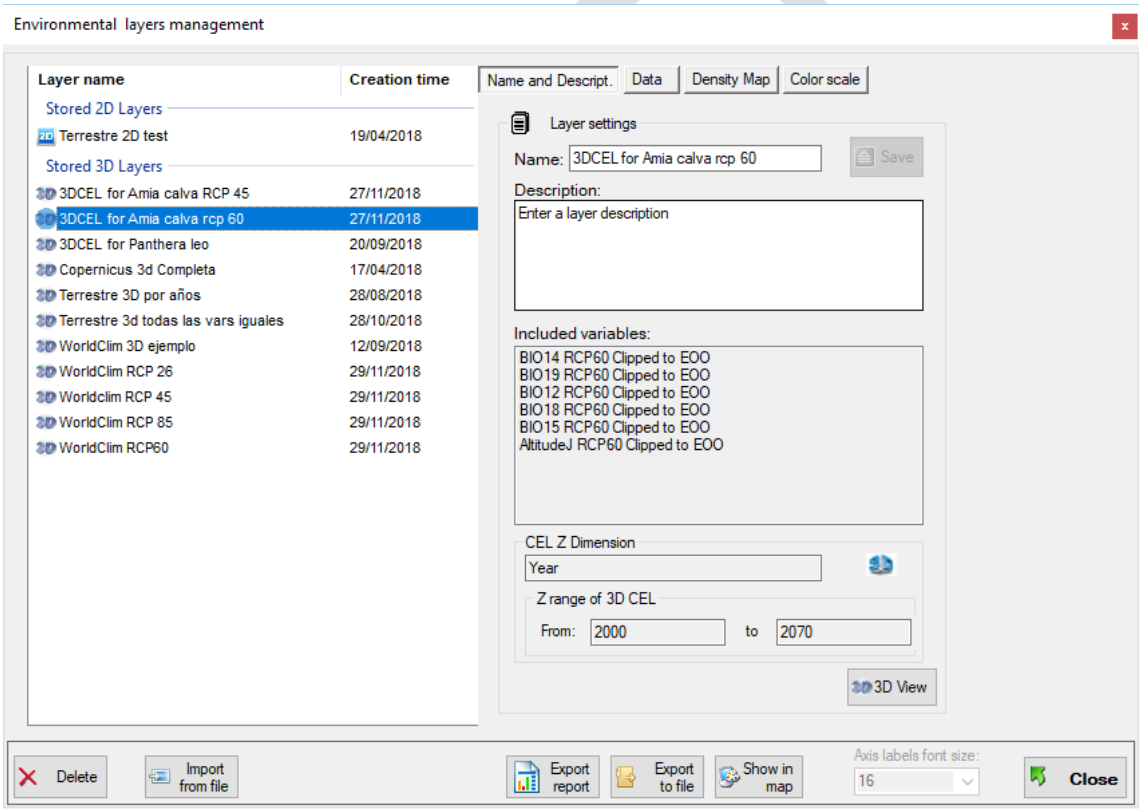
Another possibility to obtain those breakdown results by Z layer is using DataManager, which, as explained before, can generate a distinct database where the results of the NOO3D for each layer of the species distributions are stored separately.

## 2.15 Managing CEL's

You can manage the 2D and 3D CEL's you have stored going to the menu *Mapping/Niche of occurrence/Manage compounded environmental layers*. A dialog box will appear showing the currently stored 2D and 3D CEL's on the left (an error message will be shown in case there is no

stored layers). On the right, several tabs allow to view information about the CEL you select on the left list. And several buttons on the bottom allow for some actions. You can select any CEL of the list and:

- Show it in the map, as a raster , by clicking on the *Show in map* button.
- Show/edit CEL details, such as name or description, by clicking on the *Name and Descript.* tab. You can also see the variables used to generate the layer, the polar coordinates and the data distribution into this coordinates. You can also see the CEL polar coordinates graph using the *Data* tab.
- Export the CEL report. Exported report will contain several files such as an ASC raster for the X and for the Y components, and a graph of the polar coordinates with the data distribution distribution (see 7.11 for details).
- Delete a CEL that you no longer need, by clicking on the *Delete* button.
- Export or import a CEL to/from a file. This allows exchanging CEL's between ModestR users, for example.

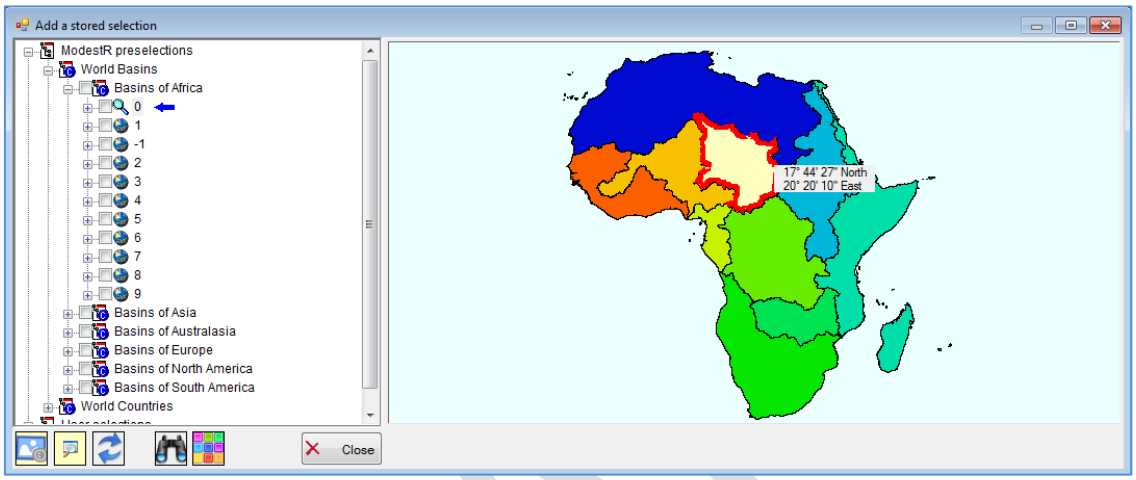



## 2.16 Working with shapes and basins

In MapMaker shapes are just helping elements that don't modify distribution data. They are not stored with a map, but just shown during a worksession. They must not be confused with shapes used in MRFinder. The shapes are the same, but the difference is their purpose, which is merely informative in MapMaker whereas in MRFinder it indicates a search area.

In MapMaker shapes will be commonly used to help the user to locate the boundaries of some specific area, such as an administrative region, a natural site, etc.

To add predefined shapes to the map, you can go to *Layers/Shapes/Add shapes* or the corresponding button of the toolbar. A dialog box will be shown where you can select the shapes to add from a tree where they are hierarchically grouped. On the right of the tree, an active map shows a preview of the shapes (when available) and allows you to browse them in an intuitive way. For example, you can select "World countries" and a world map will be shown on the right preview panel. You can then double click on Africa in this preview to go into Africa branch and see details, such as African countries. A context menu allows you to select any element in the preview to add it to the map. You can also select it in the left tree. More options are explained in section 4.5.




Some buttons on the bottom left corner provide other features such as searching by name, or color settings.

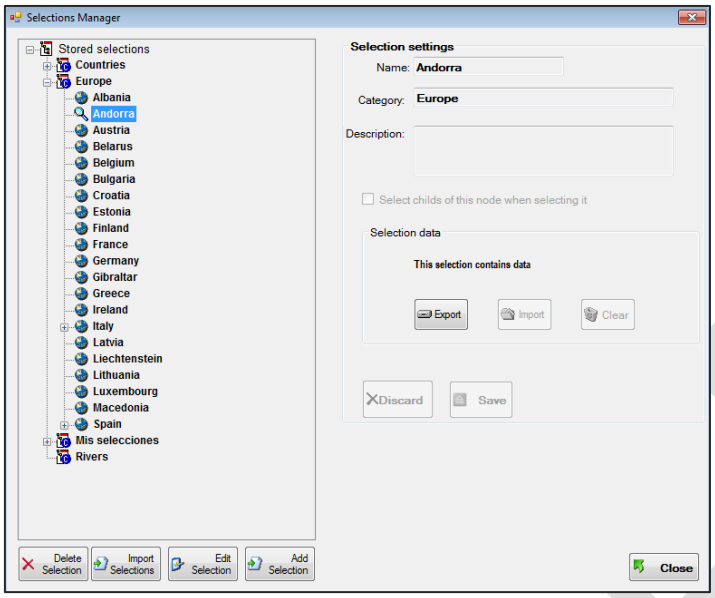
To add your owner drawn shapes, just use the *Freehand* or the *Rectangle* tool to draw any shape on the map, then click with the right mouse button and select *Add as shape*. You can add as many shapes as you want.

You can also import a shapefile as a shape going to *Layers/Shapes/Import from shapefile.* Another option is copying a species distribution map using *Edit/Copy as ModestR map*, and pasting it latter using *Edit/Paste map as shape.* In this way, the areas of the copied map will be pasted as a shape over the current map. This can be useful to use species distribution presence areas as a shapes to add, intersect or just visualize them over another species map.

You can use the *View/ Show shapes* menu item or the corresponding button of the toolbar to show/hide a shape. In the option *Layers /Shapes/Set shape transparency* you can change the transparency degree of the shapes.

From ModestR version 2.0 a collection of predefined shapes is provided with ModestR, that are displayed under the "ModestR shapes" node of the tree, but you can also add your own shapes, that will appear under the "User shapes" node of the tree. To do that go to *Add shapes* option. Initially you will meet an empty list. To import your own shapes from KML or shapefiles, you can go to *Layers/Manage shapes.* A dialog box will be shown where you can add new shapes and organize them in a tree. You can see more detail in section 4.11 about predefined shapes in MRFinder, which support the same feature.

When a predefined shape is added to the map (e.g. a country, a river basin…), its name is displayed on the mouse cursor when you place the mouse over it.



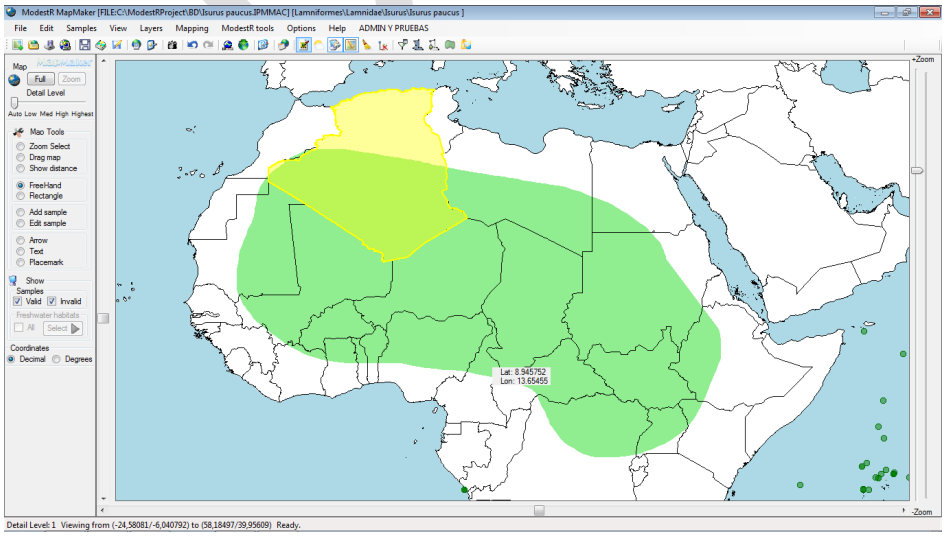

To clear shapes from the map, you can go to menu *Layers/Shapes/Clear all shapes*, or use the corresponding button of the toolbar.

You can convert a shape to a presence area going to *Layers/Shapes/Convert to presence areas.* A dialog box will ask you the valid habitats for the species, and the corresponding areas will be added to the map.

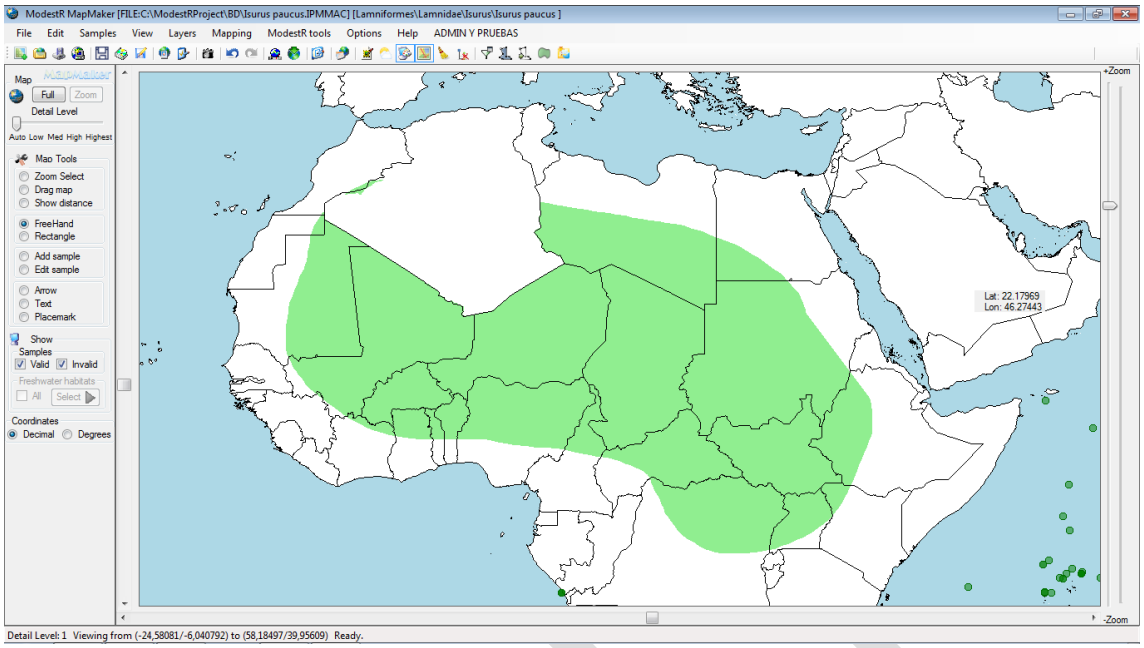
### 2.16.1 Using shapes to unfill or intersect

You can use shapes as patterns to unfill a portion of a presence area, or to calculate the intersection with a presence area.

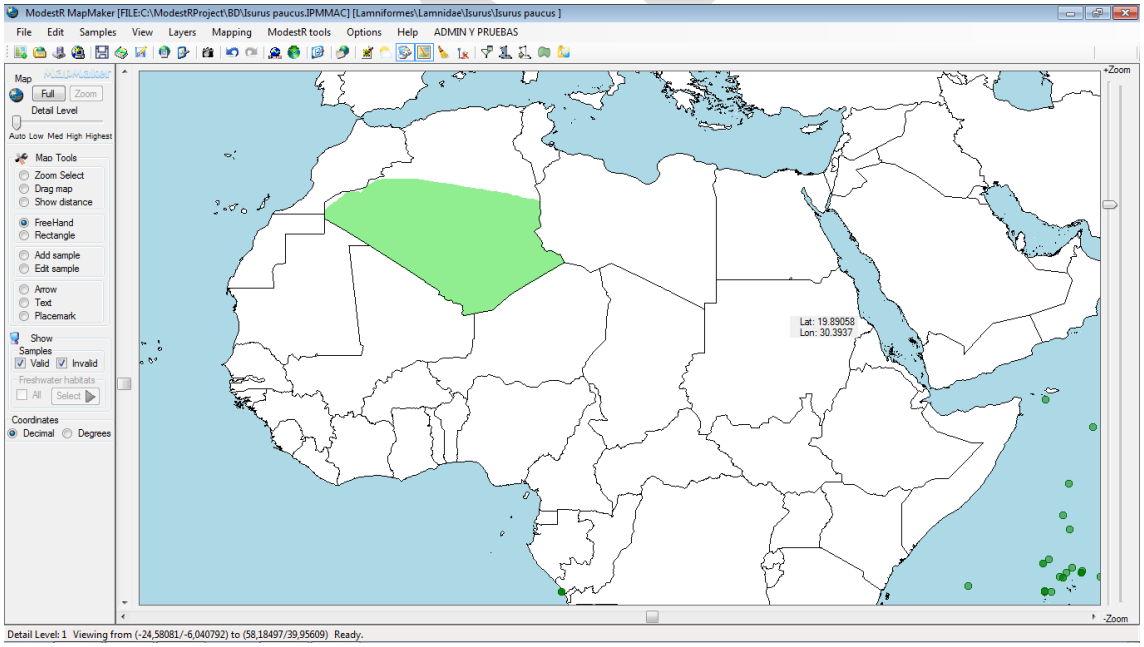
For example, in this map a presence area and next a shape have been added.



Then, going to *Layers/Shapes/Unfill using shape,* the result will be:



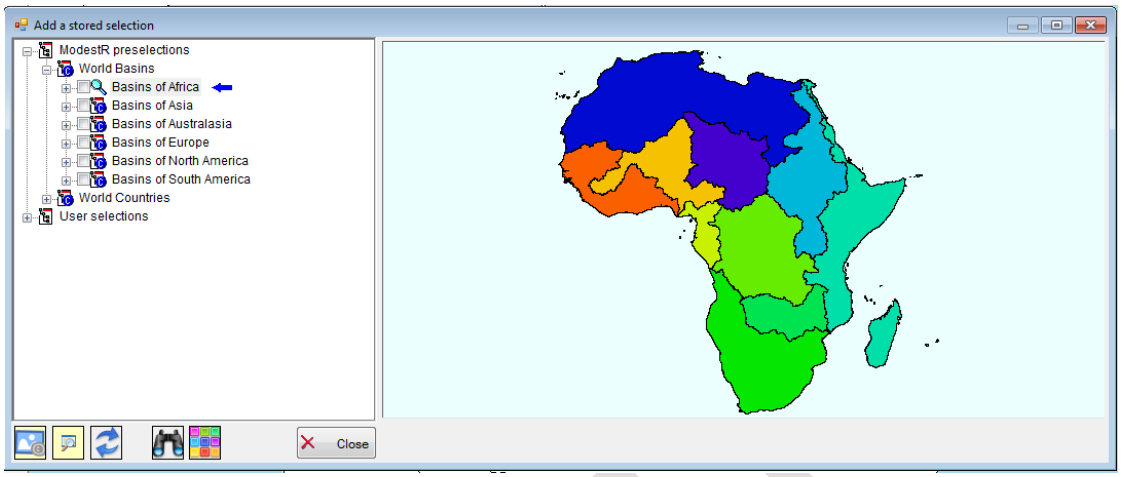

In turn, if we used *Layers/Shapes/Intersect using shape,* the result will be:



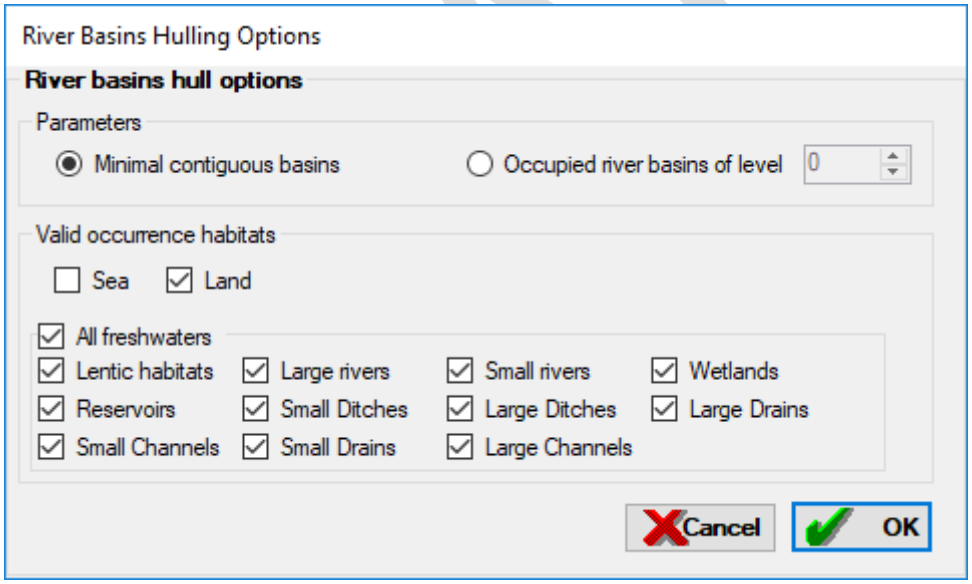

## 2.16.2 Automatically adding river basins shapes

When working with occurrence data from terrestrial species, it can be interesting to see in the map in which river basins the species is present. As pointed out above, ModestR provides a collection of predefined shapes corresponding to administrative areas for all over the world, and more recently it also includes shapes corresponding to river basins. They can be found under the node *World basins* in the tree of shapes. River basins are grouped by continent and hierarchically organized in 6 levels, from main large basins to smaller basins.

You can manually add shapes of those basins just as any other shape, going to *Layers/Shapes/Add shapes* and selecting the wanted shapes either in the left tree of in the right active map.

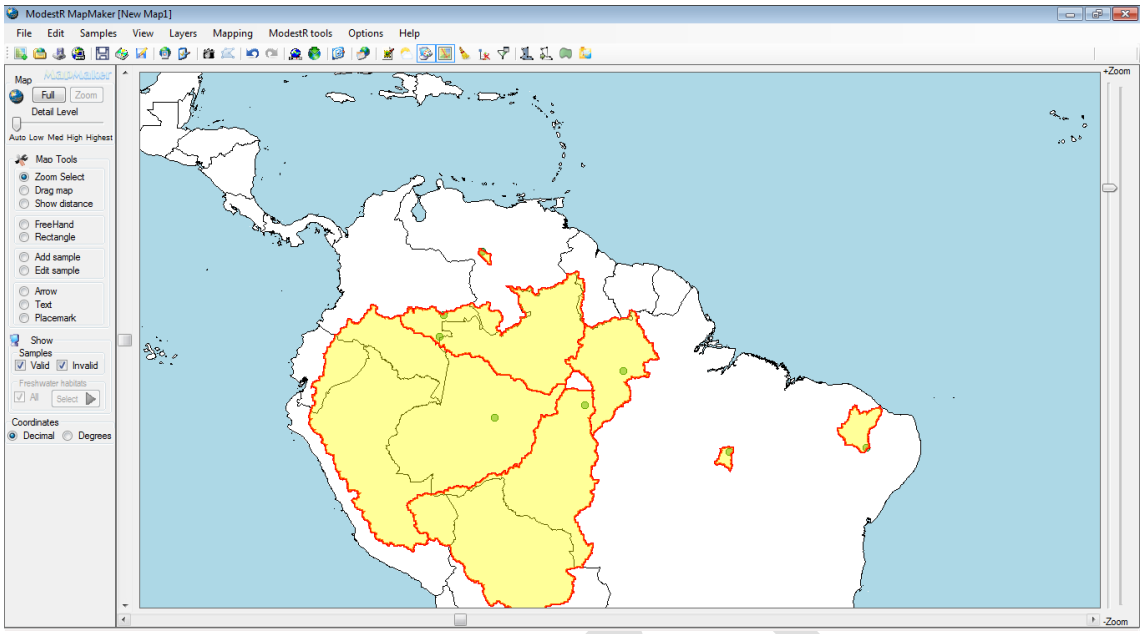



But you can also make MapMaker to automatically search and add the river basins where the species is present. Please note that this feature only works with occurrence data, but not with area data. To do that, go to *Layers/Shapes/Add river basins with occurrences*. The following dialog will be shown:




Here you can select between:

- Minimal contiguous basins where the species is present. In this case MapMaker will search for a set of basins that are contiguous, and using the lower basin level. The result in the minimal set of smallest basins which are contiguous (excepted when this is not possible, such as for species in different continents) and where the species is present.
- Occupied river basins of the level you want to add (0=higher, 5=lower). MapMaker will add to the map the shapes corresponding to the basins of the selected level where there are occurrences of the current species.

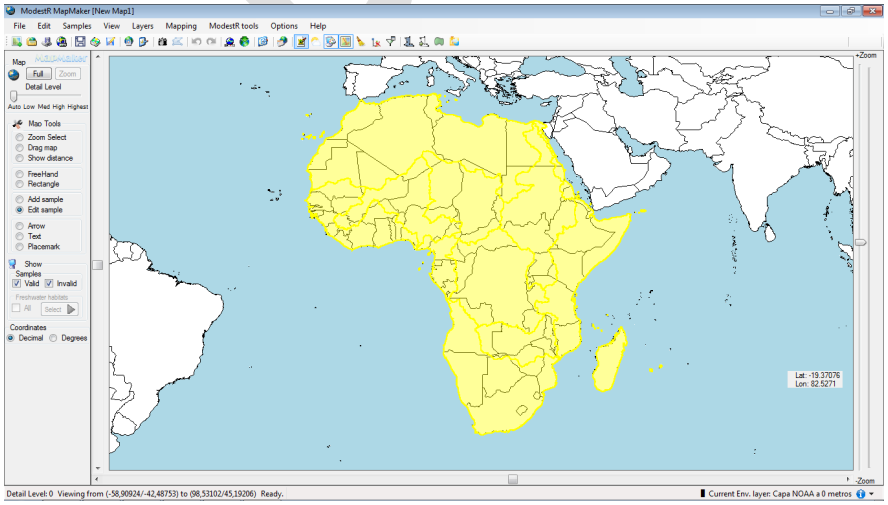



Besides visual purposes, this feature can be useful to calculate NOO2D or NOO3D for a species limiting the potential area to the river basins where the species is present, for example (see 2.12).

## 2.17 Showing 2DCEL data for a shape

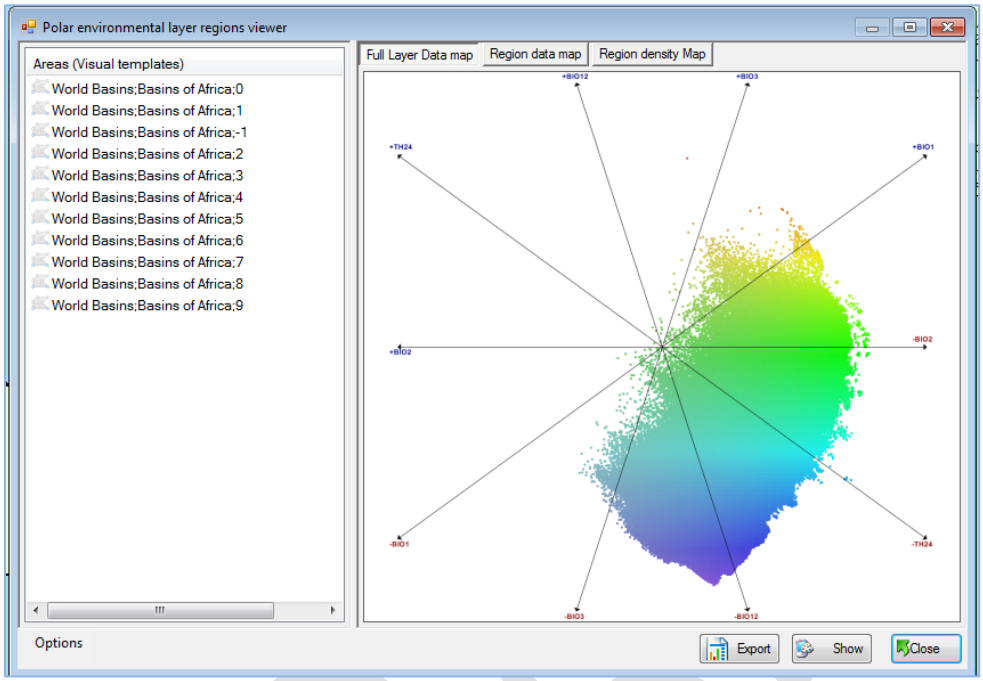
In section 2.11 2DCELs have been explained. Has seen there, a 2DCEL combines data from several environmental variables in a single polar coordinates system. Besides visualizing the whole 2DCEL in the map, ModestR also allows showing only a portion of this layer, corresponding to any geographical area.

To do that, the first step is generating and storing a 2DCEL using the wished variables, as explained in section 2.11. The next step is adding to the map the shapes that correspond to the geographical areas we want to visualize the corresponding portion of the 2DCEL. This can be done using predefined shaped, drawing your own ones, or importing them, as explained in 2.16. For this example, we added to the map the shapes corresponding to the main river basins of Africa:



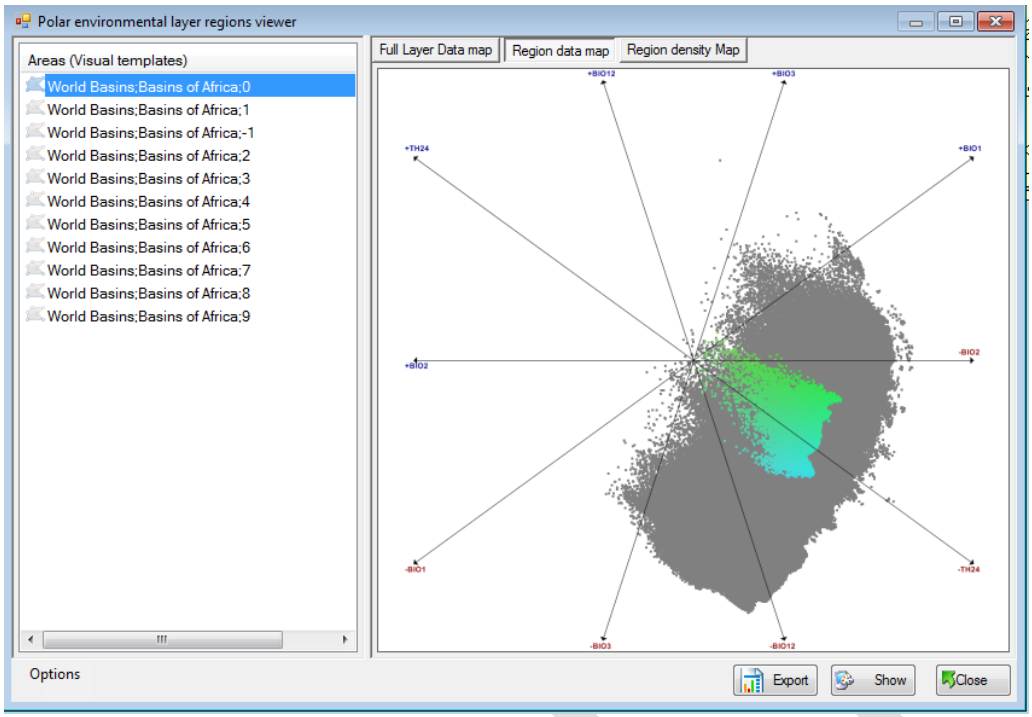
Next, the wanted 2DCEL has to be loaded, going to *Mapping/Niche of occurrence/Load compounded environmental layer* and selecting it in the list of stored layers. For this example, a layer that contains data from land areas has been used.

Next, we'll go to *Mapping/Niche of occurrence/Polar data maps for shapes*. A dialog box will appear.

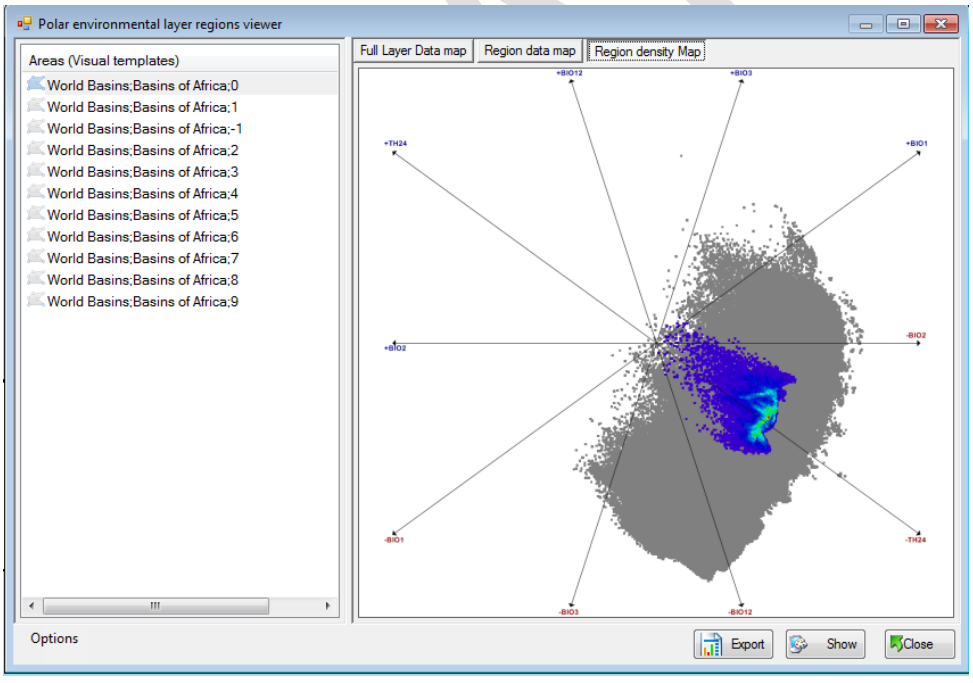



On the left side, a list of the shapes (corresponding to geographical areas) currently added to the map appears. On the right side, a graph will initially show the polar coordinates graph of the whole 2DCEL.

If we select a particular shape on the left list, after some seconds, the corresponding data map of the region will be shown on the right side. The areas of the 2DCEL that correspond to this geographical area will appear in color on this graph, while the rest of the areas appear in grey.

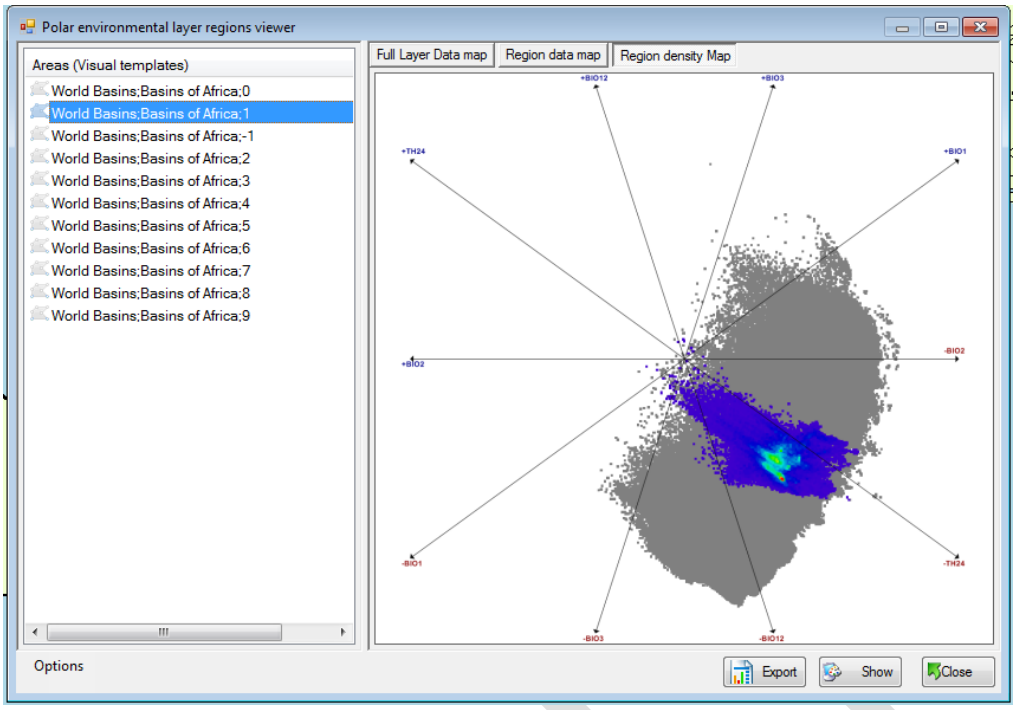



If we select the tab "*Region density map*" on the up-right tabs, we'll see the corresponding density map:




In this density map, the most frequent conditions in the selected region are highlighted in warmer colors than the less common conditions.

If we select another shape on the left list, the corresponding data and density maps will appear. Comparing the maps of different geographical areas can provide information about how different they are in terms of the variables used to create the 2DCEL we are using.




The button *Show* on the bottom-right side allows showing in the world map the graph corresponding to the one that is currently shown in the right panel.

The button *Export* exports to a folder image files for the data and density graphs corresponding to each one of the shapes of the left list.

## 2.18 Other useful features

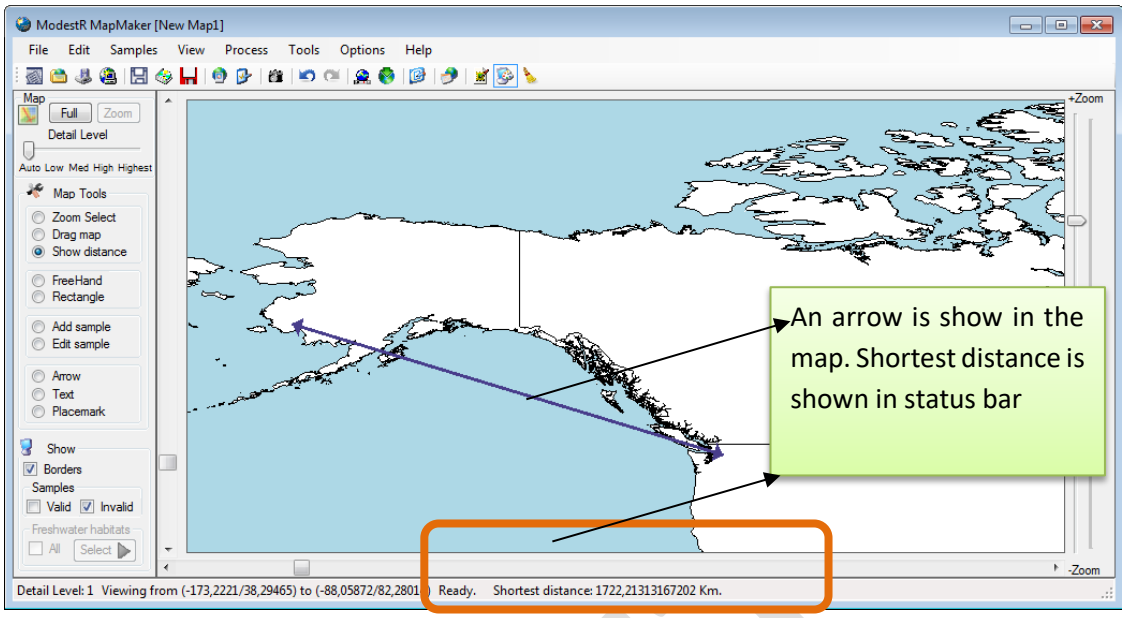
### 2.18.1 Other map elements

Besides occurrences and presence areas, other elements can be added to the map, mainly for informative purposes. On the left tool panel you'll find options to add arrows, placemarks and text to a map. The operation with those tools is very straightforward: you essentially have to select a tool and then click on the map to add the corresponding element. For arrows, you have to click on the starting point and drag until the end point, for example.

Any of those elements can be selectively deleted from a map either by selecting any area (using the *Freehand* or the *Rectangle* tools) and clicking the *Delete objects* button located on the left tool panel. Alternatively, going to the *Edit* menu, where options exist to delete in a single step all the elements of any type (arrows, placemarks…)

### 2.18.2 Querying distance between two points

You can query for the approximate distance between two points in the map using the *Show distance* tool located in the *Map tools* panel. Click on the initial point in the map the drag to the final point. The distance will be shown in the bottom status bar. Take into account that the shown distance is the shortest distance between two points.



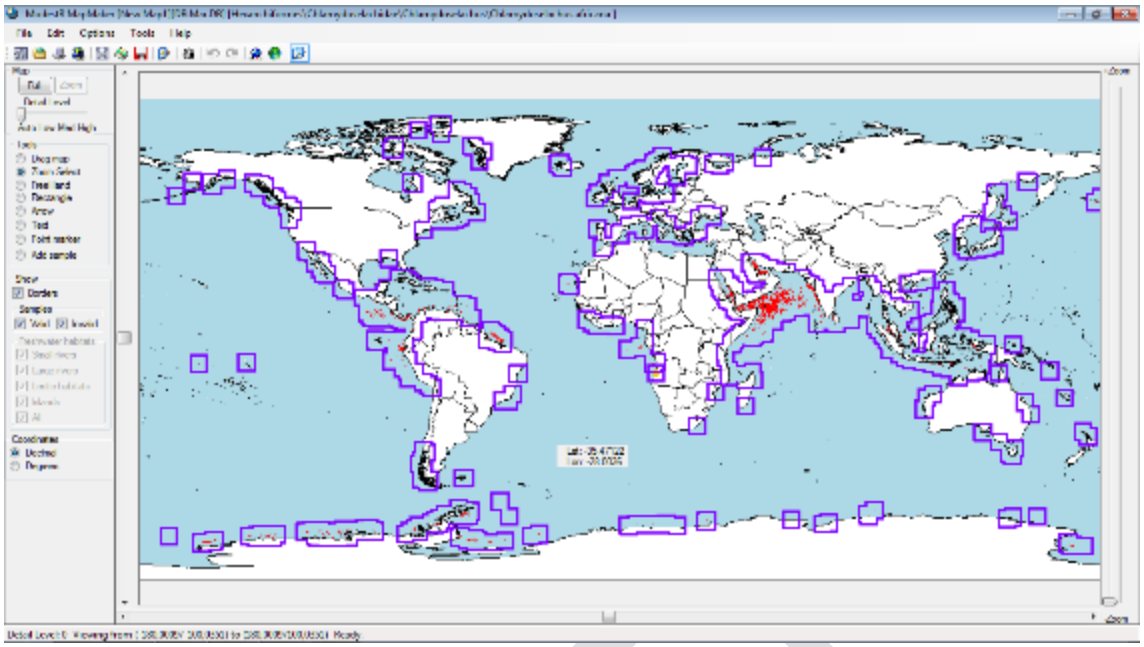

### 2.18.3 Querying a point in OpenstreetMaps or GoogleMaps

On higher zoom view, you can wish to have some help to situate, like geopolitical information. For that you can use several options:

- In the toolbar you will find a button *Open current view in GoogleMaps,* or another *Open current view in OpenStreetMaps.* Those options will open a web browser with a view quite approximate to that you have in MapMaker, using GoogleMaps or OpenStreetMaps.
- If you want to look up a precise point, you can click on this point in the map with the right mouse button, and on the contextual map select *Query current coordinates/in OpenStreetMaps* or *in GoogleMaps.* Those options will open a web browser and use GoogleMaps or OpenStreetMaps to show a marker in the exact coordinates where you clicked on.

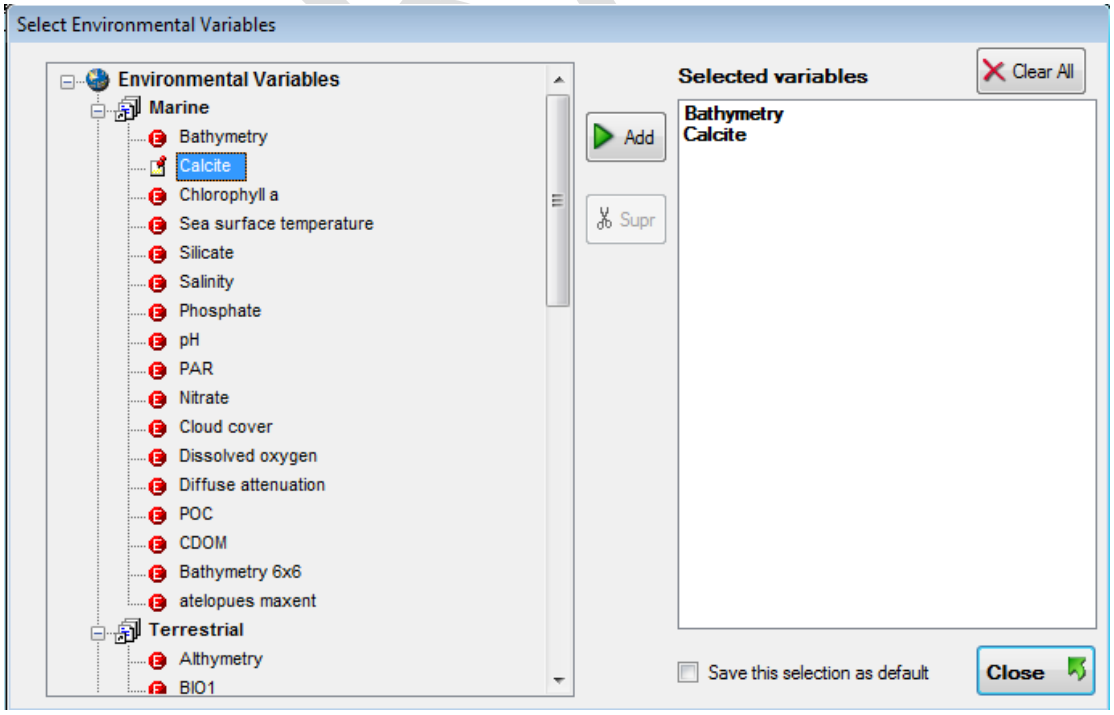
### 2.18.4 Encircling presence areas

Another useful feature is encircling presence areas. When you see a map in full extent or with low zoom level, it is not always easy to visually localize small presence areas or occurrences in the map. To help you doing this, you can ask MapMaker to encircle presence areas, using  the button in the toolbar, or the option *View/Circle presence zones* in the main menu. MapMaker will encircle all areas and occurrences of the map, as shown in the figure, making easy to visually locate all presence areas.

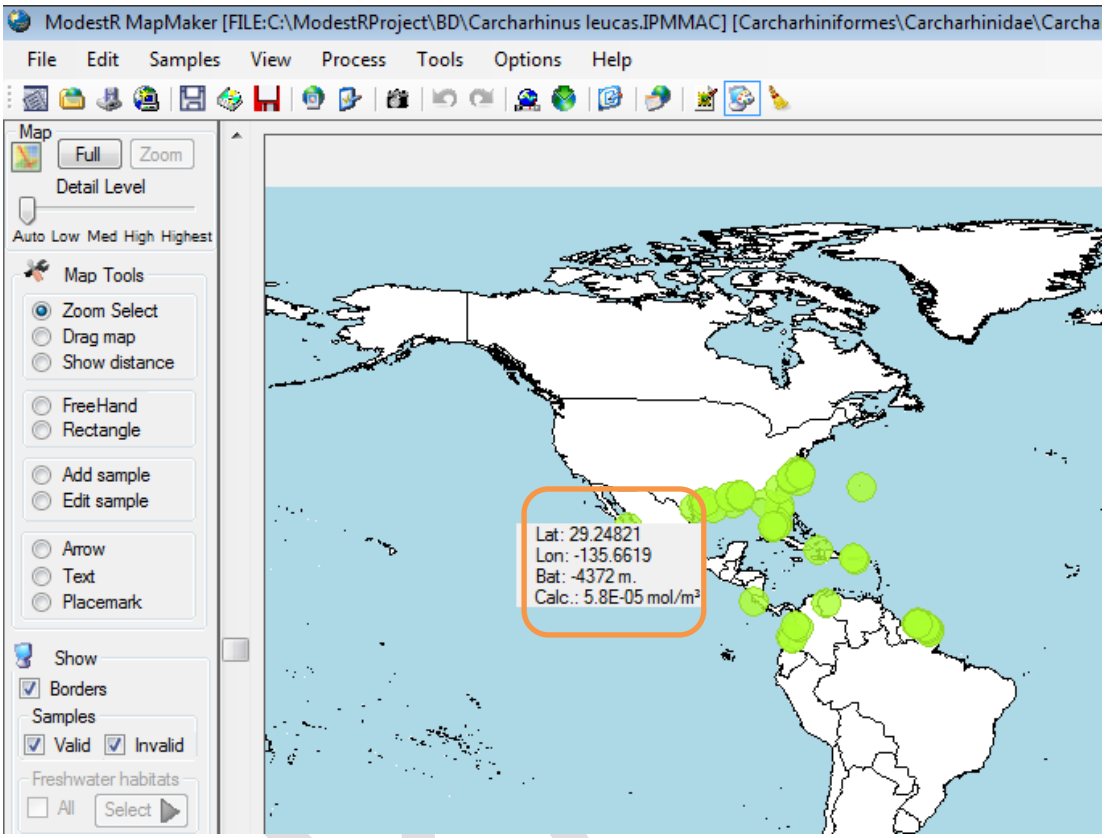


### 2.18.5 Showing environmental data at mouse cursor

If you have environmental variables integrated in ModestR (see section 6.26), you can add some of them to be shown with coordinates on the mouse cursor when moving on the map. To do that, you can use the environmental variables selection button on the toolbar or the menu *View/show environmental variables at mouse* and just select which variables to add in the tree that will be shown. We recommend not to add more that 2-3 variables, to avoid reducing performance.

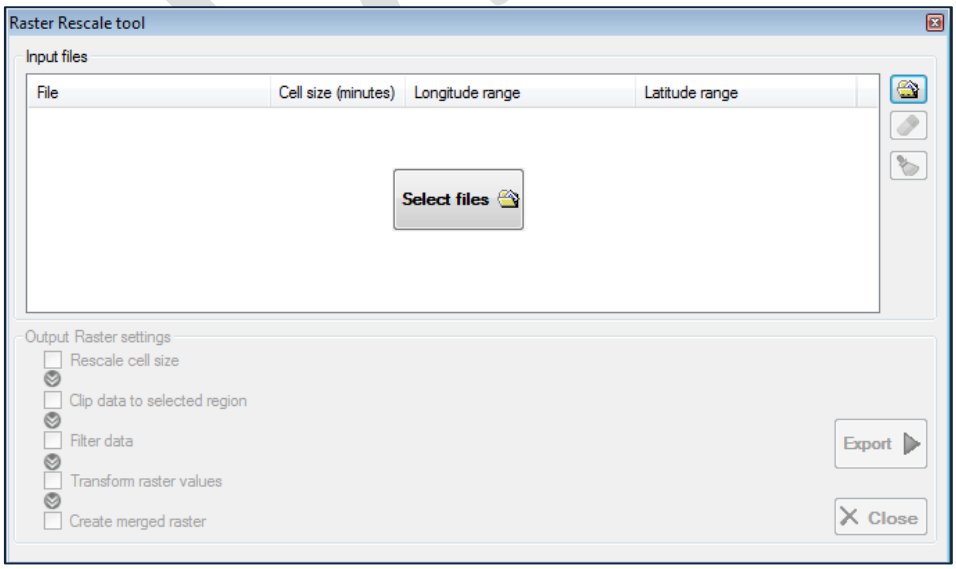


Once selected the wanted variables, they will be loaded and you can see their corresponding value on the mouse cursor when moving it on the map. If the value of one variable is missing for the current coordinates of the map (as it is usual that some variables have only data for sea areas, for example), the value will be empty.




## 2.18.6 Raster transformation tool

MapMaker includes a simple tool to rescale, clip, filter and transform rasters in ASC format. To run this tool, go to menu *Layers/Raster transformation tool* (it is also available from DataManager, in the same menu). A dialog box like the following one will be shown:

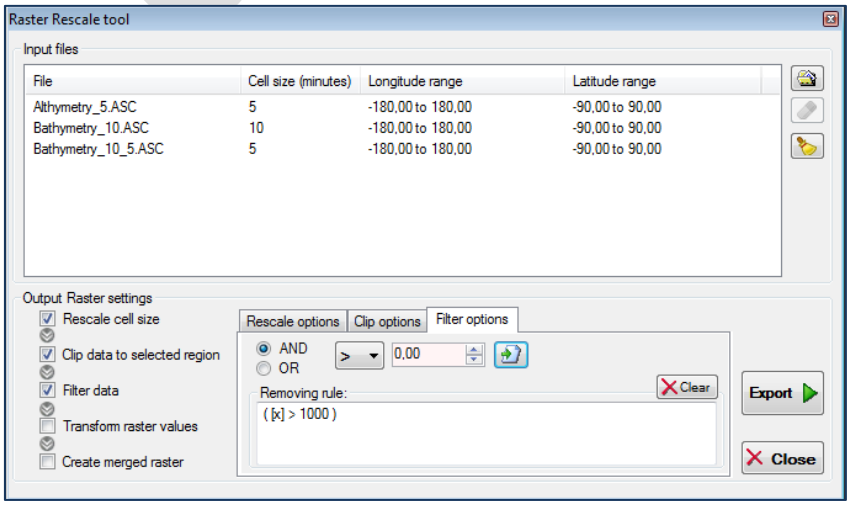


You can add the ASC files to be transformed just by dragging and dropping them on the central area, or clicking on the *Select files* button and selecting them. Take into account that *every* selected file will be transformed according to the operations that you'll select.

Next, in the output settings, you can check which operation you want to apply to selected files. Operations will be applied in the order they are shown in the dialog box, independently from the order you check them. You can check only some operations (at least one), it is not mandatory to check them all. When checking one operation, operation options will be shown in a tabbed panel. Available operations are:

- Rescale cell size: allows changing cell size of a raster. The available settings for this operation are the new cell size and the rescaling mode (mean, maximum or minimum).
- Clip data to selection region: allows clipping a raster to a particular rectangular region that you can define on the settings for this operation. Remember that you can also clip a raster to any arbitrary shape using other Modestr features (see 2.9.4 and 4.12.3).
- Filter data: allows filtering rasters to remove data values that comply with a specified rule. For example, we can add a rule such as *( [x] > 1000 )* to remove all values bigger than 1000 from the selected rasters (the x in the rule represents any value of the raster).
- Transform raster values: allows converting raster values using one of those rules:
    - Transform "no data" to zero
    - Transform any data to a specified value
    - Transform zero values to "no data"

  An *advanced* mode is also available (a vertical left tab allows switching between *Basic* and *Advanced* modes) that allows entering several transformation rules in the form *IF [X]=value THEN [X]=new value*. Those rules are applied in the order they are added to each value of the rasters.
- Create merged raster: merge all selected rasters *after* applying the previously selected transformations to them, in a new raster. This option can be useful to generate a new raster from two or more rasters using a merging operation such as the mean (the values of the merged raster will be the mean of the corresponding values of the selected rasters). The settings for this operation are the merging mode (mean, maximum, minimum, sum, difference) and the rule to apply when one of the rasters to be merged has a null value for some cell.



After configuring all settings, clicking on the *Export* button, each one of the selected input files will be processed and a new transformed ASC file for each one will be generated and saved on the selected target folder.

### 2.18.7 Merging maps

You can merge several species maps in MapMaker, but take into account that MapMaker doesn't keep information of the different species you merge. That is, MapMaker will consider all merged data as a single species map. To make maps that allow several species data to be managed and visualized as distinct datasets, you should use MRMapping instead (see chapter 5). However, this option can be useful to merge several data from a same species, or related species, such as all the species maps of a family for example, visualize and export the results, and apply any of the transformations allowed by ModestR, such as EOO estimation, NOO2D, etc.

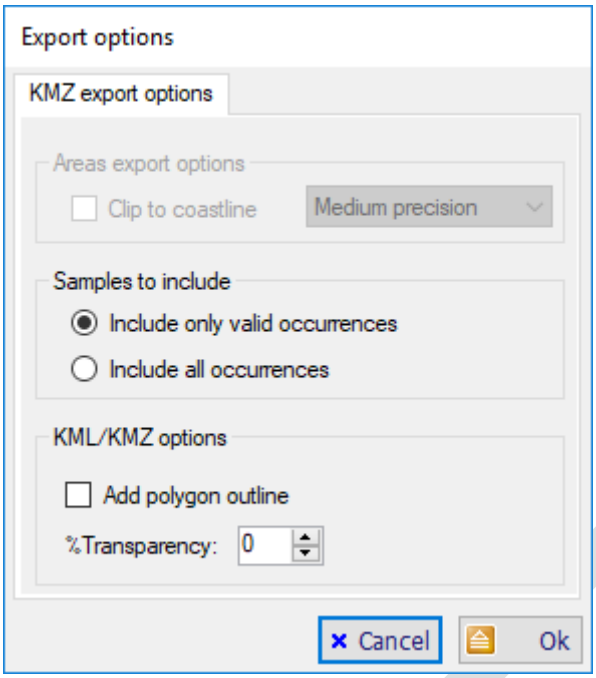
To merge several species maps there are several options:

- You can use menu *File/Merge/Standalone map file* to load a map saved as a standalone map file and merge it with the current map.
- You can use menu *File/Merge/Maps from database* to select one or more maps from the default database and merge them with the current map (default database can be changed going to *File/Select database* or to *Options/Preferences*).
- Another option is copying a species distribution map using *Edit/Copy as ModestR map*, and pasting it latter using *Edit/Paste ModestR map.*

## 2.19 Exporting data and reports

MapMaker importation and exportation formats and settings are detailed in section 7. Here we'll just point out some of the most relevant exportation featres of MapMaker:

- Export map as image: in *File/Export/Current map view as image* you can export the current map view you have in the screen to an image in JPEG or BMP format. Options such as adding a grid, scale, or quality adjust will be shown.
- Export map to other formats: in *File/Export/Species distribution* you can export current species map to other formats such as shapefile, KML, KMZ or ASC raster . Take into account that presence areas in range maps are internally stored the way the user has drawn them. To clip those area to coastline when exporting, you should enable this option.

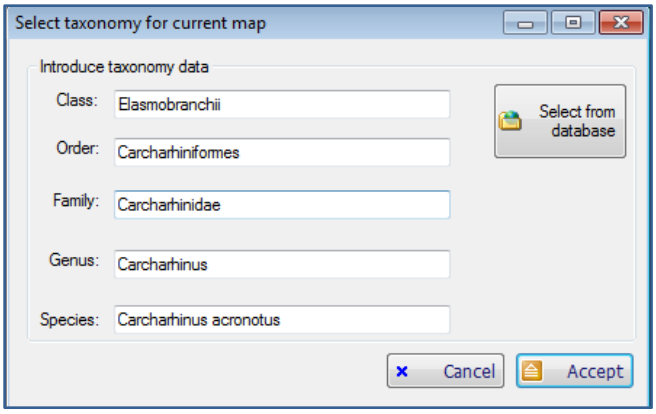



- *File/Export/ Environmental data*: this option allows exporting environmental data clipped to presence areas or shapes. It's explained in section 2.9.4.
- *File/Export/Areas summary report:* generates a report in CSV format with the areas in $km^2$ of each type of habitat where the species is present. A similar report, but for any set of species (not just of one), can be obtained using DataManager. Moreover, in MRMapping a similar report is available for any rank or arbitrary group of species (families, orders, etc…). MRFinder offers a similar feature to calculate the area of any arbitrary selection in the map.

## 2.20 Saving a MapMaker map

A map can be mainly saved in two ways:

- To a standalone map file: this option allows you to save a map in a standalone file, with extension .IPMMAC, which will only contain this map in a specific format based on XML. In this case, taxonomic information about the map is not mandatory, but it will be saved if you provide it. To enter this information before saving a map, you can use the menu item *Edit/Assign taxonomy for current map* or the corresponding button of the tool bar. A dialog box will allow entering taxonomic data or selecting it from a ModestR database (if you have configured it previously).




- To a ModestR database: using File/*Save/To ModestR maps database* option menu, that allows you to save the map associated to a species previously existing in a ModestR database. In this case taxonomic information about the map is mandatory, and you will be asked to provide it before saving the map, eventually selecting it from a tree like the explained on section 2.21. To be able to save a map in a database, you previously have to create this database and populate it with taxonomy. You can also use an existing database, like the provided in ModestR website for fish species.

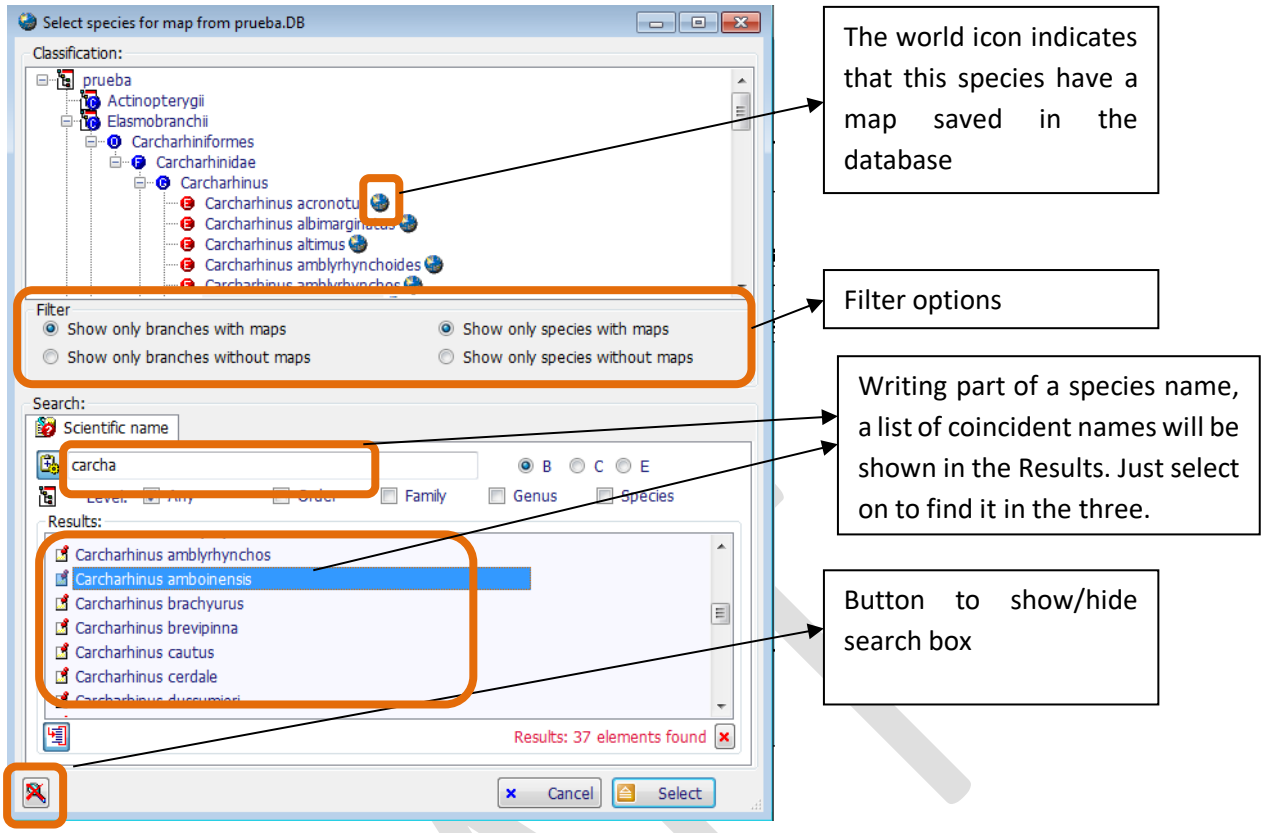
Once a map saved, either in a file or in a database, you can save it anew just by selecting the *Save* option, without having to reenter any data. *Save As* options allows you to save it in a different file.

Apart from those options that allow saving a map in a format directly readable by ModestR software, you also have several options for exporting a map to other formats in the *File* menu, such as JPEG/BMP format, shapefile, CSV data, etc. The exportable formats are explained in the section 0 of this tutorial.

## 2.21 Retrieving a map

You can retrieve a map both from a standalone file, with extension .IPMMAC. or from a ModestR database. In the first case, you just have to select the file that contains the map. In the second case, Mapmaker will display the taxonomy tree of the database (either the default database configured in the *Options/Preferences* menu, or the database you selected if no database has been selected as default).

You can directly expand nodes with the mouse and select a species with a map, which is identified with a little world icon at the right of the species. To help you finding a map, you can use the filter options that allow showing only species with maps, or only branches were there are species with maps, for examples. You can also use the search box, than can be displayed using the magnifying glass button on the bottom-left corner. In the search box you can write some part of the species you are looking for, and coinciding names will be shown in the *Results* list. Once selected a species with a map in the tree, press *Select* button to load it in MapMaker.

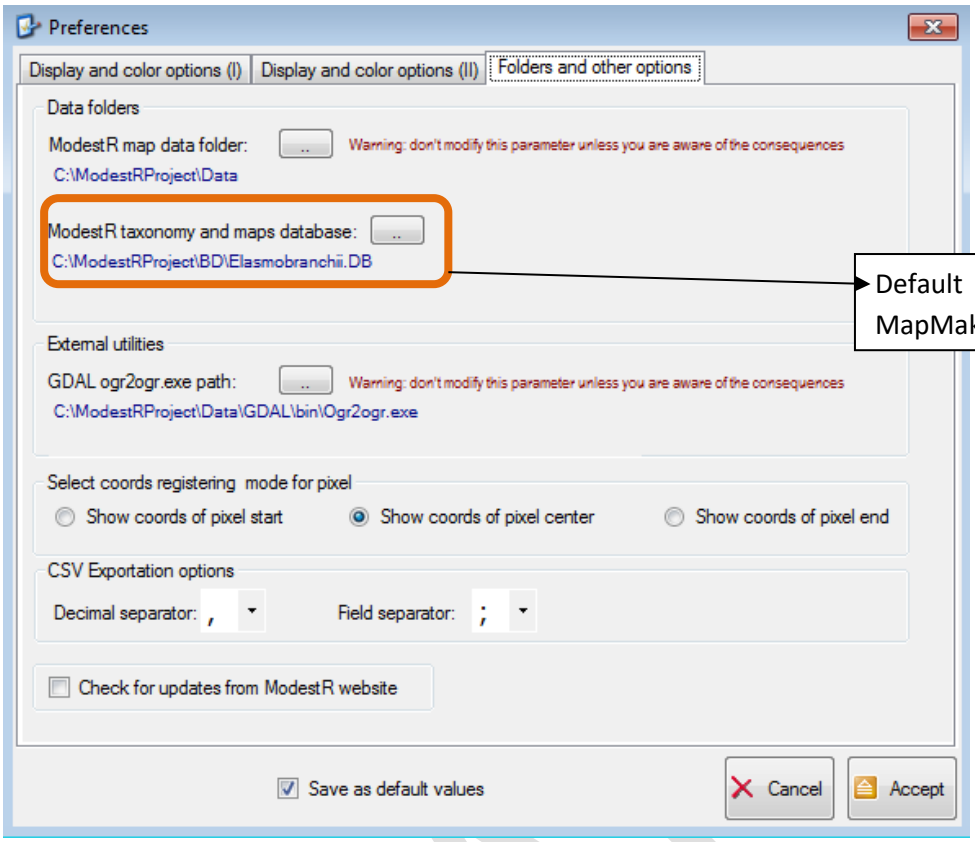



Once loaded, you can for example modify a map in MapMaker, and save it again. This way you can work on a map in different sessions.

## 2.22 Setting the default ModestR database and other preferences in MapMaker

In the previous subsections regarding MapMaker, we frequently mentioned ModestR databases. MapMaker will usually have one ModestR database configured by default, and that will be the database it will use when you select open or save a map in a database. This database can be selected in the *File/Select Database* menu item or in the dialog box that is shown selecting *Options/Preferences* menu. In this dialog box you can also set the colors for each element of the map, each kind of selection, each type of occurrence (dynamically/manually validated, and valid/invalid), and transparency percentages for some elements like shapes, occurrences or selections. Also the elements that will be visible by default in the map. If for example you don't work with freshwater ecosystems, you can unselected them to avoid MapMaker having to load and paint them, thus making it works faster.

You also have an option to select the data folder where MapMaker will search world map data. It is important that you **don't modify this parameter**, unless you have moved map data to a different folder than the default one proposed in the installation process.




Default database for MapMaker

# 3 First steps in DataManager

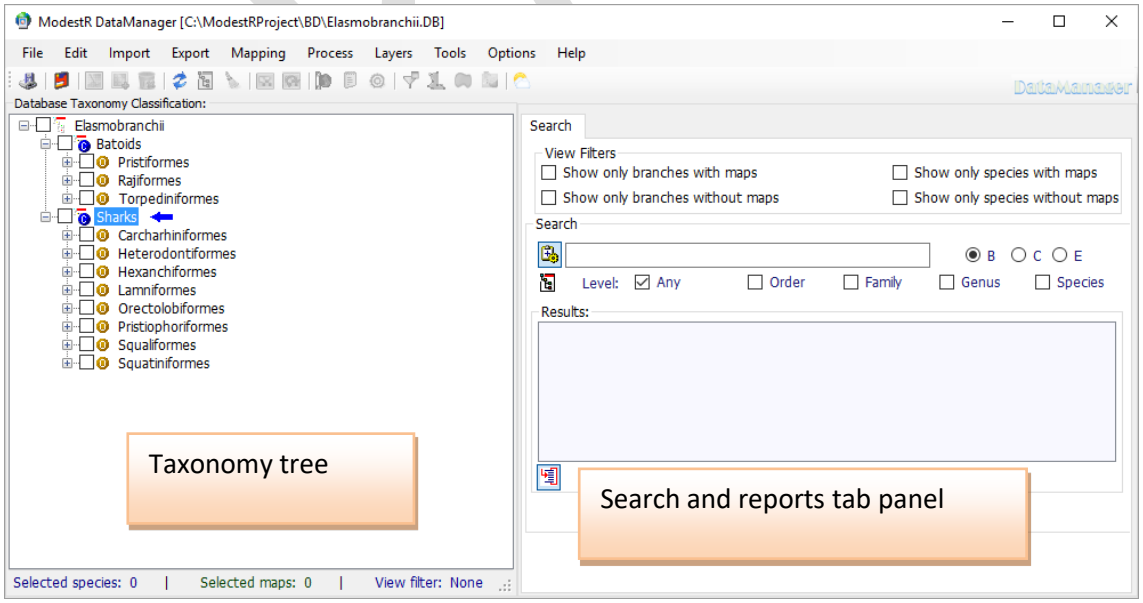
DataManager is another application of ModestR software. It is aimed to create and manage ModestR databases, which are databases containing taxonomic data and associated distribution maps. You can create a new database, or just use existing ModestR databases, like the Ipez database you can find in the ModestR website. Many of the features available in MapMaker are also available in DataManager. The main difference is that while MapMaker is designed to work with a single species distribution map, DataManager is meant to work with many maps at the same time. For example, in DataManager you can download data from GBIF or apply a data cleaning for a whole class of species in a single operation, while in MapMaker you can only do it species by species. Furthermore, DataManager can perform analyses and generate reports concerning many species, such as richness metrics, latitudinal gradients, etc. In this way, DataManager is a very powerful tool to perform studies involving hundreds or thousands species in a very simple way.

In this section we'll briefly explain the main steps to work with DataManager.

## 3.1 DataManager user interface

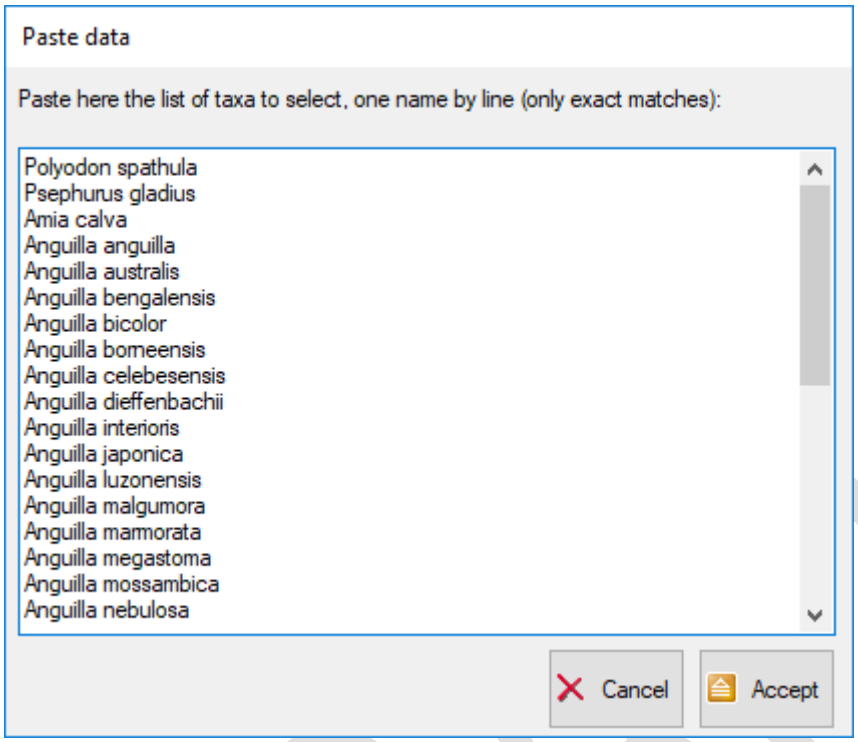
DataManager user interface is easy to use. When you open DataManager, it will try to reopen the last used database, if one. Otherwise, it simply show an empty panel. Anyway, you will see a window with a main menu, and two main parts:

- On the left half, a taxonomy tree where the taxonomic data, in a tree view, is displayed. Of course, if you don't open a database previously, no information will be displayed here.
- On the right half, a tab panel where initially, a search and filtering page is displayed. Other pages can be displayed in this tab panel, depending on the features you use.




It can be remarked that there is a checkbox on the left side of each element of the tree, to allow the user to select it. For most of the operations, you'll have to first check the nodes of the tree to be used in the operation. If you check an order, for example, all species included in this order will be used. To apply an operation to all the maps/species included in a database, you can just check the root node of the tree.

Besides selecting nodes in the species tree, you can also select any set of taxa directly entering a list of names. To do that go to *Edit/Select nodes using taxa list*. Then you can write or paste a list of taxa of any level (class, order, species…) in the dialog box and click on *Accept.* A warning will be shown if some of the introduced taxa are not found.

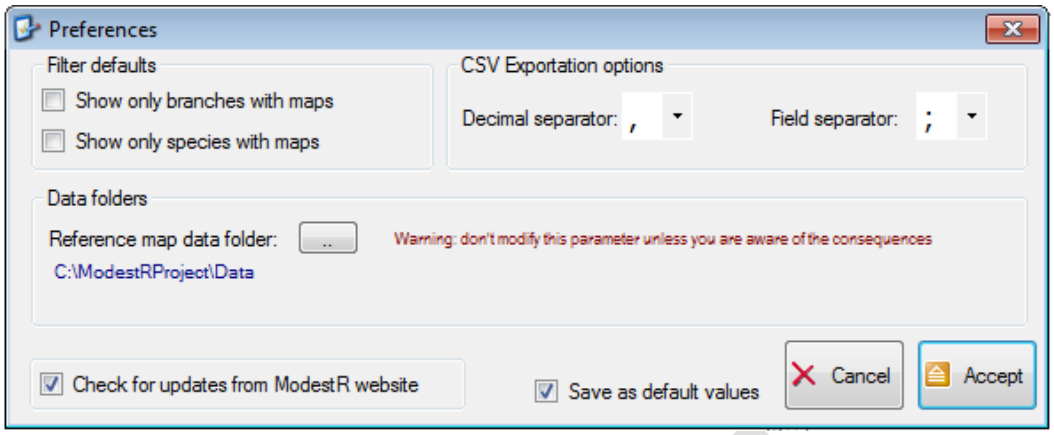



As in other ModestR applications, quite all menus, buttons and other elements have a tooltip that briefly explains their usage, and that will be shown when you place the mouse over an element. Moreover, when some list is reported on the right panel (for example, a list of exported/note exported maps), you can almost allways copy the contents of those lists to the clipboard, by clicking on them with the right button and selecting the *Copy to clipboard* option that will appear on a contextual menu.

## 3.2 Default settings in DataManager

You can set some default settings of DataManager in the *Options/Preferences* menu option. Some filter defaults, CSV format options, and other ones, can be configured.

You can also modify the data folder where DataManager will search for the reference world map, but unless you have modified the default installation options, it is unnecessary and can provoke DataManager not to run correctly.



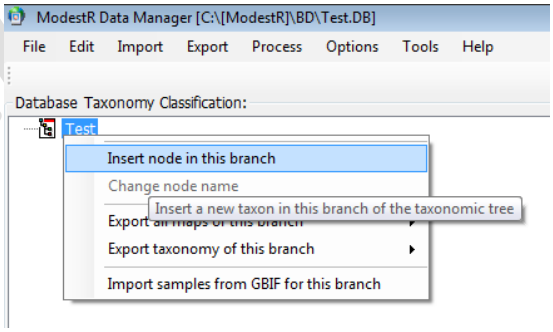

## 3.3 Creating a new database

To create a new ModestR database, you just have to use *File/New Taxonomy database* menu option. Then select the name and folder to create the database file. The ModestR database will be stored in a single file with the name you entered. So if you want to copy, send or delete a ModestR database, you just have to do it on that file. DataManager will not allow to create a database overwriting an existing file; this is to avoid accidentally overwriting files.

The new database will be created, with an empty taxonomic tree. The taxonomic levels of a ModestR database begin at the Class level, then Order, Family, Genus and Species. That is, there are 5 levels (without taking into account the root node).

To populate the database with taxonomic data, you have several options[13]:

a) Manually create each node. To do that, select a node in the tree (initially the root node) and use *Edit/Insert node in this branch* menu option, or the analog context menu item (context menu can be displayed clicking on the node with the right mouse button).
In this way you can add child nodes at the different levels of the taxonomic tree, but of course is a quite tedious option.




b) Import taxonomic data from a CSV file (the most recommended), a phyloXML file or another ModestR database. This option allows you to import existing data, maybe previously exported from other databases or catalogs. The accepted import formats are explained in sections 6.1, 6.2, 6.3. This is the preferred way to easily populate a ModestR database with your already existing data. For test purposes, you can find a CSV file with a comprehensive taxonomy of fishes, ready to be imported to DataManager, in the ModestR website (in the website go to section *Help&Doc* and search for the *Occurrence file Ipez.csv*).

[13] For test purposes, you can also download a occurrence ModestR database from the Modestr website. There are some examples available in the section *Download/Data files* of the website.

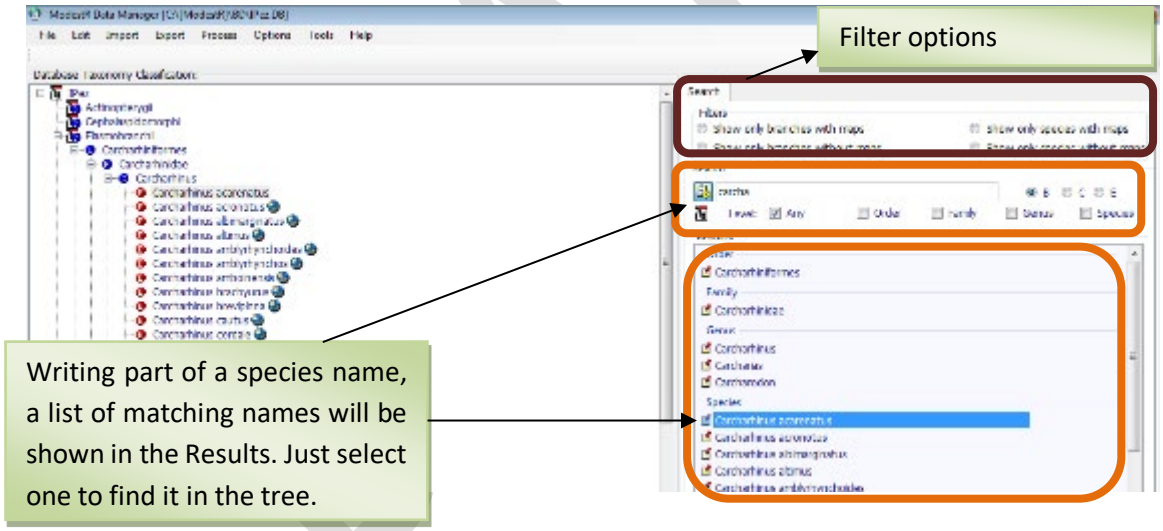
c) Import taxonomic data from ITIS database (http://www.itis.gov/). This option is explained in section 3.19.

If you only have a list of species that you want to import to a database, but not the corresponding higher levels (class, order, etc..) you can be interested in the taxonomy search tool integrated in DataManager, explained in 3.20.

Nodes can also be cut and pasted, deleted or renamed, using the corresponding options in the *Edit* menu or in the context menu displayed clicking on a node with the right mouse button.

Once populated, you will see the taxonomic data organized in a tree where you can expand or collapse the different levels clicking on the corresponding nodes. Species with a map (of course, in a newly created database no one will still have maps), are identified with a little world icon at the right of the species, like in MapMaker database view (see 2.21). To help you finding a taxon of any level, you can use the search page on the right. You can write some part of the species you are looking for, and coinciding names will be shown in the *Results* list. Clicking on one of those results, it will be selected in the tree.

When you will have maps in the database, to help you browsing contents, you can use the filter options that allow showing only species with/without maps, or only branches were there are/don't are species with maps, for examples.




## 3.4 Adding and managing maps in a database

Once a ModestR database created and populated with taxonomy, the purpose of the database is, of course, to store distribution maps associated to the species. You will have different options to do that:

- You can simply create new maps in MapMaker and directly save them in the database. To do that, you have to configure the database as the default database in MapMaker, and use the *Save/To ModestR maps database* menu item in MapMaker to store a map in the database, as explained in sections 2.20 and 2.22.
- You can import already existing maps made with MapMaker, either saved in standalone files (see section 6.5) or from another existing ModestR database (see section 6.6)

- You can create new maps by downloading data from GBIF database (Global Biodiversity Information Facility, http://www.gbif.org) or importing them from CSV files (see section 6.7).
- You can import maps from ESRI ASC probability distribution models such as the maps generated with Maxent software (see section 6.10).
- You can import distribution maps from shapefiles. DataManager is able to import a collection of maps from a set of shapefiles, or from a single shapefile which may contain several maps (such as shapefiles available in IUCN website, for example). More details about this feature can be found in section 6.11.

Of course, those options can be also used to overwrite existing distribution data with new data, depending on the importation options you use in each case. You can optionally use some of those options for adding new data to already existing maps in the database.

The difference between some of those importation options and the similar ones in MapMaker is that in DataManager they can be used to import a collection of maps at once (for example from GBIF or CSV files), whereas in MapMaker they can only used to import one single map.

After importing maps, usually a report will be shown on the right tab panel of DataManager, where imported maps and eventual errors will be reported.

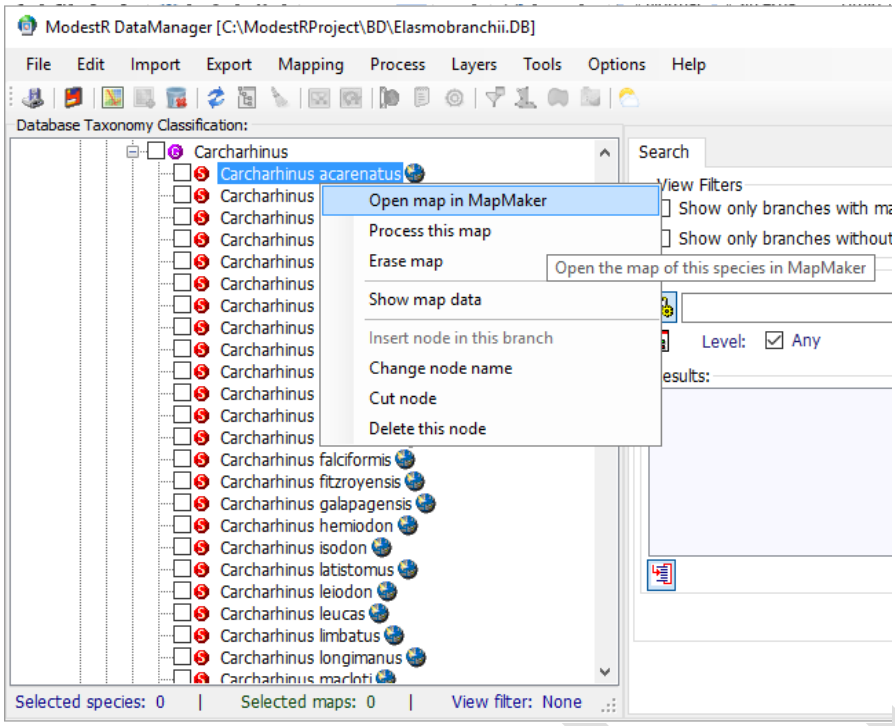
Maps can be individually deleted, simply by selecting them in the tree and using *Erase map* option in the context menu. To delete several maps you can check the corresponding nodes in the tree and use the menu option *Edit/Erase checked maps.* Be aware that if you delete a node of any level (for example if you delete a family), all maps directly or indirectly descending of this node will also be deleted.

In a ModestR database a map created and linked to a species A cannot be directly moved to a different species B. If eventually you want to do that, you should have to open the map in MapMaker, assign it to another species B (see 2.20), and save it to the database. Then you will have a copy of the map now associated to a different species B. You can finally use DataManager to delete the first map, which was linked to species A.

## 3.5 Opening a map from DataManager

Once a map is stored in a ModestR database, you can open it and modify it in MapMaker. To do that, you can just select the species in the taxonomic tree and use *Edit/Open selected map in MapMaker* menu option, or *Open map in MapMaker* option in the context menu.

Of course, you can also independently open MapMaker and load the map from the database, as explained in 2.21.



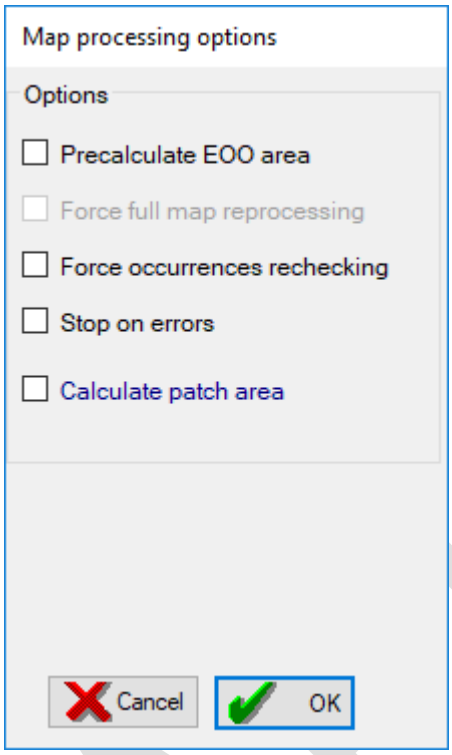

## 3.6 Processing maps

ModestR maps are basically stored in an XML-based format, either in a database or in a standalone file. This XML file contains a declarative description of the taxonomy, presence areas, occurrences and other elements of the map, but it is not directly usable to calculate some results such as presence matrix, occupied area, or metrics such as richness. That is why it is necessary to process those maps in order to *rasterize* them. That is, to convert those declarative data to a matrix of presence.

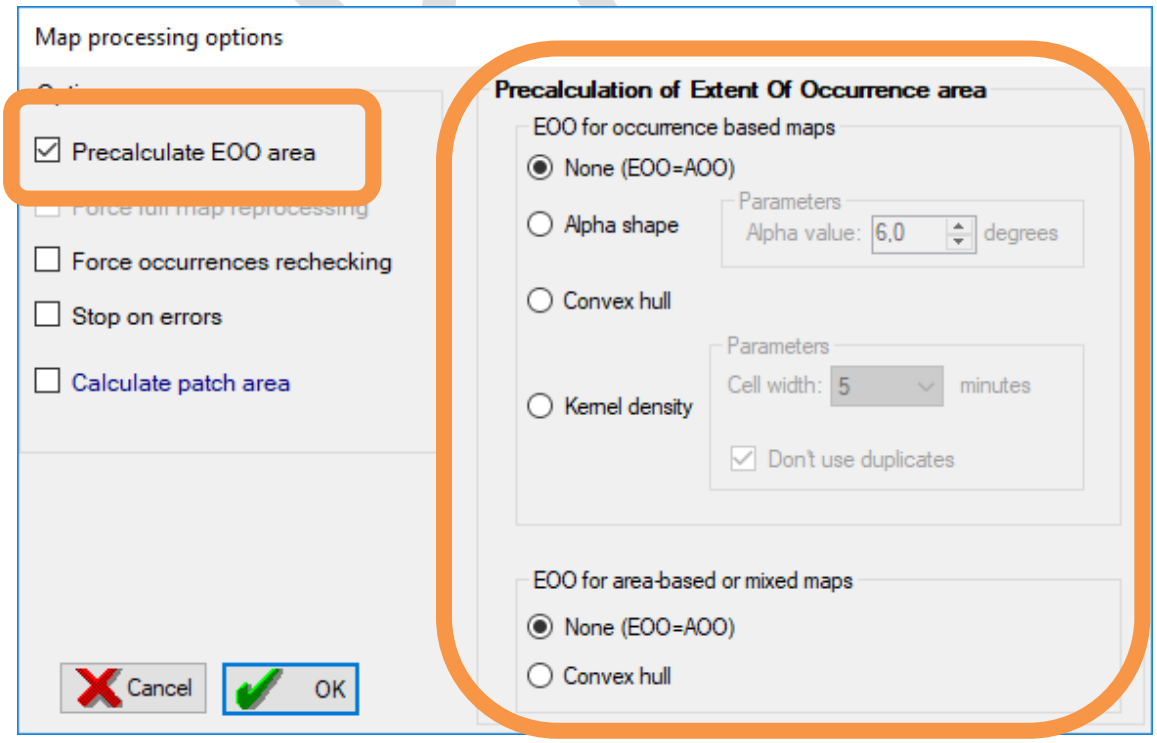
DataManager can only process maps that are stored in a ModestR database, but not those saved as standalone files (in this case you have to import them to a ModestR database). Processing feature converts each map to a presence matrix of 1'x1' cells, and stores it in the database, to avoid doing that conversion each time you need to calculate some data. Anyway, if you modify a map, or if you update the reference world map (for example when installing some ModestR updates), you should reprocess maps. In any case, DataManager will warn you about that when required.

To perform this processing, you should check the wanted branches (you can just check the root node to process all pending maps in the database) and use *Process/Process pending checked maps* menu option. That will checks and process all checked maps that need to be processed.

A dialog box will appear showing the processing options, that we describe below:




- **Precalculate EOO area**[14]: during map processing ModestR can estimate and calculate Extent of Occurrence area of each species from its current distribution map. This is a time-consuming task, therefore it is optional. If you select this option, more options will appear on the right where you can configure how this area is calculated.




[14] This EOO calculation during map processing doesn't modify the maps in any way. It's just done to obtain EOO area estimation. It's totally different from applying a hull or niche transformation to maps, as explained later in this chapter.

There are different options for occurrence maps and for area maps (range maps). Maps that contain both occurrences and areas area called mixt maps and are treated as area maps.

EOO options for occurrence maps are:

- **None**: in this case no EOO estimation is done, and EOO is merely assumed as equal to AOO (Area Of Occupancy). As assumed area of occurrences in ModestR (see section 1.1) is just a convenient estimation but not a precise value, this option is not recommended in most of the cases.
- **Convex hull:** EOO is calculated as the convex hull area of the occurrences. If there aren't not enough occurrences to calculate convex hull, it is assumed that EOO=AOO. More details about convex hull can be found in section 2.8.
- **Alpha shape**: EOO is calculated as the alpha shape hull area of the occurrences. If there aren't not enough occurrences to calculate alpha shape hull, it is assumed that EOO=AOO. More details about alpha shape hull can be found in section 2.8.
- **Kernel density estimation:** EOO is calculated as a hull area based on the density estimation (kernel smoothing) of the occurrences. If there aren't not enough occurrences to calculate density estimation hull, it is assumed that EOO=AOO. More details about density estimation hull can be found in section 2.8.

EOO options for area and mixt maps are:

- **None**: in this case no EOO estimation is done, and EOO is merely assumed as equal to AOO (Area Of Occupancy). This option is the recommended in most of the cases for area maps, because unlike occurrence maps, in area maps we can often assume that the presence data (AOO) already define the EOO.
- **Convex hull:** EOO is calculated as the convex hull area of the areas plus occurrences if any. If there aren't not enough occurrences to calculate convex hull, it is assumed that EOO=AOO. More details about convex hull can be found in 2.8.

  Regarding the possible inaccuracy of convex hulling with areas pointed out in section 2.8.1, it should be made clear that this problem will only appear in MapMaker when the user applies this feature manually. When processing maps in DataManager the maps are rasterized and only real presence area is took into account. Therefore in DataManager convex hull area will be accurately calculated.

  This option can be useful for mixt maps or for maps which we know that presence areas don't represent the EOO, so we want and estimation of the EOO to be made, but it should be used with caution and understanding well its operation.

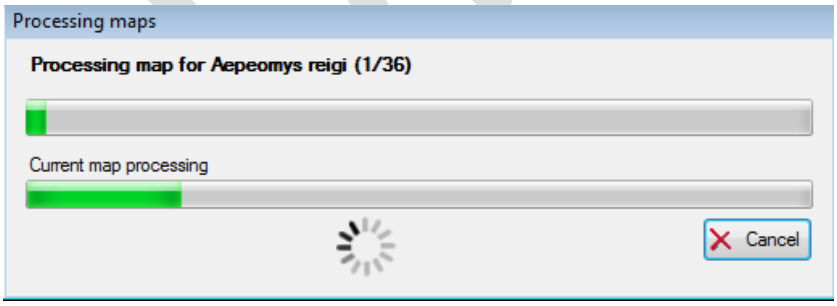
- **Other options:** besides the options described above, there are other options for map processing:
  - **Force full map reprocessing:** this option forces ModestR to fully reprocess maps. This option is not usually required, as ModestR automatically check which data has to be recalculated and which one has not been modified when

processing maps. But if some external modification has been done on a map, or some error has corrupted data, by checking this option you can ensure that maps will be correctly processed and updated.

- **Force occurrences rechecking:** this option forces ModestR to fully recheck the the habitat validity of all occurrences in the maps to be processed. This option is not usually required, as ModestR automatically check occurrences when needed. But if some external modification has been done on a map, or some error has corrupted data, you can ensure that occurrences will be rechecked by checking this option.
- **Stop on errors:** to allow you to process maps in an unattended mode, by default DataManager will not stop when it encounters a problem when processing a map, but it will continue with the next one. At the end of the process, a report with the encountered errors will be displayed. But if you prefer to stop immediately on any error, you can check this option.
- **Calculate patch area:** enabling this option, patch area and patch index are calculated for each especies (those metrics can be exported as seen in 6.17). Otherwise, those data are not calculated and then cannot be used later. This option makes map processing dramatically slower, so if you don't expect to need those data, you can disable this option to make processing quite faster. Anyway, if you disable this option now, you can reprocess maps later and enable it to force patch area calculation (use *Reprocess all maps* menu option)

It is important to point out that map processing can require a large amount of time, particularly for range maps (usually less time for occurrence-based maps). In addition, we recommend to avoid using any other ModestR application (MapMaker or MRFinder) while processing maps, to avoid resources conflicts.

During map processing a dialog box will appear showing the process progression. You can cancel the process in any moment without losing the work done until this moment. This way you can continue processing remaining maps later.




When the processing task has ended or has been canceled, a list of processed and not processed maps will be displayed on the right panel of DataManager. If there are not processed maps, it can be due just to a user cancelation, or to another kind of error that will be shown in the list, aside the species map name.

Even if usually DataManager checks for the necessity of processing maps and only process those who need it, the option exists of forcing maps reprocessing using *Process/Reprocess all checked*

*maps* menu item. This can be useful if for some reason you manually made modifications in map data, such as forcing maps to be considered as updated (see section 3.17), or in the world reference map data (which, of course, is not recommended), the way that DataManager cannot be aware of them; or if some unexpected error happened that could have corrupted data.

## 3.7 Exporting data

The utility of DataManager, besides creating and managing taxonomy and maps databases, resides in the calculation and data exportation features it offers, which are gathered in the *Export* menu option. Remember that you have first to check the nodes of the tree that will be used in any export operation. Otherwise options of the *Export* menu won't be enabled. You can select the root tree node to select the whole database.

The main exportation options of DataManager are:

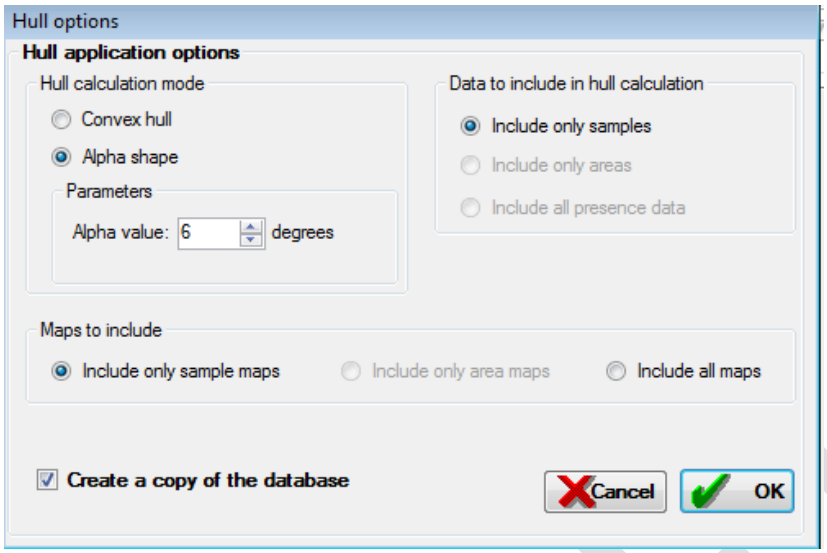
- **Taxonomy data exportation**: you can export all or part of the taxonomy data to a CSV file, as explained in section 6.12.
- **Maps exportation**: you can export species distribution maps to standalone map files, or to presence data in different formats (pseudooccurrences, valid occurrences, ASC rasters…), optionally including environmental variables. You can also export maps as JPEG images (optionally including a kernel density graph), to a Darwin Core style file, or generate a summary report.
- **Metrics and other data exportation:**
  You can also export metrics data and environmental variables. DataManager uses rasterized maps data to calculate presence, richness and other metrics (rarity, patch index…) with a specific resolution. Of course, this option requires maps to be previously processed as seen in section 3.6. Complementary to those data, you can export data coming from environmental variables data files. Options and generated output are explained in section 6.17.
  You can export maps data to Maxent or to RWizard applications[15]. Those options are explained in sections 6.13 and next ones.
  You can also export environmental data (or any other raster data) cut off to species EOO, for any rank of species. This option is explained in section 6.19.
  You can also exports latitudinal gradients for several metrics. Options and generated output are explained in 6.20.
- **Overlapping reports:**
  You can export data about geographic or environmental overlaps between any rank, from species to classes. Those features are similar to those explained for MRMapping in sections 5.8 and 5.9. The main difference is that in DataManager you can select the maps to be used directly from the whole database and group them by rank.

## 3.8 Applying convex/concave hull transformation

As seen in section 2.8, MapMaker is able to calculate the convex hull or the alpha shape area from the presence data in a species distribution map. But if you want to apply this

[15] RWizard applications are plug-in applications based on RWizard environment and the statistical package R and. Those applications will be available online in the ModestR or the RWizard websites.

transformation to numerous maps, it may be easier to do it in a single operation from DataManager in turn of doing it for each map in MapMaker. You can just check the tree branches to which you want to apply it, and select *Mapping/Apply hull transformation to checked maps /Convex-Alpha shape hull* menu option. A dialog box will be shown:




The available options are:

- **Hull calculation mode:** you can choose between convex hull or alpha shape. For alpha shape, you can select the alpha value. More details about those hull modes can be found in 2.8.

- **Data to include in hull calculation:** this option allows to select which presence data from the maps will be used for hull calculation. The possibilities are:
  - Include only occurrences: in this case presence areas, if any, are excluded. This is the only enabled option for alpha shape.
  - Include only areas: in this case occurrences, if any, are excluded.
  - Including all presence data, both occurrences and presence areas.

  Options that involve using presence areas to calculate hull require having all maps processed, which is not necessary if you choose to only use occurrences.
  Depending on the selected hull mode, some options can be disabled. For example, you cannot include areas when applying alpha shape.

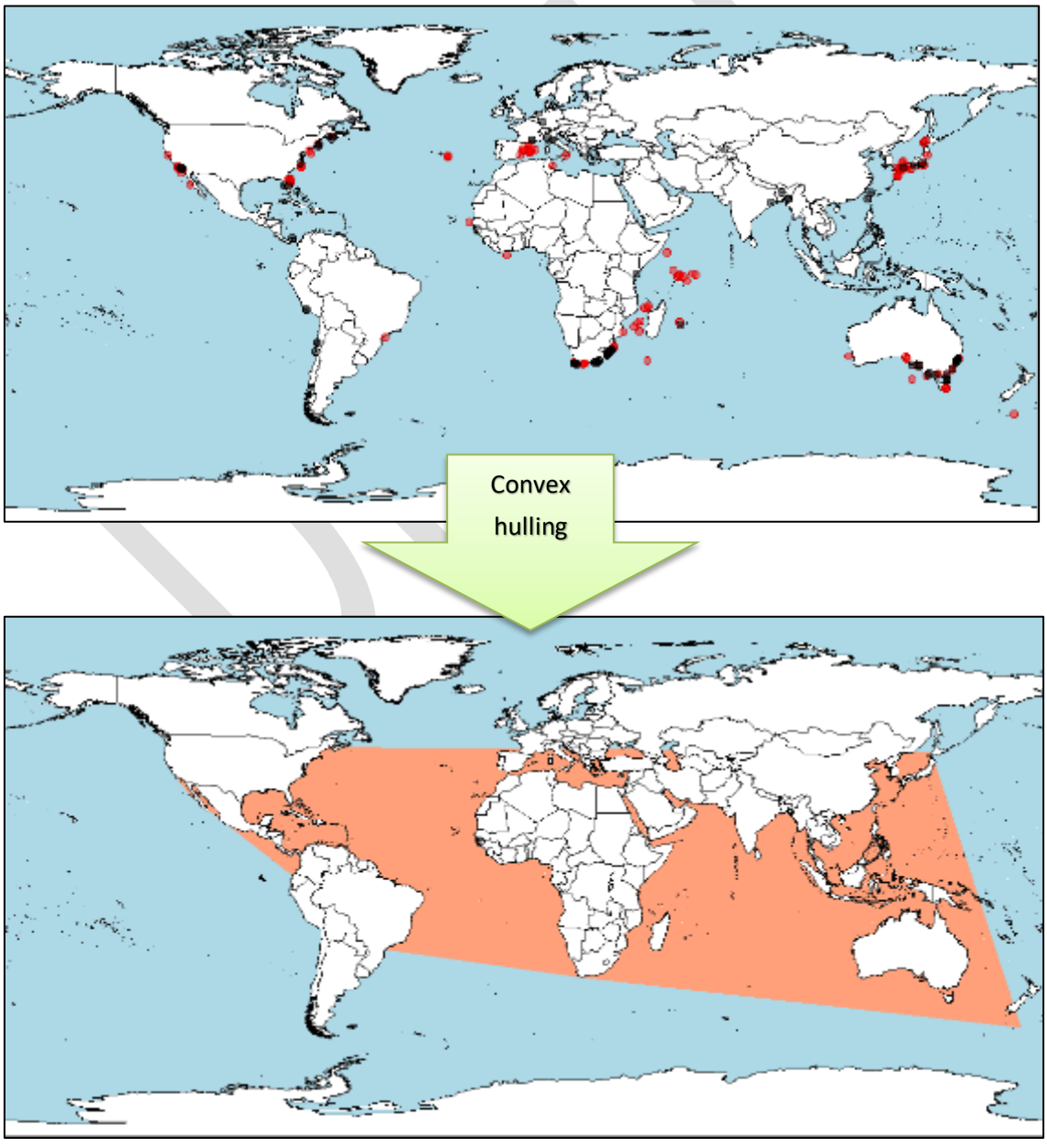
- **Maps to include:** you can select to include all maps, only occurrence maps (maps that only contain occurrences, but no presence areas), or only area maps (maps that only contain presence areas but no occurrences). Some of those options will be automatically disabled according to the selected hull calculation mode. For example area maps cannot be included when using alpha shape transformation.

- **Create a copy of the database:** as hull transformation involves irreversible map modification, it can be prudent to make a copy of the database before performing this transformation. Checking this option DataManager will ask you for a destination folder and file and will do this backup copy before modifying maps.

Hull area that will be added to the map will use as valid habitats the same ones that are currently accepted in the map. For example, if you apply hull transformation to a map that contains occurrences that have sea as valid habitat, the corresponding hull area will also have sea as valid habitat.

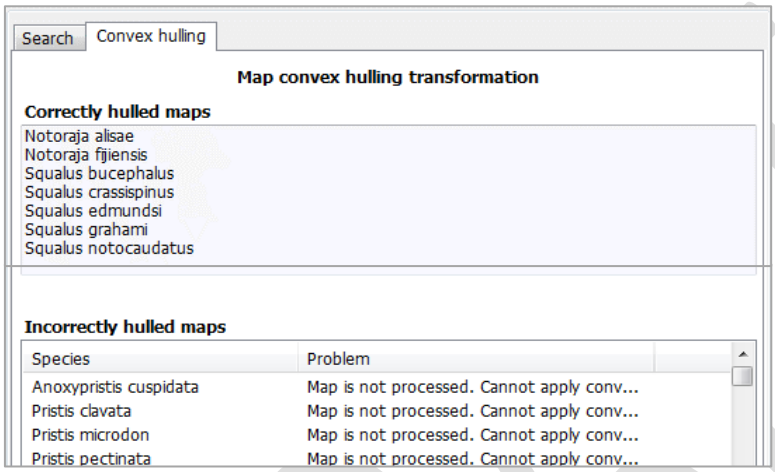
It should be stressed that unlike in MapMaker, where convex hull can only be calculated using occurrences but not areas (as explained in section 2.8), in DataManager it is possible to include areas when calculating convex hull, because it will always use rasterized data to calculate hull (therefore data where presence areas have been converted to a matrix and constrained to correct habitats for a species).

As an example, below we show a map before and after applying convex hull transformation in DataManager (occurrences have been hidden in resulting map). Has it can be seen, convex hull area is added to the map with the sea as valid habitat (so other ones such as land are excluded) because the previously existing occurrences have sea as valid habitat.



After applying hull transformation, a report will be shown in the right panel indicating the correctly transformed maps and also the incorrectly processed ones, which were not updated. A brief message will report the encountered problem that prevent processing each incorrectly processed map. The most usual problems are that the map was not processed yet (see section 3.6), or that it doesn't contains data enough to calculate hull. At least three occurrences are required to calculate convex hull, and at least four for alpha shape (but depending on their relative distances it can be impossible to calculate alpha shape).

Maps excluded because they are not of the selected kind will also be listed. For example range maps will be listed as excluded if we selected to apply alpha shape hull, which can only be applied on maps with occurrences.

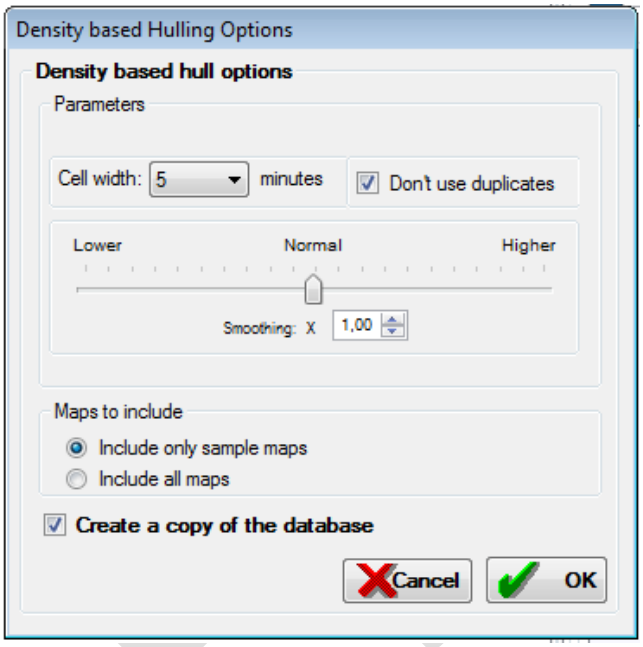



Remember that you can copy the contents of those lists to the clipboard, by clicking on them with the right button and selecting the *Copy to clipboard* option that will appear on a contextual menu.

After applying hull transformation you should process the transformed maps, as hull transformation involves a modification of the map presence data (see section 3.6).

## 3.9 Applying density based hull transformation

As seen in section 2.8, MapMaker is able to calculate the density based hull area from the presence data in a species distribution map, and add the corresponding hull area. But if you want to apply this transformation to numerous maps, it may be easier to do it in a single operation from DataManager in turn of doing it for each map in MapMaker. You can just check the tree branches to which you want to apply it, and select *Mapping/ Apply hull transformation to checked maps /Density based hull* menu option. A dialog box will be shown:



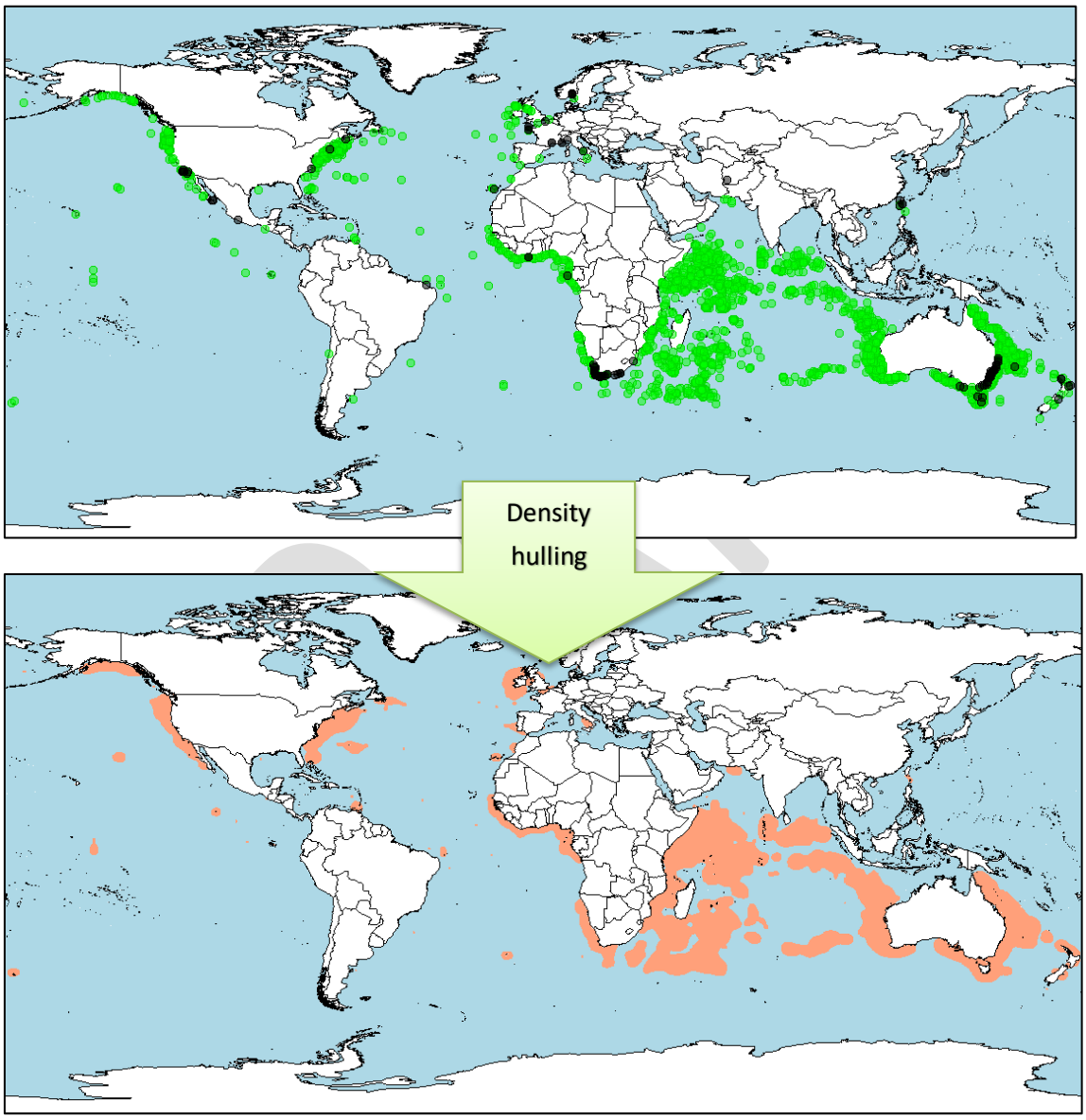

The available options are:

- **Cell width in minutes**: allows you to set the cell size to be used to calculate density. Density will be calculated for each cell of the selected size. So the smaller the cell size, the more precise is the density calculation. However, take into account that the time to perform the calculations grows geometrically as cell size diminishes. The recommended cell size for a good balance between rapidity and precision is 5 minutes.
- **Don't use duplicates**: if enabled, only distinct occurrences (usually considered until the 5$^{th}$ decimal) will be used to calculate kernel density. This avoids biasing density because there are many duplicates of some occurrences.
- **Smoothing**: smoothing factor determines how smooth the density matrix will be. The greater it is, the more the density will spread out far around presence areas. Default factor (x1) has been adjusted to return a density area with a standard deviation of the distances between cells of the presence areas near the original standard deviation of the distances between occurrences.
- **Maps to include:** you can select to include all maps or only occurrence maps (maps that only contain occurrences, but no presence areas). You should take into account that density hull can only be calculated on occurrences but not on areas. So if you select to include all maps, only the occurrences of those maps will be used to calculate hull. Therefore, maps without occurrences will be excluded.
- **Create a copy of the database:** as hull transformation involves irreversible map modification, it can be prudent to make a copy of the database before performing this transformation. Checking this option DataManager will ask you for a destination folder and file and will do this backup copy before modifying maps.

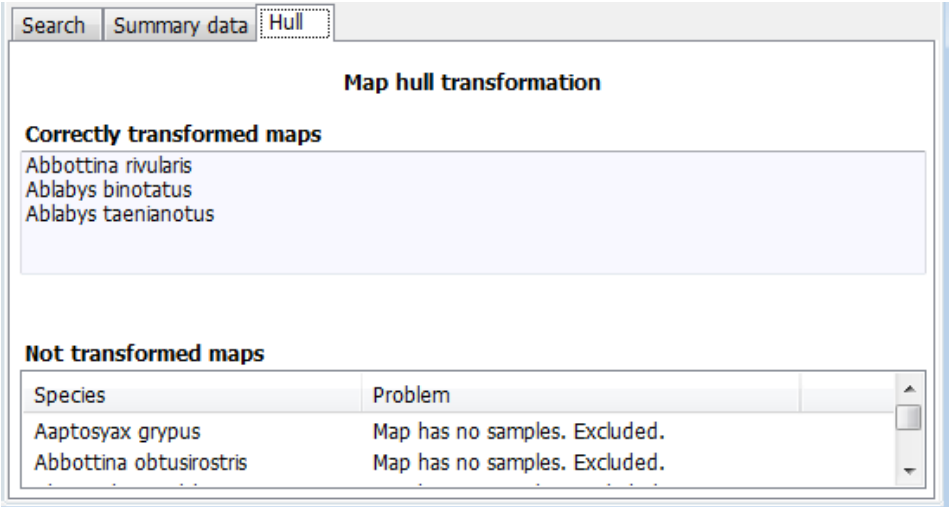
Hull area that will be added to the map will use as valid habitats the same ones that are currently accepted in the map. For example, if you apply hull transformation to a map that contains

occurrences that have sea as valid habitat, the corresponding hull area will also have sea as valid habitat.

As an example, below we show a map before and after applying density based hull transformation in DataManager (occurrences have been hidden in the resulting map). As it can be seen, density hull area is added to the map with the sea as valid habitat (so other ones such as land are excluded) because the previously existing occurrences have sea as valid habitat.



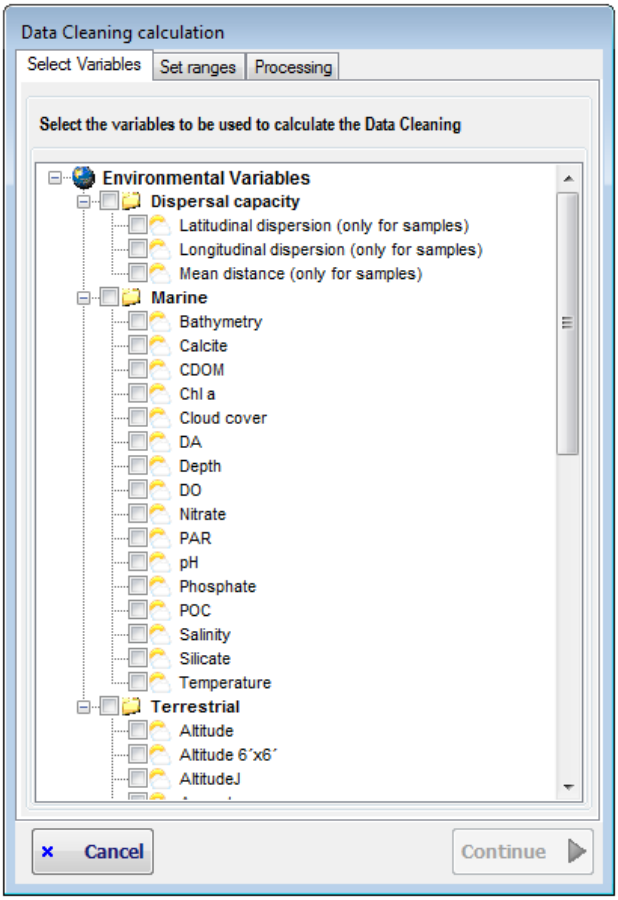

After applying hull transformation, a report will be shown in the right panel indicating the correctly transformed maps and the incorrectly processed ones, which were not updated. A brief message will report the encountered problem that prevent processing each incorrectly processed map. The most usual problems are that the map doesn't contains data enough to calculate hull. At least three occurrences are required to calculate density.




After applying hull transformation you should process the transformed maps, as hull transformation involves a map modification (see section 3.6).

## 3.10 Data cleaning in DataManager

As seen in section 2.6, MapMaker is able to perform automatic data cleaning of the presence data in a species distribution map. But if you want to apply this transformation to numerous maps in a batch mode, it may be easier to do it in a single operation from DataManager in turn of doing it for each map in MapMaker. If you are not familiar with ModestR data cleaning features, please read the section 2.6 before using it in DataManager.

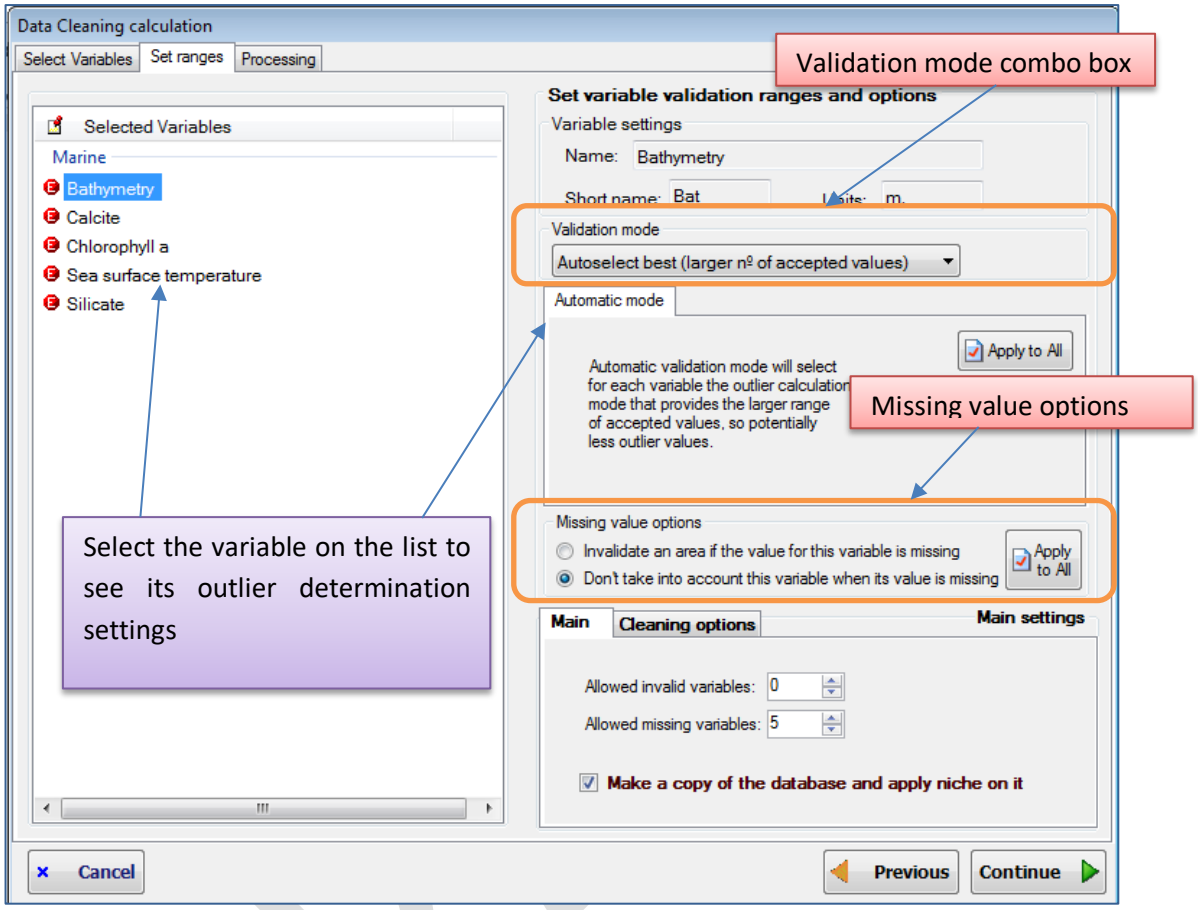



To apply data cleaning in DataManager you have to check the tree branches to which you want to apply it, and select *Mapping/ Apply data cleaning to checked maps* menu option. A dialog box will be shown. In this dialog box you can select on the tree the variables you want to use in the data cleaning. Then click on the *Continue* button to go to the next step.

In the next step you can see on the left the list of selected variables [16]. Before continuing, it is interesting to point out that many of the options and settings that are displayed on this dialog box are

[16] As you can see, unlike in Mapmaker data cleaning (see section 2.6), in DataManager current variable ranges cannot be shown, because data cleaning in DataManager will be applied to several maps, and not to a single one as in MapMaker.

autoexplained: just by placing the mouse cursor on one element, usually a short emerging explanation will be shown.




Then you can select how outliers will be detected for each variable. To do that you must select a variable on the left listbox. The corresponding settings for this variables will be shown on the right-up panel. The available settings are:

- **Validation mode:** this is the way the values for this variable on the presence zones (whether occurrences or areas) will be evaluated to determine if some occurrence/area has to be cleaning or not. Any validation mode will determine the range (minimum and maximum) of valid values for the variable, so any value that fall outside this range will be considered an outlier. The difference between different validation modes will be in the way they use to determine this range. By default, the validation mode will be set to *"Autoselect best"*. But you can manually select between the following modes:
  - ***Autoselect best**:* this mode will evaluate for each map which one of the other available validation modes returns the less outliers, and uses this mode to clean data. That is, with this mode you will always obtain the minimal number of outliers.
  - ***Quartile outliers***: this option use a validation based on the values of the Q1 and Q3 quartiles of the values took by a variable in the current presence areas/occurrences. This option will consider any value of this variable as invalid if it is outside the range [Q1 - X(Q3 - Q1) , Q3 + X(Q3 - Q1)] where X will be:
    - X=1.5 if you select to consider both mild and extreme outliers as invalid.

        - X=3 if you select to consider only extreme outliers as invalid.
        - You can also set a custom value of X in the options shown for this validation mode.
    - ***JackKnife outliers:*** this validation mode to detect outliers is described in Chapman's *Principles and Methods of Data Cleaning* (2005) and it is considered quite reliable. ModestR supports two variants of this method: the one proposed by CRIA and the one proposed in DIVA-GIS (both area described in Chapman's book).
    - ***Custom***: in this case it is the user that will manually set the minimum and maximum values of the range of valid values for a variable. Any value that fall outside this range will be considered an outlier.
- **Missing value options:** under the tabbed box that show the information for the variable, two radiobuttons allow you to select what to do when there is no value for the variable in a particular area/occurrence. You can select between two possibilities:
    - Invalidate an area/occurrence if the value for the variable is missing (that is, if the variable have a null value in this location): in this case, an area/occurrence will be cleaned (considered invalid) as soon as there is no value for the variable for this area/occurrence, no matter the values for the other variables.
    - Don't take into account this variable when its value is missing: in this case, if there is no value for the variable for this area/occurrence, this variable will be just not considered to make a decision. ModestR will check the values took by the other variables to decide about the validity of the area/occurrence.

  Take into account that those options are *per variable*. That is, you can set different options for each variable. A missing value for a variable can lead to immediately invalidate an area/occurrence, while for another variable it will not be took into account. You should decide in a per-variable basis the most adequate option.
- **Apply to all buttons:** either in the tabbed box that displays the options for the currently selected validation mode or in the missing value options, you can see a button *Apply to All*. This button can be used to replicate the corresponding settings for the current variable to all variables in an easy way.

Finally, on the right-bottom side of the dialog box, you can see a tabbed box with main settings. While settings described above are for each variable, those main settings have global effects. They are:

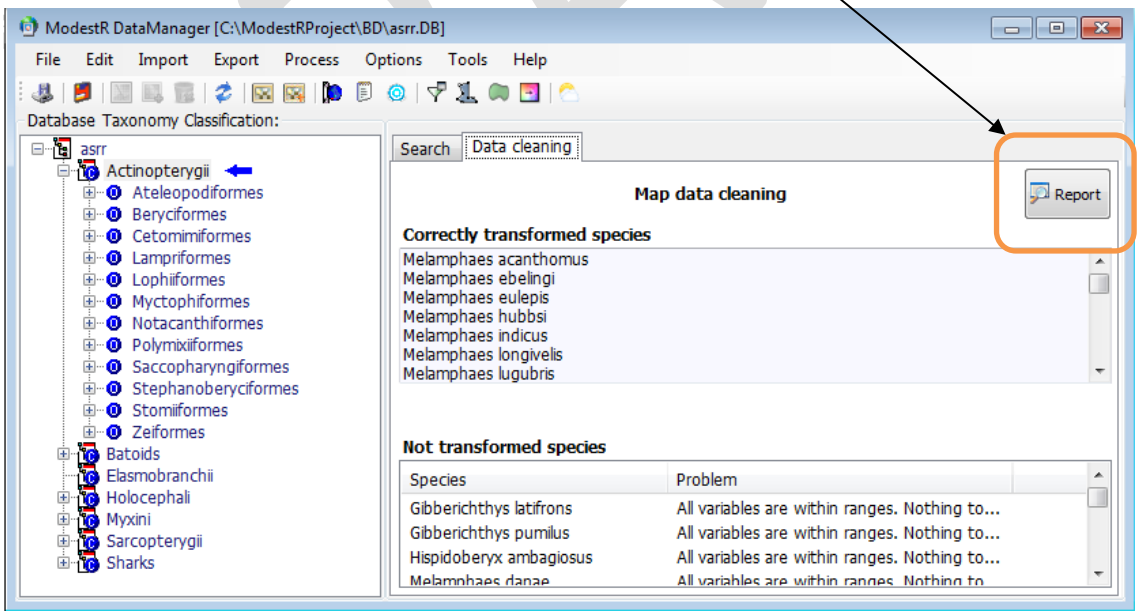
- **Allowed invalid variables:** Indicates how many variables can take invalid values, that is, outliers (according to their validation rule) before considering an area or occurrence as invalid for the species. The default value is 0, and the maximum is N-1 where N is the number of selected variables. Usually this field will be 0, because typically you will want to clean a occurrence/areas as soon as one of the selected variables takes an outlier value. But this setting provides a way to be more flexible. For example you can select 10 variables and allow that up to two variables of these take outlier values before considering a occurrence/area as invalid.
- **Allowed missing variables:** Indicates how many variables can be missing (have a null value) before considering an area or occurrence as invalid for the species. By default it will be set to N where N is the number of selected variables. That is, if all variables have

null values for a occurrence/area, the area will be considered as valid. Take into account that this parameter will be applied only if the variables with missing values are not individually configured as directly invalidating an area when their value is missing (as explained above). That is, this rule will be applied only if after applying the rules for each individual variable the occurrence/area has not already been considered as invalid.

- **Cleaning options:** in this tab page you have to select which occurrences and areas apply to (all, only already valid ones, only already invalid ones, or none). You can also select to add placemarks to validated/invalidated occurrences to make easier localizing them on the map. This tab will only be shown if the map contains occurrences.
- **Create a copy of the database:** as data cleaning involves irreversible map modification, it can be prudent to make a copy of the database before performing this transformation. Checking this option DataManager will ask you for a destination folder and file and will do this backup copy before modifying maps.

Once all those settings configured, press the *Continue* button to perform data cleaning. ModestR will then apply data cleaning to each one of the selected maps. Once finished, a summary report will be shown with the list of processed maps and the maps not processed, indicating the cause.

When occurrences cleaning has been done, a *Report* button will also allow you to save a detailed report of the cleaning operations done, in CSV format, containing a list of the cleaned occurrences for each processed species, and details about the variable ranges they didn't comply with. This report will only be generated when cleaning occurrences, and will not include details regarding areas cleaning. The reason for this is that even a relatively small region will easily contains thousands of 1'x1' cells, which results in a very long and rarely useful report.



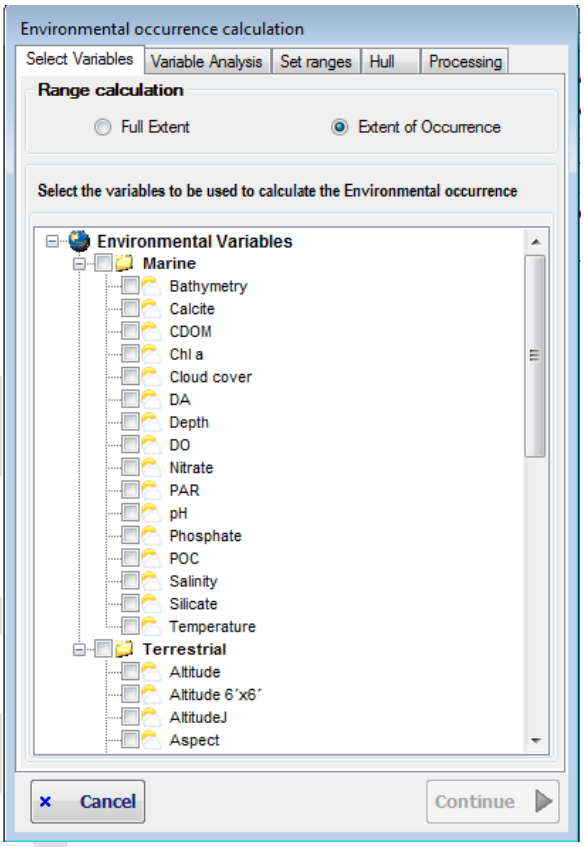

## 3.11 Applying Environmental Occurrence

Environmental Occurrence rudiments have been explained in section 2.10 where MapMaker Environmental Occurrence feature was detailed. We suggest you to do some tests in MapMaker, where results are instantly visible, to understand how Environmental Occurrence works. In

DataManager you can apply Environmental Occurrence to any taxonomic group of species in a single step.

To do that, check first the tree nodes you want to apply Environmental Occurrence to (all the database, a single class, order… etc). Then select *Mapping/ Apply Environmental Occurrence to checked maps* on the main menu.

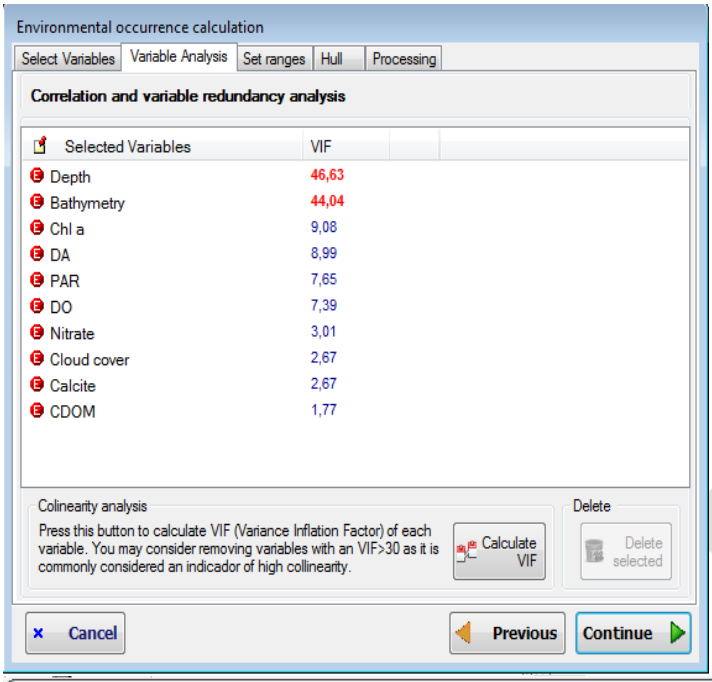
The first step is selecting the range calculation for the environmental occurrence and the EOO mode. There are two options for the range:



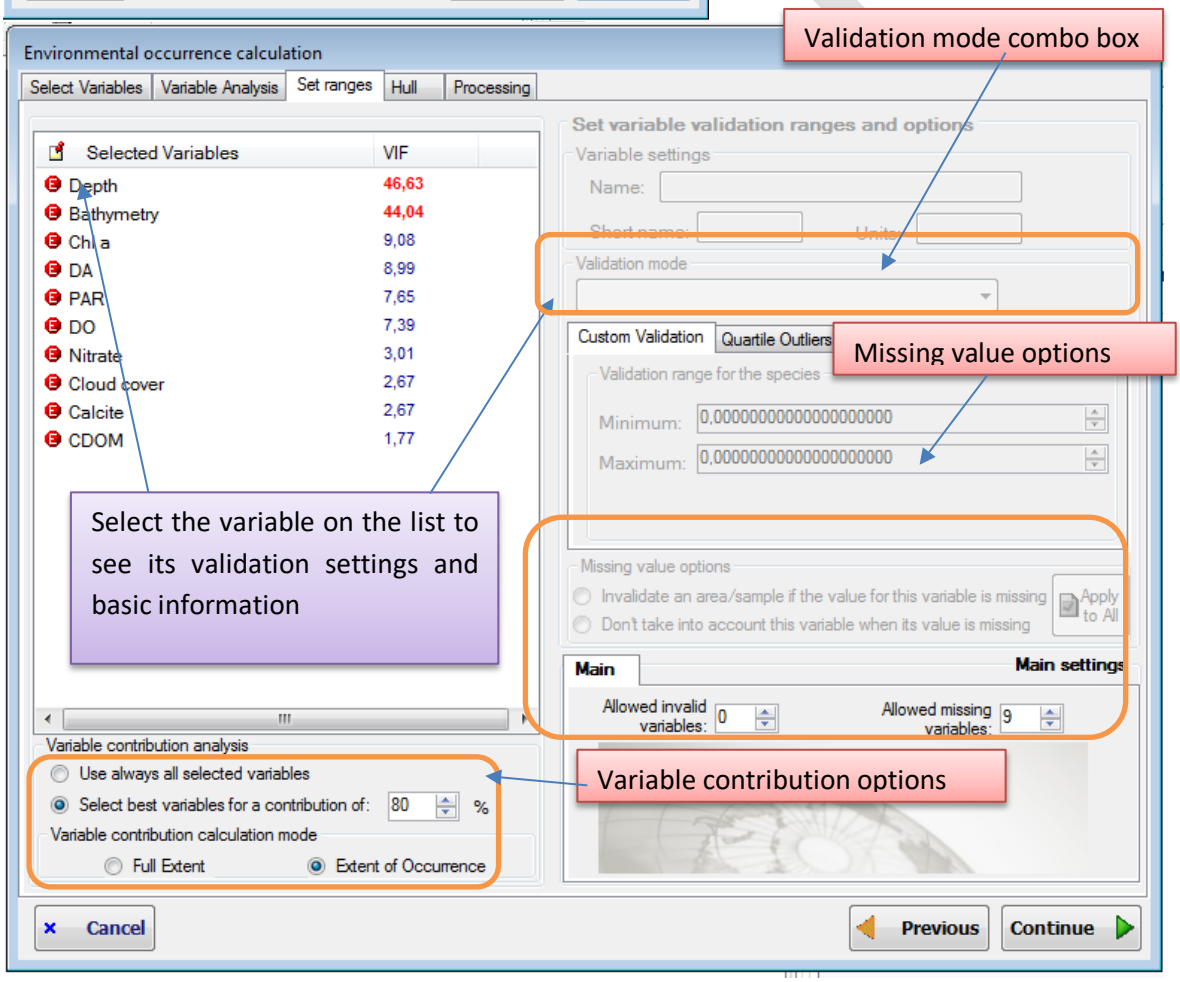

- Full extent: the environmental occurrence will be calculated for all available environmental data. For example, if you use environmental variables that contain data for all the world, the environmental occurrence will be calculated for all the world. If you use environmental variables that contain data for a single country, the environmental occurrence will be calculated for this country.
- EOO (Extent of occurrence): environmental occurrence will be calculated only for the area contained in the EOO of the species. This EOO will be calculated according to the settings you have to enter in a later step.

Next step is selecting the environmental variables that will be used to calculate Environmental Occurrence. Of course, in the ideal case, you should select the variables that affect the most the species distribution. Anyway, you can always select most variables, and ModestR will help you to choose the best ones in the next steps.

Next step allows you to optionally calculate Variance Inflation Factor (VIF). A VIF value higher than 30 is commonly considered an indicator of collinearity among the variables. Thus, you have the change choice of removing those variables significantly related and, therefore, that are redundant. But this is an optional step. If you just want to include all variables, you can simply continue. Take into account that if two variables are very highly correlated, the way that one of them if a linear function of the other one, VIF cannot be calculated. Involved variables will be marked to allow you to remove one of them if you want.




In the next step, you can select how validation will be done for each variable. To do that you must select a variable on the left listbox. The corresponding settings for this variables will be shown on the right-up panel. Before continuing, it is interesting to point out that many of the options and settings that are displayed on this dialog box are autoexplained: just by placing the mouse cursor on one element, usually a short emerging explanation will be shown. The available settings are:




- **Variable contribution analysis:** here you can select to always use selected variables to determine Environmental Occurrence, or to perform a statistical analysis for selecting those variables that are more affecting the most the distribution of each one of the species to be processed. If you select this second option, This analysis will be performed

for each species: the environmental variables selected are checked within the EOO of the species or for the full the extent where environmental data are available, according to the selected option. Then an index of instability is calculated using the fluctuation index of Dubois (1973) modified by Guisande et al. (2006). Once the instability index is calculated the percentage of contribution of each variable to the index is calculated. Those environmental variables with a higher percentage contribution to the instability index can be considered as those more affecting the most the distribution of the species. In order to include only the variables with higher contribution, you have to select a threshold percentage with the accumulative percentages of the variables.
This feature has already been explained in 2.10 for MapMaker, where this analysis can be done interactively for a species. It can be a good idea to do it first in MapMaker for a single species to better understand how it works.
Independently from Environmental Occurrence feature, a report about variable contribution for each species can also be exported using a specific exportation option of DataManager (see section 6.18).

- **Validation mode:** this is the way the values for this variable on the presence zones (whether occurrences or areas) will be evaluated to determine if some occurrence/area has to be removed from estimated EOO or not. There are two validation modes:
  - ***% Tolerance*:** the default validation mode is *% Tolerance*. It consists of using the range (minimum and maximum) of valid values for the variable, increased by a tolerance percent (by default 1%, but you can modify it), so any value that fall outside this range will be considered invalid and the corresponding area removed from the EOO.
  - ***Custom***: in this case it is the user that will manually set the minimum and maximum values of the range of valid values for a variable. Any value that fall outside this range will be considered invalid.

- **Missing value options:** under the tabbed box that show the information for the variable, two radiobuttons allow you to select what to do when there is no value for the variable in a particular area/occurrence. You can select between two possibilities:
  - Invalidate an area/occurrence if the value for the variable is missing (that is, if the variable have a null value in this location): in this case, an area/occurrence will be cleaned (considered invalid) as soon as there is no value for the variable for this area/occurrence, no matter the values for the other variables.
  - Don't take into account this variable when its value is missing: in this case, if there is no value for the variable for this area/occurrence, this variable will be just not considered to make a decision. ModestR will check the values took by the other variables to decide about the validity of the area/occurrence.

  Take into account that those options are *per variable*. That is, you can set different options for each variable. A missing value for a variable can lead to immediately invalidate an area/occurrence, while for another variable it will not be took into account. You should decide in a per-variable basis the most adequate option.
- ***Apply to all* buttons:** either in the tabbed box that displays the options for the currently selected validation mode or in the missing value options, you can see a button *Apply to*

*All*. This button can be used to replicate the corresponding settings for the current variable to all variables in an easy way.

Finally, on the right-bottom side of the dialog box, you can see a tabbed box with main settings. While settings described above are for each variable, those main settings have global effects. They are:

- **Allowed invalid variables:** Indicates how many variables can take invalid values, that is, outliers (according to their validation rule) before considering an area or occurrence as invalid for the species. The default value is 0, and the maximum is N-1 where N is the number of selected variables. Usually this field will be 0, because typically you will want to clean a occurrence/areas as soon as one of the selected variables takes an outlier value. But this setting provides a way to be more flexible. For example you can select 10 variables and allow that up to two variables of these take outlier values before considering a occurrence/area as invalid.

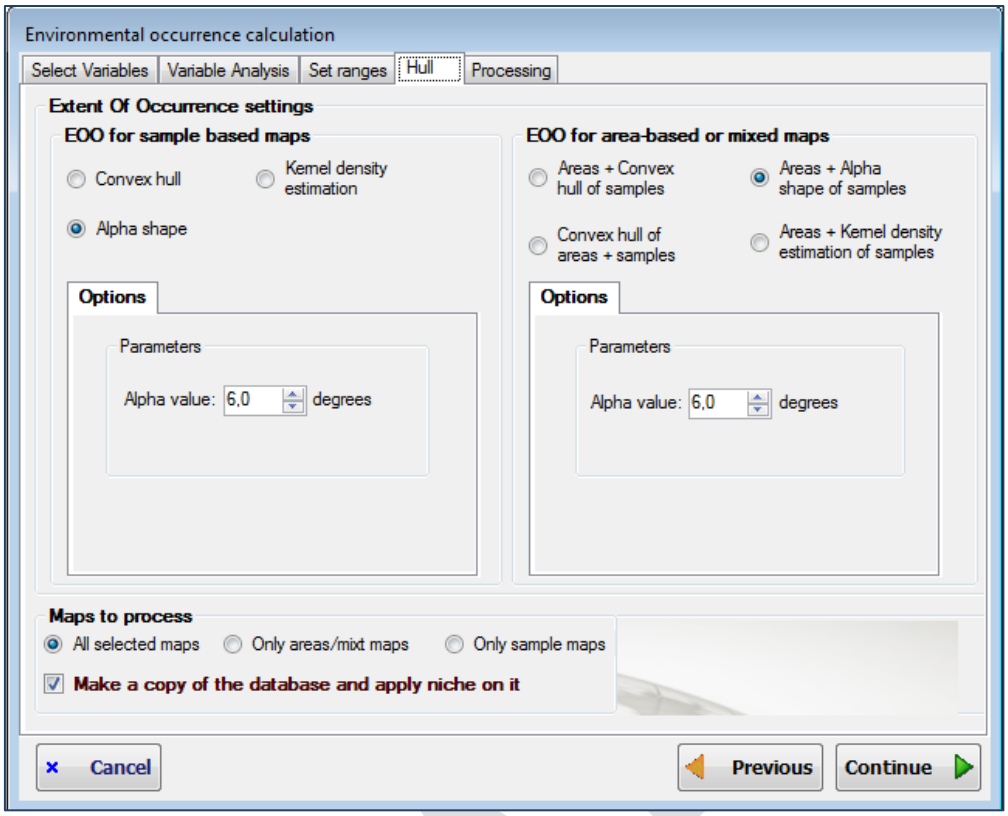
- **Allowed missing variables:** Indicates how many variables can be missing (have a null value) before considering an area or occurrence as invalid for the species. By default it will be set to N where N is the number of selected variables. That is, if all variables have null values for a occurrence/area, the area will be considered as valid. Take into account that this parameter will be applied only if the variables with missing values are not individually configured as directly invalidating an area when their value is missing (as explained above). That is, this rule will be applied only if after applying the rules for each individual variable the occurrence/area has not already been considered as invalid.

Once all those settings configured, press the *Continue* button.

In the next step, you can select how potential EOO will be calculated to each species to be processed using Environmental Occurrence feature. You should enter settings for EOO calculation if:

- you selected EOO as range calculation in the first step, or if
- you selected to apply variable contribution analysis in the previously described step

In both cases EOO is required. In any other case, EOO settings will appear disabled because it is not required.




There are different Extent of Occurrence options for occurrence maps and for maps with areas (range maps). EOO options for occurrence maps are:

- **Convex hull:** EOO is calculated as the convex hull area of the occurrences. If there aren't not enough occurrences to calculate convex hull, it is assumed that EOO=AOO. More details about convex hull can be found in 2.8.
- **Alpha shape**: EOO is calculated as the alpha shape hull area of the occurrences. If there aren't not enough occurrences to calculate alpha shape hull, it is assumed that EOO=AOO. More details about alpha shape hull can be found in 2.8.
- **Density estimation:** EOO is calculated as a hull area based on the density estimation (kernel smoothing) of the occurrences. If there aren't not enough occurrences to calculate density estimation hull, it is assumed that EOO=AOO. More details about density estimation hull can be found in 2.8.

EOO options for area and mixt maps are:

- **Areas+Convex hull of occurrences:** EOO is calculated as the presence areas already existing, plus the convex hull of the occurrences, if any.
- **Convex of areas+ occurrences:** EOO is calculated as the convex hull area of the areas plus occurrences if any.
- **Areas+Alpha shape of occurrences:** EOO is calculated as the presence areas already existing, plus the alpha shape of the occurrences, if any.
- **Areas+Kernel density estimation of occurrences:** EOO is calculated as the presence areas already existing, plus the kernel density estimation shape of the occurrences, if any.

In this step you can also select here if you want to make a copy of the database and apply Environmental Occurrence on this copy. This avoids modifying current database and then prevents from losing data.

You can select here to apply Environmental Occurrence to all types of maps, only to occurrence maps, or only to area/mixt maps (maps that contain both occurrences and areas area called mixt maps and are treated as area maps).

Once all those settings configured, press the *Continue* button. The Environmental Occurrence calculation processing will start and process each one of the selected species. The resulting estimated presence areas will be added to each distribution map. Take into account that this process can take a long time, depending on the number of maps to process and their complexity.

Once ended, a report of the whole process will be shown on the right side of DataManager, showing processed maps and not processed maps (commonly because they don't comply with conditions to apply Environmental Occurrence). If you choose to use variable contribution index to selected variables to be applied to each species, a *Report* button will be shown on the right panel, allowing to export a CSV report where the contribution index for each variable and selected variables are listed.

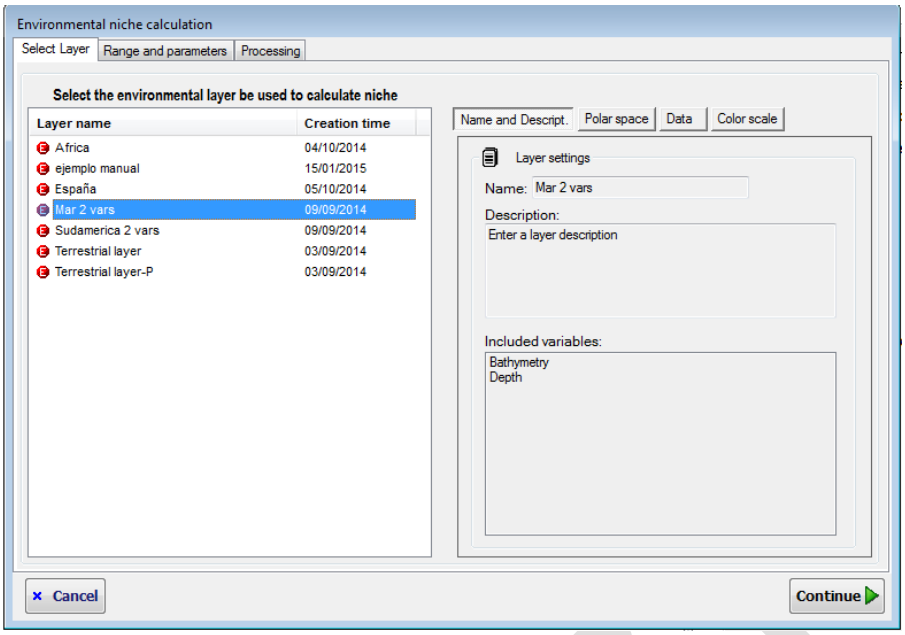
It must be said that, independently from Environmental Occurrence feature, a report about variable contribution for each species can also be exported using an exportation option of DataManager (see section 6.18).

## 3.12 Calculating 2D niche of occurrence with an existing 2DCEL

Niche Of Occurrence (NOO2D from now) has been explained in section 2.12. As for other features such as data cleaning or hull calculation, DataManager can easily calculate NOO2D for a set of maps in a single step, while MapMaker will only do it for one.

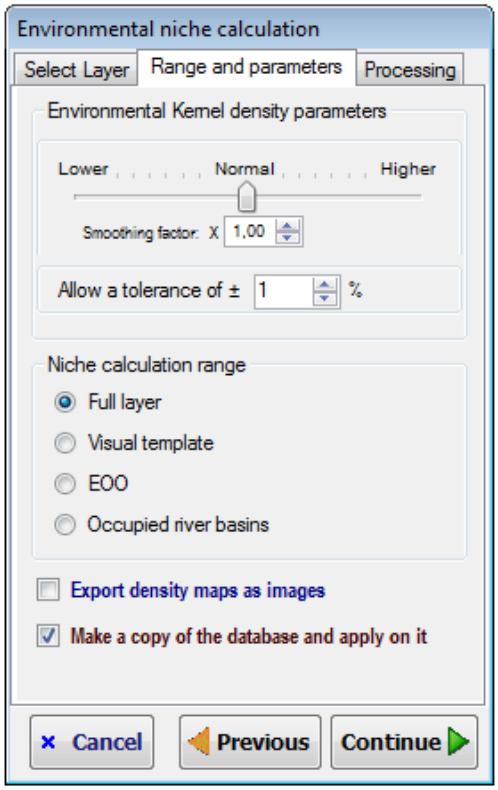
As explained in section 2.12, NOO2D relies on the concept of 2DCELs. NOO2D is intended to be applied on occurrence-based maps, as range maps already include presence areas. MapMaker allows calculating and storing 2DCELs (see section 2.11) that can be latter used to calculate NOO2D of a set of species in DataManager. It is very important to point out that, as seen before, a 2DCEL is built upon several environmental variables. And to perform niche of occurrence, those original environmental variables are needed, besides the 2DCEL itself. Consequently, to obtain accurate results, those original environmental variables must still exist, with the same names and data than when the 2DCEL was created. Errors or inaccurate results can be obtained if some variable has been modified since the 2DCEL creation. To avoid those problems as much as possible, from ModestR v.3.1 when you delete or modify an environmental variable, any 2DCEL that used this variable will be deleted.

To apply NOO2D using an already created and stored 2DCEL in DataManager, check the tree branches to which you want to apply it, and select *Mapping/ Apply Niche of Occurrence to checked maps /Using a stored layer* menu option, or the *Mapping features/ Apply Niche of Occurrence to this branch/Using a stored layer* option in the contextual menu of the tree node. A dialog box will be shown where the list of currently stored 2DCELs is displayed. You should select the layer you want to use for NOO2D and click on *Continue.*




In the next step the options dialog box is shown.

The parameters you can set are;




- Smoothing factor: determines how smooth the density matrix will be. The greater it is, the more the density will spread out far around presence areas. Default factor is x1 .
- Allow a tolerance of ± X: Apply a % of tolerance on the potential areas of presence to consider them as valid, with regard to the range between min and max values for each environmental variable involved in the currently selected 2DCEL in the presence data of the species.
- Range: determines the range or maximal area for which NOO2D will be calculated. Options are:
    - Full layer: to calculate NOO2D for all the area covered by the currently selected 2DCEL.
    - Shape: selecting this option, a right panel will appear where you'll have to select a shape (e.g. a country, a continent…). NOO2D will be calculated only within the bounds of this shape.
    - EOO: selecting this option, a right panel will appear where you'll have to select the EOO calculation method. NOO2D will be calculated NOO2D only within the bounds of the EOO (Extent of Occurrence) of the species.

- Occupied river basins: selecting this option, a right panel will appear where you'll have to select a river basin level (0=higher; 5=lower). NOO2D will be calculated only within the bounds of the river basins of this level where there are occurrences of the species. See 2.16.2 for more details of how ModestR determines the river basins occupied by a species.
- Export density maps as images: enabling this option, a NOO2D density map of each species to be processed will be generated and exported to an image file. This image corresponds to the *Density map* option explained in MapMaker (see 2.12). Image files will be saved in a folder you will have to select in the next step.
- Make a copy of the database and apply on it**:** as NOO2D involves irreversible map modification, it can be prudent to make a copy of the database before performing this transformation Checking this option DataManager will ask you for a destination folder and file and will do a copy of the database and apply NOO2D on this copy, preserving the original database.

Once parameters set, click on *Continue* button. NOO2D calculation processing will start and process each one of the selected species. The resulting estimated presence areas will be added to each distribution map. Take into account that this process can take a long time, depending on the number of maps to process and their complexity.

Once ended, a report of the whole process will be shown on the right side of DataManager, showing processed maps and not processed maps (commonly because they don't comply with conditions to calculate EOO or they don't have data in the selected range).

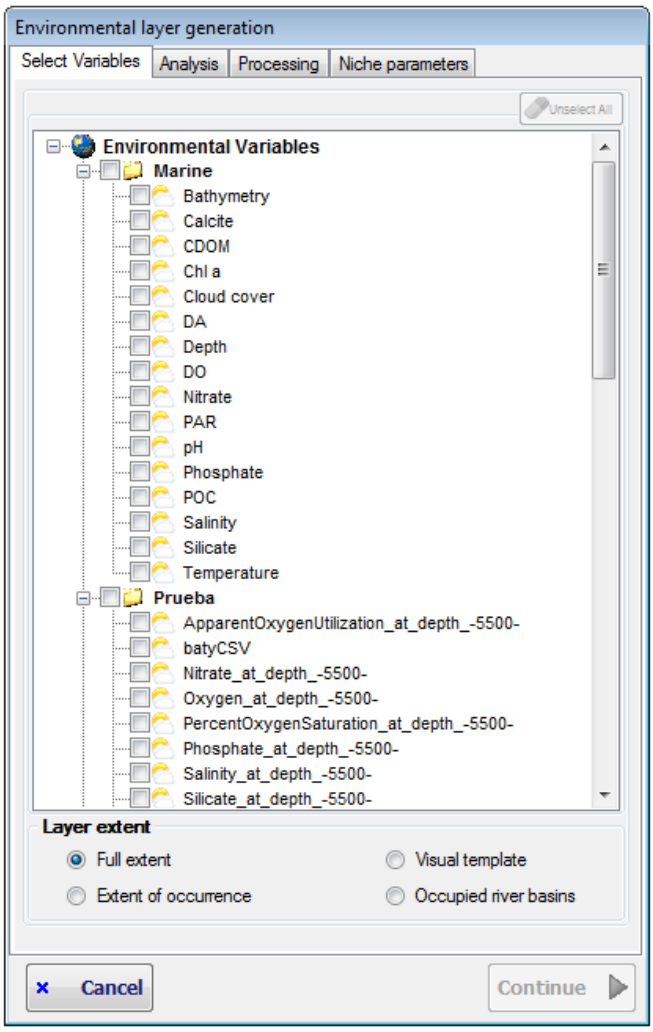
## 3.13 Calculating 2D niche of occurrence with a per-species 2DCEL

In the previous section we described how to calculate NOO2D using a previously generated and stored 2DCEL. The main drawback of this approach is that the same 2DCEL will be used for all the species for which we are calculating NOO2D. This inherently involves using the same environmental variables for all the species. If we consider that the environmental variables that influence the most a species distribution can be different from one species to another, a more convenient approach may be using a different 2DCEL to calculate the NOO2D of each species. This *ad-hoc* 2DCEL should be generated using the environmental variables that influence the most each species distribution. In MapMaker it is already possible to create and 2DCEL in this way (see section 2.11), for a specific species.

In DataManager it is possible to do the same but for a collection of species. Using this option, DataManager will determine for each species which are the variables that influence the most each species distribution, using the variable contribution approach already described in section 2.11. It will generate on the fly a 2DCEL with those variables and use it to calculate NOO2D. In this way, each species NOO2D will be calculated using the variables that influence the most the species distribution.

To apply NOO2D using an on-the-fly created 2DCEL for each species in DataManager, check the tree branches to which you want to apply it, and select *Mapping/ Apply Niche of Occurrence to checked maps/Using a per-species layer* menu option. The following dialog box will be shown, where you have to select at least two environmental variables to be used to generate the 2DCEL

for each species (details about how to integrate environmental data in ModestR are explained in section 6.26.).




In this dialog box you can also select the *Layer extent*, which can be:

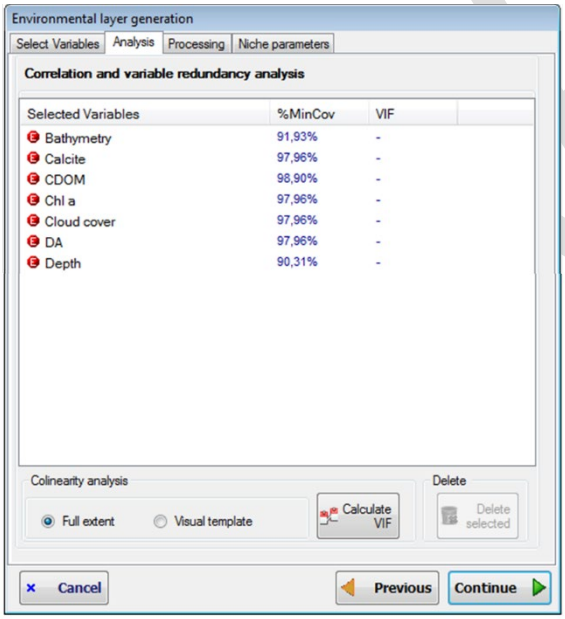
- Full extent: the layer will be calculated for the area for which selected environmental data is available. For example, if you selected environmental variables that cover the entire world, the 2DCEL to be created will also cover the entire world.
- Shape: selecting this option, a right panel will appear where you'll have to select a shape. The layer will be calculated only within the bounds of this shape.
- Extent of Occurrence: selecting this option, a right panel will appear where you'll have to select the EOO calculation method. The layer will be calculated only within the bounds of the EOO (Extent of Occurrence) of the species.
- Occupied river basins: selecting this option, a right panel will appear where you'll have to select a river basin level (0=higher; 5=lower). The layer will be calculated only within the bounds of the river basins of this level where there are occurrences of the species. See 2.16.2 for more details of how ModestR determines the river basins occupied by a species.

It has to be noted that the options explained above define the extent or bounds of the *2DCEL* to be calculated for each species; that is not the same than the bounds for the NOO2D to be calculated for the species. You can, for example, calculate the 2DCEL using as bounds a continent, and next calculate the NOO2D of the species using as bounds a specific country of this continent. Obviously, the extent of the 2DCEL inherently limits the potential NOO2D of the species. If you limit the 2DCEL to a country, for example, the NOO2D of the species will be inherently limited to this country, independently of the bounding options you use for the NOO2D. Seeing explanations about 2DCEL creation (see 2.11) and NOO2D calculation (see 2.12) in MapMaker can help correctly understanding difference between options affecting 2DCELs and those for NOO2D calculation.

Once those options set, click on *Continue* button to go to next step.

In the next step, selected variables are shown in a list. You can optionally calculate Variance Inflation Factor, by clicking on the *Calculate VIF* button. This is just optional. You can skip this step and click on *Continue* button to go to next step.

In the case you click on the *Calculate VIF* button, the VIF of each variable will be shown in the list. VIF can be useful to detect high correlated variables, and then decide to remove some variables. To remove a variable, just select in on the list and use the *Delete selected* button.

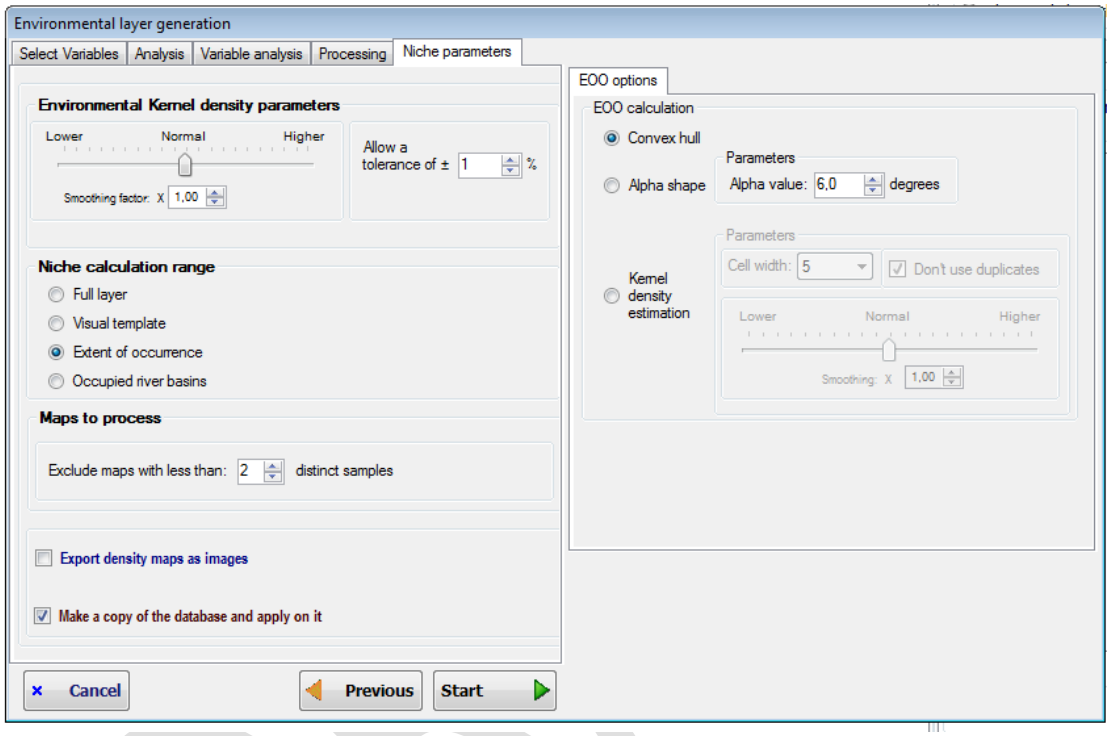



VIF can be calculated for the full extent or for a shape (given that a shape has been selected in the previous step, as explained before). Take into account that extent used to calculate VIF can be different from the extent selected for the 2DCEL. ModestR doesn't establish any limitation in this sense, so the convenience of using one or another option is left to the user. Click on *Continue* button to go to next step.

The next step will allow you to set how the contribution analysis will be done[17]. This analysis is the key step when calculating NOO2D using a per-species layer. For each species, this analysis will determine which ones of the previously selected variables influence the most the species distribution, and therefore will be used to generate the 2DCEL that will be used to calculate NOO2D. This feature has already been explained in 2.10 for MapMaker, where this analysis can be done interactively for a species. It can be a good idea to do it first in MapMaker for a single species to better understand how it works. It is also cited in 3.11.

[17] This feature has already been explained in 2.10 for MapMaker, where this analysis can be done interactively for a species. It can be a good idea to do it first in MapMaker for a single species to better understand how it works.

In a similar way to the *Layer extent* option described in a previous step, variable contribution can be calculated for the full extent, for a shape, for the occupied river basins or for the EOO of the species. Take into account that extent used to calculate variable contribution could be different from the extent selected for the 2DCEL. ModestR does not establish any limitation in this sense, so the convenience of using one or another option is left to the user. You may set the value of the field *Select best variables for a contribution of XX%*, for DataManager to select automatically the variables with the higher contribution to reach the indicated %. Click on *Continue* button to go to next step.




In the next step the options dialog box is shown. The parameters you can set are;

- Smoothing factor: determines how smooth the density matrix will be. The greater it is, the more the density will spread out far around presence areas. Default factor is x1 .
- Allow a tolerance of ± X: Apply a % of tolerance on the potential areas of presence to consider them as valid, with regard to the range between min and max values for each environmental variable involved in the currently selected 2DCEL in the presence data of the species.
- Niche calculation range: determines the range or maximal area for which NOO2D will be calculated. Options are:
    - Full layer: to calculate NOO2D for all the area covered by the currently selected 2DCEL.
    - Shape: selecting this option, a right panel will appear where you'll have to select a shape. NOO2D will be calculated only within the bounds of this shape.
    - Extent of Occurrence: selecting this option, a right panel will appear where you'll have to select the EOO calculation method. NOO2D will be calculated NOO2D only within the bounds of the EOO (Extent of Occurrence) of the species.

    - Occupied river basins: selecting this option, a right panel will appear where you'll have to select a river basin level (0=higher; 5=lower). NOO2D will be calculated only within the bounds of the river basins of this level where there are occurrences of the species. See 2.16.2 for more details of how ModestR determines the river basins occupied by a species.
- Exclude maps with less than X distinct occurrences: maps will less than X occurrences will be skipped. This option can be useful to skip maps that we consider that have too few data to calculate a meaningful NOO2D.
- Export density maps as images: enabling this option, a NOO2D density map of each species to be processed will be generated and exported to an image file. This image corresponds to the *Density map* option explained in MapMaker (see 2.12). Image files will be saved in a folder you will have to select in the next step.
- Make a copy of the database and apply on it: as NOO2D involves irreversible map modification, it can be prudent to make a copy of the database before performing this transformation Checking this option DataManager will ask you for a destination folder and file and will do a copy of the database and apply NOO2D on this copy, preserving the original database.

Click on the *Start* button to go to next step. NOO2D calculation processing will start and process each one of the selected species. The resulting estimated presence areas will be added to each distribution map. Take into account that this process can take a long time, depending on the number of maps to process and their complexity.

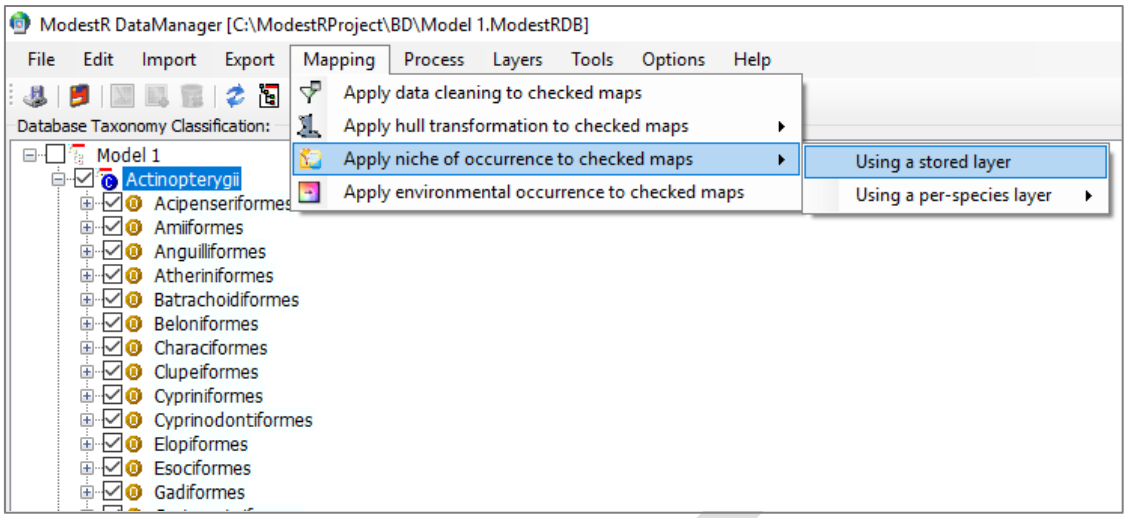
Once ended, a report of the whole process will be shown on the right side of DataManager, showing processed maps and not processed maps (commonly because they don't comply with conditions to calculate EOO or they don't have data in the selected range).

## 3.14 Calculating 3D niche of occurrence with an existing 3DCEL

NOO3D can be performed in DataManager in batch mode for any set for species of a ModestR database. The first step is opening a database and selecting the species to whom NOO3D will be applied, checking them on the tree.

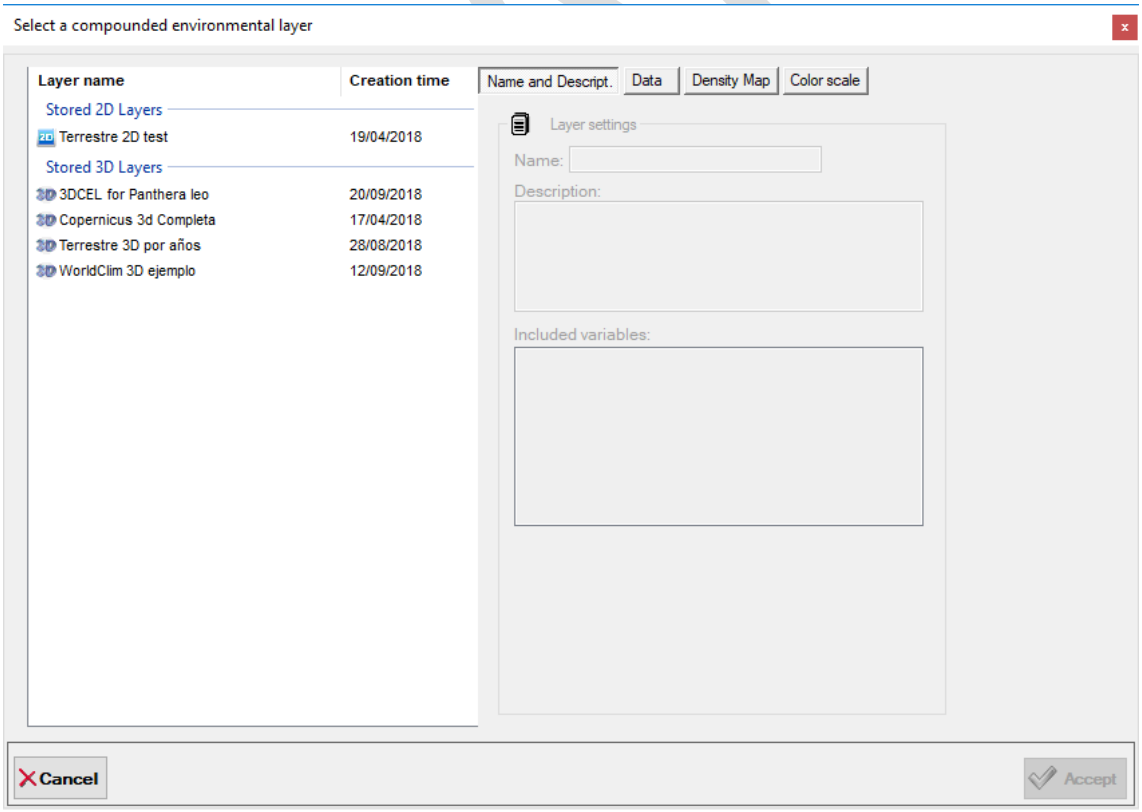
Then two options are available in the menu *Mapping/Apply niche of occurrence to checked maps*: using a stored layer[18], or using a per-species layer, in 3D.

[18] This option is common to 2D and 3D NOO. The NOO will be 2D or 3D according to the type of stored layer to be used, which can be a 2D or 3D CEL.

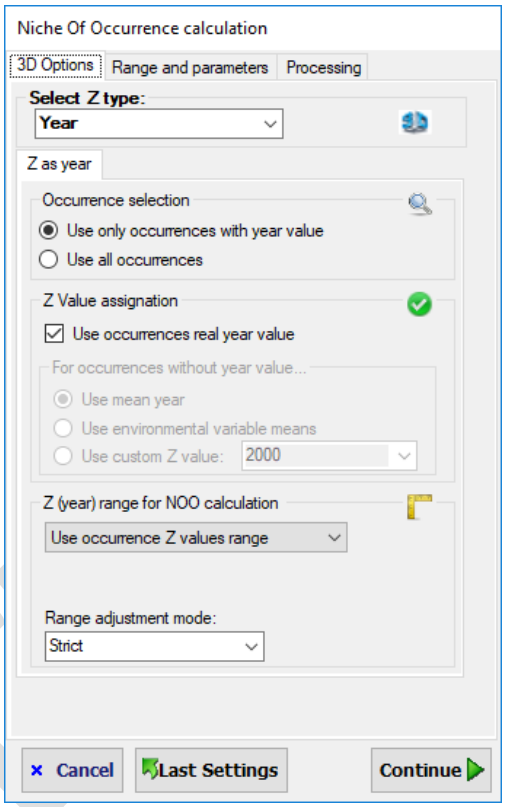



The option of using a stored layer is aimed to apply NOO3D to species using a same 3DCEL previously built and stored using MapMaker (see section 2.14). This option may be useful when we previously determined a collection of variables that are commonly explanatory for several species, and/or to obvtain easier comparable results, as they will be based on the same variables.

The first step is selecting the stored 3DCEL to be used. A list will be shown containing both 2D and 3D CEL's.



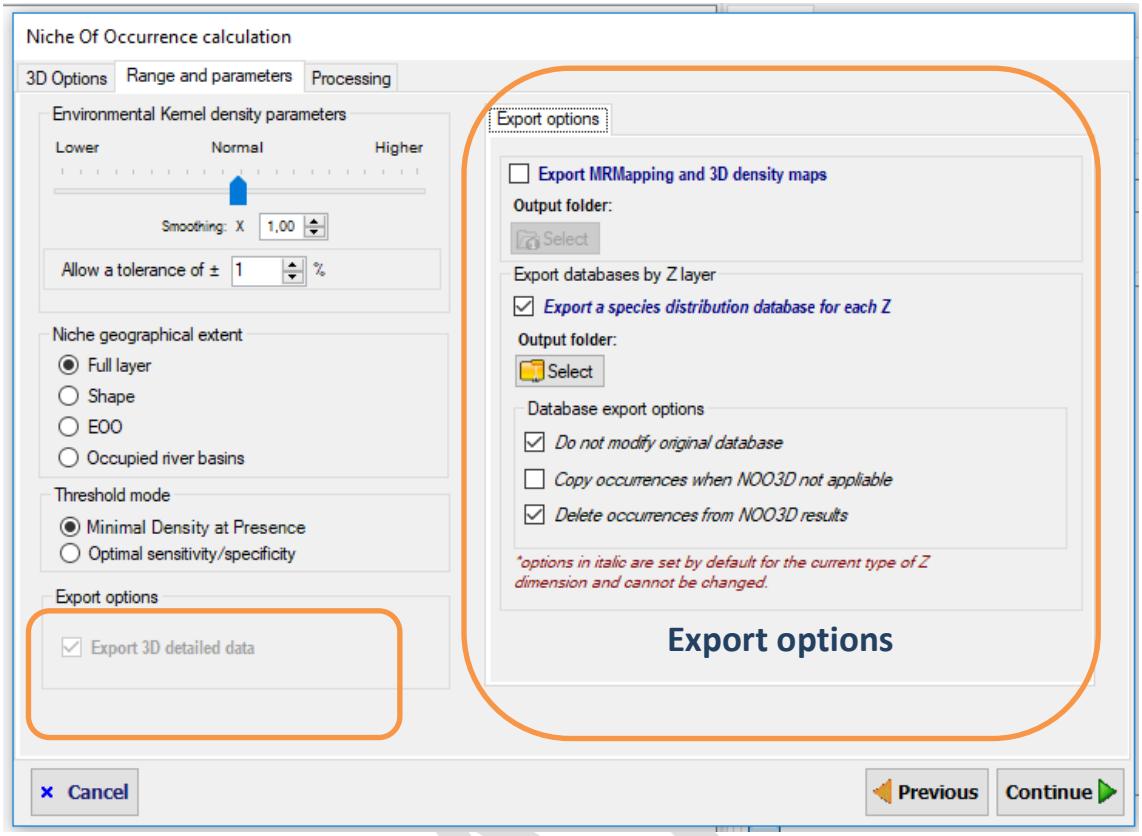
Once selected a 3D CEL, a series of parametrization steps will allow us to customize the NOO3D calculation.




The first one allows us to determine the Z type, the species occurrences to be used and the Z range. The available settings are the same than those previously explained for MapMaker in section **¡Error! No se encuentra el origen de la referencia.**. The difference, of course, is that they cannot be checked interactively to determine if they are applicable to each one of the selected species. It'll be at the calculation step that this checking will be done for each species, and any error reported (for example if a species doesn't have any occurrence with assigned Z value, etc).

It is also interesting to notice that for the second and subsequent times you apply NOO3D using the same 3DCEL a button *Last Settings* will appear, allowing you to recover the settings you used the last time.

In the next step the available options allow to customize the kernel density estimation: the kernel options; the geographical extent (spatial boundaries) where the NOO3D will be calculated; and the threshold mode to determine the species presence/absence. All of them have been already explained in section **¡Error! No se encuentra el origen de la referencia.**.




The other available options are related to data exportation about the NOO3D results. Checking the *Export 3D detailed data* checkbox a tab containing several options will be displayed on the right. The *Export 3D detailed data* can already appear as checked and disabled to avoid unchecking it, when this option is considered mandatory. This is the case, for example, when the Z dimension is temporal, such as years, because, as explained below, in this case some exportation options are necessary.

The available exportation options that appear on the right tab are:

- **Export MRMapping and 3D density maps:** this option will export a multiple-layer distribution map for each species, generated from the density map and using the selected threshold value to build a binary (presence-absence) species distribution map, where presence areas are considered the niche of occurrence of the species. The files will be exported to a folder that the user should select. Those maps for each species contain a presence area for each Z value in the selected range that can be visualized in MRMapping, which allows showing several overlapping distribution maps. It also exports 3D density maps for each species (see section **¡Error! No se encuentra el origen de la referencia.** for an example). They are polar density maps that shows the density (suitability) of each species in the environmental space of the 3DCEL. They can be visualized using the MR3DCELViewer included in ModestR (go to *Tools* menu in any of the ModestR applications to run this viewer).

- **Export a species distribution database for each Z**: When applying NOO3D to several species in batch mode, a separate database for each Z layer can also be generated. Each database will contains the distribution map of each species for each Z layer. This allows

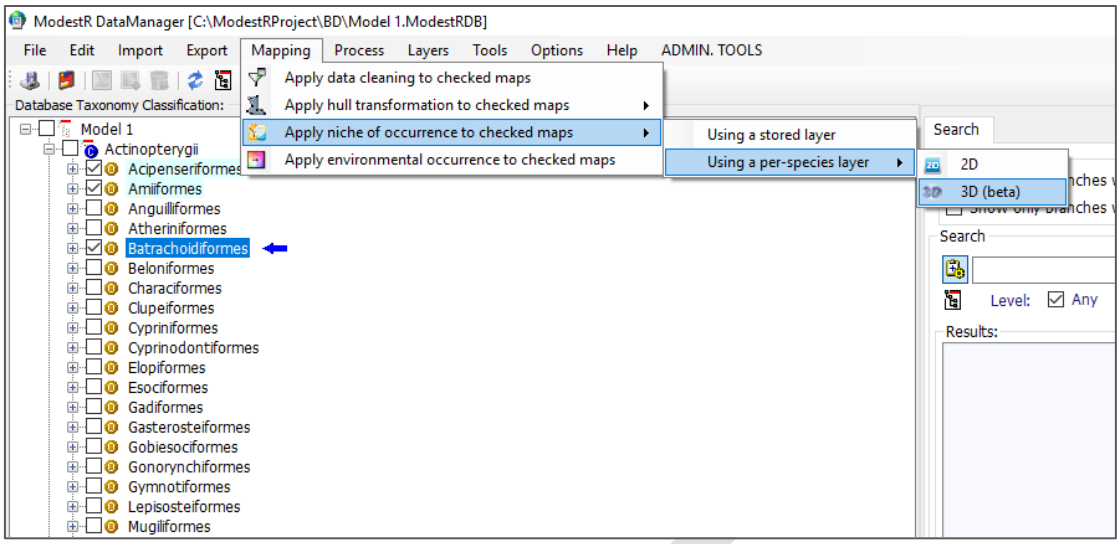
further analyses of the resulting distributions for each layer separately, such as richness, overlappings, etc. The databases will be saved in a folder that the user should select. This option is mandatory when the Z dimension is temporal, such as years, because in this case the results

There are several options when exporting a species distribution database for each Z:

- **Do not modify original database**: If this option is enabled, the original database (the currently opened one) will not be modified, but only the databases for each Z layer will be generated. This option is mandatory for example when the Z dimension is temporal. In this case we'll be interested only in the distributions for each Z layer, but it is not pertinent to add the sum of the distributions for each year to the original species map. It is optional when Z dimension is depth or another type. It can be also useful when Z is depth, to generate a distribution database for each depth but without modifying the original database.
- **Copy occurrences when NOO3D not applicable**: This option is applied when NOO3D cannot be aplied to a species for any reason (too few occurrences, selected EOO cannot be calculated, etc...). For those cases, if this option is enabled then the original occurrences of the species in each database for Z layer (this can be seen as assuming the original data we have for a species in case we can't make a prediction with NOO3D). If this option is disabled, then no species map will be added to the database for Z layer (this can be seen as not assuming any species distribution in case we can't make a prediction with NOO3D). This option is mandatorily disabled when the Z dimension is temporal.
- **Delete occurrences from NOO3D results**: This option indicates if we want to delete or to preserve the preexisting occurrences in the maps generated by NOO3D for each layer of a species where the NOO3D returned a non-empty distribution (that is, a presence area). For example, let's suppose we have a map with occurrence data for a species. Then we apply NOO3D to calculate the distribution of this species in several years (Z=year), and NOO3D returns presence areas for each year. If this option is disabled, each species map obtained for each year will contain the original occurrences of the species and the corresponding presence area obtained for this layer by the NOO3D. In turn, if this option is enabled, each species map obtained for each year will only contain the corresponding presence area obtained for this layer by the NOO3D, while the original occurrences of the species will not be included in those maps. This option is mandatory when the Z dimension is temporal.

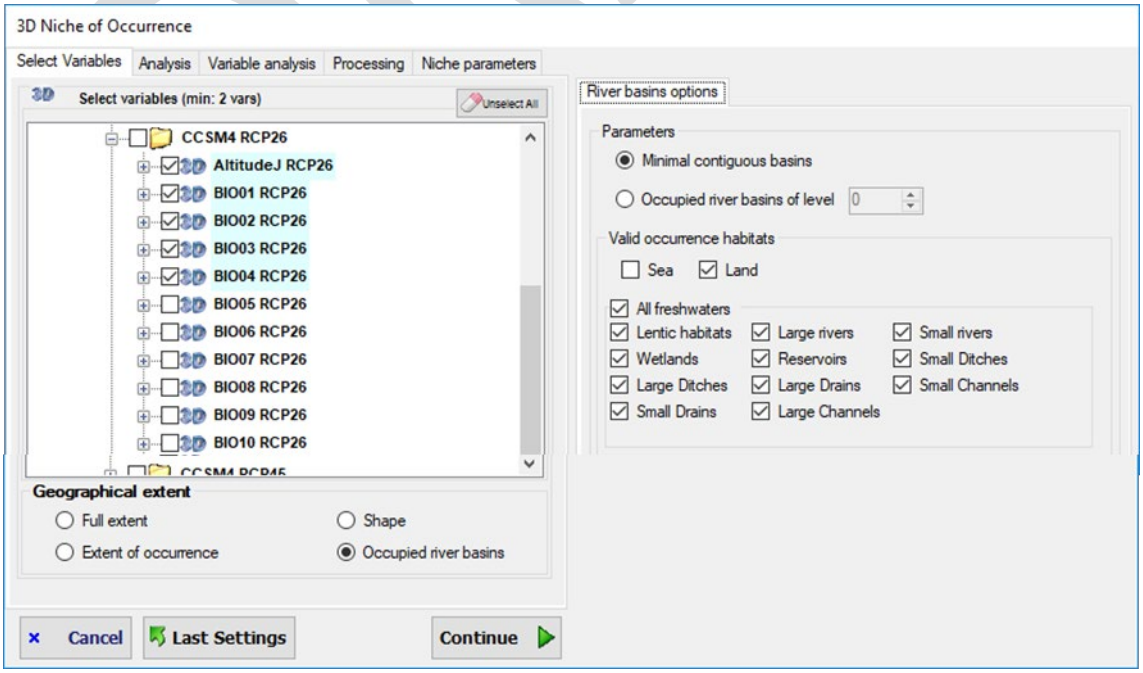
## 3.15 Calculating 3D niche of occurrence using a per-species 3DCEL

The option of using a per-species layer is aimed to apply NOO3D to a set of species using a 3DCEL specifically built on-the-fly for each species. This option may be useful when we want to use a collection of potentially explanatory variables, but selecting among them the most explanatory ones for each species, then using only those ones to apply NOO3D. Therefore this feature reproduces the processes described in sections 2.13 (creating a 3DCEL) and 2.14 (calculating NOO3D for a species) for each selected species, in a batch mode for any set of species, whereas MapMaker performs it in an interactive mode for a single species.



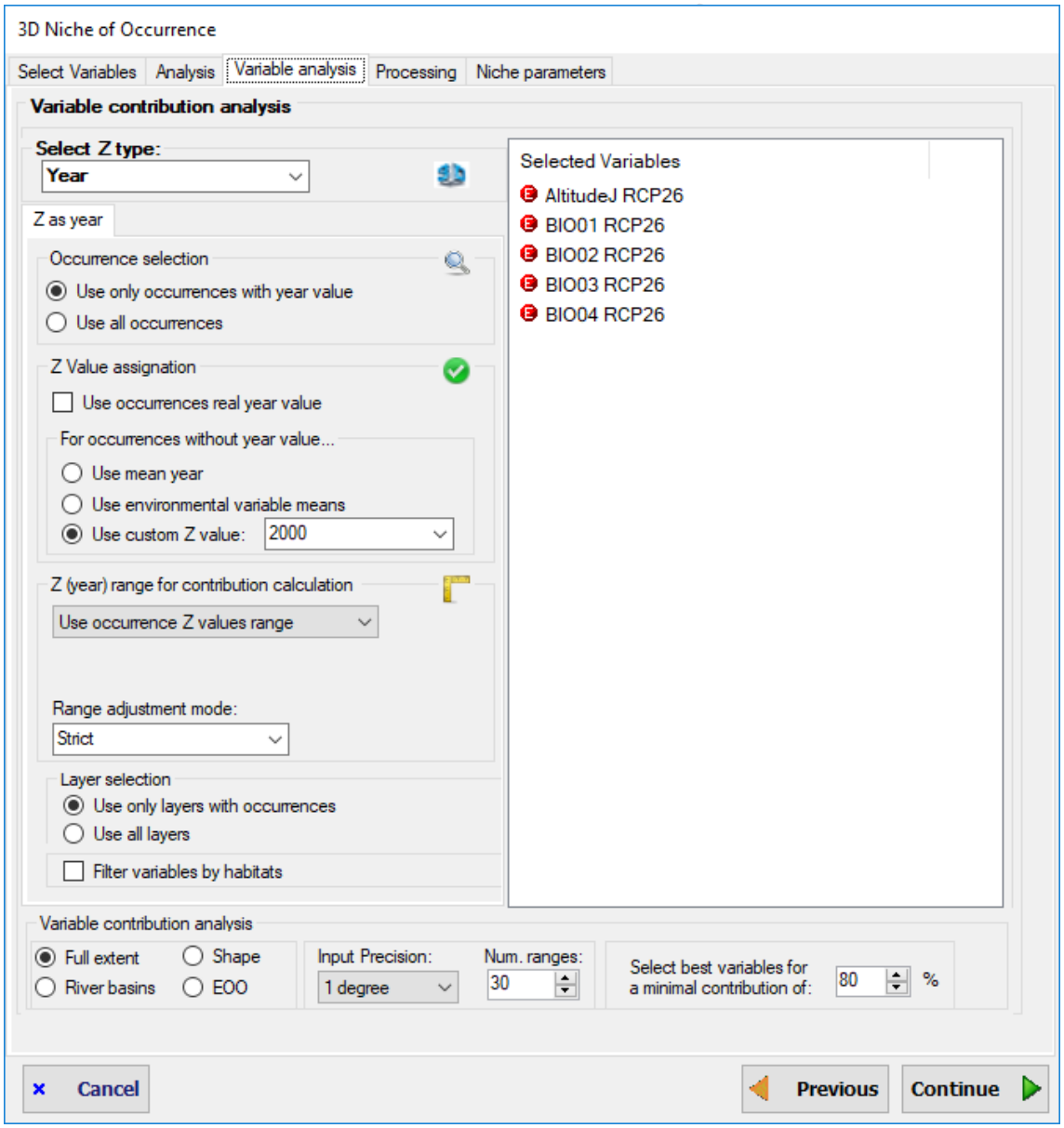

This option is available at *Mapping/Apply niche of occurrence to checked maps/Using a per-species layer/3D.* As this option involves the creation of a 3DCEL for each species, the first steps are the same than those explained in section 2.12. (Compounded Environmental Layer (3DCEL) creation in MapMaker): selecting the variables to be considered and the working area (geographical extent or GE). Of course, the difference, is that the selected GE cannot be checked interactively to determine if it is applicable to each one of the selected species. It'll be at the calculation step that this checking will be done for each species, and any error reported (for example if we selected as GE the EOO using the alpha shape method, but this EOO cannot be calculated for a particular species, etc).

It is also interesting to notice that for the second and subsequent times you apply NOO3D using the per-species option a button *Last Settings* will appear, allowing you to recover the settings you used the last time.



In the next step the redundancy among the selected variables can be optionally measured using the variance inflation factor (VIF), as already explained in section 2.12. The difference is that here the option of calculating VIF using EOO as GE is not available, because the EOO depends on each species and will only be calculated once started the NOO3D process.

In the next step we have to choose the settings to be used to calculate the contribution of each selected variable to the distribution of a species, as also explained in section 2.12. This step is relevant as the settings selected here will be used determine whose variables are the most explanatory ones for each species. Therefore those variables, that may be different for each one of the selected species, will be used to build on-the-fly a 3DCEL that will next be used to calculate the NOO3D of each species.

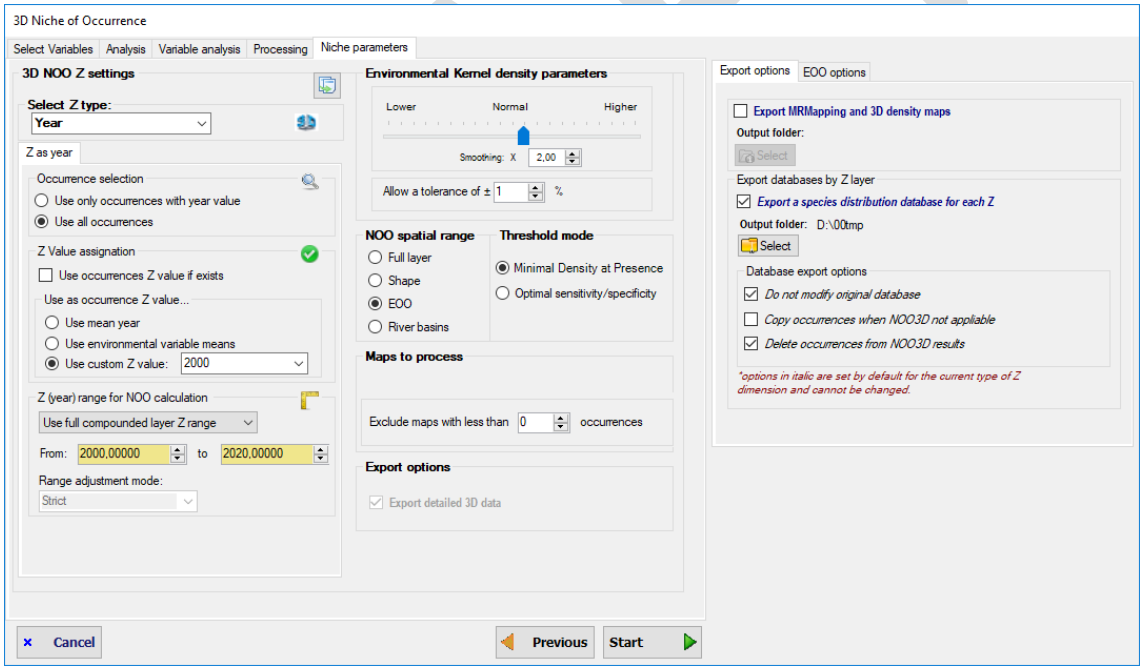



The parameters to be set in this step are already explained in section 2.12. and include setting how occurrences will be selected, which Z range and which GE will be used to calculate the contribution of each variable. Of course, the difference with the same step in MapMaker, is that here the settings cannot be checked interactively to determine if they are applicable to each

one of the selected species. It'll be at the calculation step that this checking will be done for each species, and any error reported (for example if a species doesn't have any occurrences in the selected Z range, etc).

Another noticeable difference is that while in MapMaker it is optional to calculate contributions to select most explanatory variables, here it is mandatory. This is because in another case we will be using always all the selected variables. And in this case we can just create a 3DCEL this those variables, then use it to calculate NOO3D for any set of species using the option described in section **¡Error! No se encuentra el origen de la referencia.** (NOO3D using a stored layer in DataManager). The rationale of the per-species NOO3D option in DataManager is precisely offering the possibility of using contribution calculation to select the most explanatory variables for each species in order to calculate the NOO3D, but applying it to several species in an easy way, so without having to do it species by species as it should be done in MapMaker.

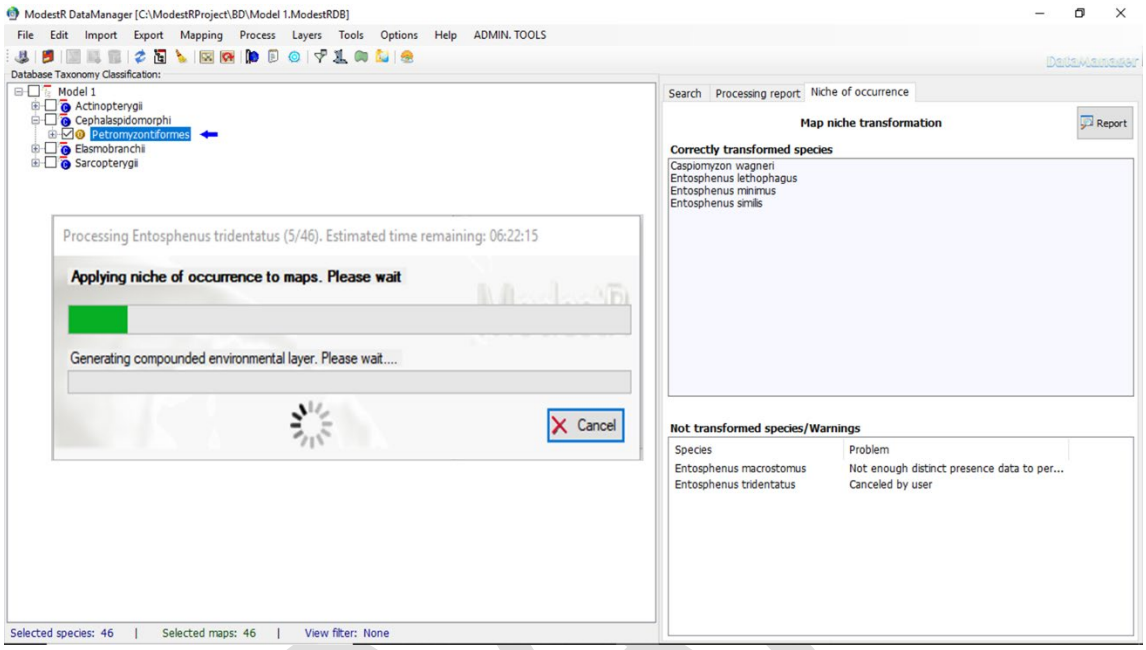
The next step allows to configure the settings corresponding to the NOO3D itself (previous settings actually determine how the variables will be selected to build a 3DCEL to be used in the NOO3D). These options include determining the Z type, the species occurrences to be used, the Z range, and the result exportation options. They have been already described in section **¡Error! No se encuentra el origen de la referencia.**.



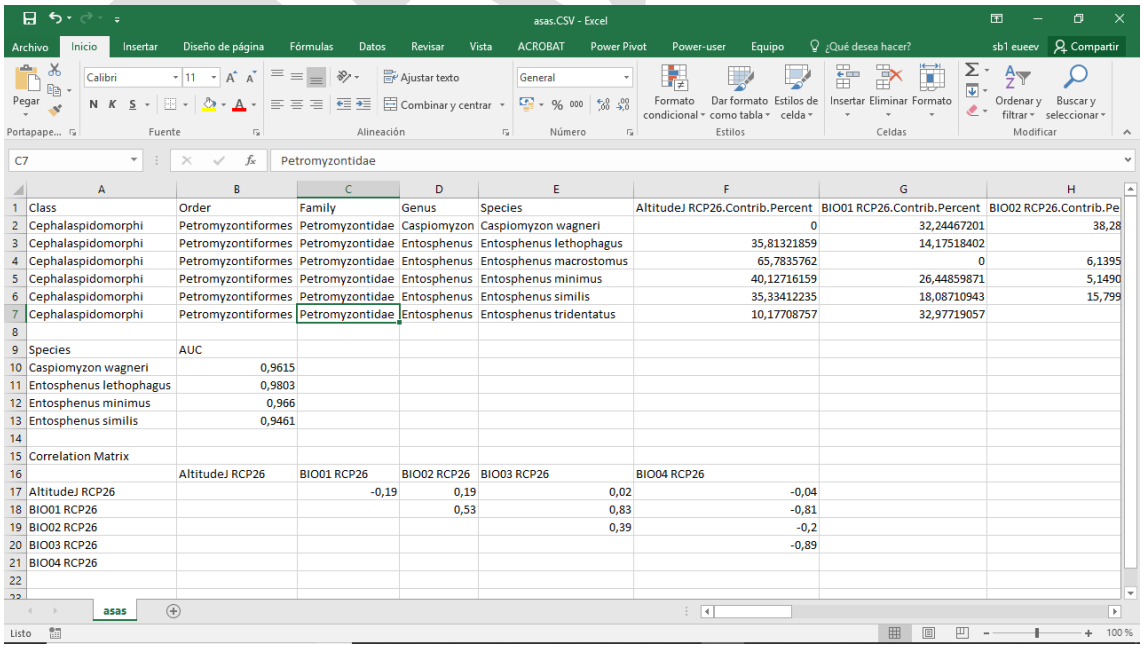

Whether using a stored 3DCEL or a per-species one, once the process started, a dialog will be shown indicating the task progress and an estimated time remaining. Take into account that this time is calculated extrapolating the time taken for the species already processed to the remaining species. Therefore it is not really reliable until a significant number of species have been processed; and even in this case, the error range is quite wide because the NOO3D involves complex iterative algorithms whose execution time cannot be accurately assessed beforehand.

Once the process completed a list of species where NOO3D was correctly applied, and another list of warnings/not applied species are shown. Both lists contents can be copy-pasted to examine them in more detail. The most common reasons why NOO3D could not be applied to a

species are the lack of enough occurrences to perform valid calculations, or to obtain the EOO[19] if this option was chosen. Other reasons are that the species occurrences are located outside of the areas covered by the selected variables, or that after applying contribution calculation only a variable appears as significantly explanatory. In this last case the origin of the problem may be that most of the selected variables are not really relevant to explain species distribution.




When using the per-species 3DCEL option, a report may also be exported in CSV format, containing the details about the contribution for each variable and each species, which explains what variables have been selected for each species.




| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | Class | Order | Family | Genus | Species | AltitudeJ RCP26.Contrib.Percent | BIO01 RCP26.Contrib.Percent | BIO02 RCP26.Contrib.Pe |
| 2 | Cephalaspidomorphi | Petromyzontiformes | Petromyzontidae | Caspiomyzon | Caspiomyzon wagneri | 0 | 32,24467201 | 38,28 |
| 3 | Cephalaspidomorphi | Petromyzontiformes | Petromyzontidae | Entosphenus | Entosphenus lethophagus | 35,81321859 | 14,17518402 | |
| 4 | Cephalaspidomorphi | Petromyzontiformes | Petromyzontidae | Entosphenus | Entosphenus macrostomus | 65,7835762 | 0 | 6,1395 |
| 5 | Cephalaspidomorphi | Petromyzontiformes | Petromyzontidae | Entosphenus | Entosphenus minimus | 40,12716159 | 26,44859871 | 5,1490 |
| 6 | Cephalaspidomorphi | Petromyzontiformes | Petromyzontidae | Entosphenus | Entosphenus similis | 35,33412235 | 18,08710943 | 15,799 |
| 7 | Cephalaspidomorphi | Petromyzontiformes | Petromyzontidae | Entosphenus | Entosphenus tridentatus | 10,17708757 | 32,97719057 | |
| 8 | | | | | | | | |
| 9 | Species | AUC | | | | | | |
| 10 | Caspiomyzon wagneri | 0,9615 | | | | | | |
| 11 | Entosphenus lethophagus | 0,9803 | | | | | | |
| 12 | Entosphenus minimus | 0,966 | | | | | | |
| 13 | Entosphenus similis | 0,9461 | | | | | | |
| 14 | | | | | | | | |
| 15 | Correlation Matrix | | | | | | | |
| 16 | | AltitudeJ RCP26 | BIO01 RCP26 | BIO02 RCP26 | BIO03 RCP26 | BIO04 RCP26 | | |
| 17 | AltitudeJ RCP26 | | -0,19 | 0,19 | 0,02 | -0,04 | | |
| 18 | BIO01 RCP26 | | | 0,53 | 0,83 | -0,81 | | |
| 19 | BIO02 RCP26 | | | | 0,39 | -0,2 | | |
| 20 | BIO03 RCP26 | | | | | -0,89 | | |
| 21 | BIO04 RCP26 | | | | | | | |
| 22 | | | | | | | | |

[19] for example alpha shape requires a certain number of occurrences not too far away from each other, depending on the alpha parameter, to be calculated

## 3.16 Batch map modification of areas or occurrences

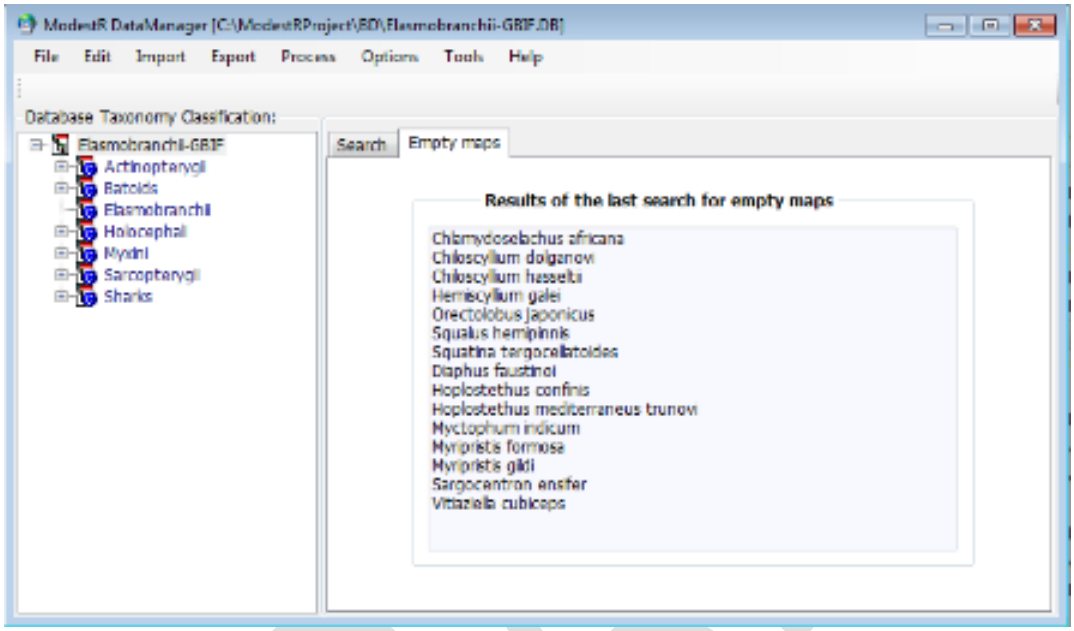
DataManager has some features that can be useful when we want to modify a set of maps in a single operation. For example to delete areas from maps, or to reassign valid habitats to occurrences of several maps. They are located in the *Edit* option of the main menu, but they can also be found on the contextual menu that appears when right-clicking on any element of the taxonomic tree:

- *Edit/ Delete areas of the checked maps:* delete presence areas of all the maps of the checked branches. Occurrences are preserved.
- *Edit/Delete occurrences of the checked maps:* delete occurrences of all the maps of the checked branches. Presence areas (polygons from range maps) are preserved.
- *Edit/Set occurrences of the checked maps/To manually validated occurrences:* set all occurrences (occurrences) of the maps of the checked branches as manually validated occurrences. That implies that all occurrences contained in a map will be considered as valid, even if they are located in habitats not valid for the species. See section 2.4 for more details about occurrence management in ModestR.
- *Edit/Set occurrences of the checked maps / To manually unvalidated occurrences:* set all occurrences (occurrences) of the maps of the checked branches as manually unvalidated occurrences. That implies that all occurrences contained in a map will be considered as invalid, regardless of the habitat they are located in. See section 2.4 for more details about occurrence management in ModestR.
- *Edit/Set occurrences of the checked maps /To habitat auto-checked occurrences:* set all occurrences (occurrences) of the maps of the checked branches as auto validated occurrences. That implies that ModestR will check if occurrences are located in habitats valid for the species to consider them valid or invalid. See section 2.4 for more details about occurrence management in ModestR.
- *Edit/Set occurrences of the checked maps /Modify occurrences valid habitats:* allow modifying the habitats that will be considered as valid for the occurrences of the selected species (presence areas of range maps will not be affected). See section 2.4 for more details about occurrence management in ModestR.
- *Edit/Set occurrences of the checked maps /Reset occurrences last modified date:* ModestR stores a date associated to the addition time or the last modification done for each occurrence of a distribution map. Even if this cannot be considered a reliable audit data, this can just help finding last changes made in occurrences of a map, particularly when several people work on the same database. When exporting occurrences from maps (e.g. exporting presence data, as explained in 6.14) this date is added to occurrence data. To reset this date for all the occurrences of several maps to a specific date, you can use the option *Edit/Set occurrences of the checked maps /Reset occurrences last modified date*.

## 3.17 Searching maps with no data

Sometimes it is useful to check if there are maps in our database that do not contain any valid data. This can happens for example if we download maps from GBIF, or we import them from

CSV files, and none of the occurrence records is classified as valid by ModestR (e.g because they are outside of the habitats of the species). Or if by mistake we save an empty map in the database. Even if an empty map or a map with no valid occurrences don't produce any problem to DataManager or MapMaker, it can be useful to be able to find them easily. To do that, use *Process/Search maps with no data* menu item in DataManager. After a moment, a list of maps with no data will be displayed in the right panel. You can even click on any species of this list to directly select it in the tree.



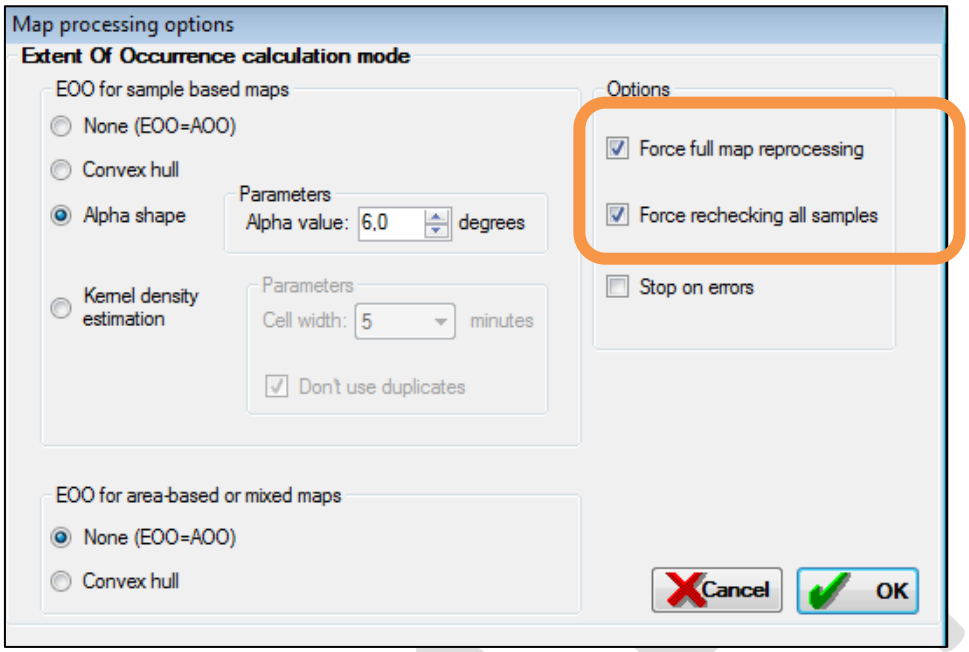

## 3.18 Forcing map updating to current map version

As explained in section 3.6 above, map processing is necessary to rasterize maps, which in turn is necessary to calculate metrics and other summary data, as well as to perform searchings with MRFinder. Map processing is necessary when you add a new map to a ModestR database, when you make modifications on a map, or also when the reference world map that ModestR software uses is updated. This last condition occurs usually because you installed an update distributed through the ModestR website. When DataManager detects that a new version of the reference world map has been installed, it will indicates that distribution maps have to be reprocessed to take into account changes in the new reference world map.

But sometimes you can be sure that some of the maps of your ModestR database don't need to be reprocessed. For example, if the modifications in the reference world map only involved freshwater, the maps of marine species are supposed not to be concerned. As map processing can be a very time-consuming task, particularly for expert-drawn maps, in those cases you can avoid having to reprocess maps by selecting the branch you want to avoid having to reprocess, and using the menu item *Tools/Force set checked maps as updated*.

This option only has to be used very carefully, and only if you clearly understand its effects. Existing processed maps will be just marked as updated, even if they were not processed using the current world reference map. This way DataManager will then consider them as already processed with the current reference world map.

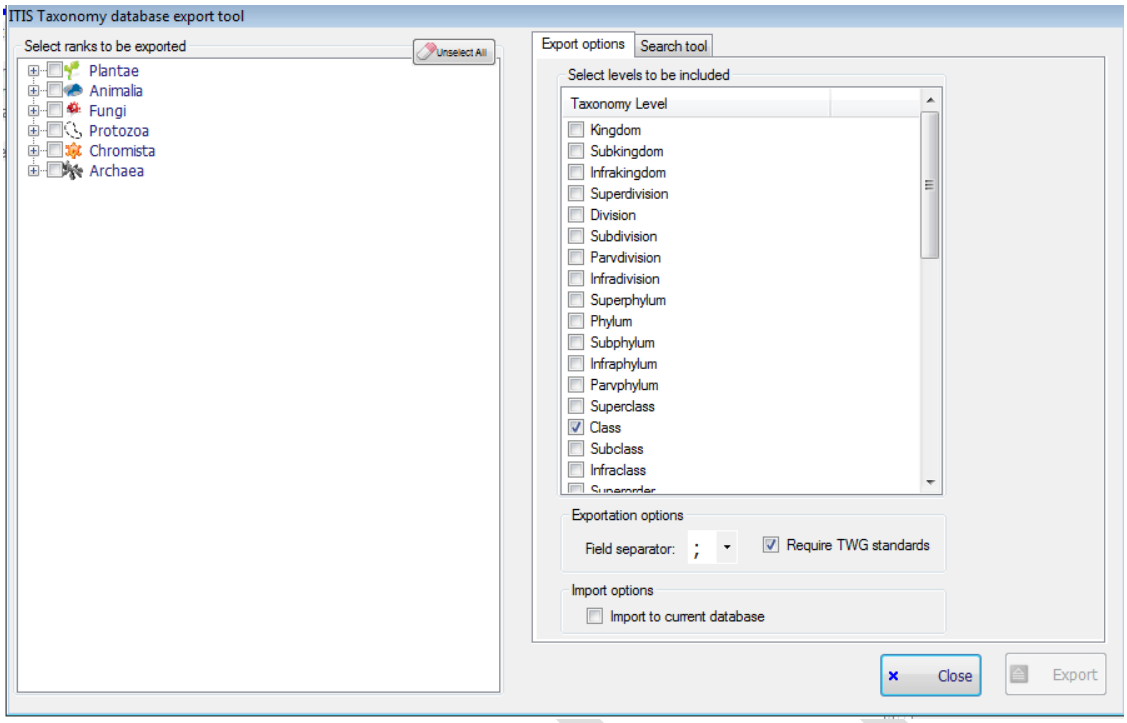
Anyway, if you incorrectly used this option, you can always reprocess all maps to come back to a correct and updated state, as explained in section 3.6 above, but forcing full map reprocessing and occurrences rechecking, to make sure that all map information is rechecked.




## 3.19 Extracting taxonomy data from ITIS database

DataManager integrates an easy to use tool to browse the ITIS taxonomy database and export data from this database to a CSV file. Those exported data can be used to subsequently import them to a ModestR database (but also for any other purpose).

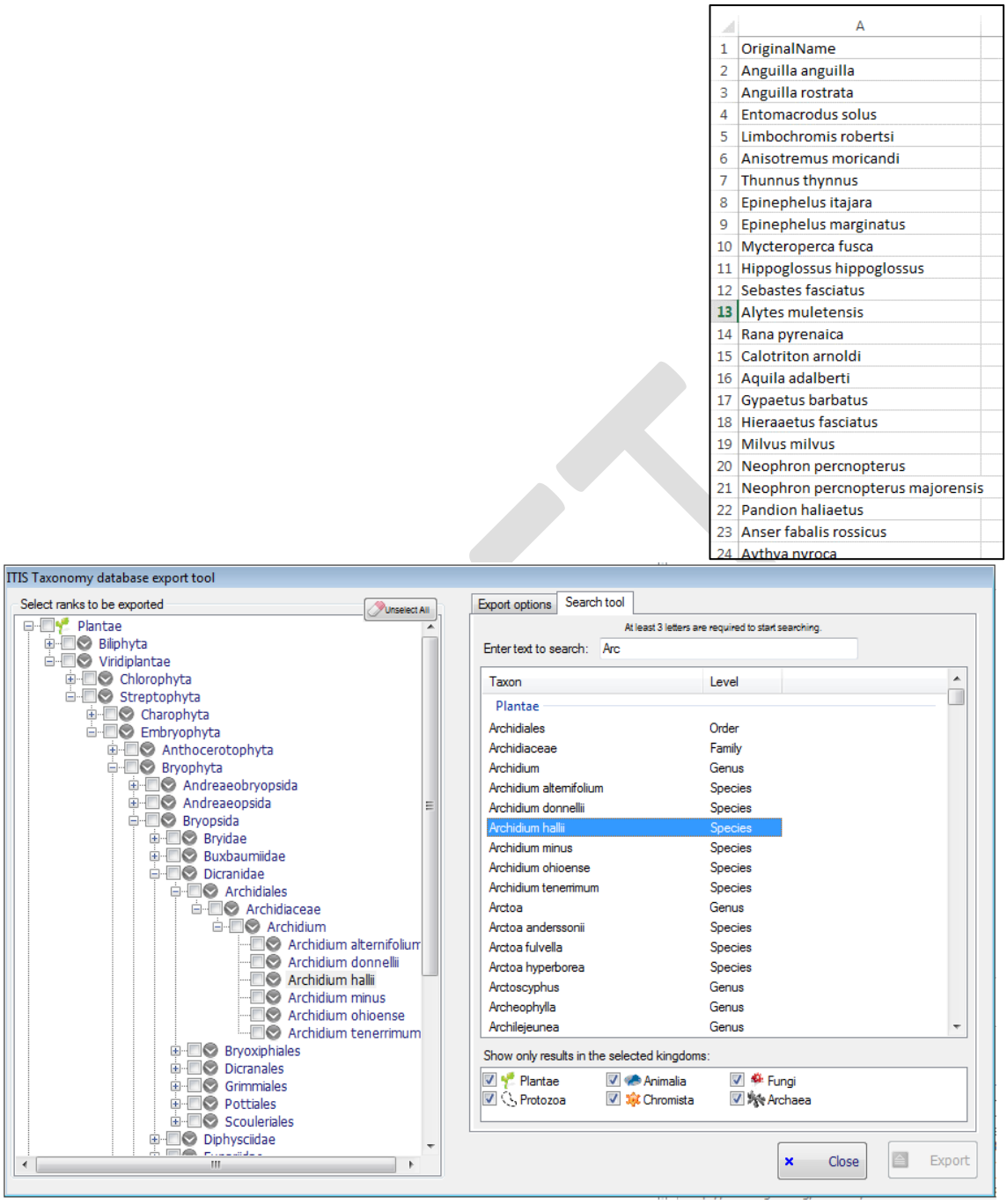
The ITIS taxonomy database can be found on the ITIS website http://www.itis.gov/ . DataManager supports browsing ITIS database in sqlite format. You can find instructions to download this database in the DataManager main menu option *Import/Taxonomy/From ITIS database/How to use ITIS database.* Once downloaded, you can browse its contents going to the menu option *Import/Taxonomy/From ITIS database/ITIS database browser tool.* You'll have to select the ITIS database file in Sqlite format. A browser will be displayed:




On the left side there is a tree where you can explore the taxonomy ranks by expanding the branches. You can also select any branch using the corresponding checkbox. All selected branches will be exported when clicking on the *Export* button located on the bottom right corner. The *Export* button will only be enabled if there is at least one branch selected to be exported.

On the right size of the dialog there are two tabs. The first one is the *Export options* tab; in this tab you can select the levels to be included when exporting data. By default, only levels supported by ModestR are selected. You can select other levels, but in this case exported data would not be suitable to be imported to a ModestR database. In this tab you can also change the field separator to be used when exporting data to CSV, and require that only taxonomic ranks that meet Taxonomic Work Group standard will be shown and exported. Finally you can enabled the option for the exported data to be imported to currently opened ModestR database. In this case, after exporting data from ITIS database to a CSV file, DataManager will import this same file to the current ModestR database.

The other tab is the *Search tool* tab. In this tab you can enter a text in a textbox, and all ranks from the ITIS database that start with this text will be shown on the list below. At least 3 characters are required to start searching. You can filter results by kingdoms using the checkboxes on the bottom. Once results shown in the list, you can click on any of them to directly find it in the left tree.




Once found the ranks you are interested in, you can select them on the left side tree. Clicking on the *Export* button located on the bottom right corner all selected branches will be exported to a CSV file. If you enabled the option for the exported data to be imported to currently opened ModestR database, DataManager will directly start the importation of this same file to the current ModestR database.

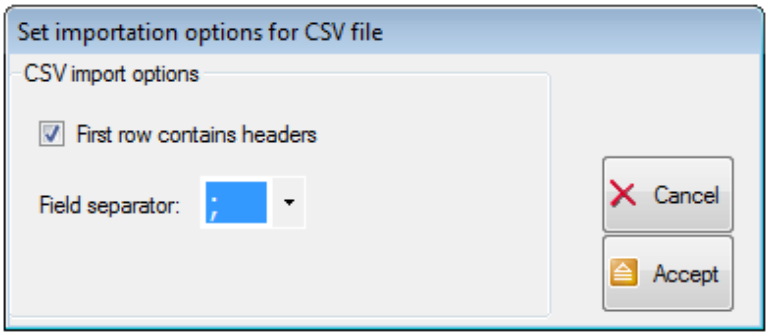
## 3.20 Taxonomy searching and checking utility

One quite frequent problem when using taxonomy data from different sources is the discrepancy between those sources. Other times, the problem is that we have for example a list of species, but not the rest of the taxonomic levels that ModestR requires to add them to a

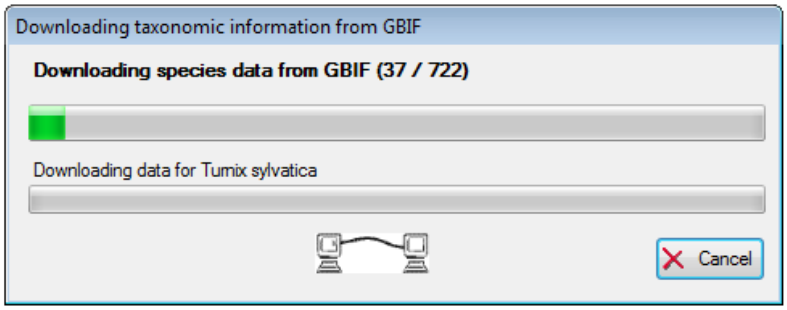
database (class, order, etc.). To try to help solving those kind of issues, a simple tool has been added to DataManager that allows searching for taxonomic data for a list of species in GBIF online database.

The first step is preparing a file in CSV format containing the list of species to be checked. This can be done with any text editor or a worksheet like Excel, that allows exporting to CSV format.

The next step is going to DataManager, and select the menu option *Import/Import taxonomy/GBIF taxonomy search tool*. A dialog box allows us to indicate if the input file contains a header (in this case the first line will be just skipped) and which separator is used in the CSV file.




After clicking on the *Accept* button, we'll have to select the folder and the filename where the results will be saved. Then DataManager will start the searching task, looking for taxonomic information for each species in GBIF online database.

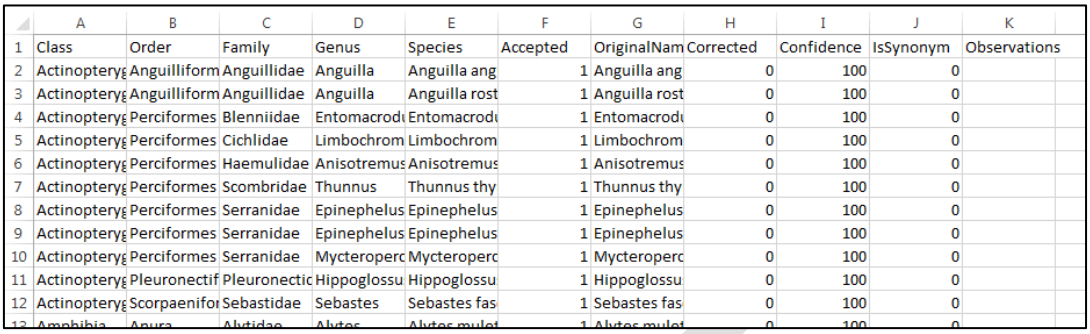



Once the task completed, a CSV file with the results will be saved. It can be easily opened with a worksheet like Excel or any other compatible application.

This file will contain the following columns:

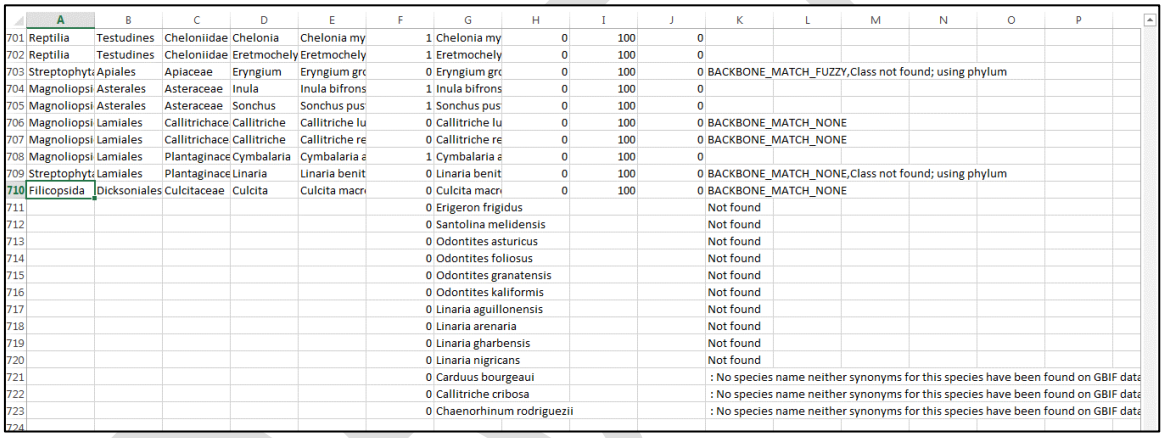
- Class, order, family, genus, species: those columns correspond the the taxonomic classification of each original species of the input file that has been found in GBIF.
- Accepted: indicates using a 1 (=yes) or a 0 (=no) if the original species of the input file is an accepted species according to GBIF.
- OriginalName: is the original species name of the input file (that can be different from the name returned by GBIF, in case of synonyms or corrections)
- Corrected: indicates using a 1 (=yes) or a 0 (=no) if GBIF returned a correction the original species of the input file is an accepted species according to GBIF.
- Confidence: when a correction has been suggested, this indicates its % confidence according to GBIF (or 100% for another case).

- IsSynonym: indicates using a 1 (=yes) or a 0 (=no) if GBIF considered that the original species name was a synonym of the returned species name.
- Observations: other observations, if any.



| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Class | Order | Family | Genus | Species | Accepted | OriginalNam | Corrected | Confidence | IsSynonym | Observations |
| 2 | Actinopteryg | Anguilliform | Anguillidae | Anguilla | Anguilla ang | 1 | Anguilla ang | 0 | 100 | 0 | |
| 3 | Actinopteryg | Anguilliform | Anguillidae | Anguilla | Anguilla rost | 1 | Anguilla rost | 0 | 100 | 0 | |
| 4 | Actinopteryg | Perciformes | Blenniidae | Entomacrod | Entomacrod | 1 | Entomacrod | 0 | 100 | 0 | |
| 5 | Actinopteryg | Perciformes | Cichlidae | Limbochrom | Limbochrom | 1 | Limbochrom | 0 | 100 | 0 | |
| 6 | Actinopteryg | Perciformes | Haemulidae | Anisotremus | Anisotremus | 1 | Anisotremus | 0 | 100 | 0 | |
| 7 | Actinopteryg | Perciformes | Scombridae | Thunnus | Thunnus thy | 1 | Thunnus thy | 0 | 100 | 0 | |
| 8 | Actinopteryg | Perciformes | Serranidae | Epinephelus | Epinephelus | 1 | Epinephelus | 0 | 100 | 0 | |
| 9 | Actinopteryg | Perciformes | Serranidae | Epinephelus | Epinephelus | 1 | Epinephelus | 0 | 100 | 0 | |
| 10 | Actinopteryg | Perciformes | Serranidae | Mycteroperc | Mycteroperc | 1 | Mycteroperc | 0 | 100 | 0 | |
| 11 | Actinopteryg | Pleuronectif | Pleuronectic | Hippoglossu | Hippoglossu | 1 | Hippoglossu | 0 | 100 | 0 | |
| 12 | Actinopteryg | Scorpaenifor | Sebastidae | Sebastes | Sebastes fas | 1 | Sebastes fas | 0 | 100 | 0 | |
| 13 | Amphibia | Anura | Alytidae | Alytes | Alytes mulet | 1 | Alytes mulet | 0 | 100 | 0 | |

At the end of this list, the species that were not found in GBIF (if any) will appear:



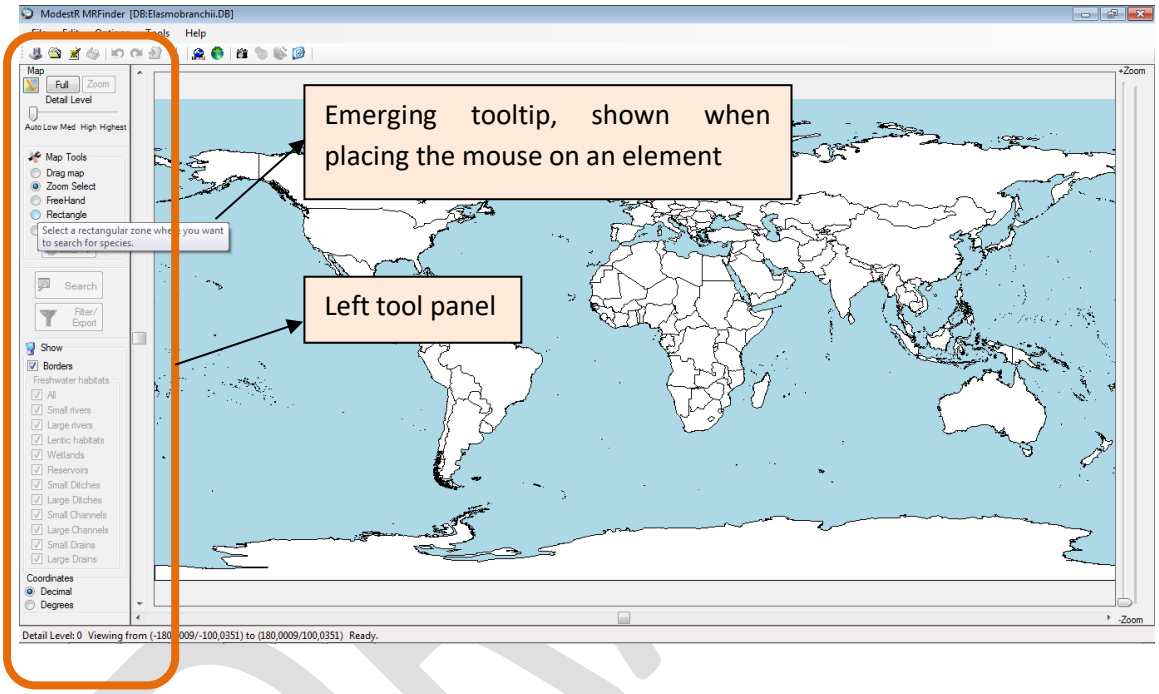
| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 701 | Reptilia | Testudines | Cheloniidae | Chelonia | Chelonia my | 1 | Chelonia my | 0 | 100 | 0 | | | | | | |
| 702 | Reptilia | Testudines | Cheloniidae | Eretmochely | Eretmochely | 1 | Eretmochely | 0 | 100 | 0 | | | | | | |
| 703 | Streptophyt | Apiales | Apiaceae | Eryngium | Eryngium gr | 0 | Eryngium gr | 0 | 100 | 0 | BACKBONE_MATCH_FUZZY,Class not found; using phylum | | | | | |
| 704 | Magnoliopsi | Asterales | Asteraceae | Inula | Inula bifrons | 1 | Inula bifrons | 0 | 100 | 0 | | | | | | |
| 705 | Magnoliopsi | Asterales | Asteraceae | Sonchus | Sonchus pus | 1 | Sonchus pus | 0 | 100 | 0 | | | | | | |
| 706 | Magnoliopsi | Lamiales | Callitrichace | Callitriche | Callitriche lu | 0 | Callitriche lu | 0 | 100 | 0 | BACKBONE_MATCH_NONE | | | | | |
| 707 | Magnoliopsi | Lamiales | Callitrichace | Callitriche | Callitriche re | 0 | Callitriche re | 0 | 100 | 0 | BACKBONE_MATCH_NONE | | | | | |
| 708 | Magnoliopsi | Lamiales | Plantaginace | Cymbalaria | Cymbalaria a | 1 | Cymbalaria a | 0 | 100 | 0 | | | | | | |
| 709 | Streptophyt | Lamiales | Plantaginace | Linaria | Linaria benit | 0 | Linaria benit | 0 | 100 | 0 | BACKBONE_MATCH_NONE,Class not found; using phylum | | | | | |
| 710 | Filicopsida | Dicksoniales | Culcitaceae | Culcita | Culcita macr | 0 | Culcita macr | 0 | 100 | 0 | BACKBONE_MATCH_NONE | | | | | |
| 711 | | | | | | 0 | Erigeron frigidus | | | | Not found | | | | | |
| 712 | | | | | | 0 | Santolina melidensis | | | | Not found | | | | | |
| 713 | | | | | | 0 | Odontites asturicus | | | | Not found | | | | | |
| 714 | | | | | | 0 | Odontites foliosus | | | | Not found | | | | | |
| 715 | | | | | | 0 | Odontites granatensis | | | | Not found | | | | | |
| 716 | | | | | | 0 | Odontites kaliformis | | | | Not found | | | | | |
| 717 | | | | | | 0 | Linaria aguillonensis | | | | Not found | | | | | |
| 718 | | | | | | 0 | Linaria arenaria | | | | Not found | | | | | |
| 719 | | | | | | 0 | Linaria gharbensis | | | | Not found | | | | | |
| 720 | | | | | | 0 | Linaria nigricans | | | | Not found | | | | | |
| 721 | | | | | | 0 | Carduus bourgeaui | | | | : No species name neither synonyms for this species have been found on GBIF data | | | | | |
| 722 | | | | | | 0 | Callitriche cribosa | | | | : No species name neither synonyms for this species have been found on GBIF data | | | | | |
| 723 | | | | | | 0 | Chaenorhinum rodriguezii | | | | : No species name neither synonyms for this species have been found on GBIF data | | | | | |
| 724 | | | | | | | | | | | | | | | | |

As the first columns of this CSV file contain the taxonomic levels (from class to species) required by a ModestR database, this kind of CSV file can be imported in DataManager to add taxonomic data to a database, as explained in section 3.3. But is important to point out that is necessary to review it before, and, particularly:

- Review and check rows where there are issues (observations column), because they can involve incoherencies (i.e. different names for a same family) that can provoke importation errors if you import those data to a ModestR database.
- Delete the species for which no taxonomic data have been found in GBIF.

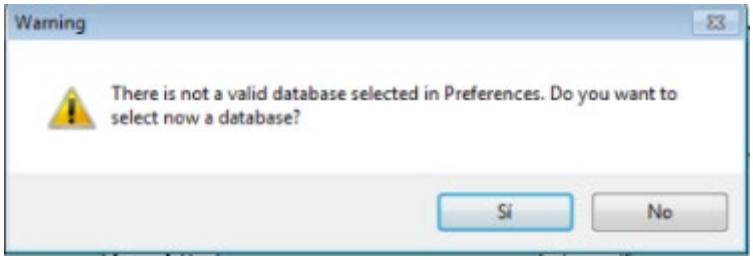
# 4 First steps in MRFinder

MRFinder is an easy-to-use software application to make the most of the data stored in ModestR databases created with DataManager. MRFinder allows for finding which species are present in any area of the world and to generate and export several data reports and statistics about those species. The MRFinder user interface is quite simple and very similar to MapMaker, as the most usual tools are always visible at the left tool panel. You also have contextual menus that will appear when you click with the right button of the mouse on the map. Also, quite all menus, buttons and other elements have a tooltip that briefly explains their usage, and that will be shown when you place the mouse over an element.




## 4.1 Default settings in MRFinder

MRFinder mandatorily has to work using the data of a ModestR database. A database can be selected using the *File/Open database* menu option (MRFinder will automatically use the last used database when running it again). If no database is selected, or the default database is not found, a dialog box will be shown when you try to work with data, allowing you to select a database to be used during this session.




Using the *Options/Preferences* menu, a dialog box will be shown where several options can be set, such as the colors of the elements and their default visibility status; delimiter and decimal separator for CSV exportation format can also be set here.

## 4.2 Moving across the world map

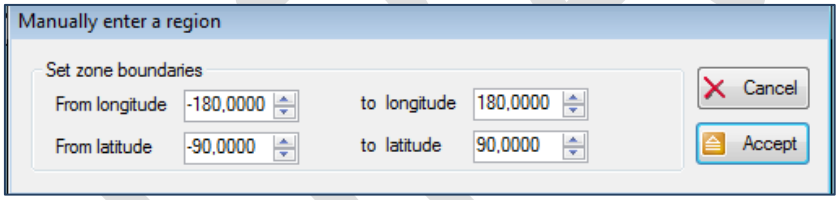
MRFinder offers the same features available in MapMaker to move across the map and zoom in and out. Please see subsection 2.1 for more details. Moreover, the same informative elements that can be added to the map in MapMaker area available in MRFinder and work in the same way. See subsection 2.18.1 for more details.

The same as in MapMaker, you have *Undo* and *Redo* options in MRFinder. Operations that can be undone in MRFinder are those that affect the selections you add to the map.

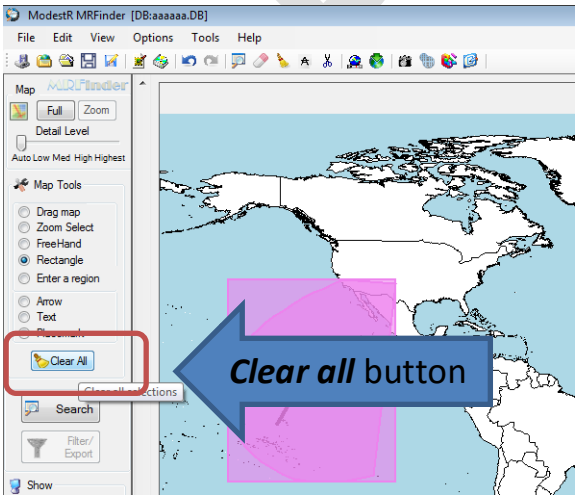
## 4.3 Adding selections to search for species

To find which species are present in an area, you simply have to select on the map the desired zone, typically using the *Freehand* or the *Rectangle* tool. It must be pointed out that *Freehand* tool allows to search for species in an irregular area, which can be a very flexible feature, but being more complex, it will take more time to process, so we recommend you to use *Rectangle* for searching in rectangular areas, and use *Freehand* only when searching in irregular areas is needed.

Another option available in the left tool panel to select an area is the option *Enter a region* to manually enter a region. In this case, a dialog box will be shown where the coordinates of a rectangular region can be entered. When accepting it, this rectangle will be shown in the map. The default coordinates that will be shown in this dialog box will be took from the current selection, if any. This way, you can make an approximate selection in the map using the *Rectangle* tool, and then select to manually enter a region, and just adjust the desired region with more precision.




You can select several areas to perform a search. So you can add several areas to the map. MRFinder allows selections that totally or partly overlap. To clear them, click on the *Clear all* button on the left tool panel.



It is important to indicate that MRFinder can search species in several selections at the same time, but that a single selection in MRFinder can contain several polygons (i.e. several areas). When the user adds selections using the *Rectangle* or the *Freehand* tools, each selection (i.e. each polygon) he/she adds is considered a different selection. But this can be changed using the *Selections editor* (explained later in this chapter) to merge several selections in one, for example. Moreover, when you add a predefined shape (see later in this chapter), this selection commonly contains several polygons. For example, if you add a predefined shape of Spain, this shape will include several polygons corresponding to all Spanish territories including islands. When searching for species, MRFinder will treat this selection as a single one, even if it is composed of several polygons. That is, distinction between selections is done by the user, and not by the fact that a selection may contain several areas. To see which logical selections are currently add to a map and manage them (delete, rename or merge them), you can use the selections editor (see section 4.10).

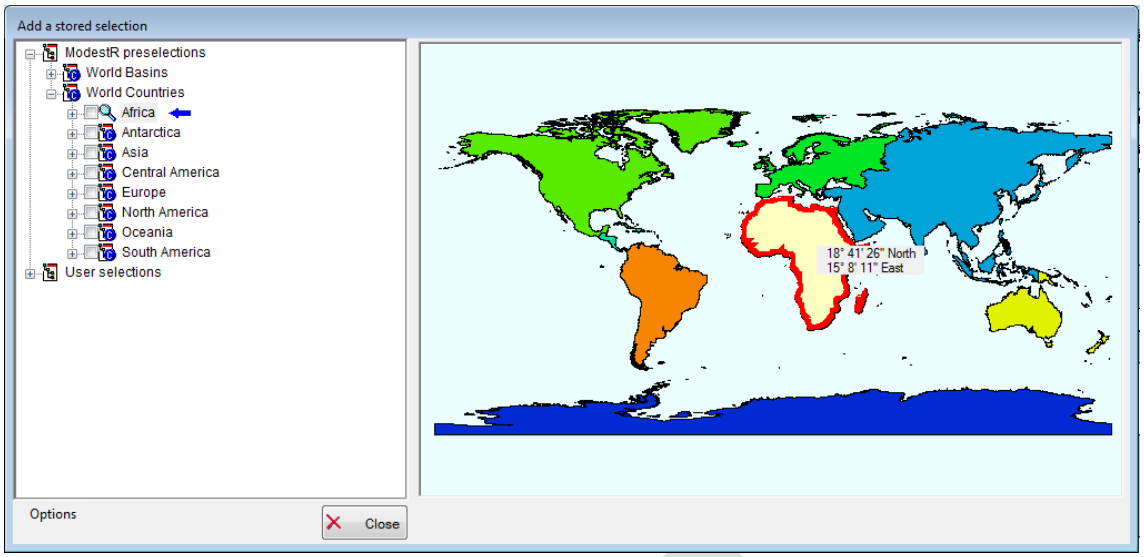
## 4.4 Importing selections

Besides manually selecting searching areas, MRFinder provides the possibility of importing existing data as selections. For example, you can import a shapefile or a KML file with one or more administrative areas. To do that use *File/Import selection* option from main menu. When importing shapefiles a dialog box will be shown where you can enter a simplification tolerance. We strongly recommend to use the default value, as it avoids some problems with shapefiles (such as autointersections) and it involves a minimal precision loss.

If you are familiar with shapefile format, you can also add importation criteria to filter which data will be imported. To do that, use the *See/hide fieldlist* button to see the list of fields contained in the shapefile. Then you can select a field to see the values it takes, and add criteria that will be used as filter when importing. You can also directly write a condition in the textbox. The syntax of the conditions is similar to the syntax of a WHERE clause in SQL.

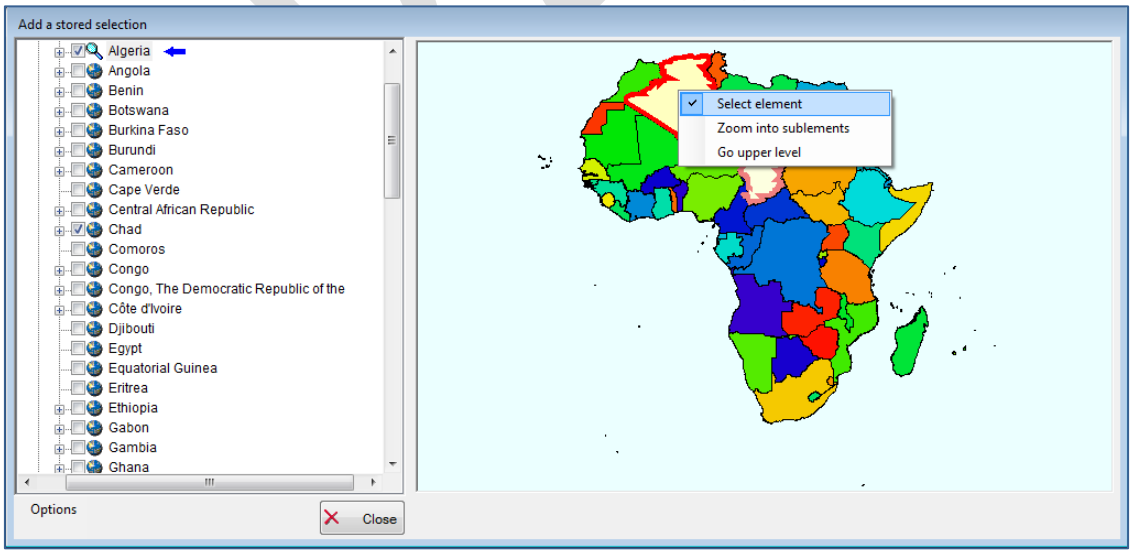

## 4.5 Using predefined shapes

To avoid having to make or to import a selection each time we want to use it to perform a search, MRFinder can store predefined shapes then you can easily reuse them. All predefined shapes can be shown going to *Edit/Add predefined shapes* menu, or the corresponding button of the toolbar. A dialog box will be shown where predefined shapes are shown in a tree. Those provided with ModestR are displayed under the "ModestR shapes" node of the tree, but you can also add your own predefined shapes, that will appear under the "User shapes" node of the tree (later in this chapter we explain how to store your own selections in this tree).



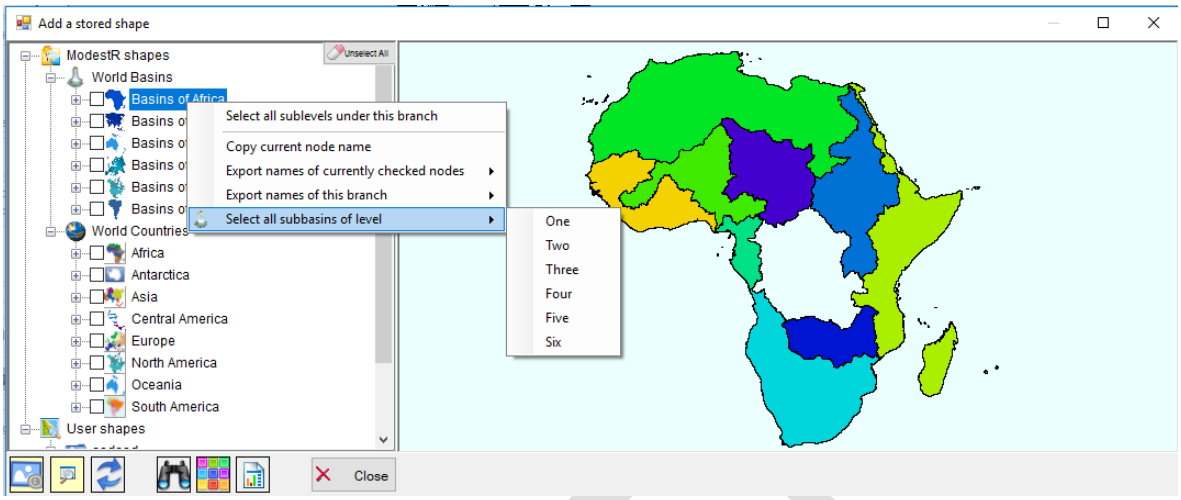

To make easier browsing predefined shape, an active map is shown by default on the right side of the dialog. This active map shows a preview (when available) of the element selected in the left tree. Moreover, this preview can be interactive in some case. For example, if you select “Word countries” in the left tree, a world map will be shown on the right panel. You can double-click on a continent to see details of this continent. As you can see, this active preview panel is synchronized with the left tree, the way that when you select an element on one of them, this element will also be selected on the other one.

To add a shape to the map, you just have to check it on the tree. You can also right-click on the left active preview and select the element in the context menu.




Some nodes of the tree can be not checkable as they are just here to group shapes. And other ones can automatically include other nodes when checked. For example a node “Europe” can automatically add all the European countries as shapes to the map. That depends on how the tree is organized. You can add one or more shapes in this way.

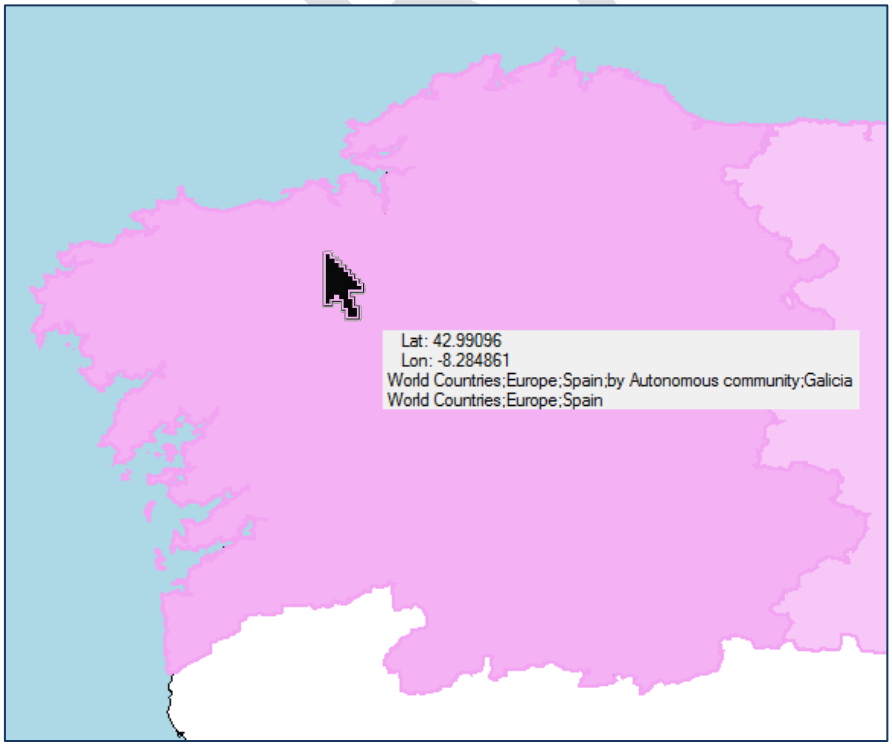
When working with basins you can also select all subbasins of a specific level from one basin directly clicking with the right button on the wanted basin (ej: Basins of Africa, or World Basins), then go to *Select all subbasins of level/* and select level.




Once selected the shapes to be added click on the *Close* button to go back to the main window. By default, MRFinder will zoom on the area where the current selections are located (this can be modified in the *Options* menu of the predefined shapes dialog box).

### 4.5.1 Querying for selections in a point of the map

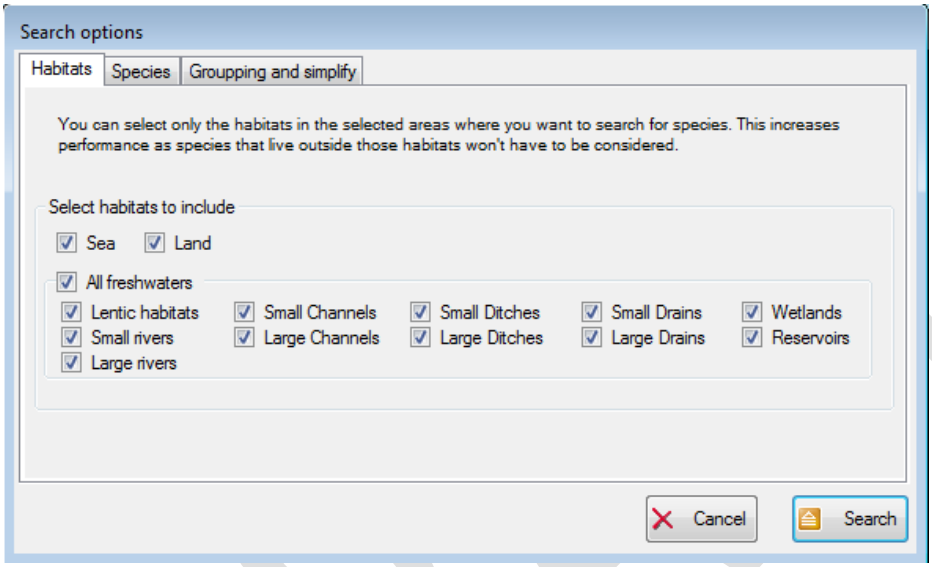
When having several selections on the world map, moreover if they overlap, it can be sometimes difficult to know which are the selections in a specific point. To query that, you have just to place the mouse in the map. If one or more selections are currently present at this point of the map, this information will appear under the information of longitude and latitude.



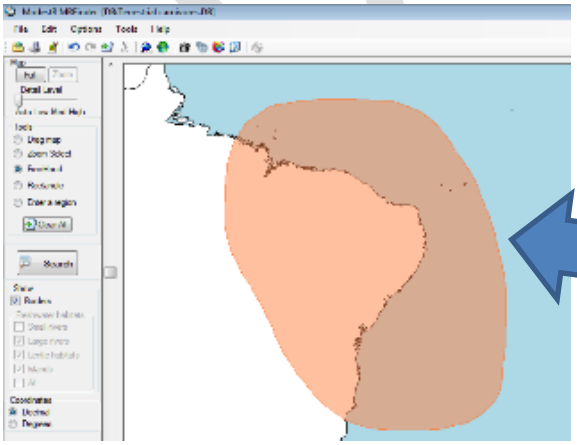
## 4.6 Search options

Once one or more search areas (selections) have been added to the map, you can click on the *Search* button that will be enabled at the left tool panel. MRFinder will use rasterized (that is, processed) maps from the currently selected ModestR database to perform the search, so it is necessary that all maps have been previously processed using DataManager to obtain accurate results (see section 3.6). MRFinder will warn you if you try to make a search using a database where not all maps have been processed, but allows you to continue anyway. In this case, only processed maps will be took into account to do the search.

Before performing the search MRFinder will display a dialog to configure search options.




In this dialog box there are 3 tabs. The first one is *Habitats* tab. There you can select the habitats that will be took into account in the search. For example if you select sea, only species that are present in sea will be checked. Besides increasing performance (as some species will be early excluded from search), this allows to focus on specific species. Moreover, it makes unnecessary to do very precise selections to perform search in specific habitats.

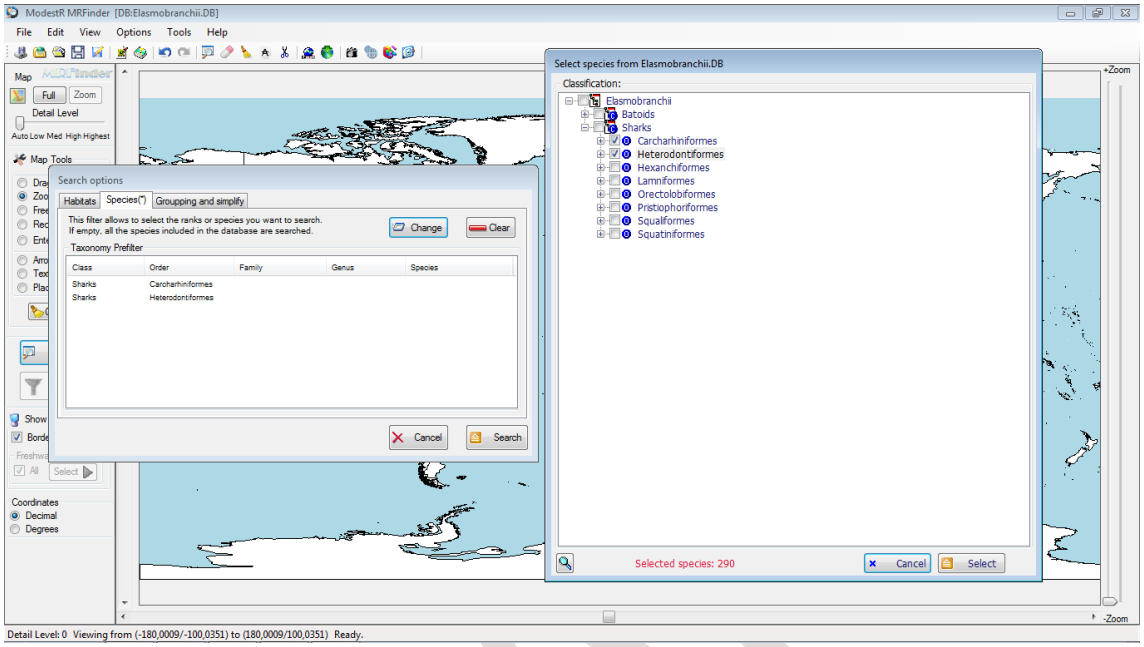


We can do an "imprecise" selection that includes for example land portions, but if posteriorly we only select to search in sea habitats, only species that are present in the sea portion included in the area selected will be found. Species only present in the land will be excluded from search.

The next tab is the *Species* tab, where you can define a previous filter of the ranks you want to search for. By default this filter is empty, meaning that the search will include all the species of any rank from the selected database. But it can happens that you only want to search for species

of some particular families, for example. In this case, you can click on the *Change* button. A tree showing all the taxonomy of the current database will appear. You can check in this tree any ranks you want. In this way, MRFinder will only search for the species of those ranks when performing the search.

You can always clear any filter using the *Clear* button in this tab.



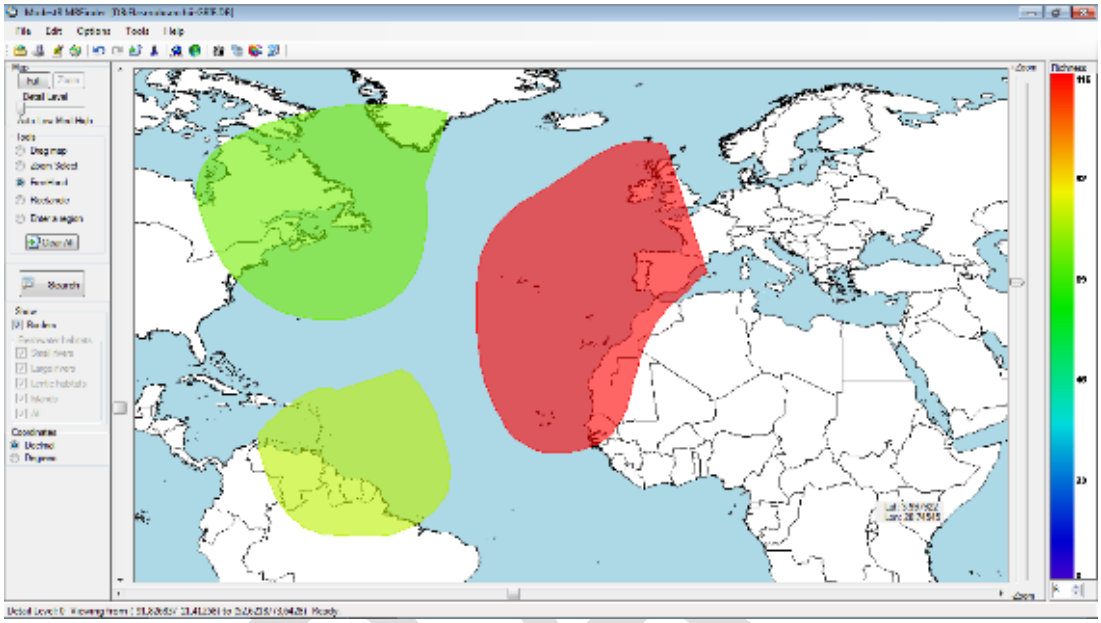

The last tab is the *Grouping and simplify* tab. There are here the group options. If you selected more than one area to search for, you can choose between a separated or a grouped search. In the first case, MRFinder will perform a search in each area, and you will be able to obtain a list of species found in each area separately. In the second one, MRFinder will perform a search in all areas at the same time, so results will contain all species present in any of the areas. The second options is obviously faster, so if you do not need the information provided by the first option, we recommend you to select this second one.

In this tab there is also the simplify option: this option will makes MRFinder preprocess selected searching areas to simplify them. This is particularly useful when using complex selections, such as some selections that can be imported from very detailed shapefiles, or complex predefined shapes (see section 4.4 for more details about predefined shapes). In most of the cases, we don't need to perform a search with more than ±1 minute of precision, and in turn a slight simplification can dramatically improve search performance. However, in order to really understand we recommend you to manually test the effects of simplifying selections using the *Simplify* tool on a selection before performing a search (see section 4.12.1). If you need a very accurate search, you can disable this option. Moreover, MRFinder also automatically disable it when selections are simple (usually selections made using *Freehand, Rectangle* or *Enter a region* tools described above) as simplification will not provide significant performance increase for those selections. Anyway, if you use this option, you can set the tolerance or level of simplification in minutes that will be applied. Again we refer you to section 4.12.1 to manually test the effects of simplifying selections using the *Simplify* tool on a selection.

## 4.7 Performing search

Once search options set in the *Search* options dialog box, and the *Search* button clicked in this dialog, MRFinder will perform the search, which can require from few minutes to much more time, if the selections are complex and there are many species in the database.

If there are more than one search area, during and after the search, a color scale will appear at the right side representing the richness or number of found species of each selected area. Each search area will take the corresponding color during the search process.



This color scale as well as the selected areas with their corresponding color will be exported if you export the map as a JPEG image (*File/Export/Current map view as image*). Scale and colors will be cleared when modifying search areas. As it is explained below, you can filter search results to select only some of the species found; it this case, the color scale and the colors of the areas will be automatically updated to take into account only filtered results.

It is important to remember, as said before (see 1.1) that, ModestR is intended to work with a resolution of 1 arcminute (1'). As a consequence, MRFinder is not designed to accurately perform searchings in areas smaller than a 1' width or a 1' height. Moreover, searching performed by MRFinder is done with a precision of ±1'. And as said above (see 1.1), ModestR assumes that the area of an occurrence is equal to the area of the cell of 1'x1' where it is located; subsequently, MRFinder can return in the search results a species with an occurrence located in a 1'x1' cell even partially included in the search area defined for the searching.

## 4.8 Filtering seach result

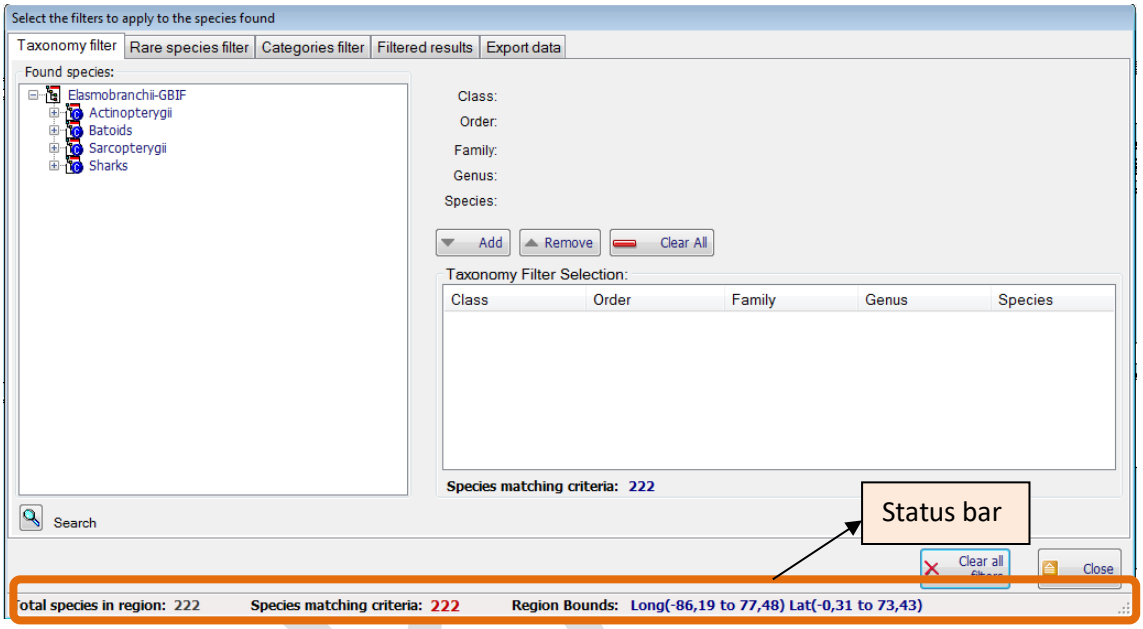
Once the search completed, if species have been found in the selected area, a new window will be shown, where two basic tasks can be performed: filtering which species among the found ones will be processed, and process those selected species to generate and export several data.

This dialog box consists for several pages that can be selected using the top tabs. Pages one to three offer different filtering options. Page four (*Filtered results*) allows you to see in any moment which species are currently selected, from the species found in the map searching,

according to the current filter settings. Last page shows the different export options, that will be applied only to the selected species after filtering.

At the bottom of this dialog box you can see a status bar. In this bar there are two counters:

- A counter labeled *Total species in region* that indicates to total of species found in the last search, including all areas selected.
- A counter labeled *Species matching criteria* indicates how much species are currently selected, according to the filters settings. Initially this counter will indicates the total of species found in the selected areas.




Following each page of this window will be explained.

### 4.8.1 Taxonomy filter page

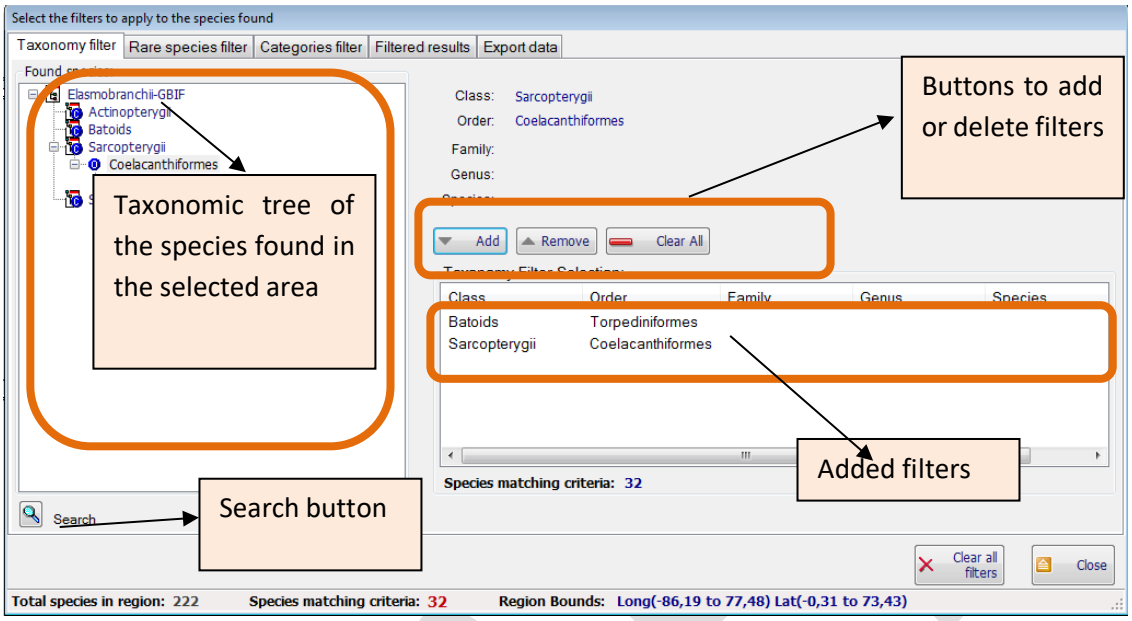
In this page you can apply filters on the search results using taxonomic criteria. On the left side of this page you can see a taxonomic tree similar to the one shown in DataManager when exploring a ModestR database (see section 3.1). The difference, of course, is that only species (and their corresponding higher taxonomic levels) found in the searching are shown in this tree. You can explore it, and also use the *Search* button to look for some taxon, in a similar way than using the search panel of DataManager (see section 3.1).

You can also see three buttons at the right-up of the tree that allow to add/remove filters. If you want to select only some taxa from the species found in the area, you can add filters just by selecting in the tree the taxon you want to include (it can be any level from class to species) and clicking on *Add* button. Filters will be added to the *Filter selection* list box.

If you add several filters they will work as an OR operation, that is, species that comply with *any* of the filters (at least one of them) will be included in later data processing. When you add a filter, the *Species matching criteria* will be updated to show how much species are currently selected. If no species comply with the filters, it will indicate zero and exportation tools will be

disabled. As said before, you can also go to *Filtered results* page to see the species that currently comply with all the filters (including the taxonomic one).

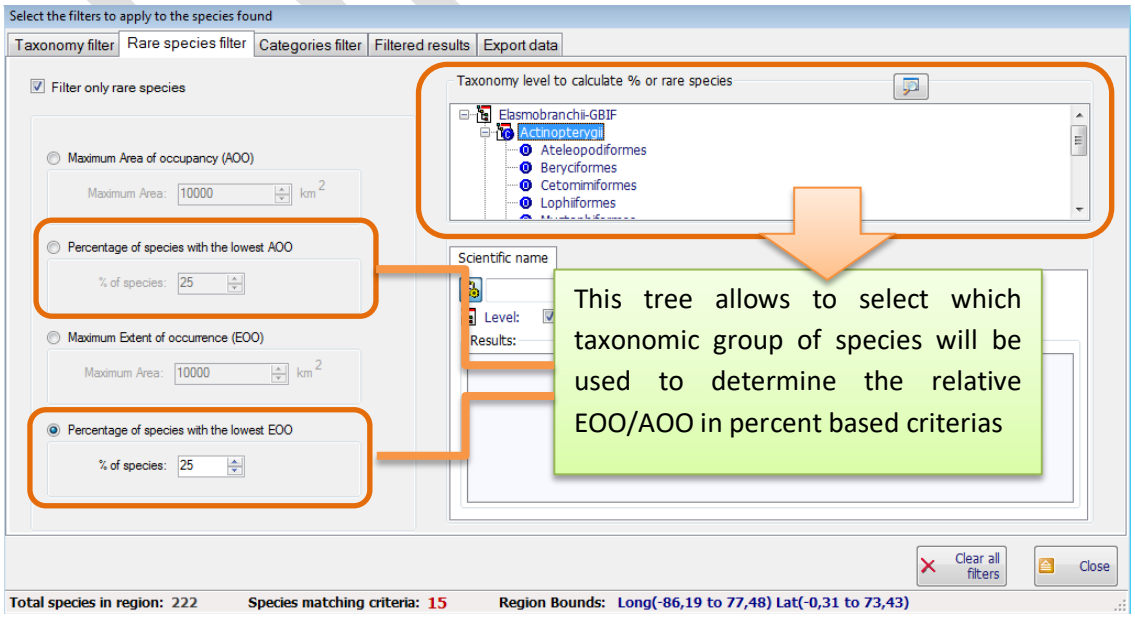
You can remove a filter o clear all using the *Remove* or *Clear filter* buttons respectively.




## 4.8.2 Rare species filter page

The second page is the *Rare species filter* page. Here you can set a filter regarding the area occupied by the species, which is particularly aimed to select endemic or rare species. This can be done checking the *Filter only rare species* checkbox to activate the filter. When using this feature, you have to set the criterion to decide when a species is considered as rare.

To do that, use the button at left of this checkbox, that will show a dialog box where you can set the criteria applied to determine if a species is rare:




The available options are:

- Maximum Area of Occupancy (AOO): a species will be considered rare if its total presence area all over the world (AOO) is less than the value you can enter in $Km^2$.
- Percentage of species with the lowest AOO: a species will be considered rare if its total presence area all over the world (AOO) is in the range of the X% species with the lowest AOO, where X is the value you enter in the corresponding field. In this case, you can also select the taxonomic level (for example a specific class or order) that will be took into account to calculate this %, just by selecting it on the tree at the right side. By default all species in the database are took into account.
- Maximum Extent of Occurrence (EOO): a species will be considered rare if its EOO (calculated when processing the species in DataManager) all over the world is less than the value you can enter in $Km^2$.
- Percentage of species with the lowest AOO: a species will be considered rare if its EOO (calculated when processing the species in DataManager) all over the world is in the range of the X% species with the lowest EOO, where X is the value you enter in the corresponding field. In this case, you can also select the taxonomic level (for example a specific class or order) that will be took into account to calculate this %, just by selecting it on the tree at the right side. By default all species in the database are took into account.

As said before, you can go to *Filtered results* page to see the species that currently comply with all the current filter settings (remember that all filters are applied to obtain those filtered results).

### 4.8.3 Categories filter page

The *Categories filter* page allows to load a categories file and use it as filter. A categories filter consists of a CSV file where the first column contains species and the second column contains a category assigned to each species.

**Example of categories file**

Aonyx congicus;Least Concern
Arctictis binturong;Vulnerable
Arctocephalus pusillus;Other
Arctogalidia trivirgata;Least Concern
Atelocynus microtis;Near Threatened
Atilax paludinosus;Least Concern
Bassaricyon alleni;Least Concern
Bassaricyon beddardi;Least Concern
Bassaricyon gabbii;Least Concern
Bassaricyon lasius;Data Deficient
…….

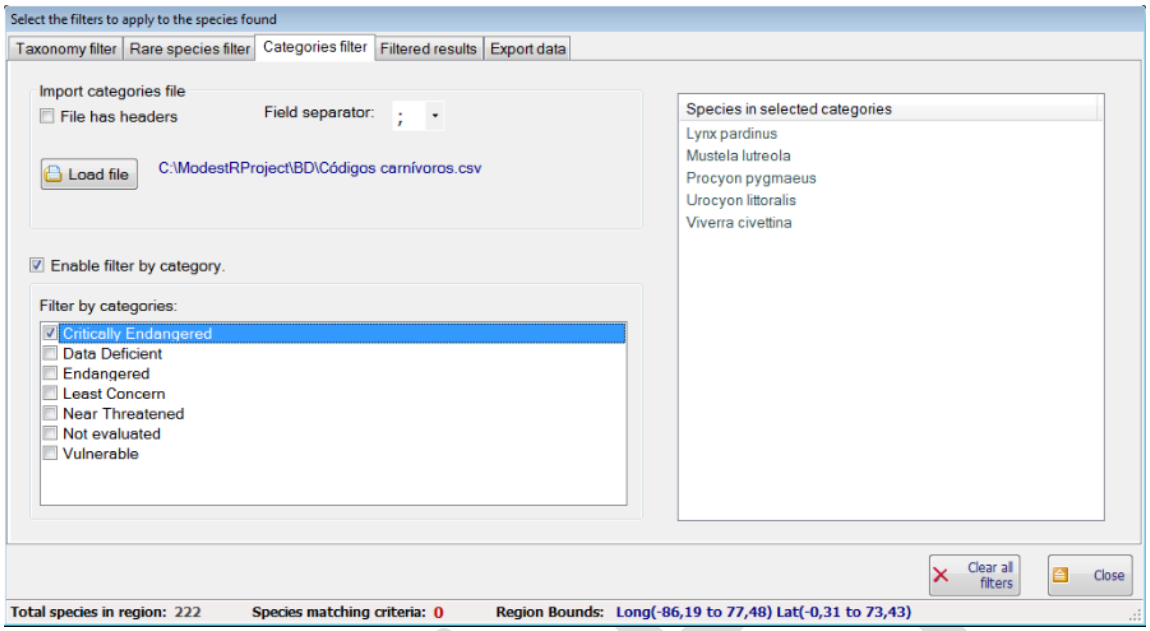
To load a categories file first select if the file has headers that have to be skipped, and the field separator used in the CSV file. Then click on the *Load file* button and select the file.

The file will be loaded and the categories will be shown on the *Filter by categories* listbox.

To apply a filter by categories, first check the *Enable filter by category* checkbox, then select which categories you want to include checking them on the *Filter by categories* listbox. The species that are included in the selected categories will appear in the right list *Species in selected categories*. Take into account that the species listed in this list are the ones that are in the imported file, independently of the species that are found in the current search.

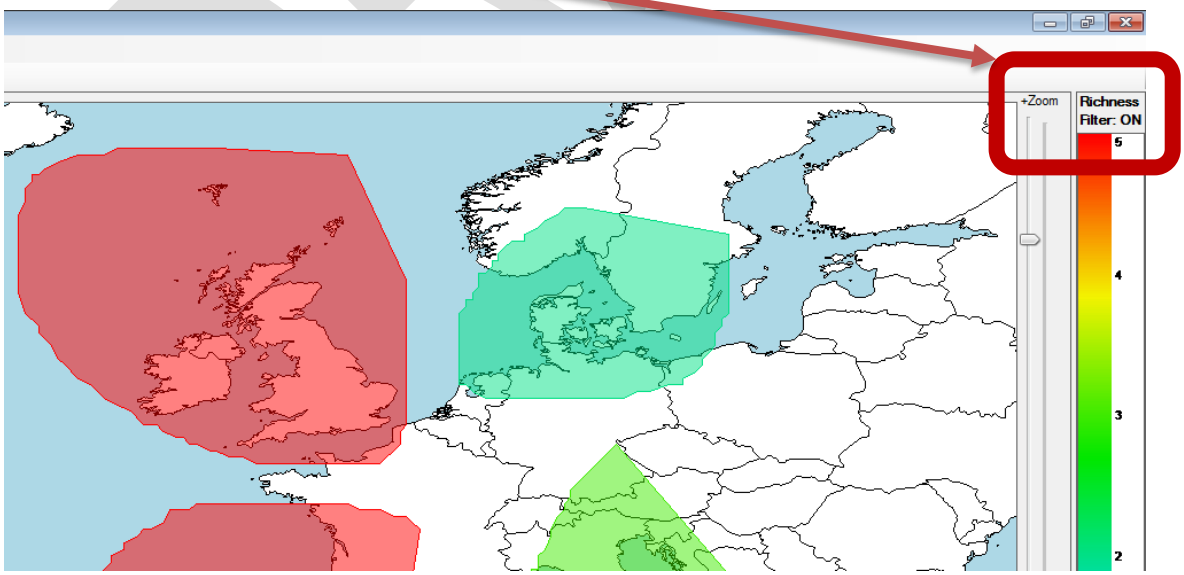
When enabling this filter and selecting categories, only species found in the current search that also are included in one of the selected categories will be included in any posterior exportation.

As said before, you can go to *Filtered results* page to see the species that currently comply with all the current filter settings (remember that all filters are applied to obtain those filtered results).

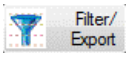


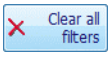

### 4.8.4 Color map and filter clearing.

Once you have set the filters you want, you can just click on the *Close* button and you will see that the colors of the areas in the map as well as the color scale will be updated to take into account only the filtered results. This is also indicated on the top of the right color scale, where a *Filter:ON* label will be shown.



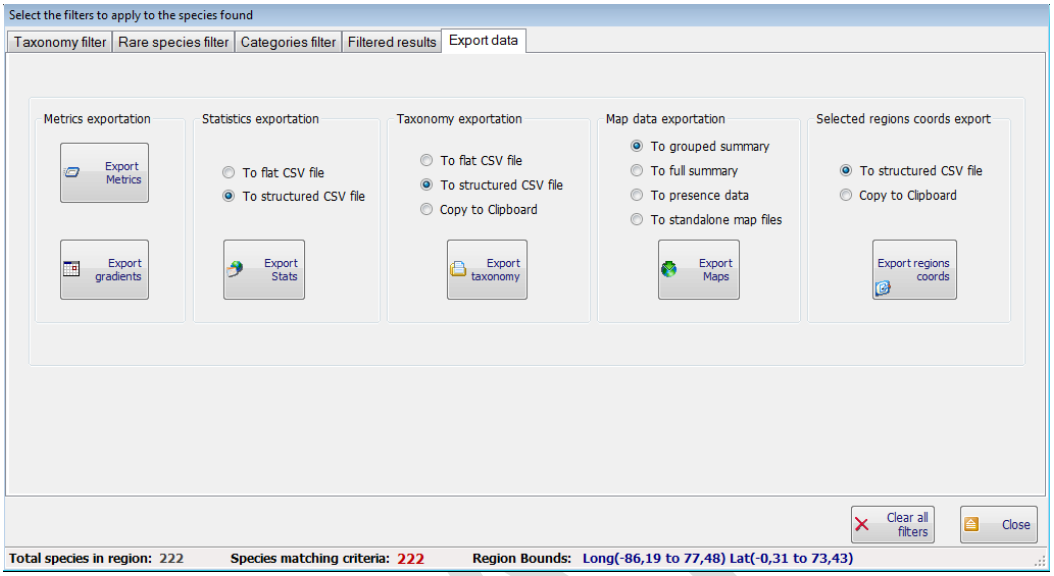

You can go back to the filtering window clicking on the *Filter/Export* button on the left tool panel, modify the filter and accept it to show the updated map as many times as you want. You can also clear all filters using the *Clear all filters* button on this window.



## 4.9 Generating and exporting data

Once filtered the species you want to include, you can go to the *Export data* page to generate and export several different data about selected species and in different formats. To select a data exportation process, you just have to click on the desired button. Some buttons are associated to different options regarding the exportation format, like for example the *Export stats* option, that will allow you to select a flat or a structured CSV format.




The available exportation options are:

- Export Metrics: generates and exports metrics about species such as richness, rarity, patch index, etc.
- Export gradients: generates and exports latitudinal gradients for several metrics.
- Export Stats: generates and exports statistical data about species like occupied area, maximum and minimum latitude and longitude, real area/reference area ratio, etc.
- Export Maps: exports selected species maps in several formats.
- Export taxonomy: exports taxonomy of the selected species in several formats.
- Export region cords: exports the coordinates of the region selected in the map.

In most of the cases the exported data will be, by default, clipped to the selected regions, either to the exact regions (such as for richness metric) or to the rectangle that circumscribes the selected areas (such as for environmental variables or areas by habitat). You can optionally disable this clipping to export data about the species found, but for the entire world, for example. If you need to clip any raster data (e.g. environmental data in raster format) to selected areas you can use the clipping feature described below in 4.12.3.

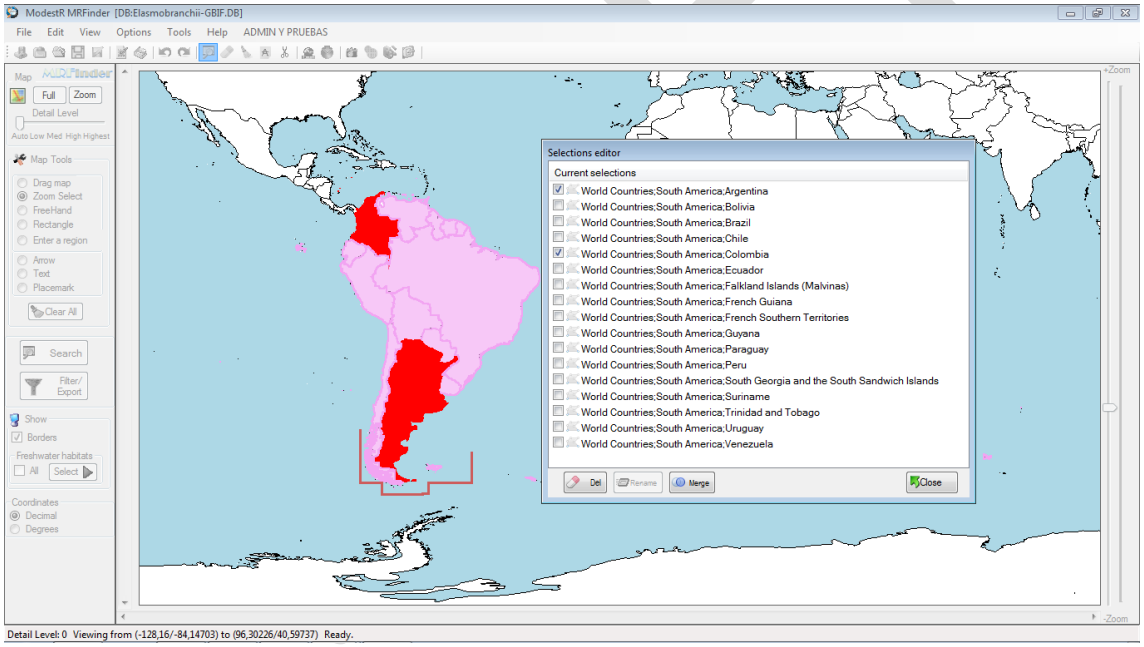
More details about those options and the exported formats can be found in section 8.

Besides exporting search results data, you can also export the current map view to an image file, using the main menu option *File/Export/Current map view as image*.

## 4.10 Selections editor

When you have added several selections to a map, you may want to remove one of them, change its name, or merge several selections in one single logical selection. Or even pinpoint one selection among many others in the map. To do that, you go to *Edit/Selections Editor.* The *Selections Editor* will be displayed. In this editor you can:

- Check any selection to highlight it in the map: use the checkbox on the left side of a selection to do that.
- Delete one or more selections: use the checkbox on the left side of the selections to select the selections to be deleted and click on the *Delete* button.
- Rename a selection: just select the wanted selection (you don't have to use the checkbox for that, just click over the selection) and click on the *Delete* or the *Rename* button.
- Merge several selections on one single logical selection: use the checkbox on the left side of the selections to select the selections to be merged and click on the *Merge* button. Remember that, as explained in 4.3, this feature allows to logically merge several selections to make MRFinder to consider them a single one regarding searching and display. But this merging does not imply any geometric union of several areas in a single one. To do this type of merging, see section 4.12.2.




## 4.11 Managing user predefined shapes

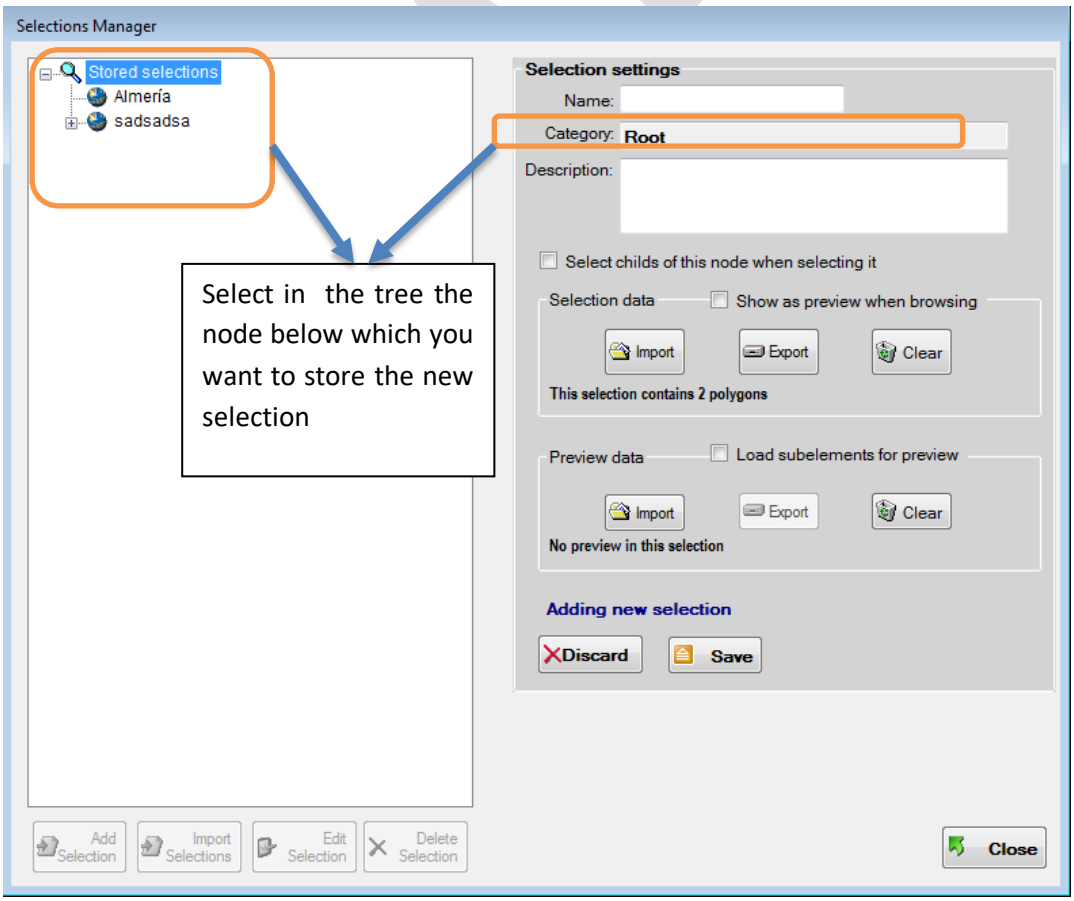
User predefined shapes are a useful tool of MRFinder that allows users to store the areas they select on the map to perform searches for later usages, as well as to import areas from existing shapefiles or KML, for example. As described before, ModestR also includes other predefined shapes such as world administrative areas, but the user cannot modify those shapes, because they are regularly updated when installing new ModestR versions. Here we are talking about the predefined shapes that you can create or import and store in the user predefined shapes database to reuse them later.

The first way to store a selection is to do it on the map and store it. You can select or import several areas. Then, when all the areas you want to store as a selection are selected in the map, click on the *Store shape as predefined* button in the toolbar or go to *File/Store as a predefined shape* menu option. A dialog box will be shown where you have to enter a name and a description for the shape, and select under which node of the shapes tree the shape will be stored.

You can also enter other settings such as:

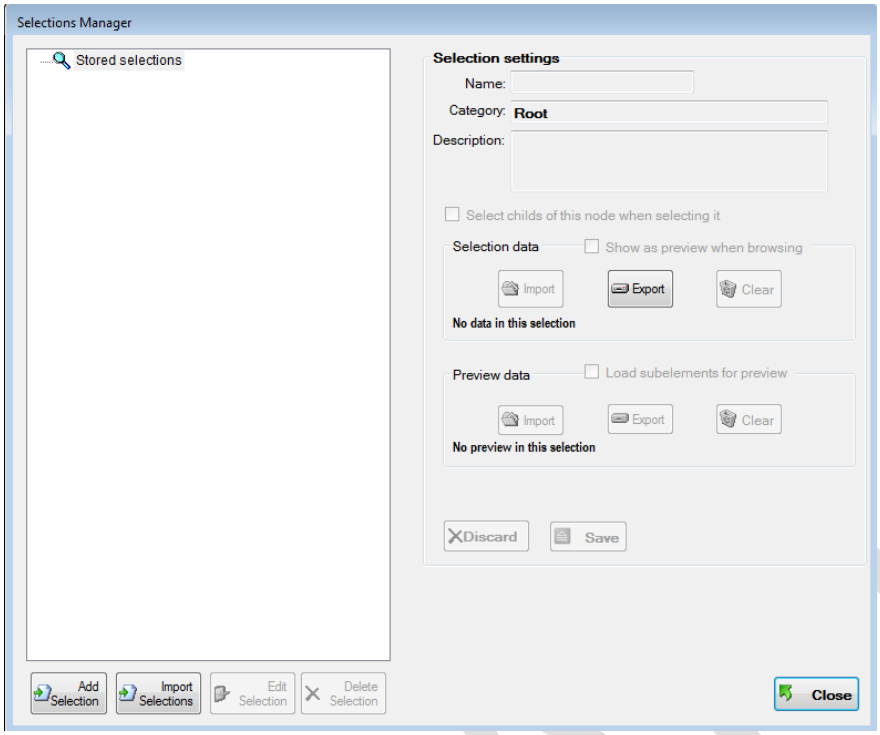
- Description: a description about the selection
- *Select childs of this node when selecting it*: enabling this option, and if you later add selection into this selection (child nodes to this node), those selections will be automatically selected when you select this node. You can see an example of that behavior when adding predefined shapes included with ModestR. When for example you select the node "Africa", all countries (*subnodes)* of Africa will be selected.
- *Import/Export/Clear* buttons let you import/export data for a selection from/to an external shapefile or clear this selection data (selection is not deleted, but it won't contain geographical data).

Once settings entered, just click on *Save* button to save the selection.

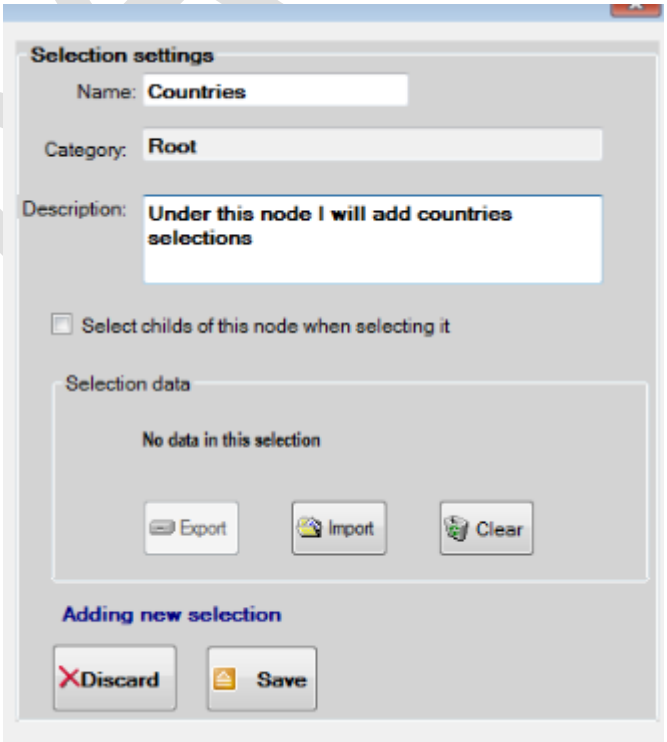



Now we will briefly explain how to manage predefined shapes using the menu *Options/Manage predefined shapes*. This will show the shapes manager window, that will be empty unless you

already stored some shapes (predefined shapes provided with ModestR are not shown here, because they cannot be modified neither managed by the user).




Selections have to be organized in a tree. So the first step can be adding a selection node. To do that, just click on the *Add Selection* button. The right panel will change its color to signal that you are in *edit* mode. Let's suppose you want to add a node named *Countries* under the *root* node of the tree. This node will not really contain data, but it's just a way to organize the selections. So just enter the name, an optional description, and click on *Save.* The node will appear on the left tree.

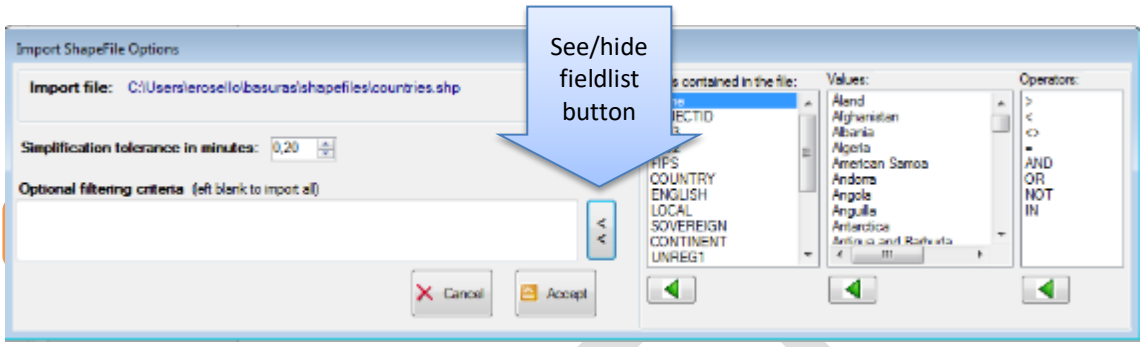



Let's assume now that you have a shapefile that contains all the countries for the world[20], and you want to import them under the *Countries* node that you have created. To do that:

- Select the *Countries* node in the tree.

[20] A file like this can be found in the DIVA-GIS web site http://diva-gis.org

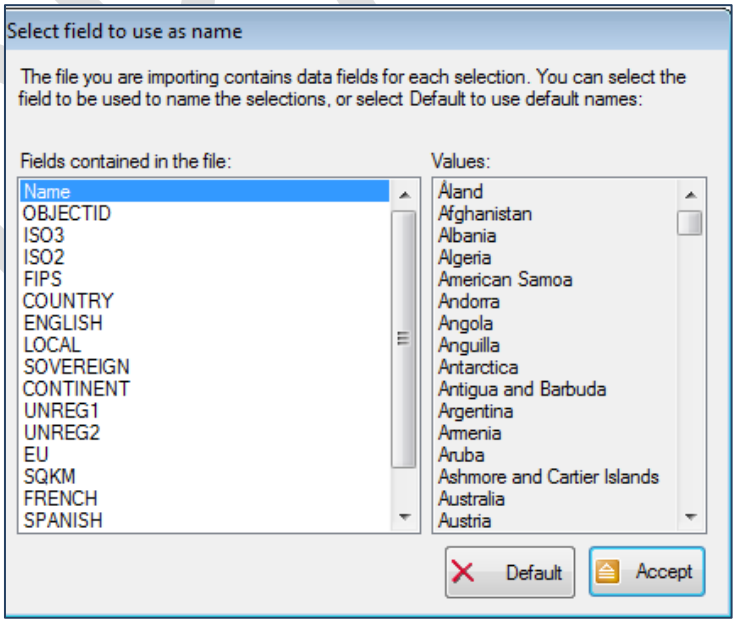
- Click on the *Import selections* button. On the dialog box, select the shapefile type, in this case, and then select the wanted file to be imported.
- ModestR will import the data from the file, and show you the importation settings dialog box (here expanded). If you are familiar with shapefile format, you can also add importation criteria to filter which data will be imported. To do that, use the *See/hide fieldlist* button to see the list of fields contained in the shapefile. Then you can select a field to see the values it takes, and add criteria that will be used as filter when importing.




You can also directly write a condition in the textbox. The syntax of the conditions is similar to the syntax of a WHERE clause in SQL.

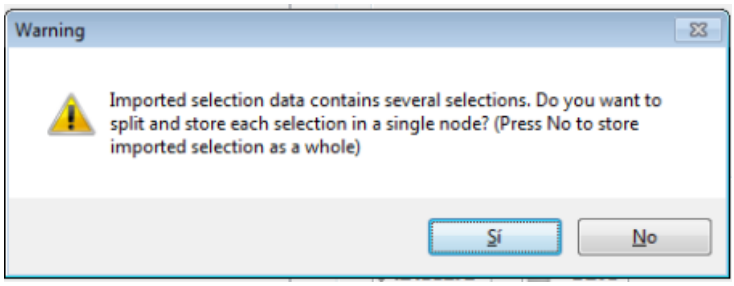
If you want to import all data, just *Accept* this dialog box.

- If the shapefile contains data fields, which is the most usual, ModestR will show you the list of fields to allow you selecting the field that will be used to assign names to the imported selections. Otherwise you can select *Default*. In this dialog box you can explore fields values clicking on the field on the left list to see its values on the right one; this makes easier to find the correct field to be used to assign names to the imported selections.

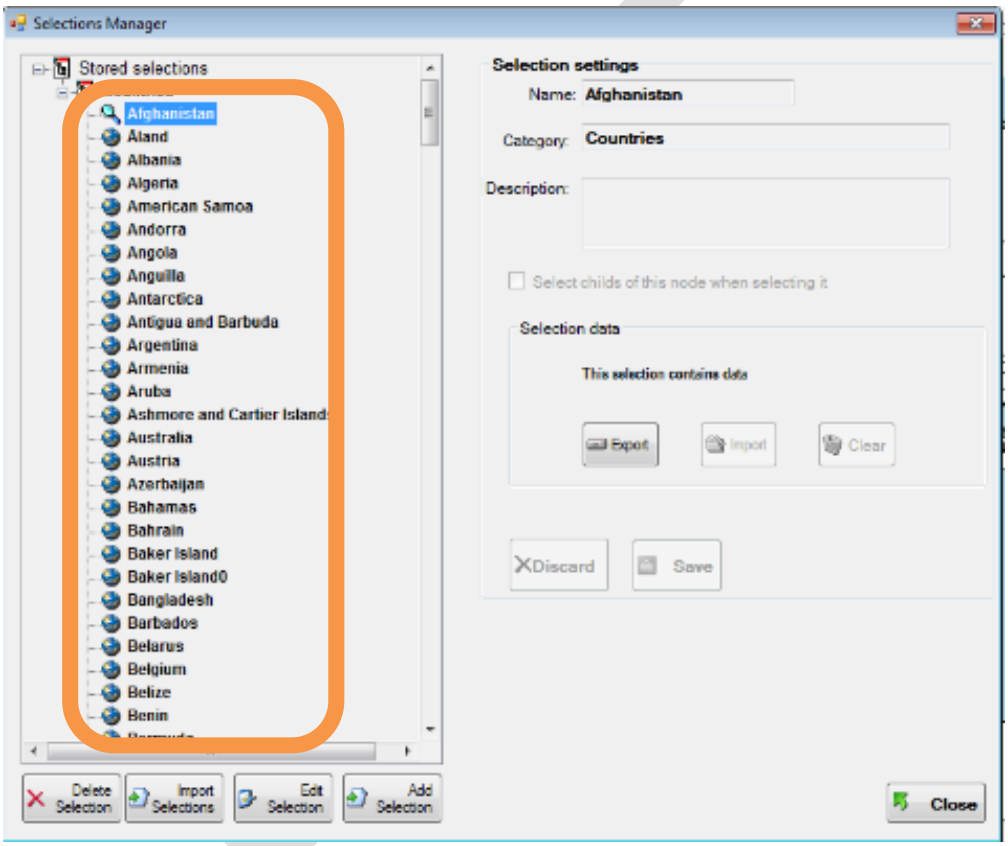



- In this example we'll suppose we choose the Name of the country field and we accepted. ModestR will import all data from the shapefile. As in this case this shapefile contains several shapes (one by country) ModestR will ask you if you want to import them as a whole, so they will be a single stored selection, or you want to split them to make each

shape to be a single selection. Of course this last one is probable the most used choice, and the one we do in this example.



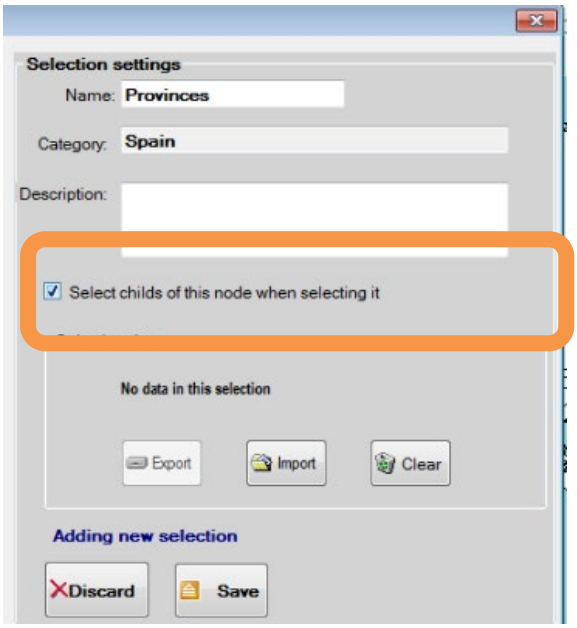

- As a result ModestR will import each country shape as a different node, under the *Countries* node:




This simple example showed how to import selections to ModestR. They could be used either as shapes in MapMaker (see section 2.16) or as predefined shapes in MRFinder (see section 4.5).

You can now for example import administrative boundaries for a country, adding them under the node of this country, following more or less the same steps of this example.

There are more useful features when importing predefined shapes. For example, you can make that when a tree node is selected, it automatically selects all the subnodes it contains. This can be useful for example if you add a node "*Provinces*" for a country, and then you add subnodes with all the provinces. If you want that when selecting the node *Provinces* ModestR add to the map all the provinces, edit the *Provinces* node and check the option *Select childs of this node when selecting it.*



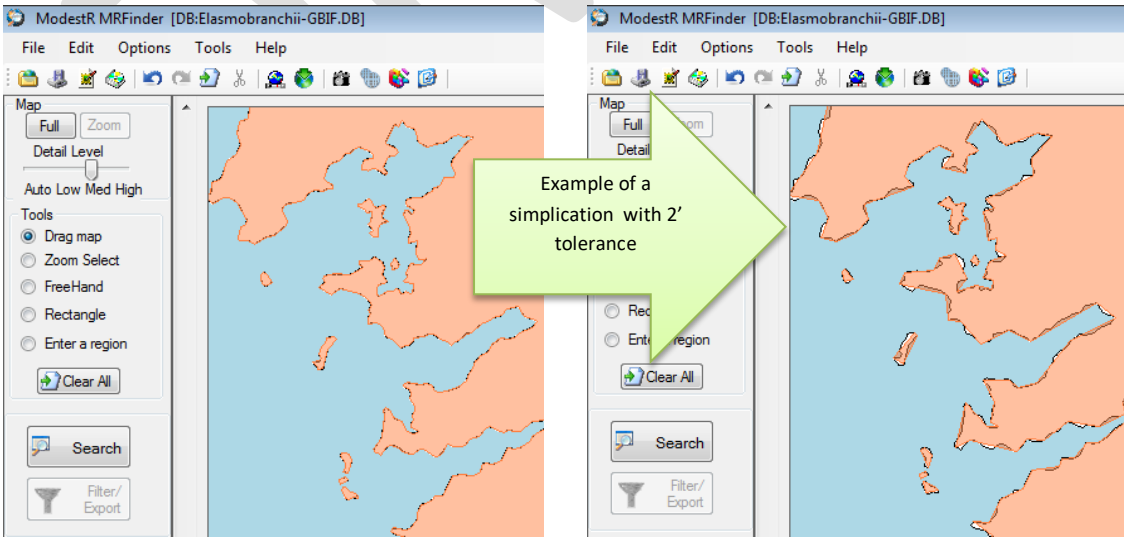

It can be also pointed out that you can delete nodes using the *Delete selection* button, but be careful because all subnodes of a node will be also deleted.

To reorganize nodes, such as moving a node to another parent, you can edit the node and select its new parent. You will see that the *Category* field of the node changes to the new parent. Then save it. You can also just drag and drop nodes to move them from one parent to another.

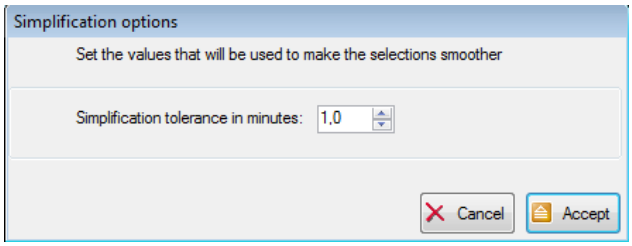
## 4.12 Other tools

### 4.12.1 Simplifying selections

MRFinder searching features are strongly affected by selection complexity, that is, the number of points (vertexes) that define the selections where to search for. Manually made selections are rarely concerned by this problem. But it can be the case for selections imported from external files such as shapefiles of KML files, whether they are directly imported to the map (see 4.4) or as predefined shapes (see 4.5). In those cases, you have the possibility of simplifying those selections. Simplification eliminates some vertexes with a given tolerance, which can be easily acceptable in most cases, but can dramatically improve MRFinder searching performance.




Both options are available in the *Edit/Simplify current selections* menu options or in the corresponding button of the toolbar. A dialog box like the shown below will be displayed, where you can set simplification tolerance:




### 4.12.2 Merging selections

If you have several selections that overlap or that are just contiguous (for example a selection for Spain and another for France), but you are interested in having a global search, and not a separated search for each selection, you can significantly improve MRFinder searching performance by merging all those selections, the way that two or mode overlapping or contiguous selections will be merged in one single selection. This option is available in the *Edit/Merge current selections* menu options or in the corresponding button of the toolbar.

It must be pointed out that this merging option has no the same effect that merging selections explained in 4.10. While this one implies a geometric operation to calculate the polygons resulting from the union of the involved selections, the one explained in 4.10 does not perform any geometric operation, but just a logical union of several selections to consider them a single one regarding searching feature and display.

### 4.12.3 Raster data clipping

You can use MRfinder to cut off raster data to any pattern. To do that, the raster file has to be previously integrated in ModestR. Details about how to integrate environmental data in ModestR are explained in section 6.26.

Then add to the map the wanted selections, either predefined ones or owner drawn, as explained in previous sections. Go to *File/Export/Environmental data for shape* option in main menu. A dialog box will show to you the available variables (those previously integrated in ModestR). Select the wanted one (it could be more than one, to generate several clipped rasters). Next selected the target folder where the file(s) will be saved, in ESRI ASC format. Be careful because any existing file with the same name will be overwritten. MRFinder will clip the rasters for each selection. That is, if there are three selections currently on the map, the selected raster will be clipped for each one of those selections. So you'll have three files resulting for each one of those clipping operations.

A report in CSV format is also generated with data about clipped areas in $km^2$ and simple statistical measures (max, min, mean, std and quartiles) for the rasters in the clipped areas.

This option is similar to the one described for MapMaker in section 2.9.4.

### 4.12.4 Calculating areas for selections

You can use MRfinder to calculate areas for any selection. This method is intended to provide an approximate result, as it is performed using a rasterization method with a resolution of 1'. Go to *File/Export/Selection area summary report* option in main menu. The generated report in CSV format contains the total area and the area of each type of habitat for each selection currently added to the map, in $km^2$. See more details in 8.14.

### 4.12.5 Export selections

MRFinder importation and exportation formats are detailed in section 8. Exportation of species data resulting from searching has been mentioned in 4.9, and exportation of clipped rasters and area data have been mentioned in 4.12.3 and 4.12.4. But MRFinder also has exportation features to export selections:

- Export map as image: in *File/Export/Current map view as image* you can export the current map view you have in the screen to an image in JPEG or BMP format. Options such as adding a grid, scale, or quality adjust will be shown.
- Export current selections to KML or Shapefile: *File/Export/Current selections/To KML* or *To Shapefile* you can export current selections to KML or Shapefile formats.
- *File/Export/ Current selections /To ASC raster*: generates a raster in ASC format, with the selected precision, where the cells where there is a selection have a value not null (usually 1).

# 5 First steps in MRMapping

MRMapping is an easy-to-use software application to build maps with data coming from several distributions. Those data will typically come from databases created with DataManager. MRMapping allows adding any number of species distribution data to a map, optionally grouping them by any rank (lass, order, family…). It also can calculate overlappings between distributions and occupied areas.

The MRMapping user interface is quite simple and very similar to MapMaker, as the most usual tools are always visible at the left tool panel. You also have contextual menus that will appear when you click with the right button of the mouse on the map. Moreover, quite all menus, buttons and other elements have a tooltip that briefly explains their usage, and that will be shown when you place the mouse over an element.

We recommend you to see the section 2, and particularly the 2.1 and the 2.18 for an introduction about moving on the world map and using some map tools.

## 5.1 Creating a map in MRMapping

MRMapping will usually work using the data of a ModestR database. A database can be selected using the *File/Select database* menu option (MRMapping will automatically use the last used database when running it again). Once a database selected, you can add species distribution data from this database to the map going to *File/Add/Maps from ModestR database*.

MRMapping will display the taxonomy tree of the database. You can directly expand nodes with the mouse. Only select a species with a map (that is, will distribution data) will be shown in the tree. To help you finding a map, you can also use the search box, than can be displayed using the magnifying glass button on the bottom-left corner. In the search box you can write some part of the species you are looking for, and coinciding names will be shown in the *Results* list. This feature is the same than the explained in 2.21 for MapMaker.

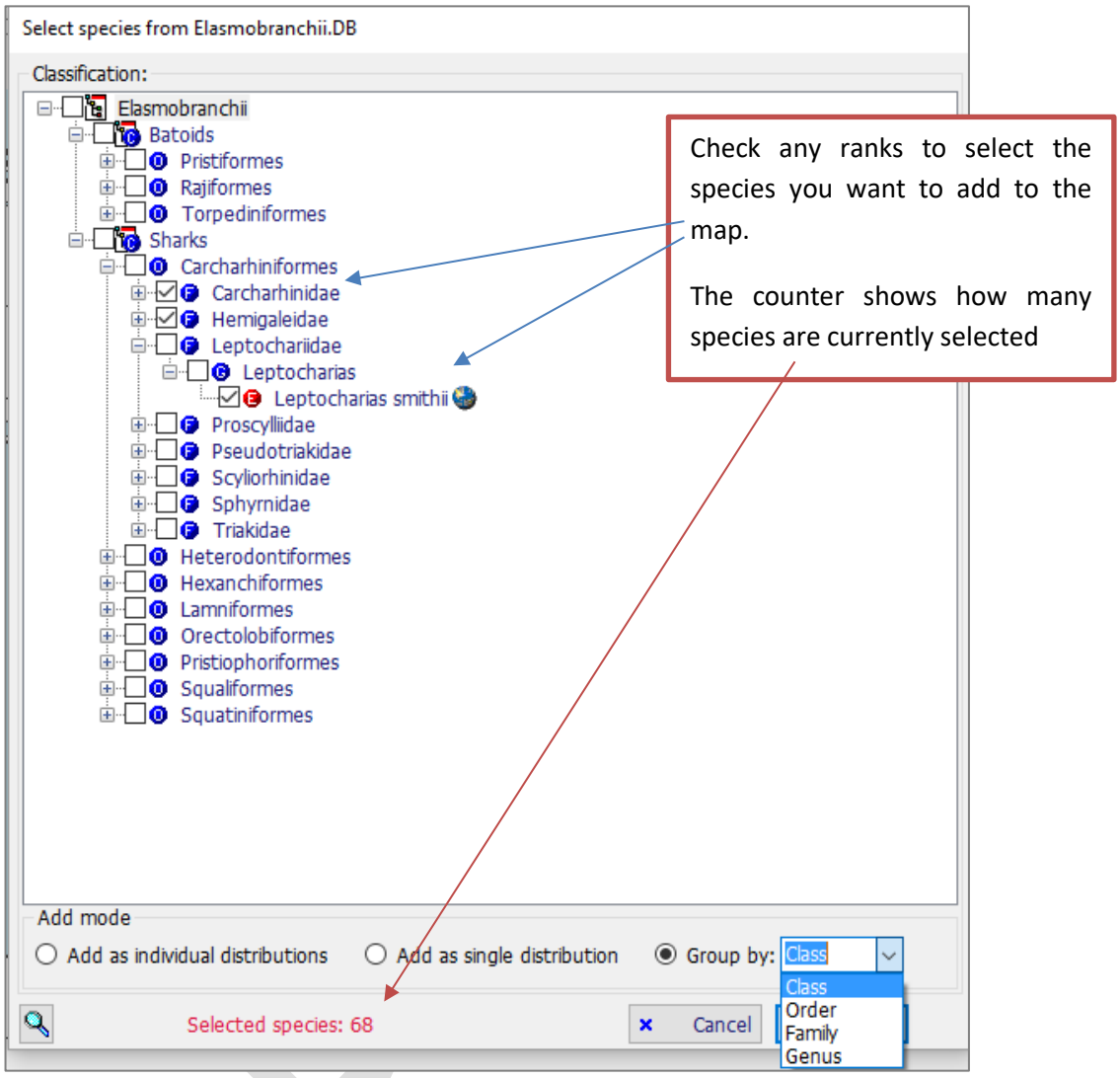
To select the species you want to add to the world map, just check the corresponding checkbox. As you will see, you can directly add all the species of a rank (e.g. a whole family) by checking it on the tree, so you don't have to check species by species. There is also the possibility of drag&drop any taxa from the tree of DataManager to MRMapping.

Then, on the bottom of the dialog box, you can select the mode the distribution data from the selected species will be added to the world map:

- Add as individual distributions: each one of the distribution maps of the selected species will be added to the world map as single separated distributions. This allows managing each one of those distributions individually, for example to change their color, delete it, etc.

- Add as single distribution: all the distribution maps of the selected species will be merged and added to the world map as a single distribution. That is, it will be identified as a single distribution with a single name and color settings, etc. You would not be able to discriminate between the species included in this distribution later on.

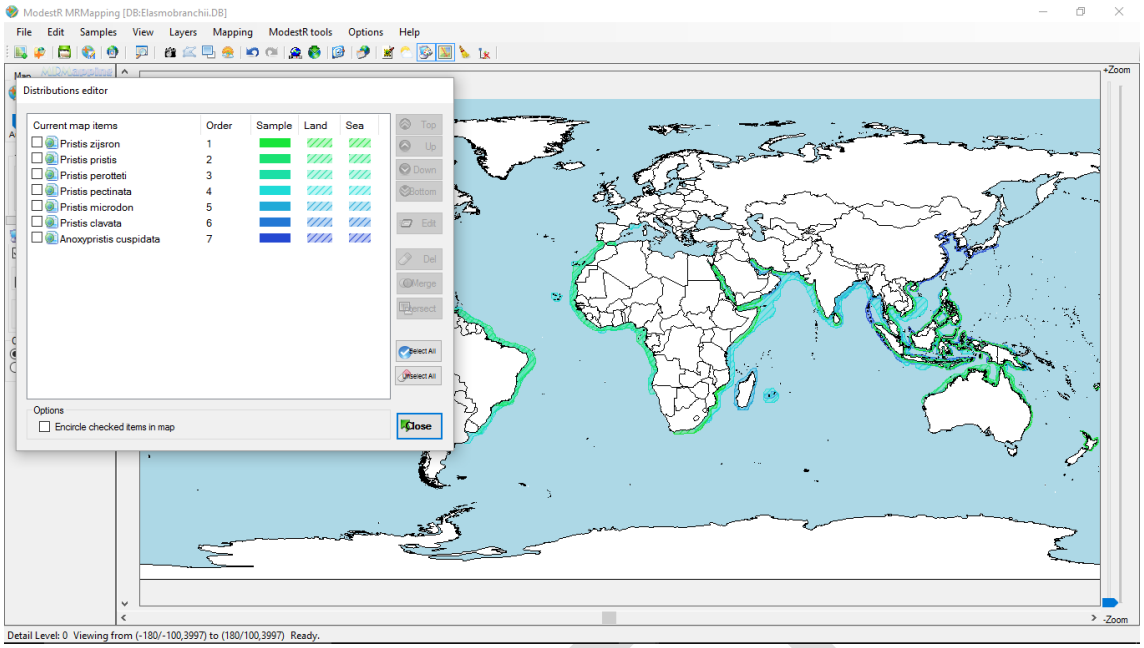
- Group by class/order/family/genus: all the distribution maps of the selected species will be grouped by the rank you select, and added to the world map as a distributions, one distribution for each different rank. You would not be able to discriminate between the species included in those distributions later on.



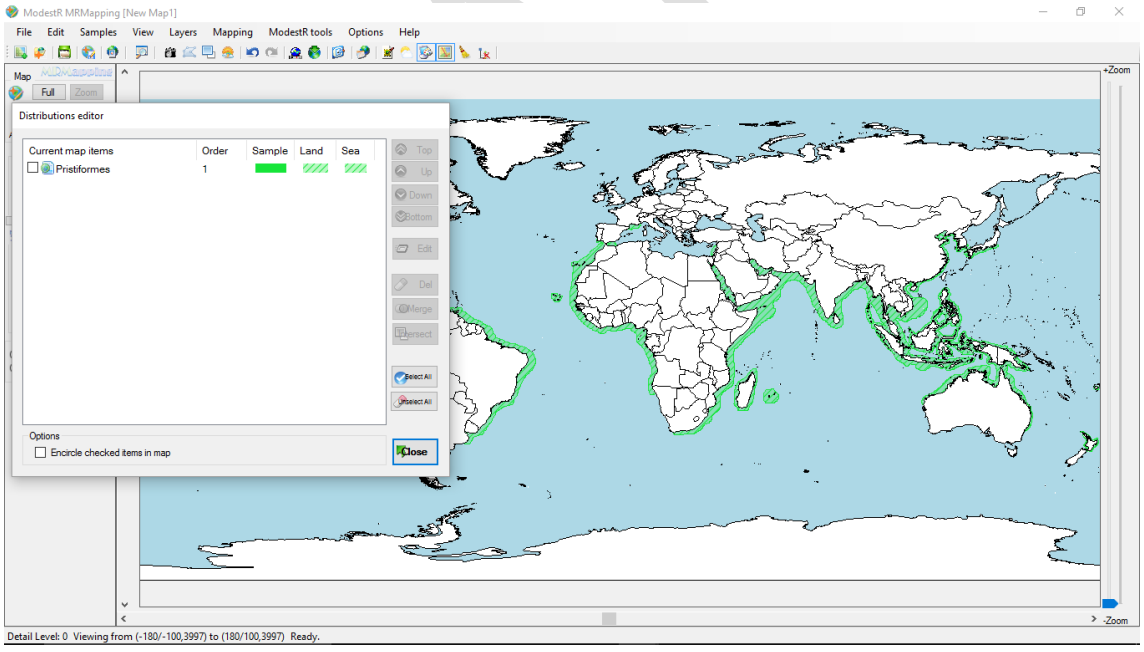

To illustrate those different options of adding data to the map next an example is described. The following maps contain the same data, corresponding to the distribution maps of seven species of the *Pristiformes* order[21].

In the first one, the data where added using the option “Add as individual distributions”. So each species distribution is added as a separated distribution, with its own name and settings:

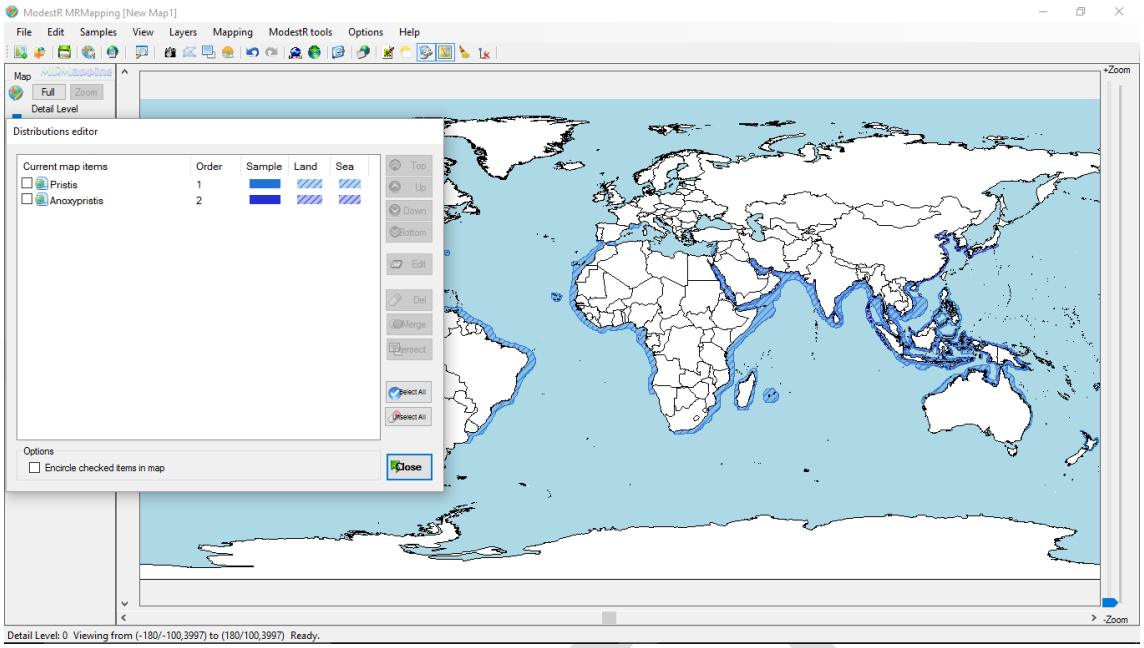
[21] You can find the *Elasmobranchii* database for ModestR from which those data came from in the ModestR website. It is available as a occurrence database that contains expert range maps for all the species of *Elasmobranchii* class.




In the second one, the same data where added using the option “Add as single”. So the data for all selected species distribution have been merged and added as a single distribution, with a single name and settings:



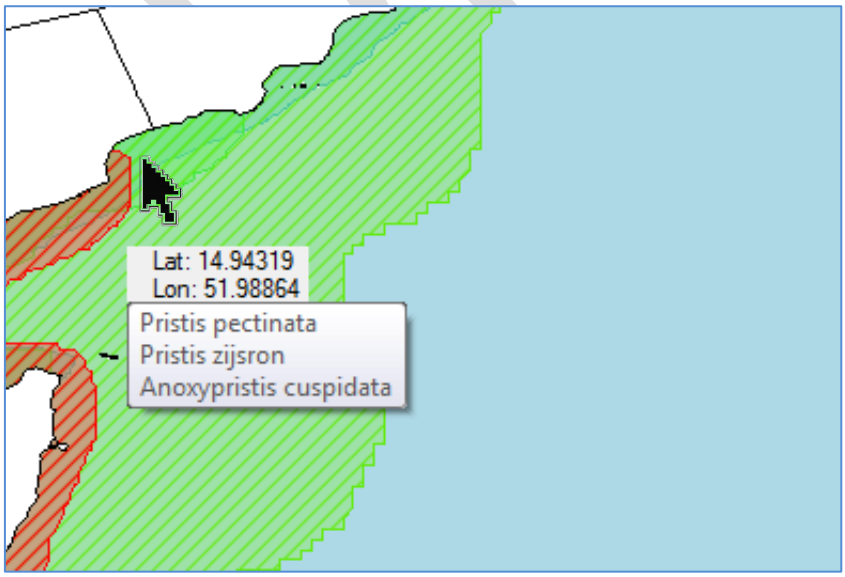

In the last one, the same data where added using the option “Group by *genus*”. So the data for all selected species distribution have been grouped by genera. Then the data of each group have been merged and added as a single distribution, with a single name and settings. As the selected species belonged to two different genera, two distributions have been added:




Independently from the option you used to add a set of distributions to the world map the first time, you can add more distributions later using a different grouping option. Moreover, you can even add species from different databases to the same world map, just by changing the selected database in the *File/Select database* menu, and going to *File/Add/Maps from ModestR database* anew.

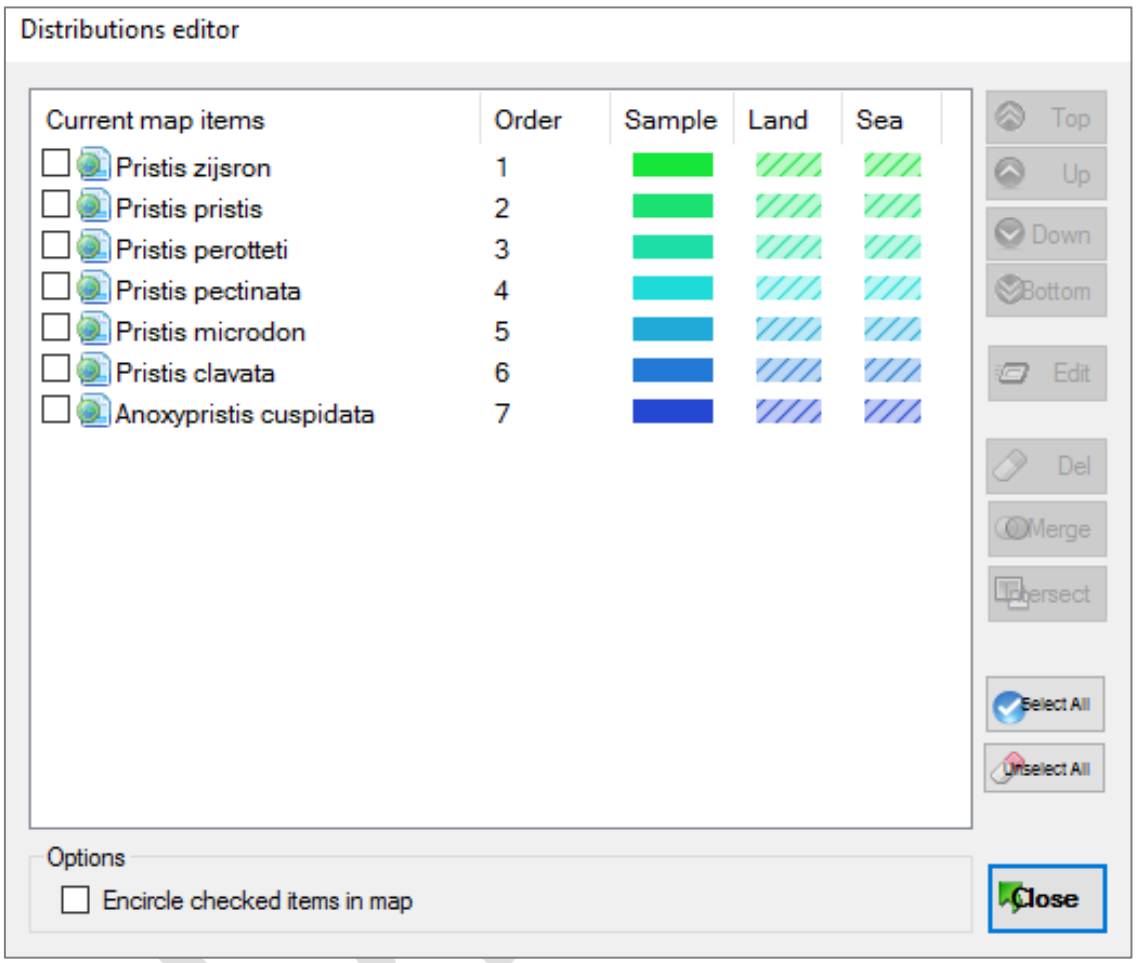
### 5.1.1 Querying for species/distributions in a point

When having several species distribution maps that overlap on the world map, it can be sometimes difficult to know which species (or other rank, if you added grouped distributions) are really present in a specific point. To query that, you have just to double click with the mouse in the map. If one or more species are currently present at this point of the map, a little tool tip will appear with this information. Of course, if you added several species but grouping them in a single distribution, you cannot obtain information about the presence of each single species.



## 5.2 Changing map settings

Once one or more distribution maps (we can also name them distributions, map items or datasets) to the world map, you can edit them and modify their settings going to the menu *Edit/Distribution list editor*. A dialog box is displayed with the list of all distribution maps (*map items)* currently added to the world map.




In this list the distributions are shown by default regarding the order they are painted in the map. That is, the last distribution of the list is the one that is painted the first, while the first one of the list is the last painted, so it is painted above all the others. That can be important because one distribution can eventually hide another that this painted below it. You can change the order the distributions are painted by selecting any distribution (not checking it, just selecting it) in the list and moving it using the buttons *Top, Up, Down* and *Bottom*.

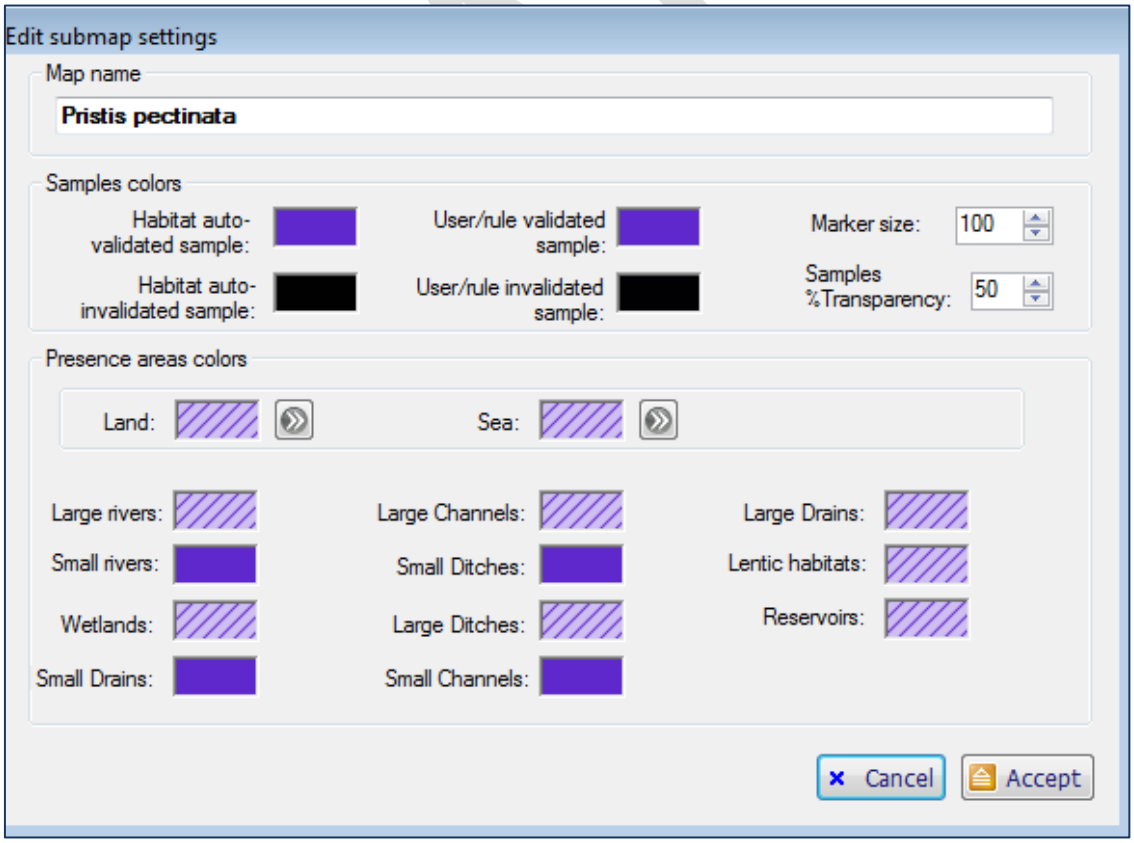
You can delete any distribution by checking it (not selecting it, but checking it using the checkbox on the left of the distribution name) and clicking on the *Del* button.

You can also merge two or more distributions, to join them in a single distribution, with a single name and single colour settings. To merge distributions, check them (not selecting it, but checking it using the checkbox on the right of the distribution name) and click on the *Merge* button. If distributions to be merged contains polygons (e.g. range maps) this operation implies geometrical calculations that can take some time, depending on the complexity of the polygons

to merge. Once completed, you'll see that the merged distributions are replaced by a new single distribution in the list, which name is by default the concatenation of the names of the distributions merged.

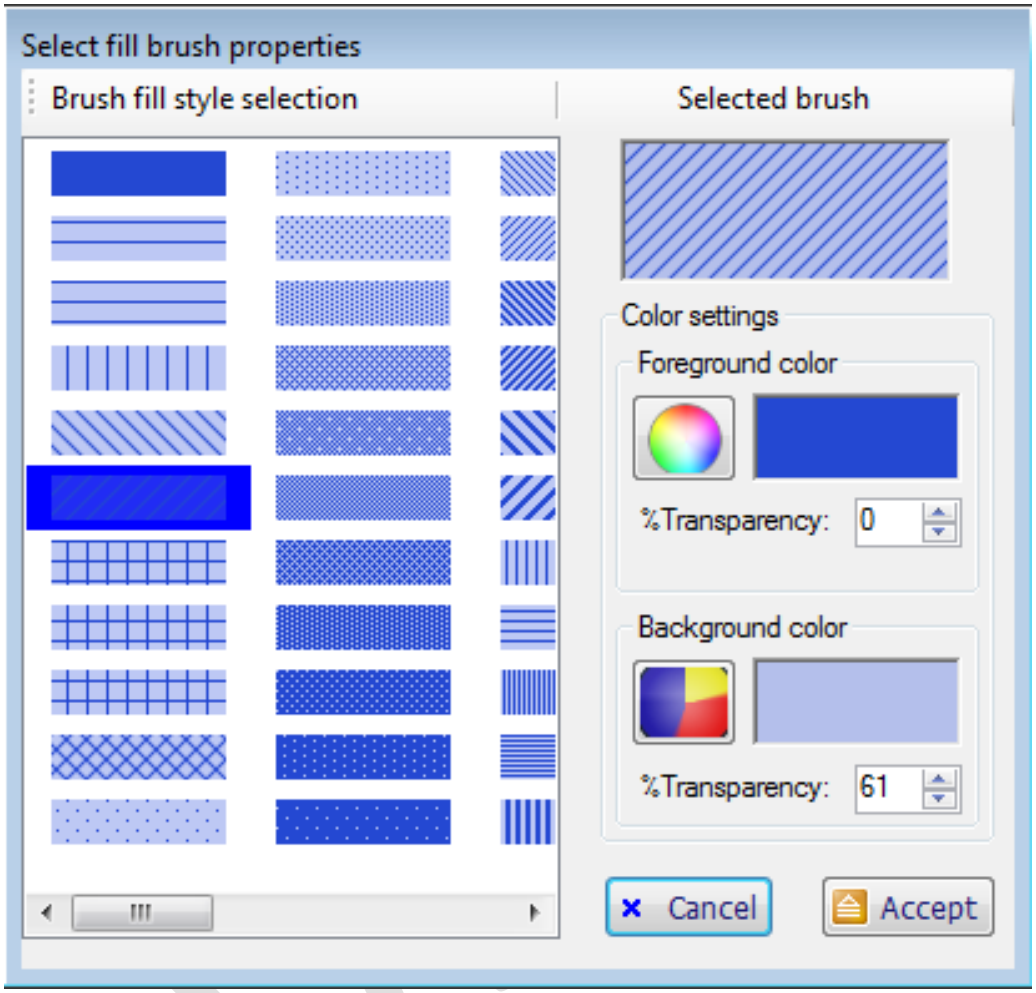
In a similar way, you can calculate the intersection of two or more distributions. The intersecting areas will be added as a new distribution with its own name and color settings. To intersect distributions, check them (not selecting it, but checking it using the checkbox on the right of the map item name) and click on the *Intersect* button. Obviously, this operation it's only useful when applied to distributions that contain polygons (e.g. range maps). This operation implies geometrical calculations that can take some time, depending on the complexity of the polygons to merge. Once completed, you'll see that the resulting intersection areas will be added as a new single distribution in the list, which name is by default the concatenation of the names of the intersecting distributions.

Finally, you can edit the name and colour settings of any distribution by selecting it (not checking it, just selecting it) in the list and clicking the button *Edit*. A dialog box will be displayed where you can change the name and the colour settings of the distribution. You can change individually the colour assigned to the presence areas for each type of habitat, as well as for occurrences. This allows for example to assign a different colour to presence areas located in the sea and to those located in the land.




To change any colour just double-click on the corresponding colour rectangle. A colour editor dialog will appear where you can change the colour.

For presence areas (range maps) the colour editor will include the possibility of changing the fill pattern, the foreground colour and the background colour, as well as the transparency for each one. It may be convenient to use a more or less high transparency for a better visualization of overlappings between several distributions.




It is worth to mention that any modifications done in the *Distributions editor* can easily be undone. Just close the *Distributions editor* and use the *Undo* button or the *Edit/Undo* menu option in the main window.

## 5.3 Adding new elements to the map

Even if the main usage of MRMapping is adding already existing distribution maps, you can eventually add presence areas or occurrences directly to the world map, the same way you can do it in MapMaker (see 2.2 and 2.3). Those elements will appear in the *Distributions editor* (seen above) under the item name *User-add operations*.

You can also add arrows, text or placemarks to the map the same way you can do it in MapMaker (see 2.18.1).

To add predefined shapes to the map, you can go to *Layers/Shapes/Add shapes* or the corresponding button of the toolbar, the same as in MapMaker (see 2.16).

Finally you can show in the map rasters (e.g. environmental data) and shapes the same way you can do it in MapMaker (see 2.9 and 2.162.18.1).

## 5.4 Importing a file of shape names

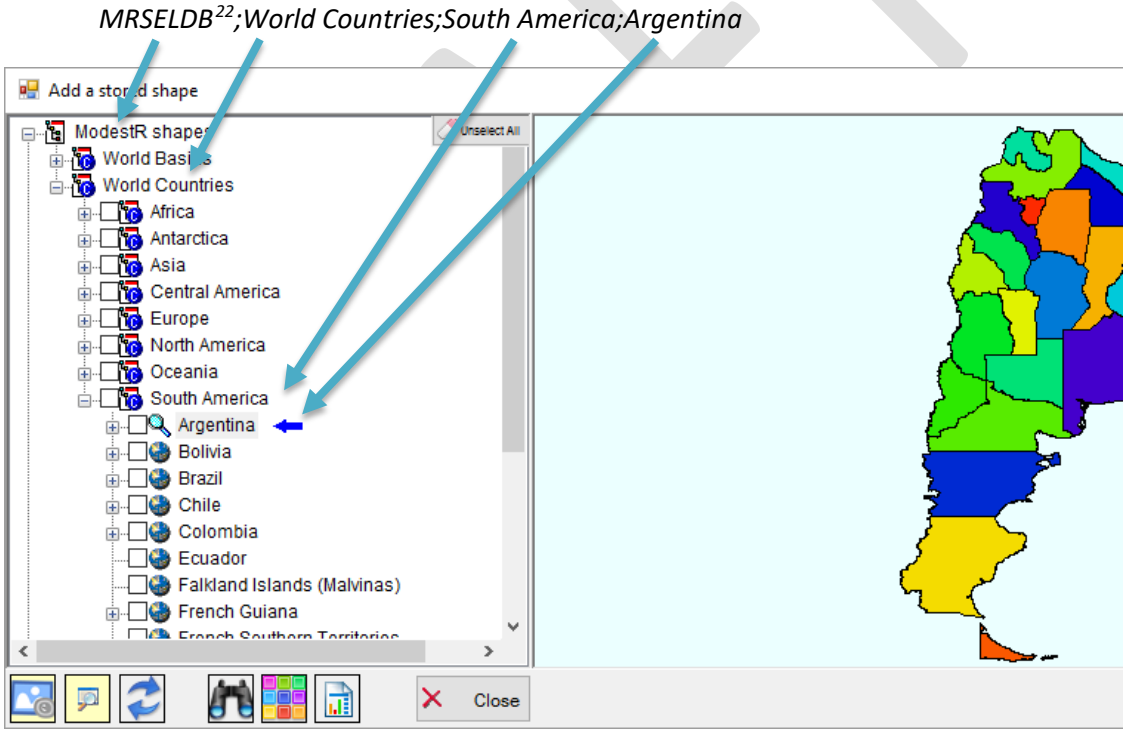
In MRMapping it is possible to import a CSV file with shape names and associated values, in order to display them using a color scale on the world map. We can, for example, import a file with country names and values of population density. The accepted names are those of the shapes that ModestR has stored in its predefined shape database (see 2.16).

An example of a file should be:

*Argentina;100*
*Panama;546*
*France; 578*

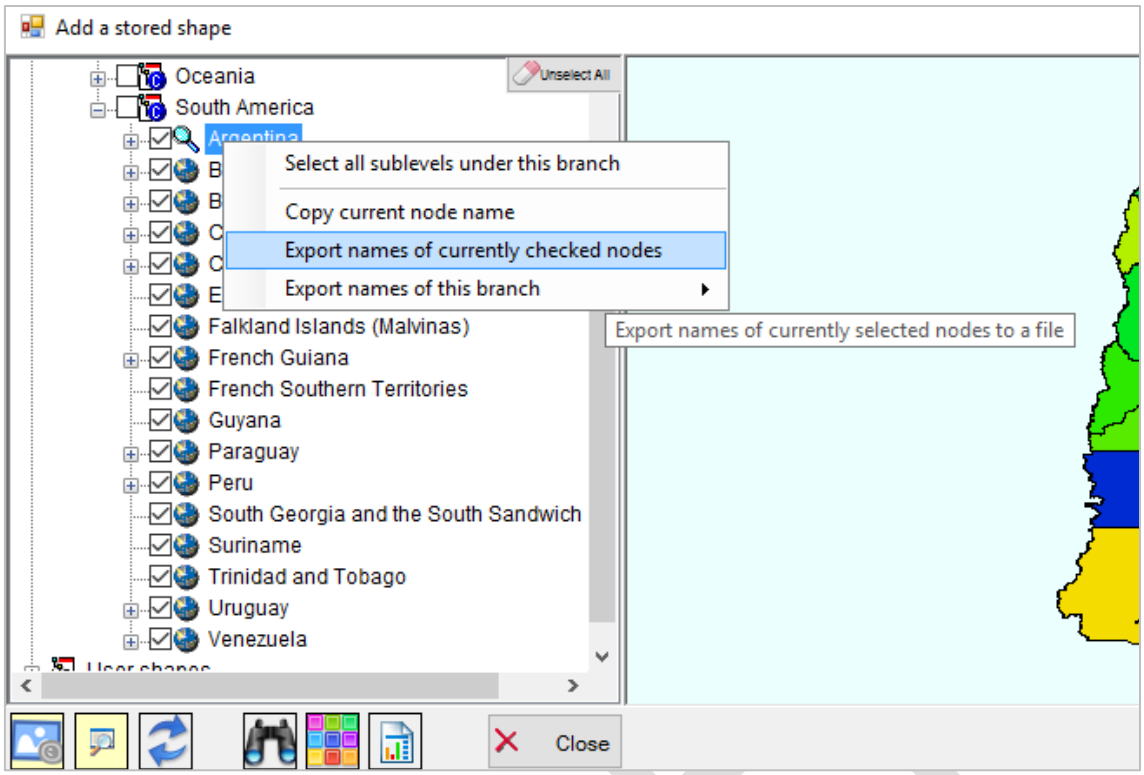
But ideally, you should use *full* shape names, that is, the names that include the full path from the root tree of the predefined shape database. This avoids eventual confusions between shapes with the same single name. For example, the full name of *Argentina* will be:

*MRSELDB*[22]*;World Countries;South America;Argentina*




Of course, it is not very handy to use this syntax. But you can directly generate a list of shape names in this format from the shape selector. To do that go to *Layers/Shapes/Add shapes.* Then, in the left tree select the nodes you want to export names. For example, we can select all South America countries. Then click with the right-button on any node and select *Export names of the currently checked nodes*.

[22] *MRSELDB* is the name associated to the integrated predefined shape database.



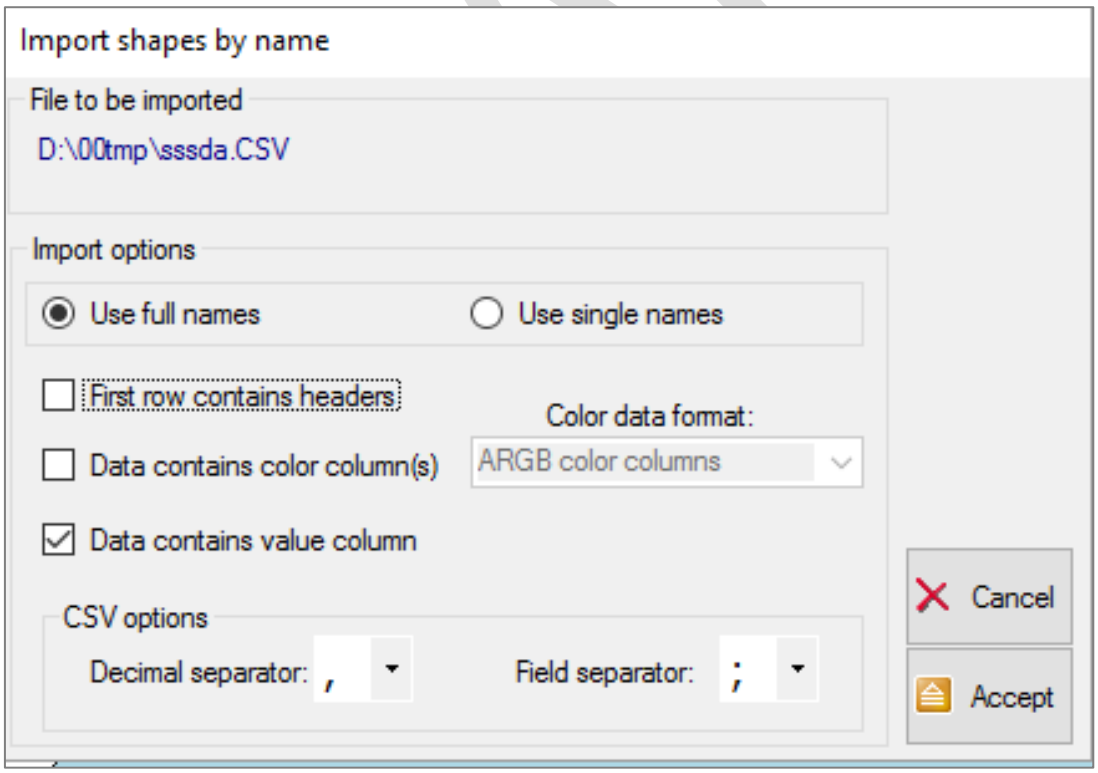

The resulting CSV file will contain that:

| |
|---|
| MRSELDB;World Countries;South America;Argentina |
| MRSELDB;World Countries;South America;Bolivia |
| MRSELDB;World Countries;South America;Brazil |
| MRSELDB;World Countries;South America;Chile |
| MRSELDB;World Countries;South America;Colombia |
| MRSELDB;World Countries;South America;Ecuador |
| MRSELDB;World Countries;South America;Falkland Islands (Malvinas) |
| MRSELDB;World Countries;South America;French Guiana |
| MRSELDB;World Countries;South America;French Southern Territories |
| MRSELDB;World Countries;South America;Guyana |
| MRSELDB;World Countries;South America;Paraguay |
| MRSELDB;World Countries;South America;Peru |
| MRSELDB;World Countries;South America;South Georgia and the South Sandwich Islands |
| MRSELDB;World Countries;South America;Suriname |
| MRSELDB;World Countries;South America;Trinidad and Tobago |
| MRSELDB;World Countries;South America;Uruguay |
| MRSELDB;World Countries;South America;Venezuela |

Using a worksheet such as Excel to edit this file it is very easy to add a new column with the values we want to assign to each country, and save it in CSV format. For example:

| | |
|---|---|
| MRSELDB;World Countries;South America;Argentina | 334 |
| MRSELDB;World Countries;South America;Bolivia | 910 |
| MRSELDB;World Countries;South America;Brazil | 649 |
| MRSELDB;World Countries;South America;Chile | 736 |
| MRSELDB;World Countries;South America;Colombia | 925 |
| MRSELDB;World Countries;South America;Ecuador | 712 |
| MRSELDB;World Countries;South America;Falkland Islands (Malvinas) | 33 |
| MRSELDB;World Countries;South America;French Guiana | 370 |
| MRSELDB;World Countries;South America;French Southern Territories | 186 |
| MRSELDB;World Countries;South America;Guyana | 765 |
| MRSELDB;World Countries;South America;Paraguay | 327 |
| MRSELDB;World Countries;South America;Peru | 945 |
| MRSELDB;World Countries;South America;South Georgia and the South Sandwich Islands | 308 |
| MRSELDB;World Countries;South America;Suriname | 835 |
| MRSELDB;World Countries;South America;Trinidad and Tobago | 743 |
| MRSELDB;World Countries;South America;Uruguay | 544 |
| MRSELDB;World Countries;South America;Venezuela | 769 |

Now we can reimport this file going to *Layers/Shapes/Import using file with names list* and selecting the CSV file. A dialog box will appear:

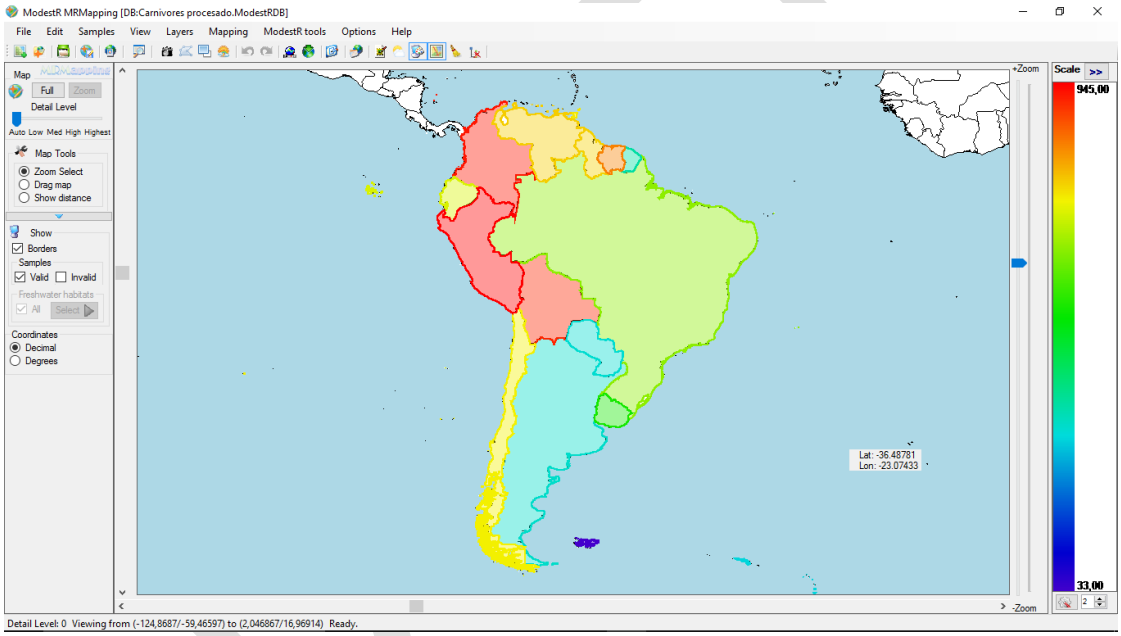



The options are:

- Use full/single names: if we are using full names as described above, we'll select the first option.

- First row contains headers: if enabled, the first row will be skipped. In the example we are describing, we'll disable this option because the file does not have headers.
- Data contains color column: if the value we added to each shape is a color in RGb or hexadecimal format, we'll enable this option. For the example we are describing, we'll disable this option because the values are not intended to be colors.
- Data contains value column: if the value we added to each shape is a quantitative value, we'll enable this option. For the example we are describing, we'll **enable** this option.
- We can optionally change the decimal and field separator according to the CSV file format we are importing.

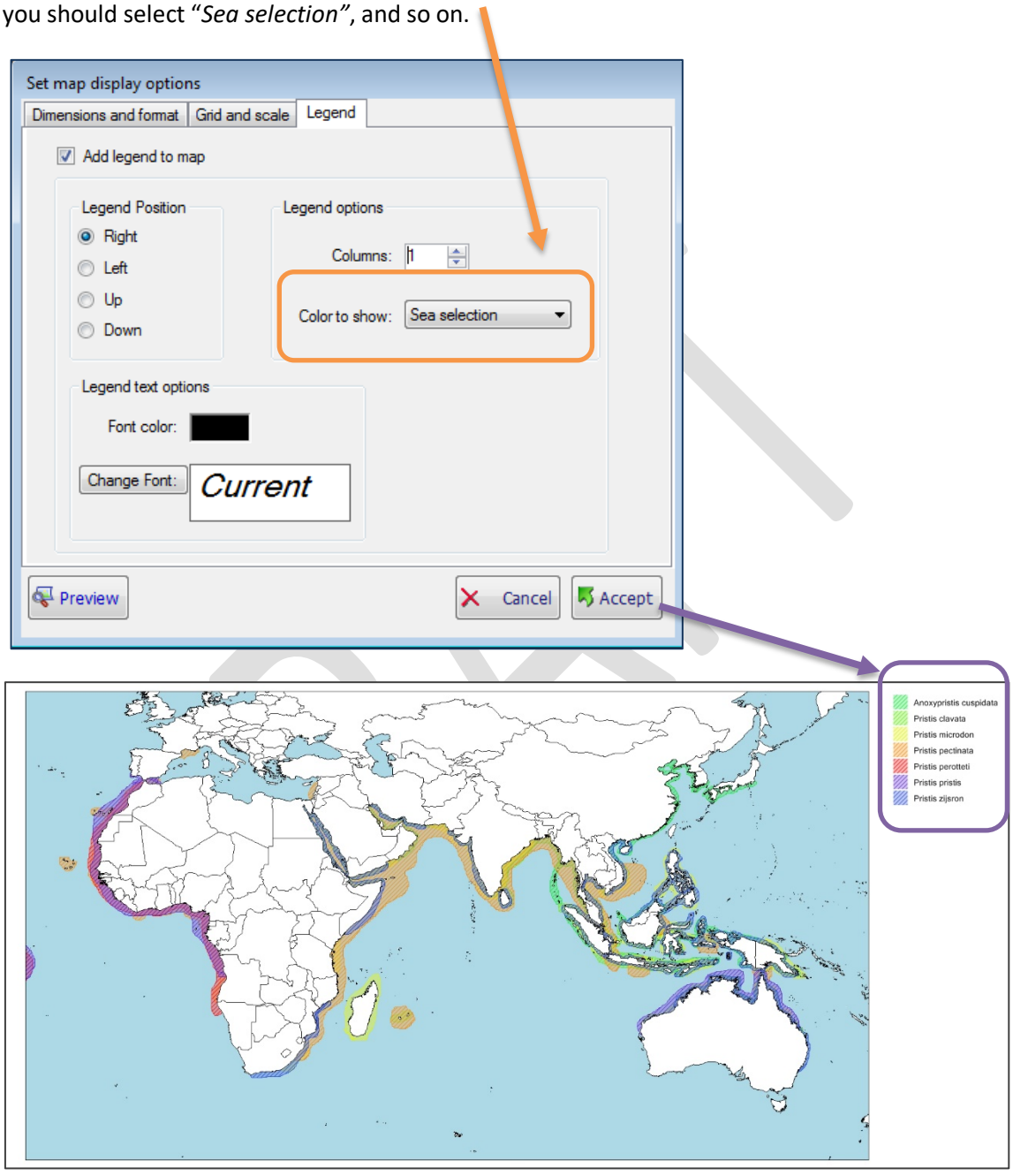
Once options set, we will click on *Accept*. MRMapping will read each row, search for the corresponding shape in the shape database and assign the corresponding value to it. This value will be used to build a color scale and show the results in the map. For the example we are describing, the result will be like that:




## 5.5 Exporting map image

To export the current map as an image in JPG or BMP format, go to *File/Export/Current map view as image*. A dialog box will appear with several tabs. In the *Grid and scale* tab you can set if you want to see a grid/scale in the exported image, each degree/minute you want this grid/scale, etc. In the *Dimensions and format* tab you can directly modify the coordinates of the portion of the map that will be exported. By default, the same view currently displayed in MRMapping will be exported. You can also select the size of the exported image. By default, MRMapping sets the image size as slightly bigger than your screen resolution. This is usually a good resolution to use the image in another document, presentation, etc. But if you want a more precise image that can be zoomed correctly, you can increase the image size. Finally, you can select between JPG or BMP format.

In the *Legend* tab you can add a legend to the map, where each distribution (dataset) name will be shown, with the corresponding color. You can select the position of the legend and the font. The *Color to show* combo box allows you to select which color will be shown in the legend for each map item. If the distributions you added to the world map contain occurrences, you should choose the color of the occurrences, while if they mainly contain presence areas in sea habitats, you should select "*Sea selection*", and so on.




## 5.6 Exporting environmental data. Raster data clipping

Similarly to MapMaker, you can export environmental data (or any raster data in general) clipped to the presence areas of each distribution (dataset) in the world map (see 2.9.4). The difference being that in MapMaker this feature allows you to clip rasters to the presence areas of a single species, while in MRMapping, as in the world map you could have added several distributions from several species, families, orders, etc., you can clip rasters to each one of those distributions.

To do that go to *File/Export/Environmental data/For each distribution*. A dialog box will show to you the available variables (those previously integrated in ModestR). We'll select the wanted one (it could be more than one, to generate several clipped files). Next, we'll have to selected the target folder where the file will be saved, in ESRI ASC format. A clipped raster will be exported for each distribution and for each selected environmental variable. Be careful because any existing file with the same name will be overwritten!

A report in CSV format is also generated with data about clipped areas in $km^2$ and simple statistical measures (max, min, mean, std and quartiles) for the environmental data in the clipped areas (see 7.9 about the similar report in MapMaker, for more details).

## 5.7 Quantifying occupied areas

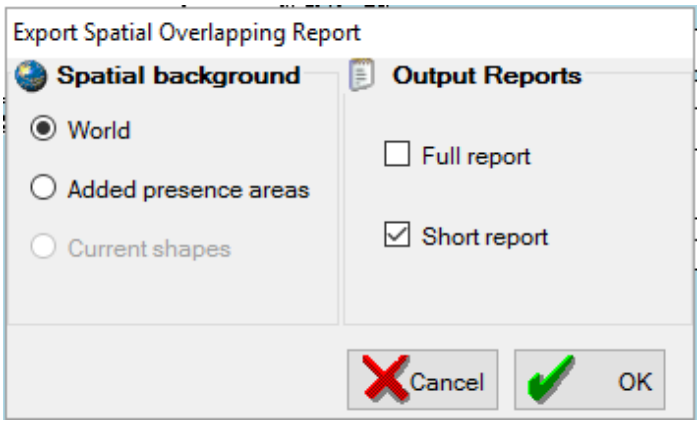
Similarly to MapMaker, you can quantify and export a report about the presence areas of each distribution (dataset) in the world map. The difference being that in MapMaker this feature allows you to measure the presence areas of a single species, while in MRMapping, as in the world map you could have added several distributions (datasets) from several species, families, orders, etc., you can measure the presence areas of each one of those distributions.

To do that go to *File/Export/Distributions/Areas summary report.* This generates a report in CSV format with the areas in $km^2$ of each type of habitat where each distribution is present. See 7.10 for details about the exported report.

## 5.8 Quantifying spatial overlapping areas

MRMapping can quantify the areas where several distributions overlap. That is, the areas where two or more species (or any other rank you added to the map, such as families, orders…) are present. To do that go to *File/Export/ Distributions/Geographic overlapping.* A dialog box will allow selecting the reference area or geographic background to be used. The available options are (1) the whole world, (2) a country, a river basin or any other desired area that can be added by the user as a shape file; and (3) the added presence areas, i.e. the extent of occurrence (EOO's) of the considered distributions.

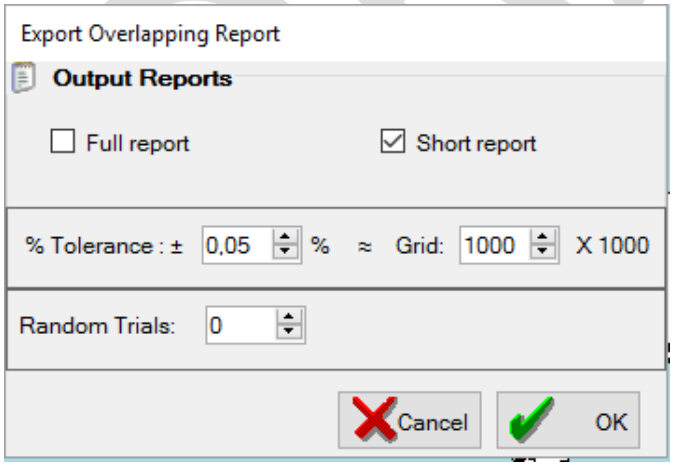
You can choose between a short report and a full report. The short report contains data about total overlapping areas and percent between each pair of distributions. The full report contains the same information, but broke down by type of habitat (e.g. land, sea, lentic waters….). See 9.1 for details about the exported report.




## 5.9 Quantifying environmental overlapping areas

The calculation of environmental overlaps in ModestR requires the usage of a 2DCEL that can be considered similar to a geographical background, translated to the environmental space. In a few words, a 2DCEL can be seen as a multidimensional compound of several environmental variables projected in a bidimensional polar coordinates system. The variables used to build an EL will obviously influence the results, and using different EL's will lead to different overlap measures. An EL can be built considering a specific geographic region selected by the user as reference area (e.g. a country, a drainage basin, the extent of occurrence of one or more species, or an arbitrary polygon). A step-by-step explanation about how to create a 2DCEL in ModestR can be found in section 2.11 .

To calculate the overlaps between each pair of environmental distributions in MRMapping the user has to go to the menu *File/Export/ Distributions/Environmental overlapping* and select the EL to be used as reference.




A dialog box allows setting the following parameters:

- Full report and/or short report: the full report includes graphs showing the environmental overlaps between each pair of ditributions, and more detailed information in the generated report file.

- Tolerance/Grid: in order to calculate overlaps, the polar coordinates system of the EL used as reference is split into a grid in which the cell size can be selected by the user, either setting the grid dimensions or a tolerance value that may oscillate from ±0.01% to ±5% of the size of the whole polar coordinates system. Overlaps are measured just by counting the cells of the grid that contain points corresponding to the distributions to be compared. The higher the tolerance the higher is the grain size used to perform calculations, then the more likely it is the degree of overlap.
- Random trials: in order to provide a supplementary assessment on the significance of the results, it is possible to compare the observed results with those of an arbitrary number of null distributions. These are generated by polling the occurrences of two environmental distributions and calculating the overlap of a number of randomly selected points equal to the original ones.

See section 9.2 for detailed explanations about the exported report.

# 6 DataManager import/export files format

In this section we briefly explain the main file formats used by ModestR DataManager to import/export data from/to a ModestR database. Remember that some exportation defaults concerning CSV format can be set in the DataManager default settings (see 3.2).

## 6.1 Importing taxonomy from a CSV file

You can import taxonomic data in CSV to a ModestR database using the *Import/Import Taxonomy/From CSV file* menu option.

Data to be imported have to be in CSV format, either separated by comma or semicolon. They can contain headers. You should indicate those options in the dialog box that will appear before importation. Each row must correspond to a species to be imported. Each row must contain the class, order, family, genus and species, in this order. Importation process is not case-sensitive.

For example, those data will be valid to be imported:

```
Elasmobranchii;Torpediniformes;Narkidae;Typhlonarke;Typhlonarke aysoni
Elasmobranchii;Torpediniformes;Narkidae;Typhlonarke;Typhlonarke tarakea
Actinopterygii;Zeiformes;Oreosomatidae;Allocyttus;Allocyttus folletti
Actinopterygii;Zeiformes;Oreosomatidae;Allocyttus;Allocyttus guineensis
Actinopterygii;Zeiformes;Zenionidae;Capromimus;Capromimus abbreviatus
Actinopterygii;Zeiformes;Parazenidae;Cyttopsis;Cyttopsis rosea
```

Taxonomy data will be imported to the currently opened database. All required taxonomic levels will be added to the database for each species contained in the CSV file. If some taxa (either a class, order, family, genus or species) already exists in the database, it will be simply skipped.

## 6.2 Importing taxonomy from another ModestR database

You can import taxonomic data from a ModestR database to the currently opened ModestR database using the *Import/Import Taxonomy/From another ModestR database* menu option. In this case, the expected format of the origin database is the original format of a ModestR database. No modifications have to be done. Just take into account that this option only imports taxonomic data, but not existing maps from the origin database.

## 6.3 Importing taxonomy from phyloXML files

You can import taxonomic data from a phyloXML file to the currently opened ModestR database using the *Import/Import Taxonomy/From phyloXML file* menu option. DataManager will search for <taxonomy> elements to import, using the <rank> element to determine the taxonomic level of the data. You have to be aware that phyloXML files not always contain comprehensive taxonomic information, so in some cases DataManager will not be able to import it. For example, a species name from a phyloXML file will only be imported if all superior levels (until class level) are also mentioned in the phyloXML file, or if the genus level already exists in the target ModestR database.

## 6.4 Importing taxonomy from ITIS database

As explained in section 3.19 DataManager provides a tool to browse and extract data from ITIS taxonomic database to a CSV file. This CSV file can be directly imported to a ModestR database as explained in section 6.1.

## 6.5 Importing Map Files from MapMaker map files

You can import map files to the currently opened ModestR database using the *Import/Import presence data to checked branches/From standalone map files* menu option. This option limits the species to be imported to those of the branches you have previously checked of the taxonomic tree. To avoid any restriction you can just check the root node, so any species of the entire database can be imported. ModestR DataManager can import map files generated with MapMaker and saved as standalone map files (files with *.IPMMAC* extension). This option will ask the user for selecting a folder, and it will import all map files from this folder to the currently opened database. Maps to be imported must contain taxonomic information (at least the species name), and the species assigned to the map has to already exist in the target ModestR database.

## 6.6 Importing Map Files from another ModestR database

You can import map files to the currently opened ModestR database using the *Import/Import from another ModestR database* menu option. In this case, the expected format of the origin database is the original format of the Modest R database. No modifications have to be done. Once selected the origin database to import maps from, a dialog box as seen in the figure will allow setting the importation options. First two options affect taxonomic data:

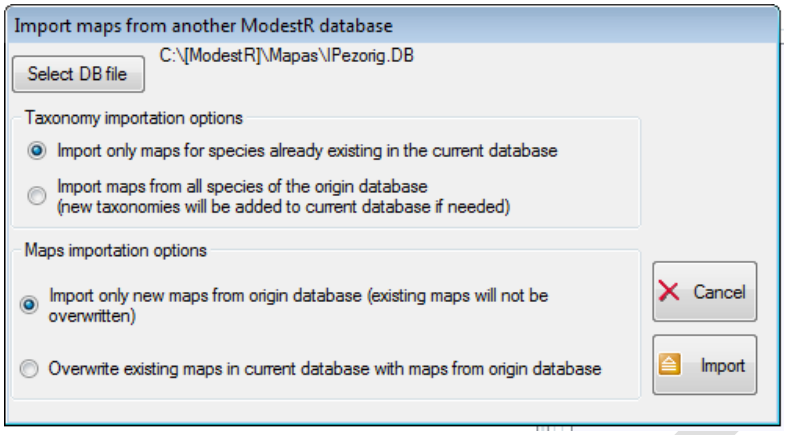
- Import only maps for species already existing in the current database: if selected, no new taxa will be added. Therefore only maps corresponding to class/order/family/genus/species that already exist in the currently opened database could be imported. If a map for the species X exists in the origin database but the species X doesn't exist in the currently opened database, the map will not be imported.
- Import maps from all species of the origin database: if selected, all maps of the origin database could be imported. Consequently, new taxonomies will be added to currently opened database if needed. So, if a map for the species X exists in the origin database but the species X doesn't exist in the currently opened database, species X will be added to taxonomic data in the target database (including lower levels if necessary, like genus, families, etc.) , and the map will be imported.

Take into account that all the superior levels are compared to determine if an imported species corresponds to an existing species. Even if they have the same name, if they differ in one superior level (for example they have a different family) two species will be considered as different during importation.

Next two options affect overwriting maps during importation:

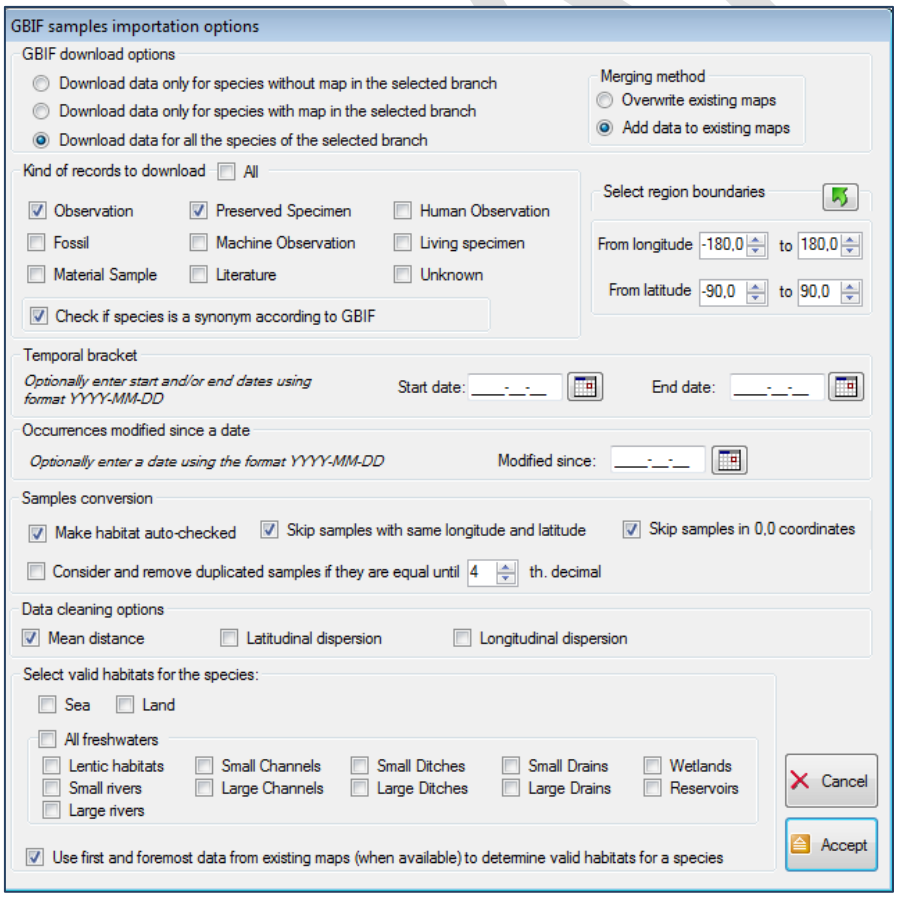
- Import only new maps from origin database: if selected, only maps for species that don't have maps in the currently opened database will be imported. Existing maps will not be overwritten. Of course, maps for species that don't exist in the currently opened database will be imported or not regarding the option selected regarding taxonomic data importation previously described.
- Overwrite existing maps in current database with maps from origin database: if selected, existing maps in the currently opened database could be overwritten if there is a map for the same species in the origin database.

You have to take care of the fact that different combinations of the selection options will affect maps importation in different ways.




## 6.7 Importing occurrences from GBIF online database

You can import occurrences and create maps downloading data from GBIF database (Global Biodiversity Information Facility, http://www.gbif.org). This can be done using the *Import / Import presence data to checked branches /From online GBIF database* menu item in the main menu. This option limits the species to be imported to those of the branches you have previously checked of the taxonomic tree. To avoid any restriction you can just check the root node, so data for all species of the entire database will be imported.



This option will download occurrence data from GBIF for each species of the checked branches, and process them regarding the options you set in the dialog box that will be shown previously. According to the settings you select, new maps can be created, existing maps can be overwritten, or new data can be added to existing maps, for example. A brief explanation of each option of this dialog box will be shown in an emerging *tooltip* if you place the mouse on it. Several options are already described in section 2.3 and in section 6.8 below. Here we summarize the most relevant ones:

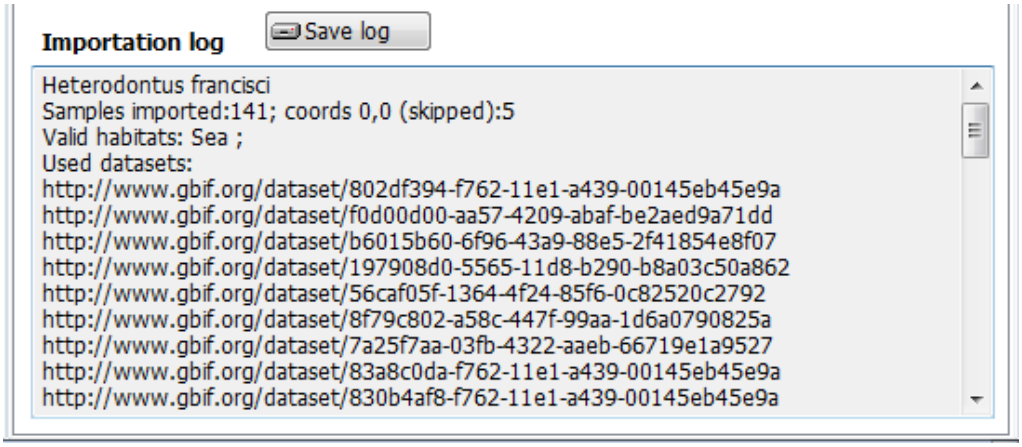
- *Download data only for species without map*: Download only data for species that still not have maps. This option avoids overwriting already existing maps.
- *Download data only for species with map*: Download data only for species that already have maps. This option is useful to add new data to already existing maps or to overwrite them with new data.
- *Merging method* can be:
  - *Overwriting existing maps with new data*: in this case, if a map already exists for a species that will be downloaded from GBIF, this map is deleted and a new map is stored only with new data downloaded from GBIF.
  - *Adding new data to existing maps*: in this case, if a map already exists for a species that will be downloaded from GBIF, new data downloaded from GBIF are added to this map. Therefore no previously existing data is deleted.
- *Check if species is a synonym according to GBIF*: indicates if you want DataManager to look for synonyms in GBIF database. In this case, if GBIF indicates that a species is a synonym of the species you are downloading, DataManager will use the main species name provided by GBIF to download data. For example, if DataManager is downloading data for species *Isurus alatus*, and this option is enabled, GBIF will indicate that this species is a synonym of *Isurus paucus*, and then DataManager will use *Isurus paucus* to download data and add them to the *Isurus alatus* map. If this option is disabled, only the name that appears in the ModestR database will be used, but no synonyms (in this case only *Isurus alatus* will be used to download data from GBIF).
- *Kind of records to download*: allows you to select the type of records you want to download from GBIF. Remember that to obtain a brief explanation of what each option and field is, just place the mouse on it to see a tooltip. However, a more complete description can be found on GBIF documentation.
- *Select region boundaries*: indicates the rectangular area to be used as boundaries (by default it is the whole world) for the occurrences to be downloaded from GBIF.
- *Temporal bracket*: indicates the start and end date of the occurrences to retrieve.
- *Occurrences modified since a date*: download only the occurrences added or modified into GBIF database since a particular date.
- *Data cleaning options:* allows selecting variables to apply a dispersal capacity cleaning once data downloaded (see section 2.6 for more details about dispersal capacity cleaning).
- *Skip occurrences in 0,0 coordinates*: don't process occurrences whose coordinates are 0º longitude and 0º latitude, as those occurrences are usually erroneous data.
- *Skip occurrences with same longitude and latitude coordinates*: skip occurrences whose longitude and latitude coordinates have the same value, as those occurrences are usually erroneous data.
- *Consider and remove duplicated occurrences if they are equal until the Xth decimal*: this allows setting when two occurrences will be considered as duplicated, thus avoiding adding them twice. If, for example, we configure this option to take into account until the 2th decimal, occurrences like (122.4567;9.4568) and (122.4598;9.4560) will be considered

duplicated. If you don't check this option, all occurrences will be imported and added to the map, even if they are duplicated.

- *Make the occurrence habitat auto-checked*: using this option an occurrence will be automatically checked by DataManager to determine if it is situated in a valid or invalid habitat, and therefore it will be considered valid or invalid with a precision of aprox. ±1 seconds[23]. You can set different colors for valid an invalid occurrences to easily distinguish them in the Options/Preferences option on the main menu of DataManager. Moreover, only valid occurrences will be used when processing and extracting data from a map (like presence or richness data). If on the contrary you don't set an occurrence as dynamic, it will be considered as manually validated/invalidated. By default, it will be valid, and you have to manually set it as invalid if needed. Therefore, we recommend you to try the dynamic validation.
- *Select valid habitats for the occurrences*: you have to select which habitats are valid for the occurrences you are importing (this option appears on the left tool panel when manually adding occurrences). You can select one or more occurrences, but it is mandatory to select at least one habitat. If you selected to make the occurrences as habitat auto-checked, those habitats will be used to validate/invalidate a occurrence. If you don't selected to make the occurrences as dynamic, this information will NOT be used, as occurrences have to be manually validated/invalidated. But it will be stored and used in case you later modify an occurrence to make it as habitat auto-checked.
- *Use first and foremost data from existing maps (when available) to determine valid habitats for a species*: if the map for a species already exists in the database, DataManager will look for data from this map (occurrences and zone selections) to determine which habitats are valid for this species. If no map exists, habitats selected above by the user will be used.

DataManager will download data and automatically transform it in ModestR maps, storing them in the database. You can configure the importation for adding downloaded occurrences to already existing maps (if any), or for overwriting existing maps. After downloading, you will see an *Importation Log* that includes a link to the information about the datasets that provided the downloaded data.

[23] This requires installing the last software updates. If you want to distinguish freshwaters, you also have to install freshwater data in ModestR (available in ModestR website). Anyway, as freshwater habitats are still being reviewed and updated, in some areas checking feature may only distinguishes between land and sea habitats.




## 6.8 Importing occurrences from CSV files

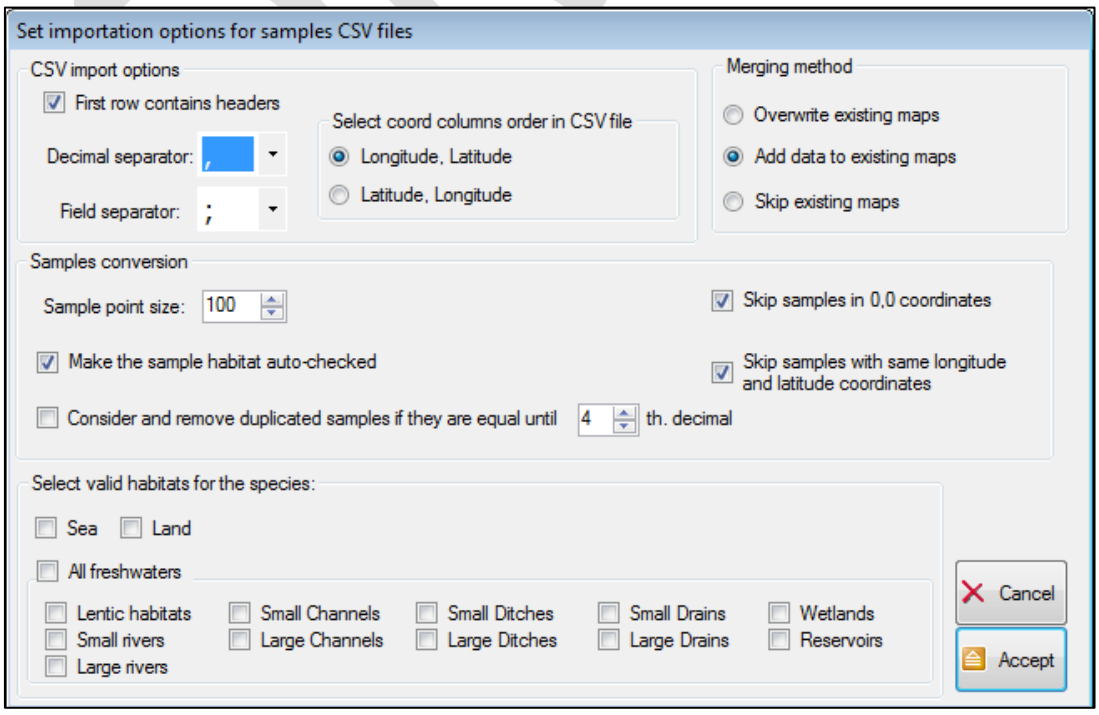
You can import occurrences and create maps importing data from CSV files containing occurrences/occurrences for one or more species. This can be done using the *Import/* Import presence data to checked branches/*From CSV occurrence files*. This option will import occurrence data from a CSV file. Expected format is a list of occurrences in the form:

Species; Longitude ;Latitude

or

Species; Latitude; Longitude

This way you can import a single file containing occurrences for several species, but only species already existing in the checked branches will be imported (that is, species in the checked branches with exactly the same name than the specified in the imported file). A dialog bow will be shown to select the importation settings. A brief explanation of each option of this dialog box will be shown in an emerging *tooltip* if you place the mouse on it.



In the dialog box you can select the correct options for the format of the CSV file to import. Most of them are the same than when importing GBIF data. Therefore you can refer to section 6.7 for an explanation of those options. The options that are specific to CSV importations are:

- The order of the coordinates in the CSV file (Longitude ;Latitude or vice versa)
- If the first row contains headers (in that case first row will be skipped).
- Decimal separator and field separator.
- Occurrence point size: select the size of the visual point that will indicate an occurrence in the map. This size has no importance when processing map. It is only a visual parameter.

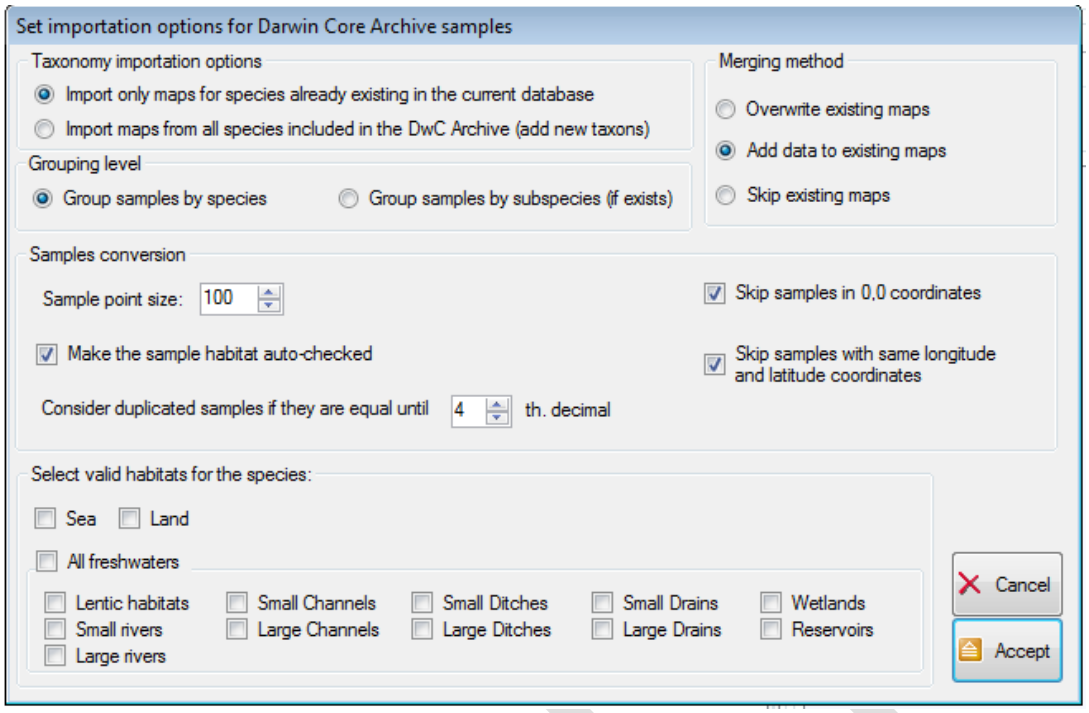
After importation, you will see the list of species imported, and an *Importation Log* that includes information about the imported data and eventual errors, like format problems, or species not imported because they are not found in the database.

## 6.9 Importing occurrences from Darwin Core Archives

You can import occurrences to a DataManager database from a Darwin Core Archive (DwC), which is a format notably used to share occurrence data through GBIF (www.gbif.org). To do that, you can use the *Import/ Import presence data to checked branches/From Darwin Core Archive of occurrences* option. In this later case, only occurrences for the selected branch will be imported.

The first step will be to select the Zip file that contains the Darwin Core Archive data. DataManager only supports Darwin Core Archives that contain a meta.xml file which describes the structure of the ocurrences file, which can be in CSV or text-delimited format. Moreover, the occurrence fields that DataManager will require to import data are *class, order, family, genus, specificEpithet* (or alternatively *scientificName*), *taxonRank, decimalLatitude*, and *decimalLongitude*. If some of these fields are empty, occurrences cannot be imported.

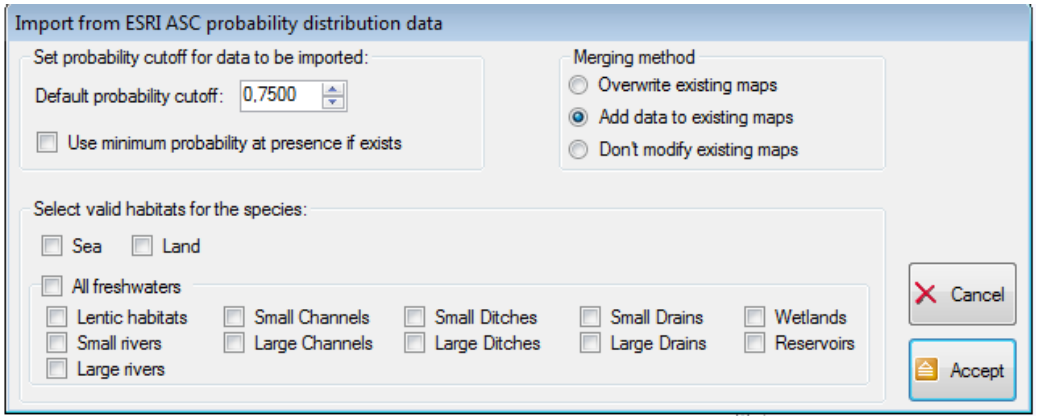
The next step is to configure the importation setting that will be shown in a dialog bog quite similar to the shown when importing from CSV occurrence files or GBIF (see sections above). The main difference is that you can select to only import occurrences from species already existing in the database, or to also import occurrences for new species. In this last case, new taxons (*class, order, family, genus, or species)* will be added to the database if needed, using the information included in the Darwin Core Archive you are importing.




## 6.10 Importing data from distribution models

You can import distribution data from distribution models as ESRI ASC raster matrices of probability, like the obtained using Maxent (Phillips & Dudík, 2008) to the currently opened ModestR database using the *Import/Import presence data to checked branches/ From ESRI ASC probability distribution data* menu option. This importation option allows to produce range maps from probability distribution data. This option limits the species to be imported to those of the branches you have previously checked of the taxonomic tree. To avoid any restriction you can just check the root node, so any species of the entire database can be imported.

You should have previously generated one or more ESRI ASC distribution files that will contain a probability matrix, where each element of the matrix represents the probability of presence for a species in a cell of a specified size (usually 5'x5'). Those are the files that DataManager can import to generate new maps or to add data to existing maps. Only filenames which name matches a species already existing in the database will be imported. DataManager will be aware of the filename format used by some software like Maxent, which uses a low dash between genus and species, so you don't need to modify it. That is, "Ateleopus_indicus.asc" will be recognized by DataManager as "Ateleopus indicus" species.



When using this importation option, you first have to select the folder where probability ESRI ASC files are stored. Then a dialog box will allow you to set several settings:

- *Default probability cutoff*: the minimal probability under which presence data will not be imported. If for example you select 0.75, only cells with a probability >=0.75 will be imported to the map.
- *Use minimal probability at presence*: this option is **only** useful if you are importing data from species that already have a occurrence-based map on the current database. In this case, this option allows you to use as probability cutoff the lowest probability value where an occurrence exists in the existing map. For example, you can have several maps with occurrences in the database for several species, then import probability distribution data files for those species to the database. Then DataManager will automatically find the probability values in the points where there are occurrences in the corresponding map, for each species and use the minimum of these values as probability cutoff to import probability data. The selected values for each species are displayed after importation. If you import data from species that don't have a occurrence-based map on the current database, default probability cutoff will be used (even if you activate the "use minimal probability at presence" option).
- *Valid habitats*: you have to select which habitats (at least one) are valid for the species data you are importing. Only areas corresponding to those habitats will be accepted and add to the map.
- *Merging method*: you can select to overwrite existing maps (if any) with new maps created using imported data, to add new data to existing maps, or to not modify existing maps, so importing only data for species that still don't have a map in the database.

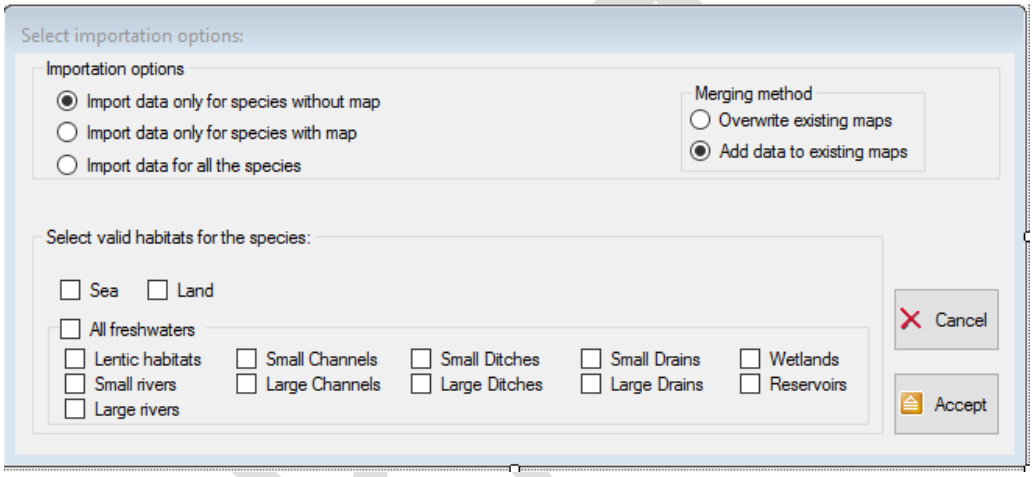
Once accepted this dialog, the importation process will start. This can be time-consuming if you import a lot of maps, or they are complex. When the process will be completed, a report showing imported maps and possible errors will be show in the right panel.

## 6.11 Importing distribution maps from shapefiles

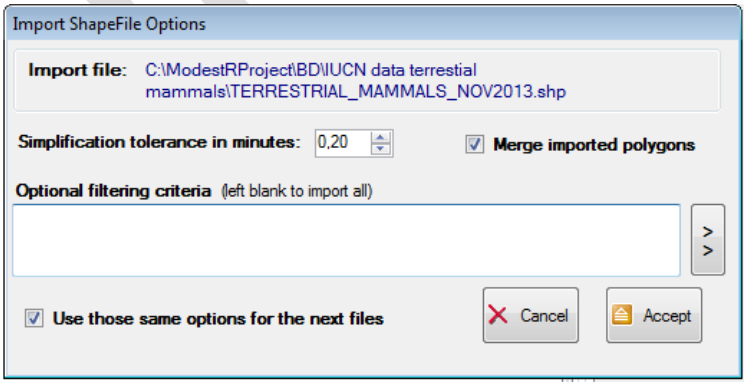
As explained in section 7.4 you can import a single shapefile to a distribution map in MapMaker. But if you want to import a set of shapefiles or a single shapefile that contains distributions maps for several species, this option can be laborious. For this situation, it can be more interesting to import shapefiles directly to a ModestR database using DataManager, as it allows importing several distribution maps in a single operation. Next we explain the steps to follow:

- First of all, you have to have a ModestR database already created and containing the corresponding taxonomy with the species you want to import. DataManager will only be able to import shapefiles that contain a data field that identifies the species names of the maps to be imported, or alternatively to use filename as species name. Those species names have to exactly match with existing species in the target ModestR. DataManager cannot import distribution maps for species that don't yet exist in the ModestR database.
- Next, go to *Import/Import Presence data to checked branches/From shapefiles* menu option . Only species included in the currently checked branches will be imported.

- DataManager will warn about the fact that this task can modify the database and suggest doing a backup. As imported data will OVERWRITE existing one, it may be recommendable to accept doing a backup to avoid data losing in case of an unexpected modification or error.
- Next, a dialog box will allow you to select one or more shapefiles to import. As in Windows Explorer, you can select several files just by using the Ctrl or Shift (get more information in Windows help).
- Then a dialog box allows you to select if you want to import data for species that don't still have a map in the database, or for those that already have it, or for all (in the last two cases you can decide if imported data will overwrite or will be added to existing maps). You shall also select which habitats will be considered as valid for all the species to be imported (when importing several species in a batch, you have to assign the same valid habitats to all species).

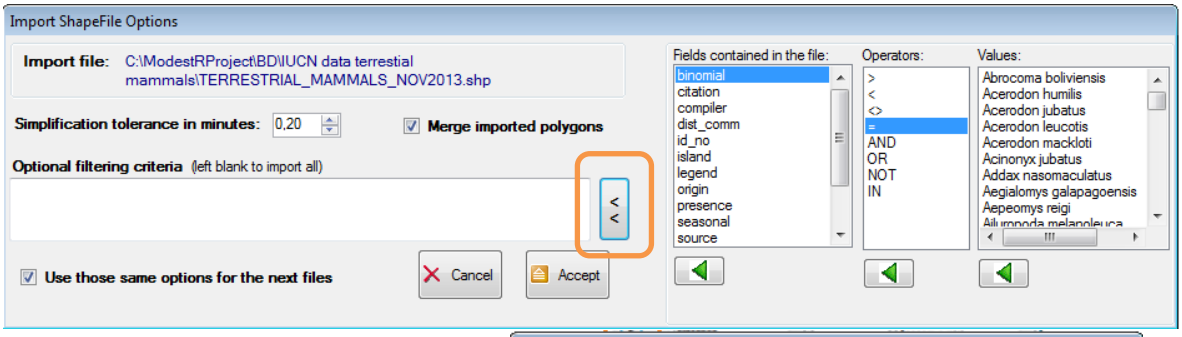


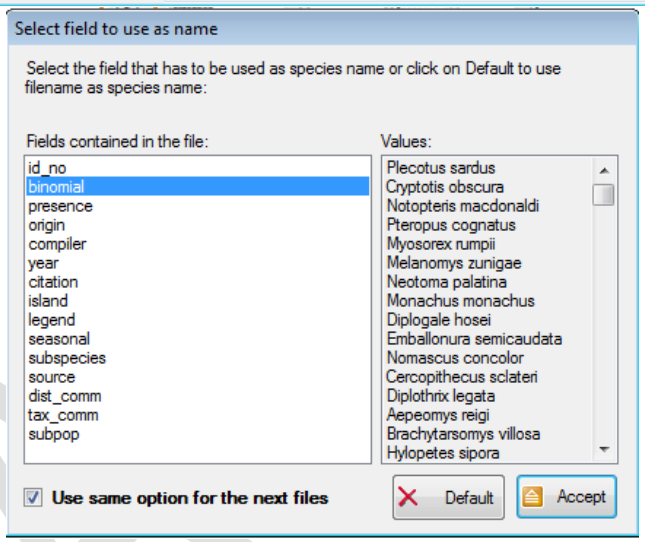

- DataManager will read the first shapefile to be imported (there may be several ones) and show to you this dialog box:




Here you have the *Merge imported polygons* option. This option will merge imported polygons for each species, so overlapping polygons will be merged and therefore data simplified. This option can cause errors when importing shapefiles, that is why it is disabled by default. The same way, DataManager will not apply by default a simplification factor to imported polygons, even if this options can positively affect performance. If you are familiar with shapefile format, you can also add importation

criteria to filter which data will be imported. To do that, use the *See/hide fieldlist* button to see the list of fields contained in the shapefile.




Then you can select a field to see the values it takes, and add criteria that will be used as filter when importing. You can also directly write a condition in the textbox. The syntax of the conditions is similar to the syntax of a WHERE clause in SQL.

Finally you can decide to use those same settings for all remaining files to be imported (if there is only one file to be imported, this option will have no effect). This option is checked by default, so it will only be shown for the first file to be imported. But if you want to use different settings for each file to be imported (for example different filtering criteria) you can unchecked it, then DataManager will show this dialog box for each file to be imported.

- Next, a dialog will be shown where you have to select the shapefile field that will be used to identify the species name. This is mandatory when importing a shapefile that contains several species maps. When importing several shapefiles where each one contains a single species map, you can alternatively use the filename as species name, by clicking on the *Default* button.
  In this dialog you can see in the left side a list of the data fields contained in the shapefile to be imported. Selecting one of those fields, in the right list you can see the values took by this field. This can help you to select the right field that contains species names.
  You can decide to use this same field as species name for all remaining files to be imported (if there is only one file to be imported, this option will have no effect). This option is checked by default, so it will only be shown for the first file to be imported. But if you want to use different fields for each file to be imported you can unchecked it, then DataManager will show this dialog box for each file to be imported.

After that last step, importation will start. You can cancel it in any moment, but species already imported until this moment will be added to the database.

At the end of the task, information about imported species, not imported and warnings will be shown on the right panel.

## 6.12 Exporting taxonomy to a CSV file

You can export taxonomic data of any branches you have checked of a DataManager database using the *Export/Export taxonomy of the checked species* menu option. You will see two options to export data to a CSV file in two different formats:

- To a CSV (flat) file: data will be exported to a CSV file using the field delimited selected in the preferences (Options/Preferences menu). First row will contain headers. Each data row will contain the class, order, family, genus and species, in this order. The *Map exists* column indicates with a 1 or a zero value if there is a map in the database for this species or not, respectively. Here is an example:

| Class | Order | Family | Genus | Species | Map exists |
|---|---|---|---|---|---|
| Mammalia | Carnivora | Ailuridae | Ailurus | Ailurus fulgens | 1 |
| Mammalia | Carnivora | Canidae | Alopex | Alopex lagopus | 1 |
| Mammalia | Carnivora | Canidae | Atelocynus | Atelocynus microtis | 1 |
| Mammalia | Carnivora | Canidae | Canis | Canis adustus | 1 |
| Mammalia | Carnivora | Canidae | Canis | Canis aureus | 1 |
| Mammalia | Carnivora | Canidae | Canis | Canis latrans | 1 |
| Mammalia | Carnivora | Canidae | Canis | Canis mesomelas | 1 |

- To a structured CSV file: data will be exported to a CSV file using the field delimited selected in the preferences (Options/Preferences menu). First row will contain headers. Next row will contain a class. Next row will contain an order of this class, and so on until species level. Once listed all species of a genus, next genus will be listed and so on. "Total Species" column will contain the total number of species in the current taxonomic level (class, order, family…). Here is an example:

| Class | Order | Family | Genus | Species | Total Species | Map exists |
|---|---|---|---|---|---|---|
| Mammalia | | | | | 6 | |
| | Carnivora | | | | 6 | |
| | | Canidae | | | 6 | |
| | | | Canis | | 6 | |
| | | | | Canis adustus | | 1 |
| | | | | Canis aureus | | 1 |
| | | | | Canis latrans | | 1 |
| | | | | Canis mesomelas | | 1 |
| | | | | Canis rufus | | 1 |
| | | | | Canis simensis | | 1 |

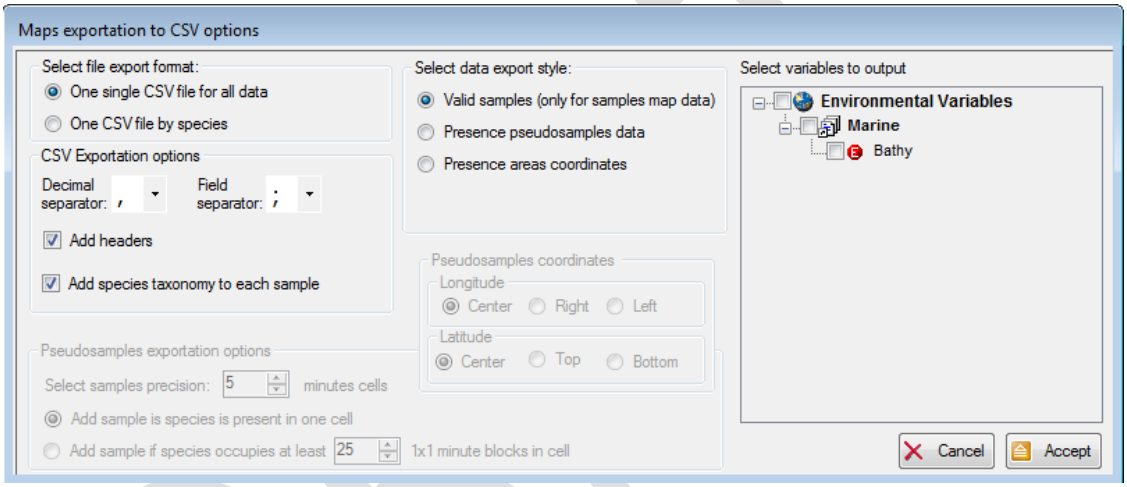
## 6.13 Exporting maps to standalone map files

You can export maps of any branches of a DataManager database to standalone map files (files with *.IPMMAC* extension) using the *Export/Export checked maps/To Standalone map files* option. The exported files can be opened with MapMaker. This option will ask the user for

selecting a folder, and it will export all maps from the currently checked branches of the database to this folder. Maps exported will contain full taxonomic information.

## 6.14 Exporting presence data to CSV data files

You can export maps of any branches of a DataManager database to CSV using the *Export/Export checked maps/ To presence data /To presence data and variables* option. This feature furthermore allows including environmental variables in the exported data. This option will show a dialog box with several selections. A brief explanation of each option of this dialog box will be shown in an emerging *tooltip* if you place the mouse on it. On the left groupbox you can select if the exported data will be exported to one separated file for each species of the selected branch, or to a single file where all species data will be consecutively stored. In the first case, you will be asked to select on file name to output, whereas in the second one you will have to select a folder where exported files will be saved. You can also set characters used as field delimiter and decimal separator.




The available exportation format options are:

- **Valid occurrences**: this option will export only valid occurrence data from maps. Therefore, it will only be useful if the maps contain occurrence data, because range data (polygonal area of presence) will be skipped. This option exports only occurrences checked as valid, either automatically or manually, without any extrapolation (on the contrary of pseudooccurrences exportation described above).
  You can add environmental variables from the right tree, if there are variables available (see section 6.26). In this case, a column will be added for each selected variable, with the value (if it exists) of this variable at the coordinates of the occurrences.
  The output will be a CSV file with this format. *CoordLastModified* column indicates the date when each occurrence coordinate has been modified (see section 3.16 for some explanations) . Taxonomy columns are optional depending of selected settings. As well as the last column of this example, which here is the Bathymetry variable:

| Class | Order | Family | Genus | Species | Longitude | Latitude | CoordLastModified | Bathymetry |
|---|---|---|---|---|---|---|---|---|
| Batoids | Pristiformes | Pristidae | Anoxypristis | Anoxypristis cuspidata | 100,095833 | 9,58749962 | 11/05/2015 0:00 | |
| Batoids | Pristiformes | Pristidae | Anoxypristis | Anoxypristis cuspidata | 75 | 10 | 11/05/2015 0:00 | -2204,7244 |
| Batoids | Pristiformes | Pristidae | Anoxypristis | Anoxypristis cuspidata | 149,600006 | -22,2830009 | 11/05/2015 0:00 | 0 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Batoids | Pristiformes | Pristidae | Anoxypristis | Anoxypristis cuspidata | 149,600006 | -22,2830009 | 11/05/2015 0:00 | 0 |
| Batoids | Pristiformes | Pristidae | Anoxypristis | Anoxypristis cuspidata | 75 | 10 | 11/05/2015 0:00 | -2204,7244 |
| Batoids | Pristiformes | Pristidae | Anoxypristis | Anoxypristis cuspidata | 146,832993 | -19,25 | 11/05/2015 0:00 | |
| Batoids | Pristiformes | Pristidae | Anoxypristis | Anoxypristis cuspidata | 140,166 | -17,4160004 | 11/05/2015 0:00 | 0 |
| Batoids | Pristiformes | Pristidae | Anoxypristis | Anoxypristis cuspidata | 80,3000031 | 13,0830002 | 11/05/2015 0:00 | 0 |
| Batoids | Pristiformes | Pristidae | Anoxypristis | Anoxypristis cuspidata | 80,3000031 | 13,0830002 | 11/05/2015 0:00 | 0 |
| Batoids | Pristiformes | Pristidae | Anoxypristis | Anoxypristis cuspidata | 80,3000031 | 13,0830002 | 11/05/2015 0:00 | 0 |
| Batoids | Pristiformes | Pristidae | Anoxypristis | Anoxypristis cuspidata | 80,3000031 | 13,0830002 | 11/05/2015 0:00 | 0 |
| Batoids | Pristiformes | Pristidae | Anoxypristis | Anoxypristis cuspidata | 136,75 | -12,1499996 | 11/05/2015 0:00 | |
| Batoids | Pristiformes | Pristidae | Anoxypristis | Anoxypristis cuspidata | 136,75 | -12,1333303 | 11/05/2015 0:00 | |
| Batoids | Pristiformes | Pristidae | Anoxypristis | Anoxypristis cuspidata | 141,516663 | -12,4333296 | 11/05/2015 0:00 | 0 |
| Batoids | Pristiformes | Pristidae | Anoxypristis | Anoxypristis cuspidata | 136,75 | -12,03333 | 11/05/2015 0:00 | 0 |
| Batoids | Pristiformes | Pristidae | Anoxypristis | Anoxypristis cuspidata | 141,5 | -12,75 | 11/05/2015 0:00 | 0 |

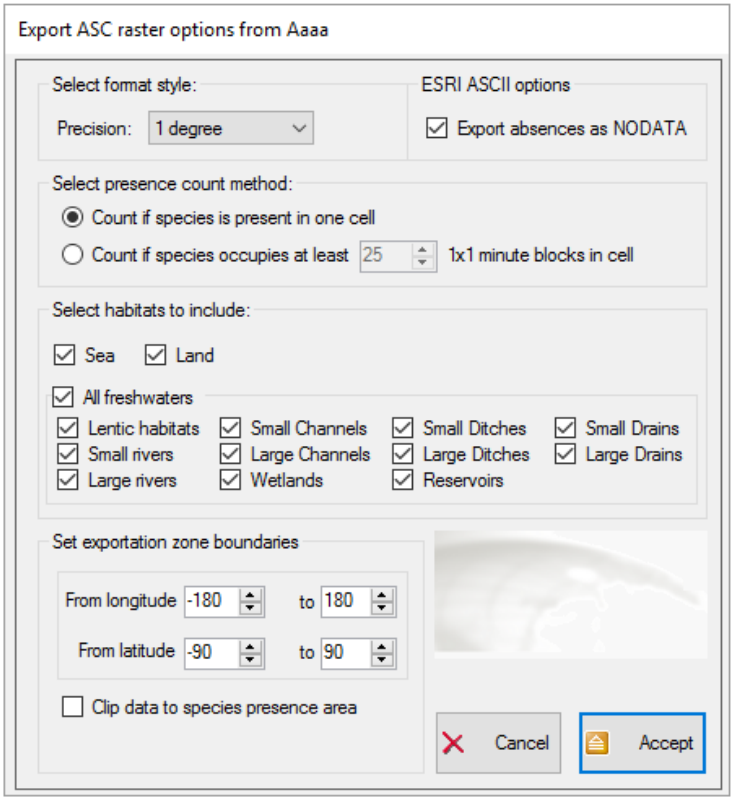
- **Presence pseudooccurrences data**: this option generates *pseudooccurrences* from rasterized species map, so basing on a 1'x1' cells precision. This requires that the maps are already rasterized (see 3.6). This option is useful to generate occurrence data from range maps where species presence is stored as polygonal zones. If the maps you have stored in the database already are based on occurrences or occurrence records (i.e. downloaded from GBIF), it's probably preferable to directly export those occurrences (see option above "Valid occurrences") in turn of generating pseudooccurrences.
  When selecting this option, *Pseudooccurrences exportation options* groupbox on the bottom will activate to allow configuring pseudooccurrences generation.
  You can also set the precision or density of the occurrences adjusting the width in minutes of the cells that will be used. If for example you select 5' cell width, at most one occurrence for each 5'x5' minutes cell on the world map will be generated, if the species is present in any point of this cell. You can also set if the occurrence will be generated with the coordinates of one corner or center of the cell. The smaller cell you select, the more numerous occurrences you have, and with the higher accuracy. But for large zones, this can result if a too large number of occurrences. You can also set if an occurrence has to be generated if the species is just present in the output cell, or in turn set a threshold of 1'x1' minutes cells occupation that has to be exceeded to generate an occurrence for this cell. For example, if you set a cell output of 5'x5', you could want that only if a species occupies at least 20% of this cell (that is 20% of 5x5= 5, so 5 cells of 1'x1' minutes) a pseudo occurrence has to be generated for this 5'x5' cell.
  You can also set how the pseudooccurrences coordinates will be exported regarding the cell they correspond to: using the longitudinal coordinate of the center, right or left corner of the cell, and so on for the latitudinal coordinate.
  You can also add environmental variables from the right tree. In this case, a column will be added for each selected variable, with the value (if it exists) of this variable at the coordinates of the pseudooccurrence.
  The exported format will be the same than for the previously described option of valid occurrences exportation, excepted for the *CoordLastModified* column, that obviously will not appear.

- **Presence areas coordinates**: this option will generate a CSV with the description of each polygonal area selected in the map as occupied by the species (Warning! This polygon will not be clipped to the selected habitat, it will contain all the coordinates selected by the user, regardless of the habitat).
  A *Outer/Hole* column indicates it the coordinates correspond to the outer limits of the polygon, or to an inner hole. The type indicates the habitat selected for this zone. A blank line separates each polygon.
  Valid occurrences contained in the map will also be exported, each one in a line, with the "FillOccurrence" type. Environmental data cannot be added when using this exportation option.

| Lon | Lat | Outer(1) /Hole(0) | Polygon | Type | Class | Order | Family | Genus | Species |
|---|---|---|---|---|---|---|---|---|---|
| -56,13 | 54,79 | 1 | 0 | Sea | Actinopterygii | Acipenseriformes | Acipenseridae | Acipenser | Acipenser brevirostrum |
| -56,13 | 55,13 | 1 | 0 | Sea | Actinopterygii | Acipenseriformes | Acipenseridae | Acipenser | Acipenser brevirostrum |
| -59,5 | 47,06 | 1 | 0 | Sea | Actinopterygii | Acipenseriformes | Acipenseridae | Acipenser | Acipenser brevirostrum |
| -59,83 | 47,4 | 1 | 0 | Sea | Actinopterygii | Acipenseriformes | Acipenseridae | Acipenser | Acipenser brevirostrum |
| -59,83 | 47,73 | 1 | 0 | Sea | Actinopterygii | Acipenseriformes | Acipenseridae | Acipenser | Acipenser brevirostrum |
| -60,17 | 47,73 | 1 | 0 | Sea | Actinopterygii | Acipenseriformes | Acipenseridae | Acipenser | Acipenser brevirostrum |
| -60,5 | 48,07 | 1 | 0 | Sea | Actinopterygii | Acipenseriformes | Acipenseridae | Acipenser | Acipenser brevirostrum |
| -60,84 | 48,4 | 1 | 0 | Sea | Actinopterygii | Acipenseriformes | Acipenseridae | Acipenser | Acipenser brevirostrum |
| -56,13 | 54,79 | 1 | 0 | Sea | Actinopterygii | Acipenseriformes | Acipenseridae | Acipenser | Acipenser brevirostrum |
| | | | | | | | | | |
| -77,98 | 33,94 | 1 | 1 | FillOccurrence | Actinopterygii | Acipenseriformes | Acipenseridae | Acipenser | Acipenser brevirostrum |
| | | | | | | | | | |
| -76,71 | 35,97 | 1 | 2 | FillOccurrence | Actinopterygii | Acipenseriformes | Acipenseridae | Acipenser | Acipenser brevirostrum |
| | | | | | | | | | |
| -67 | 40 | 1 | 3 | FillOccurrence | Actinopterygii | Acipenseriformes | Acipenseridae | Acipenser | Acipenser brevirostrum |
| | | | | | | | | | |
| -74,73 | 40,17 | 1 | 4 | FillOccurrence | Actinopterygii | Acipenseriformes | Acipenseridae | Acipenser | Acipenser brevirostrum |

## 6.15 Exporting presence data to raster files

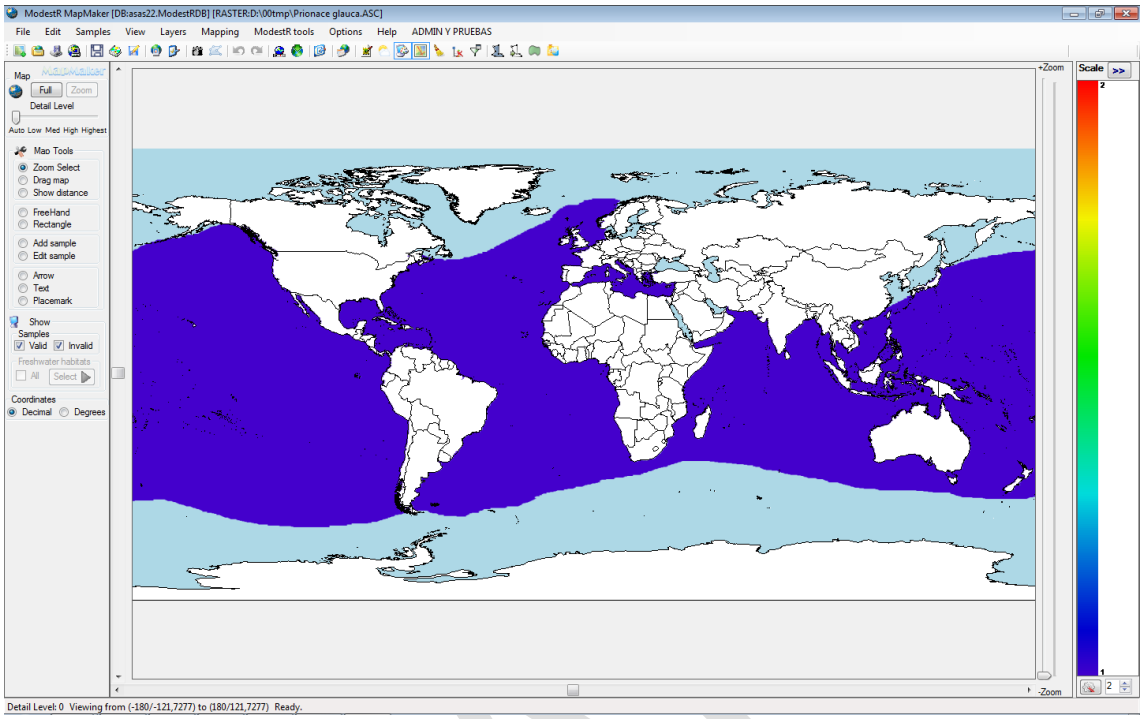
You can export maps of any branches of a DataManager database to individual rasters in ASC format, using the *Export/Export checked maps/ To presence data /To ASC raster maps* option.




- You can set the precision of the output adjusting the width in minutes of the cells that will be used. If for example you select 5' cell width, one metric value for each 5'x5' minutes cell on the world map will be generated. By default it is 1 degree.
- You can also set presence count method: count a species if it is just present in the output cell (default), or in turn set a threshold of 1'x1' minutes cells occupation that has to be exceeded to count a species in this cell. For example, if you set a cell output of 5'x5', you could want that only if a species occupies at least 20% of this cell (that is 20% of 5x5= 5, so 5 cells of 1'x1' minutes) the species has to be counted (so the value for this cell incremented).
- You can also select which habitats have to be included in presence detection. By default all habitats are took into account. If you exclude one habitat, all occurrences or zones with this associated habitat will be excluded from the presence detection.
- You can define the zone boundaries to be examined in degrees, with a minimum of a 1ºx1º zone.
- The "Export zero as NODATA" checkbox allows to select if a zero value (absence) for a cell will be exported as NODATA (usually will not appear when showing the raster in a viewer) or as zero (will be shown as a zero value).
- The "Clip data to species presence area" checkbox allows to automatically clip the ASC raster of each species to be exported to the area occupied by the species. This avoids exporting unnecessarily large rasters.

A single ASC raster file will be exported for each species distribution map in the selected branch. Cells where the species is present will be exported with a 1 value. Take into account that a huge amount of disk space can be required when using this option with many species and at a high resolution.

Remember that MapMaker can be used to visualize ASC rasters like the exported using this option. For example this may be the result of a map exported to ASC raster displayed in MapMaker.




## 6.16 Exporting occurrences to Darwin Core Archive style

You can export maps of any branches of a DataManager database to CSV data in a format that can be easily adapted to generate a Darwin Core Archive (DwC), which is a format used to share occurrence data through GBIF (www.gbif.org). To do that, you can use the *Export/Export checked maps/To Darwin Core Archive data* option. DataManager will add unique ID's for each occurrence, as required by DwC format. Exported format will contains headers for each column, and a row for each occurrence from each map exported. The format is based on the template provided by GBIF in http://tools.gbif.org/spreadsheet-processor/, with 45 data fields for ocurrences, but only some fields will be filed by DataManager, which are:

- occurrenceID: already explained above.
- class, order, family, genus: taxon names for those levels of the species
- specificEpithet: specific species name. For example, for "Isurus paucus" it will be "paucus".
- infraspecificEpithet: if you used subspecies at species level in the database, this field will be the subspecies name. For example for "Acipenser oxyrinchus desotoi" it will be "desotoi". If you used species names, it will be empty. For example, for "Isurus paucus" it will be empty.
- scientificName: full species name. For example, for "Isurus paucus" it will be "Isurus paucus"; for "Acipenser oxyrinchus desotoi" it will be "Acipenser oxyrinchus desotoi".

- taxonRank: rank of the taxon associated to the occurrences. It will be “species” by default, but if DataManager detect a species name with three words, it will use “subspecies” rank. For example, for “Isurus paucus” it will be “species”; for “Acipenser oxyrinchus desotoi” it will be “Acipenser oxyrinchus desotoi”.

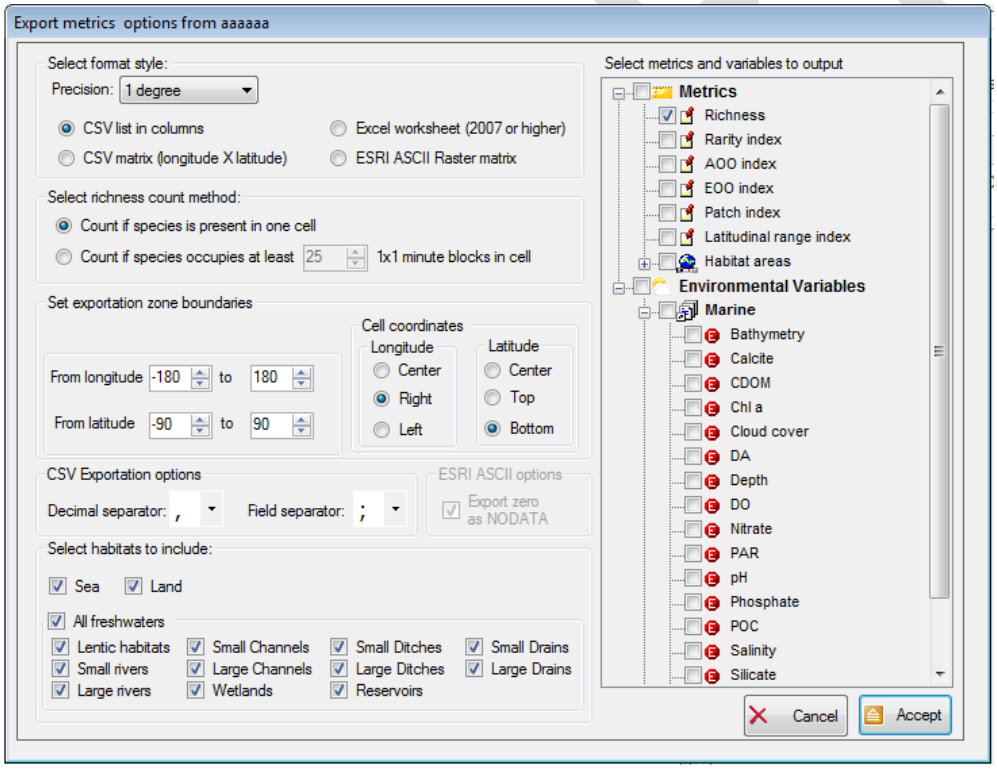
- decimalLatitude, decimalLongitude: coordinates of the occurrences.

The exported CSV file will be easily opened with a worksheet application like Excel and copied to the Excel template provided by GBIF in http://tools.gbif.org/spreadsheet-processor/, to subsequently generate a Darwin Core Archive.

## 6.17 Exporting metrics and environmental data

DataManager can calculate and export several metrics from maps, like richness data, rarity index, etc. If you have files containing environmental variables data, you can incorporate them to DataManager to make that they can be exported at the same time that metrics. The way to manage environmental variables is explained in section 6.26.

When selecting *Export/Export checked maps/To metrics and variables* in main menu, a dialog box to select exportation settings is shown:




- You can set the precision of the output adjusting the width in minutes of the cells that will be used. If for example you select 5’ cell width, one metric value for each 5’x5’ minutes cell on the world map will be generated.
- You can select which metrics will be in the output in the left selection tree:
    - *Richness:* the number of different species present in each cell of the selected dimensions. This will be equivalent to a 0/1 presence matrix if you select only one species for the output.

- *Rarity index* for each cell, as $CRI = \frac{\Sigma_{i=1}^{n} \frac{1}{AOO_i}}{n}$ where *n* is the richness of the cell, and $AOO_i$ is the Area Of Occupancy (taking into account all the world) of each species *i* present in this cell.
- *Patch index* for each cell, as $CPI = \frac{\Sigma_{i=1}^{n} P_i}{n}$ where *n* is the richness of the cell, and $P_i$ is the Patch distribution index (taking into account all the world) of each species *i* present in this cell. Patch index of each species has been previously calculated when processing maps. In subsection 6.23 you can find and explanation of patch index for each species is calculated.
- *AOO index* for each cell, as $CAI = \frac{\Sigma_{i=1}^{n} AOO_i}{n}$ where *n* is the richness of the cell, and $AOO_i$ is the Area Of Occupancy (taking into account all the world) of each species *i* present in this cell.
- *EOO index* for each cell, as $CAI = \frac{\Sigma_{i=1}^{n} EOO_i}{n}$ where *n* is the richness of the cell, and $EOO_i$ is the Extent Of Occurrence (taking into account all the world and using convex hull method) of each species *i* present in this cell.
- *Latitudinal range index:* for each cell, as $LRI = \frac{\Sigma_{i=1}^{n} \frac{1}{LR_i}}{n}$ where *n* is the richness of the cell, and $LR_i$ is the latitudinal range (that is, the higher latitude minus the lower latitude, taking into account all the world) of each species *i* present in this cell.
- *Habitats area:* the area in $Km^2$ for each kind of habitat in this cell. Take into account that this value is calculated using a raster world map of 1'x1' precision, making results only roughly approximate, particularly for rivers and small freshwater areas. Except for the *CSV list in columns* format, this option necessarily produce a multiple sheet or multiple file output, as the area occupied by each type of habitat in each cell will be outputted in a separate sheet or file.

- You can also set richness count method: count a species if it is just present in the output cell, or in turn set a threshold of 1'x1' minutes cells occupation that has to be exceeded to count a species in this cell. For example, if you set a cell output of 5'x5', you could want that only if a species occupies at least 20% of this cell (that is 20% of 5x5= 5, so 5 cells of 1'x1' minutes) the species has to be counted (so the value for this cell incremented).
- You can also select which habitats have to be included in metrics calculus. By default all habitats are took into account. If you exclude one habitat, all occurrences or zones with this associated habitat will be excluded from the metrics calculus.
- You can define the zone boundaries to be examined in degrees, with a minimum of a 1ºx1º zone. Take into account that rarity index, patch index and area index are always

calculated in a world basis, independent from the zone boundaries selected for the output.

- You can select how cell coordinates will be shown in output: using the coordinates of the center or one corner of the cell.
- You can also select any environmental variables you want to output, for the precision and boundaries, and with the same format that you selected. It must be pointed out that it is also possible to only select environmental variables to output, without including any metrics. To manage environmental variables see section 6.26.
- When exporting metrics (such as richness) in ESRI ASC format, using the "Export zero as NODATA" checkbox you can select if a zero value for a cell will be exported as NODATA (usually will not appear when showing the raster in a viewer) or as zero (will be shown as a zero value).

Finally, you can select the output format for presence/richness data, among the following ones:

- **CSV matrix (longitude X latitude):** this format outputs a matrix with coordinates of each cell at row (longitude)/column(latitude) headers, and metrics data as values for each cell of the matrix. Here we see an example of richness output for a zone with longitude from 109 to 120 and latitude from 0 to 11 and with a cell precision of 60'. Values of longitude and latitude correspond here to the right-down corner of each cell. The zero value in the first position of the first row has no signification; it is used to avoid importation problems with some applications like R. When exporting environmental data in this format, if a value is missing for a cell, an empty value will appear on the output.

| **0** | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 1 | 1 | 0 | 0 | 0 | 3 | 0 | 0 | 4 | 3 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 5 |
| 8 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 4 | 7 |
| 7 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 3 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 12 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 9 |
| 4 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 6 | 12 |
| 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 3 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 3 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |

As the CSV format does not support several pages, if you select this output format and several metrics to be outputted, you will be warned that one single CSV file for each selected metric will be generated.

- **CSV list in columns:** this format outputs a list with several columns: latitude and longitude for each cell, and a column for each selected metric. Here we see an example

where a zone with longitude from 110 to 120 and latitude from 0 to 2 with a precision of 30'.

| Latitude | Longitude | Species.Richness | Sea.Area | Land.Area | Small.Rivers. Area | Large.Rivers. Area | Lentic.Waters. Area |
|---|---|---|---|---|---|---|---|
| 2 | 110,5 | 1 | 3090,04356 | | | | |
| 1,5 | 110,5 | | 1806,41211 | 968,580384 | 315,99288 | | |
| 1 | 110,5 | 2 | | 2291,28846 | 800,403343 | | |
| 0,5 | 110,5 | 3 | | 2418,75789 | 673,404884 | | |
| 0 | 110,5 | 4 | | 2339,91575 | 752,482526 | | |

As this format requires loading all data to be outputted at the same time, if you select several variables it can require a long time. When exporting environmental data in this format, if a value is missing for a cell, an empty value will appear on the output.

- **Excel worksheet (2007 or higher):** this format will produce an XML file directly recognized and readable by Microsoft Excel 2007 or higher. This format has the advantage of being multipage, so one worksheet is generated for each metric to be outputted.
- **ESRI ASCII Raster matrix:** this format outputs the ESRI ASCII format, widely accepted by GIS software. Here we see an example where a zone with longitude from 120 to 125 and latitude from 0 to 2 with a precision of 10'.

```
ncols 30
nrows 12
xllcorner 120
yllcorner 0
cellsize 0.166666666666667
nodata_value -1
-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 4
-1 -1 -1 -1 -1 2 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 3 -1
-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 2 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
```

As the ESRI ASCII format does not support several pages, if you select this output format and several metrics to be outputted, you will be warned that one single ESRI ASCII file for each selected metric will be generated.

When exporting metrics (such as richness) in this format, using the "Export zero as NODATA" checkbox you can select if a zero value for a cell will be exported as NODATA (usually will not appear when showing the raster in a viewer) or as zero (will be shown as a zero value).

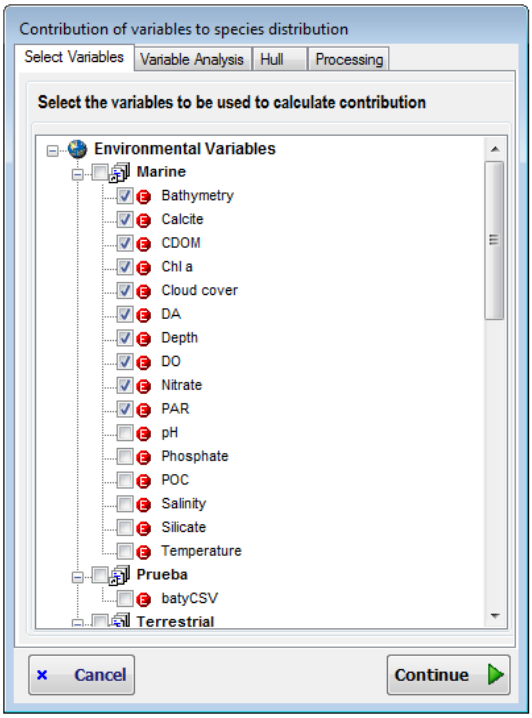
## 6.18 Exporting variables contribution report

The variables contribution report feature performs a statistical analysis for selecting those variables that are affecting the most the distribution of one or more species. This analysis will be performed for each species: the environmental variables selected are checked within a selected area (i.e. the EOO of the species, all the world, a river basin…), comparing the range of values that each variable takes where the species has a confirmed presence against the range it takes in the selected area (also called *background*). Then an index of instability is calculated

using the fluctuation index of Dubois (1973) modified by Guisande et al. (2006). Simply put, the higher the fluctuation of a variable range between the areas where the species has a confirmed presence and the background areas, the more likely this variable is affecting the distribution of the species. Once the instability index is calculated the percentage of contribution of each variable to the index is calculated. Those environmental variables with a higher percentage contribution to the instability index can be considered as those more affecting the most the distribution of the species.

This analysis is used as part of calculations such as environmental occurrence (see 2.10 and 3.11) and Niche of Occurrence using a per-species layer (see 3.13), to select the variables that affect the most the distribution of the species.

But you can also perform this analysis independently from Environmental Occurrence or Niche of Occurrence features: a report about variable contribution for any rank species can also be exported from DataManager, using the menu *Export/Export checked maps/To variables contribution report.* This same option is also available going to *Export/Export checked maps/To RWizard applications/To SPEDInstabR*, because specific R and RWizard (http://www.ipez.es/RWizard/) packages have been developed to use report data generated by this feature.

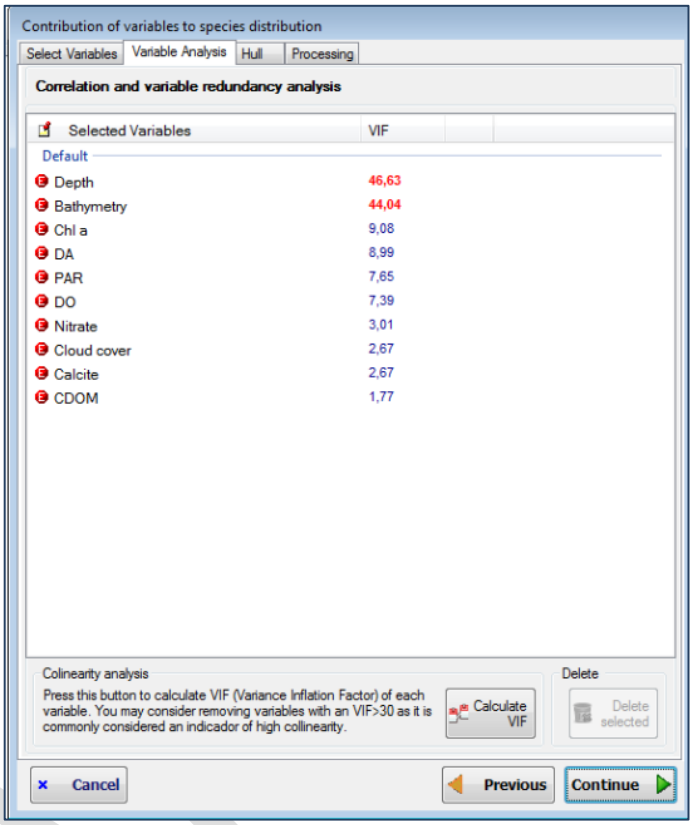



It should be signalled that even if this calculation can be performed with any kind of maps, it is clearly aimed to occurrence maps, not to area maps.

First of all you will be asked to select the folder where the report files will be saved. Then a step-by-step window will allow you to set the options to be used to calculate variable contribution.

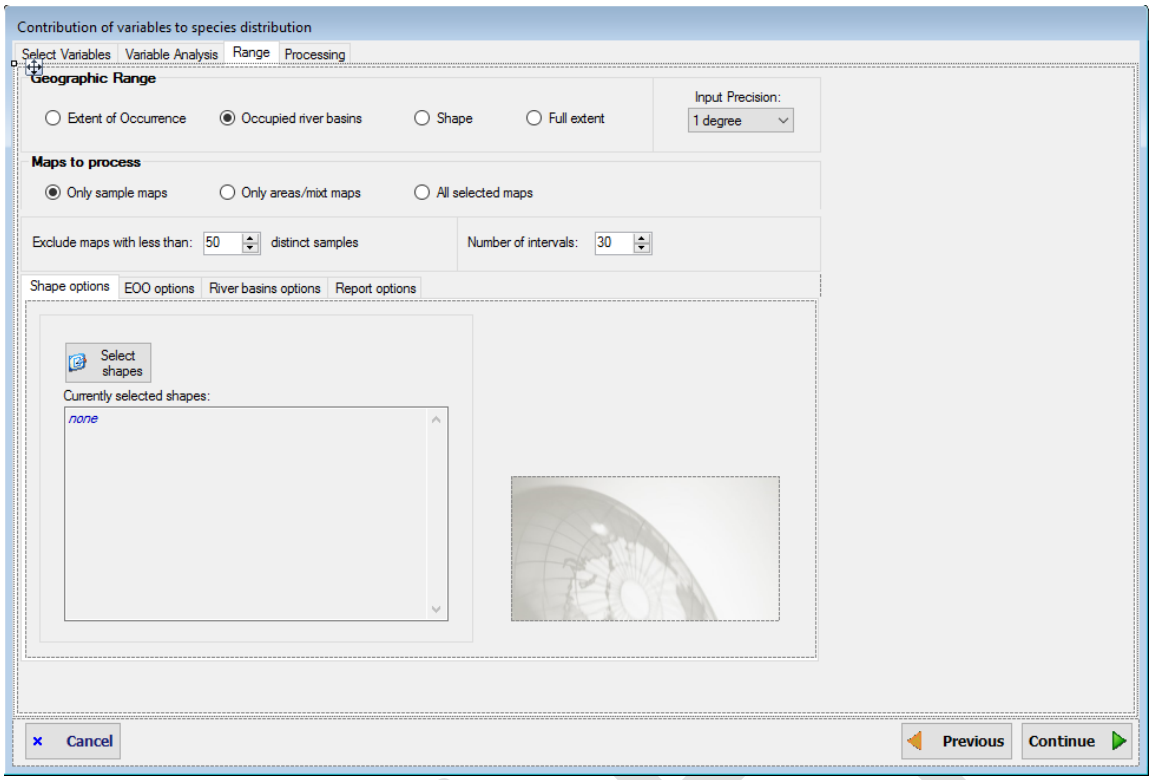
In the first step you have to select the variables for which you want to calculate their contribution index. You have to select at least 2 variables, but quite more variables will be usually selected.

Next step allows you to optionally calculate Variance Inflation Factor (VIF). A VIF value higher than 30 is commonly considered an indicator of collinearity among the variables. Thus, you have the change choice of removing those variables significantly related and, therefore, that are redundant. But this is an optional step. If you just want to include all variables, you can simply continue. Take into account that if two variables are very highly correlated, the way that one of them if a linear function of the other one, VIF cannot be calculated. Involved variables will be marked to allow you to remove one of them if you want.




Click on *Continue* button to go to next step.

In the last step, you have to select many options that determine how variable contribution will be calculated:




- **Geographic range**: this setting determines the spatial background to be used to calculate the variable contribution. Only the presences within this area will be used. The options are:
    - Extent of Occurrence: the EOO of the species will be used as background area. When selecting this option, specific options will appear in the bottom options panel that allow to select how the EOO will be calculated (alpha shape, convex hull…. see e.g. 2.8).
    - Occupied river basins: only the river basins where the species is present will be used as background area. When selecting this option, specific options will appear in the bottom options panel to select the river basin level to be used (from 0: largest basins; to 5: smallest basins). See section 2.16.2 for examples in MapMaker of how this option works.
    - Shape: one or more specific shapes will be used to delimit the area that will be used as background area. When selecting this option, specific options will appear in the bottom options panel that allow to select the wanted shape(s). Only the presences of a species within the selected shapes will be taken into account.
    - Full extent: all the available data of each variable are used. Typically, world-wide variables will be used. Therefore, the background will be the area where there are values for a variable.
- **Input precision:** cell size to be used to calculate contribution. This is usually the cell size of the variables to be used. When selecting full extent it automatically changes to 1 degree, to avoid generating a heavy load task. But the user can set the wanted precision.
- **Maps to process:** you have to select which maps, from those checked in the tree, will be processed: only occurrence maps, only areas (range) maps, or all of them. As said

before, contribution calculation is intended to be used with occurrence maps, but you can eventually use it with area maps.

- You can also select to **exclude maps** with less than a specific number of occurrences (also called occurrences). This can be useful to avoid calculating contribution for species with very few presence data.
- You can select the **number of intervals** to be used to calculate contribution.
- **Report options tab:** besides CSV files where contribution data are export you can select to export rasters in ASC format with data for each species presence (see 6.15) and a raster with the mean contribution by cell for each variable, with the selected precision (cell size).

Click on *Continue* button to start analysis.

Once analysis completed, results are exported to several CSV files:

**ContribShortReport.CSV**

The columns of this file are:

- Class, order, family, genus and species for each analysed species.
- For each one of the variables included in the analysis, there are the following columns:
    - Contrib.Percent: the relative % of contribution of this variable to the presence of the species. The sum of the Contrib.Percent of all variables should be 100%.
    - Min.Presence: the minimum detected value of this variable where the species is present (according to the used distribution data).
    - Max.Presence: the maximum detected value of this variable where the species is present (according to the used distribution data).
    - Mean.Presence: the mean value of this variable where the species is present (according to the used distribution data).
    - Min.Extent: the minimum detected value of this variable in the considered background.
    - Max. Extent: the maximum detected value of this variable in the considered background.
    - Mean. Extent: the mean detected value of this variable in the considered background.

**Contribution.CSV**

The columns of this file are:

- Class, order, family, genus and species for each analysed species.
- A column for each one of the variables included in the analysis with the relative % of contribution of this variable to the presence/absence of the species. Therefore the sum of all contributions should be 100% for each species.

**Instability.CSV**

This file contains the intermediate results obtained in order to calculate variable contribution. The columns of this file are:

- Class, order, family, genus and species for each analysed species.
- The name of the variable.
- The number of the interval. The will be as many interval rows as selected by the user to perform calculation (30 by default).
- The upper limit of the interval.
- The number of presence values within this interval.
- The number of extent values within this interval.
- The number of presence values within this interval in percentage.
- The number of extent values within this interval in percentage.
- The instability index of this interval.
- The total instability index for this variable. Therefore it will be the same for all the intervals of the variable.
- The contribution % for this variable. Therefore it will be the same for all the intervals of the variable.

Besides those CSV files that are always exported, if you selected to mean contributions rasters, a folder called *Contributions* will be created, containing an ASC raster file for each variable, with mean contribution value for each cell within the selected background. A single raster with total richness for the considered species will alse be saved into this folder.

If you selected to export presence rasters for each species, a folder called *Species* will be created, containing an ASC raster file for each species.

## 6.19 Exporting clipped rasters

In section 2.9.4 we described the raster clipping feature in MapMaker. This feature is useful when having to cut off some rasters using some single pattern. But sometimes you may want to cut off rasters using as pattern the specific EOO of a set of species. It should be too tedious having to do it species by species. Therefore, you can use DataManager raster clipping feature to do it in a single operation.

To do that, the raster files have to be previously integrated in ModestR, typically as environmental vatiables. Details about how to integrate environmental data in ModestR are explained in section 6.26.

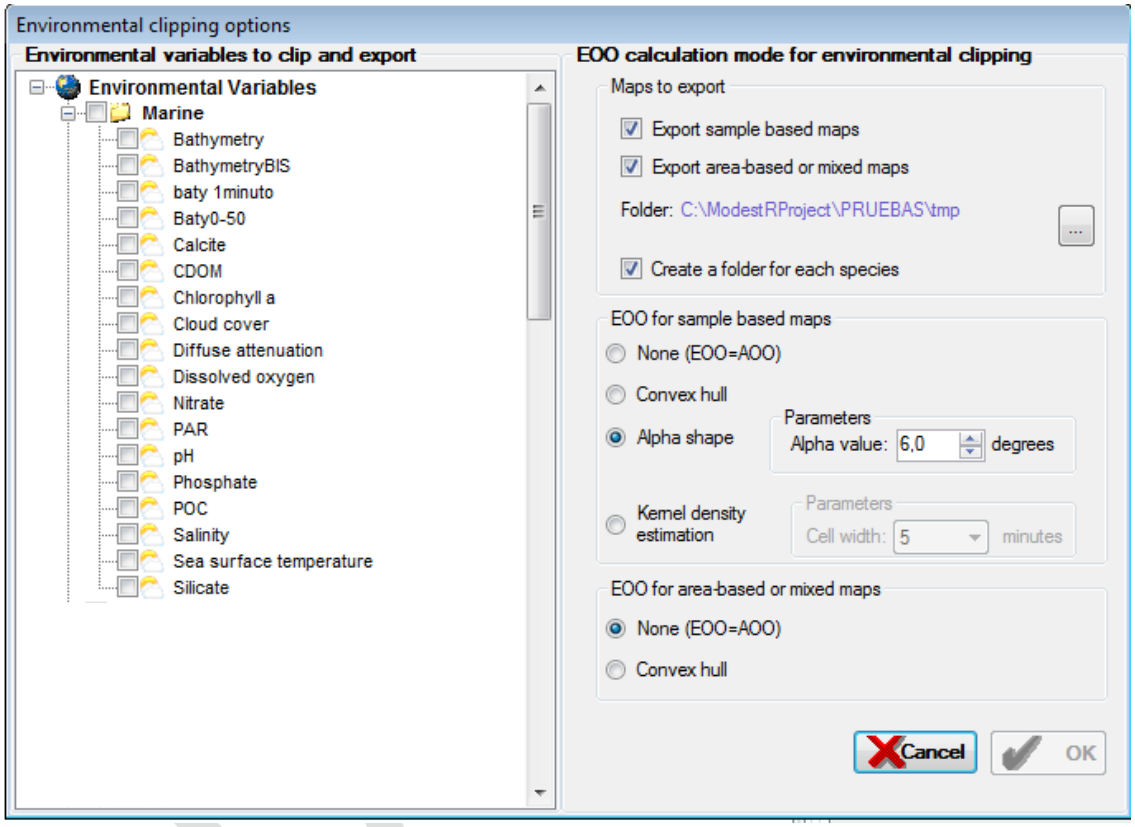
First select the wanted group (order, family…) of species to use as patterns in the DataManager taxonomy tree. Then go to *Export/Export checked maps/To clipped environmental data* in the main menu. A dialog box will appear where you will have to select the destination folder. As existing files with the same name will be overwritten, we recommend to select an empty folder.

Next, a dialog box will appear where you have to set the exportation options:

- Select the variables (raster data) you want to export in the left tree.

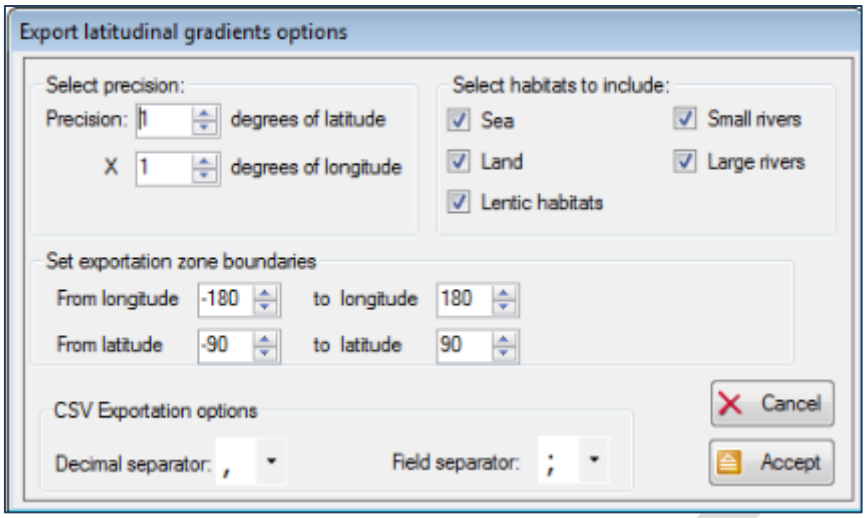
- Select which maps you want to export: only occurrence maps, only area (range maps), or both.
- Select if you want to create a subfolder where clipped raster files for each species will be stored. Those folders will be named as the corresponding species.
- Select how EOO will be calculated for each species. This EOO will be the pattern used to cut off selected rasters, for each species. More details about each EOO calculation mode can be found in section 3.6.

Once those settings set, click on the OK button and exportation will start. Take into account that this process can require quite some time, depending on the number of selected species.




## 6.20 Exporting latitudinal gradients

DataManager can calculate and export several latitudinal gradients for several metrics from maps. When selecting *Export/Export checked maps/To latitudinal gradients* in main menu, a dialog box to select exportation settings is shown:




In this dialog box you can set:

- Precision: the dimensions of the cells that will be used to calculate gradients in degrees, from 1º to 10º. Latitude precision indicates the height of the latitudinal bands for which gradients will be calculated. Longitude indicates the width of the cells that latitudinal bands will be split in, to calculate some of the metrics into the latitudinal bands. Other metrics will be calculated globally for all the latitudinal band, without taking into account longitudinal cells.
- You can also select which habitats have to be included in metrics calculus. By default all habitats are took into account. If you exclude one habitat, all occurrences or zones with this associated habitat will be excluded from the calculus.
- You can define the zone boundaries to be examined in degrees, with a minimum of a 1ºx1º zone.

Output will be generated as a CSV file with the following columns:

- **Latitude:** top limit of the latitudinal band.
- **Total.Species.Richness**: total richness of the latitudinal band This metric is calculated globally for all the latitudinal band, without taking into account longitudinal cells.
- **Mean.Species.Richness**: mean of the richness of the cells of the latitudinal band, taking cells of the longitude width in degrees indicated in the dialog box described above. That is, the richness of each cell of the longitude width and latitude height in degrees indicated is calculated for a latitudinal band, and then the mean of all those values is calculated. All cells are took into account to calculate the mean, even if they have a richness of zero.
- **SD.Species.Richness**: standard deviation of the richness of the cells of the latitudinal band, calculated in the same way than Mean.Species.Richness.
- **Mean.AOO.Index**: mean of the AOO (Area Of Occupancy) index metrics calculated for each one of the cells of the latitudinal band, taking cells of the longitude width in degrees indicated in the dialog box described above. That is, the AOO index of each cell of the longitude width and latitude height in degrees indicated is calculated for a latitudinal band, and then the mean of all those values is calculated. Be aware that only cells with a richness value bigger than zero (that is, where there at least one species

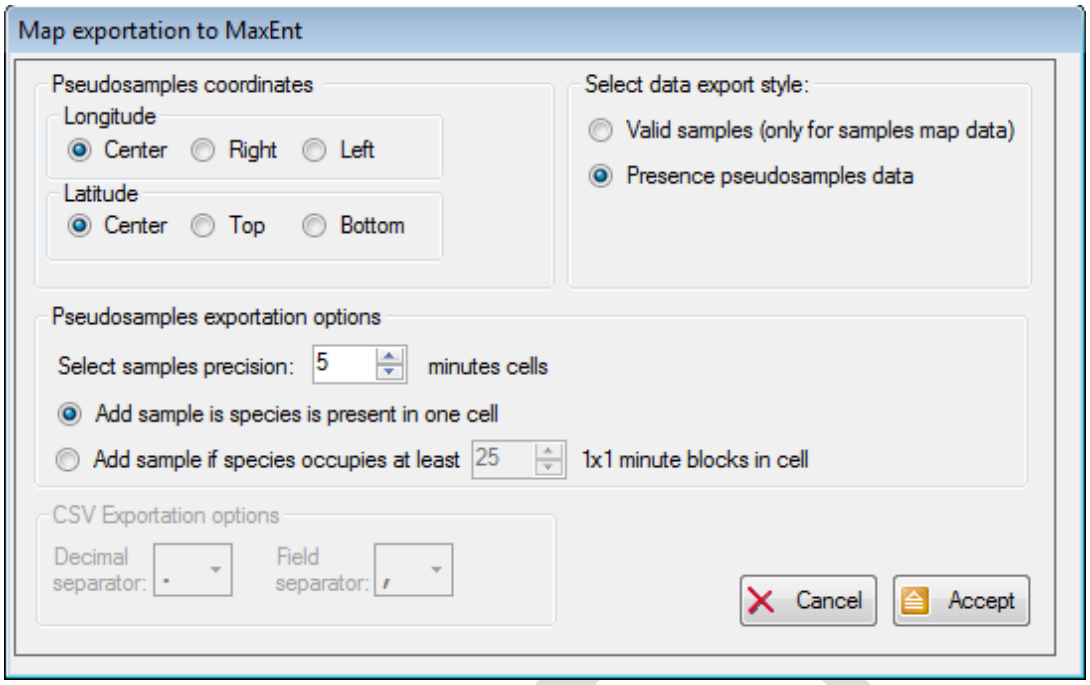
present) are took into account to calculate the mean. Details about AOO index metric can be found in section 6.17 above.

- **SD.AOO.Index**: standard deviation of the AOO index of the cells of the latitudinal band, calculated in the same way than Mean.AOO.Index.
- **Mean.EOO.Index**: mean of the EOO (Extent Of Occurence) index metrics calculated for each one of the cells of the latitudinal band, taking cells of the longitude width in degrees indicated in the dialog box described above. That is, the EOO index of each cell of the longitude width and latitude height in degrees indicated is calculated for a latitudinal band, and then the mean of all those values is calculated. Be aware that only cells with a richness value bigger than zero (that is, where there at least one species present) are took into account to calculate the mean. Details about EOO index metric can be found in section 6.17 above.
- **SD.EOO.Index**: standard deviation of the EOO index of the cells of the latitudinal band, calculated in the same way than Mean.EOO.Index.
- **Mean.Rarity.Index**: mean of the rarity index metrics calculated for each one of the cells of the latitudinal band, taking cells of the longitude width in degrees indicated in the dialog box described above. That is, the rarity index of each cell of the longitude width and latitude height in degrees indicated is calculated for a latitudinal band, and then the mean of all those values is calculated. Be aware that only cells with a richness value bigger than zero (that is, where there at least one species present) are took into account to calculate the mean. Details about rarity index metric can be found in section 6.17 above.
- **SD.Rarity.Index**: standard deviation of the area index of the cells of the latitudinal band, calculated in the same way than Mean.Rarity.Index.
- **Mean.Patch.Index**: mean of the patch index metrics calculated for each one of the cells of the latitudinal band, taking cells of the longitude width in degrees indicated in the dialog box described above. That is, the patch index of each cell of the longitude width and latitude height in degrees indicated is calculated for a latitudinal band, and then the mean of all those values is calculated. Be aware that only cells with a richness value bigger than zero (that is, where there at least one species present) are took into account to calculate the mean. Details about patch index metric can be found in section 6.17 above.
- **SD.Patch.Index**: standard deviation of the area index of the cells of the latitudinal band, calculated in the same way than Mean.Patch.Index.
- **Mean.Latitudinal.Range.Index**: mean of the latitudinal range index metrics calculated for each one of the cells of the latitudinal band, taking cells of the longitude width in degrees indicated in the dialog box described above. That is, the latitudinal range index of each cell of the longitude width and latitude height in degrees indicated is calculated for a latitudinal band, and then the mean of all those values is calculated. Be aware that only cells with a richness value bigger than zero (that is, where there at least one species present) are took into account to calculate the mean. Details about latitudinal range index metric can be found in section 6.17 above.
- **SD.Latitudinal.Range.Index**: standard deviation of the area index of the cells of the latitudinal band, calculated in the same way than Mean.Latitudinal.Range.Index.

- **Mean. AOO**: mean of the AOO of all species present in the latitudinal band (taking into account the AOO that each species occupies in all the world, independently from the range selected for the analysis). This metric is calculated globally for all the latitudinal band, without taking into account longitudinal cells.
- **SD.AOO**: standard deviation of the AOO of all species present in the latitudinal band (taking into account the AOO that each species occupies in all the world, independently from the range selected for the analysis). This metric is calculated globally for all the latitudinal band, without taking into account longitudinal cells.
- **Mean. EOO**: mean of the EOO of all species present in the latitudinal band (taking into account the EOO that each species occupies in all the world, independently from the range selected for the analysis). This metric is calculated globally for all the latitudinal band, without taking into account longitudinal cells.
- **SD.EOO**: standard deviation of the EOO of all species present in the latitudinal band (taking into account the EOO that each species occupies in all the world, independently from the range selected for the analysis). This metric is calculated globally for all the latitudinal band, without taking into account longitudinal cells.
- **Mean.Latitudinal.Range**: mean of the latitudinal range occupied by all species present in the latitudinal band (taking into account the area that each species occupies in all the world, independently from the range selected for the analysis). This metric is calculated globally for all the latitudinal band, without taking into account longitudinal cells.
- **SD.Latitudinal.Range**: standard deviation of the latitudinal range occupied by all species present in the latitudinal band (taking into account the area that each species occupies in all the world, independently from the range selected for the analysis). This metric is calculated globally for all the latitudinal band, without taking into account longitudinal cells.
- **Mean.Patch.Distrib**: mean of the patch value of all species present in the latitudinal band (taking into account the area that each species occupies in all the world, independently from the range selected for the analysis). In subsection 6.23 you can find and explanation of patch index for each species is calculated. This metric is calculated globally for all the latitudinal band, without taking into account longitudinal cells.
- **SD.Patch.Distrib**: standard deviation of the patch value of all species present in the latitudinal band (taking into account the area that each species occupies in all the world, independently from the range selected for the analysis). This metric is calculated globally for all the latitudinal band, without taking into account longitudinal cells.

## 6.21 Exporting species data to Maxent

DataManager can export species presence data in a format directly importable by Maxent. Maxent is able to import data in form of occurrences from one or more species. To export this data, select the wanted node in the tree, then go to *Export/Export checked maps/To Maxent* in main menu. A dialog box will be shown.




Options are quite similar to the ones explained in section 6.14. excepted for the absence of some options such as environmental variables or full taxonomy exportation, as those options are not suitable for Maxent. Once accepted, DataManager will generate and export a CSV file with selected species data. Consult the Maxent documentation for more details about how to import and use them in Maxent software.

## 6.22 Exporting species data to RWizard applications

DataManager can export species presence data in a format specifically designed to be imported by particular RWizard applications. RWizard (http://www.ipez.es/RWizard ) is a GUI interface for R statistical software, which can integrate a type of extensions called RWizard *Applications*.

Several of those RWizard *Applications* have already been developed to specifically use data exported form ModestR. To export data from DataManager to those RWizard applications for ModestR, select the wanted node in the tree, then go to *Export/Export checked maps/To RWizard applications* in main menu. There are currently implemented options to export data to KnowBR, MapsR, FactorsR, EnvNicheR and SPEDInstabR. Depending on the option you select, a dialog box with the required options will be shown. Depending on the case, the options are quite similar to the ones already explained to export metrics or presence data (see 6.14 and 6.17).

## 6.23 Exporting summary report

DataManager can calculate and export several metrics from maps, like richness data, rarity index, etc. When selecting *Export/Export checked maps/To summary report* in main menu, a dialog box to select output file is shown. Then a CSV file is generated containing the following columns for each select species:

- Taxonomy: *Class, Order, Family, Genus, Species*
- Total Area Of Occupancy of the species: *Total.AOO*
- Total area for each habitat type occupied by the species: *Sea.Area, Land.Area, Large.Rivers.Area, Small.Rivers.Area, Lentic.waters.Area, etc.*

- Total Extent Of Occurrence of the species: *Total.EOO*
- Minimum and maximum coordinates where the species is present : *Min.Longitude, Max.Longitude, Min.Latitude, Max.Latitude*
- Patch index [24]: *Patch.Index*. This index is calculated as:

  $$Patch.Index = total\ EOO\ of\ species\ /\ total\ convex\ hull\ area\ of\ species$$

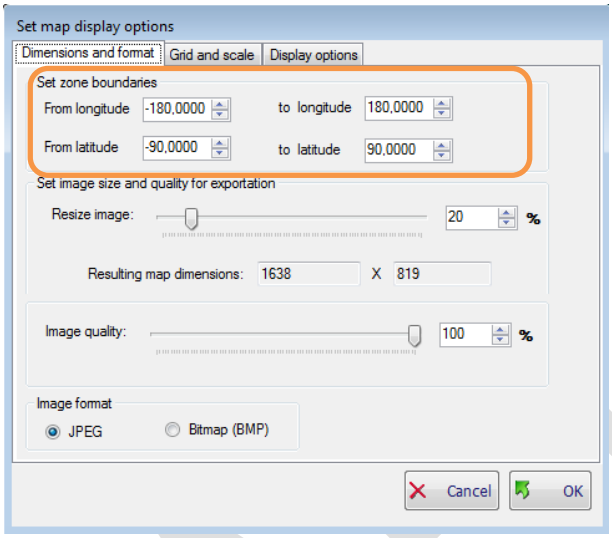
  where *EOO* is the Extent of Occurrence, calculated according to user settings (see 3.6); and *convex hull area* is the area of the convex hull calculated from all the presence data for the species (both occurrences and areas), but taking into account only habitats valid for the species. For example, is a species is only present in marine habitats, only sea areas will be took into account to calculate EOO and the convex hull area of the species.
- EOO.Mode: indicates the way the EOO has been calculated for this map. EOO mode is selected when processing maps (see section 3.6) and can be: none (EOO=AOO), ConvexHull, ConcaveHull (alpha shape) or DensityEstimation (kernel density).
- Alpha.Value: when the EOO mode is alpha shape, this value indicates the value of the alpha parameter used to calculate the EOO.
- ConvexHull.Area: area of the convex hull of the map. This value is independent from the chosen EOO mode, and it is used to calculate the patch index value, as explained above.
- HabitatAutoValidatedOccurrences: number of occurrences autovalidated by ModestR as being in a valid habitat for the species.
- ManuallyValidatedOccurrences : number of occurrences manually validated by the user (they are considered valid in any calculation independently of being placed in a valid or an invalid habitat for the species).
- TotalValidOccurrences: sum of HabitatAutoValidatedOccurrences and ManuallyValidatedOccurrences. This is the total of occurrences considered valid and took into account in any calculation about the species distribution.
- HabitatAutoInvalidatedOccurrences: number of occurrences autoinvalidated by ModestR as being in an invalid habitat for the species.
- ManuallyInvalidatedOccurrences: number of occurrences manually invalidated by the user (they are considered invalid, so not took into account in any calculation independently of being placed in a valid or an invalid habitat for the species). For example occurrences classified as invalid in data cleaning processes are considered manually invalidated samples.
- TotalInvalidOccurrences: sum of HabitatAutoInvalidatedOccurrences and ManuallyInvalidatedOccurrences. This is the total of occurrences considered invalid and not took into account in any calculation about the species distribution.
- TotalOccurrences: sum of TotalValidOccurrences and TotalInvalidOccurrences.
- HasAreas: set to 1 if the map contains presence areas (occurrences are not took into account, only range areas); zero otherwise.

Remember that characters used as field delimiter and decimal separator can be selected in the preferences (Options/Preferences menu).

[24] It is important to point out that patch index calculation has been modified after version 1.3.2 of ModestR. The current calculation used is the one described here.

## 6.24 Exporting JPEG/BMP images

DataManager can generate and export JPEG/BMP images from maps. When selecting *Export/Export checked maps/To JPEG/BMP image* in main menu, a dialog box to select the output folder will be shown. Next, an exportation settings dialog with several tabs is shown. The exported image is by default a full world map view. If you want to export a specific zoomed view, you should set the wanted boundaries in the first tab (*Dimensions and format)*.

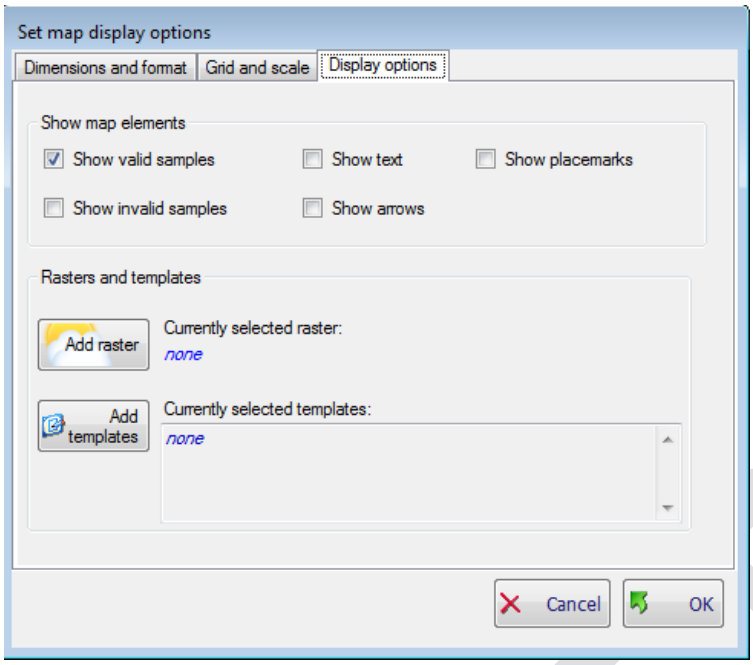



You can also modify resulting image size and quality. By default the image size is set to be slightly bigger than your screen resolution, which may be enough to use the image in a presentation, a document, etc. If you need more resolution, for example to allow zooming on the image, you can increment this size.

In the next tab, *Grid and scale*, you can add a grid and a numeric scale to the map.

In the last tab, *Display* options, you can select which elements from valid/invalid occurrences, text, arrows or placemarks will be shown on the exported map. Presence areas will be always shown. You can also add a raster to the exported maps. Rasters will typically correspond to environmental data. To add a raster, you have to have previously integrated it in ModestR, as described in section 6.26. You can find more information about raster display settings in section 2.9.1.

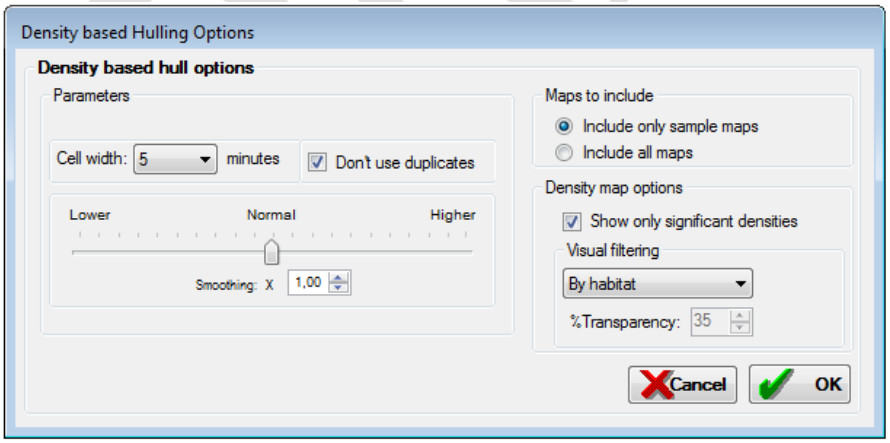
You can also add shapes to the maps to be exported. You can find more information about shapes in section 2.16 and 4.5.




Once set and accepted those options exportation will start, showing a dialog box where each exported map is shown. As this requires loading all the world map data and recreating each species map, it can take a long time to be completed, depending on the number of maps to be exported.

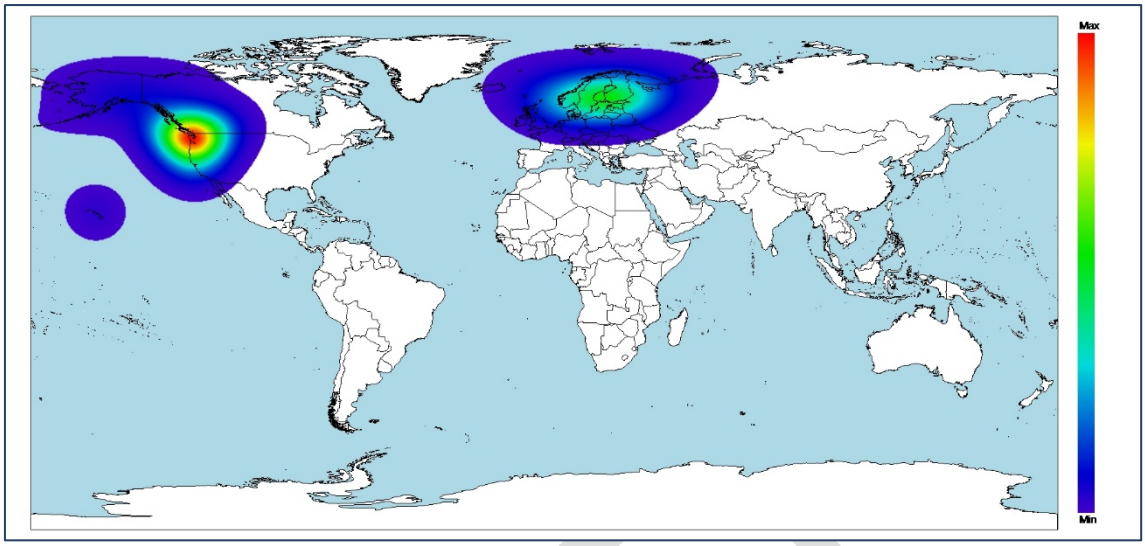
## 6.25 Exporting JPEG/BMP density map images

DataManager can generate and export JPEG/BMP images from density maps calculated for each species. An example of those maps is explained in 2.9.5 . When selecting *Export/Export checked maps/To JPEG/BMP density maps* in main menu, a first dialog box to select kernel density calculation setting. The settings here are quite the same than explained in 2.9.5.




Then another dialog box to select the output folder will be shown. Next, an exportation settings dialog with several tabs is shown, that is the same than explained in 6.24, excepted for raster and shapes settings, that are not available here. The exported image is by default a full world map view. If you want to export a specific zoomed view, you should set the wanted boundaries in the first tab (*Dimensions and format)*. Remember that in the last tab, *Display* options, you can select which elements from valid/invalid occurrences, text, arrows or placemarks will be shown

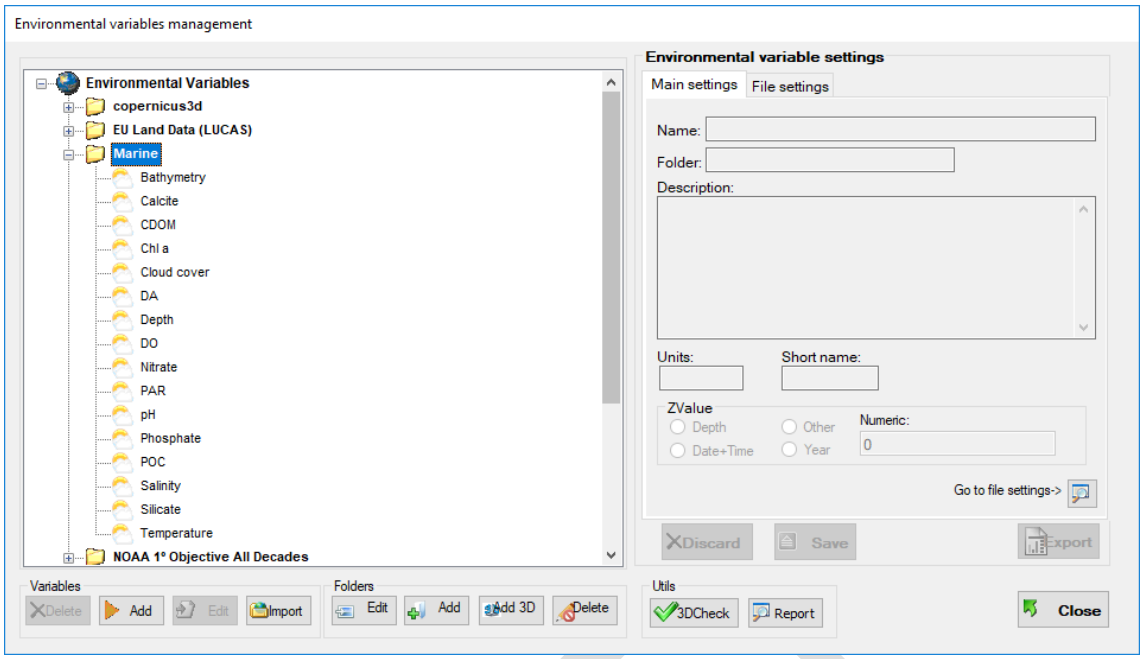
on the exported map. Below there is an example of a map exported using this feature, where occurrences have been hidden to only show density information.




## 6.26 Integrating and managing environmental variables

ModestR allows adding environmental variables, coming as data in CSV, GeoTiff, netCDF or ESRI ASC files (i.e. Bi-Oracle) to metrics exportation options (see section 6.17). This option is needed to obtain integrated outputs where metrics and environmental variables for a specific zone are included, to perform data cleaning using environmental data, to calculate niche of occurrence, to show environmental data in a map, etc.

To Add/modify/delete environmental variables you can use the *Layers/Manage environmental variables* menu in DataManager, or also in MapMaker (the same variables will be available from any ModestR application, no matter if you added them from MapMaker or DataManager). Then a dialog box will be shown. In this dialog box a tree shows the currently configured variables, ordered using their assigned folder. On the bottom-left side, you can see two groups of buttons, one for variables, and the other for folders.




If you select a variable in the tree, its settings will be shown in the right panel, in read-only mode. You can delete a variable by selecting it in the tree and using *Delete* button for variables, or the equivalent option of the contextual menu (clicking the right-mouse button on the tree). You can edit a variable by selecting it in the tree and using *Edit* button or the equivalent option of the contextual menu. The settings of the variable will be shown in the right panel, and the read-write mode will be activated to allow modifications. You can select to save or to discard changes using the corresponding buttons on the bottom.

To manually add a variable you can use the *Add* button for variables, or select an existing variable in the tree and select *Add Var using this as template* option in the contextual menu (pressing the left-mouse button).

The fields you can set for a variable are:

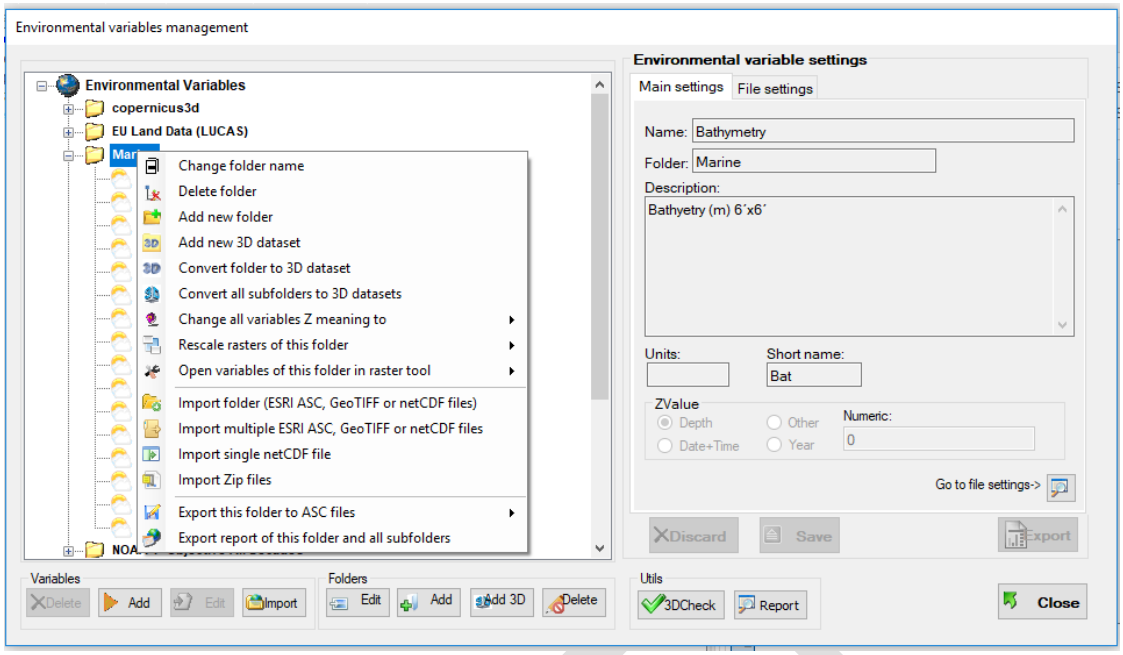
- Name: a brief descriptive name of the variable. This name will be shown in the tree.
- Folder: the folder that contains the variable. This folder will appear in the tree, with the variable hanging from it. Use folders to classify and group variables by some criteria that can make easier to find them.
- Description: Short description of the variable. This description will appear when selecting the variables in the tree. Enter a description that makes easy to understand what the variable is.
- Short Name: a short (3-4 characters) to name the variable. This name will be used in MapMaker when adding an environmental variable to be shown at the mouse cursor on the map.
- Units: this optional field allows entering the measuring units for the variable (i.e. Km., ºC, mol/l., etc…).This will be used in MapMaker when adding an environmental variable to be shown at the mouse cursor on the map.
- File: the file associated to the variable. This file has to contain data for the variable, usually for all the world. Use the *Select* button to select a file.

- Format: select the format of the variable data: CSV or ESRI ASC.
- If you select CSV settings, the variable is intended to be stored in a CSV in matrix form, where rows correspond to longitudes, and columns to latitudes. Data have to be arranged in the way that top left corner corresponds to the maximum latitude (typically 90º) and the minimum longitude (typically -180º). For this format you have to indicate the following fields:
    - Cell size in minutes: each element of the CSV matrix stores the value of the variable for a square cell of a certain size in minutes. Indicate the minutes in this field.
    - No Data value: indicate in this field the value used to indicate that there is no value for the variable for a certain cell. It is typically -9999.
    - Top corner longitude and latitude: indicates the coordinates that correspond to the first value of the file. Usually there are -180º longitude and 90º latitude.
    - First row contains headers/ First column contains headers: check/uncheck those options to set if the CSV file contains a first row/column of headers, that has to be skipped when reading variable data.
    - Decimal separator: indicates the character used as decimal part separator in the CSV file. It can be point or comma.
    - Field separator: indicates the character used as field part separator in the CSV file. It can be comma or semicolon.

If you want to add several variables which data is contained in several files in a same folder and that have the same format, you can use the *Import* button. In this case you will only have to select the kind of format (netCDF, GeoTiff or ASC) and select all the files to import or even a whole folder. The variable names will be set by default to the filenames in case of ASC files, and to the contained variables in case of netCDF files. You can later edit those variables to modify or complete the remaining fields.

But probably the easier way to import variables is to directly drag either single files or a folder and drop them on folder of the tree where you want to add them. To change a variable from one folder to another, you can drag and drop the variable over the target folder on the tree.

You can also easily add, edit or delete folders using the buttons for folders. To edit or delete a folder you must select it first on the tree. Take into account that deleting a folder, all variables within this folder will also be deleted. Many options are also available in context menus, just clicking on a variable or folder with the right button.

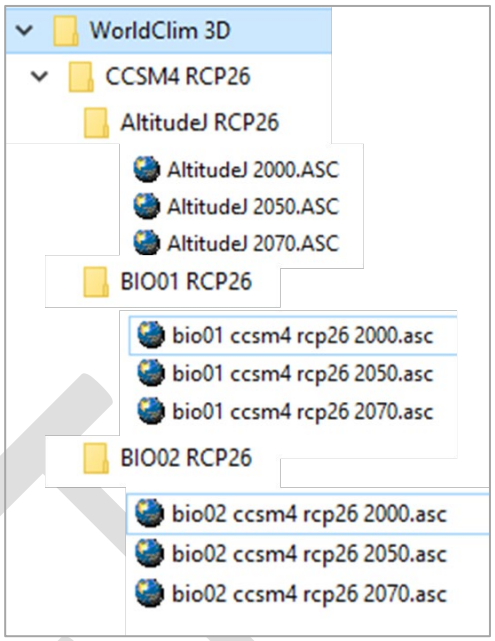



## 6.27 Integrating 3D environmental data in ModestR

Besides the usual 2D rasters (see previous section), ModestR allows integrating 3D environmental data, which can be used for example to calculate Niche Of Occurrence in 3D. A 3D environmental variable is constituted by several 2D rasters or layers, that contain the values of a same variable, for a same geographic space, and where each 2D raster corresponds to a distinct value along a third (Z) dimension. For example, several rasters (layers) representing temperature for different years (in this case Z=time); or rasters of sea salinity for different depths (in this case Z=depth).

A collection of those 3D environmental variables is needed to perform NOO3D. This can be, for example, a time series of climatic variables forecasting a scenario for the near-future, as available in WorldClim; o a set of marine environmental variables for different depths, as available in the World Ocean Atlas or Copernicus. In the first case the third, or Z dimension is the time, whereas in the last one is the depth. ModestR is agnostic regarding the meaning of the Z dimension, even if it provides default options fitted to the most usual ones such as year, date&time (a value that indicates a precise date and a time), or depth.

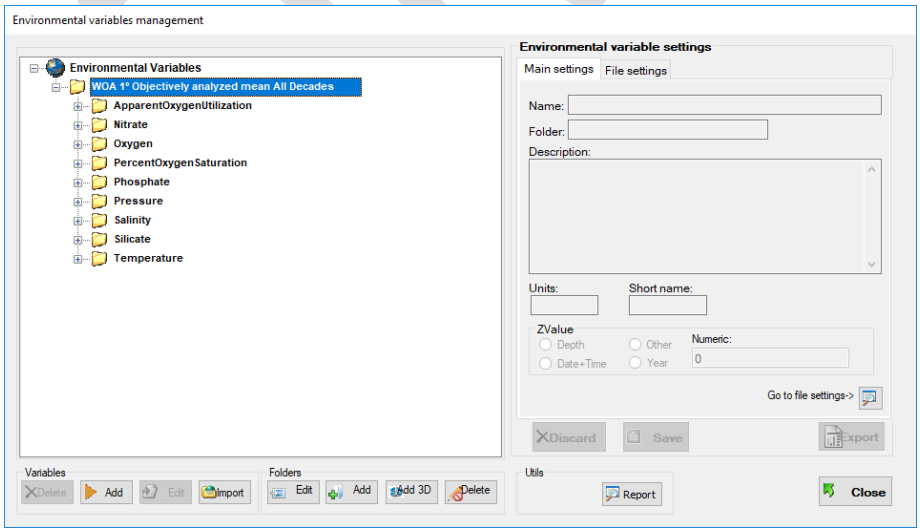
An easy way to integrate those 3D variables in ModestR is explained below:




- We assume here that you start from a collection of several raster files in .ASC or .GeoTIFF format[25]. Before importing them, we have to organize the raster files grouping them by variable in distinct folders. Thus, we'll have in each folder the rasters corresponding to the same variable, but for different Z values (i.e. for different times, or for different depths). An example can be seen in this figure, where 3 variables for different years, AltitudeJ, BIO1 and BIO2 have been grouped.

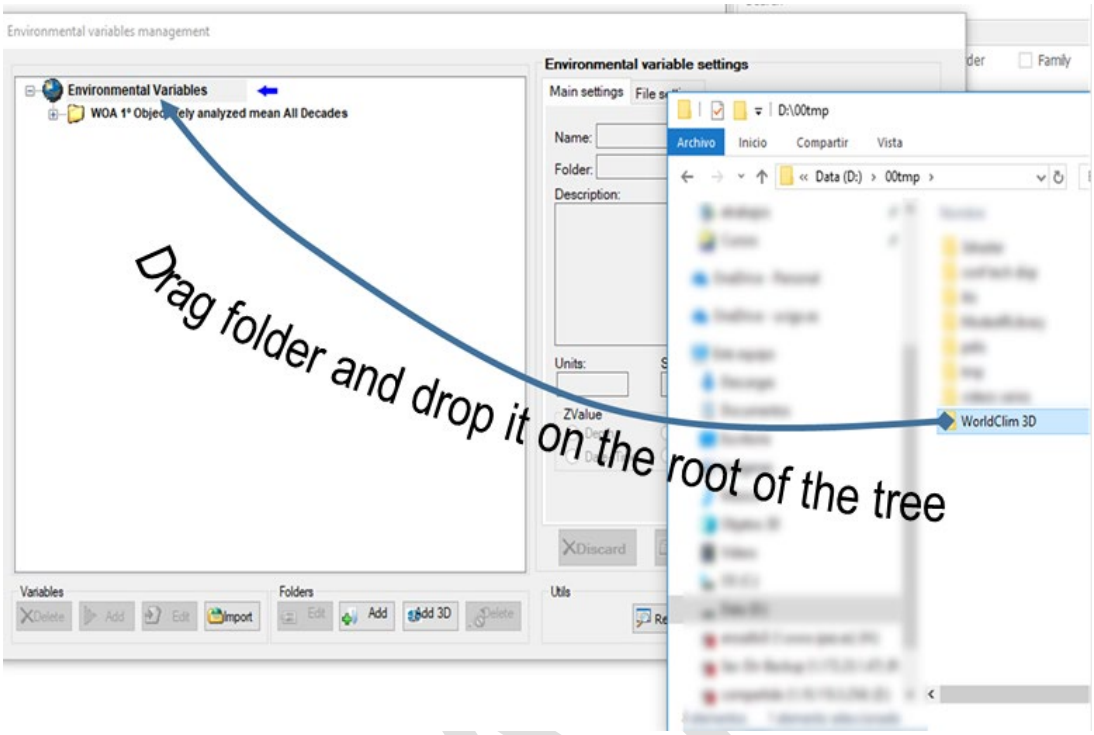
- Once the variables organized in this way, you can also avoid later tedious tasks by already adding to each file name its corresponding Z value. In the previous figure we can see that this value has been added in brackets to the end of each file name. This will prevent you from having to manually enter this value for each variable once imported in ModestR.

- Open DataManager (one of the ModestR applications) and go to *Layers/Manage environmental variables*. The variable manager will appear, showing on the left a tree where variables are organized in folders (tree will be empty the first time you access it, of course).

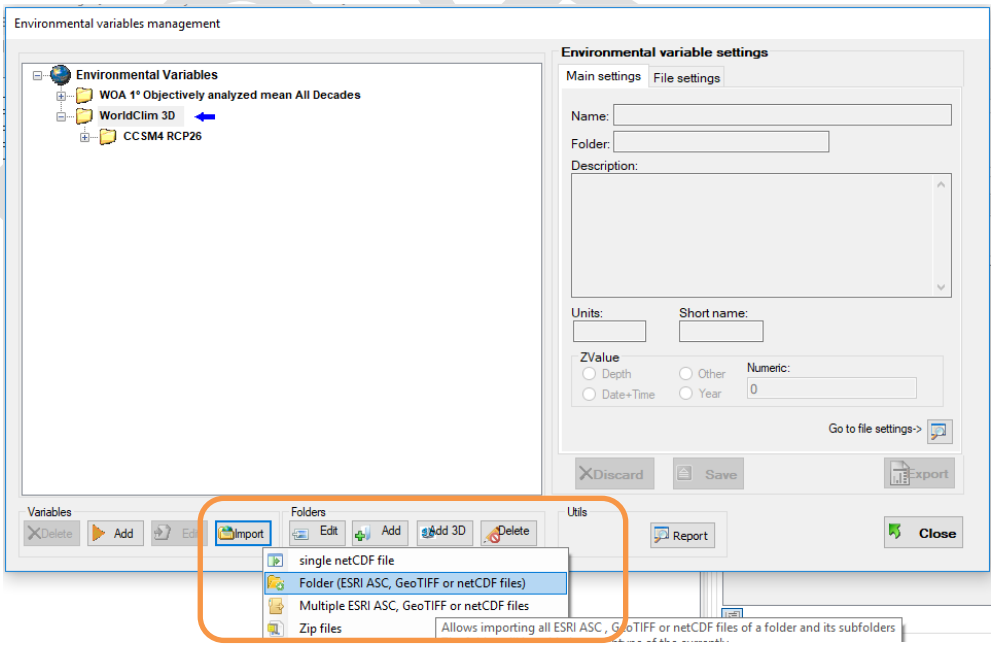



[25] Those are two formats directly supported by ModestR. netCDF format is also supported, but it can be less easy to integrate 3D variables in ModestR.

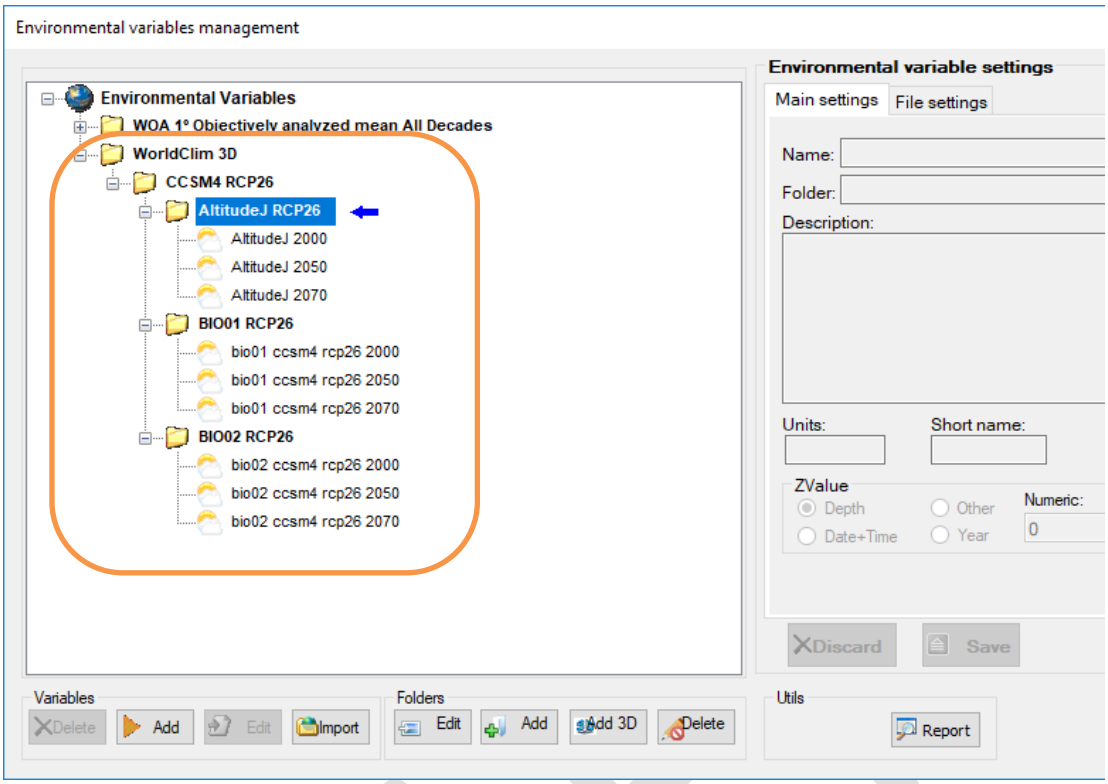
- The easiest way to import environmental data to ModestR it's just to drag&drop the folder where we previously organized the raster files to the variables tree:



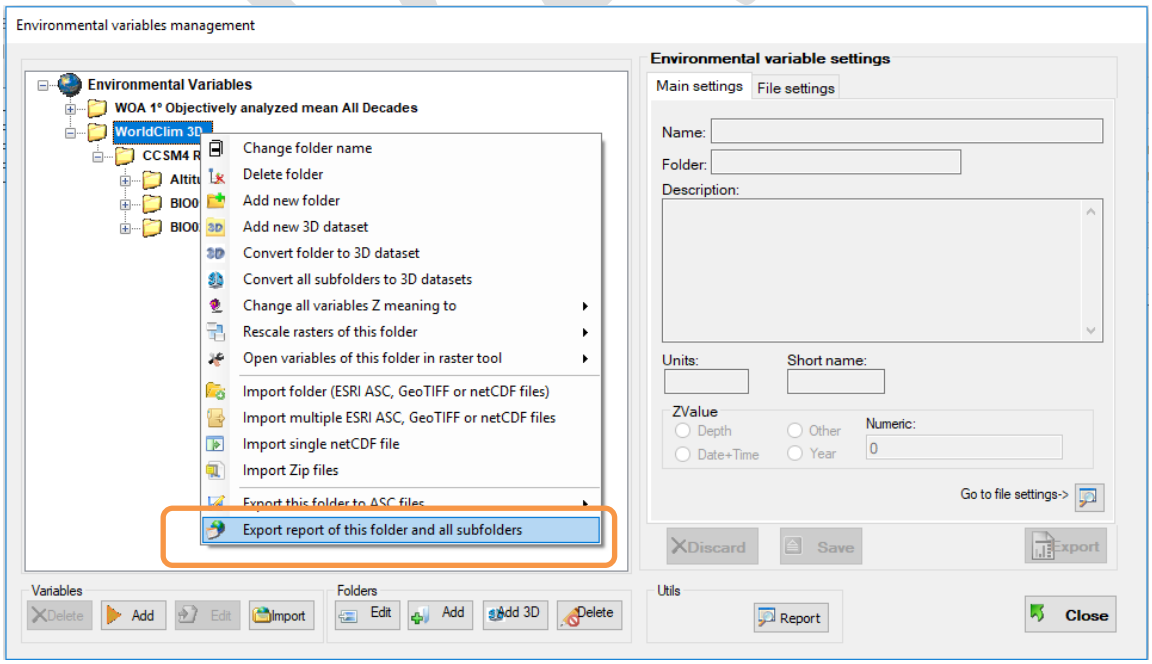

Alternatively the option Import/Folder can also be used. Selecting the main folder where variables are, all subfolders and the contained variables will be imported.




- Once the importation task ended, the variables are imported to ModestR, organized in the same way than in the imported folder.




- The next step is checking and adjusting if needed the Z value and the Z meaning of each one of the imported variables. An easy way to check those parameters is clicking with the right button on the folder that contains them and select *Export report of this folder and all subfolders.*

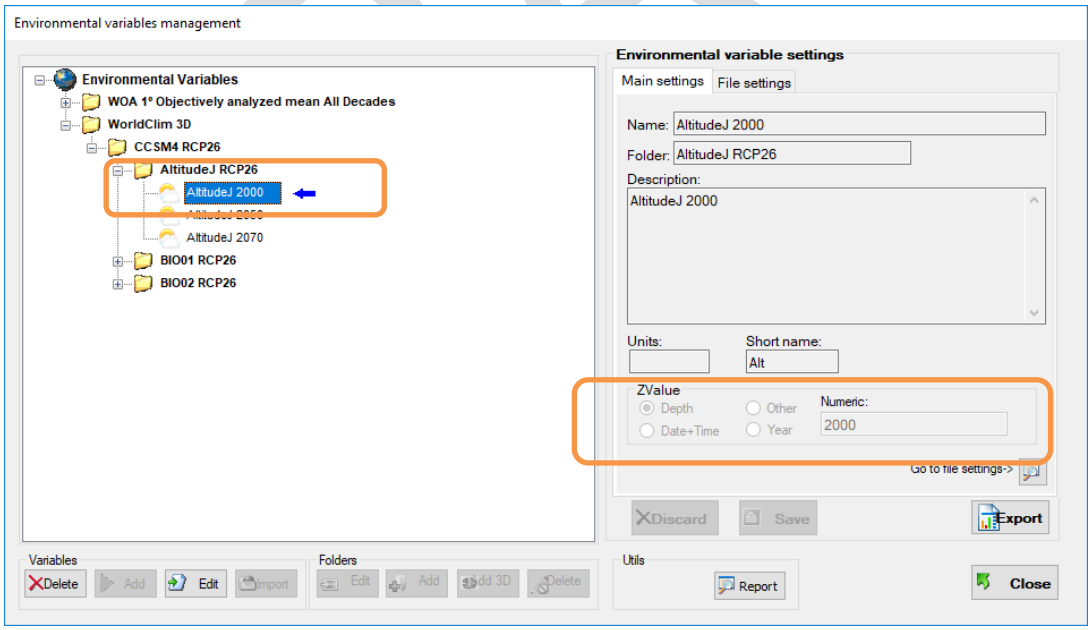


This will export a CSV file with several data about each variable. Among those data are the Z meaning and the Z value of each one. In this figure is an example of some columns of this file. As it is shown, the Z value assigned to each variable if correct, because ModestR inferred it from the filename, as explained before.

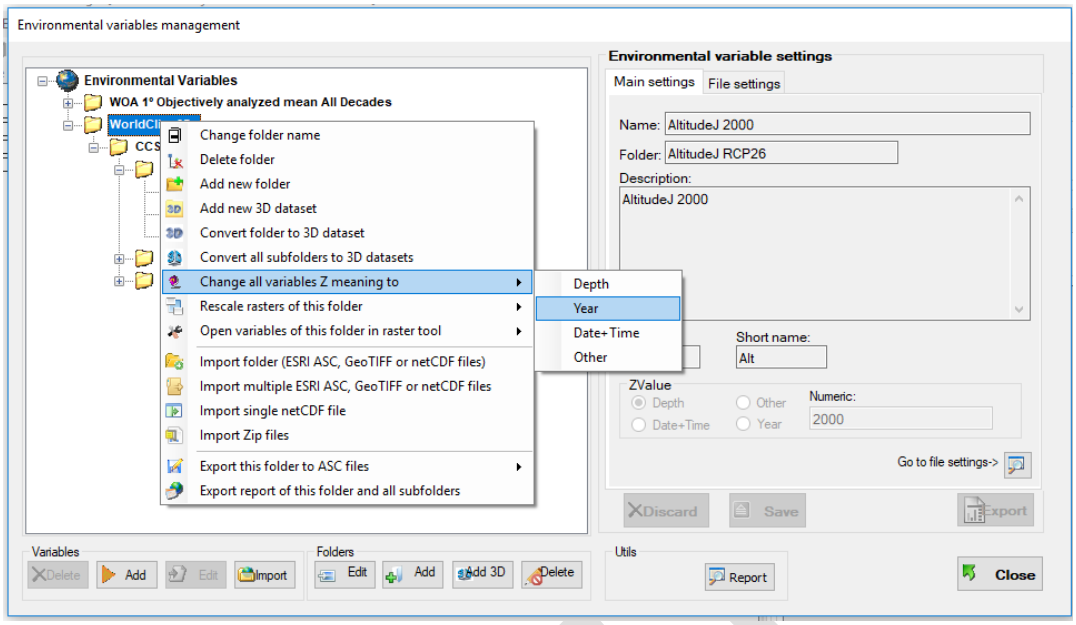
| Name | ZMeaning | ZValue |
|---|---|---|
| AltitudeJ 2000 | Depth | 2000 |
| AltitudeJ 2050 | Depth | 2050 |
| AltitudeJ 2070 | Depth | 2070 |
| bio01 ccsm4 rcp26 2000 | Depth | 2000 |
| bio01 ccsm4 rcp26 2050 | Depth | 2050 |
| bio01 ccsm4 rcp26 2070 | Depth | 2070 |
| bio02 ccsm4 rcp26 2000 | Depth | 2000 |
| bio02 ccsm4 rcp26 2050 | Depth | 2050 |
| bio02 ccsm4 rcp26 2070 | Depth | 2070 |

It is clear in this example that the Z meaning is not correct, because in this example Z represents time, not depth as it appears on the exported report.

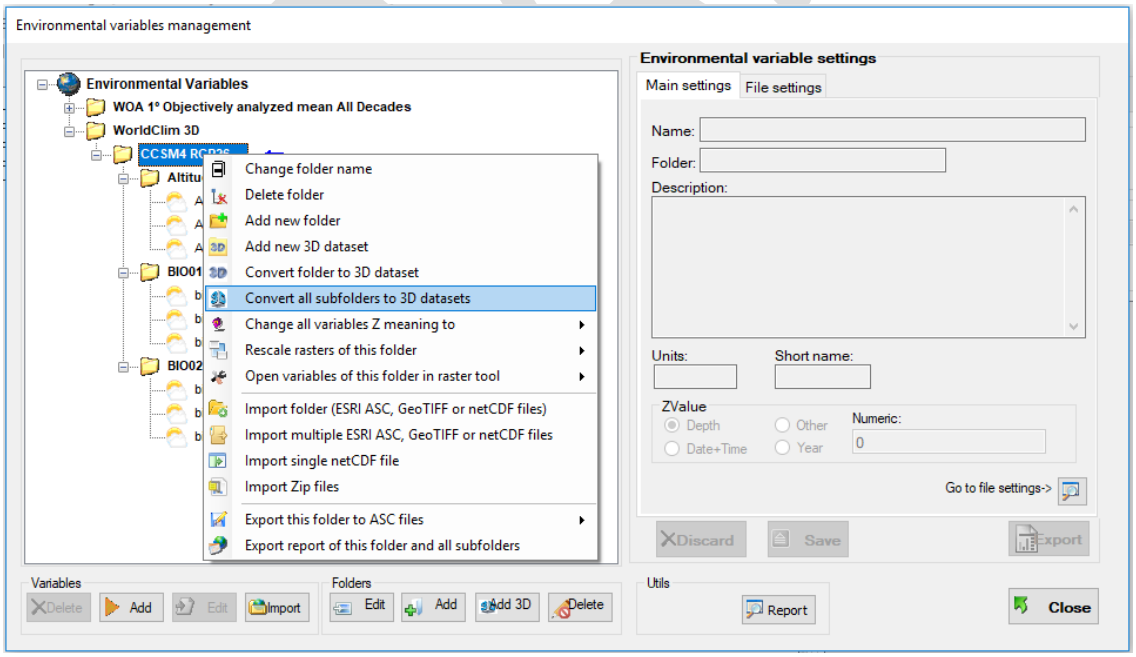
This can also be checked for an individual variable selecting it on the tree and seeing its parameters on the right panel:




- To change the Z meaning we have to select the folder where the variables are (in this example we can select the folder “WorldClim 3D” to change all variables at once). Then click with the right button and select *Change all variables Z meaning to/Year* (other meanings are available, as you can see).




- Finally, we have to indicate to ModestR that those variables are not independent from each other but composing several 3D variables. For example, we want the three BIO1 variables, one for each year, to be considered as a 3D variable. This can be done "transforming" each folder that contains the related variables in a "3D dataset".
  To do that we can select the folder containing the candidates to be transformed (in this illustrated example it'll be the folder called "CCSM4 RCP26"). Then click with the right button and select *Convert all subfolders to 3D datasets*.



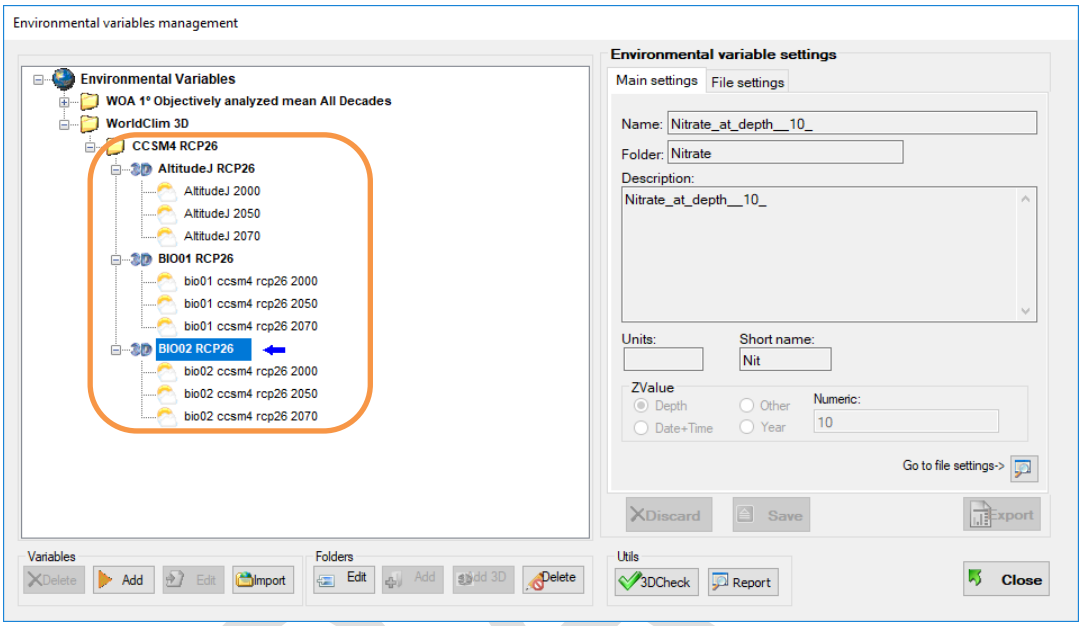

This can also be done for each individual folder containing the variables, but using the menu option *Convert folder to 3D Dataset.*

Before converting a folder in a 3D dataset, ModestR checks that it complies several rules:

✓ The folder must not have subfolders, only variables.

- ✓ All the variables in the folder should have the same precision and range. That is, all the rasters should have the same cell size and the same coordinates.
- ✓ All the variables in the folder should have the same Z meaning (such as year, depth…)
- ✓ All the variables in the folder should have a different Z value.

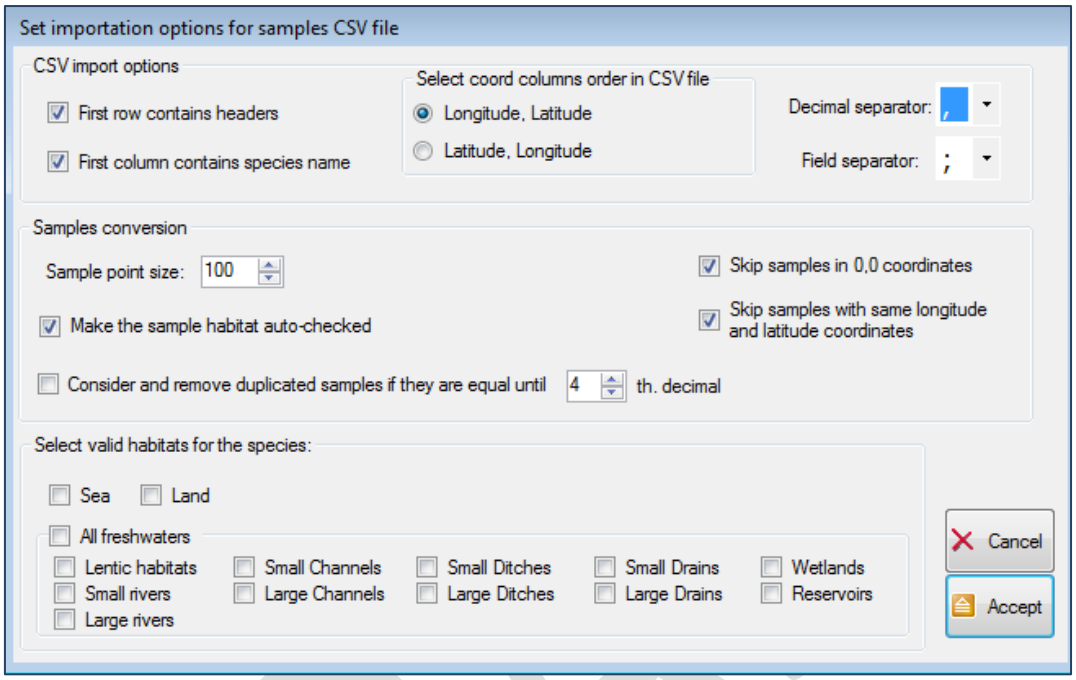
If all those criteria are fulfilled the folder will be converted to a 3D variable. This is indicated by a different icon, as shown in the next figure:



# 7 MapMaker import/export files format

## 7.1 Import occurrence data from a CSV file

You can import occurrences from a CSV file to add them to a MapMaker map. On the main menu, select *File/Import/Occurrence data from CSV*. You will have to select the file where the CSV data is located. A dialog box with the importation options will be shown:




The accepted format is a list of occurrences in the form:

[Species]; Longitude ;Latitude (*[Species] field is optional)*
or
[Species] ;Latitude; Longitude (*[Species] field is optional)*

In the dialog box you can select the correct options for the format of the CSV file to import:

- The order of the coordinates in the CSV file (Longitude;Latitude or vice versa)
- If the first columns contains the species name.
- If the first row contains headers (in that case first row will be skipped).
- Decimal separator and field separator.

More details about CSV occurrences importation are explained in section 2.3.

## 7.2 Importing occurrences from GBIF online database

You can import occurrences from the GBIF online database to a MapMaker map. On the main menu, select *File/Import/Occurrences from online GBIF database* or the equivalent button of the toolbar. You will have to write the desired species name (alternatively you can select it from a ModestR database). Then you can choose if you want MapMaker to look for synonyms in GBIF database. More details about GBIF occurrences importation are explained in section 2.3.

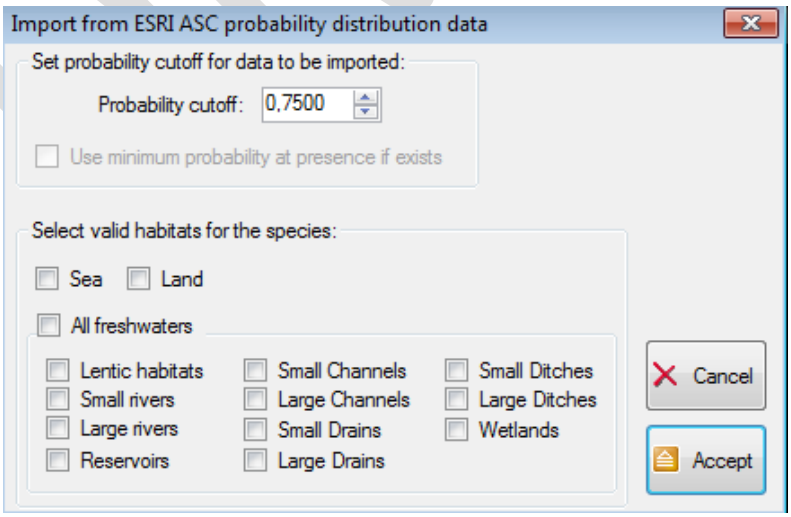
## 7.3 Importing areas from ESRI ASCII probability distribution models

You can import distribution data from distribution models with a ESRI ASC raster probability format, like the obtained using Maxent, to the currently opened ModestR map in MapMaker using the *File/Import/ESRI ASC probability distribution data* menu option. This importation option allows producing range maps from probability distribution data.

You should have previously generated one ESRI ASC file that will contain a probability matrix, where each element of the raster matrix represents the probability of presence for a species in a cell of a specified size (usually 5'x5'). Only probability matrixes with values between 0 and 1 are supported.

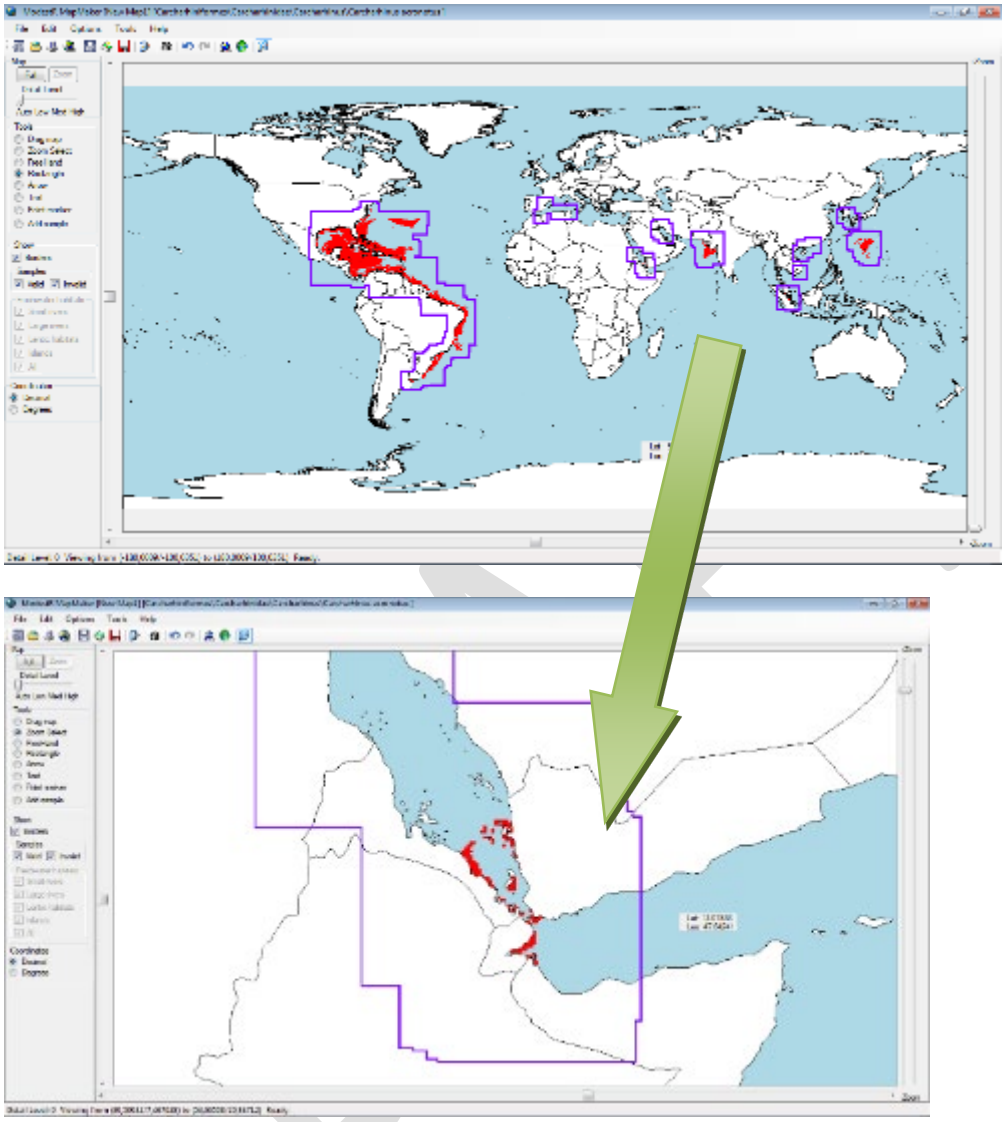
When using this importation option, you first have to select the ESRI ASC file. Then a dialog box will allow you to set several settings:

- Default probability cutoff: the minimal probability under which presence data will not be imported. If for example you select 0.75, only cells with a probability >=0.75 will be imported to the map.
- Use minimal probability at presence: this option is **only** available if you are importing data to a map where occurrences already exist. In this case, this option allows you to use as probability cutoff the lowest probability value where an occurrence exists in your map. For example, you can download occurrences to MapMaker for the species X, then import a probability distribution data file to the map, and MapMaker will automatically find the probability values in the points where there are occurrences in your current map, and use the minimum of these values as probability cutoff to import probability data. The selected value is displayed after importation.
- Valid habitats: you have to select which habitats (at least one) are valid for the species data you are importing. Only areas corresponding to those habitats will be accepted and shown in the map.




Once accepted this dialog box, the importation process will start. MapMaker will transform each cell with a probability greater or equal than the selected cutoff value to an occupied area in the map, applying the selected habitats. This way, if a cell (or some portion of it) is not located in the habitats selected as valid, it will not be shown in the map (or only the valid portion will be

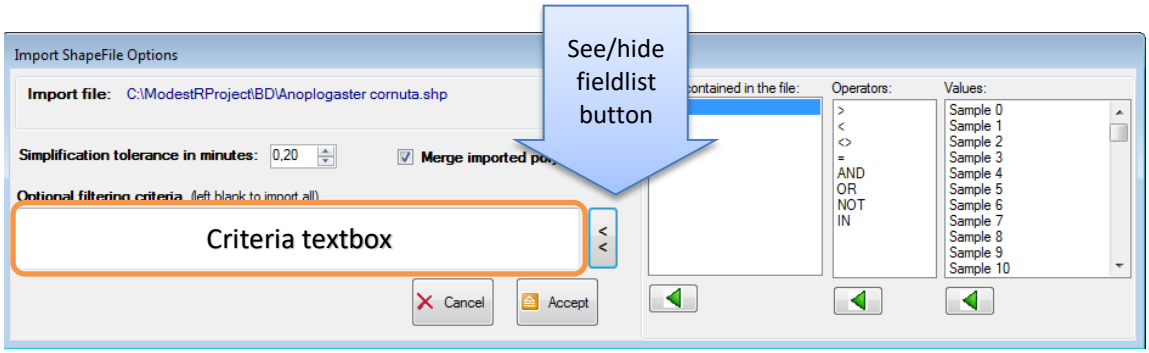
shown). It is important to take into account than some imported areas can be very small, therefore not visible in a full world map view. To allow you to localize all the imported areas, you can use the encircle presence areas feature (see subsection 2.16), as shown in this figure, where some areas are hardly visible in a full world map view, but easily located thanks to the rectangles that encircle them, and visualized zooming in.



## 7.4 Importing from Shapefiles

You can import areas or occurrences from files in ESRI Shapefile (SHP) format, which is a widely used format in GIS software. Currently, only shapefiles containing polygons or points can be imported (that is, Polygon, PolygonZ, PointZ or Point type shapefiles). Other shapefiles containing types like Polyline are not supported by now. Polygons will be converted to areas, whereas points will be converted to occurrences.

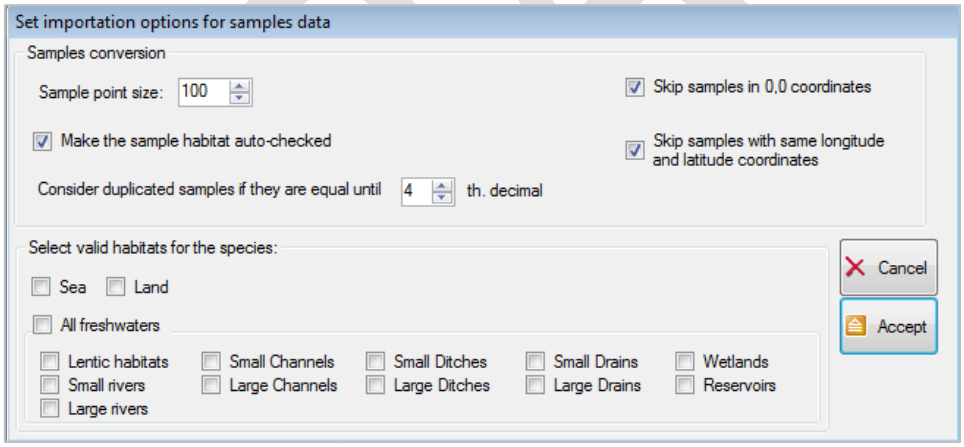
A dialog box will be shown where you can see data contents from the shapefile to be imported.




If you are familiar with shapefile format, you can also add importation criteria to filter which data will be imported. To do that, use the *See/hide fieldlist* button to see the list of fields contained in the shapefile. Then you can select a field to see the values it takes, and add criteria that will be used as filter when importing. You can also directly write a condition in the textbox. The syntax of the conditions is similar to the syntax of a WHERE clause in SQL.

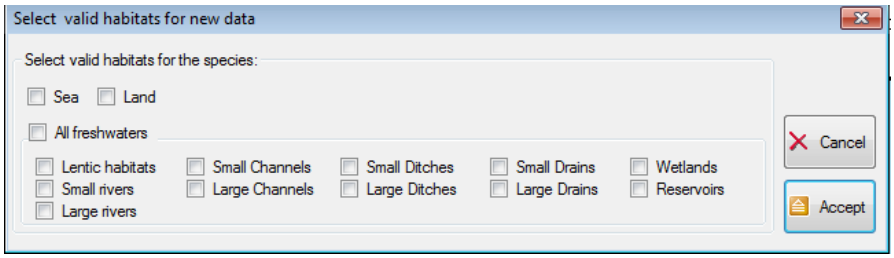
You also have the *Merge imported polygons* option, which is checked by default. This option will merge imported polygons, so overlapping polygons will be merged and therefore data simplified. But in case of errors when importing shapefiles, try to import them with this option unchecked. The same way, MapMaker will apply by default a simplification factor of 0,2 minutes to imported polygons. This options can positively affect performance, but if you really need to import data with a precision higher than this, you can set simplification tolerance to zero.

If the imported Shapefile contains points that will be converted to occurrences, a dialog box will be shown:




The options of this dialog box are the same than the shown when importing occurrences from CSV or GBIF. You can find more details in subsection 2.3.

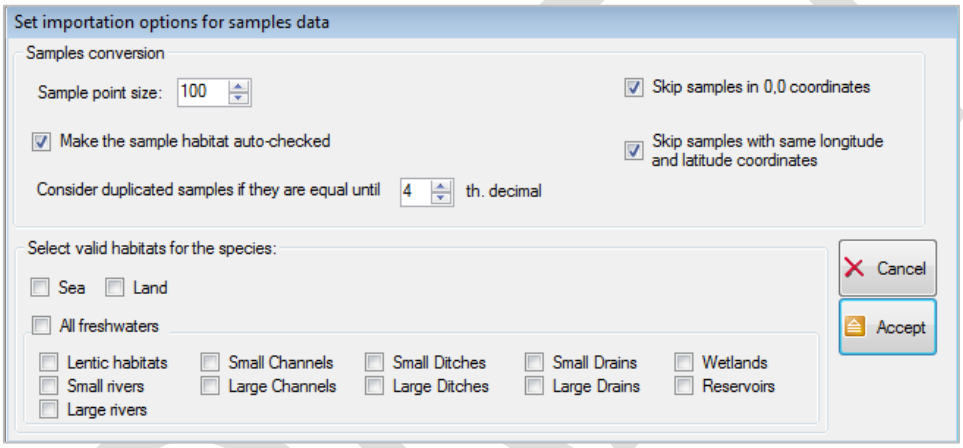
If the Shapefile contains polygons, this dialog box will be shown, where just the valid habitats for the species have to be selected. If an area is not located in the habitats selected as valid, it will not be shown in the map (or only the valid portion, if any, will be shown).




## 7.5 Importing from KML files

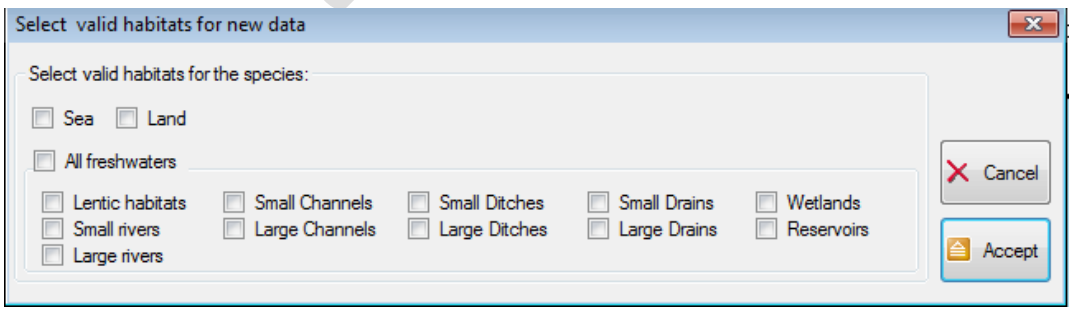
You can import areas and occurrences from files in KML format, which is a widely used format in GIS software. Currently, only polygon and point elements from a KML file can be imported. Polygons will be converted to areas, whereas points will be converted to occurrences. Besides this, as MapMaker parses KML files primarily as XML data, only KML files that comply with well-formed XML documents rules are currently supported (as KML files are XML-specialized files).

If the imported KML file contains points, which will be converted to occurrences, a dialog box will be shown:




The options of this dialog box are the same than the shown when importing occurrences from CSV or GBIF. You can find more details in subsection 2.3.

If the KML file only contains areas, this dialog box will be shown, where just the valid habitats for the species have to be selected. If an area is not located in the habitats selected as valid, it will not be shown in the map (or only the valid portion, if any, will be shown).

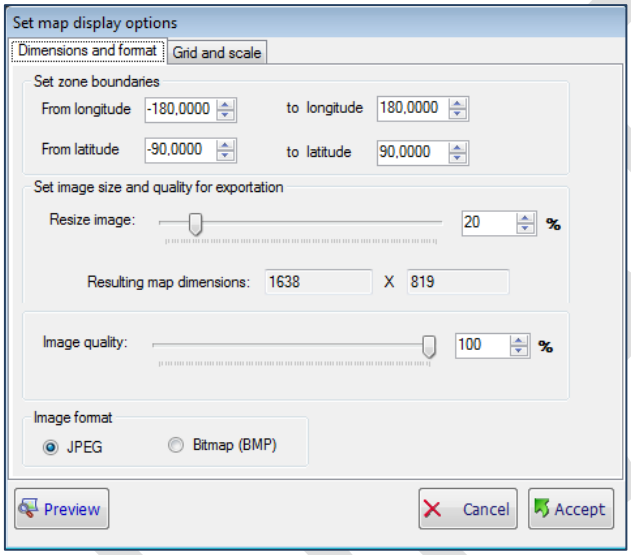


## 7.6 Export map view to an image file

You can export the map view you are currently seeing in MapMaker to a JPEG or BMP file, using the *Export current map view* button in the toolbar, or the *File/Export/Current map view as image* menu option. Next, a dialog box will appear with several tabs.

In the *Grid and scale* tab you can set if you want to see a grid/scale in the exported image, every degress/minutes you want this grid/scale, etc. In the *Dimensions and format* tab you can directly modify the coordinates of the portion of the map that will be exported. By default, the same view currently displayed in MapMaker will be exported. You can also select the size of the exported image. By default, MapMaker sets the image size as slightly bigger than your screen resolution. This is usually a good resolution to use the image in another document, presentation, etc. But if you want a more precise image that can be zoomed correctly, you can increase the image size.

Finally, you can select between JPG or BMP format.

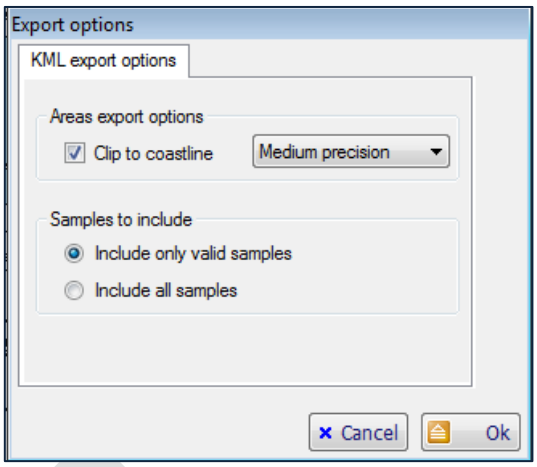



## 7.7 Export presence areas and occurrences

You can export presence areas and occurrences from the map to a file using the *File/Export* menu options. When exporting presence areas using this option, you should take into account that presence areas are actually stored in a different way than they are displayed in the map. For example, when the user selects a zone and then selects the *Sea* habitat to fill in this selection, land part will not be filled. But the data exported by default will be the coordinates of the zone selected by the user, not the exact zone filled. To export presence areas clipped to the coastline, either for land or sea areas, you should select the corresponding option, which is available when exporting to KML or shapefile. Two precision levels are proposed: medium and high. Medium precision is faster and quite precise when working with large areas. For an exact clipping to coastline, you can use the high precision option, but this can require a very long time.

Occurrences, in turn, will be exported using their exact coordinates. Only valid occurrences will be exported. The exportation formats you can select are:

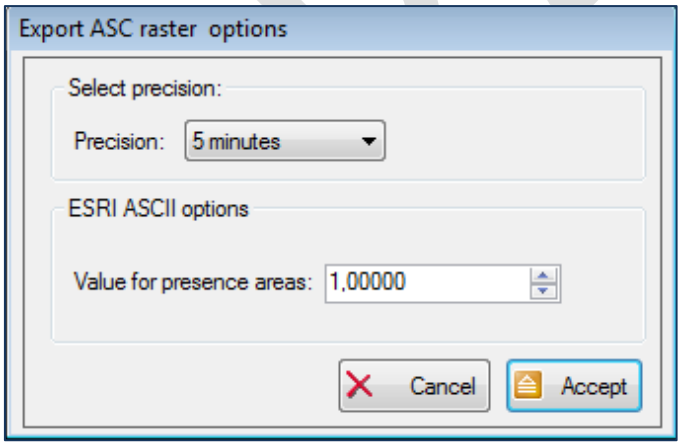
- To CSV: you can export to this format either presence areas or occurrences. See section 6.14 for details of the output formats.
- To KML: in this case a KML file will be generated, with polygons corresponding to the presence areas, and placemarks for occurrences. As explained before, you can select to clip presence areas to coastline, or just export presence areas as they where added to the map.

  KML files are widely recognized by GIS software and applications like GoogleMaps or GoogleEarth. Anyway, we have detected that some applications have problems to correctly read and interpret KML files that contain very complex polygons. In this case, shapefiles seem to work better.
- To Shapefile: this option is similar to the KML exportation, excepted that you have to select exporting presence areas or occurrences. Currently you cannot export both data to the same shapefile.

## 7.8 Export presence areas to ASC rasters

This option generates a raster in ASC format, with the selected precision, where the cells where the species is present have a value not null (usually 1, but can be selected by the used). This option is located in the menu *File/Export/To ASC raster*.




## 7.9 Export environmental data (clipped rasters)

This option allows exporting environmental data clipped to presence areas or shapes. It's explained in section 2.9.4. The exportation format is an ASC raster. It is located in the menu *File/Export/ Environmental data.*

A report in CSV format is also generated with data about clipped areas in $km^2$ and simple statistical measures (max, min, mean, std and quartiles) for the environmental data in the clipped areas. The format of this report is like that (when clipping using shapes, the report will contain a section for each shape):

| Species_1 | Total.Area | Sea.Area | Land.Area | Large.rivers. | Small.rivers. | Lentic.water | Small.chann | Large.channe | Small.ditche | Large.ditche | Small.drains | Large.drains | Reservoirs.A | Wetlands.Ar |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Species_1 | 698894,584 | 0 | 698894,584 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Species_1 | | | | | | | | | | | | | | |
| Species_1 | Variable | Num.Values | Min | Max | Mean | Std | Sum | Quartile.1 | Quartile.2 | Quartile.3 | | | | |
| Species_1 | Altitude Clip | 11077 | 1 | 2731 | 597,31868 | 410,90834 | 6616499 | 254 | 567 | 851 | | | | |
| Species_1 | BIO1 Clippec | 11321 | 0,7 | 19,2 | 13,22012 | 2,77992 | 149665 | 11,6 | 13,1 | 15,4 | | | | |
| Species_1 | BIO10 Clippe | 11321 | 7,9 | 28 | 20,72597 | 2,97527 | 234638,7 | 18,8 | 20,4 | 23,2 | | | | |
| Species_1 | BIO11 Clippe | 11321 | -5,3 | 12,9 | 6,30609 | 2,97016 | 71391,3 | 4,4 | 6 | 8,6 | | | | |

The columns are:

- Total.Area in km$^2$ of the area used to clip the rasters.
- Total area for each habitat type of the area used to clip the rasters: *Sea.Area, Land.Area, Large.Rivers.Area, Small.Rivers.Area, Lentic.waters.Area, etc.*
- For each variable clipped:
  - Number of not null values (cells) in the clipped area.
  - Minimum and maximum value of the variable in the clipped area.
  - Mean and standard deviation of the values of the variable in the clipped area.
  - Sum of the values of the variable in the clipped area.
  - Quartiles of the variable in the clipped area.

## 7.10 Export occupied areas summary report

This option generates a report in CSV format with the areas in km$^2$ of each type of habitat where the species is present. A similar report, but for any set of species (not just of one), can be obtained using DataManager. Moreover, in MRMapping a similar report is available for any rank or arbitrary group of species (families, orders, etc…). MRFinder offers a similar feature to calculate the area of any arbitrary selection in the map.

In MapMaker this options is located in the menu *File/Export/Areas summary report.*

| MapName | Total.AOO | Sea.Area | Land.Area | Large.rivers. | Small.rivers. | Lentic.water | Small.chann | Large.chann | Small.ditche | Large |
|---|---|---|---|---|---|---|---|---|---|---|
| ModestR Map | 698894,584 | 0 | 698894,584 | 0 | 0 | 0 | 0 | 0 | 0 | |

The output format is a CSV file with the following columns:

- Total Area Of Occupancy of the species: *Total.AOO*
- Total area for each habitat type occupied by the species: *Sea.Area, Land.Area, Large.Rivers.Area, Small.Rivers.Area, Lentic.waters.Area, etc.*
- Minimum and maximum coordinates where the species is present : *Min.Longitude, Max.Longitude, Min.Latitude, Max.Latitude*

## 7.11 Export CEL report

As explained in 2.11, when creating a CEL, you can export a report that contains several files about the created layer. You can export this report any time later going to menu *Mapping/Niche of occurrence/Manage compounded environmental layers*. Select the wanted layer in the list that will appear and click on *Edit details* button. A panel will be displayed on the right, where you can click on the *Export report* button.

The exported report contains the following files:

- [Name of the layer]_Axes.JPG: an image with just the axes used to define the polar coordinates.

- [Name of the layer] _ColorMap.ASC: an ASC raster where each cell contains the color value of each cell according to the CEL, as a 32 bits integer that corresponds to an ARGB color.
- [Name of the layer] _ColorScale.JPG: an image with the axes used to define the polar coordinates and the full rectangle that represents the color scale assigned to the rectangular area where there are data in the layer.
- [Name of the layer] _DataScale.JPG: an image with the axes used to define the polar coordinates and a point for each cell of the world map (or the region used to calculate the CEL) in its corresponding location in the polar coordinates system, and with its corresponding color.
- [Name of the layer] _X.ASC: an ASC raster where each cell contains the X coordinate of the cell in the polar coordinates system of the CEL.
- [Name of the layer] _Y.ASC: an ASC raster where each cell contains the Y coordinate of the cell in the polar coordinates system of the CEL.
- [Name of the layer] _Density.ASC: an ASC raster where each cell contains the density value of the cell in the polar coordinates system of the CEL.
- [Name of the layer] _DensityScale.JPG: an image with the same format than the [Name of the layer] _DataScale.JPG mentioned above, but presenting the density values. Areas with higher density (corresponding to environmental conditions more frequent) will appear in warmer colors than lower density areas.

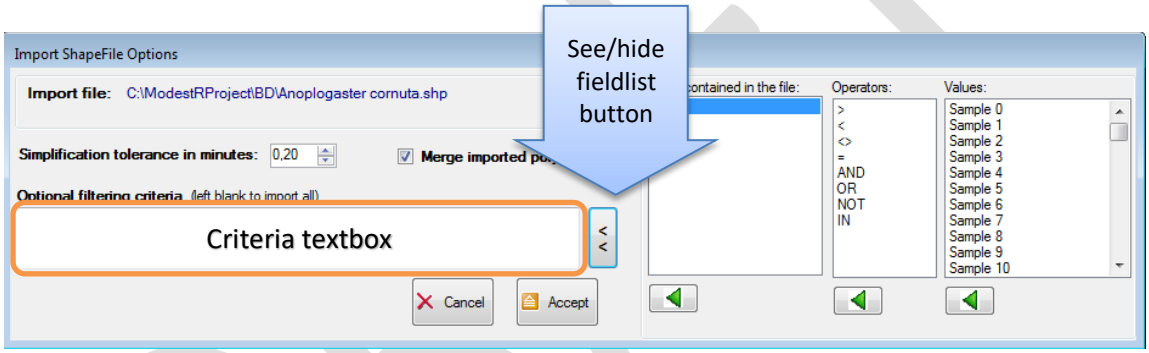
# 8 MRFinder import/export files format

In this section we briefly explain the main file formats used by MRFinder to export data from species from a ModestR database. Remember that some exportation defaults concerning CSV format can be set in the MRFinder default settings. Several MRFinder exportation options are similar to some DataManager exportation features. The difference is that MRFinder exportation options will be applied to species present in one specific area that can be additionally filtered, as explained in section 4.8, while in DataManager you just select one taxonomy branch to be exported.

## 8.1 Importing selections from Shapefiles

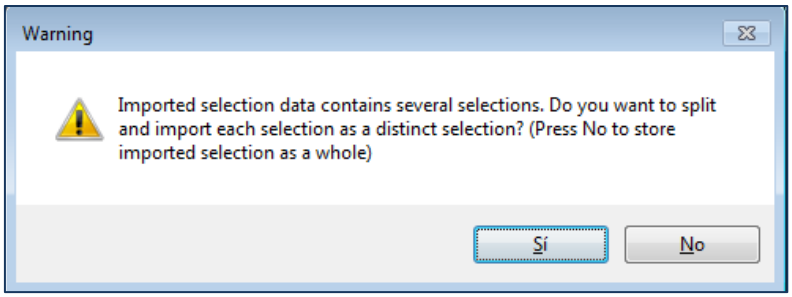
In MRFinder you can import selections to search for species from shapefiles. This is introduced in 4.4. You can import areas from files in ESRI Shapefile (SHP) format, which is a widely used format in GIS software. Currently, only shapefiles containing polygons can be imported (that is, Polygon and PolygonZ type shapefiles). Polygons will be converted to selections.

To import a shapefile go to *File/Import selection/From shapefile* menu. After selecting the file to be imported, a dialog box will be shown where you can see data contents from the shapefile to be imported.



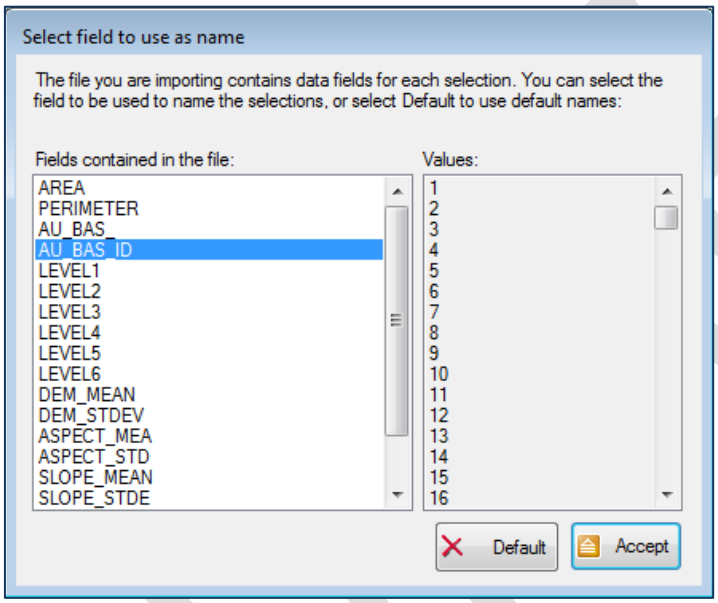

If you are familiar with shapefile format, you can also add importation criteria to filter which data will be imported. To do that, use the *See/hide fieldlist* button to see the list of fields contained in the shapefile. Then you can select a field to see the values it takes, and add criteria that will be used as filter when importing. You can also directly write a condition in the textbox. The syntax of the conditions is similar to the syntax of a WHERE clause in SQL.

If the imported shapefile contains several shapes (which is quite usual), once imported a dialog box will ask you if you want to add a single selection that include all shapes contained in the shapefile, or to split them in several selections. As explained in 4.6, as MRFinder can search for species discriminating several selections, the difference is that if you add a single selection (even if it contains several shapes), you can only obtain the species in the selection as a whole. If you add several selections, you can obtain search results grouped by selection.




If you select to split the shapes in several selections, a dialog box will show the fields contained in the shapefile. A shapefile can contain several fields with values assigned to each shape contained in the file. This dialog box allows you to see those fields and the values they take, by clicking in the wanted field. This is to help you to select which field you want to use to split and group the shapes contained in the shapefile in several selections.

For example, in the dialog box shown here as example:




If we select as field the one that is highlighted ("AU_BAS_ID"), the shapes contained in the shapefile will be grouped using this field. So all shapes with the value "1" in the field "AU_BAS_ID" will be grouped and added as a single selection to MRFinder. Then all shapes with the value "2", and so on.

## 8.2 Importing selections from KML

In MRFinder you can import selections to search for species from KML files. This is introduced in 4.4. To import a KML file go to *File/Import selection/From KML file* menu. The steps are quite the same than the described above for shapefiles (see 8.1) with the difference that the first dialog box that allows adding importation criteria when importing shapefiles will not appear when importing KML files. If the imported KML contains several shapes (which is quite usual) you will be able to select if you want to import them as a single selection or to split them, the same way as described for shapefiles (see 8.1).

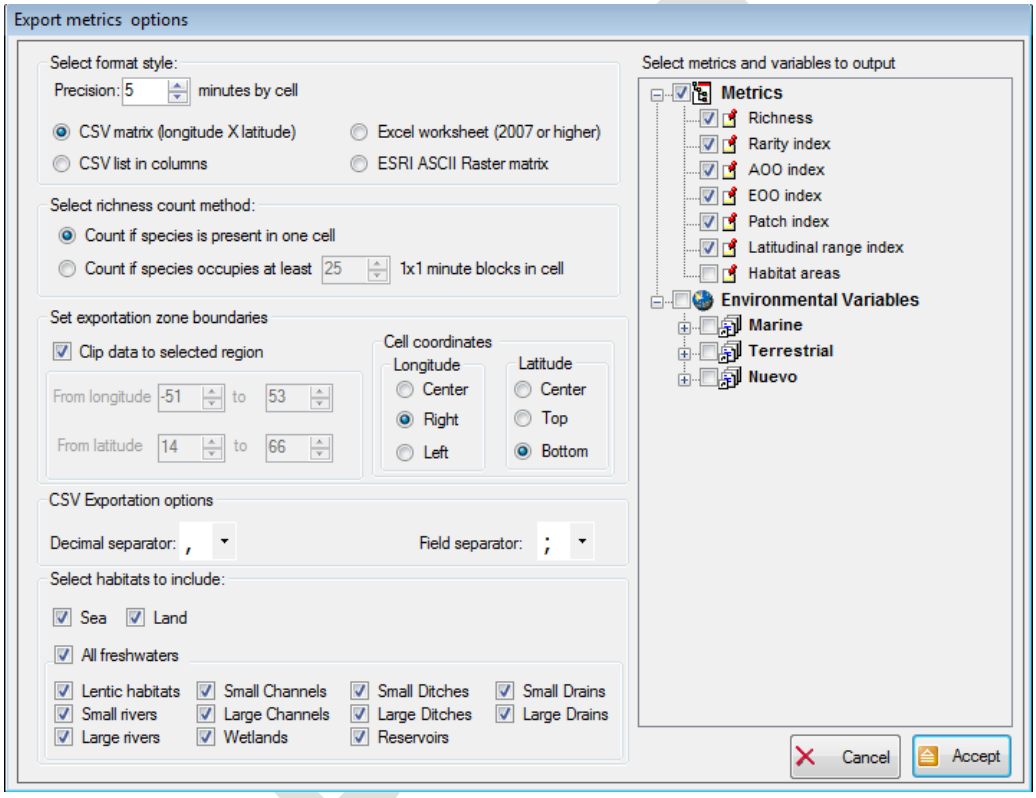
## 8.3 Importing from CSV

In MRFinder you can import selections to search for species from a CSV file. To import a CSV file go to *File/Import selection/From CSV file* menu. The expected format consists of two columns: longitude and latitude or vice versa (that can be selected by the user during importation). Each row should correspond to a point of the shape.

| Longitude | Latitude |
|---|---|
| -44,3056641 | 47,8150215 |
| -44,8905563 | 47,8150215 |
| -45,1830025 | 47,8150215 |
| -45,7678986 | 47,5225754 |
| -46,9376831 | 46,3527908 |
| -47,8150253 | 45,4754486 |
| -49,277256 | 44,3056641 |

## 8.4 Exporting metrics data

MRFinder can calculate and export several metrics from maps from selected species, such as richness data, rarity index, etc. When selecting *Export metrics* in the exportation tools of the search result window (see section 4.8), a dialog box to select exportation settings is shown, with the same options and operation as the corresponding option in DataManager. You can find more details about metrics exportation in section 6.17.




An important aspect that has to be pointed out is that while in DataManager only rectangular areas can be set to calculate and export metrics, in MRFinder irregular areas can be used, regarding the selected area in the map. The process is this:

- If you selected a rectangular area to analyse in the MRFinder map, the boundaries of this area, adjusted to the nearest integer degree, will be used by default to calculate and export metrics.
- If you selected an irregular area to analyse in the MRFinder map, the boundaries of this area will be used by default to calculate and export metrics. But the output will be a matrix adjusted to the rectangle that circumscribe the irregular area, using the nearest integer degree, but where the values corresponding to the cells outside of the selected irregular area will be set to zero.

- In any case, you can modify area boundaries disabling the *Clip data to selected region* in the metrics dialog box, and setting different rectangular boundaries.

## 8.5 Exporting latitudinal gradients

MRFinder can calculate and export latitudinal gradients for several metrics from maps from selected species. When selecting *Export gradients* in the exportation tools of the search result window (see section 4.8), a dialog box to select exportation settings is shown, with the same options and operation as the corresponding option in DataManager. You can find more details about latitudinal gradients exportation in section 6.20.

## 8.6 Exporting taxonomy to a CSV file or to clipboard

You can export taxonomic data of the selected species using the *Export taxonomy* in the exportation tools of the search result window (see section 4.8). You will see three options to export data:

- To clipboard: this option will copy the list of names of the selected species to the clipboard.
- To a CSV flat file: this option has the same operation that the corresponding option of DataManager. You can find more details in section 6.12.
- To a CSV structured file: this option has the same operation that the corresponding option of DataManager. You can find more details in section 6.12.

## 8.7 Exporting maps

You can export maps data of the selected species using the *Export Maps* in the exportation tools of the search result window (see section 4.8). You will see five options to export data:

- To grouped summary: this option exports a list of species present in each of the selected regions to search in, to a CSV file. Therefore, the first column is the region name, and the next columns correspond to summary data about each species present in this region. This option allows obtaining a list of species present in each region.
- To full summary: this option has the same operation than the corresponding option of DataManager. You can find more details in section 6.23. This option does not do provide information about the region where a species has been found. To have this information, use “To grouped summary” option.
- To presence data: this option has the same operation than the corresponding option of DataManager. You can find more details in section 6.14.
- To Maxent: this option has the same operation than the corresponding option of DataManager. You can find more details in section 6.14.
- To standalone map files: this option has the same operation than the corresponding option of DataManager. You can find more details in section 6.13.

## 8.8 Exporting statistical data

You can export statistical data of the selected species using the *Export stats* in the exportation tools of the search result window (see section 4.8). You will see two options to export data:

- To a CSV structured file: this option will generate a CSV file with the following columns:

- **Subclass, Order, Family, Genus and Species:** those columns will be shown in a structured way similar to the used when exporting taxonomy to a CSV structured file in DataManager (see section 6.12.).
- **Total.Species**: this column will count the number of species for each higher taxonomic level, in a structured way similar to the used when exporting taxonomy to a CSV structured file in DataManager (see section 6.12.).
- **Total.Rare.Species**: this column is similar to the previous one, but it will only count the number of rare species for each higher taxonomic level.
- **Rare.Species**: this column contains a "1" value for the rare species, otherwise it will be empty.
- **AOO.km2**: this column summarizes the total AOO in $Km^2$ occupied by each taxonomic level (from species to subclass), but only within the bounds of the selected area. That is, if you selected an area of 4 $Km^2$, the maximum value of this column will be 4. This column is calculated with a precision of ±1'x1'.
- **Mean.AOO**: this column summarizes the mean AOO occupied by the species of each higher taxonomic level (from genus to subclass), within the bounds of the selected area.
- **SD.AOO**: this column summarizes the standard deviation of the area occupied by the species of each higher taxonomic level (from genus to subclass), within the bounds of the selected area. If there is only one species in a higher level (for example in a genus), this value will be obviously empty.
- **Maximum.Latitude**: this column summarizes the maximum latitude where a species is present, for each taxonomic level (from species to subclass), within the bounds of the selected area.
- **Mean.Max.Latitude**: this column summarizes the mean of the previous column for each higher taxonomic level (from genus to subclass), within the bounds of the selected area.
- **SD.Max.Latitude**: this column summarizes the standard deviation of the maximum latitude where species are present for each higher taxonomic level (from genus to subclass), within the bounds of the selected area. If there is only one species in a higher level (for example in a genus), this value will be obviously empty.
- **Minimum.Latitude, Mean.Min.Latitude, SD.Min.Latitude**: those columns contain similar result than the three previous ones, but for the minimum latitude where species are present.
- **Maximum.Longitude, Mean.Max.Longitude, SD.Max.Longitude** : those columns contain similar result than the three previous ones, but for the maximum longitude where species are present.
- **Minimum.Longitude, Mean.Min.Longitude. SD.Min.Longitude**: those columns contain similar result than the three previous ones, but for the minimum longitude where species are present.
- **Real.Area.Reference.Area**: this column is the ratio between the real area occupied by a species within the bounds of the selected area, and the reference or potential area that this species could take up in this selected

area, taking into account the valid habitats for this species. For example, is a species is only present in marine habitats, only sea areas will be took into account to calculate reference area for the species. The maximum value of this column will be obviously 1 (if a species takes up all the areas with valid habitats for it, into the bounds of the selected area), but take into account that this column is calculated with a precision of ±1'x1'.

- **Mean.Real.Area.Reference.Area**: this column summarizes the mean of the *real area/reference area* ratio explained before, for all species of each higher taxonomic level (from genus to subclass).
- **SD.Real.Area.Reference.Area**: this column summarizes the standard deviation of the *real area/reference area* ratio explained before, for all species of each higher taxonomic level (from genus to subclass). If there is only one species in a higher level (for example in a genus), this value will be obviously empty.

- To a CSV flat file: this option will generate a CSV file where there are no summarize rows for higher taxonomic levels than species, so only some of the columns described above for the CSV structured file option will be exported. They are: Subclass, Order, Family, Genus, Species, Rare.Species, AOO.km2, Maximum.Latitude, Minimum.Latitude, Maximum.Longitude, Minimum.Longitude, Real.Area.Reference.Area.

## 8.9 Exporting region coordinates

You can export the coordinates of the region you selected in the map to do a search in MRFinder using the *Export region coords* in the exportation tools of the search result window (see section 4.8). You will see two options to export data:

- To clipboard: this option will copy the coordinates (longitude, latitude) of the vertexes of the selected region to the clipboard.
- To a CSV file: this option will export the coordinates (longitude, latitude) of the vertexes of the selected region to a CSV file, using the default delimiter and decimal point settings. The resulting file will look like this:

```
"Longitude";"Latitude"
-64,40827;37,33813
-66,89748;38,27158
-72,4982;39,5162
-77,16547;40,1385
-81,21043;40,1385
```

## 8.10 Export map view to an image file

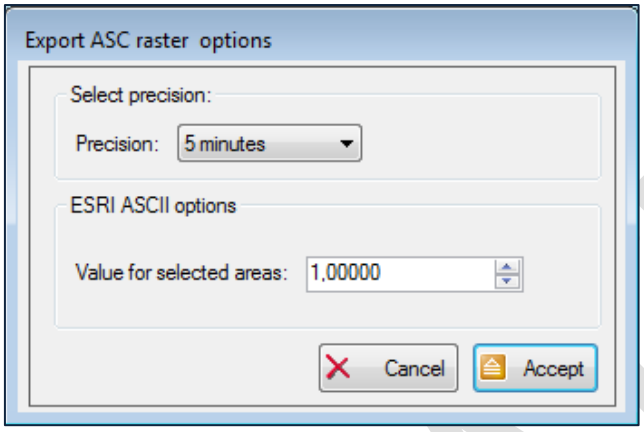
You can export the map view you are currently seeing in MRFinder to a JPEG or BMP file, using the *Export current map view* button in the toolbar, or the *File/Export/Current map view as image* menu option. Next, a dialog box will appear with several tabs. The settings and steps are the same than in MapMaker (see 7.6).

## 8.11 Export selections to shapefile or KML

You can export current selections in the map to a shapefile or a KML file using the *File/Export/Current selection* menu option, and then selecting the wanted suboption. All selections will be exported to the file in the selected format.

## 8.12 Export to ASC rasters

This option generates a raster in ASC format, with the selected precision, where the cells corresponding to the selections in the map have a value not null (usually 1, but can be selected by the used). This option is located in the menu *File/Export/Current selections/To ASC raster*.




## 8.13 Export environmental data

This option allows exporting environmental data clipped to the selections currently added to the map. It's explained in section 4.12.3. The exportation format is an ASC raster. It is located in the menu *File/Export/ Environmental data for selections.* Besides clipped rasters a report is generated as a CSV file. More details can be found in section 7.9.

## 8.14 Export selection areas summary report

This option generates a report in CSV format with the areas in $km^2$ of each selection currently added to the map. This option is located in the menu *File/Export/Selection area summary report.*

| Area | Total.Area | Sea.Area | Land.Area | Large.rivers.. | Small.rivers. | Lentic.water | Small.chann | Large.channe | Small.ditche | Large.ditche | Small.dra |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Region 1 | 12486850,2 | 7357684,37 | 5009050,18 | 12037,9188 | 13418,0881 | 40556,1719 | 641,505403 | 28,2510168 | 545,991673 | 1,24475277 | 104,004 |
| Region 2 | 210154,301 | 205277,693 | 4359,11177 | 39,7605351 | 1,98068858 | 449,092818 | 0 | 0 | 0 | 0 | |
| Region 3 | 7294616,21 | 1989837,92 | 5216552,31 | 49795,0012 | 4815,26997 | 18173,362 | 93,115466 | 2,43716662 | 24,3640554 | 0 | 14,8686 |
| Region 4 | 1429657,68 | 1429657,68 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |

The output format is a CSV file with the following columns:

- *Area*: name of the selection.
- *Total Area* of the selection
- Total area for each habitat type in the selection: *Sea.Area, Land.Area, Large.Rivers.Area, Small.Rivers.Area, Lentic.waters.Area, etc.*
- Minimum and maximum coordinates of the selection : *Min.Longitude, Max.Longitude, Min.Latitude, Max.Latitude*

# 9 MRMapping import/export files format

## 9.1 Exporting geographic overlapping areas report

In order to calculate spatial overlaps in MRMapping, once the distributions are added to the map, the user has just to select the menu *File/Export/Geographic overlapping*. A dialog box will allow selecting the reference area or geographic background to be used. The available options are (1) the whole world, (2) a country, a river basin or any other desired area that can be added by the user as a shape file; and (3) the extent of occurrence (EOO's) of the considered distributions.

Results are exported to a CSV file. To better understand the results, it is worth to explain that when performing overlap calculations, a 2x2 contingency table is internally determined from the observed values as shown below:

| | A=1 | A=0 |
|---|---|---|
| B=1 | *a* | *b* |
| B=0 | *c* | *d* |

Here A and B are a pair of distributions, while *a, b, c* and *d* are respectively the extension area in km$^2$ where both distributions overlap, where only B is present, where only A is present, and where none of the two taxonomical entities are present (therefore, *d* will depend on the selected geographical background). Another 2x2 expected contingency table is determined considering row and column marginal totals so that:

| | A'=1 | A'=0 |
|---|---|---|
| B'=1 | *a'* | *b'* |
| B'=0 | *c'* | *d'* |

In where *a'= ((a+c)*(a+b)) / (a+b+c+d), b ´= ((b+d)*(a+b)) / (a+b+c+d), c' = ((a+c)*(c+d)) / (a+b+c+d), and d' = ((c+d)*(b+d)) / (a+b+c+d).*

A brief explanation of each column of the results is provided below:

- Distribution 1 and 2: the names of the compared pair of distributions
- Total Overlap km2: the total area in km$^2$ of the region where both distributions overlap.
- Observed %Overlap Distribution 1 and 2: observed percentage of each distribution area that overlaps with the other one. For distribution 1 it is (*a/(a+c))x100*, while for distribution 2 it is (*a/(a+b))x100*.
- Expected H0 %Overlap Distribution1 and 2: expected percentage of each distribution area that would overlaps with the other one assuming the independency hypothesis ($H_0$). Thus, it is calculated in the same way than Observed %Overlap, but using the 2x2 expected contingency table. For distribution 1 it is (*a'/(a'+c'))x100*, while for distribution 2 it is (*a'/(a'+b'))x100*.
- ESP observed: observed fraction of shared presences (ESP). ESP is derived from the Sørensen similarity index (the classic formula used to compare two occurrences in a

contingency table: *2a/(2a+b+c)*). Values close to 1 reflect strong similarity while values close to 0 reflect strong dissimilarity.

- ESP H0 expected: analog to ESP observed, but calculated using the 2x2 expected contingency tables: *2a'/(2a'+b'+c')*. It is the expected ESP under the $H_0$ or independency hypothesis. Therefore, an expected ESP lower than the observed ESP can be an indication of a positive association between distributions (distributions tend to co-occur more than if they were independent) and vice versa.
- Chi-Square: Pearson's chi-square statistics calculated from the observed and expected contingency tables.
- ChiSqr P Value: p-value corresponding to the obtained chi-square statistics, applying the Benjamini-Hochberg procedure to adjust p-values for multiple comparisons. This p-value is a classical indicator of the statistical significance of the results.
- Fisher Lower MidP: p-value obtained using the Fisher's exact test lower tail mid-P variant applying the Benjamini-Hochberg correction. Besides its interest as indicator of the statistical significance of the results, it can also be used as a similarity index that provides information about the more likely direction of the association between a pair of distributions. This index ranges between [0, 1], with values close to 0 indicating strong negative association between both distribution, and values close to 1 indicating strong positive association.
- Correl Coef: this value corresponds to the widely used correlation coefficient, also called phi coefficient ($\phi$) of the observed contingency table. It is:

$$\phi = \frac{a * d - b * c}{\sqrt{(a+b) * (c+d) * (a+c) * (b+d)}}$$

  This coefficient ranges from -1 to 1, where +1 indicates total similarity/overlap; -1 indicates total dissimilarity/segregation; and 0 indicates no relationship. As expected, Fisher's test values of 0 coincides with $\phi < 0$ and Fisher's test values of 1 with $\phi > 0$.

- Odds Ratio (OR): this coefficient is classically used to assess how strongly the presence of A is associated with the presence of B in a contingency table:

$$OR = \frac{(a * d)}{(b * c)}$$

  OR range from 0 to +∞ and their classical interpretation is that when OR>1 the presence of A is considered to be associated with the presence of B; OR<1 indicates the contrary; and OR=1 indicates no relationship.

- YuleQ: Yule's Q is another coefficient used to assess the association between A and B:

$$Q = \frac{(a * d) - (b * c)}{(a * d) + (b * c)}$$

  Q range from -1 to +1. A value close to +1 indicates a strong positive association between A and B; a value close to -1 indicates a strong negative association; and Q=0 indicates no relationship.

## 9.2 Exporting environmental overlapping areas report

The calculation of environmental overlaps in ModestR requires the usage of a 2DCEL that can be considered similar to a geographical background, translated to the environmental space. In a few words, a CEL can be seen as a multidimensional compound of several environmental variables projected in a bidimensional polar coordinates system. The variables used to build an EL will obviously influence the results, and using different EL's will lead to different overlap measures. An EL can be built considering a specific geographic region selected by the user as reference area (e.g. a country, a drainage basin, the extent of occurrence of one or more species, or an arbitrary polygon). A step-by-step explanation about how to create a 2DCEL in ModestR can be found in section 2.11 .

To calculate the overlaps between each pair of environmental distributions in MRMapping the user has to go to the menu *File/Export/Environmental overlapping* and select the EL to be used as reference. A dialog box allows setting the following parameters:

- Tolerance/Grid: in order to calculate overlaps, the polar coordinates system of the EL used as reference is split into a grid in which the cell size can be selected by the user, either setting the grid dimensions or a tolerance value that may oscillate from ±0.01% to ±5% of the size of the whole polar coordinates system. Overlaps are measured just by counting the cells of the grid that contain points corresponding to the distributions to be compared. The higher the tolerance the higher is the grain size used to perform calculations, then the more likely it is the degree of overlap.
- Random trials: in order to provide a supplementary assessment on the significance of the results, it is possible to compare the observed results with those of an arbitrary number of null distributions. These are generated by polling the occurrences of two environmental distributions and calculating the overlap of a number of randomly selected points equal to the original ones.

A report of the results is exported to a CSV file. A brief explanation of each column of this file is provided below:

- Distribution 1 and 2: the names of the compared pair of distributions
- Columns from *Observed.%Overlap.Distribution1* to *YuleQ* have the same meaning as the ones previously described for geographical overlaps (see section 9.1), just in the new context of environmental overlaps. Similarly to what was explained above for geographical overlaps, for environmental overlap calculations 2x2 contingency tables are internally determined from the observed and expected value, and used to calculate the results of those columns.

In the case of the environmental space, the grids previously described, which are constructed for each distribution in order to quantify overlaps within the EL, can be straightforwardly transformed in histograms. As a cell can contain zero, one, or several points corresponding to a distribution in the environmental space, it is easy to consider each cell as a bin of the histogram. Using this approach, distance and similarity measures are also calculated between the normalized histograms *P* and *Q* of each pair of distributions (in a normalized histogram the sum

of all probabilities is equal to 1). The following measurements correspond to the columns of Table 3,

- Prob.Dist.EuclideanDistx100: Euclidean distance between the two histograms. As this column can show very small values, the obtained value was multiplied by 100 to make it more easily readable. It is:

$$EuclideanDist = 100x\sqrt{\sum_{i=1}^{n}|P_i - Q_i|^2}$$

where n=number of bins of the histograms , $P_i$ and $Q_i$ refer to the ith bin of the P and Q histograms respectively.

- Prob.Dist.BrayCurtis: the Bray Curtis similarity index (Sung-Hyuk Cha, 2007) has a range [0, 1] where values close to 0 reflect dissimilarity and values close to 1 reflect similarity. It is:

$$BrayCurtis = \frac{2x\sum_{i=1}^{n} min(P_i, Q_i)}{\sum_{i=1}^{n}(P_i + Q_i)}$$

- Prob.Dist.ChiSquare the Chi-square distance (Pele and Werman, 2010) has a range [0, 1] where values close to 0 reflect similarity and values close to 1 reflect dissimilarity. It is:

$$ChiSquareDist = \frac{1}{2}\sum_{i=1}^{n}\frac{(P_i - Q_i)^2}{(P_i + Q_i)}$$

- Prob.Dist.Bhattacharyya: the Bhattacharyya distance is another distance measure (Bhattacharyya, 1943). It has a range [0, 1] where values close to 0 reflect similarity and values close to 1 reflect dissimilarity. It is:

$$BhattacharyyaDist = \sqrt{1 - \sum_{i=1}^{n}\sqrt{P_i x Q_i}}$$

In order to provide a supplementary assessment on the significance of the results, it is possible to compare the observed results with those of an arbitrary number of null distributions, as noted previously. To a null distribution from a pair of distributions A and B with $A_p$ and $B_p$ sets of points respectively, first a set of points S is built as $S = A_p B_p$ . Then a pair of distributions A' and B' are created. A' and B' will respectively contain n$A_p$ and n$B_p$ points randomly took from S. Then the overlap between A' and B' is calculated and compared to those of A and B. Under normal circumstances, the overlap between the null distributions generated in this way will be higher than observed ones, indicating that the results of comparing A and B are more significant than if we are comparing two random distributions of the same size.

If the user has selected this option, the following columns will be included in the CSV report:

- Random.Test.Trials: number of random models generated and compared with the observed data, as selected by the user.
- Percent.Better.Than.Random.Trialst: percentage of times in which the overlap between the observed distributions are lower than the overlaps between randomly simulated distributions.

- Percent.BetterOrEquals.Random.Trials: percentage of times in which the overlaps between the observed distributions are lower or equal than the overlaps between the randomly generated ones. This measure is specifically aimed to the cases where one distribution is totally contained in another one, or two distributions are identical (i.e. there are overlaps of 100%). In those cases, as commented above, randomly generated distributions cannot have higher overlaps than the observed ones.

Besides the CSV report, the output also includes a richness graph in which the niches of all the selected taxonomical entities are overlapped (niche species richness) within the environmental space in order to visualize the environmental conditions at which a higher number of species seem to be able to maintain viable populations.

The user can optionally select to have a detailed report that additionally includes the contingency tables for each pair of distributions, detailed information about the geographical and environmental distributions, and a set of graphs showing each pair of distributions and their overlaps can be exported.
Finally, other relevant information about the environmental distributions which is included in the detailed report is:

- Distribution: the name of the distribution
- Percent.Env.Layer Cells.Presence: percentage of the cells of the EL used as reference where the distribution is present. The same as overlaps, this value will depend on the grid size selected to perform calculations, as explained above.
- NumPointsInLayer: total number of points that form the EL used as reference. This value is independent from the grid size selected.
- NumPointsInDistrib: total number of points of the EL that also belong to the distribution. This value is independent from the grid size selected.
- X.Center.Of.Mass: mean value of the X polar coordinate points of the distribution
- Y.Center.of.Mass: mean value of the Y polar coordinate points of the distribution

# 10 Checking for updates

ModestR is regularly updated to add new features. Therefore regular updates are released. You can check for updates directly from any ModestR application using *Help/Check for updates* menu option. This makes that the application connects online to ModestR website to download data and check for available updates for your installation (no data are sent from your computer to the Web). If there are new updates, a dialog box with a list of updates will be displayed.

You can also make that ModestR applications check automatically for updates, setting this options in the *Options/Preferences* menu item of any of those applications.

Updates have to be downloaded from ModestR website, and installed on your computer taking into account the path where you have installed ModestR software. Updates will try to install themselves in the default ModestR installation path (C:\ModestR). If you have modified this path when installing ModestR for the first time, you may have to set the correct path to install updates.

# 11 References

Chapman, A. D. 2005. Principles and Methods of Data Cleaning – Primary Species and Species-Occurrence Data, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen. Available online at http://www.gbif.org/orc/?doc_id=1262

Dubois, D.M. (1973) An index of fluctuations, Do, connected with diversity and stability of ecosystems: applications in the Lotka–Volterra model and in an experimental distribution of species. Rapport de sythèse III, Programme National sur l'environment Physique et Biologique, Project Mer. Commision Interministérielle de la Politique Scientifique. Liège.

Edelsbrunner, H. & Mücke E. P. (1994) Three-dimensional alpha shapes. ACM Trans. Graph. 13(1):43-72.

Guisande, C., Barreiro, A., Maneiro, I., Riveiro, I., Vergara, A.R., Vaamonde, A. (2006) Tratamiento de datos. Ediciones Díaz de Santos, 367 pp.

Jiménez-Valverde, A. 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. – Global Ecol. Biogeogr. 21: 498–507.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography, 17, 145–151.

Phillips, S. J. 2017. A Brief Tutorial on Maxent. Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on April, 10, 2018.

Phillips, S.J., Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31(2):161-175. DOI: 10.1111/j.0906-7590.2008.5203.x

Tyberghein L., Verbruggen H., Pauly K., Troupin C., Minur F., de Clerck O. 2012. Bio-ORACLE: a global environmental dataset for marine species distribution modelling— Global Ecology and Biogeog., 21: 272–281

# 12 Thematic/Feature Index

# D

# E

# F

# G

GBIF

Import occurrences

Batch importation (import occurrences for a collection of species)

Check taxonomy data using GBIF

# H

# I

Import distribution data from an SDM model (ej: Maxent)

For a single species: 2.5, 7.3

For a collection of species: 3.4, 6.10

Import distribution data from shapefiles

# J



# K



# L



# M

# N



# O

# P



# Q



# R

# S

# T



# U

# V

# W

# X

# Y

# Z