

## Supplementary material 4

### Appendix 3

Authors: Ana Manjarrés-Hernández, Cástor Guisande, Emilio García-Roselló, Juergen Heine, Patricia Pelayo-Villamil, Elisa Pérez-Costas, Luis González-Vilas, Jacinto González-Dacosta, Santiago R. Duque, Carlos Granado-Lorencio, Jorge M. Lobo.

Data type: Full description of the NOO3D procedure followed to predict the future distribution of world freshwater fishes.

---

## Niche of Occurrence 3D in ModestR

Niche of Occurrence 3D (NOO3D) is a method that extends NOO2D to generate 3D density or distribution maps for a given species using both a set of environmental layers, a set of occurrences of the species (defined by a latitude and a longitude), and a third dimension that can be time, depth, or eventually a user-defined one.

The steps to apply NOO3D in ModestR are the following:

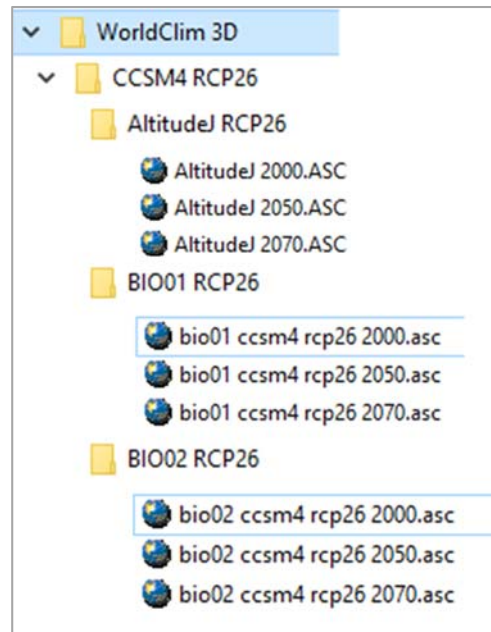
### 1. Integrating 3D environmental data in ModestR

To be able to use NOO3D, the user should first integrate what we call 3D environmental variables in ModestR. 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).

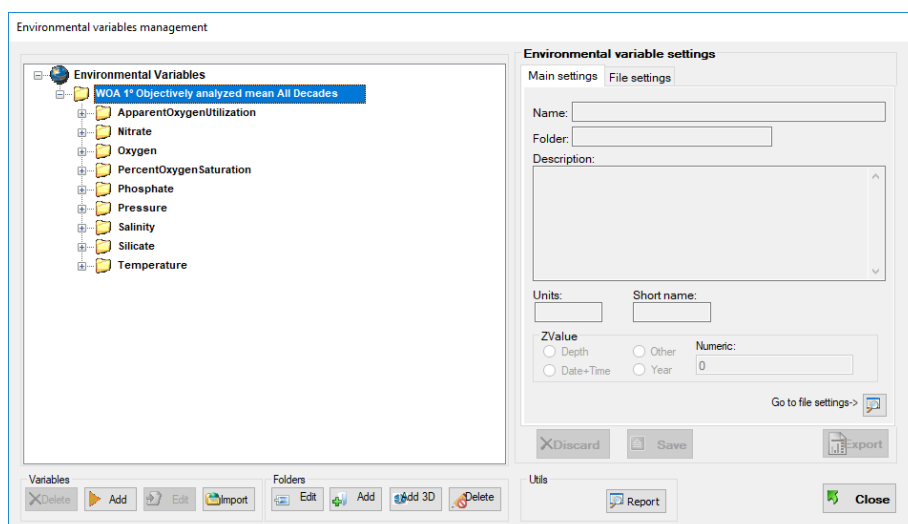
A collection of those 3D environmental variables is needed to perform NOO3D. This can be, for example, a time series of climatic variables forecasting a scenario for the near-future, as available in WorldClim; or a set of marine environmental variables for different depths, as available in the World 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<sup>1</sup>. 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.

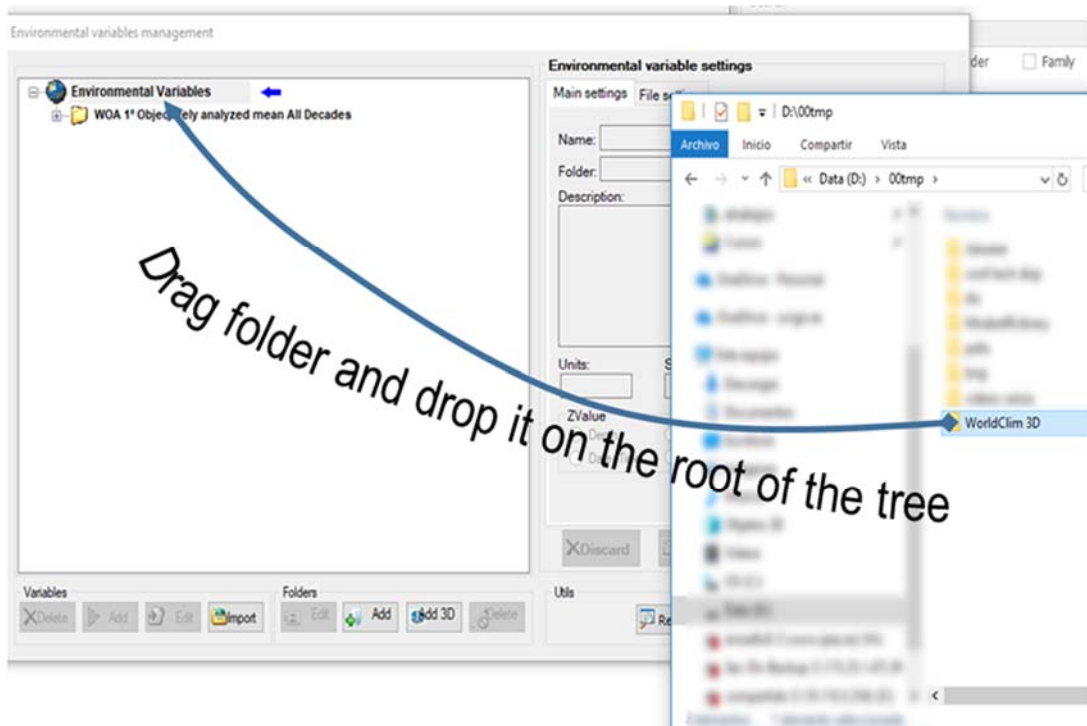


- Once the variables organized in this way, you can also avoid later tedious tasks by already adding to each file name its corresponding Z value. In the previous figure we can see that this value has been added in brackets to the end of each file name. This will prevent you from having to manually enter 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).

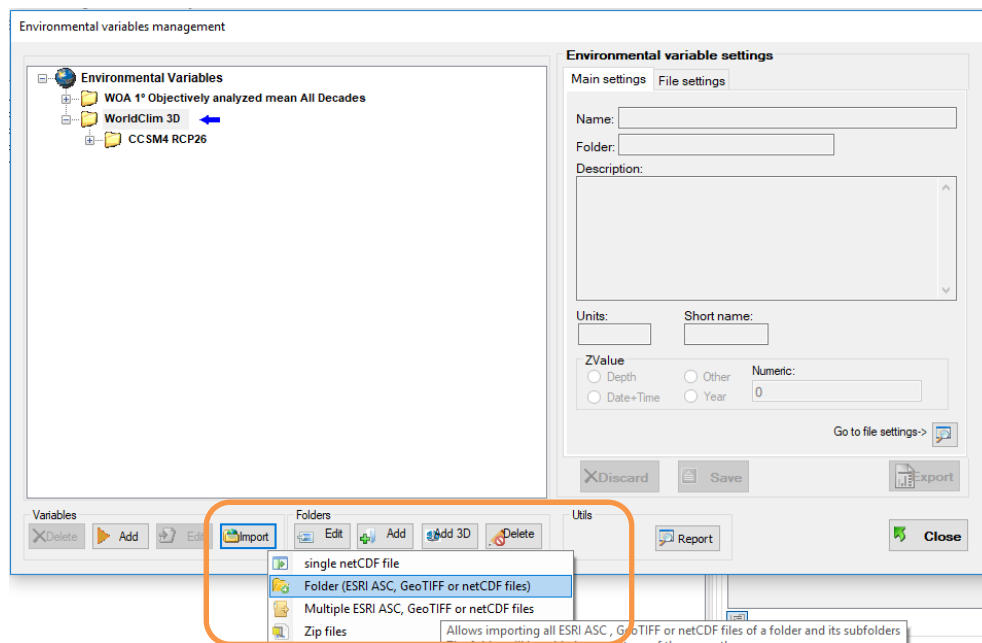


<sup>1</sup> Those are two formats directly supported by ModestR. netCDF format is also supported, but it can be less easy to integrate 3D variables in ModestR.

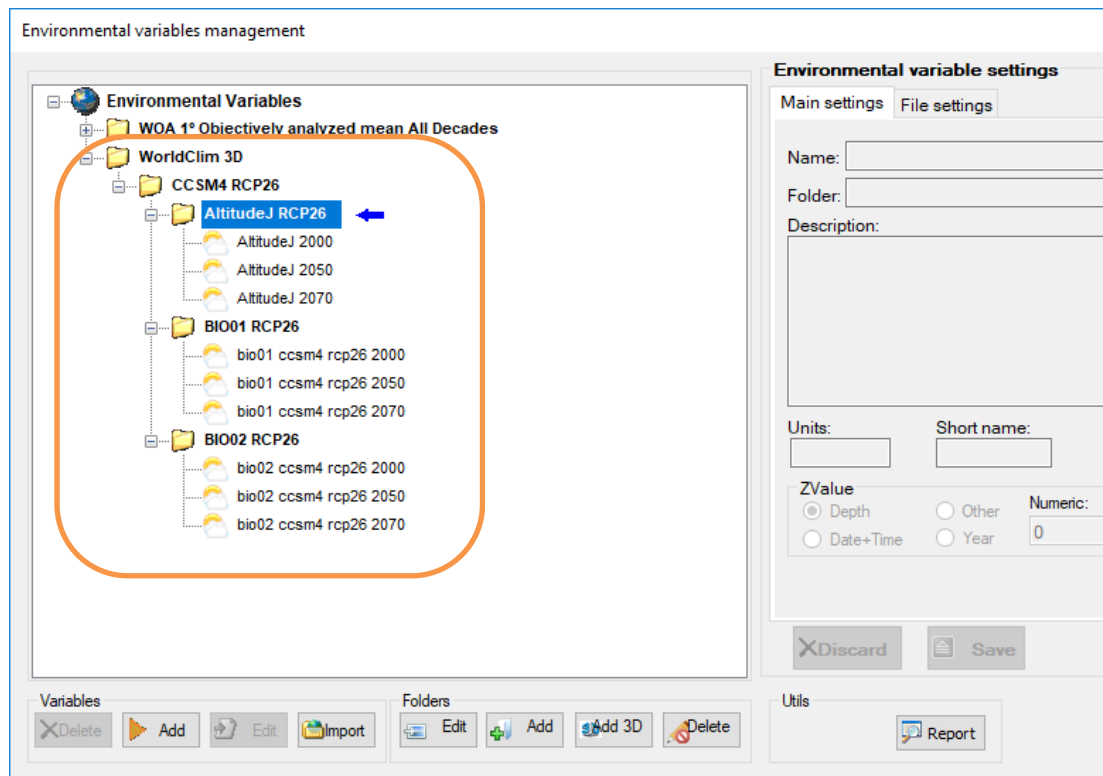
- The easiest way to import environmental data to ModestR it's just to drag&drop the folder where we previously organized the raster files to the variables tree:



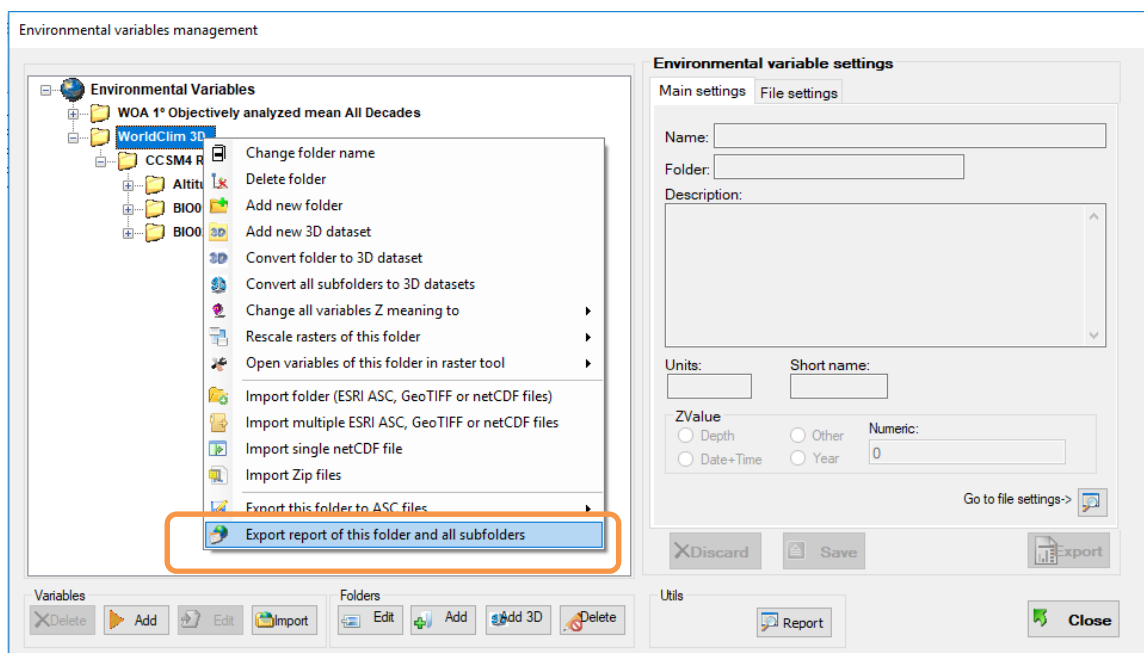
Alternatively the option Import/Folder can also be used. Selecting the main folder where variables are, all subfolders and the contained variables will be imported.



- 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*.

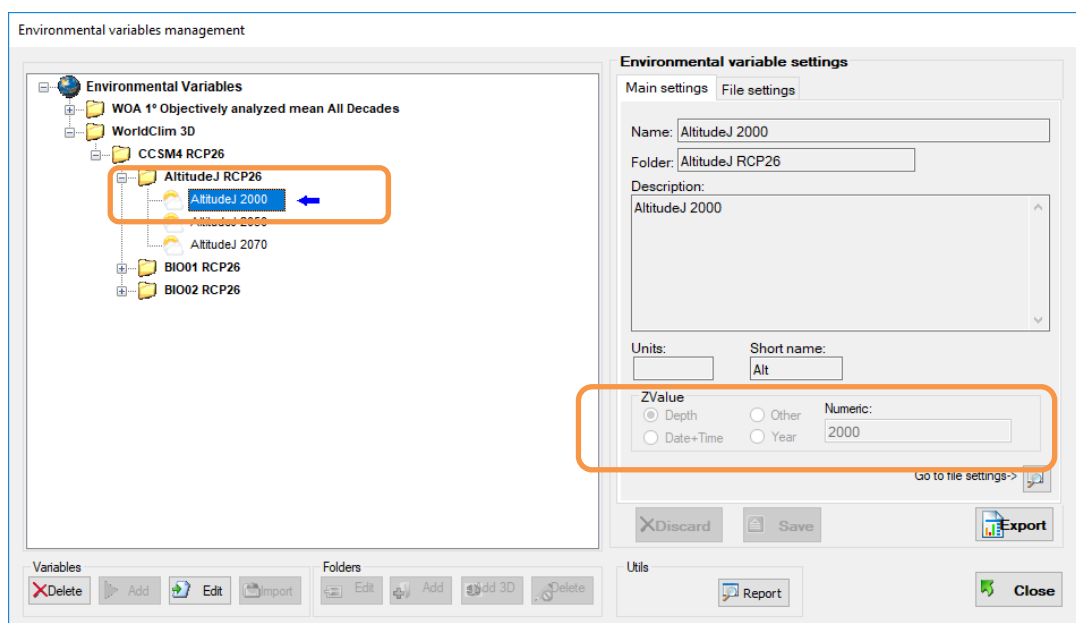


This will export a CSV file with several data about each variable. Among those data are the Z meaning and the Z value of each one. In this figure is an example of some columns of this file. As it is shown, the Z value assigned to each variable is correct, because ModestR inferred it from the filename, as explained before.

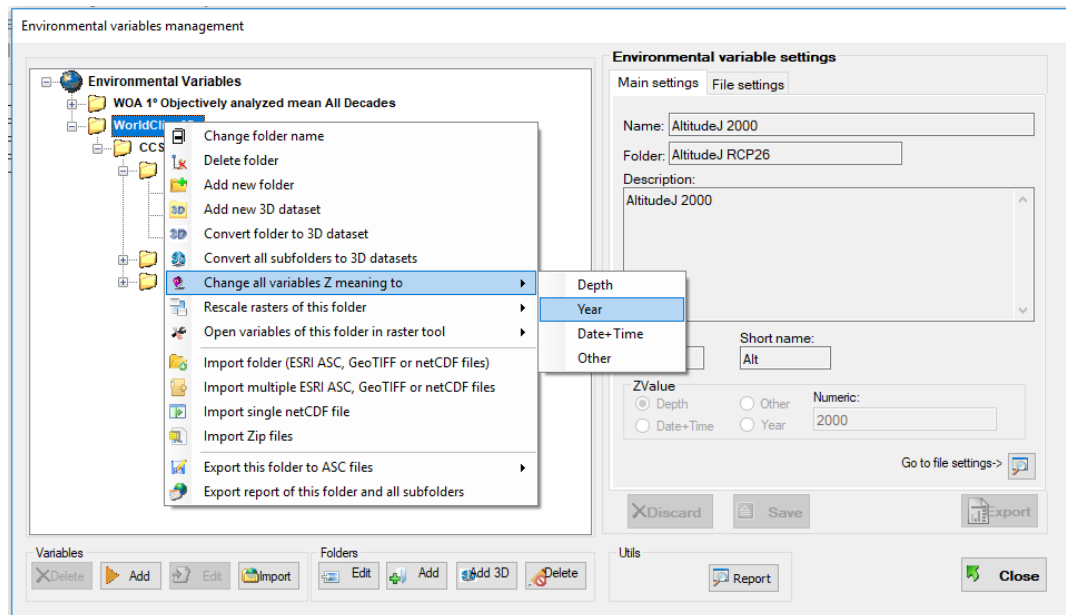
Name	ZMeaning	ZValue
AltitudeJ 2000	Depth	2000
AltitudeJ 2050	Depth	2050
AltitudeJ 2070	Depth	2070
bio01 ccsm4 rcp26 2000	Depth	2000
bio01 ccsm4 rcp26 2050	Depth	2050
bio01 ccsm4 rcp26 2070	Depth	2070
bio02 ccsm4 rcp26 2000	Depth	2000
bio02 ccsm4 rcp26 2050	Depth	2050
bio02 ccsm4 rcp26 2070	Depth	2070

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.

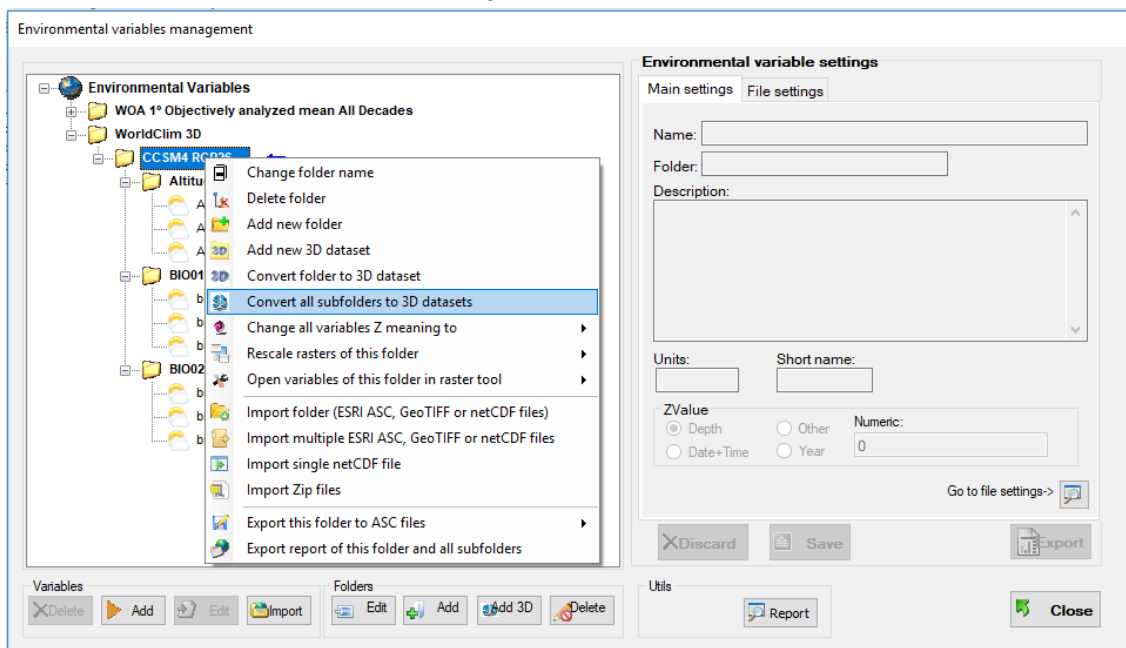
This can also be checked for an individual variable selecting it on the tree and seeing its parameters on the right panel:



- 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*.



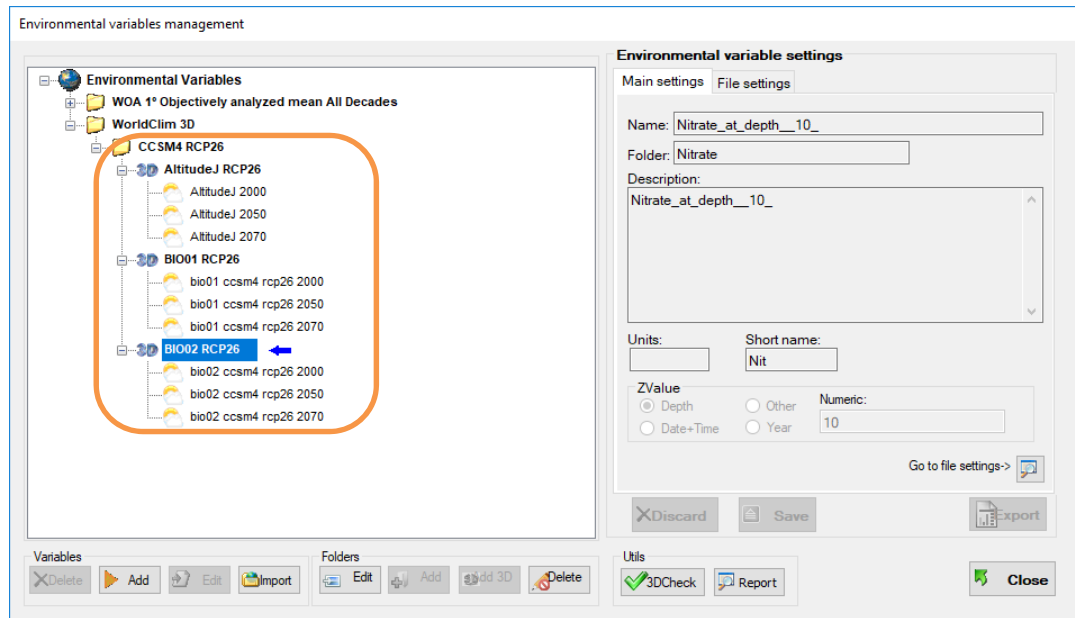
This can also be done for each individual folder containing the variables, but using the menu option *Convert folder to 3D Dataset*.

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.

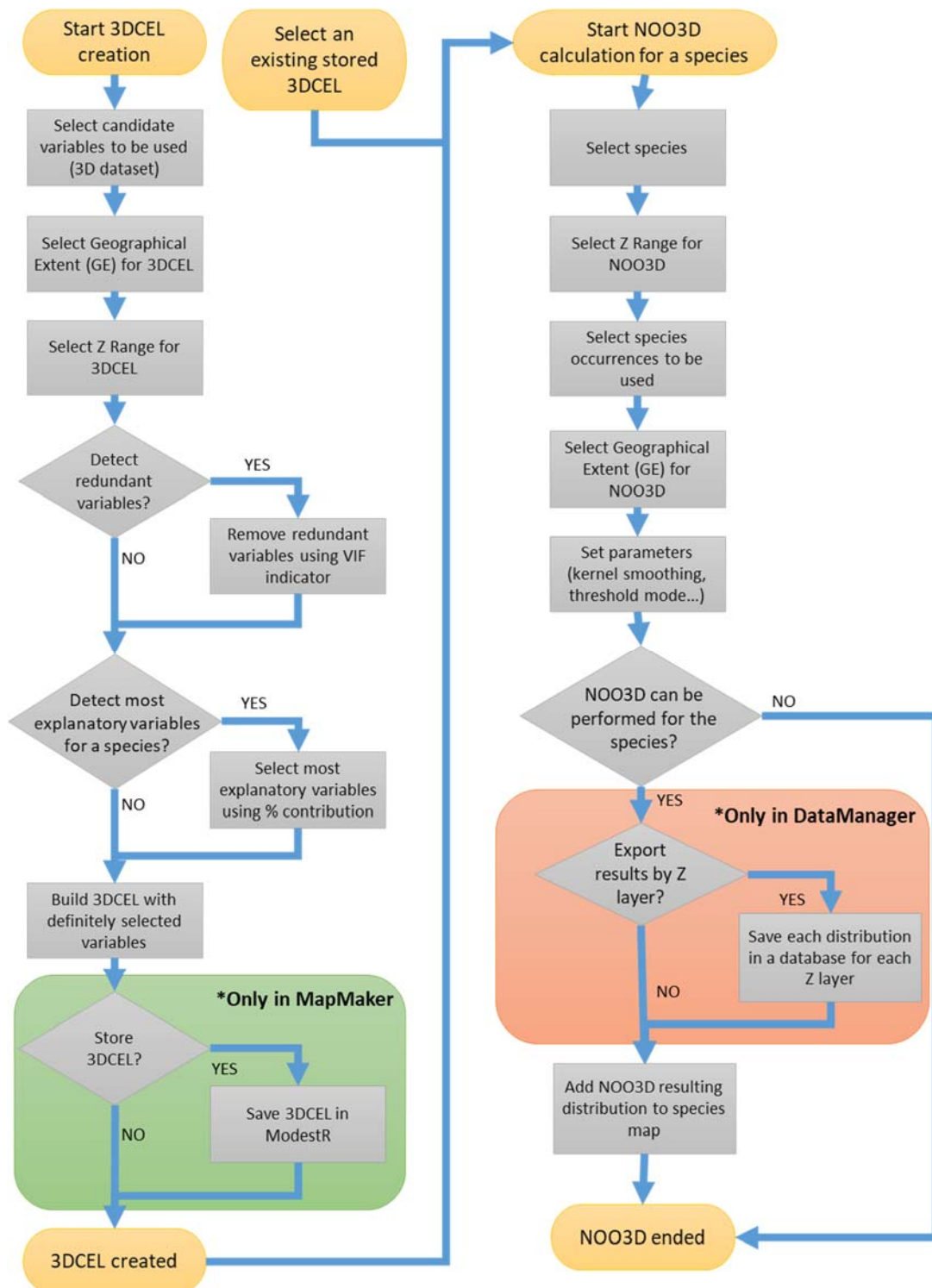
If all those criteria are fulfilled the folder will be converted to a 3D variable. This is indicated by a different icon, as shown in the next figure:



## 2. Niche of Occurrence 3D

Once integrated the wished set of variables in ModestR and defined the 3D datasets, the NOO3D can be done in two ways: interactively and for an individual species in MapMaker (see next section); or in batch mode and for any set of species in DataManager (see section 4).

An overview of the process is shown in the following workflow:



As it can be seen, the process involves two distinct subprocesses:

1. Building a 3D compounded environmental layer, or 3DCEL, which is an artifact that allows to define an environmental space using any set of variables. It should be noticed that a 3DCEL is independent from any species, as it only includes data from environmental variables. But it can be built ad-hoc for a specific species, for example using as spatial range the EOO of the species, or selecting the most explanatory variables for the species distribution to create the 3DCEL.

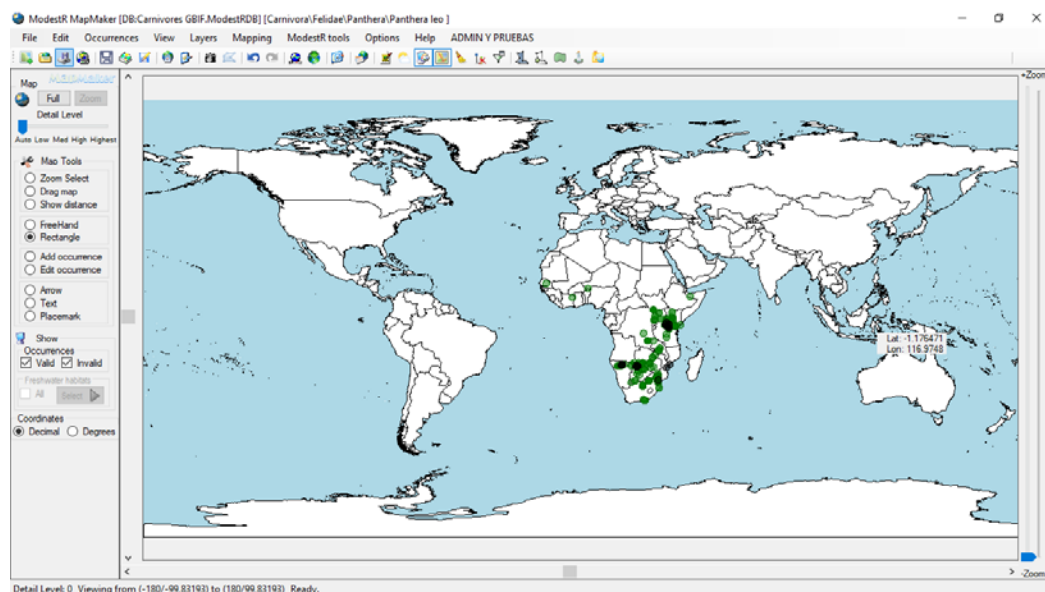


2. Calculating the NOO3D of the species, using a previously created 3DCEL as the environmental space of reference. This step properly involves species data to calculate its potential distribution in several layers. For example, it can be calculating the distribution of a species along several years (Z layers=years), or at different depths in a marine environment (Z layers=depths).

Both subprocesses can be performed sequentially, in particular when we want to create a specific 3DCEL for a species in order to next calculate its NOO3D. Or they can be performed independently, thus creating first a 3DCEL and storing it to use it later to calculate the NOO3D of one or more species. Most of the steps are the same either in MapMaker or DataManager. The main difference is that MapMaker is aimed to work in one species at a time and in an interactive way, where each step is done and the user can see its results. DataManager, in turn, is aimed to work with databases of hundreds or even thousands of species and allows launching batch processes involving any number of species, where main parameters are set once for all species and results should be reported at the end. In the next sections both options are explained.

### 3. Niche Of Occurrence 3D (NOO3D) calculation in MapMaker

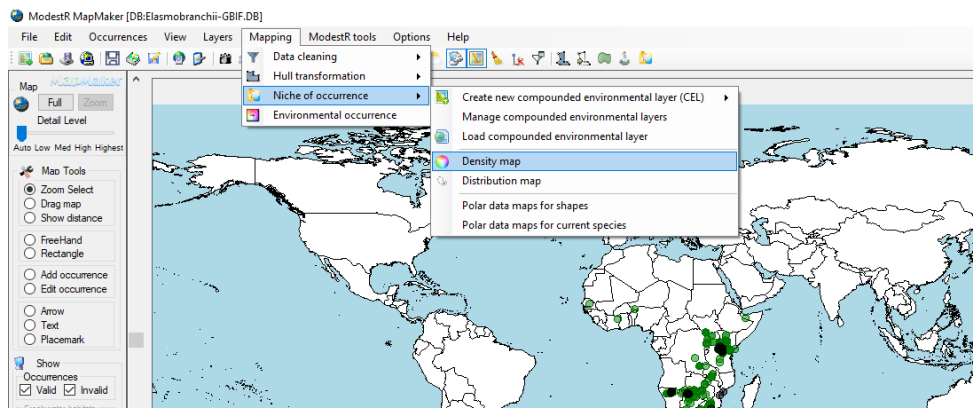
An NOO3D in MapMaker will require to have occurrence data about a species previously loaded, such as from GBIF, from a CSV file or from a preexistent ModestR database.



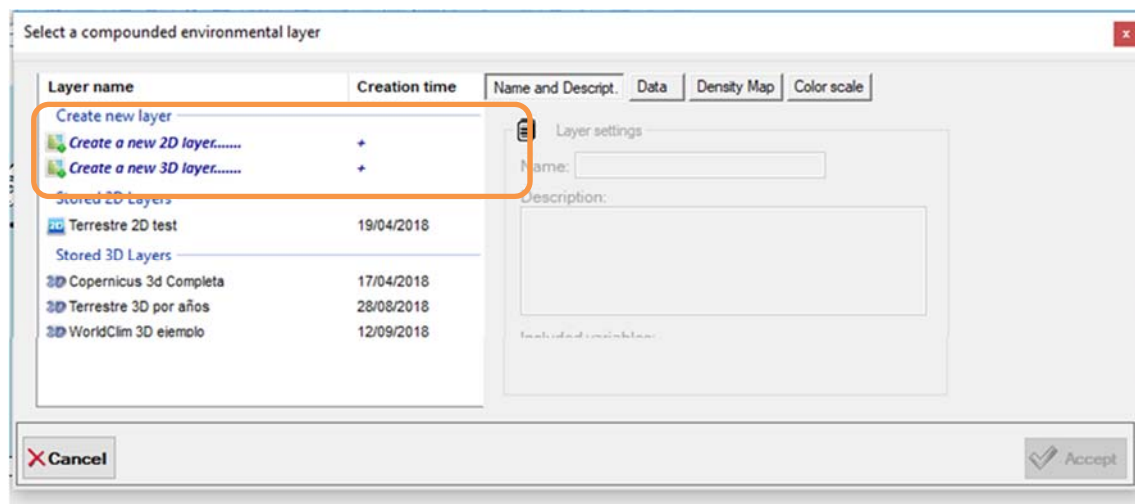
Then we can use the option *Mapping/Niche of occurrence/Density map*<sup>2</sup>.

---

<sup>2</sup> This option allows us to select an existing 3DCEL or to build a new one, and subsequently continue with a NOO3D. A 3DCEL can also be created independently from a NOO3D calculation using option *Mapping/Niche of occurrence/Create new compounded environmental layer/3D CEL*. In this case the 3DCEL can be visualized and stored for later usage, such as for an NOO3D calculation.



In the dialog that will appear, the list of stored 2D and 3D CEL's<sup>3</sup> is shown. Of course, the first time you use NOO2D or NOO3D, no stored layer will be listed. But on the top of the list two options allow to create a new 2D or 3D CEL.



For the NOO3D a 3DCEL is needed so we'll select the 3DCEL creation option. The procedure of creation is described in the next section.

### 3.1. Compounded Environmental Layer (3DCEL) creation in MapMaker

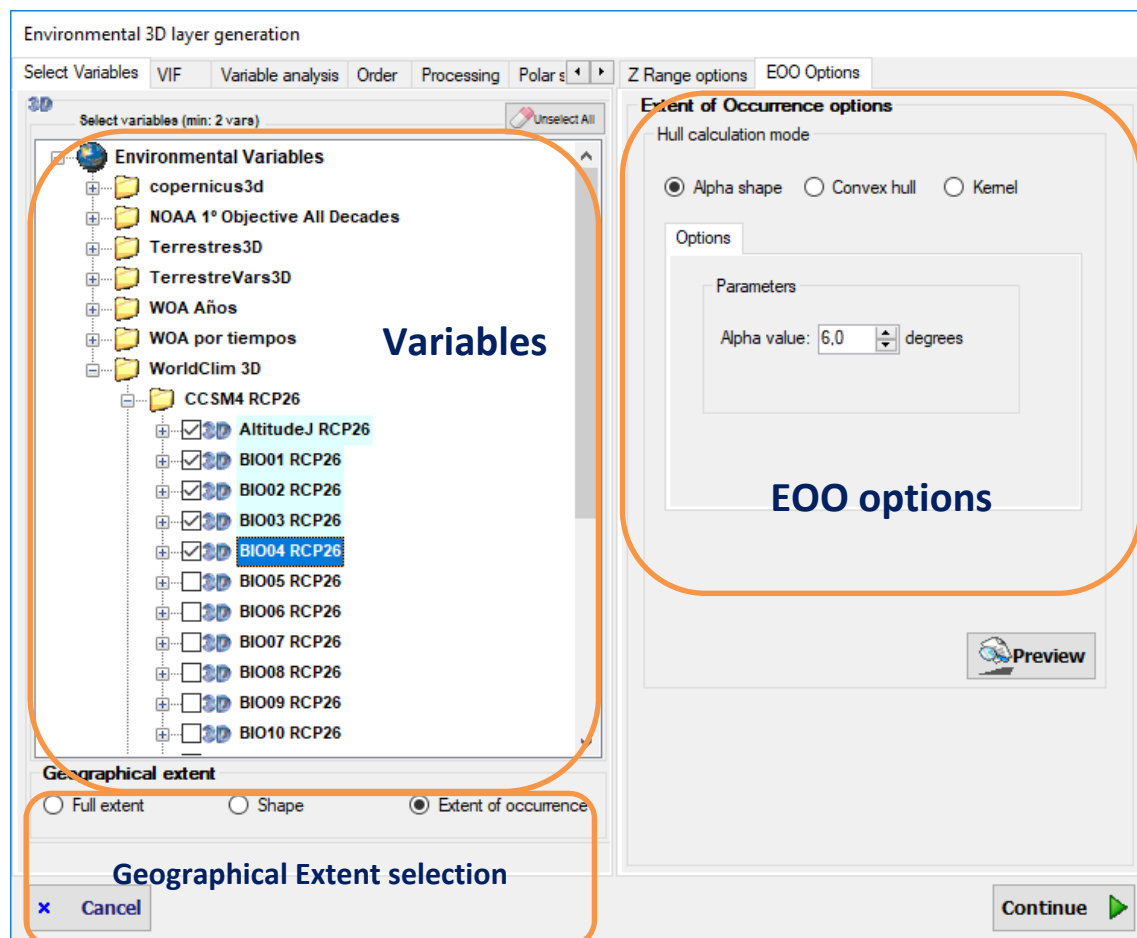
The first step to estimate the distribution of species by means of the NOO3D approach starts with the delimitation of the Geographical Extent (GE) to be used as working area. Several methods are available, such as the Extent of Occurrence (EOO) that can be calculated as an alpha shape, a convex hull, 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

<sup>3</sup> CEL: compounded environmental layers; this concept will be explained below

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, and 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.

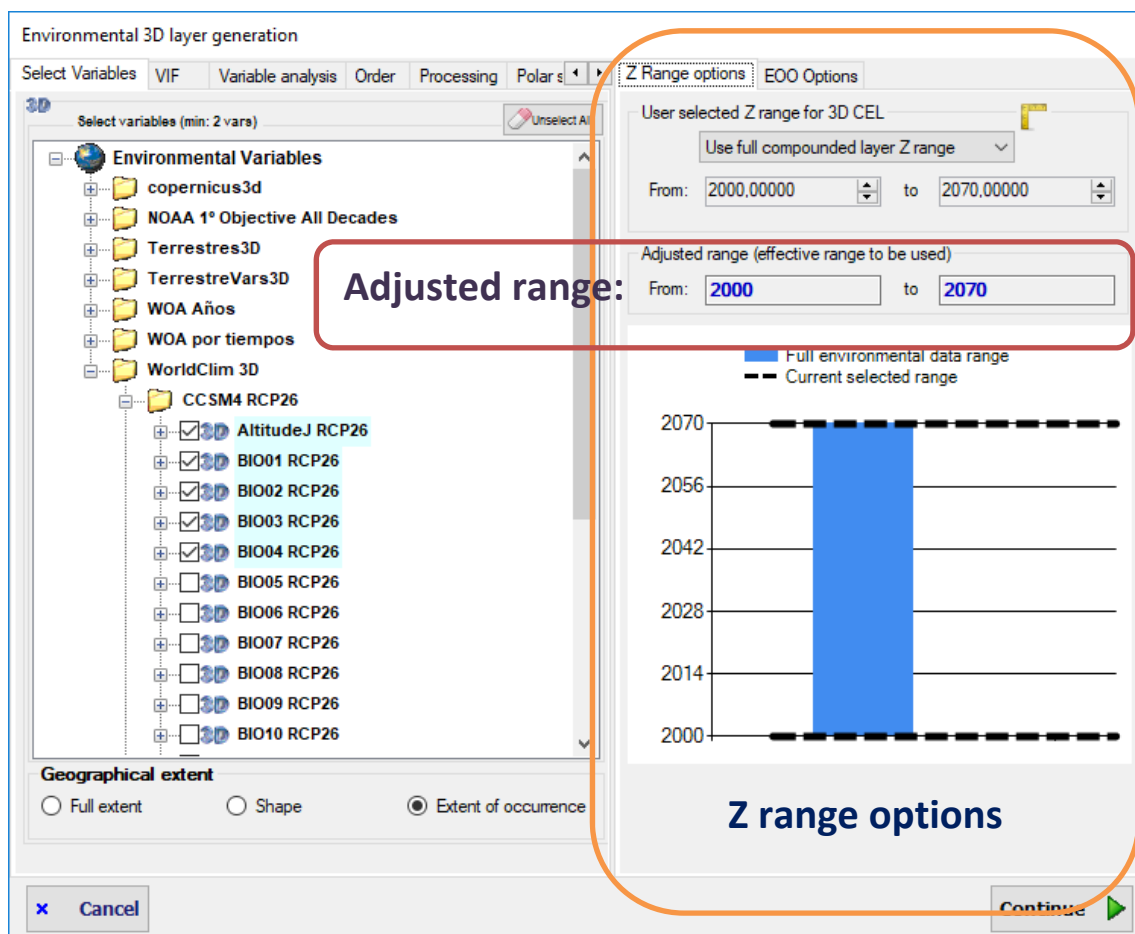
The available options are:

- 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,
- 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).
  - 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.

**Environmental 3D layer generation**

Select Variables | **VIF** | Variable analysis | Order | Processing | Polar space | Options

**Correlation and variable redundancy analysis**

Selected Variables	%MinCov	VIF
AltitudeJ RCP26	99,87%	1,94
BIO01 RCP26	99,33%	9,52
BIO02 RCP26	99,33%	3,22
BIO03 RCP26	99,35%	6,23
BIO04 RCP26	99,33%	10,05

**Colinearity analysis**

☒ Full extent
 ☐ Shape
 ☐ Extent of occurrence

Calculate VIF
 Delete selected

Cancel
 Previous
 Continue

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 wish 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 taken 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). In 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). Anyway, this step is optional and we can just go to the next step without analyzing contribution neither discard variables regarding its contribution value.

Environmental 3D layer generation

Select Variables VIF Variable analysis Order Processing Polar space Options

**Variable contribution analysis**

Select Z type:  
Year

Z as year

Occurrence selection  
☐ Use only occurrences with year value  
☒ Use all occurrences

Z Value assignment  
☒ Use occurrences Z value if exists  
 For occurrences without Z value...  
☐ Use mean year  
☐ Use environmental variable means  
☒ Use custom Z value: 2000

Z (year) range for contribution calculation  
 Use full compounded layer Z range  
 From: 2000.00000 to 2070.00000  
 Range adjustment mode: Strict

Layer selection  
☐ Use only layers with occurrences  
☒ Use all layers  
☐ Filter variables by habitats

Selected Variables	%Contrib.
AltitudeJ RCP26	34,05%
BIO01 RCP26	31,11%
BIO04 RCP26	26,53%
BIO03 RCP26	8,31%
BIO02 RCP26	0,00%

% of contribution of each variable.  
Selected variables are highlighted  
after contribution analysis  
performed

Calculation range  
☒ Full extent ☐ EOO ☐ Shape

Input Precision: 1 degree

Num. ranges: 30

☒ Select best variables for a contribution of: 80 %

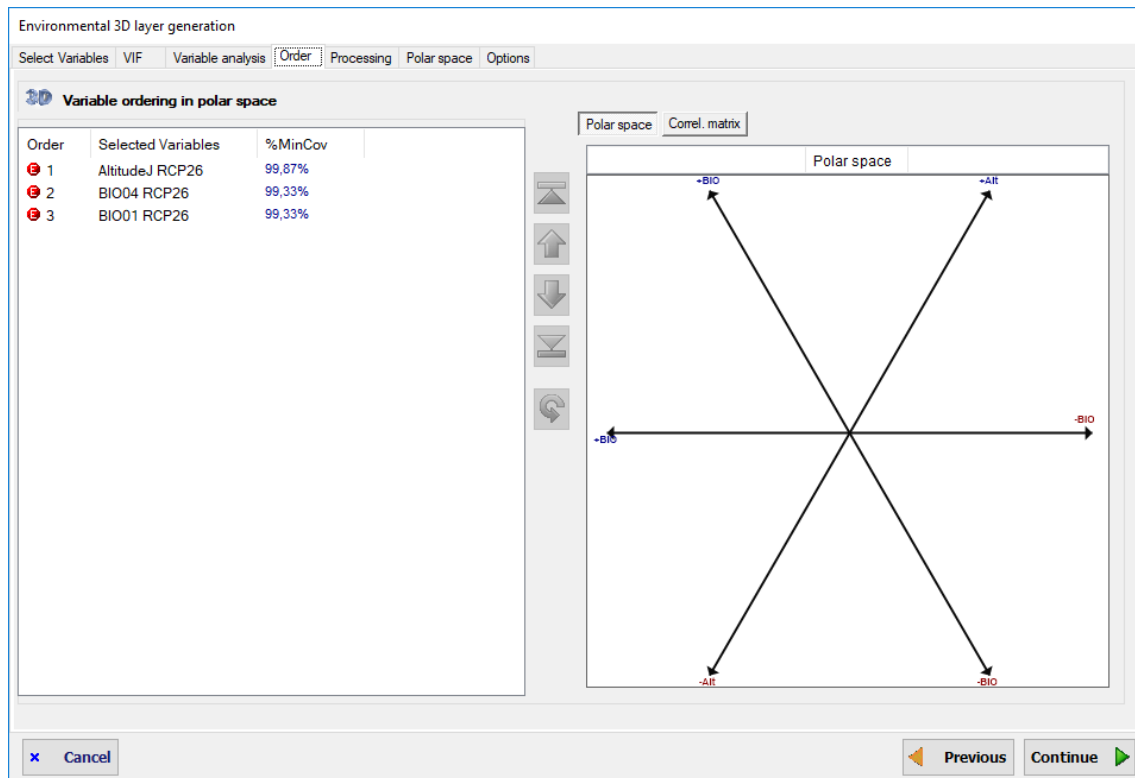
Analyse

Cancel Previous Continue

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



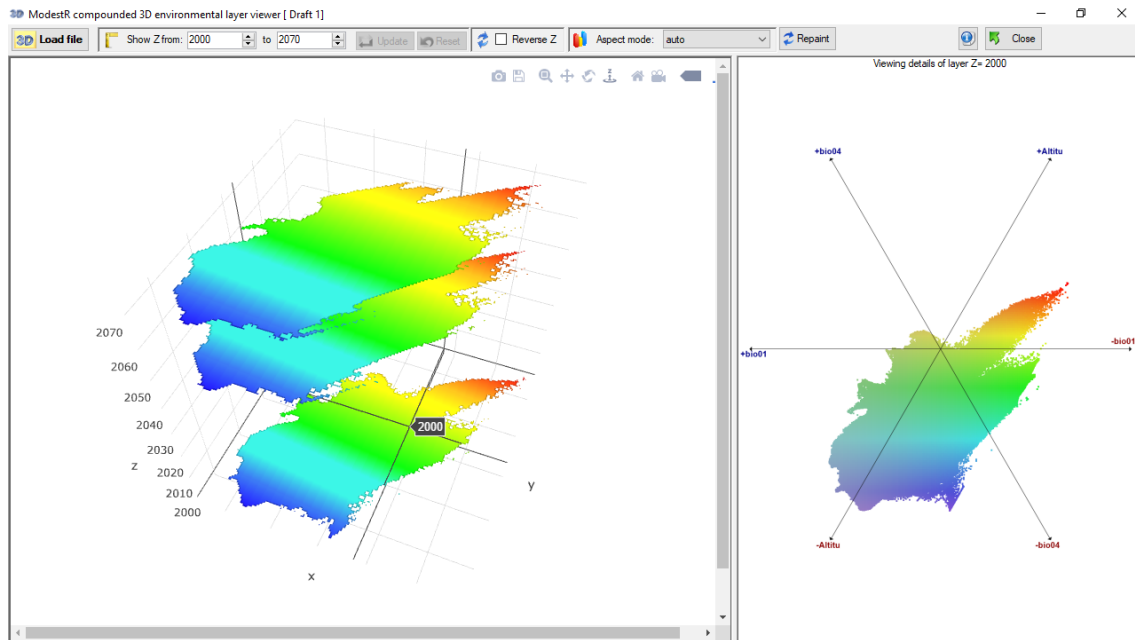
is the one which has the higher correlation with it. You can eventually manually change order by selecting a variable in the list and using the buttons located on the right of the list.



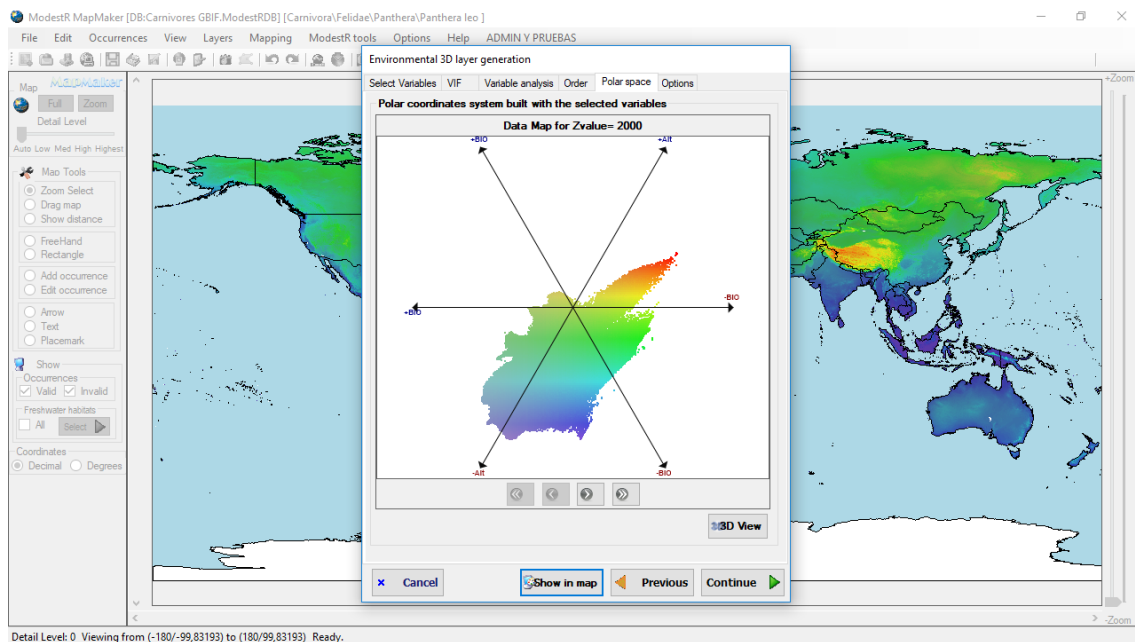
In this step, environmental layer will be calculated according to entered settings. This can take some time. To build the polar coordinate system the range of each variable is calculated taking all layers of this variable. For example, if Z is depth, every available layer of temperature for each depth 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 *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 an environmental layer, 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).

**Environmental 3D layer generation**



Select Variables   VIF   Variable analysis   Order   Polar space   **Options**




Options

 Layer identification

3D Layer name:

Layer description:

 Export report    Save 3D layer

 Cancel    Previous   Continue 

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.

### 3.2. NOO3D calculation in MapMaker

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.

Whether we are using a just generated 3DCEL or an already stored one, 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 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.

NOO 3D [3DCEL for Panthera leo]

**3D Options** Kernel options

**Select Z type:**  
Year

Z as year

**Occurrence selection**  
☒ Use only occurrences with year value  
☐ Use all occurrences

**Z Value assignment**  
☒ Use occurrences real year value  
For occurrences without year value...  
☐ Use mean year  
☐ Use environmental variable means  
☐ Use custom Z value: 2000

**Z (year) range for NOO calculation**  
Use full compounded layer Z range  
From: 2000,00000 to 2070,00000  
Range adjustment mode:  
Strict

Close Continue

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 exist. 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 also must 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.

NOO 3D [3DCEL for Panthera leo]
3D Options
Kernel options

Environmental Kernel density map

Information
Total % coverage:

Parameters
Lower Normal Higher
Smoothing: X 1,00

Visual filtering
Range

By habitat
By presence area
Show over map
Full layer
Shape
EOO

Threshold mode
Minimal Density at Presence
Optimal sensitivity/specificity

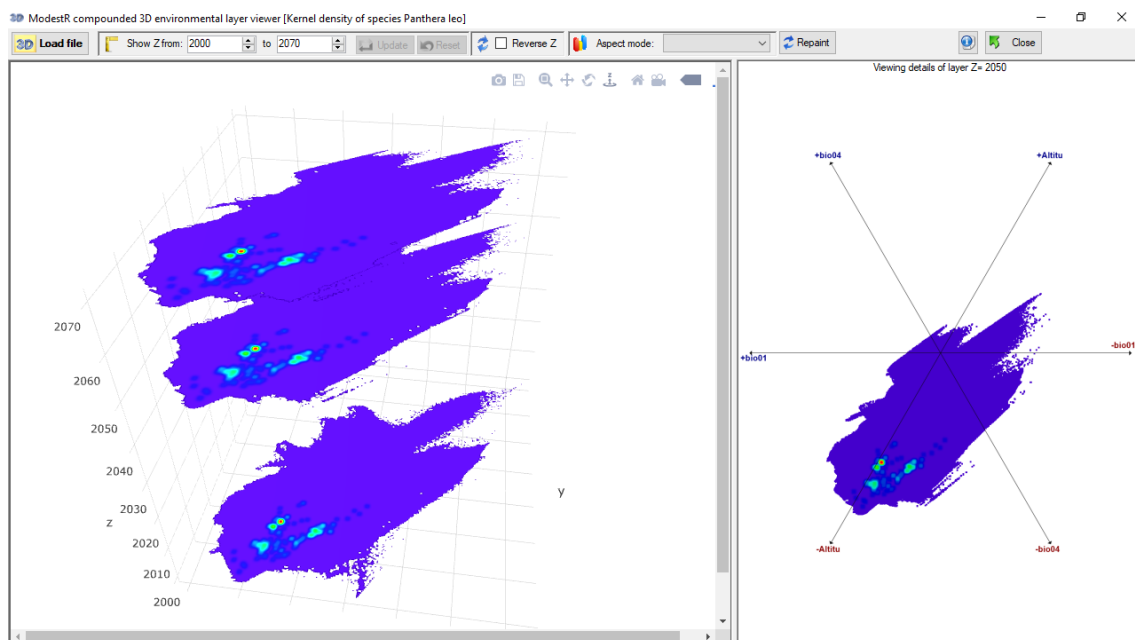
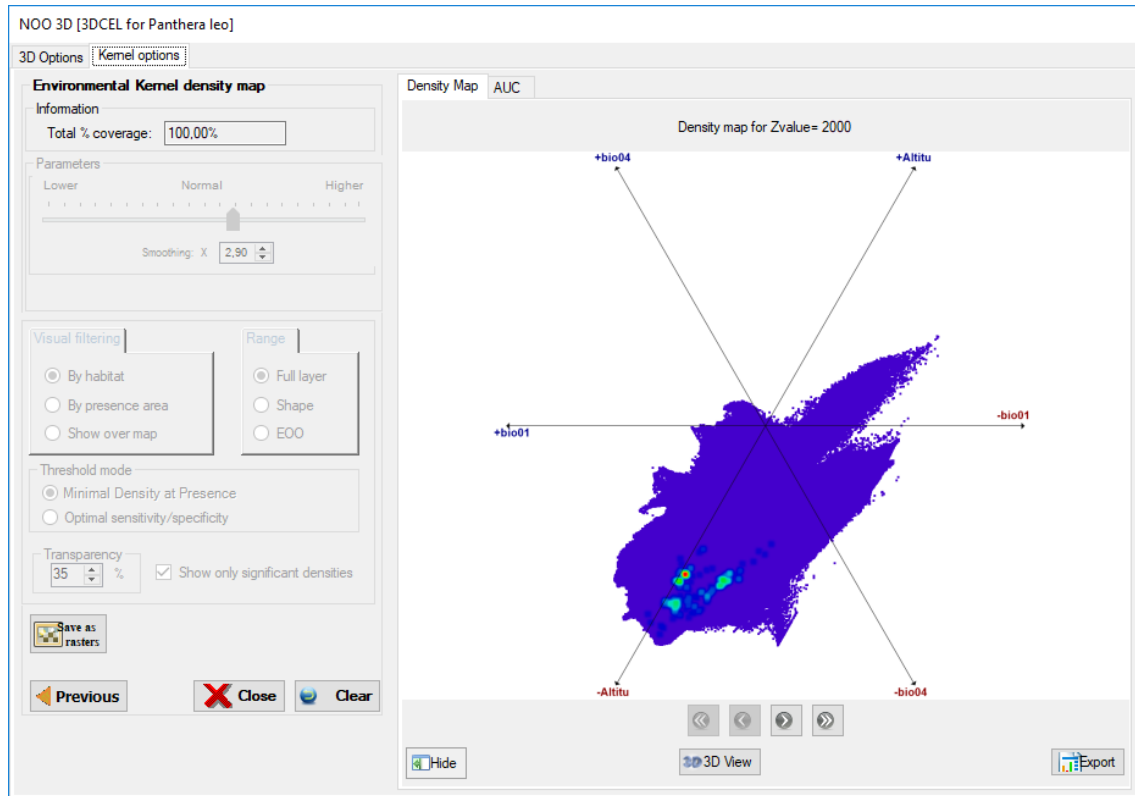
Transparency
35 %
Show only significant densities

Save as rasters

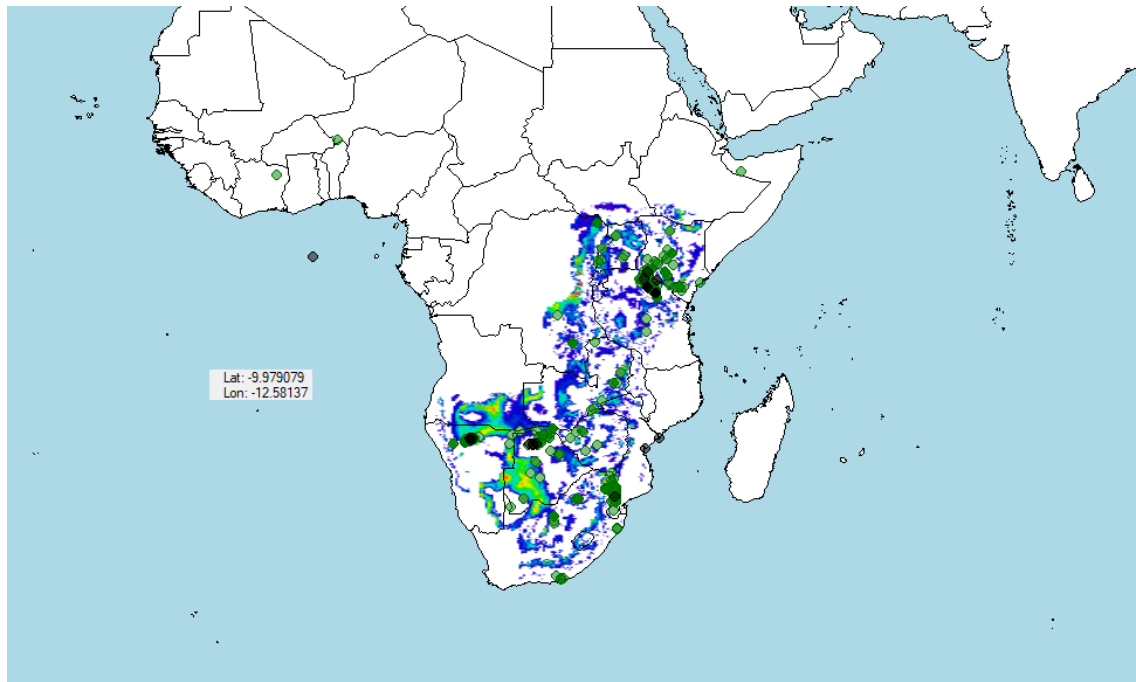
Previous
Close
Run

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

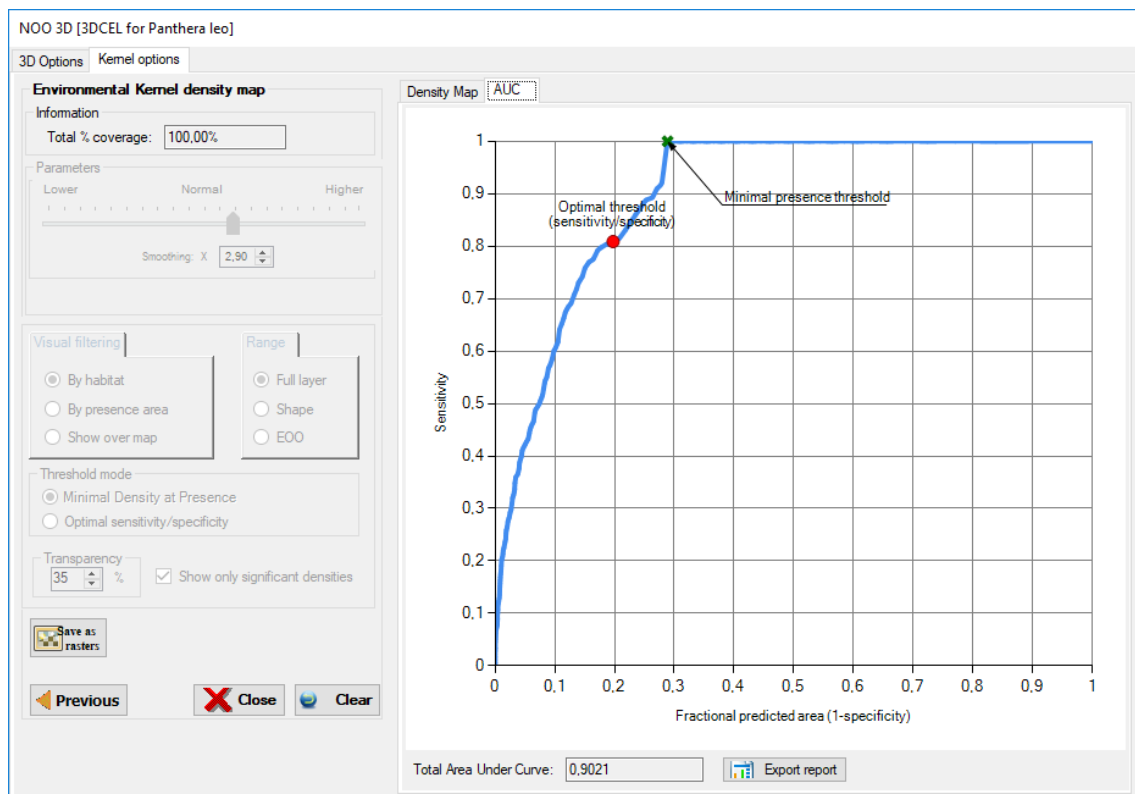
- A polar density map that shows the density (suitability) of the species in the environmental space of the 3DCEL. There is a map for each Z level, and a 3D view is also available.



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

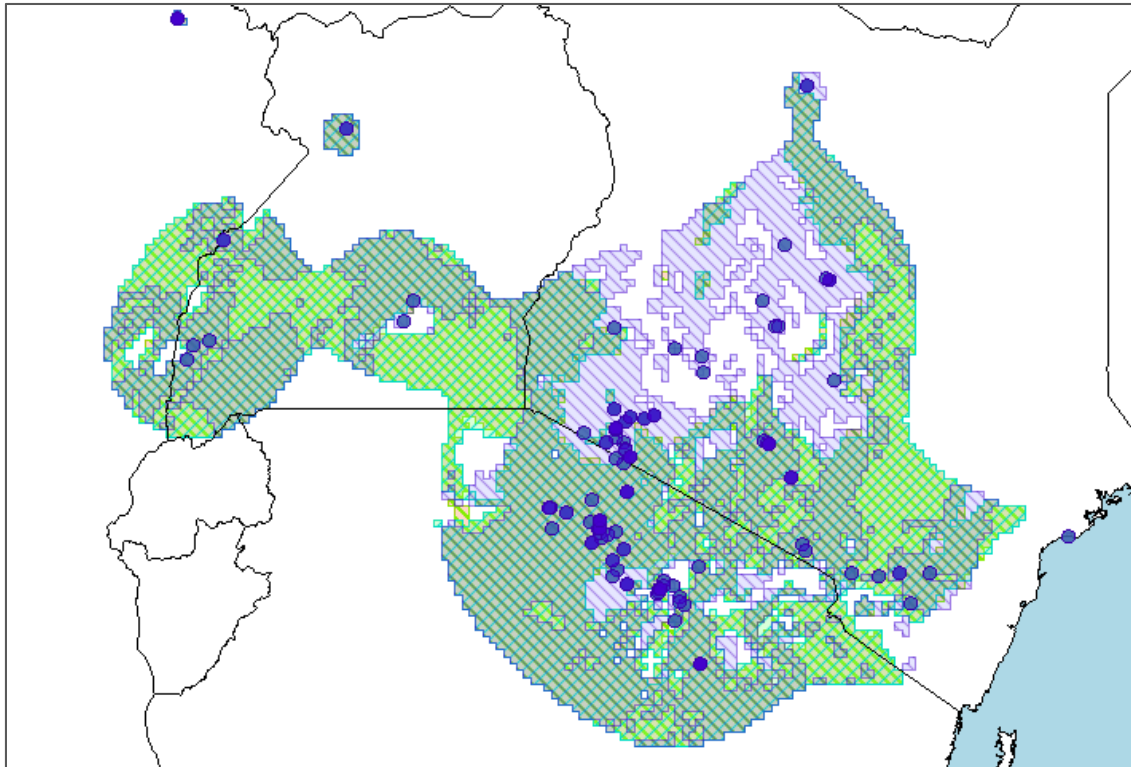


- The ROC curve and the AUC value computed from verified presences and background pseudo-absences (see Phillips et al. 2006 for more details about using ROC in presence-only species distribution context).



- 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.



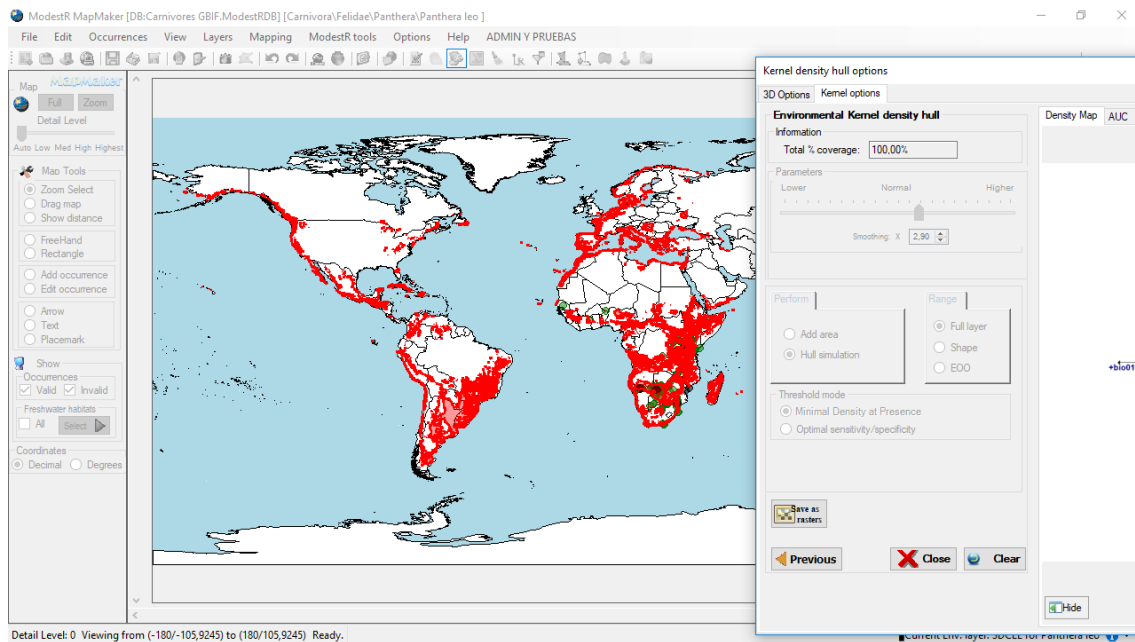
*Detail of a small area where the several distribution layers of the species for each Z layer can be seen in different colors, in MRMapping.*

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.

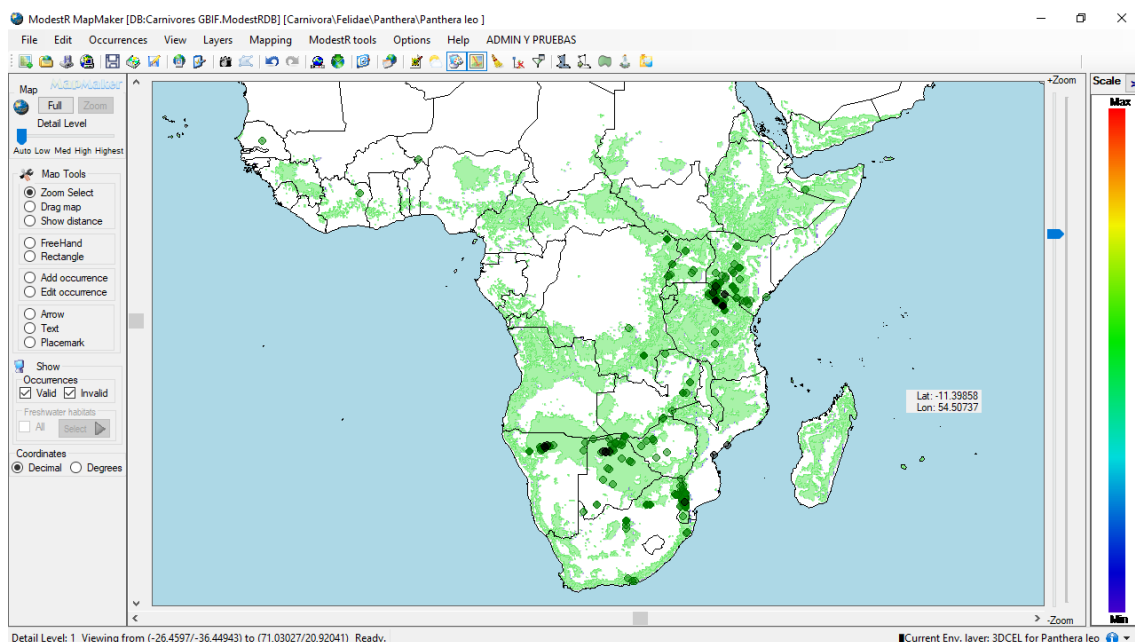
This is the process done using *Mapping/Niche of occurrence/Density map* menu option. As said before this feature provides a density map visually showing the areas where is more/less likely for the species to be present, but no map modification will be done. Using the *Mapping/Niche of occurrence/Distribution 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.

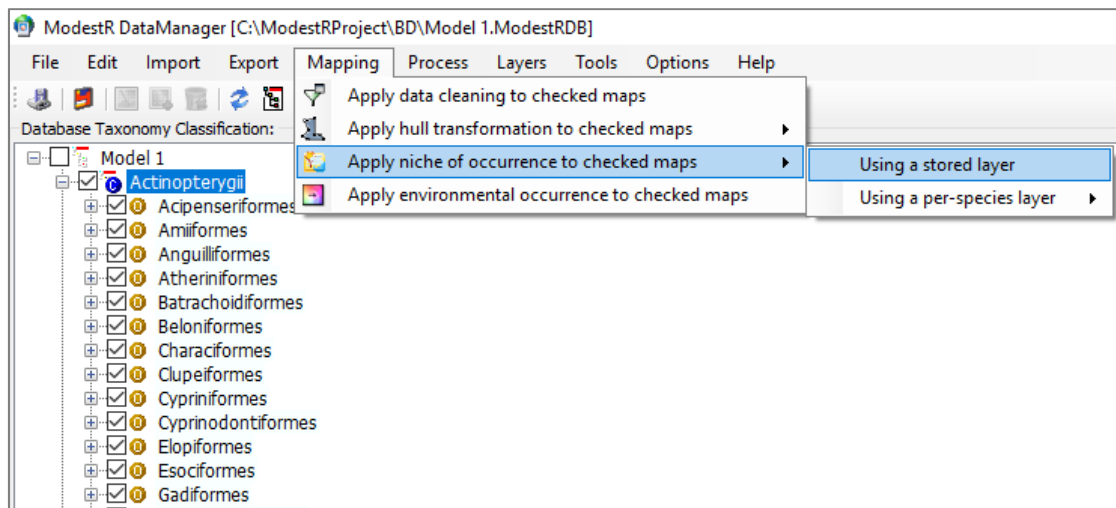
That is why both when calculating density maps or distribution maps in a NOO3D context a multiple-layer distribution map for the species is exported in the format used by MRMapping application. This map contains a distinct presence area for each Z layer. This allows showing several overlapping distribution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

#### 4. Niche Of Occurrence 3D (NOO3D) calculation in DataManager

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<sup>4</sup>, or using a per-species layer, in 3D. Both options are explained in the next sections.



##### 4.1. NOO3D using a stored layer in DataManager

This option is aimed to apply NOO3D to species using a same 3DCEL previously built and stored using MapMaker (see section 3.1). This option may be useful when we previously determined a

<sup>4</sup> 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.

collection of variables that are commonly explanatory for several species, and/or to obtain 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.

Select a compounded environmental layer

Layer name	Creation time
<b>Stored 2D Layers</b>	
Terrestre 2D test	19/04/2018
<b>Stored 3D Layers</b>	
3DCEL for Panthera leo	20/09/2018
Copernicus 3d Completa	17/04/2018
Terrestre 3D por años	28/08/2018
WorldClim 3D ejemplo	12/09/2018

Layer settings

Name:

Description:

Included variables:

Cancel Accept

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 3.2. 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.

Niche Of Occurrence calculation

3D Options Range and parameters Processing

Select Z type:

Z as year

Occurrence selection

☒ Use only occurrences with year value

☐ Use all occurrences

Z Value assignment

☒ Use occurrences real year value

For occurrences without year value...

☒ Use mean year

☐ Use environmental variable means

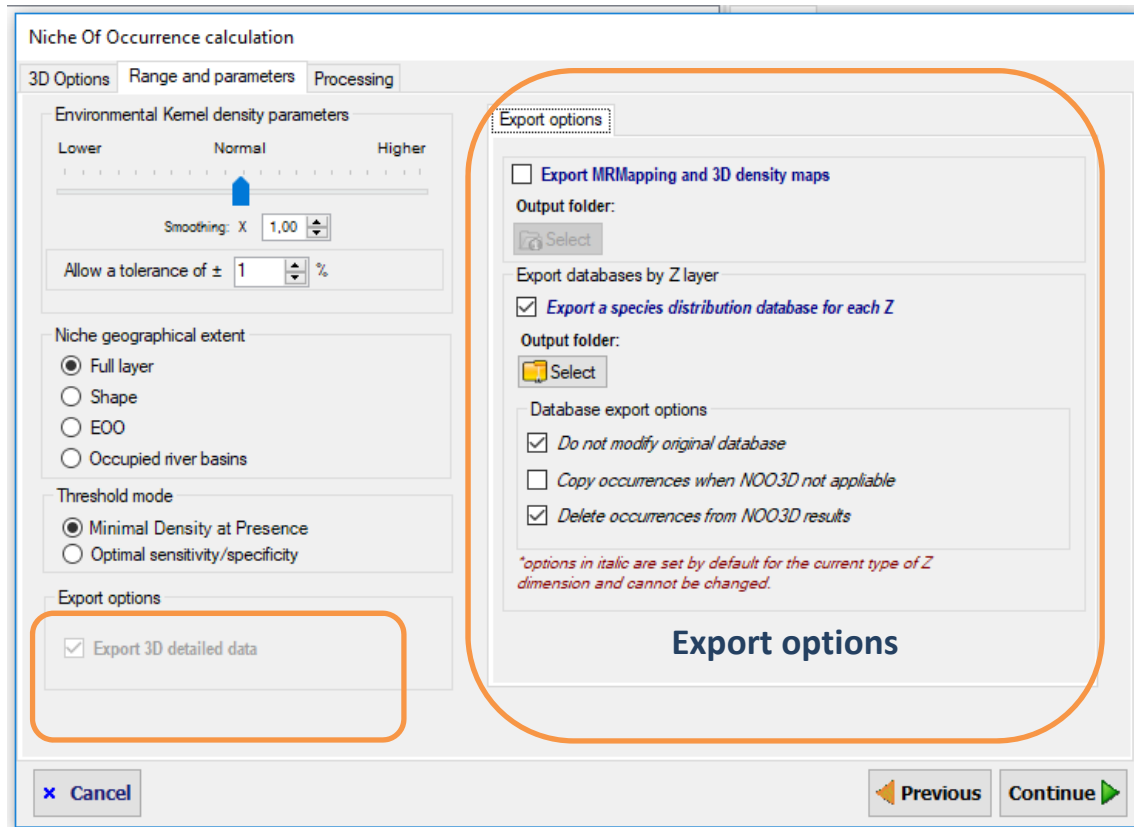
☐ Use custom Z value:

Z (year) range for NOO calculation

Range adjustment mode:

Cancel Last Settings Continue

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 3.2.



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 3.2 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 contain 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

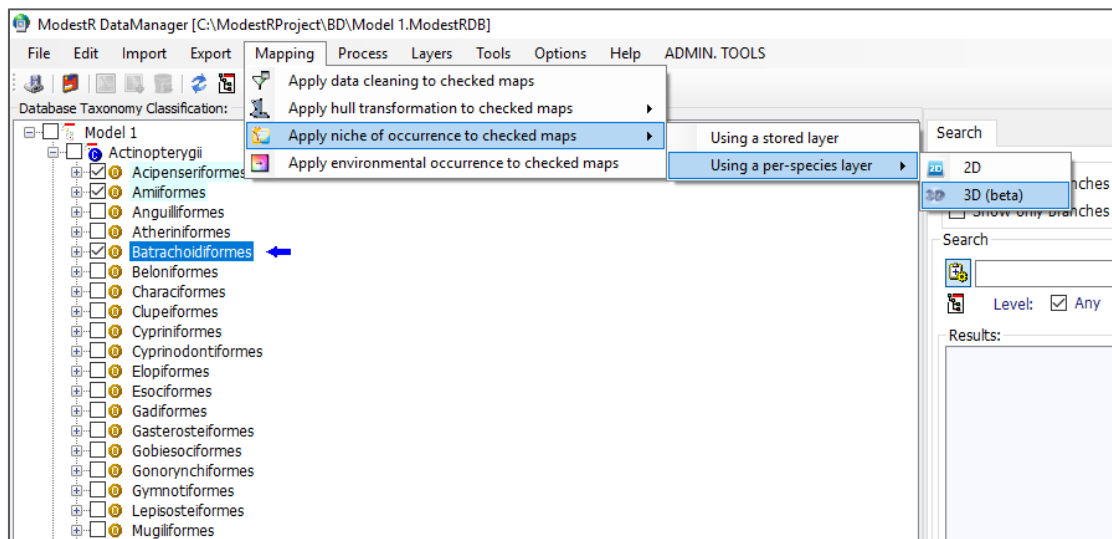
There are several options when exporting a species distribution database for each Z:

- **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 applied 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.

#### 4.2. NOO3D using a per-species layer in DataManager

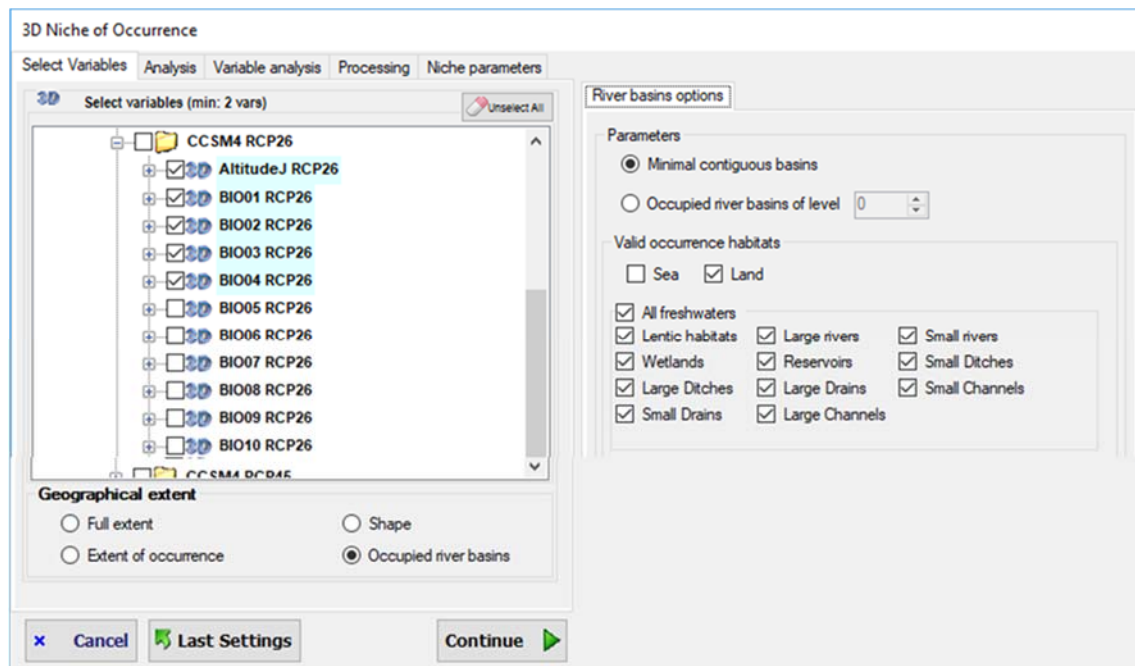
This option is aimed to apply NOO3D to a set of species using a 3DCEL specifically built on-the-fly for each species. This option may be useful when we want to use a collection of potentially explanatory variables, but selecting among them the most explanatory ones for each species, then using only those ones to apply NOO3D. Therefore this feature reproduces the processes

described in sections 3.1 (creating a 3DCEL) and 3.2 (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 3.1. (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 3.1. 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 3.1. 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.

3D Niche of Occurrence

Select Variables Analysis **Variable analysis** Processing Niche parameters

**Variable contribution analysis**

Select Z type:  
Year

Z as year

Occurrence selection  
☒ Use only occurrences with year value  
☐ Use all occurrences

Z Value assignment  
☐ Use occurrences real year value  
 For occurrences without year value...  
☐ Use mean year  
☐ Use environmental variable means  
☒ Use custom Z value: 2000

Z (year) range for contribution calculation  
 Use occurrence Z values range

Range adjustment mode:  
 Strict

Layer selection  
☒ Use only layers with occurrences  
☐ Use all layers  
☐ Filter variables by habitats

Selected Variables  
 AltitudeJ RCP26  
 BIO01 RCP26  
 BIO02 RCP26  
 BIO03 RCP26  
 BIO04 RCP26

Variable contribution analysis  
☒ Full extent ☐ Shape  
☐ River basins ☐ EOO

Input Precision: 1 degree

Num. ranges: 30

Select best variables for a minimal contribution of: 80 %

Cancel Previous Continue

The parameters to be set in this step are already explained in section 3.1. 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 with those variables, then use it to calculate NOO3D for any set of species using the option described in section 4.1 (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



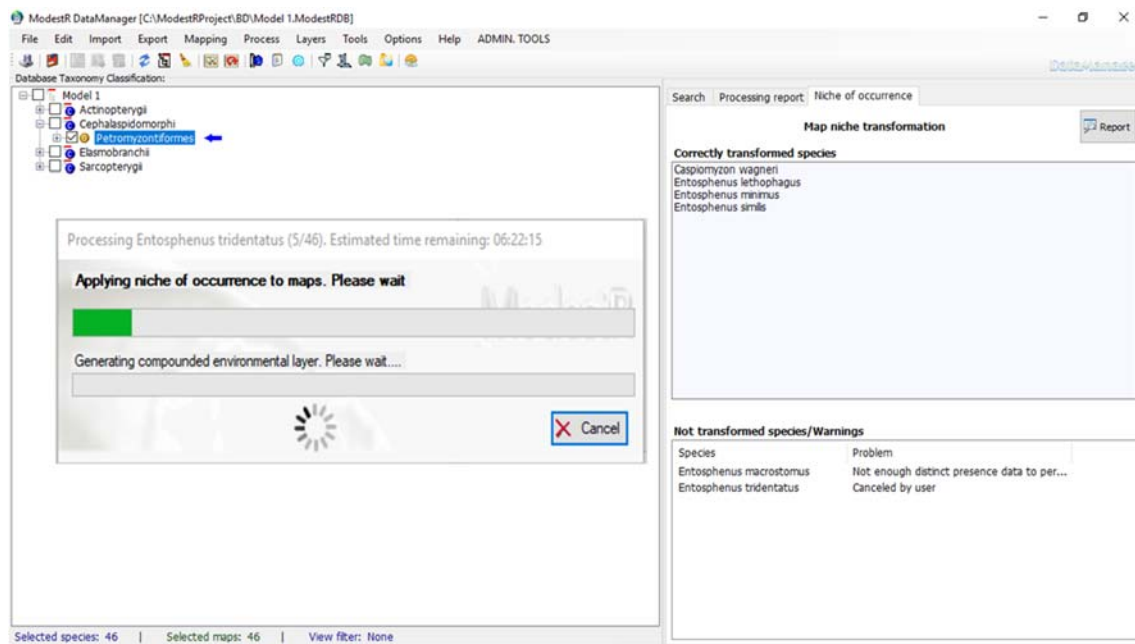
NOO3D, but applying it to several species in an easy way, so without having to do it species by species as it should be done in MapMaker.

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 4.1.

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<sup>5</sup> 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.

<sup>5</sup> 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



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.Percent	BIO03 RCP26.Contrib.Percent
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 Species	AUC							
9 Caspiomyzon wagneri	0,9615							
10 Entosphenus lethophagus	0,9803							
11 Entosphenus minimus	0,966							
12 Entosphenus similis	0,9461							
13 Correlation Matrix								
14 AltitudeJ RCP26	BIO01 RCP26	BIO02 RCP26	BIO03 RCP26	BIO04 RCP26				
15 AltitudeJ RCP26		-0,19	0,19	0,02	-0,04			
16 BIO01 RCP26			0,53	0,83	-0,81			
17 BIO02 RCP26				0,39	-0,2			
18 BIO03 RCP26					-0,89			
19 BIO04 RCP26								

## 5. References

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

Guisande, C., García-Roselló, E., Heine, J., González-Dacosta, J., González Vilas, L., García Pérez, B. J. and Lobo, J. M. (2017) SPEDInstabR: An algorithm based on a fluctuation index for selecting predictors in species distribution modeling. *Ecological Informatics*, **37**:18-23.

Pérez-Costas E, Guisande C, González-Vilas L, García-Roselló E, Heine J, González-Dacosta J, Lobo JM (2019) NOO3D: A procedure to perform 3D species distribution models. *Ecological Informatics* 54:101008.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**:231-259.