BioScore: User-manual of post-processing BioScore output

BioScore Structure

The BioScore script has a post-processing script, which performs postprocessing of the range maps calculated with the core model. These range maps include probabilities of occurrence based on the suitability of the species occurrence for the environmental conditions included in the core-model. The folder Post-processingSDMs contains all functions used for post processing the outputs generated with the core model. It contains a total of 5 core functions (Table 1). PostProcessingSDMs_system_settings is the script required for the system settings storing all the pathways and the system requirements. The PostProcessingSDMs_model_settings script is the script which contains all the specifications required by the model. By running the script 00_Initiate_PostProcessingSDMs.R the model can be ran.

Table 1: Table with scripts and their description

SCRIPTS	DESCRIPTION
Post-processingSDMs	
00. Initiate_PostProcessingSDMs	Script to initiate the full run of post-processing the BioScore output
01. PostProcessingSDMs	Script which combines all the below mentioned post-processing functions in order to run the PostProcessingSDMs model
02. RemoveUnsuitableAreas	Functions to mask unsuitable or unoccupied areas from the species range map
03. Binarisation	Functions to transform the predicted probability to binary presence-absence predictions
04. DispersalAssumptions	Functions to apply three dispersal assumptions to the future range maps, specifically no, realistic and unlimited dispersal
05. SpeciesRichness	Functions to calculate the species richness for a given scenario
PostProcessingSDMs_model_settings	Parameters to save all the system settings defined per user and model run

PostProcessingSDMs_system_settings	arameters to save all the model settings per user and model run

System requirements

BioScore runs in RStudio (version 2023.12.0+369). The PostProcessingSDMs script is usualy run in parallel on a machine with 4 CPU and 8 GiB or 8 CPU and 16 GiB for no and unlimited dispersal assumtions and 8 CPU and 64 GiB for the realistic dispersal assumption.

Model Overview and usage

To run Bioscore follow the steps below.

Step 1: Setting up file pathways

Work in the same directories as the input and output of the Core-model are stored, in order to keep all information related to the model runs together (Figure 1, README_Core).
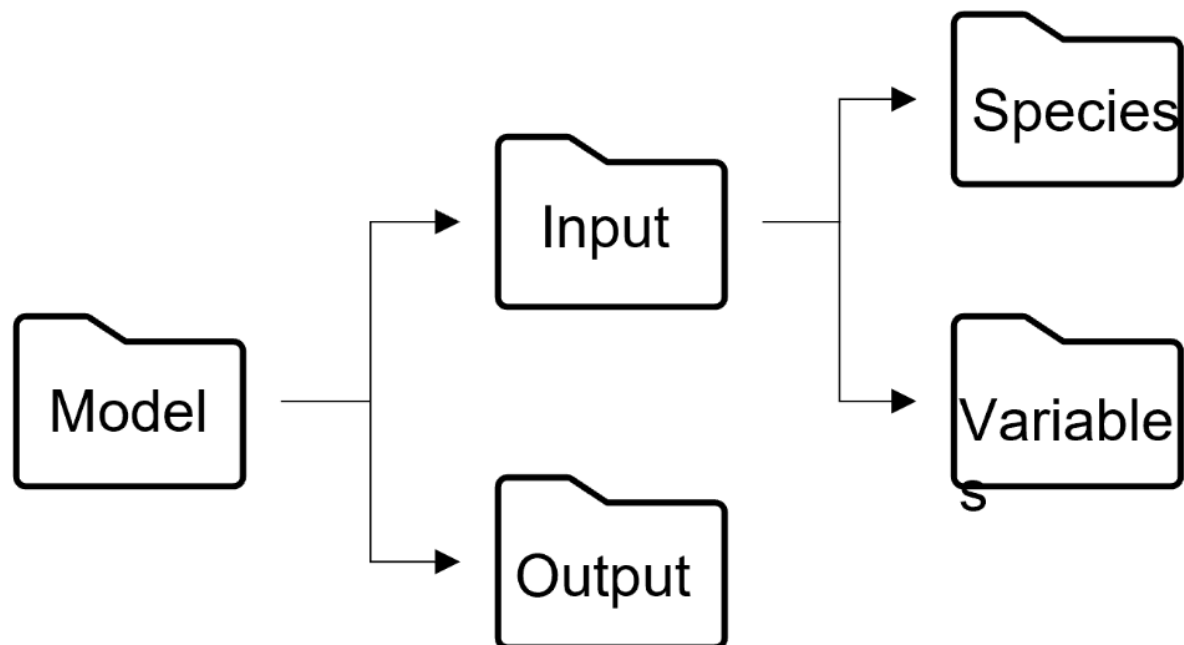

Figure 1: Directory names and structure for BioScore run

Step 2: Save the Input data to the file pathways

If the realistic dispersal assumption is used, than in the Input folder a table should be stored with the dispersal rates of all species. If the range maps will be clipped to the realised range, than the Input-folder should contain a folder that includes a raster file per species with the realised range. If the unsuitable land covers are masked from the range maps than a raster containing the unsuitable land covers should be stored in the Variables folder within the Input folder. The species range maps that will be post processed should be located in the Output folder.

Step 3: Set up the system and model settings

1. Set directories in Initiate Post processing-file

Set the file paths in the 00_Initiate_PostProcessingSDMs-file (Table 2).

Table 2: Table with directories and their description

Directory	DESCRIPTION
user_dir	directory within which all subdirectories are located
base_dir	subdirectory of user.dir, where all input data is stored and all output data can be written.
github_dir	directory of your local github repository of the BioScore scripts.

2. System settings

BioScore system configuration script, i.e. PostProcessingSDMs_system_settings.R contains all the pathways and directories for the PostProcessingSDMs script to run and find the input data (Table 3).

Table 3: Table with directories and their description

Directory	DESCRIPTION
AllMetrics_file	This is a metric file containing all the metrics from the fitted SDMs for all the species for which the SDMs were run. This file is generated in the core-model.
LU_map	optional. Land use map that is used to mask unsuitable land uses from the range maps.
species_out_dir	subdirectory of base.dir, in which all output files are stored. Always include the name of the script followed by date in the name of the output directory.
SDM_predictions	a folder containing all the range maps created per scenario as produces in the core-model.
Dispersal_data_file	optional. a table containing per species the dispersal rates.

3. Model settings

BioScore model settings, i.e. PostProcessingSDMs_model_settings.R contains all the model parameters which determine the parameters for the model runs (Table 4).

Table 4: Table with parameters and their description

Parameter	DESCRIPTION
taxo_group	Taxon group used to select species, name should matche the taxon groups mentioned in the species input data
BS_module	Either 'terrestrial' or 'wetlands'
Name	name of the person who runs the script. This is saved to the logfile.
ProjectScenarios	names of climate scenarios to be run. Scenario names should be identical to the names used when generating the range maps with the core model. Note, if DispersalAssumption is set to no or realistic dispersal, then the order of the climate scenarios in ProjectScenarios matters. In this case the current scenario should listed before the future scenarios, e.g. c("current", "ssp126").
ClipToRealisedRange	logical. whether to clip the range maps under current climate conditions to the realised range
LUMask	Set to TRUE if you want to mask unsuitable land uses from the range map
SetValuesUnderThresholdToZero	Logical. parameter for binarisation of the PoOs. Set values under cut off value to zero. The values set to SetValuesAboveThresholdToOne and SetValuesUnderThresholdToZero should be equal, in order to binarise the SDM outputs. Note, if the dispersal assumption is set to realistic dispersal, than is it highly recommended to binarise the PoOs across all scenarios in order to make the outputs across scenarios comparable.
SetValuesAboveThresholdToOne	Logical. parameter for binarisation of the PoOs. Set values above cut off value to one. The values set to SetValuesAboveThresholdToOne and SetValuesUnderThresholdToZero should be equal, in order to binarise the SDM outputs. Note, if the dispersal assumption is set to realistic dispersal, than is it highly recommended to binarise the PoOs across all

	scenarios in order to make the outputs across scenarios comparable.
DispersalAssumption	options: "no dispersal", "unlimited dispersal" or "realistic dispersal". Note, with the realistic dispersal assumption for fewer species output will be generated than for the current situation and for the other two dispersal assumptions, because of a lack of dispersal rate estimates for some species.
EvaluationMetric	either "TSS" or "ROC". Performance metric to set a minimum value to the quality of the range maps generated in the core model.
MinimalMetricValue	minimum value. related to the EvalutionMetric used.
BinarizationMetric	either "TSS", "DSS", "MCC or "F". Threshold-selection method used to binarise the range maps generated in the core-model. Only applicable if SetValuesAboveThresholdToOne and SetValuesUnderThresholdToZero are set to TRUE.
CPU	number of CPU's to use when running the model in parallel
time_periods	The years representative of the current and future scenarios. Only applicable if DispersalAssumption == "realistic dispersal".

Step 4: Run the Initiate_PostProcessingSDMs script

The Initiate_PostProcessingSDMs script runs all post processing steps. These consist of the steps mentioned below:

1. The model set loads the libraries required. It creates the directories required further to store the output from the model run. And further loads all the core functions required to run the BioScore model.
2. The create log file has functions to open a logfile and write in the logfile.
3. Loading of the species data and species selection is performed to select species with sufficiently performing SDMs.
4. Load the probability of occurrence range map as calculated by the core-model. Steps 4 to 9 are executed per species.
5. Optional: mask the areas from the range map for which the species is known to be absent. This step is only executed for maps representing the current situation.

6. Optional: mask the areas from the range map for which the land cover is known to be unsuitable.

7. Optional: transform the probability of occurrence predictions to presence-absence predictions by setting the values above and below the threshold defined in the core-model to 1 and 0, respectively.

8. Apply a dispersal assumption to the range maps, options are no, realistic or unlimited dispersal. This step is only executed for maps representing a future situation.

9. Calculate the range size.

10. Calculate the species richness.

Output data

Various outputs are created and saved on the hard disk and saved within the base_dir-directory in the folder Model/Output. Below you will find a description of the outputs in the table below (Table 5)

Table 5: Table with output files/folders and ther description

File/Folder	DESCRIPTION
Logfile_date&time	A log file containing log on the model run.
RangeMaps	A folder containing all the range maps created per scenario.
Range sizes	A table containing all range sizes of the range maps.
Species richness maps	Species richness maps per scenario.