BioScore: User-manual of the script for calculating the distribution of vegetation types

BioScore Structure

The BioScore script has a vegetation types script, which aggregates the predicted species distributions into vegetation types. The vegetation types included are 206 habitat types at level 3 of the EUNIS habitat classification, belonging to six broad categories, i.e., 1) coastal vegetation, 2) wetlands , 3) grasslands and lands dominated by forbs, bryophytes or lichens, 4) heathlands, scrub and tundra, 5) forests and other wooded land and 6) inland habitats with no or little soil and mostly with sparse vegetation. The vegetation types script trains random forests (RF) models to classify each vegetation type based on its plant species composition, using observational data from vegetation plots. Next, for each vegetation type the RF model is applied using binarised species distributions predicted with the core of BioScore as input data, to predict the distributions of each vegetation type across Europe. The folder VegetationTypes contains all functions used for predict the distributions of the vegetation types. It contains a total of 9 core functions (Table 1). BS_VegTypes_system_settings.R is the script required for the system settings storing all the pathways and the system requirements. BS_VegTypes_model_settings.R script is the script which contains all the specifications required by the model. By running the script Initiate_BS_VegTypes.R the model can be ran.

Table 1: Table with scripts and their description

Scripts	Description
Initiate_BS_VegTypes	Script to initiate the full run of vegetation types
00. BS_VegTypes	Script which combines all the below mentioned functions in order to run the vegetation types model
01. Model-setup	Functions to set up the model with the given model settings and system settings
02. FormatObservations	Functions to format observations of species to be used to fit the RF models
03. FormatSDMs_forRF	Functions to format the SDM-predictions of species to be used to apply the RF models
04. Fit_RF	Functions to fit the RF models
05. Project_RF	Functions to apply the RF models

| 06. Evaluate_RF	Functions to evaluate the performance of the RF models
07. PostProcessingVegTypes	Function to postprocess the predicted distributions of the vegetation types.
BS_VegTypes_system_settings	Parameters to save all the system settings defined per user and model run
BS_VegTypes_model_settings	Parameters to save all the model settings per user and model run

System requirements

BioScore runs in RStudio (version 2023.12.0+369). The VegetationTypes script is usualy run in parallel on a machine with 32 CPU and 256 GiB.

Model Overview and usage

To run the VegetationTypes script follow the steps below.

Step 1: Setting up file pathways

Work in the same directories as the input and output of the Core-model and PostProcessingSDMs are stored, in order to keep all information related to the model runs together (Figure 1, README_Core).
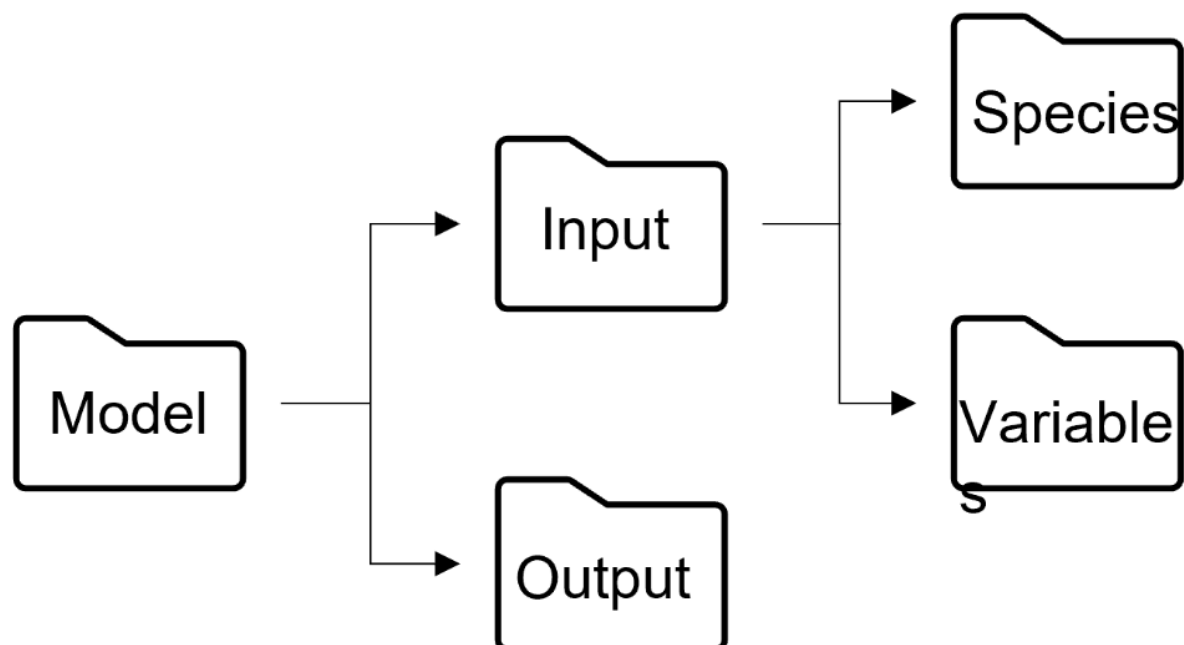
Figure 1: Directory names and structure for BioScore run

Step 2: Save the Input data to the file pathways

Input data has species and vegetation type input data (a file with species and habitat types based on the EUNIS habitat type classification per plot and a file with all plot locations) retrieved from the European Vegetation Archive (EVA; Chytrý et al. 2016) with presence and absence data. It also contains a characteristic species list of distinct habitat types based on the EUNIS habitat type classification. These input files should be identical to the input used in the core-model. Moreover, the binarised species distributions that are used as input to the vegetation types script, should be retrieved from the output-folder of the PostProcessingSDMs script.

Step 3: Set up the system and model settings

1. Set directories in Initiate vegetation types-file

Set the file paths in the Initiate_BS_VegTypes-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. BS_VegTypes_system_settings.R contains all the other pathways and directories for the vegetation types script to run and find the input data (Table 3).

Table 3: Table with directories and their description

Directory	Description
species_in_dir	file located in a subdirectory of base.dir, with .csv file species observations per plot (Fig1).
plots	file located in a subdirectory of base.dir, with all plots derived from EVA-database. This database is used in combination with the files in species_in_dir to determine the absence values (Fig1).
species_special	file located in a subdirectory of base.dir, with list of characteritstic/special group of species.

binarized_SDMs	subdirectory of base.dir, in which all output files are stored as binarised species distribution maps from the PostProcessingSDMs script.
species_out_dir	subdirectory of base.dir, in which all output files are stored. Always include thename of the script followed by date in the name of the output directory. For e.g., BS_VegetationTypes_240122. Do not forget to add the date. (Fig1)

3. Model settings

BioScore model settings, i.e. BS_VegTypes_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, either 'terrestrial' or 'wetlands'
BS_module	Should be set to "VegetationTypes"
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.
DropHabitats_plots	list of characters referring to habitat types which should be removed from the plots, e.g.inland waters, fresh water, marine water and man-made
MinYear	first year of the period for which you want to select the plots
modeltesting	Do you want to run the model for testing purposes only? Setting this to TRUE will only project the vegetation types for Northern Italy, which will speed up the running time, but will also alter model results.

Step 4: Run the Initiate_BS_VegTypes

The Initiate_BS_VegTypes script initiates the 00_BS_VegTypes script, which runs all modelling 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 model.
2. The create log file has functions to open a logfile and write in the logfile.
3. The input data for fitting the RF models are formatted. This includes selecting and formatting plots and species observation data for fitting the RF models.

4. Fitting a RF model per vegetation type, using 80% of the input data of the previous step, and evaluate the performance of these models with the remaining 20% of the data.
5. Fitting a RF model per vegetation type, using 100% of the input data, and use these models predict the distributions of the vegetation types across all scenarios, using the predicted species distributions as input.
6. Post processing of the predicted distributions of the vegetation types, specifically the selection of vegetation types with a good model performance, the binarisation of the distributions and the calculation of the range sizes.

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.
RangeSizes_scenario	Tables containing all range sizes of the range maps of the vegetation types per scenario.
Habitats_CV_AUC	Table with the area under the receiver operating characteristic curve (AUC) of the RF models predicting the occurrence of vegetation types based on the observed species compositions.
Habitats_predictions_AUC	Table with AUC of the RF models predicting the occurrence of vegetation types based on the species compositions predicted with SDMs.

References

Chytrý, M., et al. (2016). "European Vegetation Archive (EVA): an integrated database of European vegetation plots." Applied Vegetation Science 19(1): 173-180. https://doi.org/10.1111/avsc.12191