## Ecography DOI: 10.1111/ecog.01509

## van Proosdij, A.S.J., Sosef, M.S.M., Wieringa, J.J. and Raes, N. 2015.

## Minimum required number of specimen records to develop accurate species distribution models

## Appendix 3 Manual for applying the presented method to given data

Below, a brief summary is given of the methodological steps to identify the minimum number of records required to generate accurate Species Distribution Models (SDMs) using available data. This manual helps readers to use the R scripts in Supplementary material Appendices 4-9 on their own dataset by addressing the input, output and data format requirements of each script. Details on individual methodological steps are provided as comments in the scripts.

### Data preparation – vanProosdij\_et\_al\_2015\_A4\_Datapreparation.r

With this R script environmental data layers are prepared that are used throughout the analyses. Input consists of downloaded \*.asc or \*.bil files with global environmental data, here climate, soil and altitude data. By specifying the extent of the study area in the script, it becomes applicable to the researcher’s own study area and environmental data files are cropped accordingly. A Principal Component Analysis is performed on selected and standardized variables and the first two (or more if desired) PCA axes are exported as \*.asc files that form the output of this data preparation script.

### Simulated species definition – vanProosdij\_et\_al\_2015\_A5\_species\_presence.r

The simulated species definition script generates simulated species using the following input: 1) the two PCA-based \*.asc files with environmental variables from step 1, 2) the mean values for the variables of the bivariate normal response of the simulated species, and 3) the variance-covariance matrix of the bivariate normal response function. The species presence function is called from the main analysis script. The output consists of two \*.asc files with the defined probability of occurrence or habitat suitability scores for each raster cell and the defined presences and absences (1 or 0) for each raster cell.

### Main analysis – vanProosdij\_et\_al\_2015\_A7\_Africanstudyarea.r and vanProosdij\_et\_al\_2015\_A6\_Virtualstudyarea.r

The main analysis scripts run the analyses and call two additional R scripts (simulated species definition script and null model script). Here, main analysis scripts are provided for an African study area and for a virtual study area. The latter is provided for explanatory reasons only. The former can be adjusted to any other study area by manually specifying the extent of the study area. Input consists of \*.asc files with environmental variables, here the two PCA-based variables generated with the data preparation script (step 1). Sample sizes, species’ prevalence values and the number of repetitions have to be specified manually. This script calls the simulated species definition script (step 2) and the null model script (step 4) and provides input for both scripts. Output of the main analysis scripts is stored as \*.csv files and consists of 1) the minimum numbers of samples required to generate accurate SDMs based on critical levels of model accuracy and 2) model performance values to be used for plotting.

### Null models – vanProosdij\_et\_al\_2015\_A8\_nullmodel.r

The null model script is called by the main analysis script, which provides the environmental variables, sample sizes and number of repetitions for it too. The null model script computes AUC values of SDMs based on randomly drawn samples, which are then used in the main script to assess model accuracy.

### Plotting – vanProosdij\_et\_al\_2015\_A9\_plotting.r

The plotting script generates figures from the output of the main analysis script. It reads the \*.csv files with the output of the main analysis script and returns figures with legends. Lay-out and resolution of the figures can be adjusted in the script.

Additional questions can be addressed to André van Proosdij ([andrevanproosdij@hotmail.com](mailto:andrevanproosdij@hotmail.com))