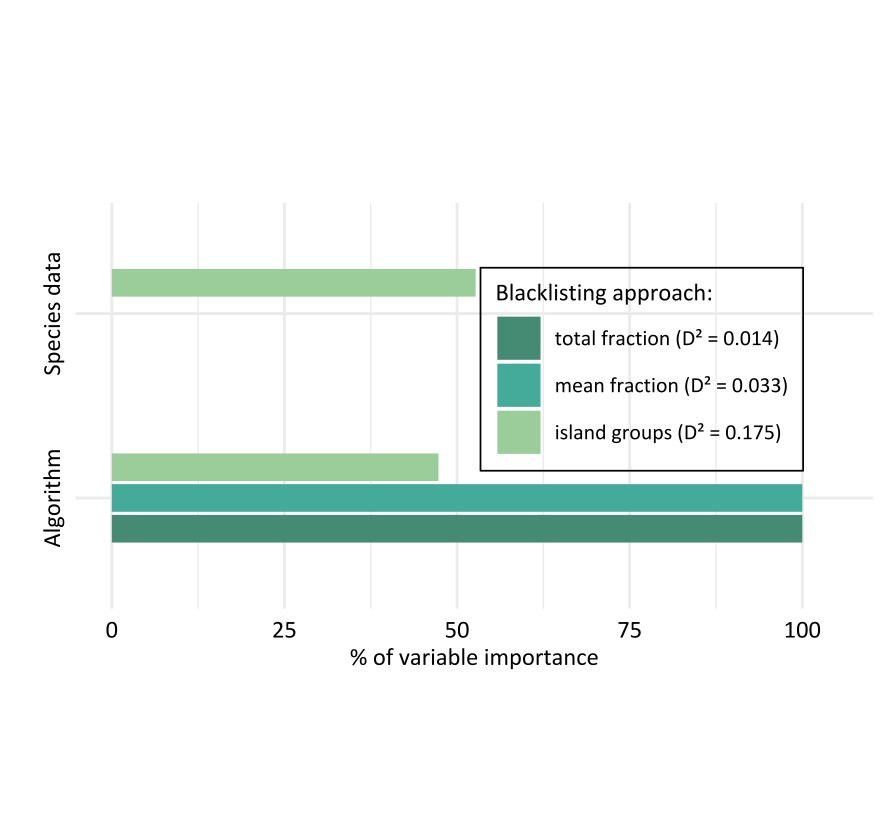
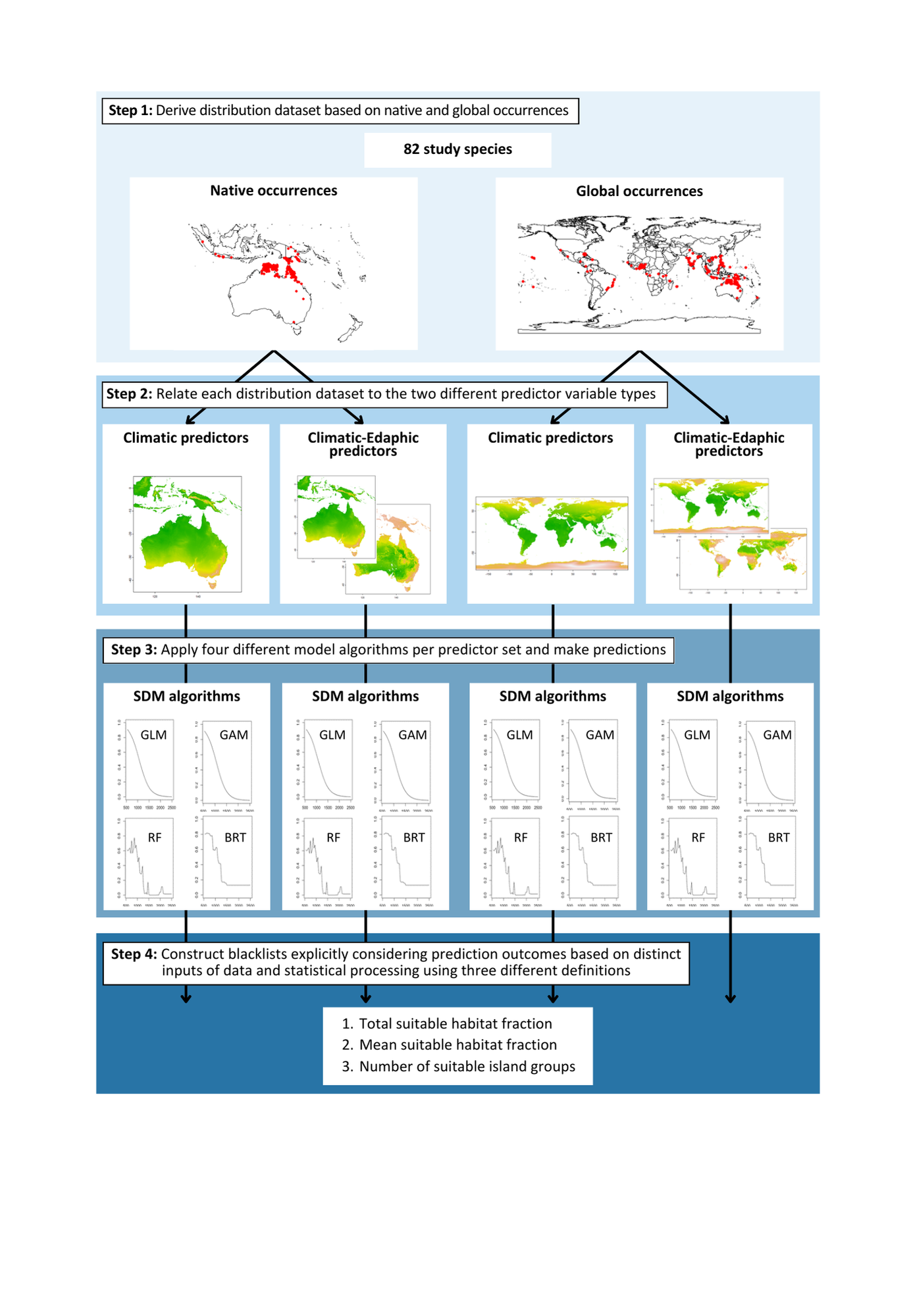
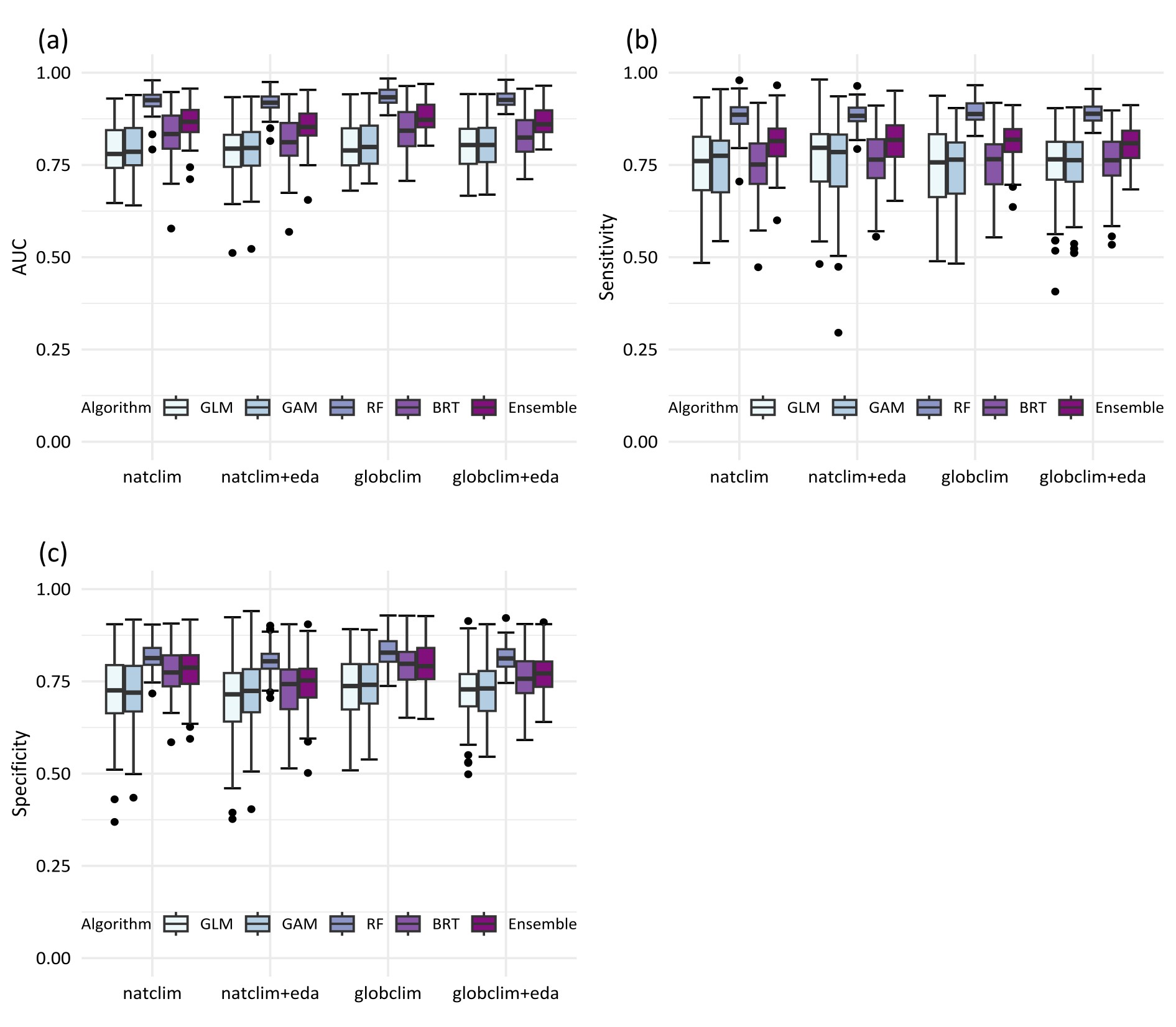
**Uncertainty in blacklisting potential Pacific plant invaders using species distribution models**

Supplementary material A

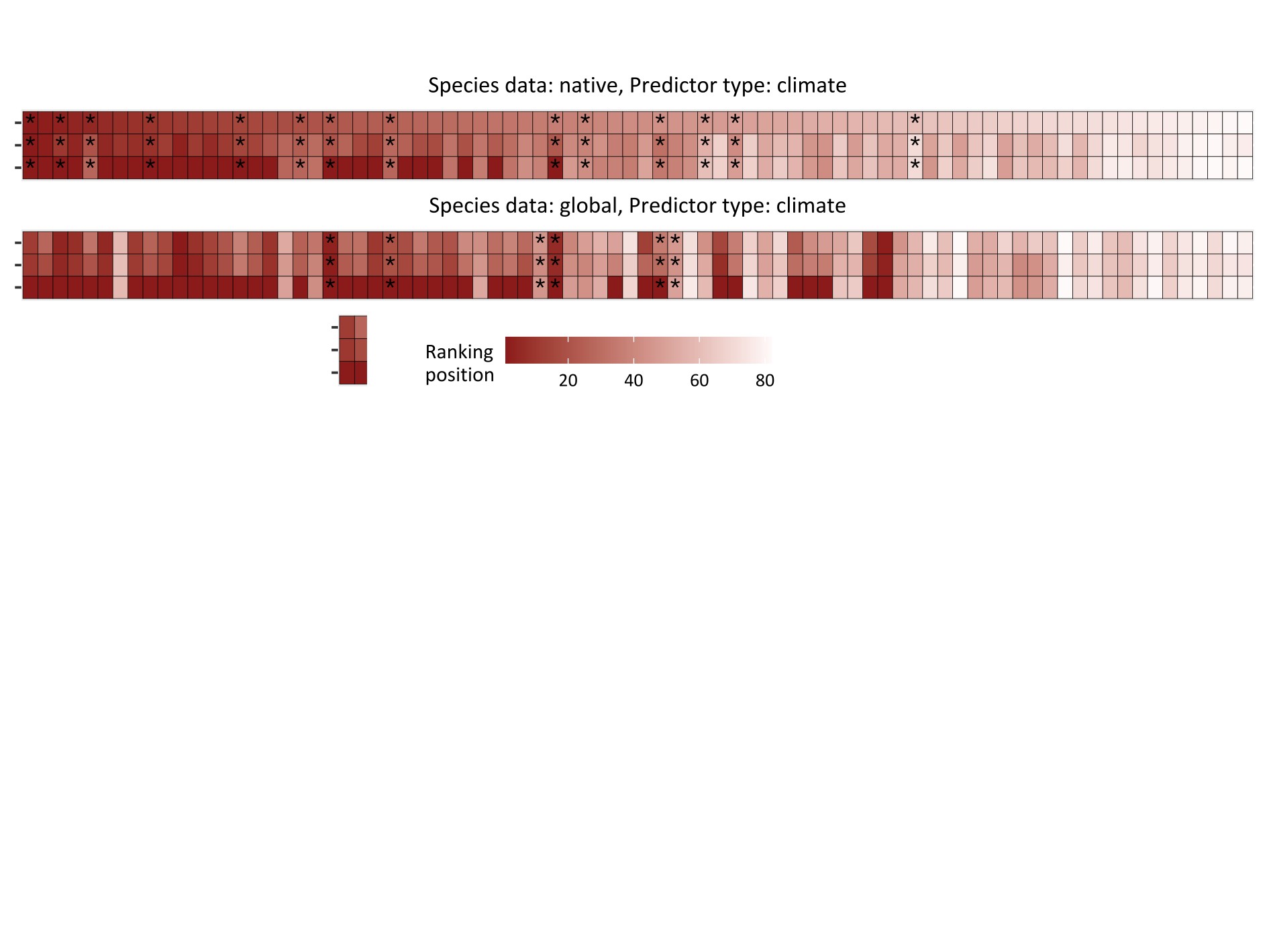
**Figure S1: Schematic representation of the workflow.** For all study species (n = 82), SDMs were fitted on opportunistic species data representative of their native or global niches, using two different predictor variable sets (climatic and edapho-climatic) and four different SDM algorithms. From the predicted habitat suitability maps, we derived blacklists using three distinct definitions. These blacklists were then used to quantify the uncertainty in ranking related to three different sources: species data, predictor type, and algorithm.



**Figure S3: Quantification of investigated uncertainty sources based on analysis of 49 Pacific island groups.** The bars display the influence in the choice of two different sources of uncertainty on the blacklist ranking outcome: the niche (global vs. native) and the algorithm (GLM, GAM, RF, BRT). Variable importance was extracted from a random forest regression relating to differences in blacklist ranking to the uncertainty factors. The values of variable importance per blacklisting approach were standardised to sum up to 100 %. The explained deviance D² of the models is noted in the legend. The models include blacklisting results of all studied non-native plant species (n = 82) based on predictions of 49 Pacific island groups.



**Figure S2:** **Model validation in terms of AUC (area under the receiver operating characteristic curve), Sensitivity and Specificity.** For all studied non-native plant species (n = 82), the range of performance measures for the four applied model algorithms and the averaged ensemble model along the four different input datasets are displayed as boxplots (a-c). Models were validated using 5-fold cross-validation. (GLM: generalised linear model; GAM: generalised additive model; RF: random forest; BRT: boosted regression tree; natclim: models based on native occurrences and climate data; globclim: models based on global occurrences and climate data; eda: models additionally considering edaphic data)



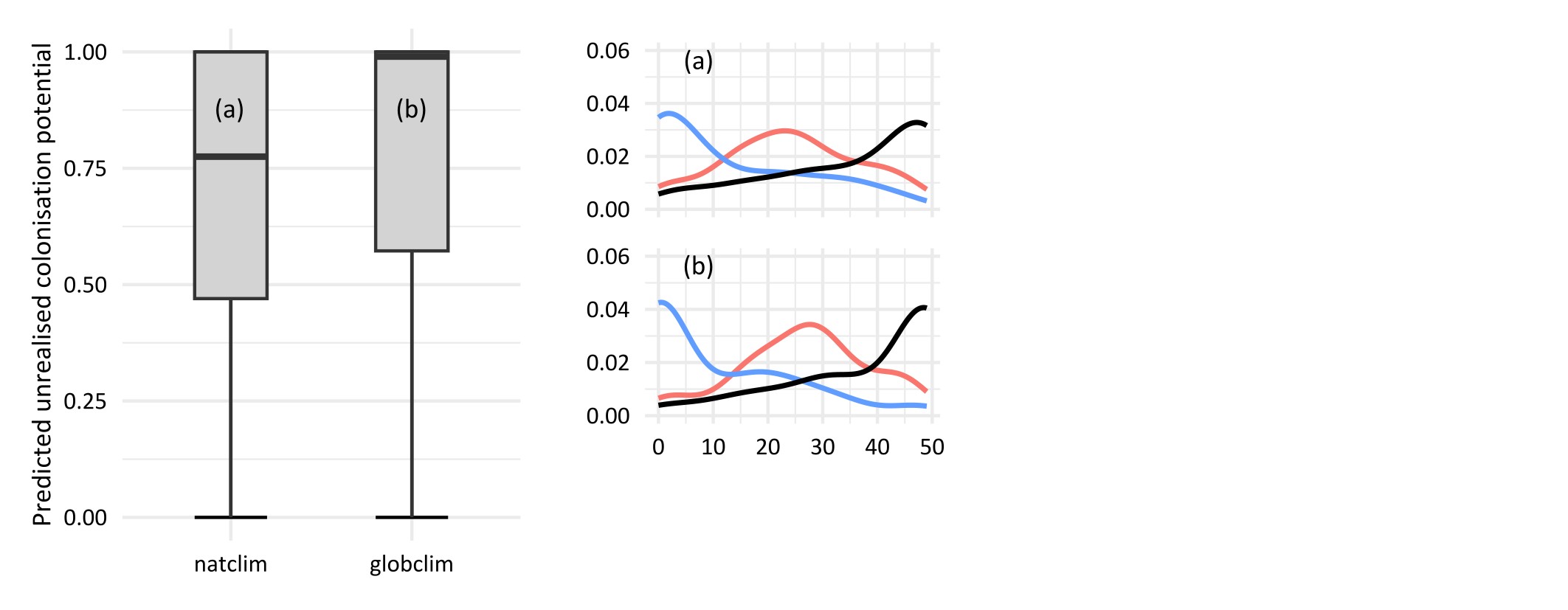
Non-native species

Total fraction

Mean fraction

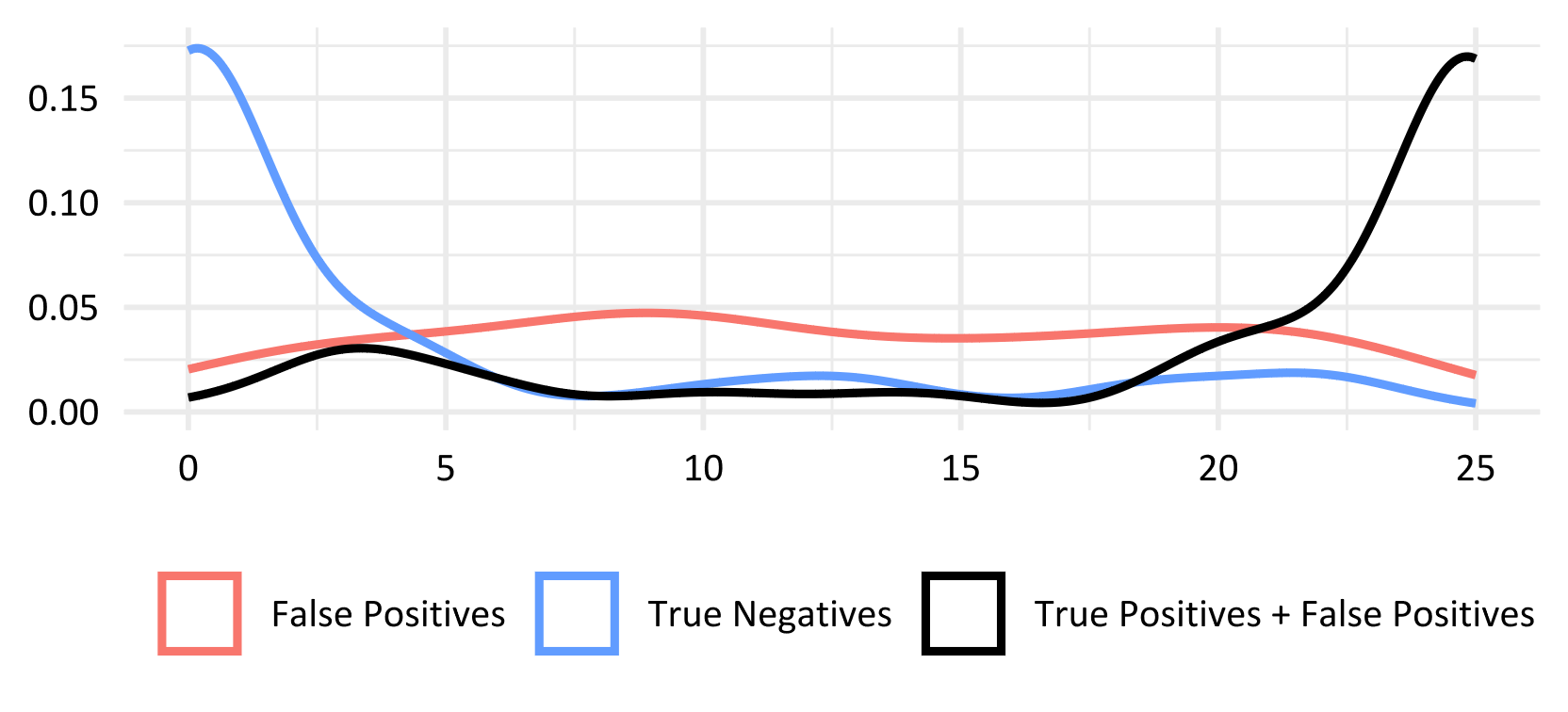
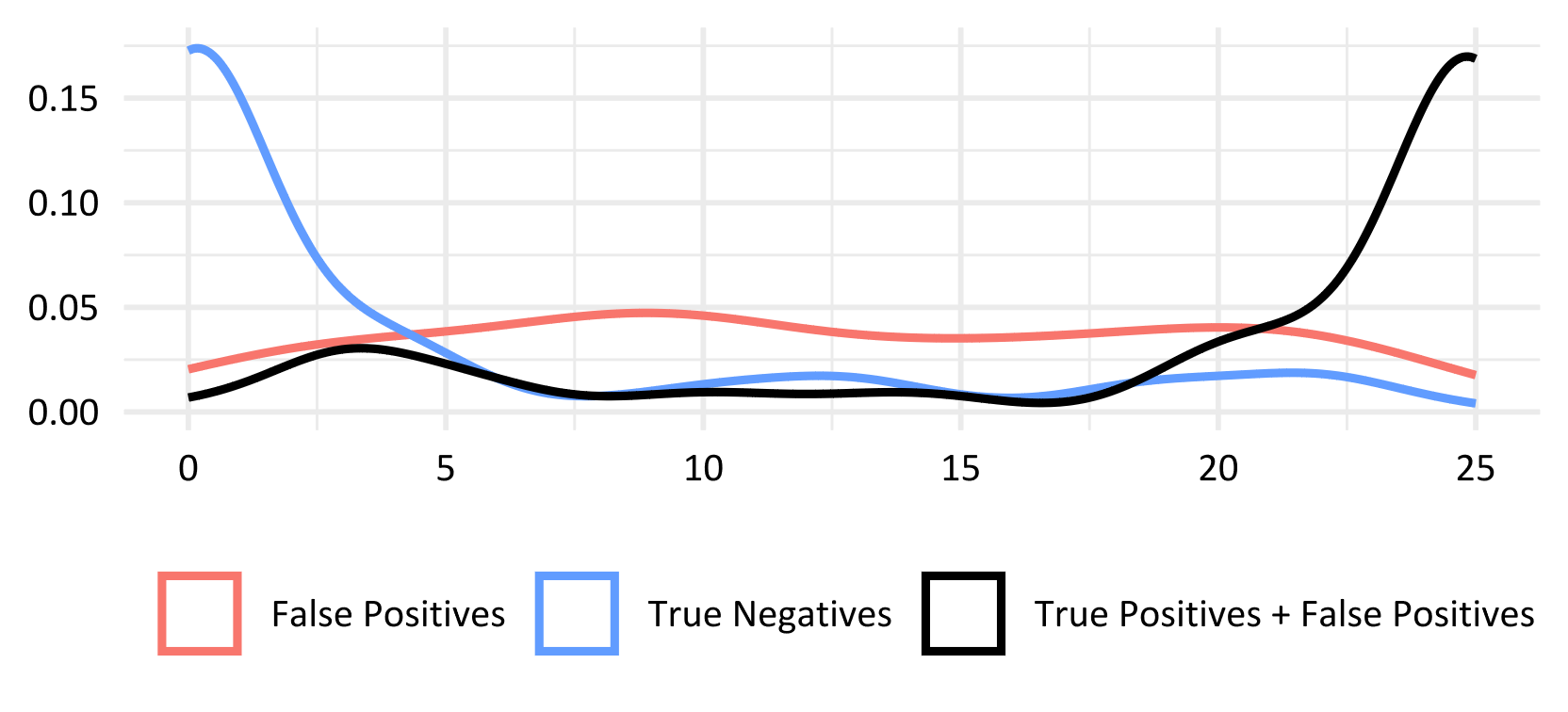
Island groups

**Figure S4: Final blacklisting based on analysis of 49 Pacific island groups.** The final blacklisting results of all three blacklisting approaches over two different input datasets are displayed. Final ranking positions were based on the predictions of ensemble models to 49 Pacific island groups solely covered by climatic data. Every non-native plant species (n = 82) is depicted along a vertical line of cells. The horizontal lines of cells per predictor set represent the three blacklisting approaches, with cell colour intensity indicating the ranking position of each study species. To enhance visual (dis)similarities of ranking positions using the number of suitable island groups as a definition, positions were recalculated incorporating gaps equivalent to the number of tied ranks. Predictive performance values of ensemble models below a TSS (True Skill Statistic) score of 0.5 are marked with an asterisk.



Density

Island groups



**Figure S5: Predicted Pacific-wide unrealised colonisation potential based on analysis of 49 Pacific island groups.** On the left, the predicted Pacific-wide unrealised colonisation potential for all studied non-native plant species (n = 82) is displayed as boxplots, based on the ensemble predictions of 49 Pacific island along the two predictor sets solely including climate data (a-b). The predicted unrealised colonisation potential per species represents the False Positive Rate defined as the number of island groups that are hitherto unoccupied by the non-native species but predicted as suitable divided by the total number of island groups where the species is currently absent. On the right, the density of island groups with false positives (predicted but absent; red line), true negatives (blue line), and all predicted presences (black line) is depicted for all study species using the same ensemble predictions (a-b). (natclim: models based on native occurrences and climate data; globclim: models based on global occurrences and climate data).