Appendix S1: Supporting Information for Kramer et al. 2017. Suitability of Laurentian Great Lakes for invasive species based on global species distribution models and local habitat

Figure S1

Golden mussel niche centrality plus occurrence points. Black crosses are locations used to fit and test the range-bagging model.

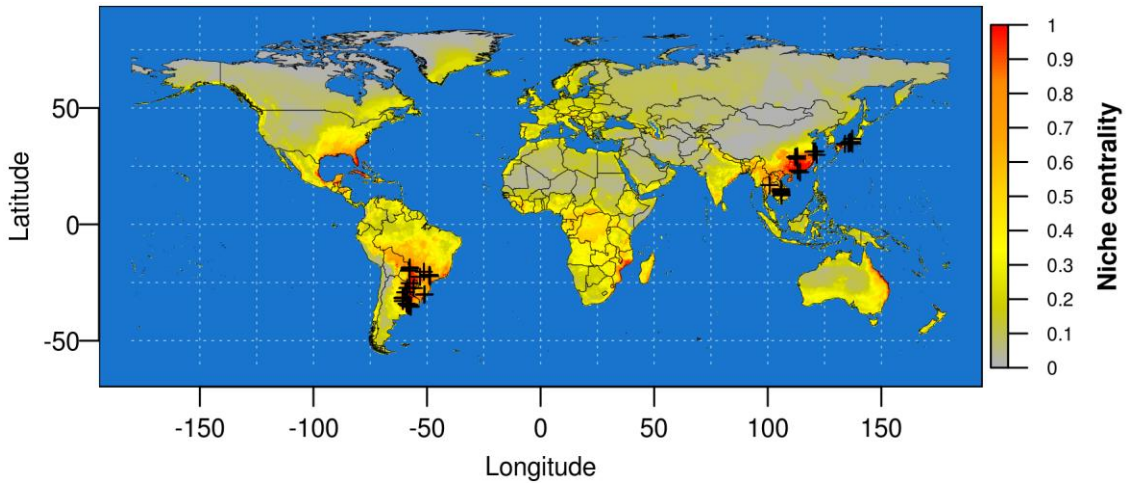

Figure S2: Killer shrimp niche centrality and occurrences. Black crosses are locations used to fit and test the range-bagging model.

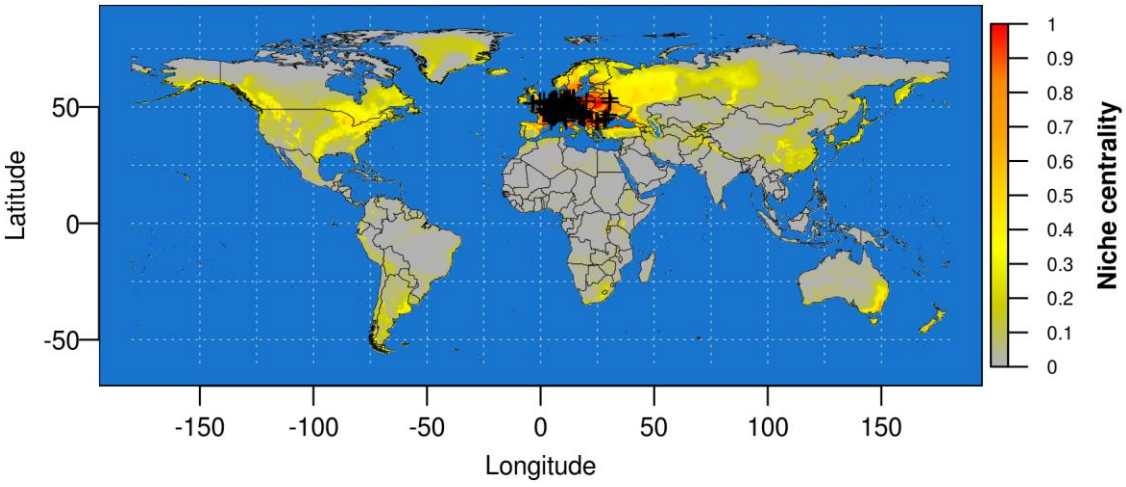

Figure S3

Northern snakehead niche centrality and occurrences. Black crosses are locations used to fit and test the range-bagging model.

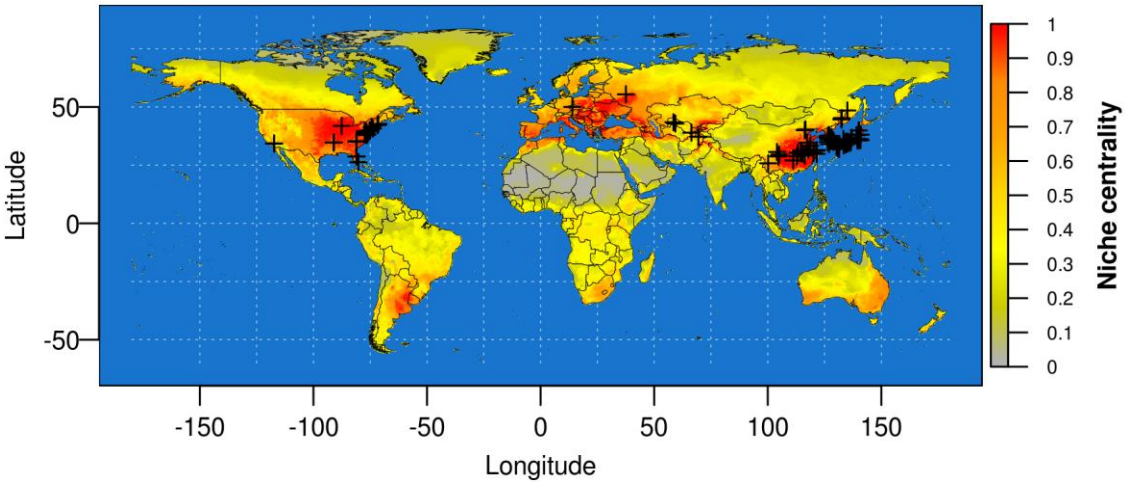

Figure S4: Flowchart showing data usage for standard and cross-validation measures of performance.

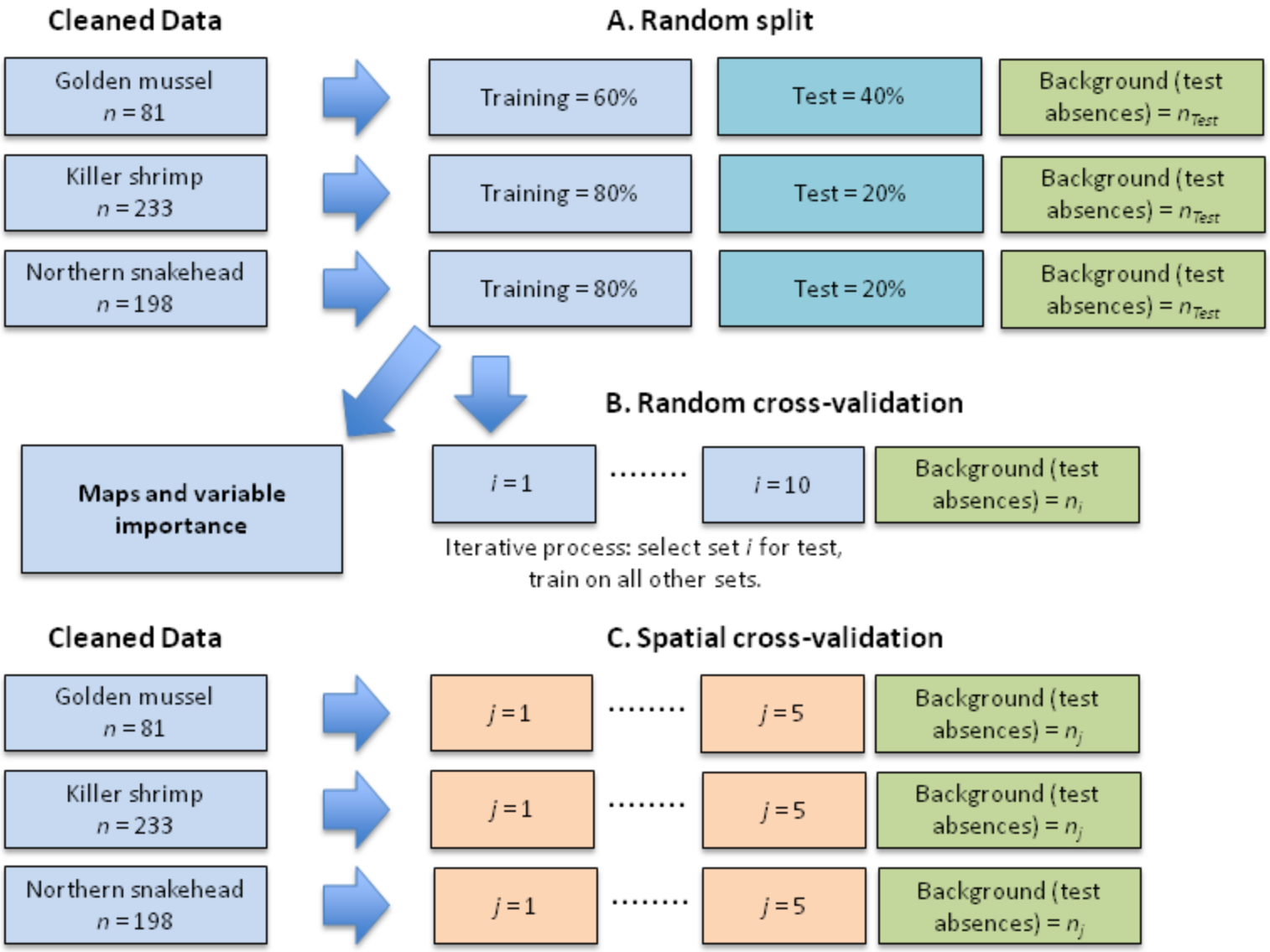
Cleaned Data
A. Random split
Golden mussel
n = 81
Killer shrimp
n = 233
Northern snakehead
n = 198
Training = 60%
Training = 80%
Training = 80%
Test = 40%
Test = 20%
Test = 20%
Background (test absences) = n_{Test}
Background (test absences) = n_{Test}
Background (test absences) = n_{Test}
Maps and variable importance
B. Random cross-validation
i = 1
i = 10
Background (test absences) = n_i
Iterative process: select set i for test, train on all other sets.
Cleaned Data
C. Spatial cross-validation
Golden mussel
n = 81
Killer shrimp
n = 233
Northern snakehead
n = 198
j = 1
j = 5
Background (test absences) = n_j
Iterative process: select set j for test, train on all other sets. Sets are split into longitudinal bands of equal n, background included. Golden mussel trained on 2 sets at a time.

Figure S5: Golden mussel model AUC. Receiver-operating characteristic for range-bagging model on a 40% hold-out test set.

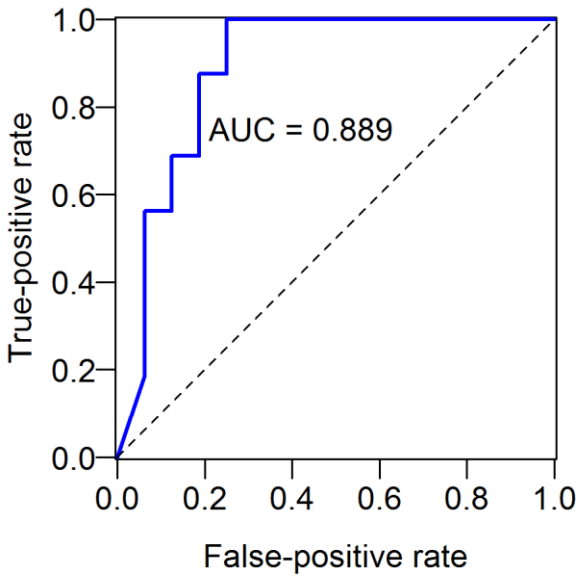

Figure S6: Performance on cross-validation folds assessed by AUC. The mean and 95% confidence interval are shown for 10-fold cross validation with random bins ("CV", dark grey) and for 5 fold cross-validation with longitudinally defined bins ("Spatial, light grey).

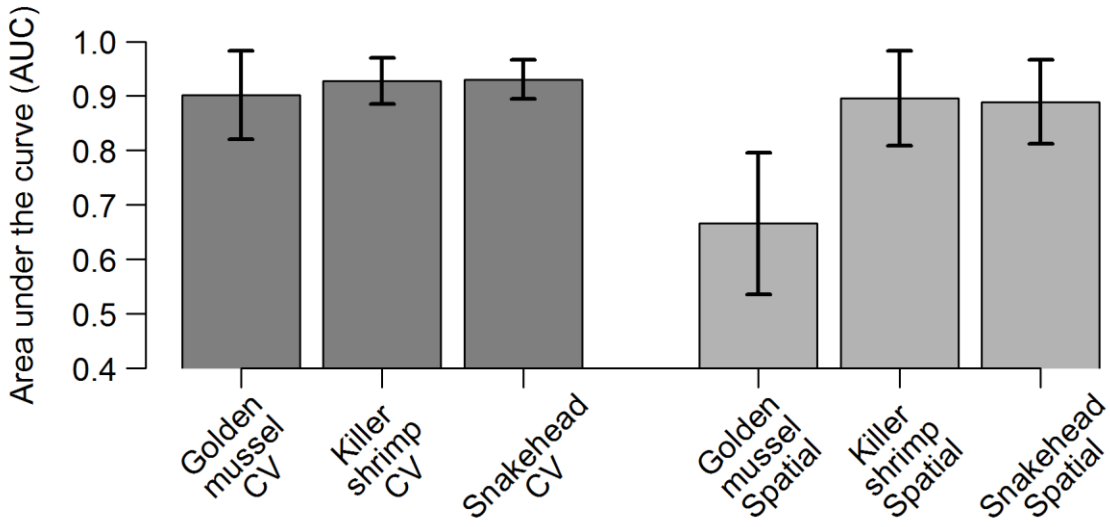

Figure S7: Performance on cross-validation folds assessed by the Boyce index. The mean and 95% confidence interval are shown for 10-fold cross validation with random bins ("CV", dark grey) and for 5 fold cross-validation with longitudinally defined bins ("Spatial, light grey).

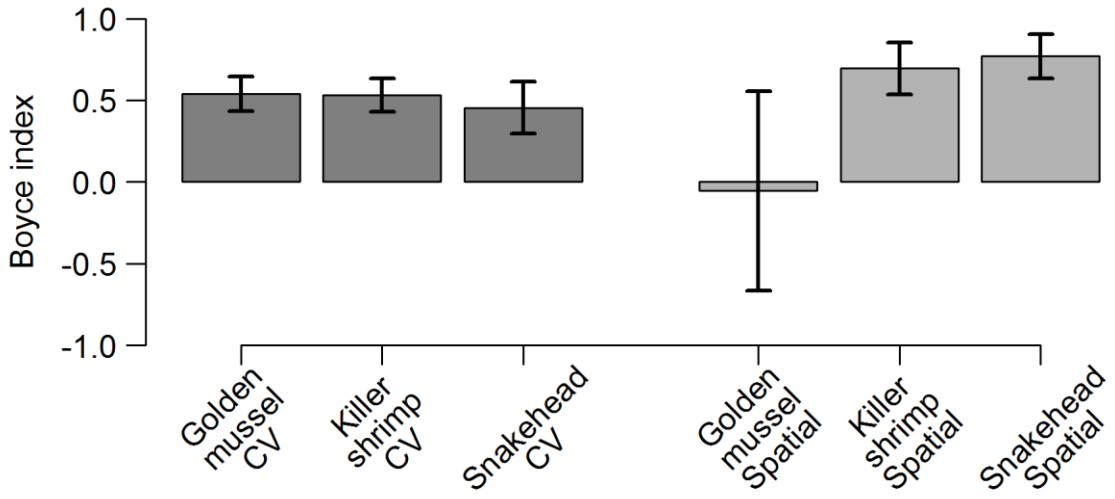

Figure S8: Variable importance plot for Golden mussel model. Relative importance of variables in the range-bagging model. Mean diurnal range was least important and precipitation seasonality most important, but there is a high degree of redundancy displayed by the even importance values.

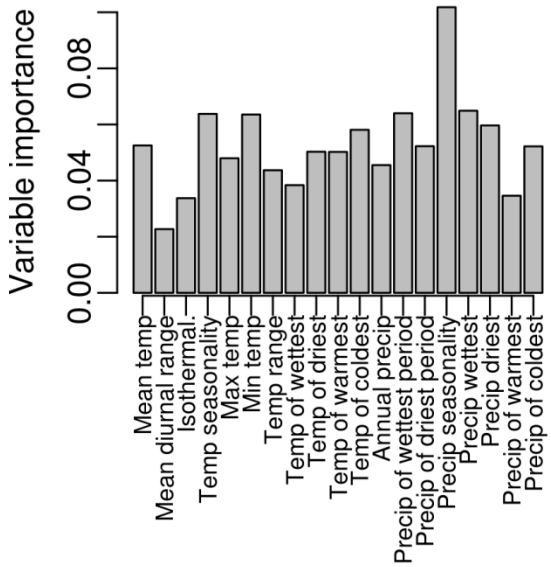

Figure S9: Killer shrimp AUC. Receiver-operating characteristic for range-bagging model on a 20% hold-out test set.

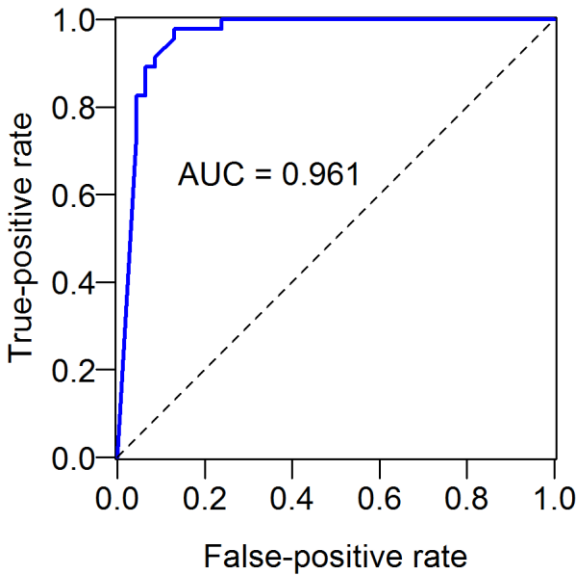

Figure S10: Killer shrimp variable importance. Relative importance of variables in the range-bagging model. Precipitation of the wettest month and wettest quarter were most important. This model also displays negative performance for several variables, while these are small in an absolute sense they are of the same relative magnitude as many of the informative variables. These negative predictors could arise from the high sampling of a limited climatic area and the resulting opportunity for points with similar covariates having opposite outcomes, or from overfitting. Overfitting may be less likely here given the high performance on the hold-out test.

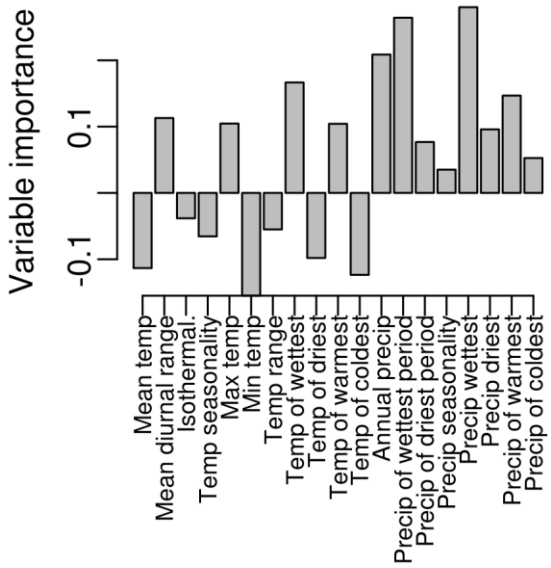

Figure S11: Northern snakehead AUC. Receiver-operating characteristic for range-bagging model on a 20% hold-out test set.

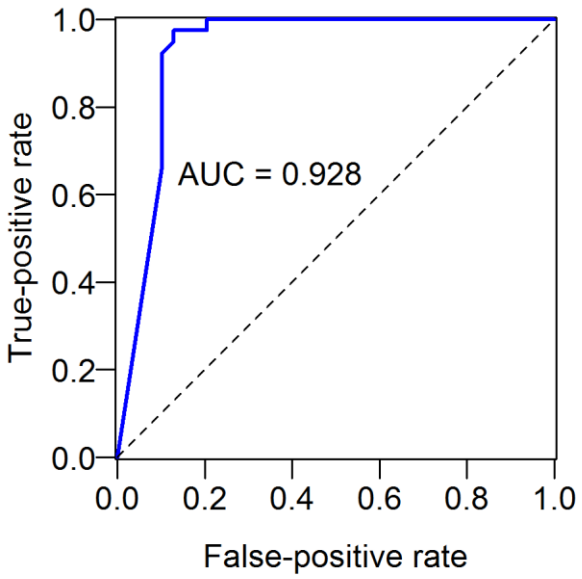

Figure S12: Variable importance for northern snakehead model. Relative importance of variables in the range-bagging model. Temperature of wettest quarter was most important and mean temperature least important, but there is a high degree of redundancy displayed by the even importance values.