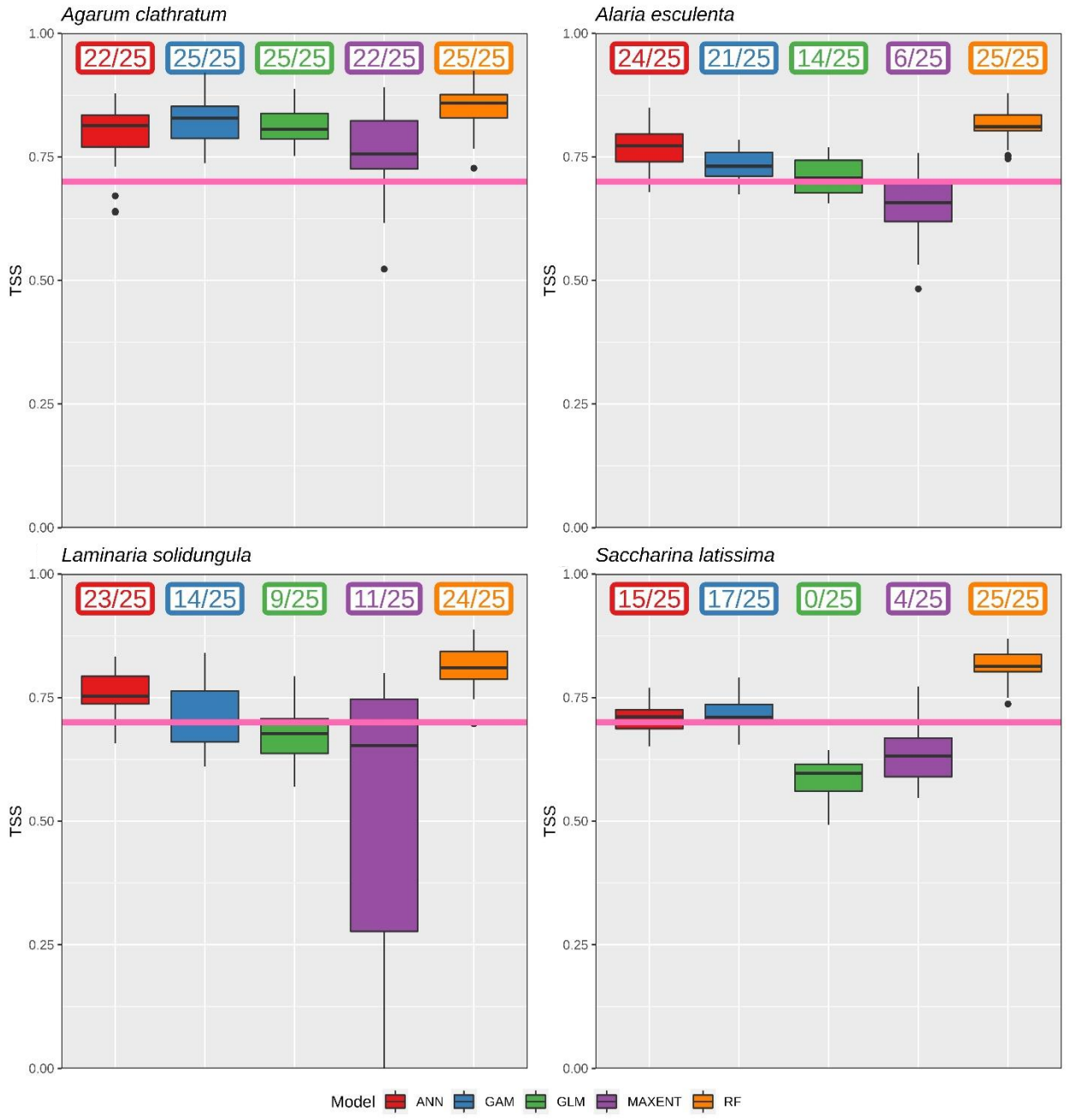
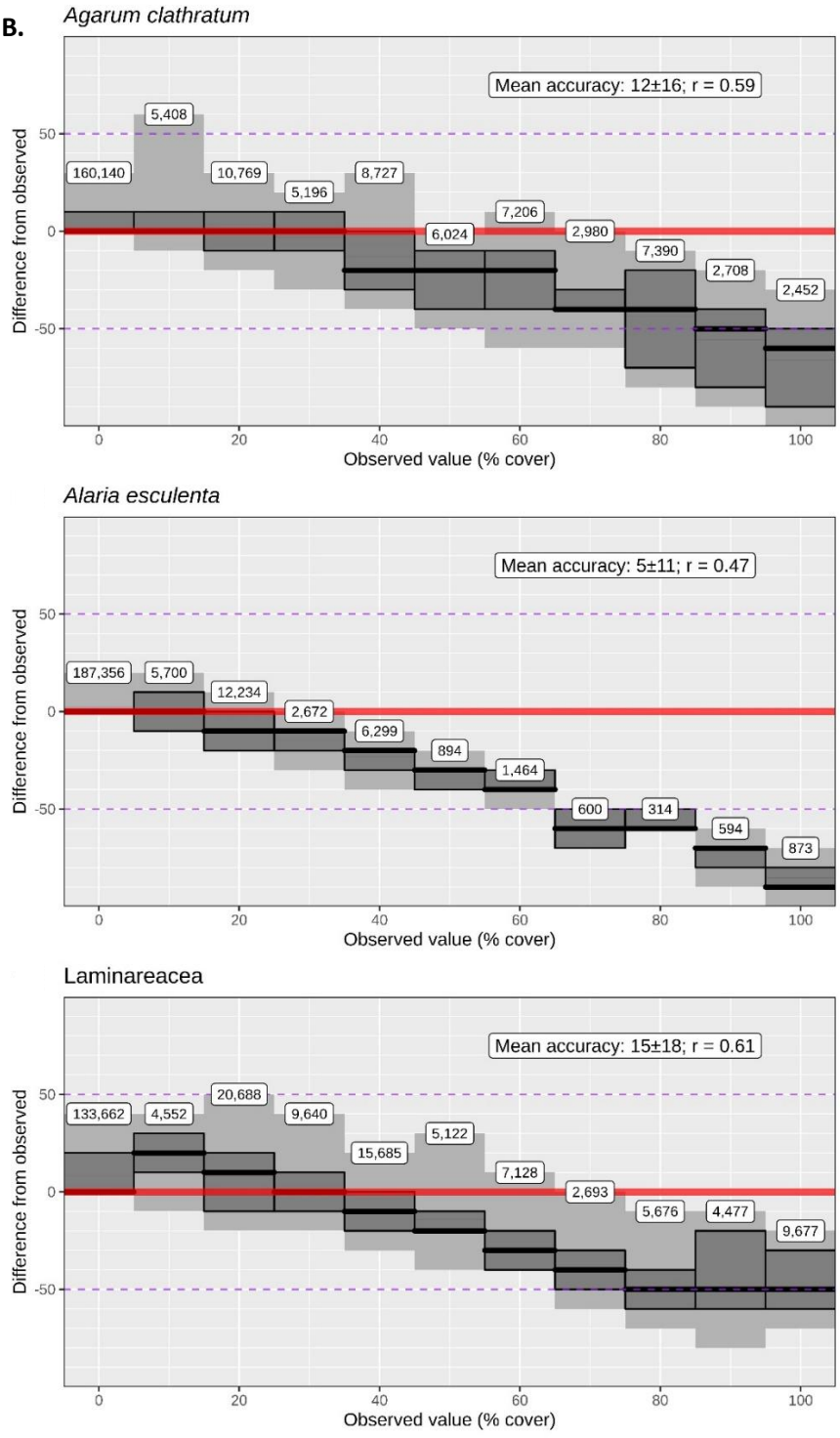
A.



Supplementary Material Figure 4.A Boxplots showing the range of scores for the various models that make up the ensembles. Panels show the scores for each species modelled: (A) *Agarum clathratum*, (B) *Alaria esculenta*, (C) *Laminaria solidungula*, and (D) *Saccharina latissima*. Model abbreviations: Artificial Neural Network (ANN), Generalized Additive Model (GAM), Generalized Linear Model (GLM), MaxEnt, Random Forest (RF). The criteria for a model passing and being used in the ensemble was a TSS score of 0.7 or greater, shown here as a horizontal pink line. The number of models that passed are shown above each boxplot.

Summary of Supplementary Material Figure 4.A: The majority of the models for most species passed this criteria, with *A. clathratum* having the fewest failed models at 6 out of 125, and *S. latissima* the most failed models at 64 out of 125. The model that most commonly failed to meet the TSS criteria was MaxEnt, with 57 fails out of 100. The RF models had the highest pass rate with only 1 fail across all four species. The single highest failure rate for a model was the GLM for *S. latissima*, for which every model failed. The lowest overall TSS scores were found for MaxEnt for *L. solidungula*. The higher the number of models that passed the test for TSS, the higher the robustness of the complete set of models that were ensembled. All GLM models failed for *S. latissima*, such that the ensemble model consisted of only four of the models, rather than five as for the other species. Overall, the ranking of models from best to least performing was: RF, ANN, GAM, GLM, MaxEnt.

B.



Supplementary Material Figure 4.B Boxplots showing the ranges in accuracy for Random Forest (RF) models used to project the observed cover values of the target species. The x axis shows the range of observed cover values in 10% steps. The y-axis shows the difference between the observed values (x-axis) and the values predicted by the RF models. The thick central black

line of each box is the median accuracy for that 10% cover group. For example, if the observed value (% cover; x-axis) is 50, and the median difference from observed (y-axis) is -20, this means that the average % cover projected by the RF models is 30% when, in reality, the observed values are 50%. The dark grey portions of the boxplots show the interquartile (25^{th} to 75^{th}) range of accuracies for projected values. The light grey portions of the boxplots show the 5^{th} to 95^{th} percentiles of the distribution. Labels above each boxplot show the overall number of values classified into each group, totaling roughly 219,000 data points (728 original data points, 30% used for validation, 1,000 RF models). The horizontal red line highlights model accuracy of 0%, meaning a perfect fit. Horizontal dashed purple lines at -50 and 50 serve as visual aids.

Summary of Supplementary Material Figure 4.B: In general, by comparing % cover results from the Random Forest with the known observed values, we see that the models tend to accurately predict values of 30% cover and lower. One important caveat to note in the results of the Random Forest is that the models tend to be very conservative in their estimation of % cover. For all species the models tend to consistently underestimate the known % cover values. Such that the results almost never return any cover values above 60%, even though observed values of 100% exist in the dataset. This means that the values projected in the results are likely much higher than reported.