Supporting Information S3

Random forest model statistics for coastal infrastructure extent in 30 global coastal locations. RMSE is the root mean squared error and values in brackets represent the range of the response variable for the model.

	% variation explained	RMSE (data range)	Obsvs. pred.
(a) Total infrastructure	54 %	39 km (478 km)	$R^2 = 0.89$
(b) Breakwalls	35 %	24.4 km (250 km)	$R^2 = 0.87$
(c) Pontoons	50.7 %	8.7 km (66 km)	$R^2 = 0.91$
(d) Wharves	54 %	7.3 km (75 km)	$R^2 = 0.92$
(e) Jetties	29.2 %	12.3 km (137 km)	$R^2 = 0.86$

To estimate the prediction variances of the random forest models infinitesimal jack-knife techniques were used This technique works by omitting each observation (Leave-One-Out) and rerunning the estimate with the remaining data. Variances were square root transformed to provide a standard deviation for each data point. The prediction variables were observed to be larger as the observed hardening increased. This is potentially due to the low number of locations with extensive hardening available for the model.

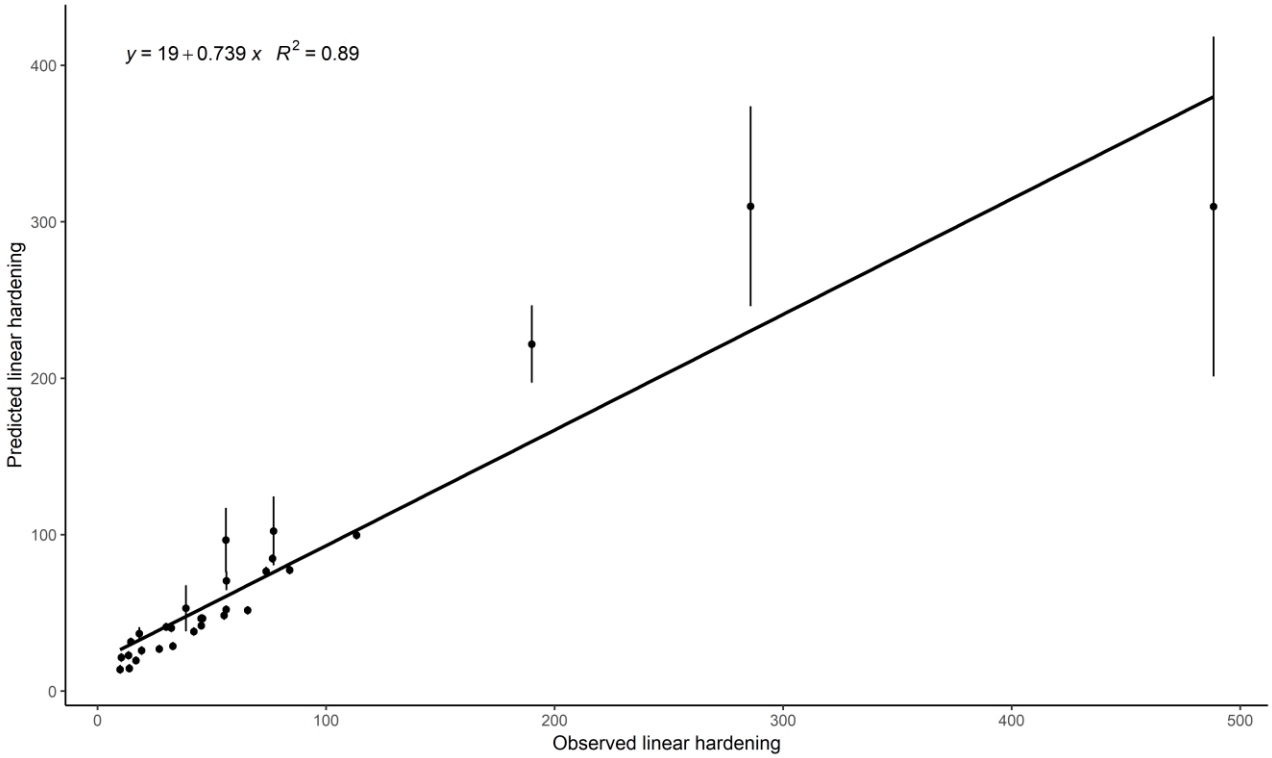

Figure 1 The observed total linear hardening vs predicted values from the Random Forest model. Standard deviations were generated using infinitesimal jackknife resampling methods.

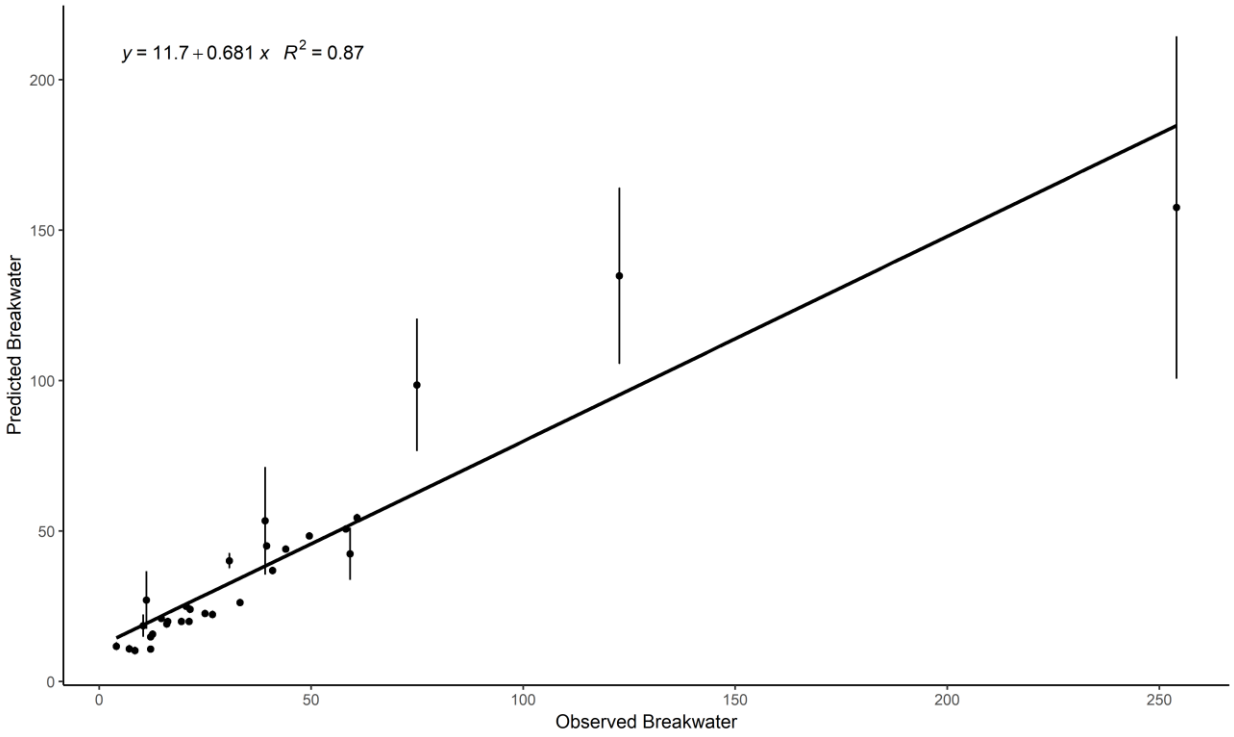

Figure 2 The observed breakwater length vs predicted values from the Random Forest model. Standard deviations were generated using infinitesimal jackknife resampling methods.

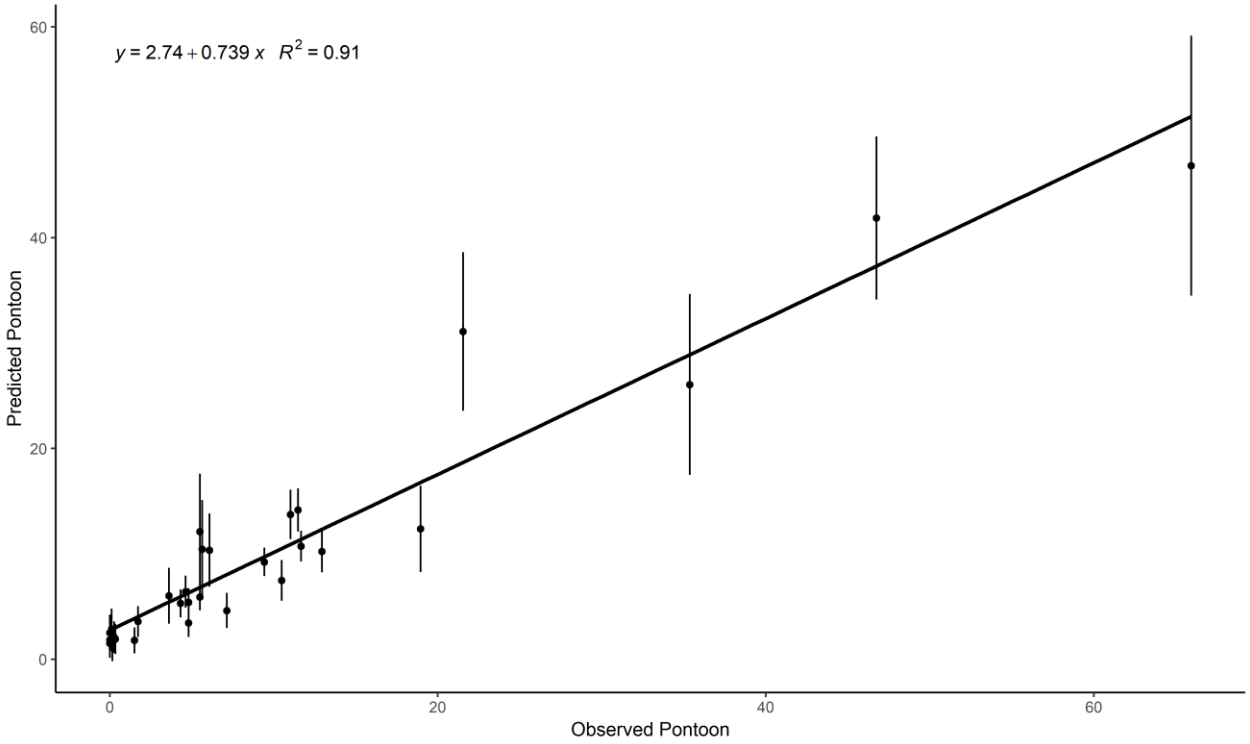

Figure 3The observed pontoon length vs predicted values from the Random Forest model. Standard deviations were generated using infinitesimal jackknife resampling methods.

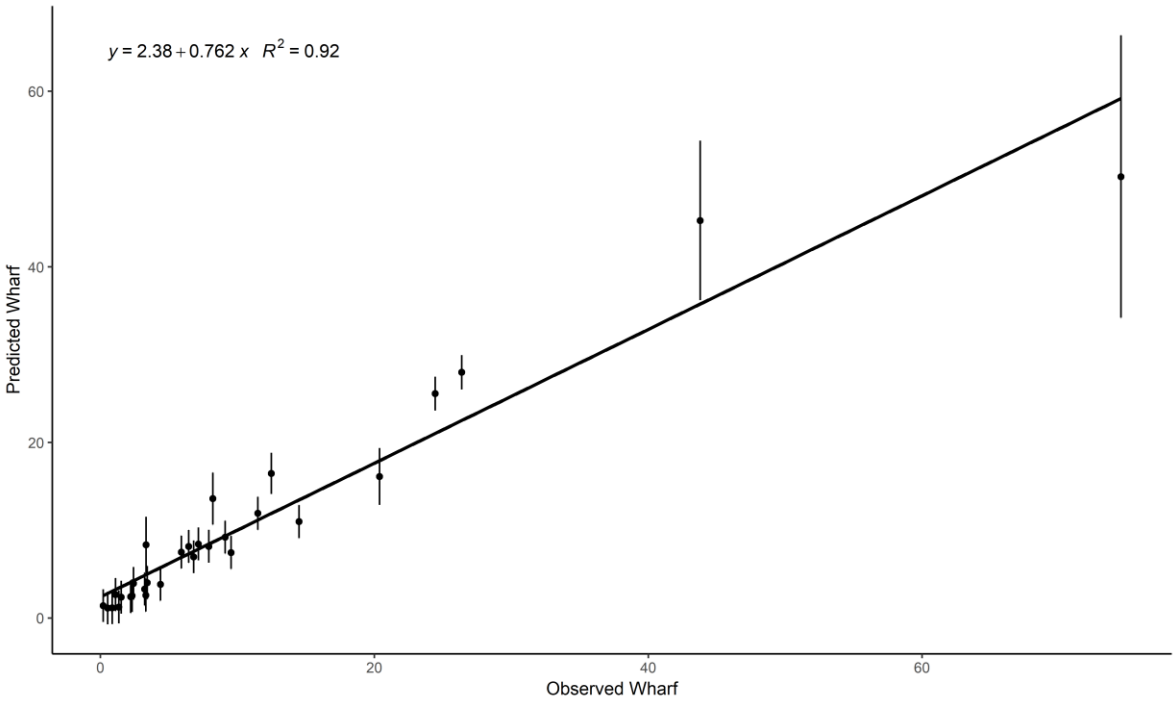

Figure 4 The observed wharf length vs predicted values from the Random Forest model. Standard deviations were generated using infinitesimal jackknife resampling methods.

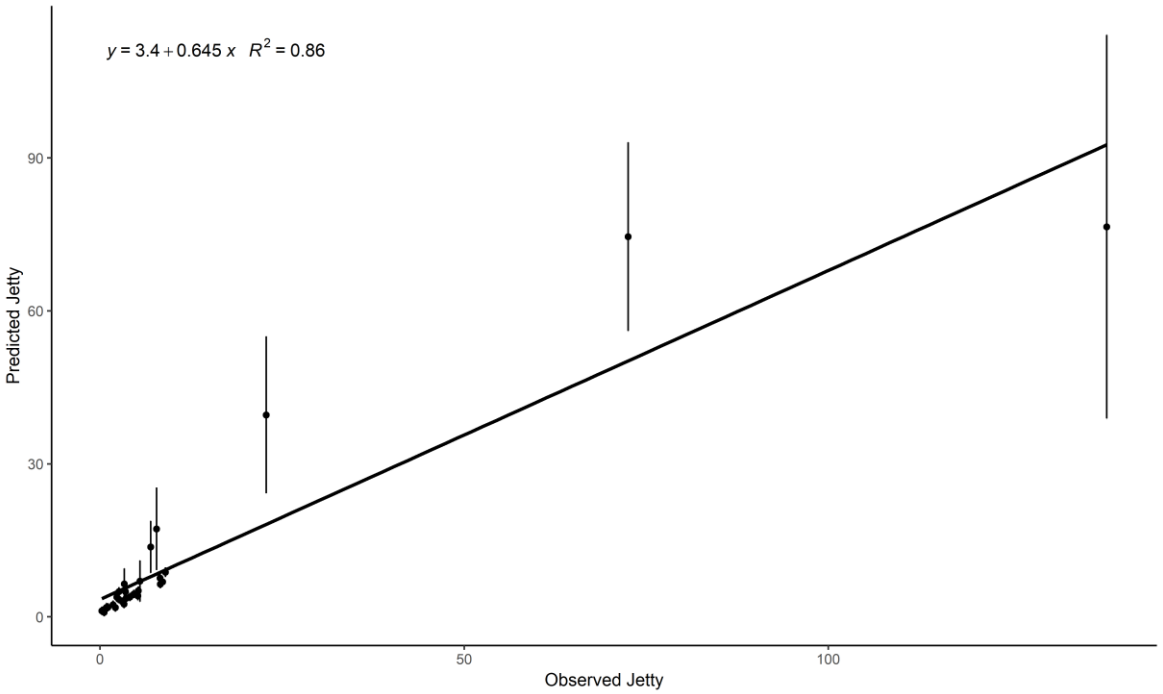

Figure 5 The observed jetty length vs predicted values from the Random Forest model. Standard deviations were generated using infinitesimal jackknife resampling methods.

The relative importance for each of the variables was different for the hardening types. For example, for the total linear hardening, ship visits, number of marinas and port revenue were the three most important variables.

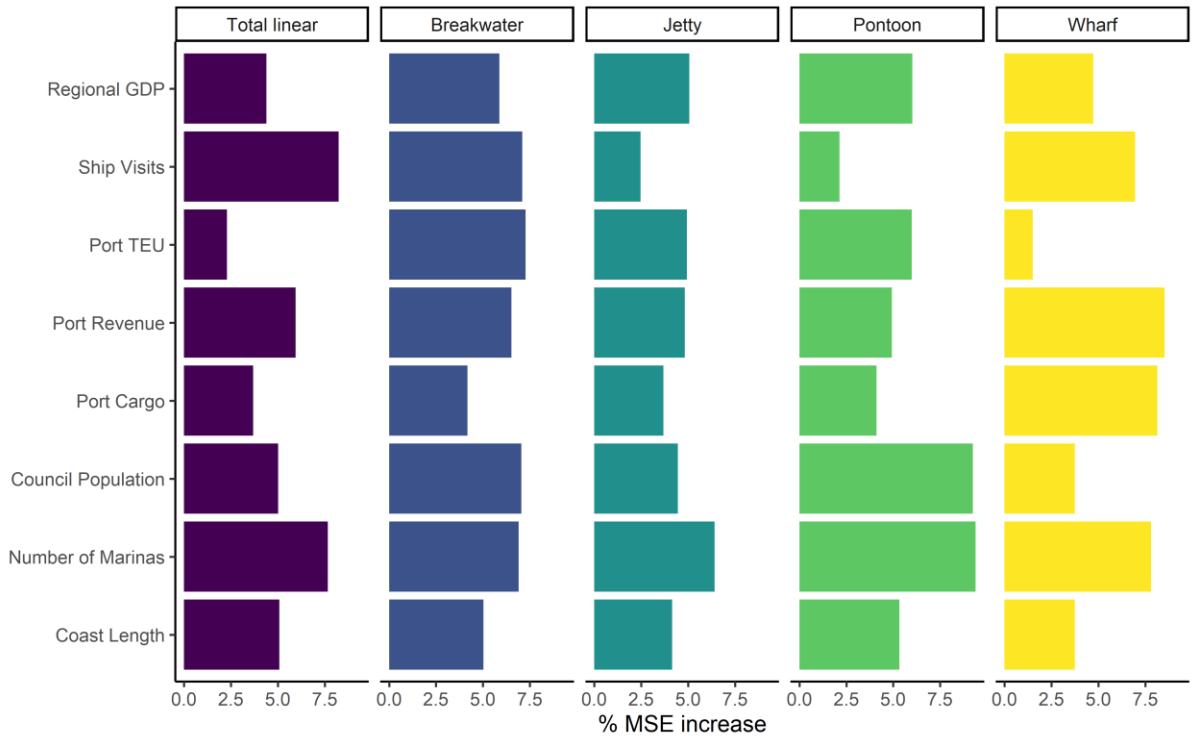

Figure 6 The relative importance of the variables in the random forest models for each of the hardening types (different colours). The importance is represented by the percent increase of mean square error (MSE) when they are removed. The higher the value the more important the variable was.