# Forecast Process

Sara Miller

November 1, 2022

## Data

The data needed to run the code is the updated file var2022_final.csv and forecasts.csv. The harvest, CPUE, and ISTI data for the var2022_final.csv file is from Andy Piston (biologist, Ketchikan office). The other temperatures variables are created by running the code satellite_data_monthly.R. The process for the temperature variables are then written up in the satellite_SST_process.Rmd file.

## Code

To create the 18 models, the code is run in the following order;
1. 1_summarize_models.R;

2. 2_diagnostics.R;

3. 2a_diagnostics.R;

4. 3_sensitivity.R; and

5. 4_retro_analysis.R

### 1_summarize_models.R

This script creates the model_summary_table1.csv, model_summary_table2.csv, model_summary_table3.csv, model_summary_table4.csv, seak_model_summary.csv, data_used_a.csv, data_used_b.csv, and a separate results_m$xx$.csv file for each model run. The columns 'model1_sim' and 'sigma' in the results_m$xx$.csv files need to be copied to the excel workbook model_summary_table_September_2022.xlsx (into each model) in the summary tables folder so that the one-step-ahead MAPE for 5 and 10 years is calculated correctly. The forecasts.csv file in the data folder is created from the results in the model_summary_table_September_2022.xlsx file. The model_summary_table5.csv file is also created from the excel workbook model_summary_table_September_2022.xlsx (although the adjusted R squared values are from the model_summary_table2.csv file). The forecast_models.png figure is also produced from this script.

### 2_diagnostics.R

This script is used to explore a best model. The outputs include model_summary_table6_*best_model*.csv. This csv files includes the residuals, hat values, cook's distance values, standardized residuals, and fitted values that are used to create the diagnostic figures catch_plot_pred_m$xx$.png, fitted_m$xx$.png, general_diagnostics_m$xx$.png, and influential_m$xx$.png. In addition, the top of the script outputs the lack of fit test (Bonferroni p-values), and the lack of fit curvature test.

## 2a_diagnostics.R

This script is used to explore an alternative best model.


## 3_sensitivity.R

This code is used to filter out certain influential years but was not used in the 2023 forecast process.


## 4_retro_analysis.R

This script creates model hindcasts for the best models and combines them with the forecasts.csv file. The result is a data frame with hindcasts (for each model) using data up to a certain year, and then the forecast for that data set. The data frame is then used to create multiple figures (e.g., year_minus_5.png) that help show how the MAPE is calculated. For example, the figure year_minus_5.png shows the hindcasts for models m1, m2, and m11 using data from 1998 to 2017 only, and the 2018 forecast based on these three models (and only using data from 1998-2017). This script also produces the MAPE_forecasts.png figure for the best (or chosen) models which are the one-step-ahead MAPE forecasts for the chosen models.