Species influence on trawl catch assemblages on the Scotian Shelf

Andrew McMillan, *M.Sc. Candidate*
Department of Geography, Simon Fraser University

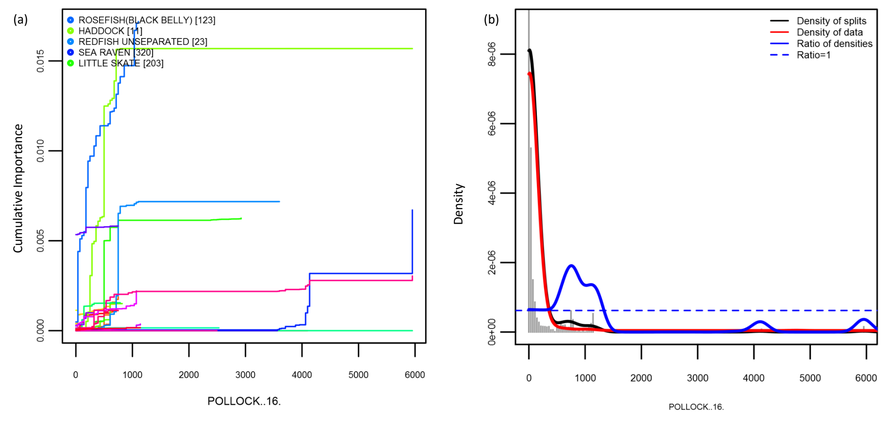
Anders Knudby, Ph.D., Assistant Professor
Department of Geography, University of Ottawa

Introduction

Identifying distributions of species commonly involves an analysis of the environmental conditions present at the locations where the species is encountered. The analysis relies on the assumption that species interactions are either sufficiently accounted for by environmental proxies or have negligible influence. Incorporating species data directly into the analysis can both assess the validity of that assumption and provide a direct representation of species interactions that can be used to identify keystone species and ultimately better model species distributions.

Random Forest (RF), an ensemble regression tree analysis, can use such a dataset to model and predict individual species distributions. RF also includes an internal variable importance assessment. Gradient Forest is an extension to the RF algorithm, using an ensemble of RF models for individual species and calculating variable importance across individual models. Predictors with high importance values have greater influence on the species assemblage in a given area.

This project aims to assess the importance of individual species catch weights as predictors of species assemblages trawled on the Scotian Shelf – in addition to the more traditionally used environmental condition predictors. Using Gradient Forest, influential species can be isolated and their distributions can be modelled as proxies for desired trawl catch assemblages.



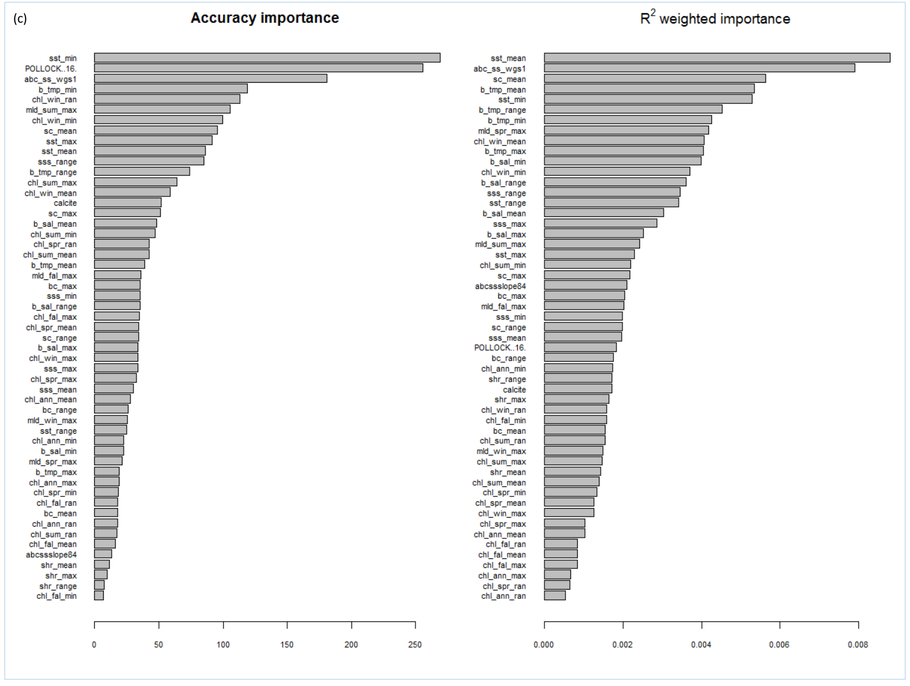


Figure 2. Cumulative importance (a) of Pollock (*Pollachius virens).* Changes of influence on different species, represented by each line, are plotted against the catch weight values of Pollock. The top five species influenced by Pollock are listed in the legend. Split densities (b) across Pollock catch weights overlaid with kernel densities of splits (black line), data (red line), and the ratio of splits to data (blue line). Overall importance (c) of all predictor variables included with Pollock averaged across all Random Forest models (Accuracy Importance) and weighted by the model fit of each Random Forest model (R^2 weighted Importance). Note that second only to minimum sea surface temperature, Pollock catch weight is the most important predictor or the species assemblage (excluding Pollock itself) on the Scotian shelf (c, left side).

Preliminary Results and Next Steps

Preliminary results are available for 30 species and include variable importance analyses for these species alongside the environmental predictors, as illustrated in *figure 2*. Until results are complete a full comparison of species predictors cannot be made, though a sense of a species' influence on catch assemblage can be gained through cumulative importance plots (*figure 2a*). The weight of these influences can be assessed by split density plots (*figure 2b*) while maintaining an overall practicality of the species predictor relative to environmental predictors (*figure 2c*).

With the completion of gradient forest models for each species predictor variable, the importance of each species must be assessed with respect to predictive power as well as potential ecological inference. To address functional redundancies amongst species predictors, catch data of species with similar functional traits will aggregated into functional groups to be used as predictor variables. If a species or species group demonstrates an influence in model output, an exploration of ecological interactions amongst species and the surrounding environment can be made.

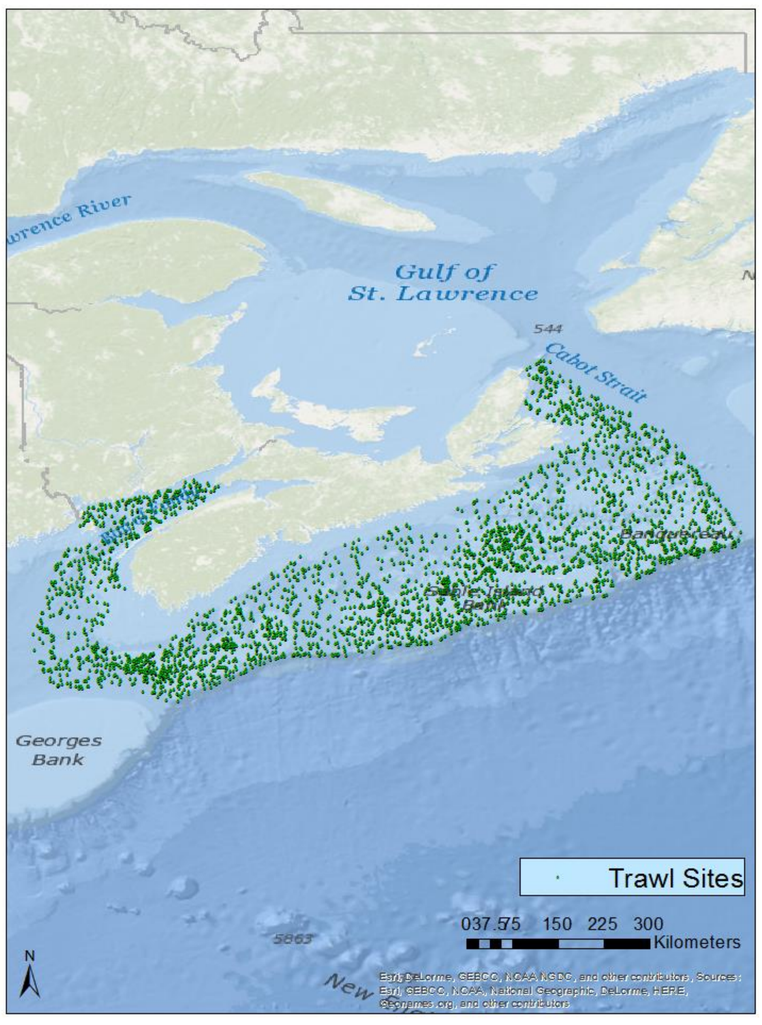


Figure 1. Sites of trawl surveys conducted by Fisheries and Oceans Canada on the Scotian Shelf between years 2000 and 2012, inclusively.

Data

Environmental Layers

The data layers representing environmental conditions in the study area were provided Fisheries and Oceans Canada (DFO). These data were derived from publicly available sources: the General Bathymetric Charts of the Ocean (GEBCO), the GLORYS v2 ocean reanalysis, and NASA's OceanColor project. The data were calculated into descriptive metrics (mean, maximum, minimum and range) with each metric as a separate data layer. In total, 52 environmental data layers were used as predictor variables in the Gradient Forest models.

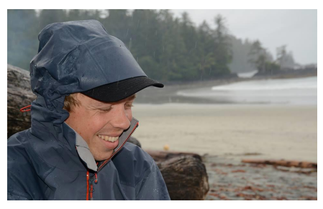
Species Layers

The data layers representing species locations and catch weights on the Scotian Shelf were collected by DFO trawl surveys from years 2000 to 2012 (*figure 1*) and consist of point data containing geographic location as well as species catch weights. Survey trawls are towed for approximately 15 minutes at depth with a speed of 3 knots, covering an estimated 0.75 nautical miles. 32 surveys were conducted over the 12 year span, collecting a cumulative of 440 species.

Methods

Each species was iteratively included with the environmental data layers as predictor variables. Predictor values were extracted at each trawl location, or site, and transformed into a site by environment matrix. The catch weights of the remaining 439 species were transformed into a site by species matrix, taking care that the site order of each matrix match exactly.

uOttawa



SFU