Boosted Regression Tree Working Paper

Adelle Molina
March 2025

Background

Gradient Boosted Regression Trees

Gradient boosted regression trees (BRT) attempt to improve model performance by fitting many models and then combining them for prediction. Boosted regression involves a combination of traditional regression or decision tree models and boosting, which builds and combines a suite of models (Elith et al., 2008). Boosting is thought to improve model accuracy and is underpinned by the concept that finding a single accurate and stationary prediction rule is more challenging than averaging across many rough "rules of thumb" (Schapire, 2003). Although they differ in theory from generalized additive models (GAMs), for example, gradient boosted regression trees often produce similar results (Leathwick et al., 2006; Lyashevska et al., 2020).

Regression trees identify regions in the predictor space with homogeneous responses to those predictors and then fit the mean response for observations in that region. Terminal nodes are regions created by splitting at various levels of the predictor variable. Individual trees are built, or grown, by repeatedly making binary splits in the predictors in a hierarchical manner. The response to one input variable depends on values of inputs higher in the tree, which essentially models interactions automatically (Elith et al. 2008). Decision tree models are fitted iteratively to a portion of the data (training data) and gradually increase emphasis on observations that are modeled poorly using a numerical optimization technique. They do so by minimizing a loss function, which represents the loss in predictive performance from suboptimal models, a process that consists of adding a new tree at each step that reduces (steps down the gradient of) the loss function. At each iteration (or new tree created) the algorithm determines the direction, or the gradient, in which it needs to move in order to improve the fit to the data (Ridgeway, 2024a). This is similar in theory to minimization of residuals, except that subsequent trees may contain quite different variables and split points because each tree is fitted to the residuals of the prior tree. This stagewise process adds trees to grow the model, "fixing" the previous tree's mistakes and then updating the fit based on contributions of the newest tree. The final model is a linear combination of many trees with the fit computed as the sum of all trees multiplied by the learning rate (Elith et al., 2008). Gradient boosted regression does not produce p values but instead computes relative influence, which is similar to an effect size. The relative influence represents the improvement in fit from splitting the tree on that predictor and is computed based on the number of times it is selected for splitting (Friedman and Meulman, 2003). The gbm packages we used (Hickey et al., 2024; Ridgeway, 2024b) implement the Friedman gradient boosting machine (Friedman, 2002, 2001) where the relative influence of a variable is computed as the frequency of it being selected for splitting, weighed by squared improvement, and averaged across all the trees. It is usually scaled to sum to 100.

Pros and Cons of GBRT

Examples of gradient boosted regression in marine fisheries have increased substantially since it was first proposed and developed in the early 2000s. Google searches for marine, fish, and gradient boosting nearly tripled between 2008-2012 and 2013-2018 (Smith, 2019). For herring specifically, Lyashevska et al. (2020) used gradient boosted regression trees to identify drivers of Celtic Sea herring growth in the first year of life, which was largely influenced by temperature and food.

Boosted trees have numerous advantages; they are unaffected by differing scales of measurement among predictors and insensitive to outliers and missing data (Elith et al., 2008). They also offer the option to use different loss functions than traditional techniques, although here we have chosen a Gaussian distribution that uses squared error as the loss function. Gradient boosted trees can model inherent, multi-level interactions between environmental variables, are generally insensitive to multi-collinearity, and can model nonlinearity between predictors and response variables (De'ath and Fabricius, 2000). Relationships between response and predictor variables and interactions between predictors do not need to be explicitly specified in the model. BRT models consistently outperform other methods and are more interpretable than other machine learning processes, such as neural networks, because they are excellent predictors and also quantify and illustrate the relationships between predictors and the response (De'ath, 2007).

Some drawbacks to BRT are that they may be less accurate than common methods and struggle to model smooth functions (Elith et al., 2008). They are also highly sensitive to small changes in subsampling proportions (Hastie et al., 2001). The biggest problem is their susceptibility to overfitting, which requires careful tuning of parameters to obtain optimal predictive performance. This includes both the hyperparameters of the model and the various stochastic, resampling hyperparameters.

Methods

Modelling process/approach

This research project formally began in 2022 before the Atlantic herring stock assessment model had transitioned from the Age Structured Assessment Program (ASAP) (Miller and Legault, 2015) to the Woods Hole Assessment Model (WHAM) (Stock and Miller, 2021). We will briefly summarize the results of those preliminary analyses, which were based on the outputs of the 2022 management track stock assessment model, and explain how they were used to further develop catered indicators to test as model covariates on the recruitment term. For these initial models, we used the R packages (R Core Team, 2023) gbm (Ridgeway, 2024) and dismo (Hijmans et al., 2023) as demonstrated in tutorials by Elith and Leathwick (2016). In the first iteration, we developed a large potential list of predictors that included many of the same processes described in the Ecosystem and Socioeconomic Profile (ESP-WP) and lags of up to 3 years. Although GBM models are generally robust to inclusion of unimportant variables, Elith et al. (2008) noted that model simplification resulted in a more parsimonious model and improved predictive performance, especially for smaller data sets. Therefore, we dropped variables that routinely had zero relative influence in subsequent model runs. The possible list was further reduced by excluding any low or near 0 relative influence variables that were highly correlated with variables that seemed to be more important. For example, if sea surface temperature (SST) had high relative influence and heat wave indices did not, we dropped the heat wave index because it is correlated with SST.

Once the WHAM model output was available, we used results from the first BRT to produce more tailored variables that were appropriate in space and time and to answer specific mechanistic hypotheses. For example, temperature, zooplankton indices, and haddock abundance proxies consistently showed high variable importance across initial model iterations. Most of the variables included in those preliminary runs used ecological production unit (EPU) spatial areas and were averaged across the entire year. Since we are interested in processes that occur during the first year of life in specific habitat areas, we used and, in some cases, developed metrics that were more realistic. In this second round of modelling, we instead used the package gbm3 (Hickey et al., 2024) as the gbm package is no longer being actively developed or maintained.

Data

A list of indicators of recruitment drivers are in the ESP-WP with information about how they were produced and why they were considered. Here we provide a brief description about which predictors were included in the various gradient boosted regression tree model runs. Logic and justification for the spatial and temporal cropping in addition to the hypothesized linkage for each predictor is also discussed in the ESP-WP and summarized in Table 5 of the ESP. We selected 2-3 potential predictors to include in model runs for each of the following categories: upper trophic level, lower trophic level, physical environment, which includes temperature and thermal threshold indicators, and climate variability. For each, we have specified which lags were used and which life stage the indicator is intended to represent.

Upper Trophic

Initial proxies for haddock egg predation were based on the stock assessment model outputs of spawning stock biomass (SSB) for both the Georges Bank (GB) and the Gulf of Maine (GoM) haddock stocks. However, in the later model runs we used the index from the Haddock-WP. Proxies for predation by gelatinous zooplankton came from the R package ecodata (Beltz et al., 2024). We used both GB and GoM cnidaria abundance anomalies because both regions are major spawning areas where eggs and early life stages are present. The two indicators we included in the final model were (Figure 1):

- Haddock predation index (1 year lag, eggs)
- Ecodata: GB & GoM cnidaria abundance anomalies (1 year lag, eggs & early larvae)

Lower Trophic

Variables of primary productivity including chlorophyll-*a* concentration and primary production from ecodata were used in the ASAP model runs. At that point, we also included a variety of zooplankton products from ecodata that were focused on the GoM and GB EPUs. Most of these variables were annual averages of either total zooplankton abundance or both large and small copepod abundance. We also used seasonal abundances for different stages of *Calanus finmarchicus*. In particular, we included just the abundance *C. finmarchicus* stage C5 in the GoM in the fall. This was the only seasonal and regional predictor included in the first model fits. Several of the VAST zooplankton indices described in the Copepod Indices-WP, were tested because they represented a more realistic spatial and temporal resolution and species groupings than those from ecodata. The two indicators we included in the final runs were (Figure 2):

- September-February (fall/winter) small copepods (1 year lag, early and late larvae)
- Spring large copepods (no lag, late larvae and juveniles)

Physical Environment

In early fits to the ASAP recruitment estimates, we used daily gridded sea surface temperature (SST) data from the NOAA optimum interpolation sea surface temperature (OISST) high resolution dataset (Huang et al., 2021), aggregated the data by EPU, computed annual averages for just GB and the GoM EPU areas, and included lags of up to 3 years. For bottom temperature (BT), we initially used the high resolution annual bottom temperature indicator from ecodata, which is also aggregated by EPU, but in the later runs, we used the gridded bottom temperature product from Du Pontavice et al. (2023). For the final models to the WHAM model outputs, for both BT and SST, we instead cropped the data to the strata occupied by herring in the fall Northeast Fisheries Science Center (NEFSC) bottom trawl survey. We used slightly different months for BT (August-December) and SST (September-December for SST) because larvae are probably not present in the surface until about a month after the beginning of spawning. Using these same datasets and literature-derived estimates of thermal optima and limits, we also explored thermal threshold based indicators of both BT and SST (see Threshold-WP for details). The final model presented here included the following temperature variables (Figure 3 & Figure 6 in Threshold-WP):

- Mean fall BT (1 year lag, embryos and yolk sac larvae)
- Mean fall SST (1 lag, pelagic larvae)
- Optimal fall SST duration (1 year lag, early and late larvae)

Climate Variability

Early runs included all of the indices of climate variability listed below, but for the final model we only included the Gulf Stream Index (GSI) and the winter North Atlantic Oscillation (NAO) indicators (Figure 4).

- Winter NAO (5 year lag, delayed parental influences)
- GSI (0 year lag, juveniles)
- Marine heatwaves (0 & 1 year lag, larvae & juveniles)
- Warm slope water proportion (1 year lag, larvae)

Parameter tuning

One of the biggest challenges of BRT, especially for new users, is tuning the various parameters. Below is a brief description of what those parameters are, what they were set to and why, and what procedures were used to select them. We also describe the stochastic hyperparameters that control the various available algorithms for resampling.

Hyperparameters:

Number of trees: GBMs often require many trees but are prone to overfitting. It is therefore important to find the optimal number of trees that minimize the loss function. Examples of processes to identify the optimal number of trees are described in Elith et al. (2008) and Death (2007). Since the stock assessment model outputs used here are relatively short (1987 to present for WHAM and 1982 for ASAP), we opted to use k-fold cross-validation to estimate the optimal number of trees, which is explained in more detail below.

Learning rate or shrinkage: This parameter controls the rate at which the boosting algorithm descends the error surface and influences the contribution of individual trees to the final fit. In practice, the learning rate should be as small as possible as allowed by the data. However, more trees are required for smaller learning rates.

Interaction.depth: Some authors prefer to call this tree size, as it determines the number of splits in the tree, while others call it tree complexity. This parameter represents interactions; a value of one creates a decision "stump", or an additive model where each tree has just two terminal nodes or one split point. A value of 2 nodes allows for a 2-way interaction between predictors. Two-way interactions are included in this model, since this is one of the advantages of boosted regression. Preliminary runs that used an additive model to reduce complexity were bumped up to 2 to allow for interactions, which is more realistic.

n.minobsinnode: The minimum number of observations needed in each node or leaf is essentially how many observations constitute the homogeneous regions in the predictor space. Since less than 40 years of recruitment model estimates are available, we concluded that the common default value of 10 is too high and have opted to use a value of 5. This means that 5 data points form a cluster that justifies a split point in the tree.

Stochastic Hyperparameters

Training fraction: Model evaluation splits up the data into a training set to grow the trees and a testing dataset to predict onto for calculating predictive performance. If the training fraction is less than 1, then only that percentage of data is used to fit the model. For this study, we used a value of 0.9 due to the smaller size of our dataset. We should note here that one of the pitfalls of gradient boosted regression is that small changes in values, such as the training fraction, can alter the results rather dramatically. Some authors suggest default training fractions between 0.4-0.6, but our dataset was far too short to justify doing so.

Bag fraction: Using a bag fraction, or the subsampling fraction, controls the amount of stochasticity. For each subsequent tree that is built from the training data, random subsets of data are selected to fit each tree. The proportion of data included at each step is determined by the bag fraction, which introduces randomness into the model fit. Run-to-run variability can be suppressed by using a value of one, while bag fractions smaller than 1 will produce similar, but different, fits when running the same model. This is beneficial because stochasticity often improves predictive performance (Friedman, 2002). Slower learning rates reduce the contribution of each individual tree to the final model and can smooth out this randomness (Elith et al., 2008 Appendix S1). Larger bag fractions also make the effects of randomness less apparent. Since our dataset is relatively small, we elected to use bag.fraction = 0.7. Defaults in the various gbm packages are 0.5 and 0.75.

Optimal number of trees: In order to avoid over-learning, one of the first steps in the model fitting process is to determine the number of trees that minimizes prediction error. This can be accomplished in different ways depending on the size of the dataset. K-fold cross-validation (Hastie et al., 2001) is recommended as the approach to determine the optimal number of trees, especially for small datasets (De'ath, 2007; Leathwick et al., 2006). Cross-validation allows one to both test the model on withheld portions of data and still use all of the data at some point in the process to fit the model (Elith et al., 2008). Cross-validation is an iterative procedure where the number of folds determines the number of data sets and model runs conducted at each iteration. This is the general procedure for small datasets. First, the data is divided into *n* folds or subsets and a boosted tree is grown for each fold. This process is repeated for many trees, the cross validated (or performance) error is calculated for each tree, and then pooled across all trees. The minimum of the cross validated (CV) error is the resulting, optimal number of trees, which is then refit to the entire dataset to explore relationships between predictors and the response variable (De'ath, 2007).

Tuning approach

Methods, tutorials, and descriptions of processes to tune boosted regression trees abound in the literature and vary slightly between studies (see Elith et al., 2008; Elith & Leathwick, 2016; De'ath, 2007). In some cases, the first step is to choose the learning rate, then use CV validation to determine the optimal number of trees as explained above. Other approaches involve beginning with a high learning rate and determining an optimal number of trees for that rate. Then tune the learning rate with fixed hyperparameters to see how the models perform. The goal is to use the smallest learning rate that makes sense with the data and does not require too many trees. Once you have landed on a learning rate (we used 0.01), you tune the other tree parameters. At this point you can attempt to lower the learning rate again to assess any improvements in model accuracy. We used a function called gbm.perf (gbm3 version 3.0) to assess model performance. The model diagnostics plot shows how the squared error loss changes as more trees are added (Figure 5). This plot has lines for the training error (black), which decreases with more iterations and will not usually increase (De'ath, 2007), the cross validation error (green), and the test error (red), which usually reaches a minimum before flattening or increasing slightly. This plot also has a vertical dashed line to indicate the estimated optimal number of trees based on the CV results (minimum of the cv error). Subsequent analyses are conducted with a model refitted to that optimal number of trees.

Results

Time series of predictors included in the final model are shown here in Figures 1-4, with the exception of the threshold duration metric (see Figure 6 in Thresholds-WP). Preliminary results from several runs using the ASAP model outputs used no herring-specific spatial "cropping" and instead focused on either GB or the GoM EPUs. The top variables always included haddock spawning stock biomass with a one year lag in either the GoM and GB (HadSSB), sea surface temperature (OISST) and/or temperature indices such as the heatwave index, and zooplankton abundance (Figure 6A). However, some of the partial dependence plots showed sigmoidal relationships, for example between recruitment deviations and total zooplankton abundance, that we believe merited further exploration. Since this metric includes a wide variety of species, some of which are predators and/or competitors, we split up the total zooplankton abundance into larger functional groups: cnidaria, large copepods, and small copepods (see State of the Ecosystem/ecodata tech doc). We also used the *Calanus finmarchicus* stage indicators (see SOE/ecodata tech doc) to include one spatially and temporally specific variable, the abundance of stage 5 Calanus copepodites in the GoM with a one year lag. Separating the zooplankton was both reasonable and important because krill (included in total zooplankton abundance) are key food items for adults, gelatinous zooplankton are intraguild predators, and larvae feed on Calanus copepodites in the fall. After the split, the patterns in the marginal effects were more clear. Increasing abundance of gelatinous zooplankton caused recruitment to decline after a certain threshold was reached, and increases in calanus copepodites had positive effects on recruitment, again, after a threshold was reached (Figure 6B).

Removing variables with 0 or very low relative influence increased the deviance explained, but only when we started model fitting with a reasonably sized list. In some of the early runs with ASAP recruitment outputs, we added up to 3 year lags to most of the possible variables, which resulted in almost as many predictors as rows in the dataset. For this reason, subsequent models only included "smart" lags relevant to the time scale of the related hypothesis. For example, large copepods did not include a lag because they represent food for juveniles and adults in the year of recruitment, while small copepods were lagged by one year because they represent the food of larvae in the fall/early winter before they become recruits. Low influence variables were the primary productivity metrics (chlorophyll-*a* and primary production), the cold pool index, and temperatures with lags greater than 1 year, although SST with lags up to 2-3 years did occasionally land among the top variables.

Once the WHAM model was finalized, we ran several iterations of BRT models with the new package (gbm3) and consolidated predictors with updated temporal and spatial scales. For all groups of indicators that had two separate metrics for GB and GoM (except for jellyfish), the final model instead included just one predictor with a more herring-specific season and geography. For example, the two haddock variables used in earlier models (Figure 7) were replaced by the singular haddock predation index from the Haddock_predation-WP in the final model (Figure 8). After tuning and tweaking the BRT parameters and updating the predictor list, the order of relative importance was fairly consistent across models. Large copepods in the spring were always the most influential variable. Predator indicators were also in the top predictors of recruitment along with climate and temperature. Alternatively, the importance of small copepods (in both the fall and in the fall-winter version of the indicator, see Copepod Indices-WP for more details) changed from being amongst the top predictors to being less important in the final model. In general, fall conditions (temperature and zooplankton, 1 year lag) were less important than unlagged variables that represent the juvenile stage (Figure 8). The top variables were large copepods in the spring, gelatinous zooplankton, haddock predation, and the GSI.

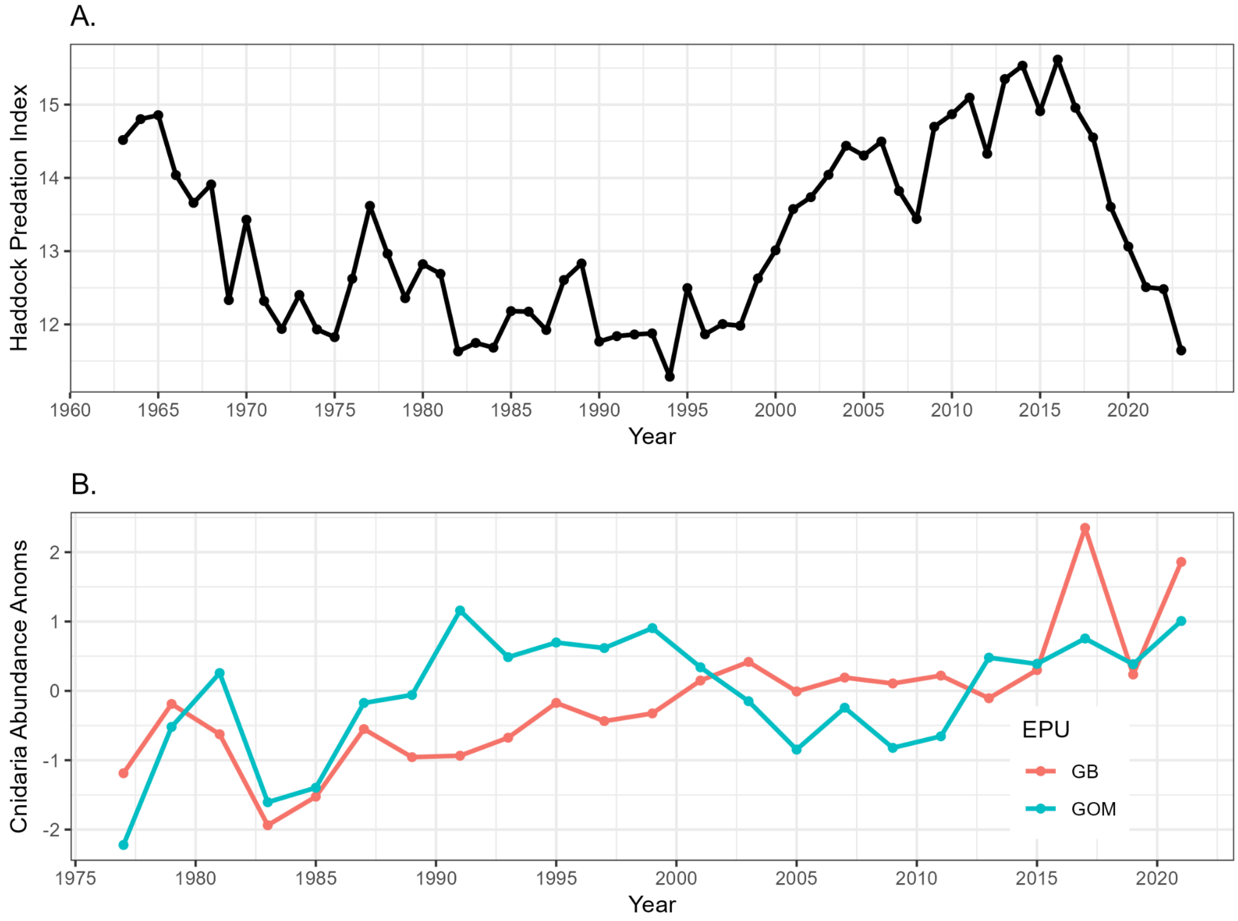
Marginal effects plots for the final model are not shown because this model included interactions, and so the relationships between a single predictor and the fitted values are more complex and difficult to visualize. These partial dependency plots did, however, indicate that the directions of some of the relationships changed between different model runs. We also uncovered a result that was opposite of our expectation; high values of large copepods corresponded with low values of recruitment. Finally, predictions from the final model using the optimal number of trees indicate a relatively good fit. We excluded the years from 2020-2023 in the final model because many of the predictors were missing values after 2020, and prediction skill for those last three years of data was always poor when they were included. This BRT model does not capture the peaks and valleys in recruitment, but it does represent the overall pattern fairly well (Figure 9). The goal of this research was to identify important processes that impact recruitment and to quantify drivers of year class success, not make perfect predictions. Therefore, using this BRT model to forecast recruitment is not advised.

Discussion

In this paper, we undertook an in-depth analysis of the factors that can affect Atlantic herring recruitment using a gradient boosted regression tree model. We identified food available for larvae and juveniles, egg and larval predation, and temperatures encountered by both larvae and juveniles as the strongest drivers. This result was consistent whether the estimated recruitment from the previous management track assessment or the current research track assessment were used. Our results identified three ecosystem variables that were tested as covariates on the recruitment term of the WHAM stock assessment model, but ultimately none were included in the final model for a variety of reasons. However, an array of processes influence herring recruitment and, while we believe we have identified a number of important drivers, there are other variables and methods that may be worthy to explore in the future (See ESP-WP). Perhaps there is no “smoking gun” to recruitment modelling and prediction and a more appropriate approach is to use some sort of variable combination technique, such as principal component analysis (PCA). Developing a multivariate index of stock health as in as in Boyce et al. (2019) might also be an avenue to incorporate multiple drivers. Finally, we recognize that the direction of some of the relationships identified here and their interactions merit further investigation.

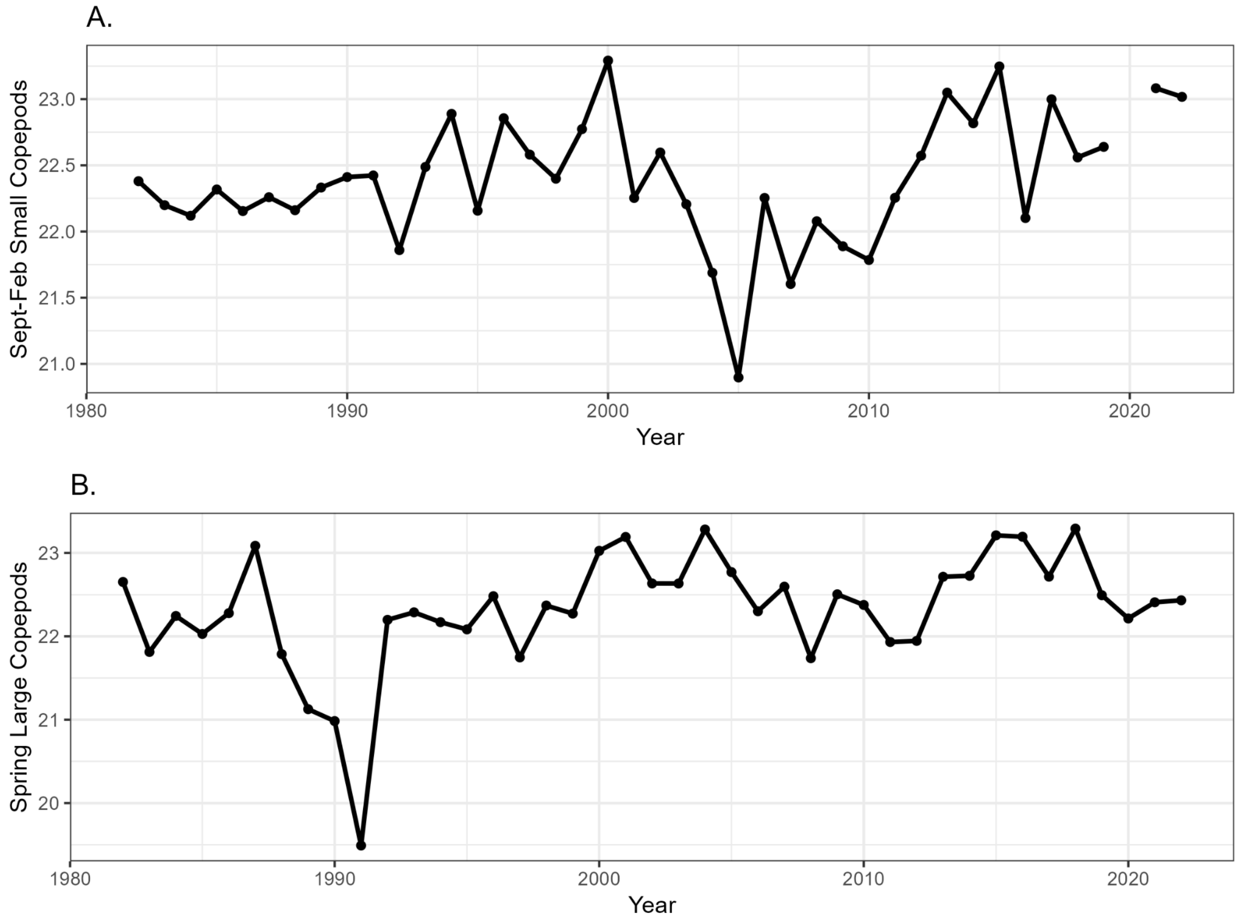
Figures

Figure 1. Upper Trophic Level



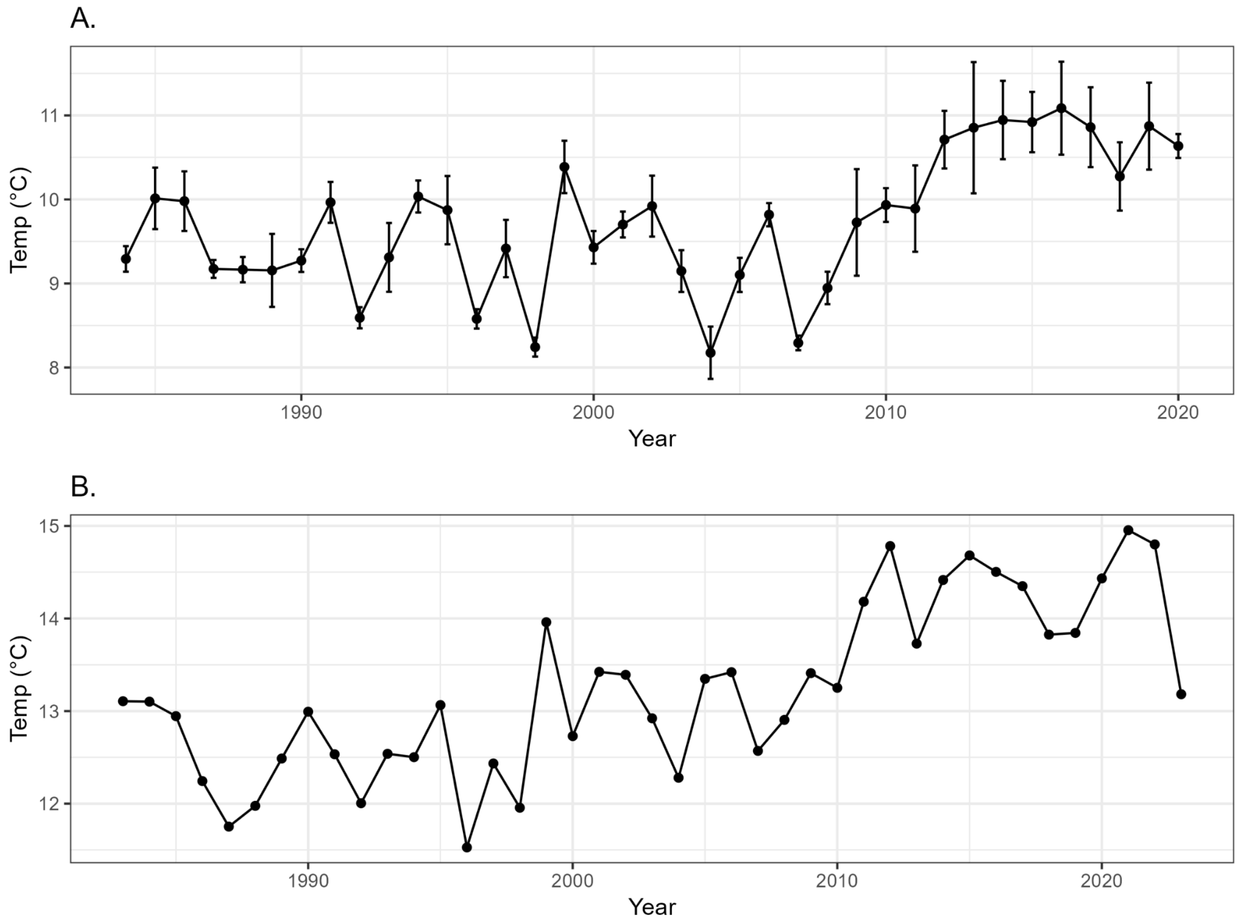
Time series of haddock predation index (A) and the cnidaria abundance anomalies (B).

Figure 2. Lower Trophic Level



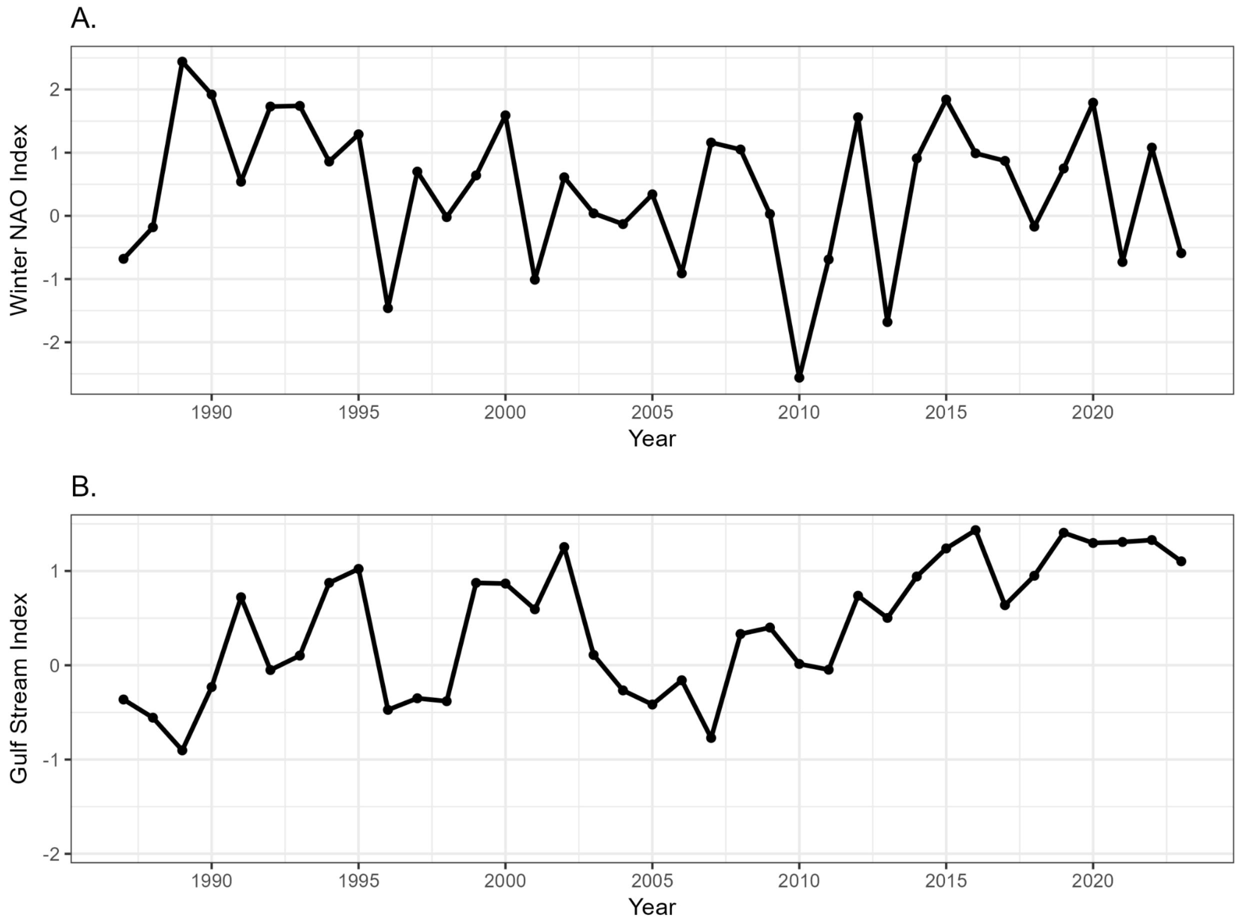
Time series of VAST zooplankton indices for small copepods between September- February (A) and large copepods in spring (B).

Figure 3. Physical Environment



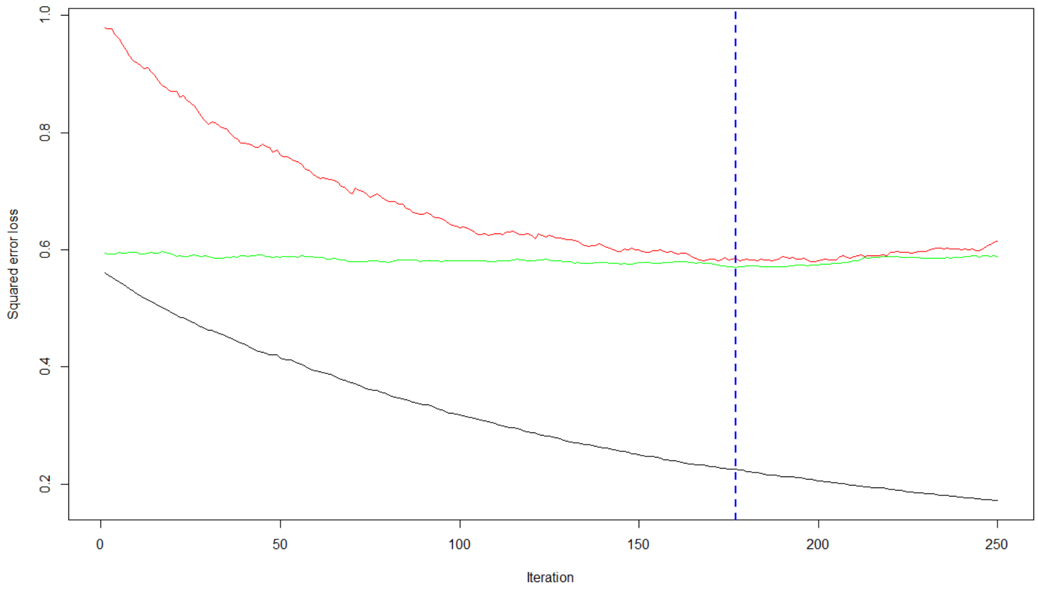
Average bottom temperature from August-December (A) and sea surface temperature from September-December (B) and in the fall NEFSC herring strata.

Figure 4. Climate Variability



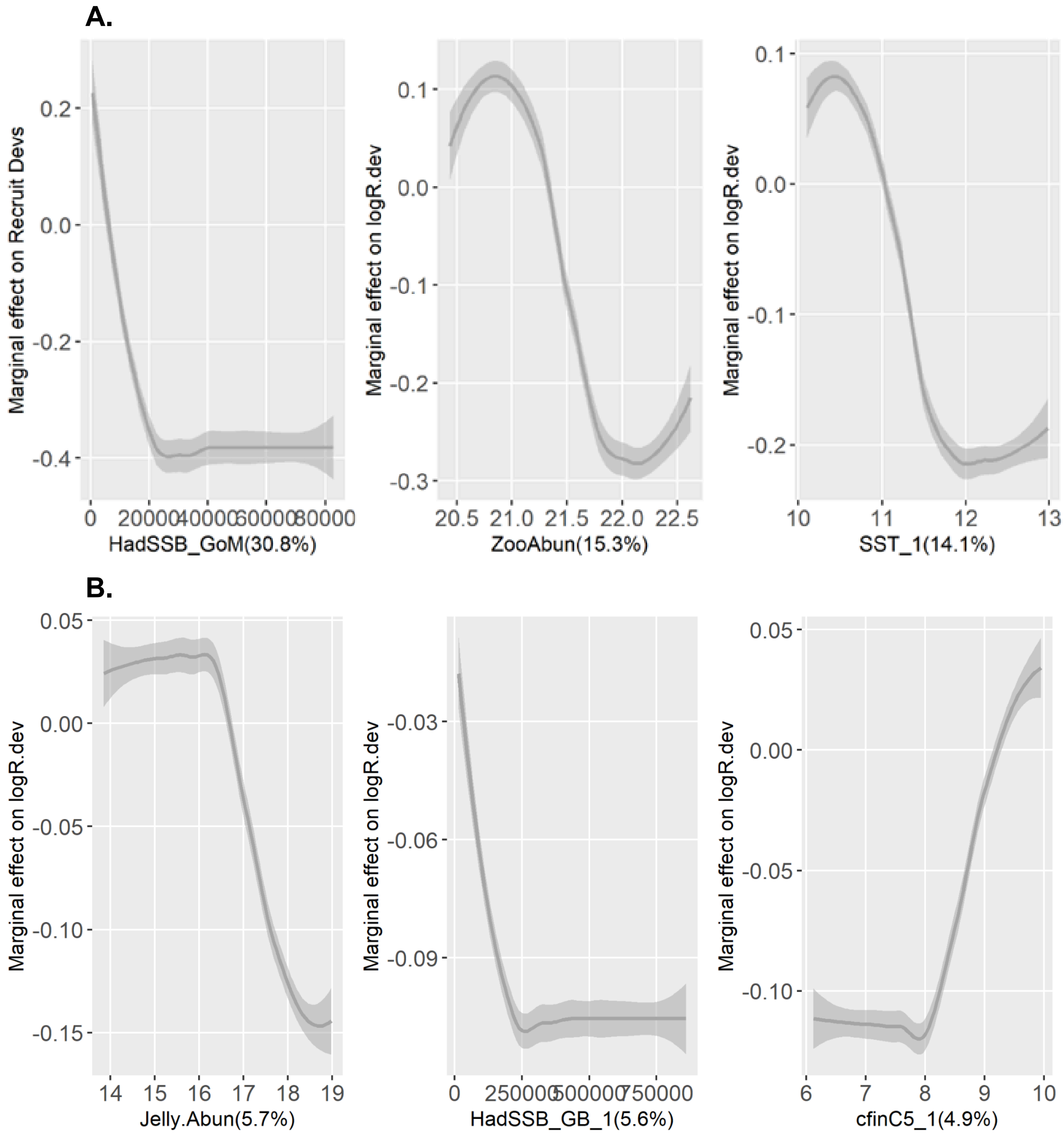
Indicators of climate vulnerability included in final model runs: Winter NAO Index (A) and Gulf Stream Index (B).

Figure 5. CV Diagnostics



Cross validation results of the chosen model shows how the prediction error (or accuracy) changes as trees are added. The black line is the training error, red is the test error, and green is CV error. Vertical dashed line indicates the optimal number of trees based on the CV results.

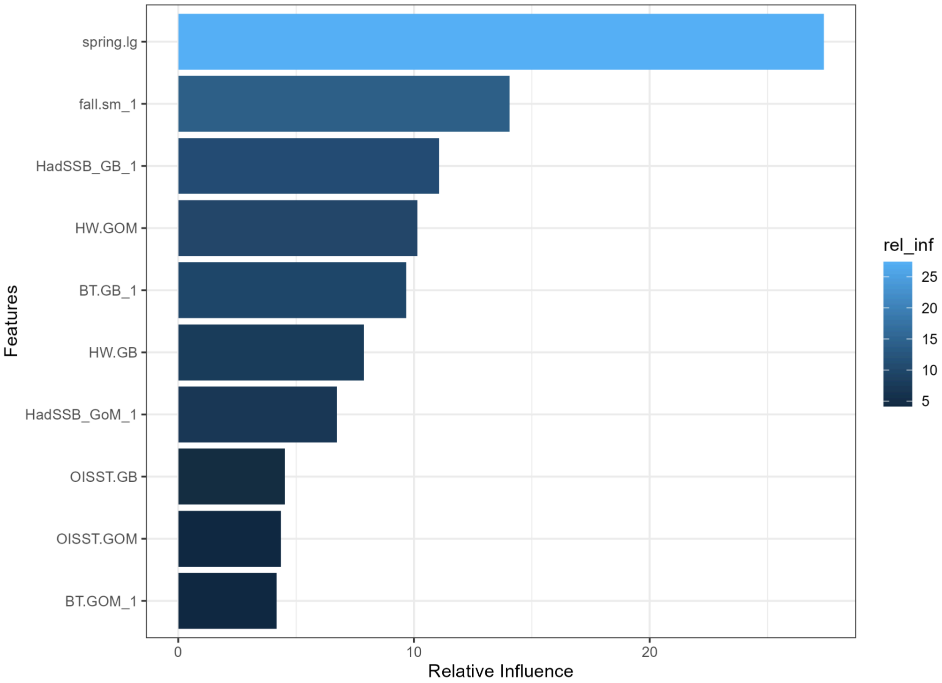
Figure 6. Example of marginal effects plot



Examples of marginal effects or partial dependence plots of the top three recruitment predictors from two different ASAP model fits.

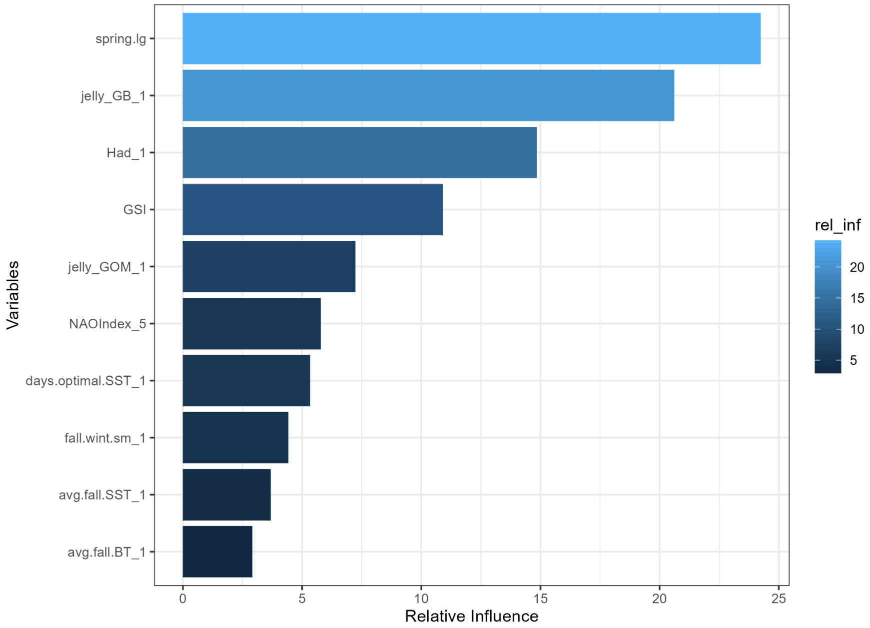
A. Top predictors in this model were haddock SSB in the GoM (HadSSB_GoM), total zooplankton abundance (Zoo.Abun), and sea surface temperature (SST_1)
B. Top predictors for subsequent models were cnidarian abundance (Jelly.Abun), haddock SSB in GB (HadSSB_GB) and the abundance of *Calanus finmarchicus* stage c5 copepodite (cfinC5_1)

Figure 7. Relative Influence



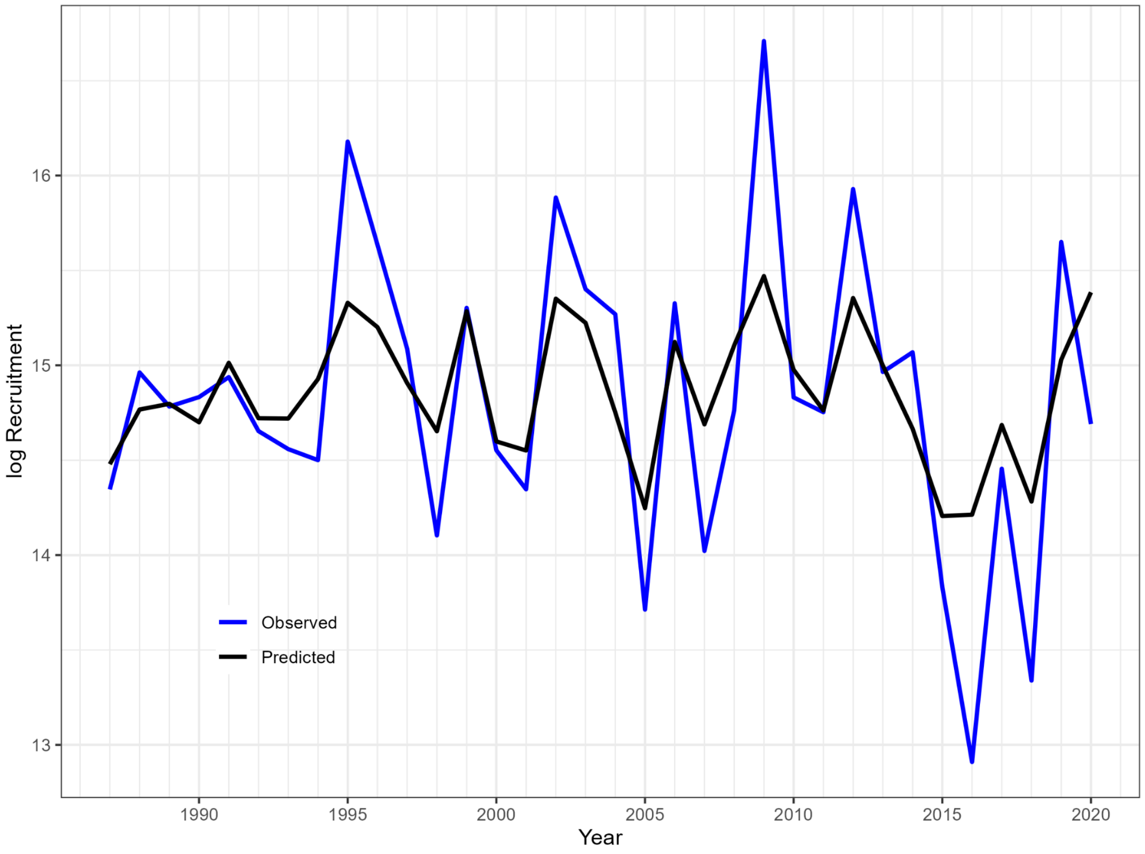
Relative influence of Atlantic herring recruitment predictors from previous model runs. Drivers: unlagged spring large copepods (spring.lg), lag 1 fall small copepods (fall.sm_1), lag 1 haddock spawning stock biomass on Georges Bank (HadSSB_GB_1), unlagged heatwaves in Gulf of Maine (HW.GOM), lag 1 bottom temperature on Georges Bank (BT.GB_1), unlagged heatwaves on Georges Bank (HW.GB), lag 1 haddock spawning stock biomass in Gulf of Maine (HadSSB_GOM_1), sea surface temperature on Georges Bank (OISST.GB), sea surface temperature Gulf of Maine (OISST.GOM), and lag 1 bottom temperature in the Gulf of Maine (BT.GOM_1).

Figure 8. Relative Influence (Final model)



Relative influence of Atlantic herring recruitment predictors from the final model.
Drivers: unlagged spring large copepods (spring.lg), lag 1 cnidaria abundance anomaly on Georges Bank (jelly_GB_1), lag 1 haddock predation index (Had_1), lag 0 Gulf Stream Index (GSI), lag 1 cnidaria abundance anomaly Gulf of Maine (jelly_GOM_1), lag 5 winter NAO index (NAOIndex_5), lag 1 duration of optimal sea surface temperature thermal window for larvae (days.optimal.SST_1), lag 1 fall-winter small copepods (fall.wint.sm_1), lag 1 fall sea surface temperature (avg.fall.SST_1), lag 1 fall bottom temperature (avg.fall.BT_1).

Figure 9. Observed vs. Predicted Recruitment



Final model fit using the optimal number of trees. Predicted (black) are the fitted values from the model and observed are the actual WHAM model recruitment estimates (log scale, blue)

References

Beltz, B., Bastille, K., Hardison, S., 2024. ecodata: Documentation of Ecosystem Indicator Reporting.

Boyce, D.G., Petrie, B., Frank, K.T., 2019. Multivariate determination of Atlantic herring population health in a large marine ecosystem. ICES Journal of Marine Science 76, 859–869.

De'ath, G., 2007. BOOSTED TREES FOR ECOLOGICAL MODELING AND PREDICTION. Ecology 88, 243–251. https://doi.org/10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2

De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81, 3178–3192.

Du Pontavice, H., Chen, Z., Saba, V.S., 2023. A high-resolution ocean bottom temperature product for the northeast US continental shelf marine ecosystem. Progress in Oceanography 210, 102948.

Elith, J., Leathwick, J., 2016. Boosted Regression Trees for ecological modeling.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77, 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Friedman, J.H., 2002. Stochastic gradient boosting. Computational statistics & data analysis 38, 367–378.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics 1189–1232.

Friedman, J.H., Meulman, J.J., 2003. Multiple additive regression trees with application in epidemiology. Statistics in medicine 22, 1365–1381.

Hastie, T., Tibshirani, R., Friedman, J.H., 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

Hickey, J., Metcalfe, P., Ridgeway, G., Schroedl, S., Southworth, H., Therneau, T., 2024. gbm3: Generalized Boosted Regression Models.

Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., 2023. dismo: Species Distribution Modeling.

Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., Smith, T., Zhang, H.-M., 2021. Improvements of the daily optimum interpolation sea surface temperature (DOISST) version 2.1. Journal of Climate 34, 2923–2939.

Leathwick, J.R., Elith, J., Francis, M.P., Hastie, T., Taylor, P., 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. Marine ecology progress series 321, 267–281.

Lyashevska, O., Harma, C., Minto, C., Clarke, M., Brophy, D., 2020. Long-term trends in herring growth primarily linked to temperature by gradient boosting regression trees. Ecological Informatics 60, 101154. https://doi.org/10.1016/j.ecoinf.2020.101154

Miller, T.J. (Timothy J., Legault, C.M., 2015. Technical details for ASAP version 4. https://doi.org/10.7289/V57W695G

R Core Team, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical

Computing, Vienna, Austria.

Ridgeway, G., 2024a. Generalized Boosted Models: A guide to the gbm package. Update 1, 1.

Ridgeway, G., 2024b. gbm: Generalized Boosted Regression Models.

Schapire, R.E., 2003. The boosting approach to machine learning: An overview. Nonlinear estimation and classification 149–171.

Smith, E.H., 2019. Using extreme gradient boosting (XGBoost) to evaluate the importance of a suite of environmental variables and to predict recruitment of young-of-the-year spotted seatrout in Florida. BioRxiv 543181.

Stock, B.C., Miller, T.J., 2021. The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time- and age-varying processes via random effects and links to environmental covariates. Fisheries Research 240, 105967. https://doi.org/10.1016/j.fishres.2021.105967