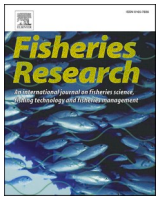
Fisheries Research 281 (2025) 107253

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres



Utilizing the random forest algorithm and interpretable machine learning to inform post-stratification of commercial fisheries data

Jason Gasper [a,*], Jennifer Cahalan [b]

[a] NOAA Fisheries, Alaska Regional Office, 907 W. 9th Street, Juneau, AK 99802, USA
[b] Pacific States Marine Fisheries Commission, 7600 Sand Point Way NE, Seattle, WA 98115, USA

ARTICLE INFO

Handled by Andre Eric Punt

Keywords:
Random Forests
Alaska Fisheries
Post-Stratification
Machine Learning
Catch Estimation

ABSTRACT

Federal groundfish fisheries off Alaska are managed based on near-real time estimates of catch generated using a combination of data from the North Pacific Groundfish and Pacific Halibut Observer Program, which deploys observers and Electronic Monitoring systems into the fisheries to sample catch, and industry-reported information. Catch is carefully monitored against limits that are based on biological constraints, quota allocations, or to control discard amounts. However, estimates of fish discarded at-sea (not retained for sale) can have large variance due to factors such as fishing behavior, species-specific vulnerability to fishing, and sample sizes. Post-stratification is a statistical approach widely used to improve the precision of catch estimates within a population because it controls for variance while also not relying on covariates known prior to sampling, which can be costly to collect or are unknown. Strategic use of post-stratification may increase the precision of estimates when compared to designs without post-stratification. However, choosing fishery characteristics to define post-strata may be elusive due to the high dimensionality of fishery data and complexity of creating post-strata that are optimized for multiple species. We propose a novel application of random forest classification and design-based estimation to explore multivariate post-stratification designs. These designs were evaluated by selecting the best performing trees from an ensemble using design-based estimation metrics. Results showed a large improvement in the precision of estimates by using the best-performing trees to label data and create post-strata. Moreover, through the use of subject matter expertise to evaluate the best performing trees, this method identified combinations of covariates that were not considered in previous estimation designs, and allows for exploration and testing of alternative post-strata designs that could be implemented in a management system.

1. Introduction

The Magnuson Stevens Fishery Management Conservation Act (2007) requires that U.S. Federal fisheries comply with a range of conservation and management measures, including the use of annual catch limits, in order to sustainably manage fisheries. Catch limits can be at the stock level where they are used for stock assessments and total allowable catch accounting towards biological limits (e.g., overfishing levels, allowable biological catch limits, annual catch limits), or at a management level where they are used for monitoring fishery-specific allocations. However, variability in catch estimates derived from data from sampling programs (e.g., observer programs and electronic monitoring) can be significant due to differences in gear types, fishing areas, seasons, target stocks, vessel operations, and bycatch species in addition to sampling variance.

Estimation methods used to manage catch in large and complex fisheries often have multiple objectives to satisfy, which complicates controlling for variation in estimates. Typically, a single estimation method, based on a complex set of code and procedures, is used to process all catch estimates in a database. Often post-stratification methods are used in combination with design-based estimation to increase the precision of estimates (e.g., Little 1991; Cochran, 1977; Williams, 1962; Hansen et al., 1953). This involves grouping similar data across multiple dimensions (e.g., area, time, species, gear) simultaneously rather than treating each dimension separately. By grouping similar data together and using a stratified estimation approach, the overall variance of the estimates will be lower than that of a simple random sample. This approach allows an estimator to be efficiently configured to the features of a population, with few assumptions, which would otherwise be constrained by the sampling design and associated

* Corresponding author.
E-mail address: jason.gasper@noaa.gov (J. Gasper).

https://doi.org/10.1016/j.fishres.2024.107253
Received 30 April 2024; Received in revised form 16 December 2024; Accepted 16 December 2024
Available online 2 January 2025
0165-7836/Published by Elsevier B.V.

costs (Holt and Smith, 1979). In such cases, post-stratification can improve the precision of estimates at a lower cost than increasing the sample size or changing sampling designs (Westfall, 2021, Cochran, 1977). These methods have been widely used in fishery catch estimation because, in addition to cost-savings, post-stratification methods are easily understood, implementable in database estimation systems, and incorporate covariates not known prior to sampling (e.g., NMFSOST, 2023; Cahalan et al., 2014; Mini et al., 2009). However, identifying groupings can be challenging when numerous potential covariates are involved, particularly regarding time and spatial areas, as well as other characteristics that may differentiate population components (e.g., gear, depth fished, species targeted, and vessel characteristics).

Post-stratification is also recognized as a model-assisted method that utilizes categorical covariate information to inform an estimator (Breidt and Opsomer, 2017; Cochran, 1977). A variety of statistical and modeling methods can be used to identify important data groupings used for post-stratification (e.g., GLM, decision and regression trees, statistical tests). The intent of this paper is not to comprehensively review all potential modeling methods, but rather to focus on post-stratification as a model assisted method that uses decision trees to identify homogeneous groupings of data to improve the precision of estimates.

Tree-based methods, such as regression trees for continuous outcomes and decision trees for categorical outcomes, are model-assisted methods used to evaluate non-parametric data for defining post-stratification, especially when there are many covariates to evaluate (e.g., Minami et al., 2024; McConville and Toth 2018; Phipps and Toth 2012). The random forest algorithm has been suggested in the literature as a potential model-assisted decision tree method that has not been well developed for identifying post-strata (McConville and Toth 2018; Breidt and Opsomer 2017; Tipton et al., 2012). In fisheries, machine learning methods like random forests are used predictively: models are built and trained on existing data, then applied to new situations to forecast outcomes. Fishery-specific examples include prediction of illegal fishing (Watson et al., 2023), prediction of population biomass using acoustic surveys (Yassir et al., 2023), and electronic monitoring where a fish species is predicted from video imagery (e.g., Salman et al., 2020). However, in this paper, we utilize machine learning to enhance the precision of an estimator by defining data groupings (post-strata) rather than focusing on predicting a specific outcome with high accuracy. This approach facilitates the creation of post-strata that helps control the precision of estimates.

Random forest is a machine learning ensemble method that constructs multiple decision trees (i.e., a forest) independently by randomizing both covariates and training data to create each tree within the forest (Ho, 1995). For classification tasks, which are the focus of this paper, inferences and predictions are based on the majority outcome from an ensemble of trees. For example, inference on covariates used for random forest modeling can be evaluated through ensemble metrics like feature importance and Shapely values (Zhang et al., 2024), or using explainers such as Local Interpretable Model-agnostic Explanations (LIME, Ribeiro et al., 2016). Individual trees within an ensemble are typically of little interest because the goal is to leverage the predictive ability of the ensemble. However, interpreting the trees within the ensemble is useful for designing post-strata because each tree represents a best fit to a subset of both features and data, making it an "expert" on sub-components of a population. Building post-strata based on these population sub-components may reduce both the within and between post-strata variation, resulting in an overall decrease in variation for a population-level post-stratified estimate. Furthermore, the defined relationships among features within the hierarchical structure of an individual decision tree is essential for constructing post-strata in our use case.

The change in precision for a population-level design-based estimate (e.g., a simple mean or ratio) can be evaluated by labeling data according to each tree's logic and applying a design-based estimator using post-stratified methods based on the labeled data. Trees with the highest precision for the population will also reflect the most influential components of the population in terms of variability. This makes them useful for designing post-strata, while also allowing investigation into how individual tree configuration and their associated covariates influence the precision of population-level estimates. Further, the stochastic processes used in random forests allow for a range of tree configurations to be identified and evaluated with design-based estimation methods.

This paper uses federal fisheries off Alaska as a case study to demonstrate the practical application of using random forests for the design of post-strata used for estimation. There are two primary study objectives: (1) use the random forest simulation algorithm and resulting tree-based models to label data, create post-strata, and investigate the precision of post-strata definitions across multiple years of fishing activity; and (2) apply subject matter expertise to the best performing trees to generalize tree-specific features for practical post-strata applications. Additionally, the proposed method serves as a preliminary step in defining potential post-strata for an Alaska fishery application, and we outline avenues for future research towards implementation for management purposes.

1.1. *Alaska fishery case study*

Federal groundfish fisheries off Alaska are managed based on near-real time estimates of catch generated using a combination of data from the NOAA Fisheries, Alaska Fisheries Science Center's (AFSC's) North Pacific Observer Program (Observer Program), which deploys observers and Electronic Monitoring (EM) systems into the fisheries to sample catch, and industry-reported information. Catch is carefully monitored against limits that are based on biological constraints, quota allocations, or to control discard amounts. The Observer Program uses a stratified hierarchical monitoring design to collect data to meet a wide range of analytic needs including catch estimation, stock assessments, protected species (e.g., marine mammals, seabirds) bycatch monitoring, ecosystem modeling, among others. As a result, deployment of observers and electronic monitoring systems into the fisheries cannot be tailored to any individual species or fishery, meaning that data are collected throughout the entire commercial fisheries through random sampling.

This paper focuses on estimating at-sea discards in the fixed-gear fisheries off Alaska using hook-and-line (HAL) or pot gear and are required to carry observers for a randomly selected portion of their trips (the fixed gear partial coverage fisheries; NMFS, 2023). The primary fishery targets for these fisheries are Pacific halibut (*Hippoglossus stenolepis*), Pacific cod (*Gadus macrocephalus*), and sablefish (*Anoplopoma fimbria*). Current estimation processes use post-stratification to control variance of the discard estimates used to monitor species-specific annual catch limits, prevent overfishing, and manage allocations that are specific to season, gear, management programs, and area. Because sampling strata in this fishery are typically large, encompassing fishing trips targeting a variety of species and areas throughout the year, estimation based on simple mean or ratio estimators can have high variance. To improve the precision of estimation, current methods use post-strata defined by NMFS reporting areas, time periods, and fishery target (predominant catch). However, these post-strata are not tailored to specific fisheries or species, and the definitions are broadly defined. Estimates are automated and updated daily in a NMFS Alaska Regional Office (AKRO) database called the Catch Accounting System (CAS, Cahalan et al., 2014). These estimates are available to fishery participants and inseason managers in near-real time, making stable and well-documented algorithms that are essential to the management of the fisheries.

In line with our objectives, this case study evaluates the utility of random forest to define post-strata to be used in estimation of at-sea discards for a broad range of species and subpopulations of fishing activity. We compare the features of the best performing trees with those currently used in CAS to provide both an assessment of whether current

CAS features are selected in the trees and to identify new features that could be integrated into the post-strata design. The model-assisted tree-based methods provide both an opportunity to simplify CAS post-strata designs and also explore many new features, including species-specific estimation and complex relationships among features.

2. Methods

Both new and currently used post-strata features in CAS were evaluated in the random forest framework. The CAS currently employs a large number of features in its post-strata design, with estimation based on available landings and at-sea observer data. Details on the post-stratification methods used in CAS are beyond the scope of this paper and can be found in Cahalan et al. (2014). However, we highlight that features used in the CAS post-stratification design are generally included in the random forest analysis. Broadly, these features include federal reporting areas and large regional areas (Gulf of Alaska, Bering Sea, and Aleutian Islands, Fig. 1), predominant catch on a fishing trip (trip targets), gear-types, and different time scales based on available data (5-week or 3-month rolling averages, or annual time periods). These features are used in combination to group landings and observer data when estimating discards, creating complex post-stratification designs that are computed based on available data, in near-real time. For example, there are two fixed gear types (pot and HAL), 22 Federal reporting areas, 19 potential trip targets, 5 management programs, and moving average time strata, all of which are combined to form separate post-strata. It's also important to note that while estimation is species-specific, the post-stratification is not, but uses different designs depending on species being broadly defined as groundfish, prohibited species, or ecosystem species (NPFMC, 2020).

In evaluating the post-strata, we introduced engineered features and new variables not currently used in CAS. These features include depth, vessel size, species-specific landing weights, spatial scales smaller than federal reporting areas, and finer details on the management programs and areas being fished. This allows evaluation of both new and existing CAS post-strata simultaneously within the random forest framework. A broad overview of processes that correspond with the analysis components and workflow is presented in Fig. 2 and described the following sections.

2.1. Data Preprocessing

Fishery data for fixed-gear (pot and HAL), including annual estimates of landings and at-sea discard, were obtained for 2018–2022 for all fishing trips that were in the fixed gear partial coverage fisheries (NMFS, 2023). Observer data and landing information were summarized for each of these fishing trips by linking the at-sea with shoreside information with a unique trip identifier. The number of annual trips used in our study ranged from 409 (2018) to 326 (2021), with the exception of 2020 which had a low of 154 trips due to COVID-19 pandemic restrictions. The number of sampled hauls ranged from 3187 (2019) to 2505 (2018), excluding 2020 which had a low of 1100 hauls, resulting in an average annual total count of approximately 70,000 records that include species caught and areas fished. Note that 2020 was included in the analysis but was omitted from the discussion on fishery characteristics due to unusual sampling and fishing situation caused by

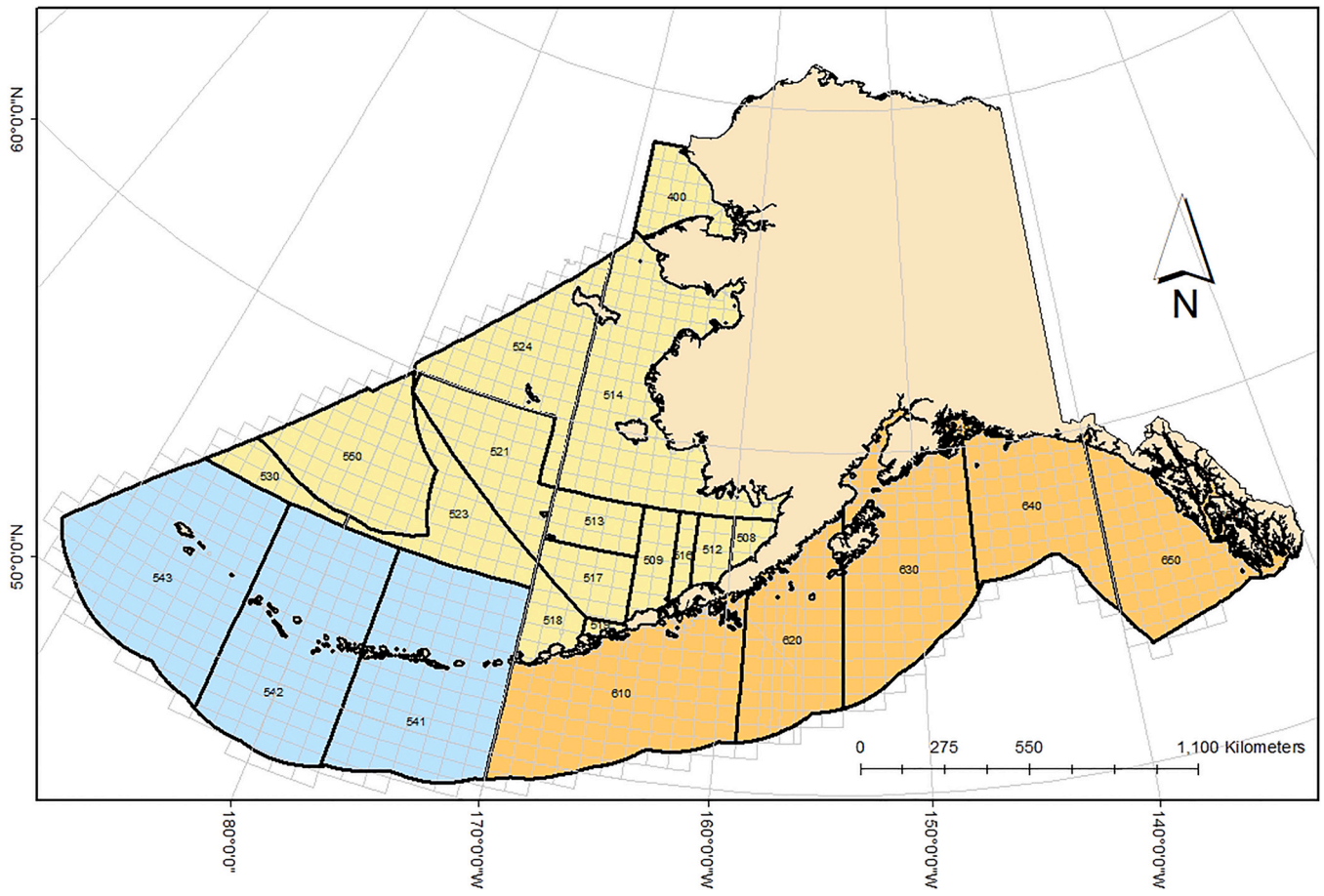


Fig. 1. Map of Alaska showing NMFS Management Reporting Areas (labeled) and ADF&G statistical areas (small grid). The region-specific color represents the following: light orange color represents the Bering Sea, light blue represents the Aleutian Islands, and orange represents the Gulf of Alaska. The Gulf of Alaska is subdivided into the western (610), central (620 and 630), and the eastern 640 and 650).

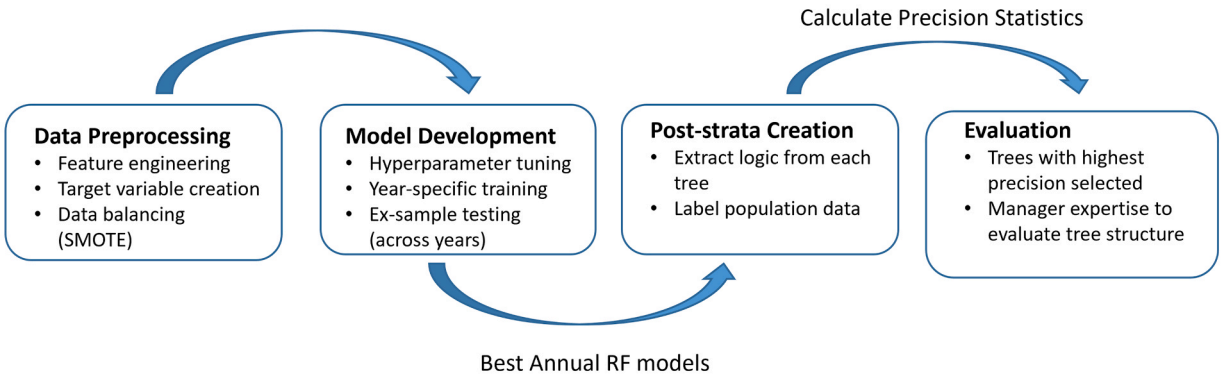


Fig. 2. Analytic workflow and development processes used to process data, fit and skill test the random forest model, label data based on tree logic, and identify potential post-strata and best performing decision trees.

COVID-19.

Multiple engineered features derived from bathymetry data, and aggregated for each Alaska Department of Fish and Game (ADF&G) statistical area (ADF&G, 2024, Fig. 1), were created. Depth was summarized for each 25 m^2 bathymetry polygon within each ADF&G area as the mean, median, and standard deviation. A scaled depth feature was also created across all depths, using the Python library Sklearn's minimum and maximum scaler (MinMax), by scaling the median depths across all ADF&G areas with fishing effort for each year, resulting in values near 1 representing the shallowest depths fished and values near 0 the deepest depths fished. The standard deviation of the depth polygons describes variation in bathymetric depth within an ADF&G area, such as steep slopes commonly found in gullies or throughout the AI where island chains have very steep slopes to deeper water relative to the low relief areas associated with the continental shelf (e.g., the BS).

Several different resolutions of management areas were included in the analysis as categorical variables. Federal groundfish Fishery Management Plan areas are broadly defined as the Aleutian Island (AI), Bering Sea (BS), and Gulf of Alaska (GOA) regions (Fig. 1). The GOA is further divided into the Western (areas 610 and 620), Central (area 630), and Eastern management areas (areas 640, 650, 659), which are consistent with the management of most groundfish species in GOA and generally align with IFQ sablefish areas, with the exception of the eastern GOA which is subdivided into west Yakutat and southeast Outside areas.

Additional data processing included transformations and the creation of new variables to enhance analysis. This included projecting latitude and longitude into meters, as well as generating indicator variables for species groups (20 species in total), management programs (IFQ, Open Access, State Managed Pacific Cod, State Managed Others, and State Managed Sablefish), and special management areas. Additional variables were created to account for vessel participation in the IFQ fishery, predominant species caught (trip target code), and temporal periods of interest (month, quarter, and first versus second half of the year).

2.2. Defining the target variable

Within each species group, discard weights were categorized into three distinct target-variable (the attribute being predicted for) classes: low, medium, and high discard weight. Catch amounts for the evaluated species are monitored by weight, so our target variable classes were based on weight rather than the number of fish. The classification was determined by analyzing the overall distribution of at-sea discard weights for each species and year evaluated (i.e., the distribution is comprised of haul and species records within a year). The 'low' class was assigned to species and hauls with less than 0.01 metric tons (mt) of discard; effectively grouping trips with no/minimal discards. The 'medium' class included discard events of 0.01 mt or more, but below the threshold where two-thirds of the species-specific discard distribution occurred. Finally, the 'high' class represented the upper third of the species-specific discard distribution; trips with high discard weights. These categories were chosen to separate the high discard amounts from the very low ("zeros") discard amounts while also providing adequate data in each category for training and validation. Identifying the "zero" category as response variable is similar to model-based estimation methods used for overdispersed distributions (e.g., zero-inflated data, mixture models, see Zuur et al., 2009, Gasper and Kruse, 2013). The categorization of discards provide a basis for post-stratification that, when used in conjunction with a design based estimator, will result in higher precision discard amounts (Thompson, 2012). By fitting the model to these discard categories, we can identify a set of covariates and relationships among covariates, in a decision tree structure, that define post-strata with lower within group (i.e., post-stratum) heterogeneity than between groups (i.e., among post-strata).

The dataset displayed an imbalance among classes, indicating highly unbalanced data among the target-variable classes, with approximately 85 % of the records belonging to the lowest discard class (<0.01 mt). Class imbalance is caused by certain classes being over-represented in a training dataset, potentially causing bias in the model during training. Given the high amount of class imbalance, Synthetic Minority Oversampling Techniques (SMOTE, Chawla et al., 2002) were explored to address the class imbalance during model training. These commonly employed techniques were implemented using the Python library Imblearn (Lemaitre et al., 2017) which includes methods for support vector machines (SVM), borderline, K-Means, and nominal-categorical (NC) oversampling methods. Oversampling techniques were applied to each year independently to avoid contaminating test data, which could occur if synthetic data were generated using combined years. Based on the testing results, which compared F1 scores (see Eq. (1)) from fitted random forest models for each year against an out-of-sample testing group spanning multiple years, the SMOTE SVM (with default parameters) method was the preferred method for addressing class imbalance. While this method did not explicitly investigate decision boundaries for the synthetic samples, which would be difficult due to the dimensionality of these data, it did provide a measure of the impacts of not balancing the data.

2.3. Model training and testing

Between year variability in fisheries characteristics, including catch composition and discard amounts, is difficult to predict for future years; hence post-strata definitions need to be effective in grouping trips with similar discard amounts across a range of potential fishery changes. To ensure that selected models identified post-strata definitions that were robust across years, training for the random forest was conducted on

each year of data, with testing done on all out-sample years. For example, a model trained on 2019 fishery information was tested on the combined 2018 and 2020–2022 fishery data and the effectiveness of data groupings (i.e. resultant post-stratification) in reducing variance was evaluated for each species and out-of-sample year separately. Alternatively, models could have been trained on multiple years and tested on individual out-of-sample years. However, this approach would have been less sensitive to significant changes in the fishery due to the combined years used in training. By training on individual years and testing across all out-of-sample years, we were able to evaluate model performance over time while also gaining the detailed insights needed to investigate the tree structure in relation to each year of fishing activity. This allowed an assessment of changes in fishing practices through time and how well a tree performed relative to those changes.

Hyperparameter tuning was accomplished using k-fold cross validation and searching the parameter space using the Python Sklearn package's GridSearchCV function (Pedregosa et al., 2011) to assess tree depth (number of levels), ensemble forest size (number of trees), and the maximum number of features selected for each training iteration. The GridSearchCV function searches across a predefined hyperparameter space to find an "optimal" model fit. The leaf size (minimum number of samples for each leaf) was set to a minimum of approximately 10 trips, thereby maintaining a reasonable sample size for variance calculation and mitigating the risk of overfitting (Westfall et al., 2011). Tree depth tuning was limited to a range of 2–12 levels to avoid overly complex trees that complicate interpretation of individual trees and potentially result in overfitting.

The performance of the random forest model, following hyperparameter tuning, was assessed using F1 scores, precision (P) and recall scores (R), and graphical inspection. The F1 scores assesses the predictive ability of the model and is a combination of P, which measures how many of the "positive" predictions (i.e., class membership predictions) were correct, and R which measures how many of the positives in the dataset (true class memberships) were correctly identified. F1 scores were calculated for the training and testing results for the ensemble for each target feature (class) and across all target features (classes). Additionally, plots were investigated across the grid search space to view the sensitivity in model performance based on the number of features selected, tree depth, and ensemble size. The F1 scores for each class, j (low, medium, high), calculated for each target variable class using the following equation: $F1_j = 2\frac{P_j R_j}{P_j + R_j}$, where

$$P_j = \frac{True\ Positives_j}{True\ Positives_j + False\ Positives_j} \text{ and}$$
$$R_j = \frac{True\ Positives_j}{True\ Positives_j + False\ Negatives_j} \tag{1}$$

The weighted F1 scores for calculation cross classes were calculated as follows:

$$F1_{weighted} = \frac{1}{N}\sum_{j=1}^{3} n_j F1_j \tag{2}$$

where n_j is the number of observations in class j, and N is the total number of observations across all classes. A weighted F1 score was chosen over a non-weighted version due to the class imbalance in the testing data, noting that the training data was balanced using SMOTE and thus the weighted F1 and non-weighted F1 are equivalent.

2.4. Evaluation of post-strata

Post-strata definitions were created based on the final selected model, using the tree leaf covariates to define individual strata and label the population data. For post-stratification to be effective, fishing trips that are similar relative to the estimation need (at-sea discard weight) will be grouped together in a post-stratum. To assess whether post-strata are more similar than the population as a whole (the entirety of the sampling strata), the post-stratified population variance of the species-specific mean discard per trip was estimated, omitting sample-variance terms such as the finite population correction (FPC), and compared to the variance for the non-post-stratified population. A lower between-trip variance for post-stratified data would indicate that incorporating post-stratification through tree-based definitions can reduce the variance of estimates relative to non-post-stratified estimation.

To estimate the overall post-stratified mean discard per trip, the proportion of the population trips in each post-stratum (N_p/N) is used to weight the post-stratum-specific means (Eq. (3)), where ($Y_{ip} = \mu_{ip}$) is the known mean discard weight per trip for the post-stratum $p = 1, 2, \ldots P$, N is the known total number of trips in the population, N_p is the known number of trips in post-stratum p, *and* i indexes trips ($i = 1, 2, \ldots, N$). Subscripting for sampling strata has been omitted for simplicity in Eqs. (1) through (4). Trips were assigned to each post-strata and the between-trip variance was calculated for the predefined suite of species. Note that the post-stratified mean for a sampling stratum is based on the population as a whole (N trips) and is equal to the non-post-stratified mean for that stratum (see Thompson, 2012; Cochran, 1977).

$$\overline{Y}_{post-stratified} = \sum_{p=1}^{P}\frac{N_p}{N}\overline{Y}_p = \sum_{p=1}^{P}\frac{N_p}{N}\frac{\sum_{i=1}^{N_p}\overline{Y}_{ip}}{N_p} = \frac{1}{N}\sum_{p=1}^{P}\sum_{i=1}^{N_p}Y_{ip} = \frac{1}{N}\sum_{i=1}^{N}Y_i = \overline{Y} \tag{3}$$

The post-stratified population variance is the weighted sum of the variances for each of the post-strata (Eq. (4)), where the within post-stratum variance is the best unbiased estimator of variance given that the true population variance, σ^2, is known (see Casella and Berger, 2002; Holt and Smith, 1979).

$$\begin{aligned}\mathrm{var}\left(\overline{Y}_{post-stratified}\right) &= \sum_{p=1}^{P}\left(\frac{N_p}{N}\right)S_p^2 = \sum_{p=1}^{P}\left(\frac{N_p}{N}\right)\frac{\sum_{i=1}^{N_p}(Y_i-\overline{Y}_P)^2}{N_P} \\ &= \left(\frac{1}{N}\right)\sum_{p=1}^{P}\sum_{i=1}^{N_P}(Y_i-\overline{Y}_P)^2\end{aligned}$$

$$\mathrm{var}\left(\overline{Y}_{post-stratified}\right) = \left(\frac{1}{N}\right)\sum_{p=1}^{P}\sum_{i=1}^{N_P}\left(Y_i - \frac{\sum_{i=1}^{N_p}Y_{ip}}{N_p}\right)^2 \tag{4}$$

Because the between-post-stratum variability does not contribute to this post-stratified variance, the post-stratified variance does not simplify to the variance of the mean for the non-post-stratified mean (Eq. (5)). For this same reason, if post-strata are chosen appropriately, the post-stratified variance can be expected to be less than non-post-stratified variance.

$$\begin{aligned}\mathrm{var}\left(\overline{Y}_{post-stratified}\right) &= \left(\frac{1}{N}\right)\sum_{p=1}^{P}\sum_{i=1}^{N_P}\left(Y_i - \frac{\sum_{i=1}^{N_p}Y_{ip}}{N_p}\right)^2 \neq \frac{\sum_{i=1}^{N}(Y_i-\overline{Y})^2}{N} \\ &= \mathrm{var}(\overline{Y})\end{aligned} \tag{5}$$

To assess the effectiveness of the post-stratification, the difference between the post-stratified and non-post-stratified variance estimates was computed, with each variance estimated using the best unbiased estimator given that σ^2, is known (see Casella and Berger, 2002).

Finally, to evaluate the impact of post-stratification on variance estimates and compare the variance savings across numerous species with varying discard characteristics, we assessed each tree's difference relative to the true variance of the mean for the non-post-stratified data. The

relative change in precision (%RP, Eq. (6)) for each testing year and species (or species group) was calculated by comparing the variance of the mean for a preferred tree (i) to those of the species or species group (k) for a given testing year (t). This result provides a measure of how much the preferred tree reduces the variance compared to the unstratified group. This measure is the percent difference between the CVs of the post-stratified and unstratified groups, where negative values indicate a gain in precision due to post-stratification. The best performing trees (preferred trees) were defined as those that had the lowest median %RP across all testing years.

$$\%RP = \frac{\sqrt{\mathrm{var}(\overline{Y}_{i,k,t})}*\frac{1}{\overline{Y}_{i,k,t}} - \sqrt{\mathrm{var}(\overline{Y}_{k,t})}*\frac{1}{\overline{Y}_{k,t}}}{\sqrt{\mathrm{var}(\overline{Y}_{k,t})}*\frac{1}{\overline{Y}_{k,t}}}*100 = \left(\frac{\sqrt{\mathrm{var}(\overline{Y}_{i,k,t})}*\frac{1}{\overline{Y}_{i,k,t}}}{\sqrt{\mathrm{var}(\overline{Y}_{k,t})}*\frac{1}{\overline{Y}_{k,t}}} - 1\right)*100 \quad (6)$$

3. Analysis and results

3.1. Model training and testing

Models trained across a range of tree depths showed greatest improvements to skill testing at tree depths up to 10, with most years showing less sensitivity for greater tree depths (Fig. 3). Tree depth was constrained to 7 for the random forest used for final testing and prediction because while allowing trees a greater depth resulted in improvements in training fits, the greater depths resulted in large increases in tree complexity that would not be implementable in the CAS. Further, the increased complexity beyond a depth of 7 yielded only small improvements in training F1 scores, typically less than 10 %, which did not warrant the added tree complexity (Fig. 3). The number of selected features at tree depth 7, based on grid search performance, was 30 for 2018 and 2020–2022, and 20 for 2019. Differences between F1 scores and whether 20 or 30 features were selected were generally small (~1 % difference in F1 scores), whereas inclusion of only 10 features in the hyperparameter setting generally resulted in substantially lower F1 scores (Fig. 3).

Upon examination of the model training performance across target variable classes (discard amounts of low, medium, or high), the models demonstrated strong training performance at a tree depth of 7, with F1 scores falling in the range of 0.56–0.71 (Fig. 3). Notably, the lowest scores were observed in the year 2020, which coincided with the COVID-19 pandemic. During this time, fisheries and observer coverage experienced unusual patterns, likely contributing to the dip in model fit performance (lowest training accuracy score was 0.59, Table 1). Additionally, the minimum leaf size on individual trees was restricted to approximately 10 fishing trips to reduce the potential for overfitting. This decision was based on findings in Westfall et al. (2011) that post-stratum sample size should be at least 10 to provide a stable estimate of means. Restricting the tree depth to 7 levels resulted in leaf sample sizes generally not being close to the 10-trip minimum defined as a constraint on the minimum number of records in a leaf.

Skill testing of the trained random forest ensemble demonstrated an overall high accuracy of classification (55–81 %, Table 1) and F1 scores (0.64–0.82, Table 1), suggesting reasonable predictive performance relative to the entire testing population. However, it is important to note that the predictive performance varied depending on the class of the target variable (i.e., a high, medium, low discard amount). High F1 scores for the out-of-sample data were observed for the 'low discard' class (0.71–0.91, Table 1), indicating that the ensemble performed well in predicting this class, and this class accounted for approximately 85 % of the testing records (unbalanced). In contrast, F1 scores were notably lower for the 'medium' and 'high discard' categories (0.13–0.25, Table 1), showing limited class-membership predictive performance in these categories. These skill testing patterns remained consistent regardless of which year was used for training, including 2020 which exhibited the lowest skill testing results. The poor skill testing scores in the 'high' and 'medium' categories reflect heterogeneity in discard

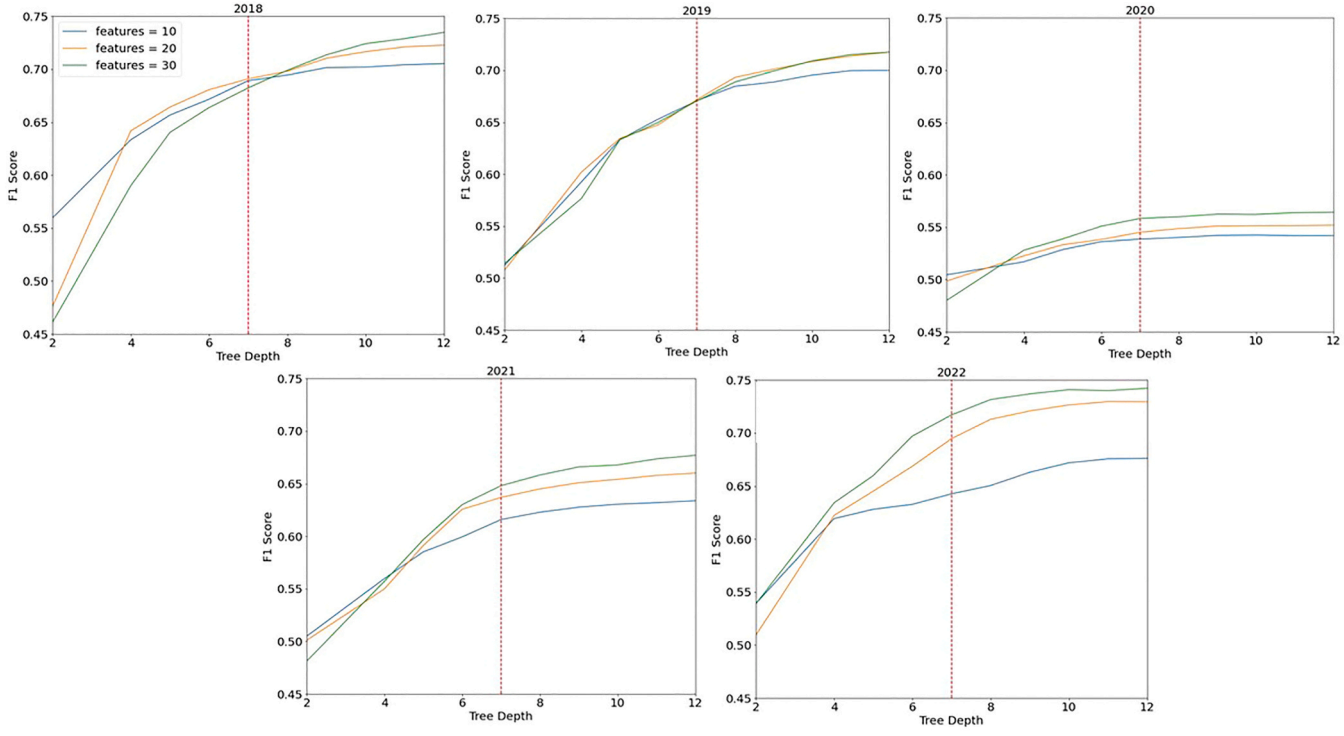


Fig. 3. Grid search results comparing F1-scores across the number of training features and tree depth, with the preferred depth of 7 denoted by the red dotted vertical line.

Table 1
Summary of skill testing for random forest models, with training and testing F1 scores specific to the target variable classes of low, medium, or high.

Training Year	Accuracy		Testing F1				Training F1			
	Testing	Training	All	(low)	(medium)	(high)	All	(low)	(medium)	(high)
2018	0.73	0.72	0.80	0.88	0.22	0.20	0.70	0.72	0.70	0.69
2019	0.81	0.70	0.82	0.91	0.14	0.25	0.70	0.65	0.74	0.70
2020	0.58	0.59	0.66	0.73	0.16	0.17	0.59	0.62	0.50	0.64
2021	0.55	0.68	0.64	0.71	0.17	0.18	0.68	0.71	0.64	0.68
2022	0.68	0.74	0.74	0.82	0.13	0.22	0.72	0.78	0.71	0.72

amounts across spatial areas, gear types, and time. In contrast, low discard amounts were generally associated with specific fishery characteristics, such as species, depth fished, and certain gear types, often where the discard amount was very low for those species that were retained (i.e., to be sold). However, while detail is provided on the underlying data associated with each class, the logic used to label data and define post-strata is not specific to these individual class outcomes, but rather the overall decision tree structure. This detail is provided for context with how well the model performed on specific classes for training and testing.

3.2. *Efficacy of post-stratification*

An increase in precision was observed across nearly all trees in the forest when combined over testing years (Fig. 4). The median precision gain for the forest across all species and on testing data was approximately 55 %, with the lower quartile of the forest distribution among most species showing precision gains of more than 75 %. Within the forests, some trees showed low to no gains in precision (as indicated by the long right tail in the density plot in Fig. 4). This is expected and indicates that the selection of features and training data during the random forest fitting process resulted in trees optimized for subsets of the population that contributed minimally to the overall variation.

The precision gain for the best performing tree across all species was substantial, with the upper interquartile range (IQR) limit generally below −65 %. Overall, the IRQ upper and lower values (and "whiskers") had relatively small ranges. Thus, the overfitting did not create a situation where precision results were widely varying such that we would conclude there was poor generalization to out-sample years. Variability in precision gain among species was observed among the preferred trees, as shown by the differences in the IQR of the CV ratios across at testing years and species (Fig. 4). The less commonly discarded species demonstrated smaller gains in precision (sablefish and arrowtooth), whereas commonly discarded species showed higher gains in precision (skates, halibut, Pacific cod, sharks, and sculpins). The distribution of all trees (light blue in Fig. 4) showed a long right skew for all species, which indicates that there are many trees that would have shown substantial improvement in precision, and the selected best performers generally performed above the distribution median. The target variable included all species and thus we did not optimize the model for a specific species, but the best trees performed well across multiple species for most years.

3.3. *Fishery patterns*

Consultation with in-season managers on the tree structures revealed that, although tree structures and features varied across years, there was substantial overlap in the fishery behaviors being described. Managers were able to identify specific fisheries within the tree structures even though the branch logic was different among trees. For example, Federal regulations supporting IFQ programs governing Pacific halibut (primarily targeted using HAL gear) and sablefish (targeted using both HAL and Pot gear), and non-IFQ harvest of Pacific cod, create the observed

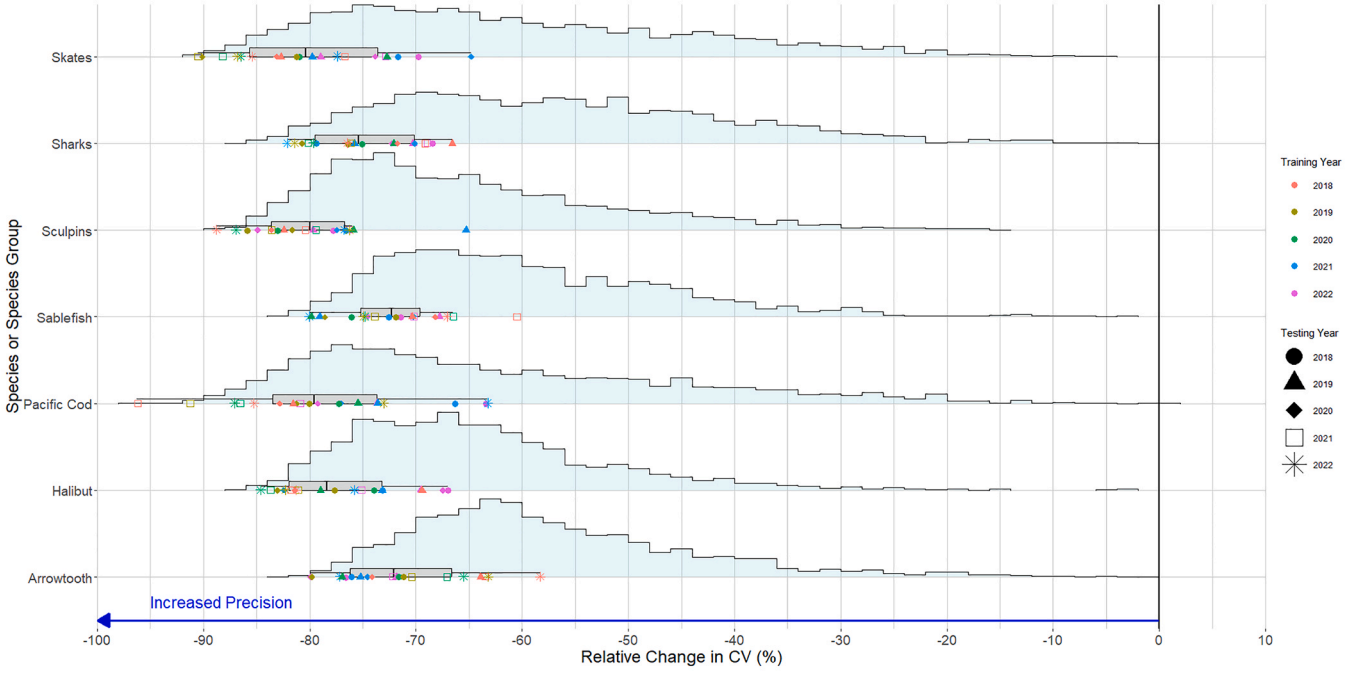


Fig. 4. The distribution of the relative percent change in the precision (%RP) for all trees (blue histogram) and preferred trees (boxplot and points) across all testing years. The points for the preferred tree are colored based on the year of data used for training and the shape is specific to the testing year of data. The relative change in CV (x-axis) is defined in Eq. 6.

patterns in the trees across years. The interaction among these three major fisheries was captured within the features and logic of the decision trees (Fig. 5, Table 2). The features included in the first five levels of the decision trees included gear type fished (Pot or HAL); metric tons (mt) of Pacific halibut landed (LWHLBT); mt tons of Pacific Cod landed (LWCod), IFQ program flag (IFQ); length of vessel overall (LOA); depth or slope of an ADF&G area (AvgDpth, scaled depth, σDpth, WtrIn, WtrOut); total mt of all species landed (LW); longitude (Lng); latitude (Lat); mt of sablefish landed (LWSable); mt of groundfish landed (RetGF); landing amounts (mt) of deep water rockfish species such as Thornyhead Rockfish (LWThds), other rockfish (LWRck), and shortraker rockfish (LWSR), and region specific landings (LWLskt, LWOctp). Date breaks corresponding to management events were also noted by managers upon review of the trees.

Managers were able to explain an import shift in the tree structures due to the sablefish fishery moving from HAL to pot gear. The first

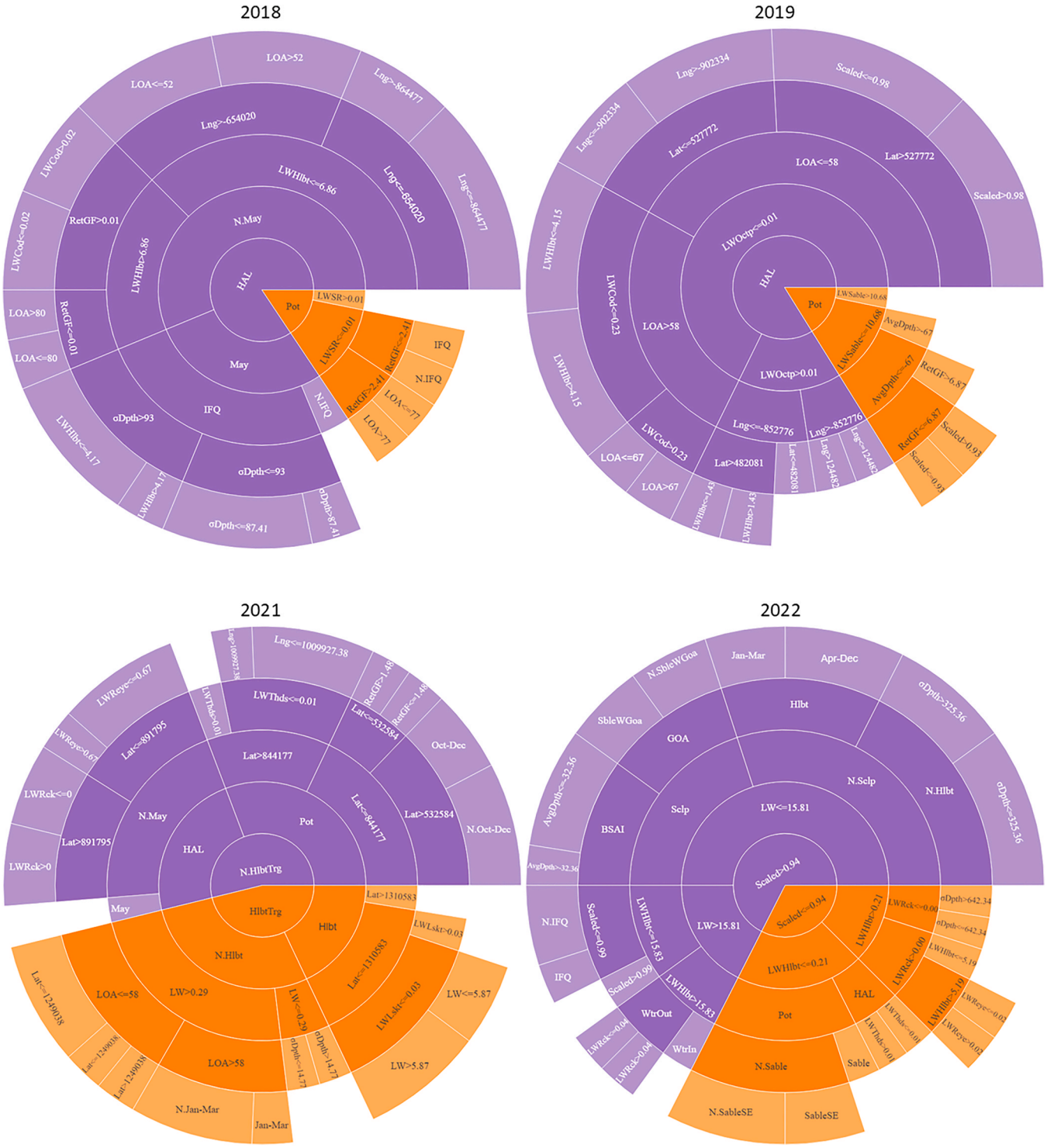


Fig. 5. The first 5 levels of the top-performing decision trees for the periods 2018–2019 and 2021–2022. Each tree corresponds to the year in which it was trained, with the 2020 tree excluded due to atypical sampling and fishery behavior resulting from the COVID-19 pandemic. A detailed description of the features is contained in Table 2; note that N. indicates "not" (e.g., N.HlbtTrg indicates "Not Pacific Halibut Target"), and the color indicates the 1st branch.

Table 2
Descriptions of features shown in Figs. 5 and 6. Note that in the text and Fig. 5, a "N." indicates "not" (e.g., N.HlbtTrg indicates "Not Pacific Halibut Target"), and colors indicate the 1st branch. Moreover, abbreviations that are species-specific indicate a tree path for that specific species. Features with an asterisk are those currently included in the CAS post-strata.

Feature	Description	Feature	Description
AvgDpth	Average Depth (meters)	LWRck	Landings (mt) of Rockfish (*Sebastes* spp.)
BSAI*	Bering Sea and Aleutian Islands	LWSable	Landings (mt) of Sablefish (*Anoplopoma fimbria*)
GOA*	Gulf of Alaska	LWSR	Landings (mt) Shortraker Rockfish (*Sebastes borealis*)
HAL*	Hook-and-Line Gear	LWThds	Landings (mt) Thornyhead Rockfish (*Sebastolobus alascanus*)
HlbtTrg*	Pacific halibut (*Hippoglossus stenolepis*) target	Pot*	Pot Gear
Hlbt	Pacific Halibut	RetGF	Retained Groundfish (mt)
IFQ	Individual Fishing Quota	Sable	Sablefish
Lat	Latitude (Meters)	SableAI	Sablefish AI Area
Lng	Longitude (Meters)	SableSE	Sablefish SE GOA Area
LOA	Vessel Length Overall (ft)	SbleWGOA	Sablefish Western GOA Area
LW	Landings (mt) of all Catch	Scaled	Min/Max Scaled Depth
LWCod	Landings (mt) of Pacific Cod (*Gadus macrocephalus*)	Sclp	Sculpin
LWHlbt	Landings (mt) of Pacific halibut	WtrIn	State of Alaska Waters
LWLskt	Landings (mt) of Longnose Skate (*Raja binoculata*)	WtrOut	Federal Waters
LWReye	Landings of Rougheye Rockfish (*Sebastes aleutianus*)	σDpth	Standard Deviation of Depth (m)

branch feature for trees in 2021 and 2022 was not a specified gear type, as was the case in 2018–2019, but instead was Pacific halibut target (HlbtTarg) or scaled depth fished. The features selected for trees in 2021 and 2022 represent a change in gear types fished for sablefish from HAL gear to primarily slinky pot gear. Slinky pots are soft-sided pots that can be fished individually or strung together on a line (longlined), generally have very low discard amounts, and can be fished on smaller vessels than the traditional, larger, hard-sided pots. The adoption of slinky pots also brought about a significant transformation in the use of gear types for the IFQ versus Pacific cod fisheries. The pot gear type specified in the data was primarily focused on Pacific cod in 2018 and shifted to target both Pacific cod and sablefish by the 2021–2022 period. However, Pacific cod fishing activity among vessels is evident in the 2021–2022 tree structure by the vessel length breaks occurring within the range of 51–58 feet. In addition, the seasonality of the Pacific cod fishery was captured in the tree structures, with seasonal fishing for HAL catcher vessels January to March and September to October.

A key finding from this study is that most features used in the decision tree are not currently included in CAS, and many attributes used in CAS post-strata, such as trip targets, federal reporting areas, and monthly breaks, were not selected in the best performing decision trees. The decision tree logic was generally simpler than the criteria used in CAS. Most importantly, of the covariates currently employed in CAS, only halibut target, gear type, and FMP area are included in the decision tree logic. The spatial breaks identified in the trees generally corresponded to larger areas than currently used in CAS, with GOA areas being more regional, rather than the federal reporting area post-strata criteria currently in used in CAS (Fig. 6 versus Fig. 1). Inseason managers indicated these regional areas correspond to fishery allocations and fleet activity. For example, all trees showed breaks between the BS, AI and the western GOA, with the far southern BS included in the AI/western GOA (Fig. 6). The GOA showed breaks delineating the western from Central and eastern GOA regions (Fig. 6). While the 2021 decision tree did not show an explicit branch in the Central GOA, it did show a branch for landed longnose skates conditioned on fishing activity that targeted Pacific halibut. Longnose skates are generally only landed in ports servicing the Central GOA fisheries (Kodiak, Seward, Homer) and thereby represent both a species and a spatial component (and specific fishing fleet) for the Central GOA. The 2021 tree also defined spatial areas by latitude rather than longitude or pre-defined spatial areas (Fig. 6). These breaks correspond to similar management areas as the other trees, with the Central, western, and eastern GOA areas being delineated, and the southern BS delineated from the GOA and AI. Hence, although the specific characteristics identified in selected trees varied across years, these different traits effectively identified the same fishing activities and showed that post-strata definitions built from one tree are applicable to multiple out-years.

Finally, depth was an important feature identified that is not currently used in CAS for estimation. This feature is associated with fishing activity in the nearshore, continental shelf, or slope of the continental shelf and is connected to both fishing activity and species distribution (e.g., Laman et al., 2018). The 2022 tree selected the inshore versus offshore water variable, which indicates State of Alaska waters versus waters only in the federal Exclusive Economic Zone (EEZ, i.e., 3 or less nautical miles from mean high tide line versus up to 200 nautical miles from mean high tide line for the EEZ). State of Alaska waters encompasses a greater proportion of ADF&G statistical areas with shallow waters versus deeper waters found in ADF&G areas corresponding to the EEZ. The standard deviation of depths ($\sigma Dpth$) across an ADF&G area was also commonly selected as an important feature. This feature was in the lower branches for all trees and is likely associated with localized discard situations, such as the AI, gullies, or steeper terrain within an area or fishery defined by the higher-level branches. Scaled depth was selected for 2021 and 2022 trees, with branch definition based on breaks ranging from 0.85 (2021) to 0.99 (multiple trees). The lower cutoff of 0.85 corresponds with a median depth of approximately 620 m in 2021 (interannual range from 500 to 700 m), and a depth break of 0.99 ranges from a median depth of 36 m (2020) to 58 m (2018).

4. Discussion

This paper has identified an approach to investigating post-strata for estimation of catch and bycatch from observer data collected under a complex hierarchical sample design. Estimates of catch are used to manage quota in near-real time and are used by a diverse group of stakeholders. To accommodate this, a post-stratification design capable of integration into an application across multiple years is essential, thus avoiding the need for significant annual adjustments or model-based estimations. This analysis effectively combined interpretable machine learning approaches with expert opinion on post-strata design to provide a methodology to meet data management requirements in context with an assessment of precision gains across years and species.

The methods presented provide a set of high performing trees that, in combination with expert opinion from inseason managers, can be used to identify plausible sets of decision tree models and associated post-stratification criteria. Within the hierarchy of the trees, the selected features were able to isolate fishery activity, and thus group fishing trips with similar catch characteristics. Despite variation among trees, common themes emerged that are useful in implementing a post-stratification design: (1) post-stratifications generally resulted in large increases in precision compared with non-post stratified data; (2) features that are not currently considered in CAS estimation were commonly selected in the best performing decision trees; and (3) tuning choices played a crucial role in tree selection, with the primary objective being the selection of simpler trees for interpretation by experts, along with the restriction of minimum leaf sizes and constraining tree depth to mitigate overfitting. Although the selected features were consistent across years, the tree structures varied annually, necessitating expertise

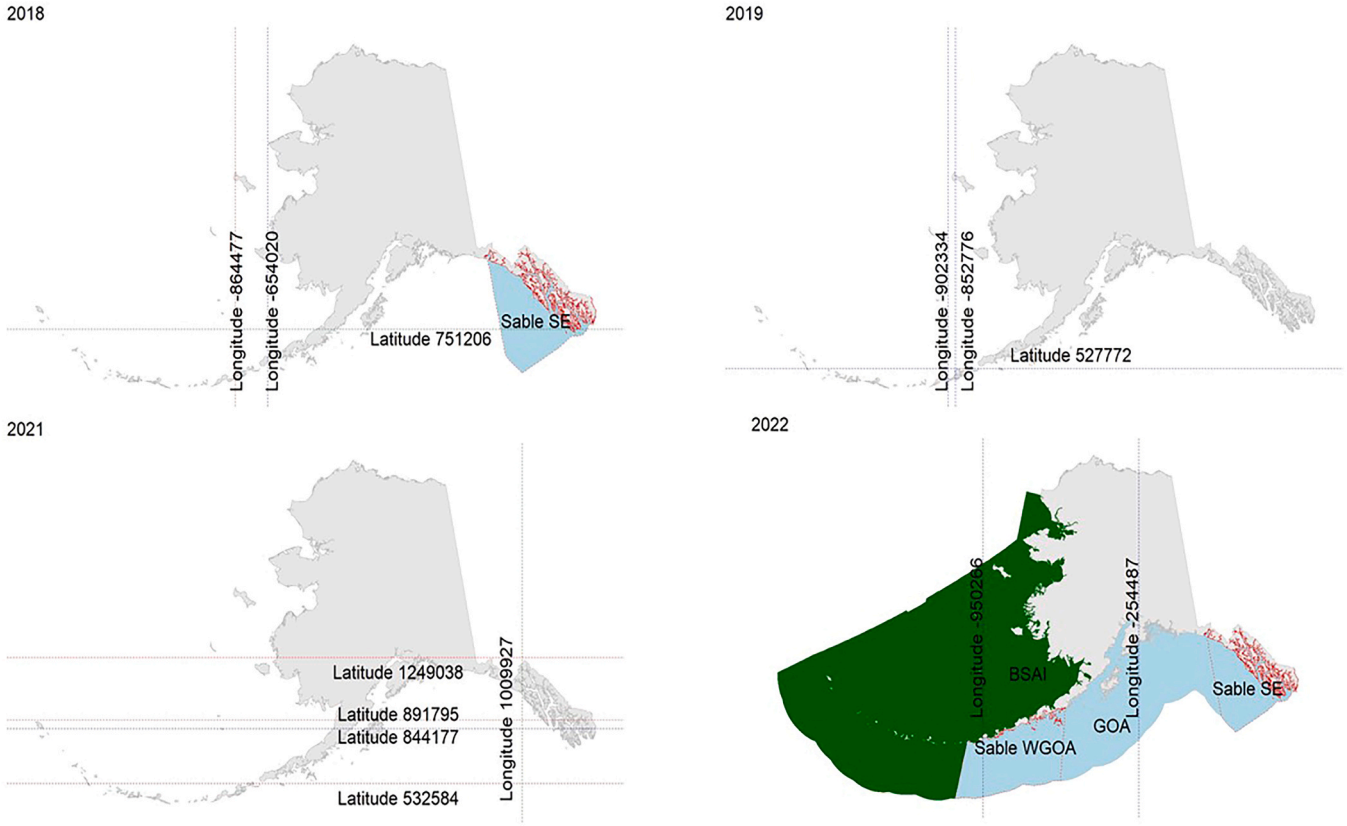


Fig. 6. Spatial breaks associated with branches of the preferred trees. Green color indicates the BSAI region and blue color indicates the GOA region. See Table 2 for acronym definitions. Results from 2020 are excluded due to atypical sampling and fishery behavior resulting from the COVID-19 pandemic.

to interpret the models and develop actionable post-stratification. This approach was effective for data exploration and finding important patterns that resulted in gains in precision but, from a practical application perspective, further adjustments would be required to implement it within CAS.

The increase of precision among the best trees varied by species and year. This variation is expected as the trees were trained on a single year of data and tested on multiple out-of-sample years rather than being optimized for a single species. However, the range of percentage differences among the best-performing trees remained consistent across years. Our use of out-of-sample testing for non-training years allowed for evaluation of the stability of methods across a short time series. Trees that are able to consistently perform well when compared with the population of all trees are likely to be more robust against typical year-to-year variation in the fishery. Notable differences in structures in trees among were observed among training years, and subject matter expertise would be required to settle on a single structure suitable for management. Maintaining stable methods is an important requirement for implementing logic that defines post-stratification procedures such as those used in the CAS (Cahalan et al., 2014).

Model training on individual year and testing across all out-sample years allowed inseason expertise to identify important long-term shifts in fishing behavior such a new gear type being used for regulatory changes (e.g., rockfish retention). If model training had been conducted using multiple years of data and testing on a single out-of-sample year, shifts in fishing behavior may not have been as apparent because those shifts would have been incorporated as variability during model training. For example, in the years 2018–2020, gear emerged as a primary differentiator at the first level of classification. However, in 2021 and 2022, gear was integrated into branches differently than earlier years, indicating a change in how it influenced the classification of vessel- and trip-specific data. Inseason managers were able to determine that this pattern was caused by the increased use of slinky pots for sablefish, resulting in the break between pot and HAL gear becoming less of a determinate for IFQ fishing versus other targets (e.g., Pacific cod). This important change can be incorporated into future CAS post-stratification work. Another key feature included in all trees was the amount of fish landed, which also revealed changes in fishing behavior related to regulatory changes and is not considered in current CAS post-stratification. For example, a prohibition on discarding rockfish species in federal waters off Alaska was enacted March 23, 2020 (85 FR 9687). This regulatory change maybe apparent in the 2021 and 2022 decision trees, with the landing weight of rockfish species being important features during those years and not earlier years. The landing weight of rockfish species as well as other species reflected in the trees is likely a surrogate for fishing effort, locations and depth fished, and gear types fished, all of which influence fishery discard characteristics.

Implementing this method required making decisions about hyperparameters to reduce the number of structural-related considerations needed for management purposes. A key decision is tree depth and how post-strata size is considered for each leaf or branch. Overfitting the data is a problem with decision trees and, in our case study, overfitting could lead to post-strata with few trips and poor ability to generalize to multiple years, with the potential for post-strata to be created with few vessels. Counter to this, we found the cross-validation associated with the reduction in variance to provide consistent results on the validation data (Fig. 4); however, further refinement of the post-strata is likely required for use in management. Sample size variability is also an important issue that was not considered in this study and would influence the precision of estimates and effectiveness of any post-strata design due to the expansion of the sampled to unsampled portion of the population. Thus, the method proposed here is a first step to define potential post-strata. Simulation testing could be an additional (second) step to determine the feasibility of certain post-strata alongside subject

matter expert review of decision trees, and to determine if strata were created that are too focused (possibly a consequence of overfitting the medium and high categories). Leveraging subject matter expertise to generalize and prune the best performing decision trees into potential post-strata designs for CAS would enable a comparison and assessment of simplified tree designs. For example, the spatial components identified generally aligned with larger regional areas that could be identified by subject matter experts. This alignment could be used to generalize branches or prune the decision tree. This process would also influence sample size allocations among post-strata. In this way, details associated with the machine learning results (i.e., decision trees) can be simplified, tested, and tailored to a practical application.

The random forest algorithm is a "greedy" algorithm in that it makes locally optimal choices at each iteration with the hope of finding a global optimal solution, although it does not guarantee that a global optimum comprised of optimal trees are included in the forest (Xin et al., 2022). Future work could improve the choice set of trees by exploring Rashomon sets of trees (Semenova et al., 2022, Fisher et al., 2019, Rudin, 2019) that comprise a set of sparse nearly-optimal trees that could be evaluated for precision using methods in this paper. Identification of a Rashomon set(s) would allow the precision associated with a statistic to be calculated on individual trees from a nearly optimal ensemble of trees. Decisions about forest size, tree depth, and penalties based on leaf or branch sizes could still be made to restrict fitting, but fewer of these decisions may be required if the overall ensemble performs better than a random forest due to the nearly-optimal set of sparse trees. However, expertise on the individual tree models will remain a critical component method for pruning trees and creating a post-stratification design suitable for management.

In conclusion, our study introduces a method for designing post-stratification definitions based on covariates identified through a combination of random forest techniques and the expertise of subject matter experts, specifically fishery managers in our case. This method is especially applicable when high dimensionality makes identifying important post-strata design features difficult. This method also allows exploration of the post-stratification design by evaluating decision trees in context with the fishery and post-strata sample size and provides a method for incorporating subject-matter expertise to identify important relationships among covariates. Through consideration of potential designs, this method is a tool for incorporating expert opinions to guide machine learning and random forest approaches rather than relying only on inaccessible modeling details that are common with machine learning approaches.

CRediT authorship contribution statement

Jason Gasper: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jennifer Cahalan:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jennifer Cahalan reports financial support was provided by Pacific States Marine Fisheries Commission. Jason Gasper reports a relationship with NOAA Fisheries Alaska Regional Office that includes: employment. Jennifer Cahalan reports a relationship with Pacific States Marine Fisheries Commission that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Cathy Tide (NMFS Alaska Regional Office) and, in memory of Steve Lewis (NMFS Alaska Regional Office) for providing bathymetric information used in the spatial analysis. We also thank Josh Keaton (NMFS Alaska Regional Office) and Mary Furuness (NMFS Alaska Regional Office) for their expertise on inseason management and interpretation of the decision trees. We thank Dr. Cindy Tribuzio, Jennifer Mondragon, and anonymous reviewers who greatly improved the manuscript. This work was funded in part by NOAA Award NO. NA22NMF4370332 to the Pacific States Marine Fisheries Commission. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the authors and do not necessarily reflect the views of NOAA, the Department of Commerce, or Pacific State Marine Fisheries Commission.

Data Availability

The data that has been used is confidential.

References

Alaska Department of Fish and Game (ADF&G). 2024. Information by Fishery: Shellfish and Groundfish Statistical Areas. Available at ⟨https://www.adfg.alaska.gov/index.cfm?adfg=fishingCommercialByFishery.statmaps⟩.

Cahalan, J., Gasper, J., Mondragon, J.(2014) Catch sampling and estimation in the federal groundfish fisheries off Alaska, 2005 edition. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-AFSC-286, 46 pages. Document available: ⟨http://www.afsc.noaa.gov/Publications/AFSC-TM/NOAA-TM-AFSC-286.pdf⟩.

Casella, G., and Berger, R.L. (2002). Statistical inference (2nd ed.). Thomson Learning. ISBN-13: ↱ 978-0534243128.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority oversampling technique. J. Artif. Intell. Res. 16, 321–357.

Cochran, W.G., 1977. Sampling techniques (third ed.). John Wiley & Sons, Inc. ISBN-13: ↱ 978-0471162407.

Fisher, A., Rudin, C., Dominici, F., 2019. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. J. Mach. Learn. Res. 20 (177), 1–81.

Gasper, J.R., Kruse, G.H., 2013. Modeling of the spatial distribution of Pacific spiny dogfish (*Squalus suckleyi*) in the Gulf of Alaska using generalized additive and generalized linear models. Canadian Journal of Fisheries and Aquatic Sciences 70 (9), 1372–1385. https://doi.org/10.1139/cjfas-2012-0535.

Hansen, M.H., Hurwitz, W.N., Madow, W.G., 1953. Sample survey methods and theory, I & II. John Wiley & Sons. ISBN: 978-0-471-00628-2.

Ho, T.K., 1995. Random decision forests, in: Proceedings of 3rd International Conference on Document Analysis and Recognition, 1(1), pp. 278–282.

Holt, D., Smith, T.M.F., 1979. Post stratification. J. R. Stat. Soc. Ser. A (Gen.) 142 (1), 33–46. https://doi.org/10.2307/2344652.

Laman, E.A., Rooper, C.N., Turner, K., Rooney, S., Cooper, D.W., Zimmermann, M., 2018. Using species distribution models to describe essential fish habitat in Alaska. Canadian Journal of Fisheries and Aquatic Sciences 75 (8), 1230–1255. https://doi.org/10.1139/cjfas-2017-0181.

Lemaitre, G., Nogueira, F., Aridas, C.K., 2017. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. J. Mach. Learn. Res. 18 (17), 1–5.

Little, R.J.A., 1991. Post-stratification: a modeler's perspective. J. Am. Stat. Assoc. 88 (423), 1001–1012. https://doi.org/10.2307/2290792.

Minami, M., Lennert-Cody, C.E., 2024. Regression tree and clustering for distributions, and homogeneous structure of population characteristics. J. Agric. Biol. Environ. Stat. https://doi.org/10.1007/s13253-024-00631-z.

Mini, K.G., Kumaran, M., Jayasankar, J., 2009. On use of post-stratification for estimating the marine fish landings. Indian J. Mar. Sci. 38 (4), 464–469. Available at http://eprints.cmfri.org.in/6/.

National Marine Fisheries Service Office of Science and Technology (NMFSOST). 2023. Marine Recreational Information Program Survey Design and Statistical Methods for Estimation of Recreational Fisheries Catch and Effort. Silver Spring, MD.

NMFS. (2023). 2024 Annual Deployment Plan for Observers and Electronic Monitoring in the Groundfish and Pacific halibut Fisheries off Alaska. National Oceanic and Atmospheric Administration, 709 West 9th Street, Juneau, Alaska 99802. Available at: ⟨https://www.fisheries.noaa.gov/resource/document/2024-annual-deployment-plan-observers-and-electronic-monitoring-groundfish-and⟩.

NPFMC (2020). Fishery Management Plan for Groundfish of the Bering Sea and Aleutian Islands. North Pacific Fisheries Management Council. 1007 West Third, Suite 400 Anchorage, Alaska 99501. Available at: ⟨https://www.npfmc.org/wp-content/PDFdocuments/fmp/BSAI/BSAIfmp.pdf⟩.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Duchesnay, E., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.

Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why Should I Trust You?": Explaining the predictions of any classifier, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations. https://doi.org/10.18653/v1/N16-3020.

Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. 1, 206–215. https://doi.org/10.1038/s42256-019-0048-x.

Salman, Ahmad, Siddiqui, Shoaib Ahmad, Shafait, Faisal, Mian, Ajmal, Shortis, Mark R., Khurshid, Khawar, Ulges, Adrian, Schwanecke, Ulrich, 2020. Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system. ICES J. Mar. Sci. 77 (4), 1295–1307. https://doi.org/10.1093/icesjms/fsz025.

Semenova, L., Rudin, C., Parr, R., 2022. On the existence of simpler machine learning models, in: ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT), 2022. https://doi.org/10.1145/3531146.3533232.

Thompson, S.K., 2012. Sampling (third ed.). John Wiley & Sons Inc. ISBN: 9781118162934.

Watson, J.T., Ames, R., Holycross, B., Suter, J., Somers, K., Kohler, C., Corrigan, B., 2023. Fishery catch records support machine learning-based prediction of illegal fishing off US West Coast. PeerJ 11, e16215. https://doi.org/10.7717/peerj.16215.

Westfall, J.A., Lister, A.J., Coulston, J.W., McRoberts, R.E., 2021. Realized and potential efficiency for post-stratified estimation in a national forest inventory. Can. J. For. Res. 51, 1450–1457. https://doi.org/10.1139/cjfr-2020-0379.

Westfall, J.A., Patterson, P.L., Coulston, J.W., 2011. Post-stratified estimation: within-strata and total sample size recommendations. Can. J. For. Res. 41, 1130–1139. https://doi.org/10.1139/x11-031.

Williams, W.H., 1962. The variance of an estimator with post-stratified weighting. J. Am. Stat. Assoc. 57, 622–627. https://doi.org/10.1080/01621459.1962.10500550.

Xin, R., Zhong, C., Chen, Z., Takagi, T., Seltzer, M., Rudin, C., 2022. Exploring the whole rashomon set of sparse decision trees. Adv. Neural Inf. Process. Syst. 35, 14071–14084. PMID: 37786624. Available at: https://proceedings.neurips.cc/paper_files/paper/2022/file/5afaa8b4dd18eb1eed055d2d821b58ae-Paper-Conference.pdf.

Yassir, Anas, Jai Andaloussi, Said, Ouchetto, Ouail, Mamza, Kamal, Serghini, Mansour, 2023. Acoustic fish species identification using deep learning and machine learning algorithms: a systematic review. Fish. Res. 266. https://doi.org/10.1016/j.fishres.2023.106790.

Zhang, T., Guo, H., Song, L., Yuan, H., Sui, H., Li, B., 2024. Evaluating the importance of vertical environmental variables for albacore fishing grounds in the tropical Atlantic Ocean using machine learning and Shapley additive explanations (SHAP) approach. Fish. Oceanogr. https://doi.org/10.1111/fog.12701.

Zuur, Alain, Ieno, Elena N., Walker, Neil, Saveliev, Anatoly A., Smith, Graham M., 2009. Mixed Effects Models and Extensions in Ecology with R. 2009th ed. Statistics for Biology and Health. Springer, New York, NY.