ICES Journal of Marine Science, 2022, **0**, 1–12
DOI: 10.1093/icesjms/fsac219
Original Article

ICES CIEM
International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

Predicting bycatch of Chinook salmon in the Pacific hake fishery using spatiotemporal models

Philip L. Shirk[1,*], Kate Richerson[2], Michael Banks[1] and Vanessa Tuttle[3]

[1]The Cooperative Institute for Marine Resources Studies, Oregon State University, Hatfield Marine Science Center, 2030 SE Marine Science Drive Newport, OR 97365, USA
[2]Northwest Fisheries Science Center, Newport Research Station, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2032 SE OSU Drive Newport, OR 97365, USA
[3]Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd E, Seattle, WA 98112, USA

[*] *Corresponding author: tel: +1-541-867-0100;* e-mail: philip.shirk@gmail.com.

Fisheries bycatch is a global problem, and the ability to avoid incidental catch of non-target species is important to fishermen, managers, and conservationists. In areas with sufficient data, spatiotemporal models have been used to identify times and locations with high bycatch risk, potentially enabling fishing operations to shift their effort in response to the dynamic ocean landscape. Here, we use 18 years of observer data from the Pacific hake (*Merluccius productus*) fishery, the largest by tonnage on the US West Coast, to evaluate our ability to predict bycatch of the commercially, culturally, and ecologically important Chinook salmon (*Oncorhynchus tshawytscha*). Using multiple approaches (regression models, tree-based methods, and model averages), we tested our ability to predict bycatch at weekly and yearly timescales and found that spatiotemporal models can have good predictive ability. Gradient boosting trees (GBTs) and model averages typically had higher performance, while generalized linear models and generalized additive models (without interaction terms) did less well. Using a GBT model to remove 1% of hauls with the highest predicted bycatch reduced the bycatch-to-hake ratio by 20%. Our results indicate that spatiotemporal models may be a useful forecasting tool that can help fishing operations avoid bycatch while minimizing losses from target catches.

Keywords: bycatch, Chinook salmon, fisheries observer data, hake, predictive modelling, whiting.

Introduction

Bycatch, or the incidental capture of non-target individuals, presents an ongoing challenge for fishery management and conservation. Globally, bycatch represents a significant proportion of total catch (~10% in recent decades; Zeller *et al.*, 2018) and threatens many rare and ecologically important species (Lewison *et al.*, 2014). In response to the economic and ecological costs of bycatch, many countries have implemented policies to reduce incidental take and discards, including landing obligations, closed areas or seasons, bycatch caps, and gear modifications. In the United States, federal law requires that bycatch and bycatch mortality be minimized to the extent practicable, and fisheries management plans must establish a standardized reporting methodology to assess bycatch. Accounting for bycatch is a key part of ecosystem-based fishery management, which requires considering both the direct and indirect impacts of fishery operations on target and non-target species (Gilman *et al.*, 2014).

A key factor limiting our understanding of bycatch and its effects is a lack of data (Komoroske and Lewison, 2015). However, the Pacific Coast groundfish fishery on the US West Coast offers a unique opportunity to investigate bycatch because all sectors are subject to observer or electronic monitoring. One of the species regulated under the Pacific Coast groundfish fishery, Pacific hake (*Merluccius productus*, also called Pacific whiting), is one of the largest fisheries by tonnage on the US West Coast, with landings averaging 258000 tonnes over the past 10 years and a total economic impact of $279 million in 2018. The stock is managed through the bilateral Pacific Whiting Agreement between the United States and Canada, which allocates an annual quota across the two countries. The US hake fishery also has room to grow, as it has averaged only 75% attainment of the target catch over the last 10 years (Johnson *et al.*, 2021). Though the ratio of bycatch to hake in this fishery is overall relatively low (typically <2%), the fishery regularly encounters species of concern, and the high-volume nature of the fishery means that the total incidental catch may be significant. Since 2002, the hake fishery has accounted for approximately two-thirds of the total observed bycatch of salmon off the US West Coast (Richerson *et al.*, 2020). Regulations prevent salmon caught as bycatch from being sold commercially or otherwise retained for human consumption, though they may be donated to food banks. All vessels targeting hake in the US carry scientific observers or electronic monitoring equipment, so virtually all catches are sampled and any bycatch is recorded.

Bycatch of Chinook salmon (*Oncorhynchus tshawytscha*; hereafter Chinook) presents a particular management and conservation challenge because Chinook are culturally, economically, and ecologically important. In addition to being the subject of valuable commercial and recreational fisheries, nine Chinook populations are listed as threatened or endangered under the United States Endangered Species Act, and many runs are expected to face additional challenges as marine and freshwater conditions continue to change (Muñoz *et al.*, 2015; Shelton *et al.*, 2020). The west coast directed ocean

Received: July 8, 2022. **Revised:** November 11, 2022. **Accepted:** November 17, 2022

Chinook fisheries have faced sharply restricted seasons in recent decades in order to protect stocks with low predicted returns, and multiple federal fisheries disasters have been declared due to low Chinook availability. Consequently, incidental catch of Chinook in other fisheries is subject to close scrutiny, and has led to a number of management measures and voluntary industry-led bycatch avoidance practices.

Efforts to minimize bycatch in the hake fishery include fleet communication measures, bycatch quotas, and a catch-share programme (Holland and Martin, 2019). The three sectors targeting hake (motherships, catcher-processors, and shoreside) each have a set of strategies and agreements for avoiding bycatch. The mothership sector formed a cooperative in 2011 that has a number of internal rules, including bycatch hotspot closures, restrictions on fishing at night, the use of test tows, and the relocation of catcher vessels delivering to individual motherships if bycatch rates exceed a threshold (Holland and Martin, 2019). The catcher-processor sector also formed a cooperative in 1997 and has internally agreed to a number of bycatch avoidance techniques, though the extent to which they use closed areas is less clear (Holland and Martin, 2019). In the shoreside sector, most vessels act as cooperatives and use a risk pool to manage the risk of exceeding low quotas of bycatch species. This sector also uses a number of bycatch avoidance methods, including information sharing, night fishing restrictions, closed areas, and hotspot closures (Holland and Martin, 2019). In all three sectors, the implementation of bycatch avoidance strategies is facilitated by Sea State Inc., which analyses data from onboard observers to generate daily information on bycatch hotspots, cautionary areas, and closed areas that is then disseminated to the fleet (Little *et al.*, 2015; Holland and Martin, 2019). The spatial distribution of observed Chinook bycatch and hake catch rates is shown in Figure 1.

Under the National Marine Fisheries Service's (NMFS) biological opinion on take of salmon in the Pacific Coast groundfish fishery, a total take of 20000 Chinook is allowed for the entire groundfish fishery (for comparison, the non-tribal targeted ocean Chinook fishery landed ∼13000–900000 fish annually between 2000 and 2020, with a generally declining trend; PFMC, 2021). Of this incidental take amount, 11000 Chinook are apportioned to the hake fishery and 5500 to the non-hake fishery, with a "reserve" of 3500 fish that can be accessed by either sector (Matson and Erickson, 2018). If the threshold of 11000 fish is exceeded or projected to be exceeded, conservation measures like closed areas may be implemented, and a complete closure of the fishery may occur under certain circumstances (50 CFR § 660.60, 2022). Chinook bycatch in the hake fishery has varied over time and exceeded 11000 individuals in 2005 and 2014 (Richerson *et al.*, 2020), with the latter resulting in a closure of the fishery shoreward of the 100 fathom (183 m) depth contour. The interannual variation in Chinook bycatch does not appear to be a direct result of fluctuating ocean abundances, as bycatch does not appear to be clearly linked to indices of Chinook abundance (Matson and Erickson, 2018). Reducing bycatch could help the hake fishery avoid closures and potentially facilitate an expanded fishing effort to utilize more of the full target catch of hake.

Efforts to model bycatch in fisheries have advanced considerably in the past decade as spatiotemporal modelling options have become more widely available (Ward *et al.*, 2015; Otto *et al.*, 2016; Eguchi *et al.*, 2017; Stock *et al.*, 2019, 2020).

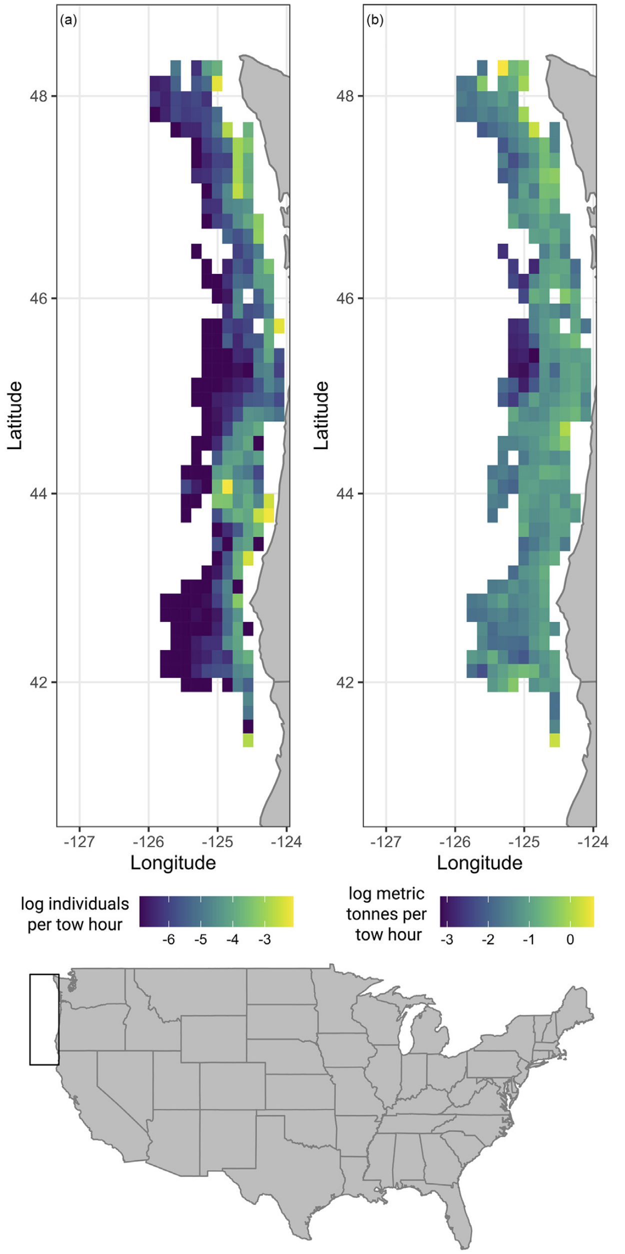

Figure 1. The spatial distribution of (a) observed Chinook salmon bycatch per unit effort and (b) observed Pacific hake catches per unit effort, 2002–2020. To maintain confidentiality, any spatial cells containing data from fewer than three vessels are not shown.

The explicit incorporation of spatial correlation in modelling approaches is valuable because bycatch is often spatially correlated (Lewison *et al.*, 2009). Models that account for spatial correlation are typically more accurate than those that ignore it (Dormann, 2007; Stock *et al.*, 2019). A common method of partially accounting for spatial correlation in marine fisheries is to bin observations into grid cells or strata and work with average values. However, binning discards information on the fine-scale variation in the response and predictor variables, and models based on binned data typically perform worse than geostatistical models (Shelton *et al.*, 2014; Thorson *et al.*, 2015). Spatial correlations in bycatch are also likely to change over time, particularly as climate change alters environmental conditions and species' phenologies. In addition to improving model fit, the spatiotemporal patterns from models may be of interest themselves. They can be used to identify hotspots of bycatch, shifts in the distribution of bycatch, and unmodelled covariates that may help to explain variation in bycatch.

Here, we use spatiotemporal models to investigate trends in the bycatch of Chinook in the hake fishery and to test our ability to forecast future bycatch. In particular, we want to know how predictable bycatch is at varying temporal scales and to identify correlates of bycatch that are useful in predicting bycatch risk. A better understanding of Chinook bycatch across space and time may help inform bycatch avoidance and management measures, as well as increase the ability of the fishery to maximize the catch of hake.

Methods

Data

Our analysis focuses on the US West Coast limited-entry hake fishery, which is responsible for the vast majority of hake catches (Somers *et al.*, 2020; Johnson *et al.*, 2021). Within the limited entry fishery are three sectors: at-sea catcher-processors, at-sea motherships (with their catcher vessels), and shore-based catcher vessels (Warlick *et al.*, 2018). The two at-sea sectors are monitored through the At-Sea Hake Observer Program (A-SHOP); the shore-side sector is monitored through the West Coast Groundfish Observer Program (WCGOP) and the West Coast Region, National Marine Fisheries Service. In recent years, many shoreside vessels have begun using electronic monitoring in lieu of at-sea observers; this monitoring programme is administered by the Pacific States Marine Fisheries Commission. To avoid adding further complexity to an already complex analysis, we do not include data from vessels carrying electronic monitoring in this analysis. For a summary of observer coverage and electronic monitoring coverage in both the at-sea and shoreside sectors, see Somers *et al.* (2020).

We used haul-level A-SHOP ($n = 26674$ catcher-processor and $n = 19050$ mothership catcher vessel hauls) and WCGOP ($n = 7840$ shoreside hauls) observer data from 2002 to 2019 to evaluate our ability to predict Chinook bycatch in the hake fishery using spatiotemporal models. Though both the at-sea and shoreside sectors target hake and carry observers or electronic monitoring, there are differences in fishing and data collection methods across these sectors. All vessels in the at-sea hake fishery carry two observers, and virtually all hauls are sampled for species composition. Typically, 50% of each haul is sampled by an on-board observer. In addition to information on retained and discarded catch, data include information on the time, duration, location, and depth of each haul. For Chinook salmon, at-sea hake observers also collect biological data, including length, weight, sex, adipose fin status, a tissue sample, and a coded-wire tag scan. Our analysis used all recorded Chinook, regardless of adipose fin status (an indicator of wild versus hatchery stock). Details on observer data collection methods in the at-sea hake sectors can be found in NWFSC (2021a).

In contrast to the at-sea sector, where hauls are sampled individually by on-board observers, the shoreside hake sector is a maximized retention fishery. This means that, though all vessels carry an observer or electronic monitoring equipment, catch is not typically sorted at sea. Instead, bycatch is quantified at the trip level after it is landed, and one trip typically consists of 2–3 hauls. Although observers record haul information (e.g. overall total catch; time; depth; location; duration); haul-level bycatch is typically estimated proportionally to the observed retained catch of each haul after the catch is sorted shoreside. In other words, the species composition of all hauls within a trip is assumed to be the same. Biological data is also collected shoreside by catch monitors. For details on observer data collection methods in the shoreside hake fishery, see NWFSC (2021b).

Models

We followed the general approach of Stock *et al.* (2020), which includes the application of hurdle models (Cragg, 1971) to account for the large number of zeros in the data and incorporates a spatial component in each model. Hurdle models consist of two component models: the first predicting the probability that there was any bycatch, and the second predicting the amount of bycatch for hauls that contained bycatch.

We fitted ten model structures to the data to predict whether hauls contained bycatch (i.e. the first component of the hurdle model) (Table 1). Additionally, we tested two model averages for comparison because model averages often outperform individual component models in predictive capability (Dormann *et al.*, 2018). The ten models can be split into two categories: those based on linear models and those based on decision trees. The linear-based models include generalized linear models (GLM) and generalized additive models (GAM), both of which assume a binomial distribution to predict the presence of bycatch. Individual models differed in how they incorporated covariates (including latitude and longitude) (Table 1). The 3D space–time spline smoother in GAM 3 (Wood, 2017) was complicated by the way in which we performed cross-validation. For all three methods of cross-validation (see below for details), GAM 3 allowed the spatial smooth to vary through time by creating a separate smoother in each year. The complication for GAM 3 models is that independent 2D spatial smooths for each year cannot be used to make predictions for years that were not represented in the training dataset, which is precisely what the yearly time-series cross-validation was designed to do. Therefore, GAM 3 models in the yearly time series cross-validation included a 3D space–time spline smoother where time was a continuous variable, which allows extrapolation to years that were not in the training data. To maintain tractability, we reduced the dimension of the basis of the two spatial terms in the spline ($k = 40$), forcing the spatial smooth to vary more slowly through space. The pseudocode for the model formulas can be found in Table 1.

The tree-based models included random forests (RF) and gradient boosted trees (GBT), both of which aggregate predictions across numerous individual decision trees. Random forests average across component trees. GBT are fit sequentially, with each additional tree fitting to the residuals of the previous tree (Elith *et al.*, 2008). For the first component of the hurdle model, tree-based models used binomial classification trees. In highly imbalanced datasets (few observations of a particular class), decision trees may do a poor job of predicting the rarer class (Chawla *et al.*, 2004). We included two RFs to account for imbalance in our dataset, which had four times as many hauls without Chinook bycatch as with bycatch. While RF 1 fit all hauls in the dataset, RF 2 subsampled an

Table 1. Descriptions of the models comprising the hurdle model.

Model	Model description	Family or type	Link	Formula pseudocode*
GLM 1	GLM with linear covariates	Binomial /Gamma	Logit Log	bycatch ∼ sector + year + ToD + DoY + duration + depth_fishing + depth_bottom + slope_bottom + sst_anomaly + cuti + lat + lon
GLM 2	GLM with second-order polynomials for continuous covariates	Binomial/ Gamma	Logit Log	bycatch ∼ sector + year + ln(ToD) + ln(ToD)^2 + ln(DoY) +ln(DoY)^2 + ln(duration) + ln(duration)^2 + ln(depth_fishing) + ln(depth_fishing)^2 + ln(depth_bottom) + ln(depth_bottom)^2 + ln(slope_bottom) + ln(slope_bottom)^2 + ln(sst_anomaly) + ln(sst_anomaly)^2 + ln(cuti) + ln(cuti)^2 + ln(lat) + ln(lat)^2 + ln(lon) + ln(lon)^2
GAM 1	"Low" complexity GAM with linear covariates and 2D smooth over latitude and longitude	Binomial /Gamma	Logit Log	bycatch ∼ s(lon, lat, k = 100) + sector + year + ToD + DoY + duration + depth_fishing + depth_bottom + slope_bottom + sst_anomaly + cuti
GAM 2	"Medium" complexity GAM with low-order smooths on covariates and 2D smooth over latitude and longitude	Binomial /Gamma	Logit Log	bycatch ∼ s(lon, lat, k = 100) + sector + year + s(ToD, bs = "cr," k = 4, fx = TRUE) + s(DoY, bs = "cr," k = 4, fx = TRUE) + s(duration, bs = "cr," k = 4, fx = TRUE) + s(depth_fishing, bs = "cr," k = 4, fx = TRUE) + s(depth_bottom, bs = "cr," k = 4, fx = TRUE) + s(slope_bottom, bs = "cr," k = 4, fx = TRUE) + s(sst_anomaly, bs = "cr," k = 4, fx = TRUE) + s(cuti, bs = "cr," k = 4, fx = TRUE)
GAM 3	"High" complexity GAM with smooths over covariates and 3D smooth over space and time	Binomial/ Gamma	Logit Log	bycatch ∼ s(lon, lat, k = 25, by = year) + sector + year + s(ToD, bs = "cr," k = 4) + s(DoY, bs = "cr," k = 4) + s(duration, bs = "cr," k = 4) + s(depth_fishing, bs = "cr," k = 4) + s(depth_bottom, bs = "cr," k = 4) + s(slope_bottom, bs = "cr," k = 4) + s(sst_anomaly, bs = "cr," k = 4) + s(cuti, bs = "cr," k = 4)
RF 1	Random forest using all data	Classification /Regression	-	bycatch ∼ sector + year + ToD + DoY + duration + depth_fishing + depth_bottom + slope_bottom + sst_anomaly + cuti + lat + lon, num.trees = 1 500, mtry = 3, replace = FALSE
RF 2	Random forest with downsampling of over-represented class (classification only)	Classification	-	bycatch ∼ sector + year + ToD + DoY + duration + depth_fishing + depth_bottom + slope_bottom + sst_anomaly + cuti + lat + lon, num.trees = 1 500, mtry = 3, replace = FALSE, sample.fraction = rep(nmin.prop/3, 2)
RF 3	Random forest with oversampling of the under-represented class and downsampling of the over-represented class (classification only)	Classification	-	bycatch ∼ sector + year + ToD + DoY + duration + depth_fishing + depth_bottom + slope_bottom + sst_anomaly + cuti + lat + lon, num.trees = 1 500, mtry = 3
GBT 1	Gradient boosting trees with default tuning parameters	Classification /Regression	-	bycatch ∼ sector + year + ToD + DoY + duration + depth_fishing + depth_bottom + slope_bottom + sst_anomaly + cuti + lat + lon, nrounds = 1 000, max.depth = 3, booster = "gbtree," params = list(eta = 0.1, subsample = 0.7)
GBT 2	Gradient boosting trees with DART	Classification /Regression	-	bycatch ∼ sector + year + ToD + DoY + duration + depth_fishing + depth_bottom + slope_bottom + sst_anomaly + cuti + lat + lon, nrounds = 1 000, max.depth = 3, booster = "dart," params = list(eta = 0.15, subsample = 0.7, sample_type = "uniform," normalize_type = "tree," skip_drop = 0.25, rate_drop = 0.5)

*Parameter descriptions:
sector = Fishing sector (mothership, catcher-processor, shoreside)
year = Year (treated as a factor except in the yearly cross-validation)
ToD = Time of day
DoY = Day of year
duration = Haul duration
depth_fishing = Fishing depth
depth_bottom = Bottom depth
slope_bottom = Bottom slope
sst_anomaly = Sea surface temperature anomaly (temporally lagged where appropriate)
cuti = Coastal upwelling transport index (temporally lagged where appropriate)
lat = Latitude
lon = Longitude
nmin.prop = Proportion of hauls belonging to the smaller category (with/without bycatch) in the training data

equal number of observations from each class (hauls with and without bycatch). This reduces the sample size for fitting each tree. To avoid reducing sample size as much, RF 3 employs the synthetic minority over-sampling technique (SMOTE; Chawla *et al.*, 2002) to artificially increase the number of hauls containing bycatch and then also subsamples an equal number of observations from each class. If imbalance in our dataset is problematic, we expect RF 2 and RF 3 models to outperform RF 1 and potentially the GBT models. If imbalance is not an issue, then limiting the sample sizes and creating artificial data may weaken the predictive accuracy of RF 2 and RF 3. Note that because of the sampling methods used, RF 2 and RF 3 are only applicable to classification.

Model GBT 1 employed a standard GBT, which can have excellent predictive power but can also easily overfit the training data. GBT models are strongly influenced by trees near the beginning of the sequence, while trees near the end of the sequence impact fewer data points and may be fitting the model to noise in the data. One method to reduce overfitting in GBT is to randomly drop individual trees via DART (Dropouts meet Multiple Additive Regression Trees) (Rashmi and Gilad-Bachrach, 2015). In DART, if no trees are dropped, the model is identical to a standard GBT. If all trees are dropped, DART is identical to a random forest. Model GBT 2 employed DART with a dropout rate of 50%.

The second component of the hurdle model, predicting abundance of bycatch in hauls that contained bycatch, included the same models as the first (binomial) component of the hurdle model, with the exception of RF 2 and RF 3, which are only applicable to classification, not regression. Furthermore, GLMs, GAMs, and the GBT models in the second (abundance) component of the hurdle model implemented gamma regression with a log link; tree-based models used regression trees rather than classification trees.

For both components of the hurdle model, we calculated the first model average (AVG 1) as the unweighted mean of the predictions (on the response scale) of all models (ten for the first/binomial component; eight for the second/abundance component). We calculated the second model average in the same way, except that we excluded model GAM 3 because the definition of GAM 3 changes for models fit with yearly time-series cross-validation and because some GAM 3 models had very poor fit.

Response and predictor variables

We chose to model the number of Chinook as our response variable. Ultimately, the number of Chinook and the amount of hake are both relevant, but using the ratio between the two is not necessarily the best solution for two reasons. First, covariates may influence Chinook and Hake populations differently. Second, if hake occur at low densities in an area without Chinook, models of the Chinook-to-hake ratio would classify the area as an ideal fishing location in spite of the high effort that would be required to catch hake.

As predictor variables for Chinook bycatch, we included haul and environmental characteristics. Haul characteristics included fishing sector (at-sea catcher-processor; at-sea mothership; shoreside); year; haul duration; day of year; time of day; fishing depth; and location (latitude and longitude, projected to the USA Contiguous Equidistant Conic projection). The characteristics of the hauls were included in the A-SHOP and WCGOP data. We chose to include haul duration as a covariate rather than an offset term because we did not want to force a 1:1 relationship between duration and bycatch (Stock *et al.*, 2019). Environmental variables included ocean depth, sea surface temperature (SST) anomaly, upwelling, and bottom slope, all of which were interpolated to the location and time of each haul. Justification for, and sources of, the environmental variables are as follows:

Ocean depth is correlated with many variables that likely influence the presence of hake and Chinook, including temperature, light, predator and prey abundances, and the direction and strength of currents. When available, we used the bottom depth as measured by fishermen and recorded in the A-SHOP data. When the bottom depth was not recorded (as was the case for all 7840 hauls in the WCGOP dataset and 74 of 45724 hauls in the A-SHOP dataset), we used bathymetry data from the National Geophysical Data Center's US Coastal Relief Model (NOAA National Geophysical Data Center, 2003a, b).

The slope of the seafloor is also an important component of habitat for hake and their prey (Mackas *et al.*, 1997). We calculated the bottom slope from the bathymetry dataset using the "terrain" function in the raster package (Hijmans, 2020) for R.

Temperature is related to a host of biological processes (Angilletta, 2009) and to species distributions (Stuart-Smith *et al.*, 2017). We use the SST anomaly (difference from the 1971 to 2000 mean) from the NOAA 1/4° daily Optimum Interpolation SST V2 dataset (https://www.psl.noaa.gov/data/gridded/data.noaa.oisst.v2.highres.html). Using temperature anomalies (rather than absolute temperature) is common in the marine sciences and has the advantage of removing seasonal cycles. Sea surface temperature anomaly has also been shown to be related to the marine distributions of both hake (Malick *et al.*, 2020) and Chinook (Shelton *et al.*, 2020).

Upwelling alters a host of physical and biological characteristics by bringing cool, nutrient-rich waters to the surface, which supports abundant plant and animal life. Upwelling can also influence the distributions of some fish (Sato *et al.*, 2018). We used the coastal upwelling transport index (CUTI) (Jacox *et al.*, 2018) as a metric of upwelling.

Model assessment

We performed cross-validation to assess the models' predictive ability because forecasting bycatch may be a useful management tool. Cross-validation splits the data into training and testing datasets. Each model was fit using only the training data. Then each model's predictive ability was assessed by comparing the response values observed in the testing dataset to those the model predicted for the testing dataset. We performed three sets of cross-validation, representing three potential management scenarios. The first, *k*-fold cross-validation, splits the dataset into an arbitrary number (k) of evenly sized subsets, or folds. We used tenfold cross-validation. Each fold is successively held out as a testing dataset while the model is fit to all other folds collectively. The result is that testing data are randomly distributed among training data along all covariate axes. Model predictions are interpolations, not extrapolations to new covariate spaces. *The k*-fold cross-validation represents a management scenario where static regulations are implemented once and apply in perpetuity. This is most appropriate in a static system where relationships among predictor and response variables remain constant; however, this method may result in overly optimistic model evaluations when the data are temporally structured or have other internal dependence structures.

In order to account for the temporal structure in our data and to test realistic predictive ability, we used blocked time-series cross-validation for the second and third cross-validations (Bergmeir and Benitez, 2012; Roberts *et al.*, 2017). The second cross-validation set we performed was time-series cross-validation, in which each testing dataset consisted of all hauls in the week immediately following the training dataset. In other words, we trained the models on observations from the block of time leading up to each week in the dataset, then made predictions for that week, which we compared to the true values. We refer to this set of cross-validation as weekly cross-validation, and it essentially tests the model's ability to predict one week in the future based on observations prior to that week. We excluded hauls in the first week of each fishing season from the testing datasets because they did not immediately follow the training dataset (the hake fishery is closed from January to mid-May each year). We repeated weekly cross-validation twice with different amounts of training data. The two training datasets included all hauls that occurred before, but in the same year as, the testing dataset and an additional 4 or 12 years of prior hauls. Because testing datasets are always in the future of training datasets, model predictions are always extrapolations along the time axis. Predicting into the future also makes models more vulnerable to misleading predictions if relationships among predictor and response variables change through time. Because these models made predictions into the future, they were fitted using 7-day lagged measures of SST anomaly and CUTI (i.e. the prediction of bycatch on 15 June 2018 used SST anomaly and CUTI from 8 June 2018). This cross-validation method represents a management (or voluntary bycatch avoidance) scenario with highly dynamic rules that change on a weekly basis and is most appropriate in a system that changes quickly as well.

The third cross-validation set is also time-series cross-validation, but with testing datasets comprised of all hauls in the year immediately following the training dataset. In other words, we trained the models on observations from the block of time prior to each year in the dataset, then made predictions for that year. We refer to this set as yearly cross-validation. As with weekly cross-validation, we repeated yearly cross-validation twice with different amounts of training data: either 4 or 12 years, and models used 1-year lagged SST anomaly and CUTI (i.e. prediction of bycatch on 15 June 2018 used SST anomaly and CUTI from 15 June 2017). As previously mentioned, yearly cross-validation required modification of the GAM 3. This cross-validation method represents a management scenario with annual updates to regulations and is most appropriate in a system where regulations are difficult to change quickly and/or systems change very slowly and are predictable many months in advance.

We assessed models' fit in each cross-validation set using the area under the receiver operating curve (AUC) for binomial classification models and the root mean square error (RMSE) for models of abundance (both the second half of the hurdle models and the full hurdle models). AUC ranges from 0 to 1 and represents the probability of correctly ranking a randomly drawn positive case above a randomly drawn negative case. A value of 1 represents perfect classification accuracy; a value of 0.5 represents very poor classification accuracy. RMSE is a very common metric of models' predictive accuracy and is in the same units as the response variable. RMSE weights over- and under-estimates equally and is sensitive to outliers.

For each model type in each cross-validation set and each length of the training dataset, we combined testing datasets to get predicted bycatch presence for each haul other than those in the first week of the fishing season each year, from 2014 to 2019. Predictions were limited to those years because models trained with 12 years of training data did not make predictions for any earlier years. We then calculated a single AUC value for each combination of model, cross-validation set, and training dataset. This allows for a direct comparison among modelling approaches because AUC is calculated using exactly the same hauls. We calculated the RMSE for the second half of the hurdle models similarly but used predicted abundance rather than predicted presence and only included the subset of hauls from 2014 to 2019 that contained bycatch. We calculated RMSE for the full hurdle models using the same hauls that we used to calculate AUC. In addition to calculating AUC and RMSE across all predicted hauls, we also calculated AUC and RMSE separately for each fold (for the *k*-fold cross-validation), week (for the weekly cross-validation), and year (for the yearly cross-validation). We then used Tukey's honestly significant difference (HSD) test to evaluate mean differences in the fold-, week-, or year-level metrics across models (Stock *et al.*, 2020). For the weekly cross-validation comparison, we removed the ∼5% of weeks that had <50 hauls, because AUC is poorly estimated for small sample sizes (Hanczar *et al.*, 2010).

Finally, to aid interpretation of each model's predictive ability, we calculated the reduction in bycatch that could be achieved by removing hauls from the test data that the full hurdle models predicted would contain the most bycatch. We note that this estimate is overly optimistic because a more realistic management scenario would be to relocate the fishing effort rather than remove it entirely.

Code to download the publicly available data, run the models, perform cross-validation, and visualize results is available at https://github.com/kricherson-NOAA/hake-salmon-bycatch.

Results

Model performance

For both the binomial (Figure 2) and abundance (Figure 3) components of the hurdle model, *k*-fold cross-validation resulted in lower RMSE and higher AUC compared to weekly and yearly cross-validation. For the full hurdle model, *k*-fold cross-validation also generally resulted in lower RMSE compared to the temporally blocked cross-validations (Figure 4). The further models predict into the future, the lower their predictive ability. For *k*-fold cross-validation, more complex linear-based models generally had a higher AUC and a lower RMSE compared to simpler linear-based models. This relationship largely holds for the weekly cross-validation set but collapses for yearly cross-validation. The decline in performance of complex linear-based models in the yearly cross-validation is exemplified in the most complex linear-based model, GAM 3. Tree-based models consistently outperformed GLMs and GAMs in *k*-fold cross-validation. Linear-based and tree-based models performed similarly in weekly and yearly cross-validation, though tree-based models were

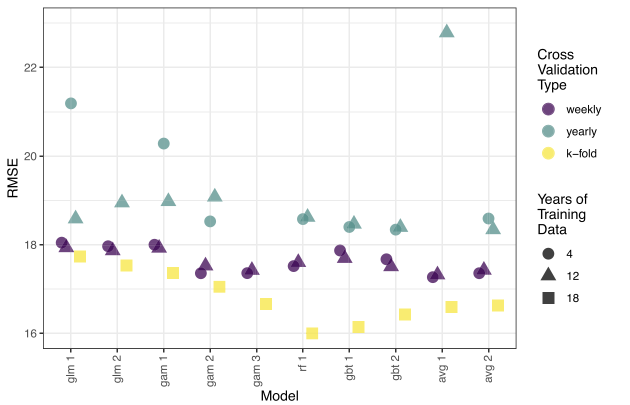

Figure 2. Comparison of different models' ability to predict the presence of bycatch in hauls between 2014 and 2019. A higher area under the receiver operating curve (AUC) is better. For information on individual models, see Table 1. Models generally increase in complexity from left to right. Avg 1 is the average of the ten models to the left of it. Avg 2 excludes model GAM 3.

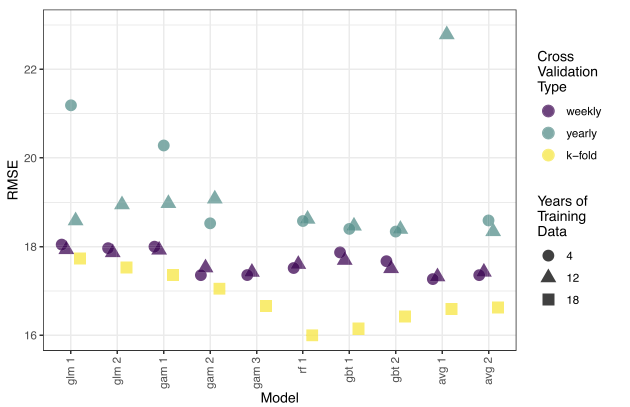

Figure 3. Comparison of different abundance models' ability to predict the abundance of bycatch in hauls between 2014 and 2019 that contained bycatch. A lower root mean squared error (RMSE) is better. For information on individual models, see Table 1. Models generally increase in complexity from left to right. Avg 1 is the average of the ten models to the left of it. Avg 2 excludes model GAM 3. The RMSE values for GAM 3 in the yearly cross-validation set were cropped out of the plot because they were orders of magnitude larger than the other values due to a number of hauls with very high predicted bycatch.

generally slightly better. Increasing the amount of data in the training set had little impact on model performance in the weekly cross-validation set but improved performance in the yearly cross-validation set. Model averages performed very well overall, performing better than any single model in the weekly and yearly cross-validation but not in the *k*-fold cross-validation. Among the RF models, downsampling was helpful for time-series predictions, but SMOTE was the worst-performing RF model for all cross-validation sets, suggesting that imbalance was not problematic in our dataset or that the synthetic data that SMOTE created did not adequately represent true data.

When comparing differences across fold-, week-, or year-level metrics, differences in RMSE across models were not significant at the $p = 0.05$ level according to Tukey's HSD test. However, for the AUC comparison, the models with the highest AUC were often significantly different from those with

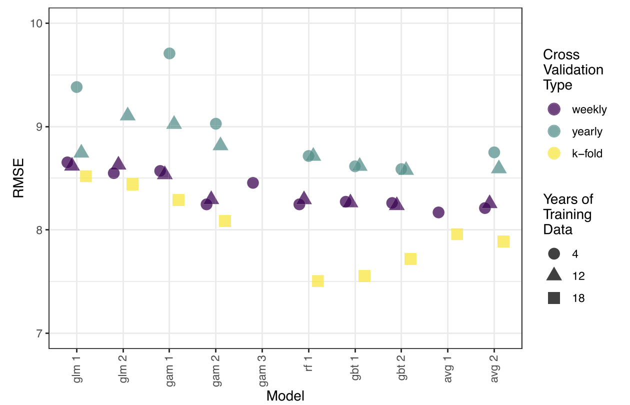

Figure 4. Comparison of different hurdle models' ability to predict the abundance of bycatch in hauls between 2014 and 2019. A lower RMSE is better. For information on individual models, see Table 1. Models generally increase in complexity from left to right. Avg 1 is the average of the ten models to the left of it. Avg 2 excludes model GAM 3. RMSE values for GAM 3 in the yearly and *k*-fold cross-validation sets were cropped out of the plot because they were orders of magnitude larger than the other values due to a number of hauls with very high predicted bycatch.

the lowest, particularly for the *k*-fold cross-validation (Supplementary Figures S25–S27). We noted that for GAM 3, a small number of hauls were sometimes associated with unrealistically high predicted bycatch, resulting in high RMSEs for the GAM 3 *k*-fold hurdle and yearly cross-validation (Figures 3–4, Supplementary Figures S25–S27).

Correlates of bycatch

Covariates in linear and tree-based models generally showed similar patterns in our marginal effect plots (see supplementary materials). There were higher levels of bycatch at night (Supplementary Figure S12), towards the end of the fishing season (Supplementary Figure S13), at shallower fishing depths (Supplementary Figure S15), in shallower water (Supplementary Figure S16), in more anomalous ocean temperatures (Supplementary Figure S18), and in stronger downwelling (Supplementary Figure S19). The spatial components of the models (see supplementary materials) suggested much higher bycatch in the Strait of Juan de Fuca and moderate increases closer to shore in northern Washington and southern Oregon/northern California (Supplementary Figure S5). The models also suggested higher bycatch in the southwest corner of our study area, but there are relatively few hauls from that area. Results from models fit to the full dataset can be found in Supplementary Tables S1–S5. Supplementary Table S1 includes information on overall model fit (e.g. % variance explained), and Supplementary Tables S2–S4 contain information about individual covariates (e.g parameter estimates and significance, where appropriate).

Reduction in bycatch

Removing a small proportion of hauls that are predicted to have the most bycatch had a large influence on the overall ratio of bycatch to hake (Figure 5). Using weekly cross-validation on the GBT 1 model trained with 4 years of data, removing just 1% of hauls reduced the bycatch-to-hake ratio by 20%. We show this model because it reduced the bycatch-to-hake ratio the most in the temporal cross-validations, though

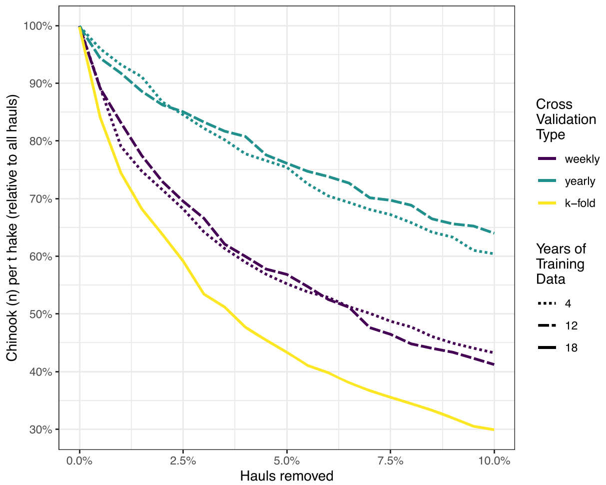

Figure 5. The reduction in Chinook-to-Hake ratio achieved by removing hauls that the full hurdle models predicted to have the most bycatch. Only the GBT 1 model is shown. See supplementary materials for a comparison of models. Note that training datasets in *k*-fold cross-validation included a random subset of hauls across all 18 years in the total dataset, whereas weekly and yearly cross-validation training datasets included all hauls in the 4 or 12 years of training data.

results were similar for GBT 2, which also performed well. Other models generally showed smaller reductions in bycatch (∼5–15% reductions in bycatch when 1% of hauls were removed; Supplementary Figure S22). Changes in the bycatch-to-hake ratio were driven almost entirely by changes in bycatch. Hake catch declined in direct proportion to hauls removed (i.e. removing 1% of hauls reduced hake catches by 1%, $R^2 = 0.9996$; analysis not shown), implying a lack of correlation between hake catches and Chinook bycatch. The hauls removed in these calculations were not uniformly distributed through time (more hauls were removed in some years than others). When calculated on a yearly basis, the reductions in bycatch-to-target ratio varied considerably, particularly for time-series cross-validation (Supplementary Figure S24). In some years (e.g. 2015, 2019), all three methods of cross-validation resulted in similar reductions in the ratio of Chinook-to-hake, while in other years (e.g. 2017, 2018), the relative performance of the different types of cross-validation closely resembled that seen in Figure 5.

Discussion

Overall, the models we tested were effective at identifying hauls with large amounts of bycatch. Fleet communication measures, bycatch quotas, and catch-share programmes (Gilman *et al.*, 2006; Somers *et al.*, 2018; Holland and Martin, 2019) have already reduced bycatch levels, but we show that further improvement may be possible. Using model-based predictions to redistribute fishing effort from areas with high expected bycatch to areas with low expected bycatch is better than simply moving fishing effort away from areas where bycatch is observed because the model-based approach helps to avoid moving fishing effort to other areas of high bycatch (Smith *et al.*, 2021). This is congruent with other efforts to model bycatch, which generally show that relatively minor modifications to fishing effort are likely to result in large declines in bycatch (Lewison *et al.*, 2009; Otto *et al.*, 2016; Stock *et al.*, 2020; Smith *et al.*, 2021), and in some cases the modifications to fishing behaviour may be revenue neutral for fishers (Otto *et al.*, 2016). We note that the benefits of predicting bycatch and adjusting fishing effort in response may not be consistent across time. For example, when the bycatch-to-target ratio was calculated on an annual basis, there was considerable variation across time, with much larger relative reductions in bycatch in some years compared to others (Supplementary Figure S24). Thus, if forecasting is implemented, it is realistic to expect that the results may vary.

Failing to account for temporal structure can lead to an underestimation of prediction error, so we used temporally blocked cross-validation to test our models. As expected, models' predictive ability declines the further into the future they make predictions. When interpolating predictions in *k*-fold cross-validation, tree-based models outperformed linear-based models by a substantial margin (at least as measured by AUC), but when extrapolating predictions with time-series cross-validation, the tree-based models performed only slightly better than the top-performing linear-based model, and this was not statistically significant when comparing across block-level cross-validation metrics (Supplementary Table S5, Supplementary Figures S25–S27). Becker *et al.* (2020) also found that their BRT model outperformed their GAM model in predicting the presence/absence of cetacean species, but the same BRT model performed worse at predicting future presence/absence. Similarly, Stock *et al.* (2020) showed that random forests were more sensitive to spatial extrapolation than some of their linear-based models when predicting bycatch in west coast fisheries. More generally, the decline in predictive ability of tree-based models when extrapolating results in space or time may be a result of complex models being overfit to training data, though we note that the linear-based models also had lower performance when predicting further out in time. Overfitting can often be reduced with careful tuning parameter selection, but this will not guarantee better predictions of the future; it may simply decrease the difference in predictive ability between interpolations and extrapolations. To avoid deceptively optimistic estimates of models' predictive ability, it is important to design cross-validation schemes that accurately reflect the way in which models will be used to make predictions.

Model averages performed well, particularly for time-series cross-validation, but were not better than the best individual models. Overall, tree-based models outperformed linear-based models (lower RMSE in testing data; Figure 4). GBT models with and without DART performed similarly, particularly in time-series cross-validation. However, we note that our study compared models with both differing model structure (e.g. linear versus non-linear covariates) and modelling techniques. Thus, we emphasize that both model structure and model type can influence performance and that our study does not disentangle them.

We chose our statistical methods to test the models' predictive accuracy. Therefore, we urge caution in using these models to determine the strength or statistical significance of the relationships between individual correlates and bycatch (see Tredennick *et al.*, 2021, for an overview of different modelling purposes and their limitations). Our methods do allow for useful generalizations regarding individual correlates of bycatch, particularly when validated by other studies finding similar results.

Our models generally agree that bycatch increases in anomalous temperatures. The same trend has been observed in other fisheries (e.g. bycatch of eulachon in the pink shrimp fishery; Ward *et al.*, 2015). However, in large temperature anomalies, different models suggest different relationships, with linear-based models suggesting more bycatch in warmer temperatures and GBT models the opposite. Temperature is likely affecting Chinook and hake differentially, with potentially varying relationships through time and/or space. Otto *et al.* (2016) did not find a relationship between SST and the catch of one Chinook stock (a metric of Chinook distribution) in the Chinook fishery, but they note that fishermen use SST to help locate Chinook, so their small sample sizes may have limited their ability to identify a relationship between temperature and Chinook catch. In addition, preferential sampling may result in differing apparent relationships between salmon catches (or bycatch) and temperature in targeted versus bycatch fisheries. In a much larger study of the ocean distribution of Chinook, Shelton *et al.* (2020) showed that different Chinook stocks have different relationships between their ocean distribution and the SST anomaly, and that those relationships can vary through time. Similarly, Malick *et al.* (2020) found a spatially variable relationship between adult hake distributions and subsurface sea temperature anomaly, with higher hake biomass off the coast of Washington and northern Oregon in cooler temperatures, but the relationship weakens and potentially reverses off the coast of southern Oregon and northern California. Therefore, the relationship between temperature and Chinook bycatch in the hake fishery is likely more complex than our linear models can capture.

Over the most common hake fishing depths (100–400 m), all models show decreasing bycatch with depth when other covariates are held constant at mean values (Supplementary Figure S15). That relationship could change for tree-based models with other combinations of covariate values. Otto *et al.* (2016) found the maximum catch probability of a single Chinook stock at ~150 m, but the amount of Chinook caught (given that some were caught) kept increasing with increasing depth. In spite of this, the mean capture depth for the stock was 72 m, suggesting that fishers target shallower depths. Indeed, the average capture depths across the 23 Chinook stocks reported by Otto *et al.* (2016) were all between 50 and 100 m. Teahan *et al.* (Teahan *et al.*, In Prep.) found that adult Chinook were caught in the Chinook fishery at deeper depths in warmer temperatures and in late summer, with vertical distribution also varying by Chinook stock and the amount of upwelling. However, mean capture depths across stocks and years in the targeted Chinook fishery were <25 m, so it is unclear whether any of these factors would increase Chinook abundance at the depths common in the hake fishery.

Fishers have some agency to change when, where, and how they fish, potentially altering bycatch in ways that our models fail to incorporate. All potential reductions in bycatch suggested by these models are based on the assumption that fishers do not alter their behaviour in ways that are not already accounted for in the models. For example, the models do account for fishing location (and thus spatial bycatch avoidance measures that were in place when the data were collected) but do not account for gear configurations, the use of bycatch exclusion devices, or the use of test hauls to look for bycatch species in an area before conducting longer hauls. Incentivizing fishers to reduce bycatch is often very efficient, demonstrating the effect that fishers' choices can have on bycatch (Abbott *et al.*, 2015; Somers *et al.*, 2018; Sugihara *et al.*, 2018). Ultimately, determining the amount of bycatch reduction these models can facilitate would require real-world validation.

One limitation of our modelling approach is that we exclude the first week of each fishing season from the calculations of model performance metrics. We do this to accommodate the weekly cross-validation but must acknowledge that excluding the first week of the fishing season, which is one of the busiest fishing weeks of the year in the hake fishery, potentially limits the inference of our models. However, we believe that our models do reflect achievable reductions in Chinook bycatch. From 2014 to 2019 (the years included in our summary statistics), 9% of hauls (1868 of 20007) were in the first week of a fishing season, but only 2.5% of bycatch was in the first week (788 of 30804 Chinook). Thus, our results (Figures 2 and 5) are relevant to 97.5% of the total Chinook bycatch and to the parts of the season when there is more scope to avoid bycatch (i.e. when bycatch ratios are higher). Furthermore, both the yearly and the *k*-fold cross-validation models provide some ability to predict bycatch in the first week of the fishing season, albeit roughly half of their predictive ability to reduce bycatch in later weeks (Supplementary Figure S23).

There are three ways in which our models could be improved to make them more useful to hake fishers and more relevant to Chinook conservation. First, incorporating the catch of hake into models would be an important step towards identifying areas that are not only likely to result in less Chinook bycatch but are also likely to result in more hake catch. As we noted earlier, switching to the ratio of Chinook-to-Hake as the response variable in our models fails to take into account the absolute amounts of either species, which is critically important. An alternative would be to model hake independently of Chinook bycatch and combine the results of both models, giving the expected absolute amounts of each species and the ratio between the two. Second, our models fail to account for costs to fishers associated with efforts to avoid Chinook bycatch. Bioeconomic models that account for costs (e.g. increased fuel, decreased hake) as well as benefits (i.e. decreased Chinook bycatch) would be useful in further expanding the utility of our models. Third, effective Chinook conservation warrants the incorporation of stock-specific information. Komoroske and Lewison (2015) identify uncertainty in the population-level effects of bycatch as one of the primary impediments to understanding bycatch and its effects. Currently, our models treat Chinook as a single population (yet they are not), in which case less bycatch of Chinook is always preferred. However, different stocks of Chinook tend to have different marine distributions (Satterthwaite *et al.*, 2015; Otto *et al.*, 2016; Shelton *et al.*, 2020), and genetic analysis indicates that there are latitudinal patterns in the stock composition of Chinook bycatch (Moran *et al.*, 2021). It is possible that hauls with relatively large amounts of Chinook bycatch have mostly fish from abundant Chinook stocks, while some hauls with relatively small amounts of Chinook bycatch may pose the largest threat to rare stocks. Identifying ways of reducing the bycatch of rare stocks would provide considerably more conservation value than do the results of our current models.

In addition to these model improvements, there are a number of additional modelling approaches that could be evaluated for their ability to predict Chinook bycatch in the hake fishery. For example, if the binomial and positive components of the hurdle model are not independent, using the Tweedie distribution may be advantageous (Stock *et al.*, 2019). Tweedie models may also have the advantage of reducing the unrealistic predictions noted above for a small number of hauls. Similarly, zero-inflated Poisson and/or negative binomial models could also reasonably be applied to these data. However, given that zero-inflated models often perform similarly to hurdle models in simulations, these models may not greatly improve predictive performance (Feng, 2021). In the context of linear-based models, additional interactions between terms could be considered, both to improve predictive ability and potentially to provide further inference about potential drivers of bycatch. To help avoid adding too many predictor variables when adding interaction terms, LASSO regularization (Tibshirani, 1996) can be used to perform automated variable selection. Our observation that tree-based models generally outperform linear-based models may be at least partially attributable to the fact that tree-based models can account for interactions. Thus, comparing them with linear-based models both with and without interaction terms could add additional insight about why some models may do better at predicting bycatch. Finally, another potential approach would be to use spatial time-series models based on latent Gaussian (Markov) random fields. Using latent Gaussian random fields as spatial random effects allows for forecasting, and this approach has shown utility in modelling bycatch patterns (Yan *et al.*, 2021), though it may not necessarily outperform other methods like random forests (Stock *et al.*, 2020). Finally, neural networks have shown predictive promise in fisheries applications (e.g. Núñez *et al.*, 2018; Chen *et al.*, 2021) and may also be useful in modelling bycatch.

Identifying methods to reduce Chinook bycatch, especially for threatened stocks, is likely to become even more important in the future as salmon populations face a changing and increasingly variable climate. In spite of conservation efforts, many Chinook populations continue to decline, with negative consequences for ecosystems, fisheries, and fishing communities. Though bycatch has not been identified as a primary threat to any Chinook populations where it has been studied (e.g. *Witherell et al.*, 2002; Ianelli and Stram, 2015; NMFS, 2017; Cunningham *et al.*, 2018), the regulation of harvest (including bycatch) is an important management lever influencing ocean salmon survival. In addition, legal mandates to protect threatened species, rebuild overfished stocks, and minimize bycatch mean that fisheries targeting abundant stocks (e.g. hake) may be constrained by impacts on non-target species like Chinook. Methods developed here could potentially be applied to other fisheries; for example, the Bering Sea walleye pollock (*Gadus chalcogrammus*) trawl fishery shares many characteristics with the west coast hake fishery, including the challenge of avoiding Chinook bycatch (Stram and Ianelli, 2015). More broadly, some data-rich fisheries are taking advantage of high-resolution environmental, tracking, and fishery data to inform dynamic management strategies, which show promise in reducing bycatch while maintaining target catches (Hazen *et al.*, 2018; Welch *et al.*, 2019; Pons *et al.*, 2022). The approach presented here shows how observer and environmental data, in combination with spatiotemporal models, can be used to forecast bycatch at realistic timescales that can enhance avoidance of a protected bycatch species while minimizing opportunity costs for the fishery.

Acknowledgements

We thank the observers in the At-Sea Hake Observer Program and West Coast Groundfish Observer Program for collecting the data that made this project possible. We thank Eric Ward, Brian Stock, and two anonymous reviewers for their helpful comments on this manuscript.

Supplementary material

The following supplementary material is available at *ICESJMS* online:
shirk_et_al_supplement.docx.

Conflict of interest

The authors declare no conflicts of interest.

Funding

This work was supported by the NOAA Fisheries Office of Sustainable Fisheries and administered through the Cooperative Institute for Marine Resources Studies at Oregon State University. This support is gratefully acknowledged.

Author contributions

K.R. and M.B. conceived of the project and obtained funding. V.T. oversaw data collection and curation. P.S. conducted data analyses. P.S. and K.R. created visualizations and wrote the original manuscript draft. All authors contributed to manuscript edits.

Data availability

The data used in this analysis are subject to confidentiality requirements set forth in the Magnuson-Stevens Act (MSA) at Section 402(b), 16 U.S.C. 1881a(b). This means that raw data is not available to the public except as described in the MSA. For example, an aggregated dataset that prevents public disclosure of the identity or business of any person can be requested from the Fisheries Observation Science programme at the Northwest Fisheries Science Center. The text of the MSA is available at https://www.fisheries.noaa.gov/resource/document/magnuson-stevens-fishery-conservation-and-management-act.

References

50 CFR § 660.60—Specifications and Management Measures. 2022. United States Code of Federal Regulations. https://www.ecfr.gov/current/title-50/chapter-VI/part-660/subpart-C/section-660.60 (Accessed 1 December 2022).

Abbott, J. K., Haynie, A. C., and Reimer, M. N. 2015. Hidden flexibility: institutions, incentives, and the margins of selectivity in fishing. Land Economics, 91: 169–195.

Angilletta, M. J. 2009. Thermal Adaptation: A Theoretical and Empirical Synthesis. Oxford University Press, Oxford. 289pp.

Becker, E. A., Carretta, J. V., Forney, K. A., Barlow, J., Brodie, S., Hoopes, R., Jacox, M. G. *et al.* 2020. Performance evaluation of cetacean species distribution models developed using generalized additive models and boosted regression trees. Ecology and Evolution, 10: 5759–5784.

Bergmeir, C., and Benitez, J. M. 2012. On the use of cross-validation for time series predictor evaluation. Information Sciences, 191: 192–213.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. 2002. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16: 321–357.

Chawla, N. V., Japkowicz, N., and Kotcz, A. 2004. Editorial: special issue on learning from imbalanced data sets. ACM SIGKDD Explorations Newsletter, 6: 1–6.

Chen, T., Zhong, L., Zhou, N., and Hoppe, D. 2021. Catch weight prediction for multi-species fishing using artificial neural networks. *In* 20th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, Pasadena, CA. pp. 1545–1552.

Cragg, J. G. 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica, 39: 829–844.

Cunningham, C. J., Westley, P. A., and Adkison, M. D. 2018. Signals of large scale climate drivers, hatchery enhancement, and marine factors in Yukon river Chinook salmon survival revealed with a Bayesian life history model. Global Change Biology, 24: 4399–4416.

Dormann, C. F. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Global Ecology and Biogeography, 16: 129–138.

Dormann, C. F., Calabrese, J. M., Guillera-Arroita, G., Matechou, E., Bahn, V., Bartoń, K., Beale, C. M. *et al.* 2018. Model averaging in ecology: a review of Bayesian, information-theoretic, and tactical approaches for predictive inference. Ecological Monographs, 88: 485–504.

Eguchi, T., Benson, S. R., Foley, D. G., and Forney, K. A. 2017. Predicting overlap between drift gillnet fishing and leatherback turtle habitat in the California current ecosystem. Fisheries Oceanography, 26: 17–33.

Elith, J., Leathwick, J. R., and Hastie, T. 2008. A working guide to boosted regression trees. Journal of Animal Ecology, 77: 802–813.

Feng, C. X. 2021. A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. Journal of Statistical Distributions and Applications, 8: 8.

Gilman, E., Passfield, K., and Nakamura, K. 2014. Performance of regional fisheries management organizations: ecosystem-based governance of bycatch and discards. Fish and Fisheries, 15: 327–351.

Gilman, E. L., Dalzell, P., and Martin, S. 2006. Fleet communication to abate fisheries bycatch. Marine Policy, 30: 360–366.

Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M., and Dougherty, E. R. 2010. Small-sample precision of ROC-related estimates, Bioinformatics, 26: 822–830.

Hazen, E. L., Scales, K. L., Maxwell, S. M., Briscoe, D. K., Welch, H., Bograd, S. J., Bailey, H. *et al.* 2018. A dynamic ocean management tool to reduce bycatch and support sustainable fisheries. Science Advances, 4: eaar3001.

Hijmans, R. J. 2020. Raster: geographic data analysis and modeling. https://CRAN.R-project.org/package=raster (Accessed 1 December 2022).

Holland, D. S., and Martin, C. 2019. Bycatch quotas, risk pools, and cooperation in the Pacific whiting fishery. Frontiers in Marine Science, 6: 600.

Ianelli, J. N., and Stram, D. L. 2015. Estimating impacts of the pollock fishery bycatch on western Alaska Chinook salmon. ICES Journal of Marine Science, 72: 1159–1172.

Jacox, M. G., Edwards, C. A., Hazen, E. L., and Bograd, S. J. 2018. Coastal upwelling revisited: ekman, bakun, and improved upwelling indices for the U.S. West Coast. Journal of Geophysical Research: Oceans, 123: 7332–7350.

Johnson, K. F., Edwards, A. M., Berger, A. M., and Grandin, C. J. 2021. Status of the Pacifc Hake (Whiting) Stock in U.S. and Canadian Waters in 2021. Prepared by the Joint Technical Committee of the U.S. and Canada Pacific Hake/Whiting Agreement, National Marine Fisheries Service and Fisheries and Oceans Canada. http://web.archive.org/web/20210406173057/https://media.fisheries.noaa.gov/2021-03/hake-assessment-post-srg-final_20210303.pdf (Accessed 1 December 2022).

Komoroske, L. M., and Lewison, R. L. 2015. Addressing fisheries bycatch in a changing world. Frontiers in Marine Science, 2: 1–11.

Lewison, R. L., Soykan, C. U., and Franklin, J. 2009. Mapping the bycatch seascape: multispecies and multi-scale spatial patterns of fisheries bycatch. Ecological Applications, 19: 920–930.

Lewison, R. L., Crowder, L. B., Wallace, B. P., Moore, J. E., Cox, T., Zydelis, R., McDonald, S. *et al.* 2014. Global patterns of marine mammal, seabird, and sea turtle bycatch reveal taxa-specific and cumulative megafauna hotspots. Proceedings of the National Academy of Sciences, 111: 5271–5276.

Little, A. S., Needle, C. L., Hilborn, R., Holland, D. S., and Marshall, C. T. 2015. Real-time spatial management approaches to reduce bycatch and discards: experiences from Europe and the United States. Fish and Fisheries, 16: 576–602.

Mackas, D. L., Kieser, R., Saunders, M., Yelland, D. R., Brown, R. M., and Moore, D. F. 1997. Aggregation of euphausiids and Pacific hake (*Merluccius productus*) along the outer continental shelf off Vancouver Island. Canadian Journal of Fisheries and Aquatic Sciences, 54: 2080–2096.

Malick, M., Hunsicker, M., Haltuch, M., Parker-Stetter, S., Berger, A., and Marshall, K. 2020. Relationships between temperature and Pacific hake distribution vary across latitude and life-history stage. Marine Ecology Progress Series, 639: 185–197.

Matson, S. E., and Erickson, D. L. 2018. Analysis of West Coast groundfish fisheries for the 2017 biological opinion on endangered species act-listed salmon. *In* NOAA Technical Memorandum NMFS-OSF-7, 93pp. NOAA Fisheries, Silver Spring, MD.

Moran, P., Tuttle, V. J., Bishop, S., and LaVoy, L. 2021. Compositional forecasting of Chinook Salmon Evolutionarily Significant Units in bycatch for Pacific Hake fisheries. bioRxiv. https://www.biorxiv.org/content/10.1101/2021.11.29.470462v1 (Accessed 1 December 2022).

Muñoz, N. J., Farrell, A. P., Heath, J. W., and Neff, B. D. 2015. Adaptive potential of a Pacific salmon challenged by climate change. Nature Climate Change, 5: 163–166.

(NMFS) National Marine Fisheries Service. 2017. Endangered Species Act (ESA) Section 7(a) (2) Biological Opinion; Reinitiation of Section 7 Consultation Regarding the Pacific Fisheries Management Council's Groundfish Fishery Management Plan. National Marine Fisheries Service West Coast Region, Portland, OR.

NOAA National Geophysical Data Center. 2003a. U.S. Coastal Relief Model Vol.7—Central Pacific. NOAA National Centers for Environmental Information, Silver Spring, MD.

NOAA National Geophysical Data Center. 2003b. U.S. Coastal Relief Model Vol.8—Northwest Pacific. NOAA National Centers for Environmental Information, Silver Spring, MD.

Núñez, R. A. C., de la Peña, M., Irigollen, A. F., and Rodriguez, M. R. 2018. Deep learning models for the prediction of small-scale fisheries catches: finfish fishery in the region of "Bahia magadalena-almejas". ICES Journal of Marine Science, 75: 2088–2096.

(NWFSC) Northwest Fisheries Science Center. 2021a. At-Sea Hake Observer Program, 2021 Sampling Manual. Northwest Fisheries Science Center, Seattle, WA.

(NWFSC) Northwest Fisheries Science Center. 2021b. West Coast Groundfish Observer Program, 2021 Training Manual. Northwest Fisheries Science Center, Seattle, WA,

Otto, S. A., Simons, S., Stoll, J. S., and Lawson, P. 2016. Making progress on bycatch avoidance in the ocean salmon fishery using a transdisciplinary approach. ICES Journal of Marine Science, 73: 2380–2394.

(PFMC) Pacific Fishery Management Council. 2021. Review of 2020 Ocean Salmon Fisheries. Pacific Fishery Management Council, Portland, OR.

Pons, M., Watson, J.T., Ovando, D., Andraka, S., Brodie, S., Domingo, A., Fitchett, M. *et al.* 2022. Trade-offs between bycatch and target catches in static versus dynamic fishery closures. Proceedings of the National Academy of Sciences, 119: e2114508119.

Rashmi, K. V., and Gilad-Bachrach, R. 2015. DART: dropouts meet multiple additive regression trees. In Artificial Intelligence and Statistics (pp. 489-497). PMLR, arxiv.

Richerson, K. E., Somers, K. A., Jannot, J. E., Tuttle, V. J., Riley, N. B., and McVeigh, J. T. 2020. Observed and Estimated Bycatch of Salmon in

U.S. West Coast Fisheries, 2002–19. U.S. Department of Commerce, NOAA Data Report NMFS-NWFSC-DR-2020-04. https://repository.library.noaa.gov/view/noaa/27559 (Accessed 28 April 2021).

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S. *et al.* 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40: 913–929.

Sato, M., Barth, J. A., Benoit-Bird, K. J., Pierce, S. D., Cowles, T. J., Brodeur, R. D., and Peterson, W. T. 2018. Coastal upwelling fronts as a boundary for planktivorous fish distributions. Marine Ecology Progress Series, 595: 171–186.

Satterthwaite, W. H., Ciancio, J., Crandall, E., Palmer-Zwahlen, M. L., Grover, A. M., O'Farrell, M. R., Anderson, E. C. *et al.* 2015. Stock composition and ocean spatial distribution inference from California recreational Chinook salmon fisheries using genetic stock identification. Fisheries Research, 170: 166–178.

Shelton, A. O., Thorson, J. T., Ward, E. J., and Feist, B. E. 2014. Spatial semiparametric models improve estimates of species abundance and distribution. Canadian Journal of Fisheries and Aquatic Sciences, 71: 1655–1666.

Shelton, A. O., Sullaway, G. H., Ward, E. J., Feist, B. E., Somers, K. A., Tuttle, V. J., Watson, J. T. *et al.* 2021. Redistribution of salmon populations in the northeast Pacific ocean in response to climate. Fish and Fisheries, 22: 503–517.

Smith, J. A., Tommasi, D., Welch, H., Hazen, E. L., Sweeney, J., Brodie, S., Muhling, B. *et al.* 2021. Comparing dynamic and static time-area closures for bycatch mitigation: a management strategy evaluation of a swordfish fishery. Frontiers in Marine Science, 8: 630607.

Somers, K. A., Pfeiffer, L., Miller, S., and Morrison, W. 2018. Using incentives to reduce bycatch and discarding: results under the West Coast catch share program. Coastal Management, 46: 621–637.

Somers, K. A., Jannot, J. E., and Richerson, K. 2020. Estimated Discard and Catch of Groundfish Species in the 2018 U.S. West Coast Fisheries. Northwest Fisheries Science Center (U.S.), Seattle, WA.

Stock, B. C., Ward, E. J., Thorson, J. T., Jannot, J. E., and Semmens, B. X. 2019. The utility of spatial model-based estimators of unobserved bycatch. ICES Journal of Marine Science, 76: 255–267.

Stock, B. C., Ward, E. J., Eguchi, T., Jannot, J. E., Thorson, J. T., Feist, B. E., and Semmens, B. X. 2020. Comparing predictions of fisheries bycatch using multiple spatiotemporal species distribution model frameworks. Canadian Journal of Fisheries and Aquatic Sciences, 77: 146–163.

Stram, D. L., and Ianelli, J. N. 2015. Evaluating the efficacy of salmon bycatch measures using fishery-dependent data. ICES Journal of Marine Science, 72: 1173–1180.

Stuart-Smith, R. D., Edgar, G. J., and Bates, A. E. 2017. Thermal limits to the geographic distributions of shallow-water marine species. Nature Ecology & Evolution, 1: 1846–1852.

Sugihara, G., Criddle, K. R., McQuown, M., Giron-Nava, A., Deyle, E., James, C., Lee, A. *et al.* 2018. Comprehensive incentives for reducing Chinook salmon bycatch in the Bering Sea walleye pollock fishery: individual tradable encounter credits. Regional Studies in Marine Science, 22: 70–81.

Teahan, A., Bellenger, M. R., Miller, S., and Banks, M. In prep. Marine heatwaves influence depth preferences of Chinook salmon (*Oncorhynchus tshawytscha*). MSc thesis, Oregon State University, Corvallis, OR.

Thorson, J. T., Shelton, A. O., Ward, E. J., and Skaug, H. J. 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES Journal of Marine Science, 72: 1297–1310.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58: 267–288.

Tredennick, A. T., Hooker, G., Ellner, S. P., and Adler, P. B. 2021. A practical guide to selecting models for exploration, inference, and prediction in ecology. Ecology, 6: e03336.

Ward, E. J., Jannot, J. E., Lee, Y.-W., Ono, K., Shelton, A. O., and Thorson, J. T. 2015. Using spatiotemporal species distribution models to identify temporally evolving hotspots of species co-occurrence. Ecological Applications, 25: 2198–2209.

Warlick, A., Steiner, E., and Guldin, M. 2018. History of the West Coast groundfish trawl fishery: tracking socioeconomic characteristics across different management policies in a multispecies fishery. Marine Policy, 93: 9–21.

Welch, H., Hazen, E.L., Bograd, S.J., Jacox, M.G., Brodie, S., Robinson, D., Scales, K.L. *et al.* 2019. Practical considerations for operationalizing dynamic management tools. Journal of Applied Ecology, 56: 459–469.

Witherell, D., Ackley, D., and Coon, C. 2002. An overview of salmon bycatch in Alaska groundfish fisheries. Alaska Fishery Research Bulletin, 9: 53–64.

Wood, S. N. 2017. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC Press, Boca Raton, FL. 476 pp.

Yan, Y., Cantoni, E., Field, C., Treble, M., and Flemming, J. M. 2022. Spatiotemporal modeling of bycatch data: methods and a practical guide through a case study in a Canadian Arctic fishery. Canadian Journal of Fisheries and Aquatic Sciences, 79: 148–158.

Zeller, D., Cashion, T., Palomares, M., and Pauly, D. 2018. Global marine fisheries discards: a synthesis of reconstructed data. Fish and Fisheries, 19: 30–39.

Handling Editor: Shijie Zhou