Canadian Science Publishing

ARTICLE

Improving forecasts of sockeye salmon (*Oncorhynchus nerka*) with parametric and nonparametric models

Daniel Ovando, Curry Cunningham, Peter Kuriyama, Christopher Boatright, and Ray Hilborn

Abstract: Accurate forecasts of sockeye salmon (*Oncorhynchus nerka*) in Bristol Bay, Alaska, play an important role in management and harvesting decisions for this culturally and ecologically vital species. We used a suite of parametric and nonparametric models to assess the frontiers in forecast accuracy of Bristol Bay sockeye salmon possible given currently available data. In retrospective performance testing individual models were capable of reducing pre-season forecast error at the river system level by on average 15% relative to a benchmark model. We used an ensemble modeling approach to produce pre-season forecasts based on historical performance of individual models. This ensemble model reduced river system forecast error by 13% on average in 5 of the 7 evaluated river systems, though it increased forecast error by 39% on average in the remaining 2 systems. We found potential for modest improvements in forecast accuracy across a variety of scales. However, all tested models failed to accurately predict certain periods in the historical salmon return patterns, indicating that further forecast improvements likely depend on novel data rather than more flexible models.

Résumé : Des prédictions exactes concernant le saumon rouge (*Oncorhynchus nerka*) dans la baie de Bristol (Alaska) sont importantes pour la prise de décisions de gestion et de récolte pour cette espèce d'importance vitale des points de vue tant culturel qu'écologique. Nous utilisons un ensemble de modèles paramétriques et non paramétriques pour évaluer les frontières de l'exactitude possible des prédictions concernant le saumon rouge de la baie de Bristol étant donné les données disponibles. L'évaluation rétrospective de la performance montre que des modèles individuels sont en mesure de réduire l'erreur des prédictions établies avant la saison à l'échelle du réseau hydrographique d'en moyenne 15 % par rapport à un modèle de référence. Nous employons une approche de modélisation ensembliste pour produire des prédictions pré-saison reposant sur la performance passée de modèles individuels. Ce modèle ensembliste réduit l'erreur de prédiction pour le réseau hydrographique d'en moyenne 13 % pour 5 des 7 réseaux évalués, la rehaussant toutefois de 39 % en moyenne pour les deux autres réseaux. Nous notons un potentiel d'améliorations modestes de l'exactitude des prédictions à différentes échelles. Tous les modèles évalués n'ont toutefois pas réussi à prédire avec exactitude certaines périodes des registres passés des retours de saumons, ce qui indique que l'amélioration future des prédictions dépendra vraisemblablement de nouvelles données et non de modèles plus souples. [Traduit par la Rédaction]

1. Introduction

Animal populations exhibit complex dynamics driven by interactions with many aspects of their ecosystem. Predicting the outcomes of these dynamics is a critical task of natural resource management. Forecasts of future abundance are often used to set fisheries regulations, vessel operators may make decisions about alternative fisheries based on predicted abundance, and industries and communities use forecasts to inform long-term and short-term investment plans in staffing and production capacity. The past two decades have seen explosive progress in the ability of modern "computer age" (Efron and Hastie 2016) parametric and nonparametric models to improve prediction, revolutionizing fields such as financial modeling, weather forecasting, and medicine. However, these predictive methods are still uncommon in applied ecological forecasting (Peters et al. 2014). We use the ecologically and economically critical sockeye salmon (*Oncorhynchus nerka*) populations of Bristol Bay, Alaska, to assess the potential of a range of predictive modeling techniques to improve pre-season forecasts and identify frontiers in forecast accuracy achievable given currently available data.

Sockeye salmon are semelparous, born in fresh water where they spend the first one or more years of their lives. Eventually, these fish migrate to the ocean, where they remain until returning to their natal river systems to spawn and then die. Sockeye salmon exhibit life history variation in the number of years they spend in these freshwater and oceanic phases, representing distinct "age groups". Following conventions in the salmon literature, we denote age groups here by the format "years spent in fresh water. years spent in the ocean". For example, a fish in the 1.2 age group spent one winter post-hatching in fresh water before migrating to sea two years after it was spawned, and two winters in the ocean before returning to fresh water to spawn. The Bristol Bay sockeye salmon fishery is primarily made up of salmon from seven different river systems, each of which is managed as a separate stock (Fig. 1).

The commercial salmon fishery in Bristol Bay, Alaska, is the single largest sockeye salmon fishery in the world (Steiner et al. 2011). The estimated wholesale value of the Bristol Bay commercial sockeye

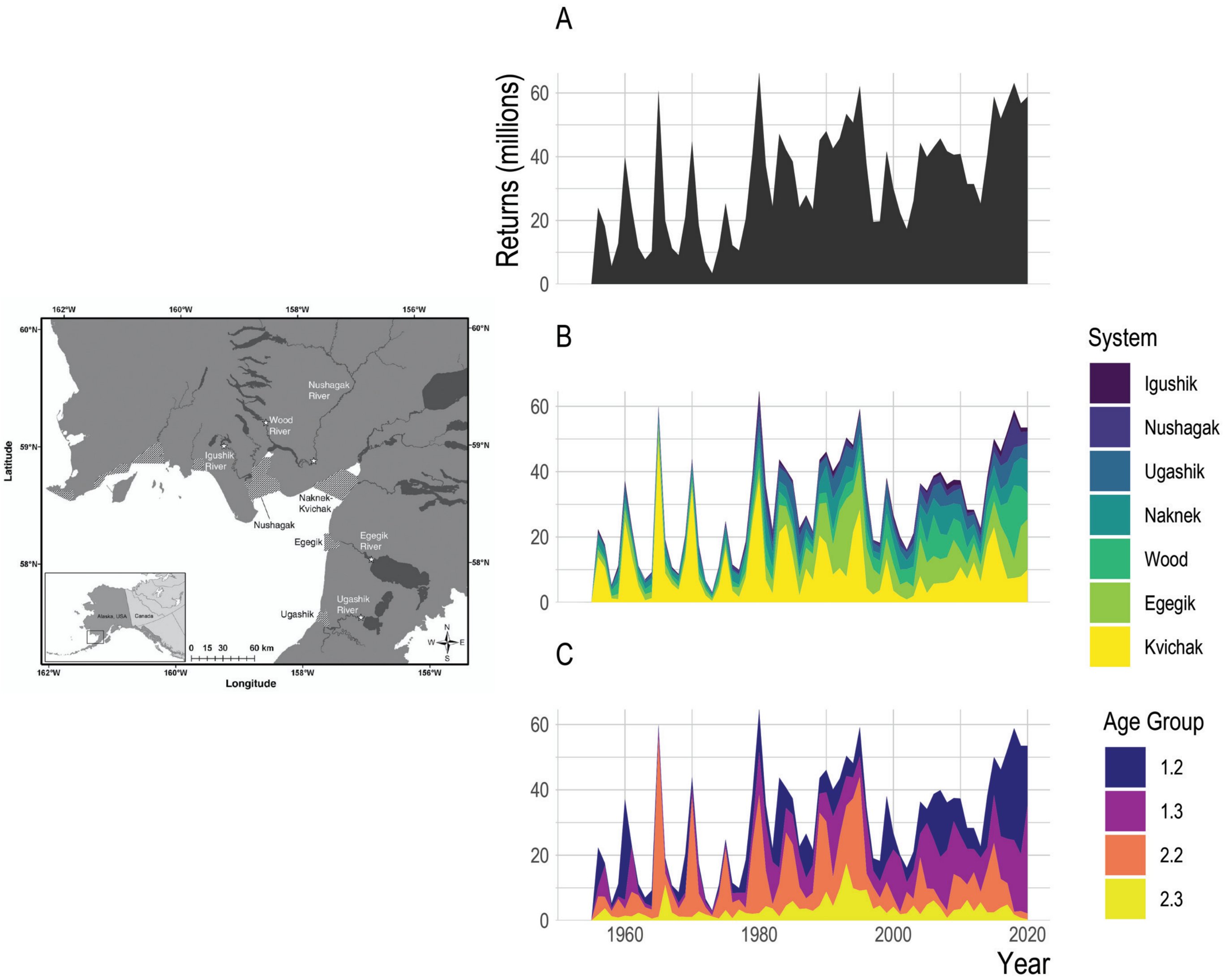
Received 20 October 2021. Accepted 11 January 2022.

D. Ovando and C. Boatright. School of Aquatic and Fishery Sciences, University of Washington, Box 355030, Seattle, WA 98195, USA.
C. Cunningham. College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, 17101 Point Lena Loop Road, Juneau, AK 99801, USA.
P. Kuriyama. NOAA Fisheries Southwest Fisheries Science Center, 8901 La Jolla Shores Dr., La Jolla, CA 92037, USA.
R. Hilborn. School of Aquatic and Fishery Sciences, University of Washington, Box 355030, Seattle, WA 98195, USA; Center for Sustaining Seafood, University of Washington, Box 355030, Seattle, WA 98195, USA.

Corresponding author: Daniel Ovando (email: danovan@uw.edu).

Published at www.cdnsciencepub.com/cjfas on 26 January 2022.

Fig. 1. Annual total abundance of returning sockeye salmon (*Oncorhynchus nerka*) to Bristol Bay, Alaska (A), by river system (B), and by age group (C). Numbers indicate millions of salmon. Age group is formatted by "years spent in fresh water.years spent in ocean". Map adapted from Cunningham et al. (2019). [Colour online.]



harvest was US$390 million in 2010, providing approximately one-sixth of the total value of all United States seafood exports (Knapp et al. 2013). The value of the Bristol Bay fishery has continued to grow, reaching US$508 million in 2017 (McDowell Group 2018). Salmon returning to Bristol Bay also provide vital food security for subsistence-dependent Alaskan communities and are critical vectors of marine-derived nutrients that support vibrant freshwater habitats (Naiman et al. 2002; Schindler et al. 2003). Sustainable management of the Bristol Bay salmon fishery depends in part on the accuracy of pre-season forecasts for salmon abundance, which inform development and implementation of in-season harvest strategies and successful operation of subsistence and commercial fisheries. While in-season forecasts updated as evidence for the strength and timing of a run accrue are vital for management of many salmon stocks, pre-season forecasts are also important for planning by the processing industry, as a basis for identifying the appropriate level of supplies, equipment, and personnel necessary to process the annual harvest. As such, the accuracy of salmon forecasts has a direct influence on the profitability and efficiency of the salmon industry as a whole, particularly for stocks with a shorter harvest window relative to the time needed to plan and adjust fishing and management actions.

The Fisheries Research Institute (FRI) at the University of Washington and the Alaska Department of Fish and Game (ADFG) have been providing pre-season forecasts for the annual abundance of sockeye salmon returning to the major river systems and fishing districts of Bristol Bay since at least 1967 (C. Cunningham, personal communication.). While the exact statistical methods used for FRI and ADFG forecasts have evolved over time, throughout their history they have primarily been based on the relationship between the abundance of successive age classes of salmon returning in different years. While these traditional forecast methods have been useful in guiding decisions by fishers, processors, and managers alike, improvements in the accuracy and precision of pre-season forecasts would represent a valuable advance.

There exists some "frontier" of maximum predictive ability contained in the available data used to make forecasts. However, a model might perform far below the predictive frontier attainable given available data if it is severely misspecified (i.e., if the assumptions of the model do not properly reflect reality). Identifying the best predictive model in an ecological setting has conventionally been achieved by manual construction and comparison of competing models via some information criteria. However, this can be a cumbersome process, particularly as the number of covariates and subsequently the number of potential interactions and nonlinearities increases.

Alternatively, nonparametric methods such as machine learning excel at identifying and exploiting potentially complex correlations between variables in a system. Parametric statistical methods often restrict themselves to simplified (e.g., linear) and often nondynamic representations of natural systems, both for analytical tractability and to facilitate heuristic understanding of underlying processes. Typically, these parametric statistical methods are concerned with explicitly estimating and interpreting model parameters rather than solely forecasting responses, such as population size (Beyan and Browman 2020; Malde et al. 2020). We would expect nonparametric methods to show substantial improvement in predictive power when the "true" underlying system linking observed variables and outcomes differs dramatically from the simplified representations of the system approximated by more parametric statistical approaches. In the case of salmon, we know that interannual variation in run sizes is affected by a wide range of ecosystem variables, including spawning success, river conditions, oceanic predator and prey abundance, and competition with other salmonids (Connors et al. 2020). By reducing the potential for predictive model misspecification, nonparametric models that essentially seek to "learn" the best model structure for the sole purpose of maximizing out-of-sample predictive performance can provide a test of the predictive information content of the available data themselves. Utilizing multiple parametric and nonparametric modeling approaches allows us to develop a clearer picture of this frontier in predictive ability achievable given a certain set of data.

Salmon forecasting has traditionally relied on cohort or "sibling" regression methods, in which the return abundance of an older age class is predicted by the abundance of younger age classes, returning in prior years but originating from the same river system and brood year. For example, the return abundance of four-year-old fish are predicted by the returns of three-year-old fish observed in the previous season. There are good reasons for this practice: trends in sibling abundance integrate across many environmental factors affecting salmon survival and returns. If a particular cohort suffers from poor environmental conditions, increased competition with conspecifics, or a greater abundance of predators, the demographic impacts of these changes will be reflected in the return abundance of younger age classes from the same cohort (i.e., originating from the same brood year) that experienced similar environmental conditions or resource availability, and by extension survival, as juveniles.

However, the standard sibling regression method does have shortcomings, most notably the underlying assumption of consistency in the relationship between the abundance of different age classes and stability in the maturation schedule (i.e., the probability of salmon maturing and returning to fresh water to spawn after a given number of years in the ocean). For example, if environmental conditions cause members of a cohort of salmon to spend more time at sea than in previous years, a sibling regression might underpredict the number of future returns. In addition, sibling regression requires accurate observations of the return abundance for younger sibling age classes, limiting the performance of these models in predicting returns of younger salmon for which few or no siblings (i.e., returning younger age classes) have yet been observed.

We hypothesize that directly incorporating data on candidate potentially time-varying factors influencing and correlated with salmon return size, rather than relying on sibling returns alone, may help improve forecast performance given the complex dynamics of salmon populations. However, these variables are likely to have complex, nonlinear, and nonstationary effects on salmon populations, potentially obscuring their value from conventional parametric statistical approaches with user-defined parameters, structures, and error distributions. To explore this possibility, we used a suite of four methods together with a panel of data on salmon populations and environmental conditions in Bristol Bay, Alaska, to explore what if any improvements in forecast skill could be achieved. These models included two machine learning methods, a random forest (rand_forest) (Breiman 2001; Wright and Ziegler 2017) and a boosted regression tree (boost_tree) (Chen et al. 2020), empirical dynamic models (edm) (Sugihara and May 1990; Ye et al. 2020; Munch et al. 2020), and dynamic linear models (dlm) (Pole et al. 1994; Petris et al. 2009). We compared each model to the performance of a lag(1) model in which the predicted returns for a given age group and river system in a year are equal to the observed returns for that that age group in that river system in the prior year. We also evaluated the performance of a model ensemble that weights predictions from individual ensemble members (alternative predictive model types) based on recent performance, and compare this with observed performance from the FRI forecast, a benchmark forecast that utilizes a qualitative ensemble approach based on evaluation of recent performance for alternative models within the ensemble.

Other studies have incorporated various time-varying parametric and nonparametric models, including versions of the models used here, for salmon forecasting (Holt and Peterman 2004; DFO 2018; Vêlez-Espino et al. 2019; Yi et al. 2019). Our study builds on this literature not by seeking to establish whether one type of model performs inherently better than others, but by examining the ability of a suite of these models to collectively improve salmon forecasting by leveraging correlations both within and among river systems and age classes in Bristol Bay. In doing so, we demonstrate how collections of parametric and nonparametric models can be used to identify frontiers in forecasting ability available in a given dataset.

2. Materials and methods

All code and data needed to fully replicate our results are publicly available at https://github.com/DanOvando/salmon-forecast-paper/. We describe critical details of each our main methods here. All analyses were conducted in the R programming language (R Core Team 2021).

The general structure of our methods are as follows:

1. Individual models for each river system and age group were fit to historical data.
2. Retrospective performance of individual model was assessed using one-step-ahead predictions (e.g., model fit to data through 1999 and used to predict return abundance in 2000) over the period 2000–2020.
3. Comparison of performance from individual model types against a benchmark "lag(1)" prediction model in which the forecast for next year is simply the observed returns in the previous year.
4. Individual models were aggregated into a statistical ensemble model based on their historical performance against the lag(1) benchmark.
5. The statistical ensemble model was then compared to a more qualitatively constructed ensemble model, in which researchers manually select individual models from an evolving suite of methods based on recent (20-year) performance. This is the method historically used to generate Bristol Bay salmon forecasts, although the individual prediction models within the selection suite have changed over time. As such this method provides a status quo benchmark to which individual models and statistical ensembles may be compared.

2.1. Data

2.1.1. Salmon returns

The primary data behind this analysis are historical numbers of sockeye salmon by age group returning to each of the seven Bristol Bay river systems considered here (Fig. 1). We included data from 1963 through 2020, omitting pre-1963 data as that year marks a major change in the data collection methods. We generated forecasts for the four most prominent age groups in the data, the

Table 1. Environmental and salmonid datasets available to machine learning models.

Name	Description	Source
Pacific Decadal Oscillation	Mean PDO index between May–August in year cohort entered ocean	JISAO
Sea level pressure	Median Bristol Bay SLP between May–August in year cohort entered ocean	ERDDAP ICOADS
Sea surface temperature	Median Bristol Bay SST between May–August in year cohort entered ocean	ERDDAP HadISST
Wind stress	Median Bristol Bay wind stress between May–August in year cohort entered ocean	ERDDAP ICOADS
Pink and chum abundance	Natural origin returns of pink and chum salmon	Ruggerone and Irvine 2018

Note: JISAO, Joint Institute for the Study of the Atmosphere and Ocean; ERDDAP, National Ocean and Atmospheric Administration's Environmental Research Division's Data Access Program; ICOADS, International Comprehensive Ocean–Atmosphere Data Set; HadISST, Hadley Centre Global Sea Ice and Sea Surface Temperature.

youngest of which being the 1.2 age group. However, we include data from younger age groups as candidate covariates. For example, returns of 1.1 fish in the year 2000 are used to generate forecasts of 1.2 fish in the year 2021, even though we never forecast 1.1 fish explicitly. In this manner all forecasts generated by our model can be based at least in part on previous observed siblings.

The dynamics of the Bristol Bay salmon runs changed dramatically from their historical patterns starting in the 1980s (Fig. 1). We chose to include data from before and after this change, rather than fitting to data from the more recent regime only, as exploratory analyses found better forecast accuracy resulting from inclusion of the full dataset. Since each of the models used here have the capacity for time-varying parameters, including data from before and after 1980 allows in theory for the model to leverage shared patterns across the two regimes while also theoretically learning about changes in patterns over this time period. All models were fit on the raw unit-scale (i.e., not log-transformed) returns, as we found better performance with this route than through log-transformation.

2.1.2. Additional covariates

Along with sockeye salmon returns, we also included several environmental and salmonid datasets as potential covariates (Table 1). Environmental data included the strength of the Pacific Decadal Oscillation (PDO), sea surface temperature, sea level pressure, and wind stress. Each of these variables were included as the mean or median of the values of that index over the Bristol Bay area between May and August of the year in which the cohort being forecasted would have entered the ocean. Our assumption here is that this early oceanic period represents a critical stage in the survival of sockeye. We tested treating gridded values of environmental covariates over space and time as predictors (rather than aggregating to the Bristol Bay-wide value), in theory allowing the models to learn which locations and times were the most useful predictors, but found this approach to perform poorly, likely due to the sample size available. Environmental datasets were queried from the NOAA Environmental Research Division's Data Access Program (ERDDAP) portal using the rerddap package in R (Chamberlain 2019). We also included as candidate covariates natural origin returns of pink (*Oncorhynchus gorbuscha*) and chum (*Oncorhynchus keta*) salmon from a range of North Pacific stocks, pulled from Ruggerone and Irvine (2018). While all of the models used in this paper are capable of including all of the covariates included in Table 1 in some manner, only the machine learning models made use of these data.

2.2. Machine learning models

We evaluated two different machine learning models: a random forest (rand_forest, implemented through the ranger package in R; Wright and Ziegler 2017), and boosted regression trees (boost_tree) through the xgboost package (Chen et al. 2020). A recurrent neural network implemented through tensorflow (Allaire and Tang 2020) through the keras interface (Allaire and Chollet 2020) was also tested but was found to perform poorly relative to the other methods and to be extremely computationally intensive and as such was not included in the main analysis. Random forests are ensembles of regression trees, which make predictions by selecting nested splits of variables and mapping the mean level of the dependent variable at the terminal nodes of each tree. Boosted regression trees are similar to random forests, but have mechanisms in place that actively update the model to address data points that the model is struggling to fit (Elith et al. 2008). For all machine learning methods, within a model fit the model selects splits/transformations/coefficients to minimize the root mean squared error (RMSE) of predictions for data withheld from the fitting process by the algorithm.

Both the random forest and boosted regression tree models had access to the same data. These data included transformations of both the environmental and salmonid data in Table 1 and the historical return data. For the environmental and salmonid data, we calculated the cumulative mean value for each variable experienced by the cohort in question during its oceanic phase. For the historical return data, the predictors for a given cohort are all the observations of that cohort in previous years across all river systems. For example, if the model is currently predicting the 1.3 age group, the return covaraites would be the returns of 1.2 fish in all river systems in the prior year, the returns of 1.1 fish in all river systems the year before that, and so on. We also calculated the number of spawners that produced each cohort and used that as a predictor. Some data were missing for the earliest years in the data (e.g., spawning numbers for cohorts born before 1963), and these were imputed from the most recent years with available data for that river system.

We fit versions of each model separately for each age group in each major river system. We tested versions of the models that fit the age groups and river systems simultaneously, but did not use this approach, as it performed worse than the individual approach, likely due to our limited sample size. When fitting models at the level of age groups by river system, data were first split into rolling training and testing sets. For example, if the goal is to forecast returns in the year 2001, all data prior to 2001 were used as the training data, and all data post-2000 were set aside as the "testing" data. We then split each of the training sets for performance testing (e.g., all data before the year 2000 if the year 2001 is to be forecasted) into a series of analysis and assessment splits for tuning purposes. Given the time series nature of the data, we generated these analysis and assessment splits in a rolling manner. For example, for the first split, we used the first 70% of the training data as the analysis data to fit a model and the remaining 30% of the training data as an assessment split to evaluate the performance of that model. For the next split we used the first 75% of the training data for the analysis split and 25% for the assessment split, and so on. These analysis and assessment splits were used to tune nuisance parameters common to all machine learning models, for example the minimum node size of fitted trees (see the online Supplementary Table S3[1] for a complete list of tuning parameters). We fit each of our assessment splits across a grid of potential parameter values and selected the set of tuning parameters that minimized the RMSE of the predictions on the assessment splits (see computational environment available at https://github.com/DanOvando/salmon-forecast-paper/ for detailed steps in this process).

[1] Supplementary data are available with the article at https://doi.org/10.1139/cjfas-2021-0287.

Once the optimal set of tuning parameters for each training set were selected, we then fit the final model using all the training data with those tuned parameters, and used that model to predict the returns in the testing set. This process was repeated for each forecasting year. All relevant data transformations were prepared only on training/analysis splits (e.g., means and standard deviations for centering and scaling) and then applied to testing/assessment splits.

2.3. Dynamic linear models

To date methods for forecasting sockeye salmon abundance in Bristol Bay and throughout Alaska have generally relied on the relationship between the abundance of different age classes from the same cohort, or originating from the same brood year, but returning to breed in subsequent years at different ocean ages. Foundational to the predictability of sibling relationships is the assumption that the ratio of returns by age class remains stable across time. In a context of a linear model, for example, we can model returns as $\hat{R}_t^{1.3} = \alpha + \beta R_{t-1}^{1.2}$, where $\hat{R}_t^{1.3}$ is the predicted return abundance of the older (1.3) age class and $R_{t-1}^{1.2}$ is the observed abundance of the same cohort returning in the prior year after one fewer years in the ocean (i.e., age group 1.2). Under a classic sibling regression, the assumption is that the estimated parameters α and β remain constant across time. However, there are multiple conditions under which both the average return abundance of a particular age class or the ratio of abundances among age classes might change over time. For example, if the average maturation schedule (i.e., the probability that an individual will mature after 2 vs. 3 years in the ocean) changes in response to natural or anthropogenic selection, the assumption of a stationary parameter is violated. Alternatively, if average marine mortality experienced by salmon changes as a result of large-scale climate, ecosystem, or trophic shifts, this should be reflected by changes in both parameters of the regression model.

To better represent the dynamic nature of sibling or cohort relationships over time and improve predictive performance, we implement dynamic linear models (DLMs). DLMs are a class of regression models where the values of regression coefficients are permitted to evolve over time, rather than remain static (Pole et al. 1994; Petris et al. 2009). DLMs were fit to available data using a single predictor age class (one fewer year in the ocean, returning the prior year), and allowing for evolution of both the slope and intercept parameters over time, as follows:

$$\hat{R}_t^{1.3} = \alpha_t + \beta_t R_{t-1}^{1.2} + \epsilon_t$$

Both regression parameters are described by a random walk (i.e., $\alpha_t \sim \text{Normal}(\alpha_{t-1}, \sigma_\alpha^2)$ and $\beta_t \sim \text{Normal}(\beta_{t-1}, \sigma_\beta^2)$), and errors were assumed normally distributed ($\epsilon_t \sim \text{Normal}(0, \sigma_\epsilon^2)$). DLMs were implemented using the Multivariate Autoregressive State-Space Modelling (MARSS) package (version 3.10.12) in R (Holmes et al. 2012, 2020). The full time series (brood year 1963 forward) of age and river system specific abundances reconstructed by Cunningham et al. (2019) were for model fitting. For example, to predict the abundance of the 1.2 age class returning to the Wood River system in 2010, the DLM model was fit to available data 1963–2009, with the 1.1 age class in prior years assumed a priori to be the most informative sibling abundance predictor. Given the random walk structure of these dynamic linear models, it is implicitly assumed that both the average abundance and the empirical relationship between age classes for the terminal forecast year are most similar to values for those parameters observed in the recent past. This is in contrast to the machine learning and empirical dynamic modeling approaches that are more flexible in this regard.

2.4. Empirical dynamic modeling

Empirical dynamic modeling (EDM) is a nonparametric approach to characterize ecological dynamics and generate forecasts. The approach is predicated on Takens' theorem, which states that a single time series and a number of lags (dimension; E) are representative of overall system dynamics (Takens 1981; Sugihara and May 1990). Different types of EDM have identified causal relationships in ecological systems (Sugihara et al. 2012) and improved forecast skill in Fraser River sockeye salmon (Ye et al. 2015). See Munch et al. (2020) and Chang et al. (2017) for more general overviews of EDM. We used the software package rEDM (Park et al. 2021) for our analysis.

We focused on multiview embedding form of EDM (Ye and Sugihara 2016) to predict Bristol Bay sockeye returns. We predicted out-of-sample river and age-class-specific returns for 2000–2020. The idea behind multiview embedding is that there are potentially many valid reconstructions of system dynamics, and evaluating possible different combinations may improve performance. The top multiview embedding was identified with river-specific data with a maximum number of dimensions (E) of two. Multiview embedding selects models based on the within-sample fits. So to predict, say, Kvichak 2.2 returns in the year 2000, we subset data through 1999 for Kvichak 1.2, 1.3, 2.3, and 2.2, then selected the multiview embedding that had the highest within-sample predictive skill. We evaluated embeddings with maximum dimensions up to $E = 4$, although this increase did not consistently result in improved within-sample predictive skill, perhaps due to noise in the data.

We present results from multiview embedding but we also evaluated additional EDM approaches. These included multivariate simplex, multivariate sequentially locally weighted global linear maps (s-map), and composite libraries for prediction to the salmon return data. These methods require identifying the dimensionality (E) of a time series and constructing an attractor (a time series and its E-lagged coordinates). Leave-one-out prediction identifies the best E of a time series. We used E values ranging from 1 to 10, found the E-nearest neighbors (based on Euclidean distance) from the observation of interest, and calculated a predicted value by averaging the E-nearest neighbors. The best E had the highest correlation between observed and predicted values. S-maps is an extension of simplex that has the addition of a weighting parameter (theta, θ), which modifies the strength of nearest neighbor weighting ($\theta = 0$ weights nearest neighbors equally; $\theta > 0$ means stronger weighting of nearest neighbors) (Sugihara et al. 1994). Across these tests, the multiview method with $E = 2$ was the best-performed, and as such is what we present here.

2.5. Performance metrics

We did not conduct formal statistical tests of model fit or performance. Parameters of conventional statistical models might be assessed in terms of statistical significance, and models compared via some form of information criterion. However, neither the machine learning or the empirical dynamic modeling methods have formal estimates of uncertainty or likelihoods and as such do not produce measures of statistical significance around individual forecasts and cannot be compared using information criteria such as AIC (Akaike 1974) scores. Accordingly, we judged model performance by the point estimates of SRMSE produced by each model across the retrospective horizon 1990–2020 SRMSE measures the performance of each model relative to a lag(1) model, a conventional benchmark model for time series modeling (Hyndman and Koehler 2006; Ward et al. 2014).

RMSE is calculated as

$$\text{RMSE}_m = \sqrt{\frac{1}{I}\sum_{i=1}^{I}(y_i - f_{i,m})^2}$$

where *i* represents an observation of numbers of returning salmon *y* and the forecast for those numbers *f* by a given model *m*. In the manner of mean absolute scaled error (MASE; Hyndman and Koehler 2006), we scaled each model's RMSE for a given resolution by the RMSE of a lag(1) model for the same resolution (for example at the river system level).

$$\mathrm{RMSE}_{\mathrm{lag}(1)} = \sqrt{\frac{1}{I}\sum_{i=1}^{I}\left(y_i - y_{i,\mathrm{lag}(1)}\right)^2}$$

and SRMSE for model *m* is then

$$\mathrm{SRMSE}_m = \frac{\mathrm{RMSE}_m}{\mathrm{RMSE}_{\mathrm{lag}(1)}}$$

MASE is commonly used to judge the accuracy of predictions derived from time series models, since it compares the error of a given model to the error expected by a simple model in which the predictions in a given time step are equal to the observed values in the last time step (a lag(1) model). We used SRMSE instead of MASE to reflect the use of the forecast. MASE considers an error of ten to be twice as bad as an error of five. In the context of salmon forecasting, our primary objective is to avoid massively over or under estimating the pre-season forecast. SRMSE penalizes large errors more than small errors, helping select models that avoid the kinds of large errors that are most problematic for the task of managing salmon populations. Bias is also of importance in judging a forecast, and we include summaries of bias performance in the online Supplementary Materials.[1]

An SRMSE of one means that a model has predictive performance equal to that of the lag(1) model. An SRMSE greater than one indicates that a given model performs worse than the lag(1) benchmark, and an SRMSE less than one that a model performs better than the lag(1) benchmark (Hyndman and Koehler 2006). We also calculated the predictive R^2 for each of our relevant results.

2.6. Testing regime

All models were compared based upon one-step-ahead forecast skill, defined by SRMSE. Each of the evaluated models generate forecasts at the resolution of age group and river system in a given year. Forecasts for a given year are produced by a model trained on all years after 1963 and prior to the year for which a forecast is desired. This is performed in a rolling fashion, such that for example forecasts for the year 2018 are produced by a model trained on data from 1963 to 2017, the 2019 forecast by a model trained on data from 1963 to 2018, and so on. One-step-ahead performance skill was preferred over simple leave-one-out cross validation because it better aligns with the context of pre-season forecasting (i.e., data in hand through the current year are used to predict the next), and should be expected to more appropriately reflect true forecast uncertainty in the presence of periodic regime shifts in salmon production and the potential for unmodeled autocorrelation. Each method has its own ways of tuning and validating the model, but all such steps are performed using only the training data: all data for the forecast year are held out until the final prediction.

Predictive performance of candidate models was calculated by generating one-year-ahead forecasts for each target river system by age class combination, as a rolling window from the year 2000 to 2020. This method for quantifying forecast performance is most applicable to the context of this ecological forecasting problem as each candidate model is trained on data up to, but not including, the prediction year. Even though each model generates predictions at the resolution of age group and river system, we generally compare model performance at coarser resolutions (for example all age groups summed within a river system). In those cases, we first aggregated the total returns at the resolution in question (e.g., summing all observed and forecast returns across all age groups for a given river system) and then calculated the SRMSE based on those aggregated data. This allowed the best model to differ based on the scale of the predictions. For example, the model that performed best when measured at the level of age group and river system may not be the model the performed best in terms of total system returns.

2.7. Ensemble models

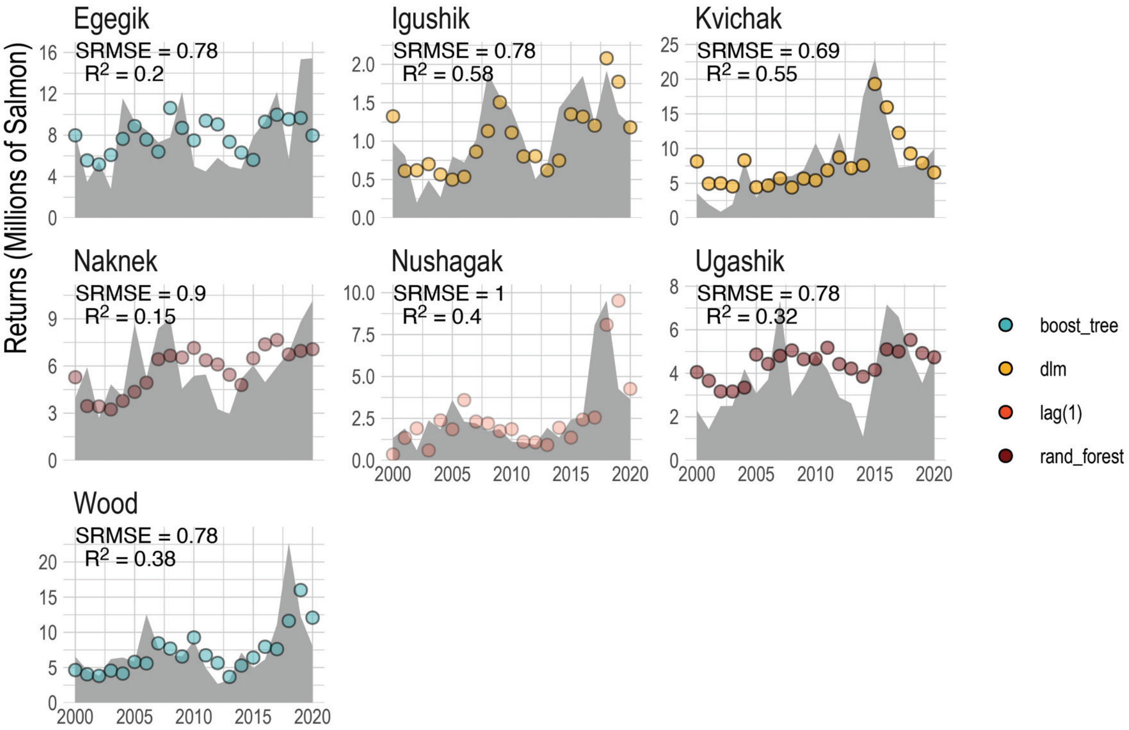
The chosen testing regime allowed us to compare the retrospective predictive power, defined by SRMSE, of individual models at a variety of spatial resolutions. However, scientists must make a decision each year as to which models to use for particular forecasts, and there is no guarantee that past model performance will predict future model performance. A substantial body of literature suggests that creating "ensemble" models that weight individual models to create a single composite prediction can outperform any one individual model (Dietterich 2000; Araújo and New 2007; Anderson et al. 2017*b*). To assess the ability of this idea to assist in annual model selection and weighting, we compared two different ensemble models: a purely statistical ensemble constructed by a random forest and a mixed-methods ensemble model published as the FRI forecasts.

The random forest ensemble model was updated each year by evaluating the performance of different models in the past and creating a prediction for the current time step based on the performance of component models (the ensemble members) in the prior time steps. For the random forest ensemble, we predicted the total returns by river system as a function of the predictions by river system and age group from each individual candidate model type. Along with the four models estimated in our results, the random forest also had access to the baseline lag(1) model, in case the data suggest that in fact this baseline model is preferable to or in some way complements the four statistical models included. A conventional ensemble might be constructed by taking an AIC-weighted mean of forecasts of each of the candidate models for a particular river system. By constructing an explicit "model-of-models" ensemble through a random forest, we allow the choice of model weighting to vary depending on the performance of different models in different river systems and time periods (Anderson et al. 2017*b*).

The FRI forecast is a mixed-methods ensemble model manually constructed by FRI scientists, which has used various models throughout the years to arrive at pre-season forecasts for each river system based on the recent performance of the component models. The FRI forecast for a specific river by age class combination was traditionally constructed by AIC-weighting across candidate linear sibling regression models. Candidate linear models predict returns of the target age class using returns of one or two younger age classes seen in prior years as predictors, but unlike the DLM model explored here, assume regression coefficients are time-invariant. These candidate predictive models were fit on two alternative time series, 1963 onward and 1980 onward, to account for broad-scale shifts in average Bristol Bay salmon population productivity following the shift in the PDO in the late 1970s. Natural-scale and log-transformed transformations were both fit for all models. Since 2013, the FRI forecast ensemble has been constructed by comparing the performance of the linear and log-linear AIC-weighted sibling models, random forest models, dynamic linear sibling models, boosted regression trees, and simple autoregressive integrated moving average (ARIMA) time series models and selecting the model with the lowest residual error in predictions for the target stock–age group across the most recent 20-year time horizon. All models in the historical FRI ensemble used only data from within a single river system, but across multiple age classes, to generate predictions (i.e., age-specific time series of Nushagak River returns were never used to forecast Wood River returns despite their spatial proximity).

The FRI ensemble forecast values were pulled from the historical pre-season forecasts as published. For the random forest ensemble, we follow a similar routine to that employed for the individual (i.e., river- and age-specific) random forest model. We compiled the pre-

Fig. 2. Observed (grey ribbons) and predicted (points) numbers of returning sockeye salmon to primary sockeye-producing river systems in Bristol Bay, Alaska. The colour of the points corresponds to the best-performing model in terms of scaled root mean squared error (SRMSE); point transparency reflects the SRMSE of the best-performing model, noted in the top left corner of each panel along with the R^2 value. [Colour online.]



season forecasts by river system and age group for each of the candidate models going back to 1991. The ensemble sought to predict the observed total returns by river system using the returns by river system and age class produced by each of the candidate models. For the years 2000 to 2020, we performed a series of rolling model fits, where individual forecasts and observed returns before the testing year was held out for training (and analysis and assessment splitting and model tuning) and then used fit the ensemble model, which was then evaluated on the testing year. The held-out one-year-ahead predictions of the ensemble model in each time step were then compiled to create the historical series of ensemble forecasts at the river system level.

3. Results

3.1. Individual model forecasts

3.1.1. River system forecasts

Management of Bristol Bay sockeye salmon operates at the river system level, with in-season fishery managers regulating allowable fishing effort on a daily basis to meet annual escapement goals for each river (Fried and Hilborn 1988; Cunningham et al. 2019). For each river system, we selected the individual model with the lowest SRMSE over the years 2000 to 2020 as the model of choice for that river system. On average the best-performing method reduced the SRMSE in pre-season run forecasts at the river system level by 15%, with a minimum improvement of 2% and a maximum of 28%, relative to the performance of the historical published pre-season FRI forecasts.

River systems varied in both the lowest SRMSE achieved and in the model that produced the best performance. At least one model was able to out-perform or equal a simple lag(1) benchmark model in each of the river systems except for the Nushagak, with the dlm model achieving a SRMSE of 0.69 at the top end in the Kvichak River system. The dlm, boost_tree, and rand_forest models were selected as the best-performing candidate in at least one river system (Fig. 2). R^2 values at the system level ranged from a low of 0.15 in the Naknek River to a high of 0.58 in the Igushik River.

3.1.2. Age group forecasts

While total river system returns are the primary metric of interest to the Bristol Bay sockeye fishery, the age composition of the returns are also important given their influence of the average size, and therefore price for each salmon harvested, and options for processed product forms. As such we also examined the ability of our tested models to generate predictions at the age group level. In retrospect, different models performed best for each of the four age groups considered, and at least one model was able to improve substantially on a lag(1) model in all age groups (Fig. 3; SRMSE < 1 in all cases, with at minimum a 20% improvement over the lag(1) model). R^2 values at the age group level ranged from a low of 0.19 in the 2.3 age group to a high of 0.71 in the 1.2 age group.

3.2. Ensemble forecasts

In theory the individual models tested here were capable of improving pre-season forecast accuracy over the years 2000–2020 when viewed retrospectively. However, scientists must make annual decisions as to which models to use and how to weigh their predictions. To approximate this process, we selected the top-performing (in terms of SRMSE) rolling ensemble model (either the FRI or the random forest model-of-models ensemble) for each of the main river systems. In 5 of the 7 evaluated river systems, the random forest ensemble produced the preferred ensemble, improving on the FRI forecast by on average 13%, with the FRI forecasts being preferable of the two ensembles in the remaining 2 river systems, outperforming the random forest ensemble by 39%. R^2 values of the best ensemble model at the river system level ranged from a low of 0.12 in the Wood age group to a high of 0.52 in the Kvichak age group (Fig. 4).

Fig. 3. Observed (grey ribbons) and predicted (points) numbers of sockeye salmon within each age group returning to Bristol Bay, Alaska. Age group refers to "years spent in fresh water.years spent in ocean". Colour corresponds to the best-performing model in terms of scaled root mean squared error (SRMSE); transparency reflects the SRMSE of the best-performing model, noted in the top left corner of each panel along with the R^2 value. [Colour online.]

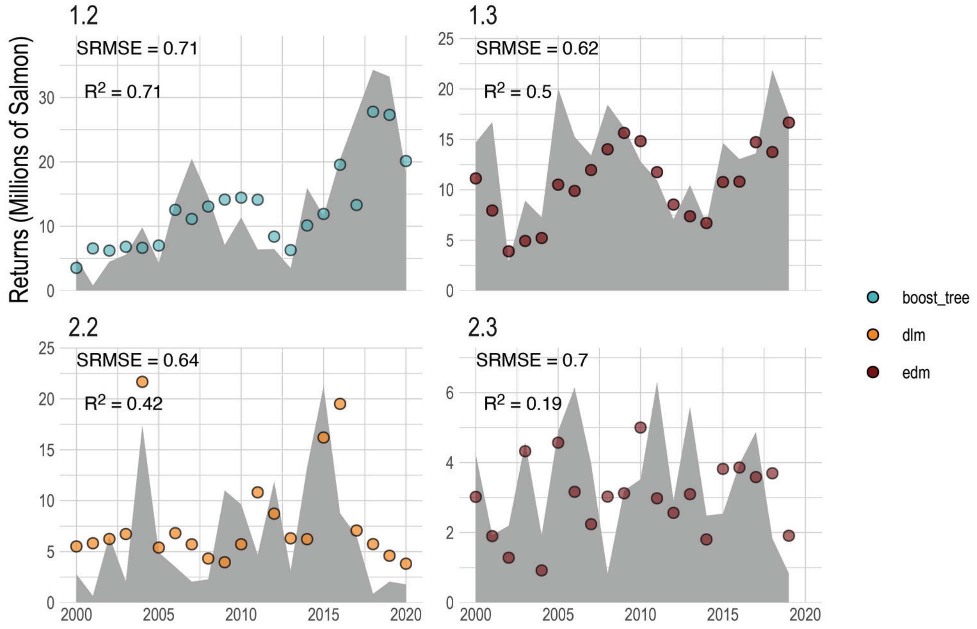
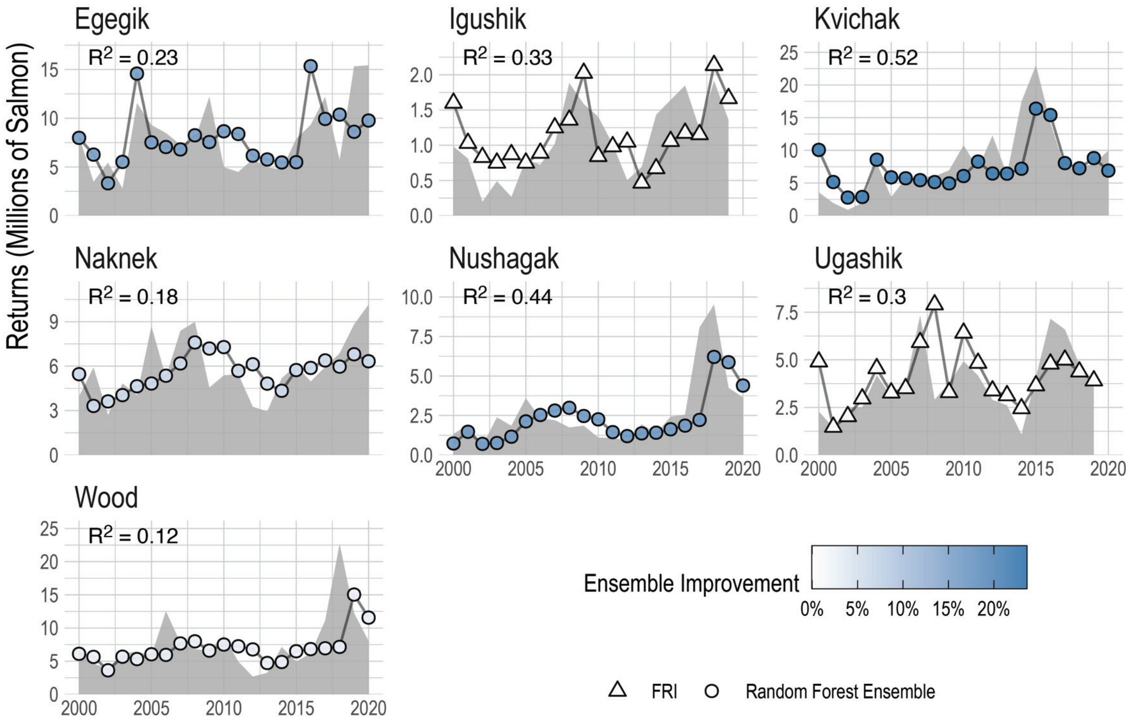


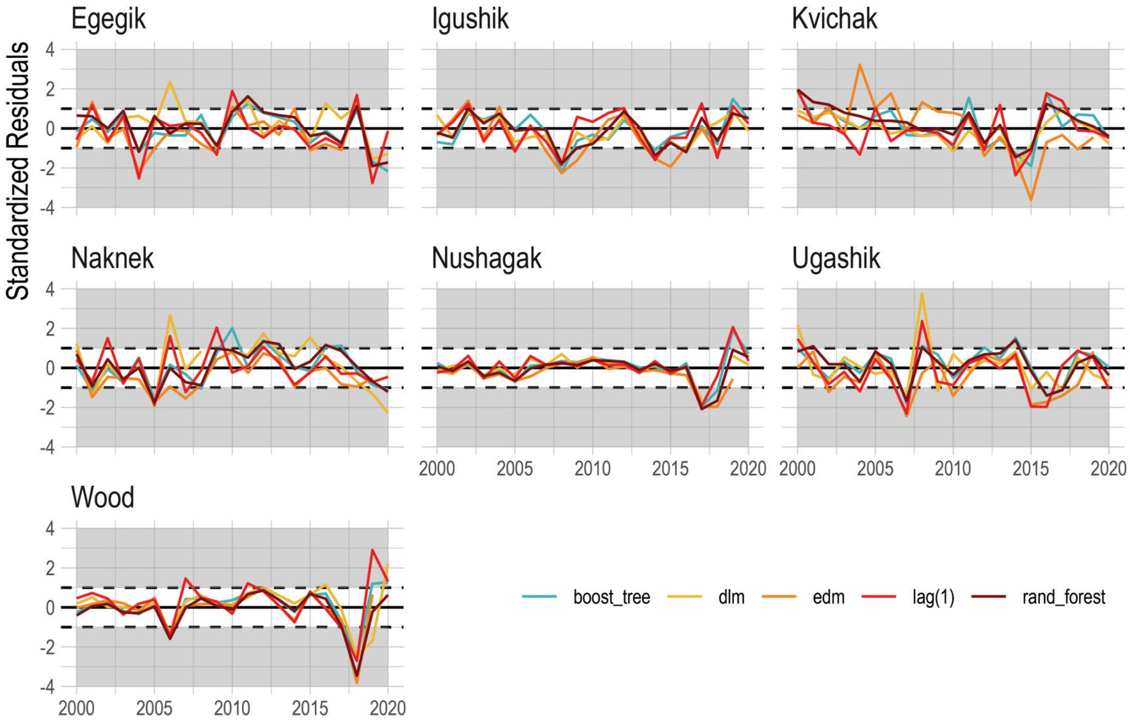
Fig. 4. Performance of candidate ensemble models. Shape of points indicates which ensemble model had the lowest scaled root mean squared error (SRMSE). FRI refers to the published forecasts by the Fisheries Research Institute. The random forest ensemble is an ensemble model constructed by random forest made out of candidate model forecasts. The forecast from the best-performing ensemble is plotted and denoted by point shape. Colour of points shows the percent improvement of the ensemble model relative to the published FRI forecast. R^2 value noted in text in top left corner of each panel. [Colour online.]



3.3. Frontiers in performance

The underlying assumption of such an ensemble strategy is that the information needed for an accurate forecast is present in the data, and the key is finding the combination of individual models that are best able to identify and leverage that information. However, no model can find information that simply is not present or succeed if it is based upon data that is subject to overwhelming observation or process error. Examining trends in the

Fig. 5. Centered and scaled annual residuals (forecast returns minus observed returns) by river system and model over time. Grey bands indicate areas more than one standard deviation from the mean residuals for a given system. Years in which all the lines are within a grey band indicate periods where all the models struggled to provide reasonable forecasts. [Colour online.]



annual residuals by model and river system shows clear patterns. In some years and river systems, all models perform similarly well, indicating that the information needed for a good forecast was present and detectable by each of the models (e.g., Nushagak River before 2015). In other years, only particular models performed well, while others struggled, indicating that information needed for a robust forecast was present but only some models were able to accurately identify the underlying relationship, highlighting the value of ensemble methods (e.g., Naknek River between 2005 and 2010). However, in other years and river systems, all models struggled, for example the Wood River in 2018 and the Kvichak River in 2014. This provides evidence that the information needed to generate a robust forecast in those years was simply not present in the data that were available at the time (Fig. 5).

Our residual analysis suggests that in some instances we simply may need to collect different data for inclusion in the forecast model if we hope to improve forecasts. For example, none of our models were able to predict the massive spike in returns to the Wood River system in recent years (Fig. 5), indicating that a signal of the process resulting in an increase in salmon survival was not among the suite of predictors explored. Conversely, all of the models performed reasonably well over most of the history of the Nushagak River, except for the most recent years. This may be explained by the relative lack of variation in historical returns to the Nushagak River prior to 2017, allowing both parametric and nonparametric models to perform equally well.

We can use the results of our most recent estimated boosted regression tree model to examine the relative importance of different included data streams in improving forecast skill (Fig. 6). While these importance scores cannot be interpreted in the same manner as regression coefficients, they give us a sense of where we might look for new data to inform prediction. Across all river systems, prior returns in that system were an important predictor (and in many systems past returns in other river systems were also a useful predictor).

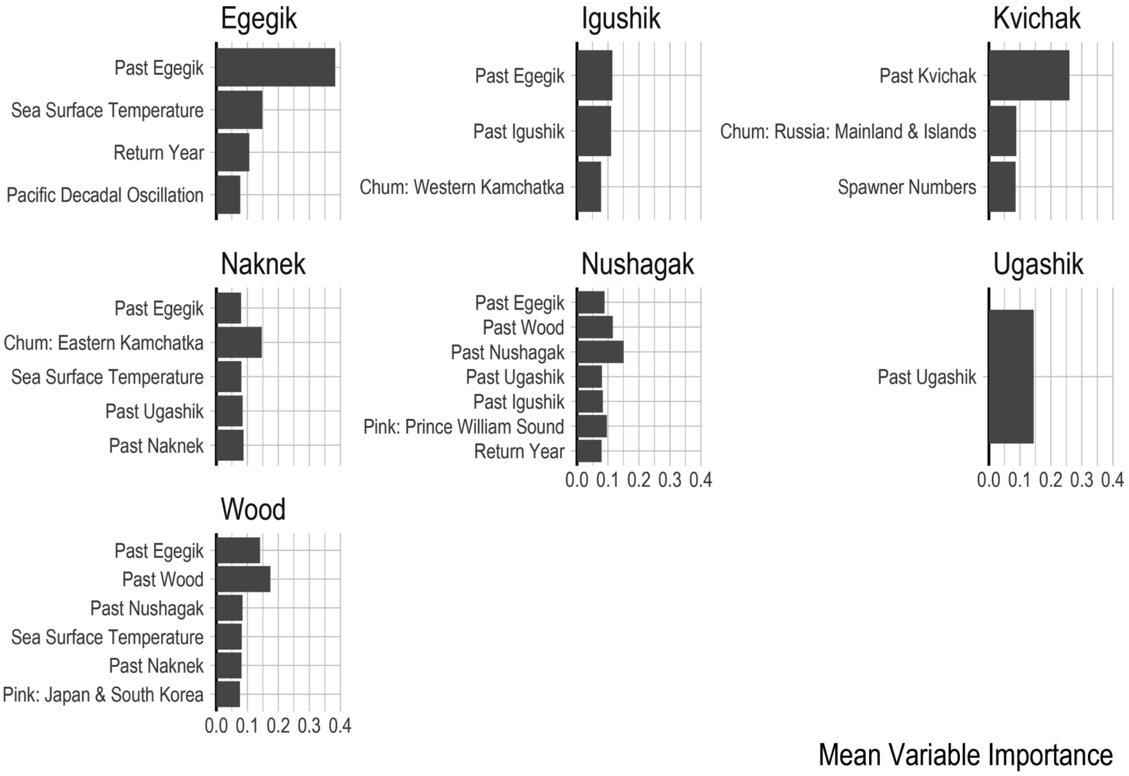
4. Discussion

While our tested methods made meaningful improvements in forecast accuracy in many cases, no one model type stood out as a clear winner, highlighting the value of multi-model inference in ecological forecasting. Viewed in retrospect individual models tested here were able to make substantial improvements in forecast accuracy (Figs. 2–3). However, the best retrospective model over the years 2000–2020 varied widely by age group and system, presenting a challenge for decision makers charged with picking which model to use for a particular forecast. Ensemble models such as the random forest ensemble (i.e., a “model-of-models”) constructed here can help users separate out the signal from the noise in historical model performance, which in this case resulted in modest improvements in forecast still in the majority of river systems evaluated in this study.

Our results cannot be interpreted as a generalized assessment of the relative strengths or weaknesses of the types of models evaluated here. Model performance is a complex function of the suitability of a given model for the task at hand, the data made available to it, and a wide range of design decisions. This is reflected in the diversity of models classified as the best performer depending on the specific question being asked of them in our analysis. Our claim is not that boosted regression trees for example are inherently best at predicting the dynamics of the Wood River system, but rather that under these particular conditions the boosted regression tree happened to work best. Attempts to classify more general “best” models for salmon forecasting would require considering a broader range of empirical and simulated states, as well as increased standardization of the design decision process.

Using multiple types of highly flexible parametric and nonparametric models can provide insight into whether historical limits to forecast skill were likely due to limitations in the information content of the available data or from simply not finding the best model to apply to the data at hand. While we were able to improve forecast skill of Bristol Bay sockeye salmon in some

Fig. 6. Mean variable importance across all river systems of variables with importance scores greater than 0.075.



instances, in particular years and systems all tested models performed poorly. These events may reflect changes in the effect of currently observed data (i.e., a violation of the assumption that the past correlation between a variable and salmon returns will apply in the future) or may be indicative of the underling effect of an unobserved variable. The former case may be resolved by simply giving the model more years on which to train or through explicit techniques for modeling outlier events in the manner of Anderson et al. (2017*a*). The latter case can only be resolved through the inclusion of new data that contains information on the previously omitted process.

That forecasts for individual river systems can be improved by treating historical returns in other river systems as predictors, as evidenced by the machine learning models, is an important finding. The historically used largely parametric salmon forecast methods have largely focused on relationships among age classes within single river systems in isolation. While perhaps not surprising given the juvenile salmon from multiple river systems enter the same area of the eastern Bering Sea during approximately the same season and likely experience similar survival conditions at ocean entry, this result suggests that sharing age-specific return abundance information among salmon stocks and river systems within Bristol Bay can inform and improve predictive performance.

In addition to the return abundance of salmon from the home and neighboring river systems, we found that oceanographic variables including mean sea surface temperature and sea surface air pressure throughout the spatial and temporal range of the oceanic phase of these salmon were informative predictors for some river systems. As reported by Connors et al. (2020), in some instances the abundance of other salmon species (chum salmon (*Oncorhynchus keta*) in western Kamchatka and northern British Columbia, pink salmon (*Oncorhynchus gorbuscha*) in Prince William Sound) proved important predictors of Bristol Bay salmon return abundance (Fig. 6). Going forward, data on freshwater conditions, interspecies competitors, and the size structure of the salmon populations may prove useful in improving forecast accuracy.

Takens' theorem suggests that the dynamics of a variable, in this case salmon returns, can be reconstructed simply by the lags of that variable, potentially obviating the need to collect the right covariates to provide an accurate forecast (Munch et al. 2020). While this may be true given sufficient sample size, temporal coverage, and lack of observation error, all of these conditions rarely hold in ecological forecasting. Case in point, the EDM models tested here were unable to accurately predict many events in the Bristol Bay sockeye return history (Fig. 5), indicating that the attractor constructed out of the lagged returns alone did not have the right information needed to forecast particular events. In these cases, collection and use of relevant covariates may help the model make improved predictions than are possible given lagged returns alone.

Traditional pre-season forecast methods for sockeye salmon returning to Bristol Bay and throughout Alaska have often assumed that relationships among age classes are static over time. However, there is increasing recognition of time-varying relationships between Alaskan salmon production and sea surface temperature (Litzow et al. 2018) and large-scale oceanographic processes including the PDO (Litzow et al. 2020*a*, 2020*b*). Given evidence for the dynamic nature of salmon–climate relationships, it should not be surprising that salmon abundance forecast relationships should also exhibit temporal variability. While not informed by environmental data and only leveraging information from a single river system, the DLM approach was found to exhibit superior performance in several river systems and the 2.2 age class. It seems reasonable that the flexible nature of the DLM approach to capture time-varying dynamics in both average abundance, and the ratio among age classes permits an indirect accounting for the dynamic salmon–environment processes that are increasingly recognized.

Forecast methods historically employed by the FRI involved evaluation of a suite of alternative forecast models in each year, and selection of a preferred model and data time series on which to train the model (i.e., 1963 onward or after the observed shift in the PDO in 1980), for each salmon stock by age class combination based on forecast bias and precision over the recent 20-year period. While the FRI forecast has always been primarily based on the relationship between the abundance of age classes from the same cohort among successive years, the suite of forecast models explored as part of the FRI forecast has evolved over time. In recent years new methods have been added to the forecast model suite, including autoregressive integrated moving average (ARIMA) models, boosted regression trees, Bayesian indicator variable methods, and dynamic linear models. The manual model selection process at the heart of the FRI ensemble approach has proven effective over time at identifying candidate models for forecast groups (stock-by-age) that best leverage patterns within individual time series (i.e., ARIMA, DLM), weighting candidate predictor age classes (i.e., Bayesian indicator variable methods) and nonlinear relationships between the return abundance of age classes for a stock in prior years (i.e., boosted regression trees). However, despite the observed value in comparing performance of alternative forecast model types inherent in the FRI forecasting approach, significant forecast errors have occurred. The range of models historically used by the FRI only leveraged data for sibling age classes from the same river system, and the potential for human error in the manual model selection process cannot be overlooked. We demonstrate here that statistical ensemble approaches, such as the random forest ensemble, present a viable complement to more "human-based" ensemble approaches.

However, the relative performance of the random forest ensemble model to the historical FRI forecast cannot be construed as a broader result about the relative performance of "statistical" vs. "manual" models. The random forest ensemble utilized a range of models and datasets that were not all available to the historical FRI forecasts, and it is entirely possible that had the FRI had access to those same models in the past, they would have produced more similar results to the random forest ensemble. What our results do show is that addition of new model types and data does provide potential to improve on the historical FRI methods such as they were.

We demonstrate here how parametric and nonparametric modeling approaches can provide improvements in ecological forecasting. Ward et al. (2014) also explored the use of models similar to those used here in the context of ecological forecasting of time series data from natural populations. They, however, found that the sorts of tools explored here generally performed worse than simple autoregressive models while being substantially more computationally intensive. In contrast we found that our lag(1) benchmark model was outperformed or equaled by one or more of our models across nearly every resolution we evaluated. What might explain this difference? First, Ward et al. (2014) specifically designed their study around making predictions of future population size solely based on historical population size, while the forecast methods we explore here were informed by the abundance of multiple salmon age classes or stocks and in some cases by environmental conditions and the abundance of other salmon species. Second, Ward et al. (2014) did find that more complex models such as random forests and neural networks performed well for some salmon populations, particularly those characterized by regular cyclic behavior. Our results are broadly consistent then with the findings of Ward et al. (2014) in this respect.

However, Ward et al. (2014) did find that more complex models performed poorly relative to their baseline random walk model in salmon stocks that exhibited less cyclic behavior. In contrast, we found near universal improvements over our baseline model to some degree across all river systems and age groups, including those that either do not seem to exhibit cyclic behavior or have experienced a break from past cycles in recent years (Fig. 1, Supplementary Fig. S9[1]). The machine learning methods explored here have access to much more data than the historical returns alone though, including environmental conditions and abundance of other salmonids. In addition, the machine learning methods were able to leverage correlations in returns across multiple age groups and river systems (Fig. 6). While we have access to over 50 years of data, longer than some of the series reported in Ward et al. (2014), our sample sizes are still minute compared to the sample sizes in most applications of machine learning methods, indicating that these methods can still be used with the relatively small sample sizes often encountered in forecasting the population dynamics of harvested species. These differences of long time series, use of cross-system and age group correlations, and inclusion of environmental covariates may explain the ability of the more complex models tested here to outperform benchmark lag(1) models even in systems without an obvious cyclical pattern.

One of the primary advantages of parametric statistical approaches that make explicit assumptions about data-generating processes, and by extension error structure, is that they provide estimates of the degree of uncertainty associated with a model coefficient or a prediction. Nonparametric machine learning methods are powerful in that they are able to learn complex predictive correlations within data, but a key limitation of these methods is that they generally do not provide estimates of uncertainty for their predictions. We cannot therefore provide 95% confidence intervals or other conventional metrics of uncertainty around many of our forecasts, though the SRMSE values provide an estimate of the historical error in the forecasts of each model. The distribution of Pearson's residuals for each of the models do not exhibit any clear differences across models (see Supplementary Materials[1]).

5. Conclusion

The field of ecology is generally concerned with developing theories and evidence for why ecosystems are structured and behave the ways they do. This pursuit of heuristic understanding can lead to construction of interpretable models that provide insight about system dynamics, but limited predictive power. However, for specific application in areas such as pre-season salmon abundance forecasts, the objective is solely to obtain accurate and precise predictions one year into the future. We designed and optimized our models solely around predictive power, and while some methods such as empirical dynamic modeling and dynamic linear models can provide both insight and predictive skill, the machine learning methods tested here (boosted regression trees and random forests) are focused on prediction alone, with limited scope to improve ecological insight. In the case of natural resources management that often depends on making decisions today based on predictions about the future, prediction-focused methods such as those presented here can present substantial opportunity. Here we show that incorporating multiple predictive models into a statistical ensemble was able to provide some meaningful improvements in the pre-season forecast accuracy of Bristol Bay sockeye salmon.

Accurate forecasts are a crucial part of natural resource management, a task made increasingly challenging by climate change. Our gains in forecast accuracy for the economically and ecologically critical Bristol Bay sockeye salmon fishery demonstrate the ability of parametric and nonparametric models to make meaningful improvements in short-term predictive ability for the abundance of natural populations faced with a rapidly changing environment. By combining multiple model types, we are able to identify likely frontiers in forecast performance given currently available data. However, even for this relatively robust dataset, we were fundamentally unable to predict the returns of particular river systems and age classes in certain years. The collective failure of multiple methods in specific time steps and locations helps clarify instances in which the only likely path to meaningful forecast improvement is collection or incorporation of additional data, while also highlighting the potentially irreducible impact of observation error on the limits of forecast performance. It is critical that we allocate resources to both the advancement of predictive modeling methods in ecology and to the hard work of collecting the data from the natural world that are the foundation of any successful forecasting efforts.

Contributors' statement

D.O., C.C., and P.K. conducted the analyses. All authors contributed to the development of the manuscript.

Funding statement

Funding for this study was provided by the Bristol Bay Regional Seafood Development Association, by the Bristol Bay Seafood Processors, and by Douglas and Joyce McCallum.

Data availability statement

All data, code, and package dependencies needed to fully reproduce our results are publicly available at www.github.com/danovando/salmon-forecast-paper.

Acknowledgements

We thank two anonymous reviewers for their helpful suggestions and edits.

References

Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. **19**(6): 716–723. doi:10.1109/TAC.1974.1100705.

Allaire, J., and Chollet, F. 2020. Keras: R interface to 'keras'. Available from https://CRAN.R-project.org/package=keras.

Allaire, J., and Tang, Y. 2020. Tensorflow: R interface to 'TensorFlow'. Available from https://CRAN.R-project.org/package=tensorflow.

Anderson, S.C., Branch, T.A., Cooper, A.B., and Dulvy, N.K. 2017*a*. Black-swan events in animal populations. Proc. Natl. Acad. Sci. U.S.A. **114**(12): 3252–3257. doi:10.1073/pnas.1611525114. PMID:28270622.

Anderson, S.C., Cooper, A.B., Jensen, O.P., Minto, C., Thorson, J.T., Walsh, J.C., et al. 2017*b*. Improving estimates of population status and trend with super-ensemble models. Fish Fish. **18**(4): 732–741. doi:10.1111/faf.12200.

Araújo, M., and New, M. 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. **22**(1): 42–47. doi:10.1016/j.tree.2006.09.010. PMID: 17011070.

Beyan, C., and Browman, H.I. 2020. Setting the stage for the machine intelligence era in marine science. ICES J. Mar. Sci. **77**(4): 1267–1273. doi:10.1093/icesjms/fsaa084.

Breiman, L. 2001. Random forests. Mach. Learn. **45**(1): 5–32. doi:10.1023/A:1010933404324.

Chamberlain, S. 2019. Rerddap: General purpose client for 'ERDDAP' servers. Available from https://CRAN.R-project.org/package=rerddap.

Chang, C.-W., Ushio, M., and Hsieh, C. 2017. Empirical dynamic modeling for beginners. Ecol. Res. **32**(6): 785–796. doi:10.1007/s11284-017-1469-9.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., et al. 2020. Xgboost: Extreme gradient boosting. Available from https://CRAN.R-project.org/package=xgboost.

Connors, B., Malick, M.J., Ruggerone, G.T., Rand, P., Adkison, M., Irvine, J.R., et al. 2020. Climate and competition influence sockeye salmon population dynamics across the Northeast Pacific Ocean. Can. J. Fish. Aquat. Sci. **77**(6): 943–949. doi:10.1139/cjfas-2019-0422.

Cunningham, C.J., Anderson, C.M., Wang, J.Y.-L., Link, M., and Hilborn, R. 2019. A management strategy evaluation of the commercial sockeye salmon fishery in Bristol Bay, Alaska. Can. J. Fish. Aquat. Sci. **76**(9): 1669–1683. doi:10.1139/cjfas-2018-0133.

DFO. 2018. Pre-season run size forecasts for Fraser River Sockeye (*Oncorhynchus nerka*) salmon in 2018. DFO Can. Sci. Advis. Sec. Sci. Res.

Dietterich, T.G. 2000. Ensemble methods in machine learning. *In* Multiple classifier systems. Springer, Berlin, Heidelberg. pp. 1–15. doi:10.1007/3-540-45014-9_1.

Efron, B., and Hastie, T. 2016. Computer age statistical inference: Algorithms, evidence, and data science. Cambridge University Press, New York.

Elith, J., Leathwick, J.R., and Hastie, T. 2008. A working guide to boosted regression trees. J. Anim. Ecol. **77**(4): 802–813. doi:10.1111/j.1365-2656.2008.01390.x. PMID:18397250.

Fried, S.M., and Hilborn, R. 1988. Inseason forecasting of Bristol Bay, Alaska, sockeye salmon (*Oncorhynchus nerka*) abundance using Bayesian Probability Theory. Can. J. Fish. Aquat. Sci. **45**(5): 850–855. doi:10.1139/f88-103.

Holmes, E.E., Ward, E.J., and Wills, K. 2012. MARSS: Multivariate autoregressive state-space models for analyzing time-series data. R J. **4**(1): 11–19. doi:10.32614/RJ-2012-002.

Holmes, E.E., Ward, E.J., Scheuerell, M.D., and Wills, K. 2020. MARSS: Multivariate autoregressive state-space modeling. Available from https://CRAN.R-project.org/package=MARSS.

Holt, C.A., and Peterman, R.M. 2004. Long-term trends in age-specific recruitment of sockeye salmon (*Oncorhynchus nerka*) in a changing environment. Can. J. Fish. Aquat. Sci. **61**(12): 2455–2470. doi:10.1139/f04-193.

Hyndman, R.J., and Koehler, A.B. 2006. Another look at measures of forecast accuracy. Int. J. Forecasting, **22**(4): 679–688. doi:10.1016/j.ijforecast.2006.03.001.

Knapp, G., Mouhcine, G., and Goldsmith, S. 2013. The economic importance of the Bristol Bay salmon industry. Institute of Social and Economic Research, University of Alaska Anchorage.

Litzow, M.A., Ciannelli, L., Puerta, P., Wettstein, J.J., Rykaczewski, R.R., and Opiekun, M. 2018. Non-stationary climatesalmon relationships in the Gulf of Alaska. Proc. R. Soc. B Biol. Sci. **285**(1890): 20181855. doi:10.1098/rspb.2018.1855. PMID:30404879.

Litzow, M.A., Hunsicker, M.E., Bond, N.A., Burke, B.J., Cunningham, C.J., Gosselin, J.L., et al. 2020*a*. The changing physical and ecological meanings of North Pacific Ocean climate indices. Proc. Natl. Acad. Sci. U.S.A. **117**(14): 7665–7671. doi:10.1073/pnas.1921266117. PMID:32205439.

Litzow, M.A., Malick, M.J., Bond, N.A., Cunningham, C.J., Gosselin, J.L., and Ward, E.J. 2020*b*. Quantifying a novel climate through changes in PDO–climate and PDO–salmon relationships. Geophys. Res. Lett. **47**(16): e2020GL087972. doi:10.1029/2020GL087972.

Malde, K., Handegard, N.O., Eikvil, L., and Salberg, A.-B. 2020. Machine intelligence and the data-driven future of marine science. ICES J. Mar. Sci. **77**(4): 1274–1285. doi:10.1093/icesjms/fsz057.

McDowell Group. 2018. Sockeye Market Analysis, Spring 2017. (Online.) Available from https://www.bbrsda.com/s/Spring-2017-Sockeye-Market-Report-FINAL.pdf.

Munch, S.B., Brias, A., Sugihara, G., and Rogers, T.L. 2020. Frequently asked questions about nonlinear dynamics and empirical dynamic modelling. ICES J. Mar. Sci. **77**(4): 1463–1479. doi:10.1093/icesjms/fsz209.

Naiman, R.J., Bilby, R.E., Schindler, D.E., and Helfield, J.M. 2002. Pacific salmon, nutrients, and the dynamics of freshwater and riparian ecosystems. Ecosystems, **5**(4): 399–417. doi:10.1007/s10021-001-0083-3.

Park, J., Smith, C., Sugihara, G., and Deyle, E., 2021. rEDM: Empirical dynamic modeling ('EDM'). Available from https://CRAN.R-project.org/package=rEDM.

Peters, D.P.C., Havstad, K.M., Cushing, J., Tweedie, C., Fuentes, O., and Villanueva-Rosales, N. 2014. Harnessing the power of big data: Infusing the scientific method with machine learning to transform ecology. Ecosphere, **5**(6): art67. doi:10.1890/ES13-00359.1.

Petris, G., Petrone, S., and Campagnoli, P. 2009. Dynamic linear models With R. Springer, Dordrecht, New York.

Pole, A., West, M., and Harrison, J. 1994. Applied Bayesian forecasting and time series analysis. Chapman and Hall, New York.

R Core Team. 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/.

Ruggerone, G.T., and Irvine, J.R. 2018. Numbers and biomass of natural- and hatchery-origin pink salmon, chum salmon, and sockeye salmon in the north Pacific Ocean, 1925. Mar. Coast. Fish. **10**(2): 152–168. doi:10.1002/mcf2.10023.

Schindler, D.E., Scheuerell, M.D., Moore, J.W., Gende, S.M., Francis, T.B., and Palen, W.J. 2003. Pacific salmon and the ecology of coastal ecosystems. Front. Ecol. Environ. **1**(1): 31–37. doi:10.1890/1540-9295(2003)001[0031:PSATEO]2.0.CO;2.

Steiner, E.M., Criddle, K.R., and Adkison, M.D. 2011. Balancing Biological sustainability with the economic needs of Alaska's sockeye salmon fisheries. N. Am. J. Fish. Manage. **31**(3): 431–444. doi:10.1080/02755947.2011.588917.

Sugihara, G., and May, R.M. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature, **344**(6268): 734–741. doi:10.1038/344734a0. PMID:2330029.

Sugihara, G., Grenfell, B.T., May, R.M., and Tong, H. 1994. Nonlinear forecasting for the classification of natural time series. Philos. Trans. R. Soc. Ser. A Phys. Eng. Sci. **348**(1688): 477–495. doi:10.1098/rsta.1994.0106.

Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., and Munch, S. 2012. Detecting Causality in Complex Ecosystems. Science, **338**(6106): 496–500. doi:10.1126/science.1227079. PMID:22997134.

Takens, F. 1981. Detecting strange attractors in turbulence. *In* Dynamical Systems and Turbulence, Warwick 1980. *Edited by* D. Rand and L.-S. Young. Springer, Berlin, Heidelberg. pp. 366–381. doi:10.1007/BFb0091924.

Vélez-Espino, L.A., Parken, C.K., Clemons, E.R., Peterson, R., Ryding, K., Folkes, M., and Pestal, G. 2019. ForecastR: Tools to automate forecasting procedures for salmonid terminal run and escapement. *In* Final Report submitted to the Southern Boundary Restoration and Enhancement Fund, Pacific Salmon Commission. Pacific Salmon Commission, Vancouver, B.C.

Ward, E.J., Holmes, E.E., Thorson, J.T., and Collen, B. 2014. Complexity is costly: A meta-analysis of parametric and non-parametric methods for short-term population forecasting. Oikos, **123**(6): 652–661. doi:10.1111/j.1600-0706.2014.00916.x.

Wright, M.N., and Ziegler, A. 2017. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Softw. **77**(1): 1–17. doi:10.18637/jss.v077.i01.

Ye, H., and Sugihara, G. 2016. Information leverage in interconnected ecosystems: Overcoming the curse of dimensionality. Science, **353**: 922–925. doi:10.1126/science.aag0863. PMID:27563095.

Ye, H., Beamish, R.J., Glaser, S.M., Grant, S.C.H., Hsieh, C., Richards, L.J., et al. 2015. Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. Proc. Natl. Acad. Sci. U.S.A. **112**(13): E1569–E1576. doi:10.1073/pnas.1417063112. PMID:25733874.

Ye, H., Clark, A., Deyle, E., and Munch, S. 2020. rEDM: Applications of empirical dynamic modeling from time series. (Online.) Available from https://ha0ye.github.io/rEDM/index.html.

Yi, X., Hawkshaw, M., Patterson, D., Hourston, R., and Chandler, P. 2019. How Fraser River Sockeye Salmon recruitment was affected by Climate Change: A model study. *In* State of the physical, biological and selected fishery resources of Pacific Canadian marine ecosystems in 2018. *Edited by* J.L. Boldt, J. Leonard, and P.C. Chandler. Fisheries and Oceans Canada, Sidney, B.C. pp. 214–217.