©Copyright 2023

Emily Sellinger

Characterizing and forecasting fish recruitment in a changing world

Emily Sellinger

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2023

Committee:

André Punt, Chair

Cody Szuwalski

Richard Methot

Mark Scheuerell

Program Authorized to Offer Degree:

Quantitative Ecology and Resource Management

University of Washington

Abstract

Characterizing and forecasting fish recruitment in a changing world

Emily Sellinger

Chair of the Supervisory Committee:
André Punt
Quantitative Ecology and Resource Management

Recruitment, the entry of young fish into a stock, is an important process in population dynamics models, which form the basis for many stock assessments. As such, recruitment has been a focus of research among fishery scientists for over a hundred years. Recruitment is often assumed to be a function of the spawning biomass of a stock. However, quantifying this relationship is challenging. The first chapter of this thesis examines the prevalence of a detectable influence of spawning biomass on recruitment for the stocks included in the RAM Legacy Stock Assessment database. We found most stocks (57%) did not have a detectable relationship between spawning biomass and recruitment over the observed ranges of spawning biomass. Environmental factors appear to play a larger role in recruitment variation. Furthermore, many of the stocks we examined had evidence of a regime shift (46%). Our results highlight the need to develop effective methods to model and forecast large variations in recruitment over time. The second chapter of this thesis evaluates the forecast performance of six methods on groundfish recruitment. The methods varied in the level of parameterization and operational use. We found that forecast performance depends on the time period, the performance metric, and the characteristics of the time series. Our

results indicate an advantage in using non-parametric forecast methods, especially for mid-term projections.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

LIST OF TABLES

Table Number Page

ACKNOWLEDGMENTS

This thesis could not have been completed without the help of a fair number of people. First, I would like to thank my family and my partner. They all have provided me with so much love and support throughout my time in grad school. They take the time to cheer me up when I feel down and listen to me vent when I feel stuck. Staying motivated would have been impossible without them. Next, I would like to thank my friends and colleagues in the SAFS and QERM departments, and, in particular, the Punt Lab. I am constantly inspired by my lab mates. They are incredible scientists, and even more incredible people. I am deeply honored to have had the opportunity to get to know and work alongside them. I would also like to thank my committee, - Cody, Rick, and Mark. I deeply appreciate the time they have spent over the last few years providing me with feedback and guidance. They all have been an invaluable resource. Finally, I would like to thank my advisor, Andre. It should be physically impossible for someone as busy as Andre to make time for his graduate students, but he always does. I have learned so much from him over the years. His passion for the field is infectious. It reminds me of what interested me in quantitative ecology in the first place.

DEDICATION

This thesis is dedicated to my family and friends. Without their support, I would be lost.

INTRODUCTION

The National Oceanic and Atmospheric Administration (NOAA) was mandated to create national standards to which fisheries in the United States must comply following the passage of the Magnuson-Stevens act in 1976 (PFMC 2020). National Standard 1 specifies that fisheries should be managed with the goal of achieving an optimum yield (NS 1 2021). Stock assessments are used by fisheries scientists and regional fishery management councils to determine the optimum yield of a stock, while preventing overfishing (NS 1 2021). Modern stock assessments seek to estimate the population processes governing a stock's dynamics: natural mortality, growth, recruitment, and selectivity (Punt, 2023). Recruitment is defined as the number of fish entering a fishery or the population (the latter often age-0 fish). Estimating recruitment is a necessary component of age- and size-structured assessment models. However, recruits are often not observed in catches or surveys. Furthermore, the amount of recruitment can be influenced by a variety of biological and environmental factors. The effects of these two types of factors have made providing accurate forecasts of recruitment elusive (Subbey et al., 2014). Despite the challenges associated with estimating and forecasting recruitment, it continues to be an integral part of stock biomass projections (Kell et al., 2016). This thesis aims to contribute to the vast body of research on recruitment.

Classically, recruitment was assumed to be a function of spawning biomass. Parametric stock-recruitment relationships such as the Beverton and Holt (1957) and Ricker (1954) models were used to estimate and predict recruitment. These models assume that recruitment is a function of the spawning biomass. However, the primary driver in the stock-recruitment relationship has been a source of debate. Myers and Barrowman (1996) suggested reduced spawning biomass was associated with reduced recruitment while Gilbert (1997) argued that instead of spawning biomass primarily driving recruitment, environmental factors played a larger role than spawning biomass in determining recruitment. For example, environmental conditions could lead to lower recruitment, which results in lower spawning biomasses. More recent analyses of stocks in the RAM Legacy Stock database have found spawning biomass has little detectable impact on recruitment over the observed range of spawning biomass for many stocks (Szuwalski et al., 2015), lending credence to Gilbert (1997) hypothesis.

Many studies have sought to incorporate environmental covariates into the stock-recruitment relationship, with varied success (Basson, 1999; Deyle et al., 2018). Haltuch et al. (2019) reviewed studies that incorporated environmental variables into recruitment forecasting. The authors found environmental drivers are most likely to be useful for species that recruit at younger ages and have bottlenecks on survival during the juvenile stages (Haltuch et al., 2019). A simulation study conducted by De Oliveira and Butterworth (2005) found that an environmental index needs to explain more than 50% of the variation in recruitment for there to be measurable differences in the performances of alternative management procedures. Environmental effects on recruitment have also been modeled indirectly, using time-varying parameters (Xu et al., 2018) or hidden Markov models (Munch and Kottas, 2009; Tang et al., 2021). Stock-recruitment relationships that account for temporal change in expected recruitment are attractive, given the presence of regime changes in stocks (Szuwalski et al., 2015). The first chapter of this thesis seeks to contribute to the literature on the evidence for stock-recruitment relationships by characterizing the robustness of the findings of Szuwalski et al. (2015) to the addition of new data and the use of alternate analysis methods. Specifically, it addresses the hypotheses that recruitment is primarily driven by changes in spawning biomass and is not time-varying. Time series from the RAM Legacy Stock Assessment database were extracted to characterize the form of the stock-recruitment relationship. Two methods are used to determine the nature of the relationship between spawning biomass and recruitment over the observed range of spawning biomass, 1) Spearman's correlation and 2) detrended cross-correlation analysis (DCCA). The presence and location of recruitment regime shifts are identified using three methods, 1) the pruned exact linear time (PELT) algorithm, 2) Bayesian change point detection, and 3) multivariate autoregressive state-space (MARSS) models. Finally, we compare simulated and observed time series data to determine what characteristics best aid detection of a stock-recruitment relationship.

Due to the difficulty in calibrating stock-recruitment relationships for many stocks, identifying suitable alternatives has been an area of interest. Forecast methods differ in the amount and quality of data needed, the assumptions made by the models, and the ease of implementation. Furthermore, different methods can have variable performance in terms of short- and long-term predictions (Van Beveren et al., 2021).

The challenges of parameterizing stock-recruitment relationships have led to increased testing of semi- and non-parametric methods. Semi- and non-parametric methods for forecasting recruitment are based on a variety of modeling approaches. For example, Evans and Rice (1988) used a probability distribution to predict recruitment in place of a functional relationship with spawning biomass, while Cardinale and Arrhenius (2000) used generalized additive models (GAMs) to analyze the recruitment of Baltic cod, and Sun et al. (2009) evaluated artificial neural networks for Norwegian spring-spawning herring. Empirical dynamic modeling (EDM), described by Sugihara and May (1990), is also a popular non-parametric prediction technique. EDM accommodates nonequilibrium and nonlinear dynamics (Ye et al., 2015). Simplex projection and S-map, two EDM techniques were used to forecast sockeye salmon recruitment in the Fraser River (Ye et al., 2015). The EDM forecasts outperformed a standard Ricker model in eight of the nine salmon stocks examined. An analysis conducted by Munch et al. (2018) on 185 stocks in the RAM Legacy Stock Assessment database found that EDM forecasts had lower prediction error than classic stock-recruitment relationships. The EDM method used by Munch et al. (2018) was able to take advantage of recent observations, as opposed to the single spawning biomass input required for classic stock-recruitment functions. Also, fishery scientists and managers are interested in methods that are robust to changing environments, which are expected under climate change (Subbey et al., 2014). Identifying forecasting techniques that perform well within systems that undergo regime shifts may allow for more sustainable harvests.

The second chapter of this thesis therefore evaluates the performances of several common and novel forecasting techniques in the short- and long-term, with a focus on data-rich US West Coast and North Pacific groundfish stocks. Specifically, we test a suite of forecast methods that vary in operational use across regions and the level of parameterization: 1) Beverton-Holt stock-recruitment relationship, 2) an autoregressive-1 function, 3) mean recruitment, 4) a hidden Markov sampling procedure, and 5) simplex projection, 6) a PELT changepoint sampling procedure. We evaluate each forecast method using two different metrics and compare performance to several characteristics of the stocks' time series. Finally, we suggest appropriate recruitment forecast methods based on the region of the stock, the period of the desired forecast, and the characteristics of the stock's time series.

Chapter 1

THE ROBUSTNESS OF OUR ASSUMPTIONS ABOUT RECRUITMENT: A RE-EXAMINATION OF MARINE RECRUITMENT DYNAMICS WITH ADDITIONAL DATA AND NOVEL METHODS

1.1 Abstract

Recruitment is the number of young fish entering a population and is an essential process in age-structured stock assessment models. Many of the models on which stock assessments are based assume some level of influence of spawning biomass on recruitment, and estimation of this functional relationship has been a focus of substantial research. However, recruitment is also influenced by environmental fluctuations that induce autocorrelated patterns, trends, and shifts in deviations from the stock-recruitment relationship. Consequently, previous meta-analyses of stock-recruitment relationships have had trouble calibrating the relationship. We revisited the findings of Szuwalski et al. (2015) to determine if their results are robust to the addition of data to the RAM Legacy Stock Assessment database and choice of analysis method (the PELT algorithm, MARSS models, and Bayesian change point detection). Since the publication of Szuwalski et al. (2015), the number of stocks in the database has doubled. We determined the primary influence of spawning biomass on recruitment and examined the recruitment time series for regime shifts for 432 stocks. Our results indicated that 57% of stocks did not have a significant correlation between spawning biomass and recruitment over the observed biomasses. Environmental conditions played a larger role in recruitment variation than spawning biomass. The presence, location, and number of regime shifts in recruitment time series was highly dependent on the detection method, with Bayesian change point detection identifying the fewest regime shifts. Despite the sensitivity

to the detection method, 46% of stocks without a significant correlation between spawning biomass and recruitment are estimated to have experienced at least one regime shift as determined by the PELT algorithm Our analyses suggest that effective methods for modeling and forecasting large variations in recruitment over time are needed, particularly given that climate change is predicted to impact the frequency and magnitude of regime shifts.

1.2 Introduction

The relationship between recruitment, the number of fish entering a fish stock (often age-0 fish), and spawning biomass has been of interest to fishery scientists and managers for decades (Subbey et al., 2014). Assessments for many stocks in the United States are based on age-structured population dynamics models, which often use spawning biomass as a proxy for reproductive potential (Kell et al., 2016). The stock-recruitment relationship has a direct impact on the estimation of maximum sustainable yield (MSY). Understanding the relationship, if any, between spawning biomass and recruitment is necessary for the estimation of reference points and hence management advice.

The primary driver of recruitment strength has been a source of debate. Myers and Barrowman (1996) suggested reduced spawning biomass was associated with reduced recruitment. Gilbert (1997) argued that instead of spawning biomass driving recruitment, reductions in recruitment resulted in lower spawning biomasses. Since Gilbert (1997)'s publication, multiple studies have found that spawning biomass explains the variation in recruitment for only some stocks (Cury et al., 2014; Munch et al., 2018; Szuwalski et al., 2015). Szuwalski et al. (2015) examined common assumptions about recruitment drivers using the RAM Legacy Stock Assessment database (Ricard et al., 2012). Out of the 224 stocks examined, 61% had no statistically significant positive correlation (p-value 0.05) between spawning biomass and recruitment. Instead, environmental factors likely had a larger effect on recruitment than spawning biomass (Szuwalski et al., 2015). The use of maximum sustainable yield (MSY) reference points is questionable without a detectable stock-recruitment relationship, as MSY estimation is dependent on the stock-recruitment relationship. This influence of the environment on recruitment is pressing given 71% of the stocks without a detectable stock-recruitment relationship in Szuwalski et al. (2015) exhibited changes in the average recruitment, or regime shifts. Regimes, or shifts in the ecosystem to an alternate state of nature, result in a change to the average recruitment (Collie et al., 2004; Scheffer et al., 2001; Szuwalski et al., 2015). Regime changes are thought to be linked to variation in temperature, salinity, currents, or upwelling (Jiao, 2009).

Current approaches to modeling recruitment often assume no large, sudden temporal variation, as would be observed under regime shifts (Perälä et al., 2017). However, climate change is predicted to increase the frequency of regime shifts (Hughes et al., 2013), which could lead to biased projections for stock biomass (Perälä et al., 2017). Many methods have been used to detect regime shifts post-hoc (e.g., Andersen et al., 2009; Crépin et al., 2012). These methods are usually tested in isolation and often with a hypothesized location of a past regime shift. Investigating how these methods compare to each other, specifically when used naively, can further our understanding of the ability to detect future regime changes.

This study characterizes the robustness of the findings in Szuwalski et al. (2015) to the addition of new data and the use of alternate analysis methods. Time series from the RAM Legacy Stock Assessment database were extracted to characterize the form of the stock-recruitment relationship (driven by spawning biomass, driven by the environment, monotonically increasing or dome-shaped, etc.) The presence and location of regime shifts were identified using three methods. Finally, we compared simulated and observed time series data to determine what characteristics best aid detection of a stock-recruitment relationship.

1.3 Methods

1.3.1 Data

Recruitment and spawning biomass time-series data were extracted from the RAM Legacy Stock Assessment database (`doi:10.5281/zenodo.4824192`). This database compiles stock assessment results for the world's fisheries. For this analysis, only stocks with at least 20 years of contiguous time-series data were included. Stocks with a recruitment time-series derived directly from a deterministic stock-recruitment relationship were eliminated. In total, 432 stocks representing 182 species in 15 Food and Agriculture Organization (FAO) regions met our criteria (Fig. 1.1), and 253 of these stocks were new to the database since the publication of Szuwalski et al. (2015). Of the 224 stocks included in the original analysis, 45 were removed from the database or modified (i.e., split or merged with other stocks) and 179 were unmodified, except for the addition of new data. Supplementary Appendix 1 contains a comparison of the stocks included in both analyses.

1.3.2 Stock-recruitment relationships

Two methods were used to determine the nature of the relationship between spawning biomass and recruitment over the observed range of spawning biomass, (a) Spearman's correlation and (b) detrended cross-correlation analysis (DCCA). Stocks with a significant positive correlation between spawning biomass and recruitment were categorized as influenced by spawning biomass. Stocks without a significant correlation between spawning biomass and recruitment were considered environmentally-driven. Spearman's correlation differs from the traditional Pearson's correlation by comparing the ranks of the variables (here spawning biomass and recruitment) rather than absolute values (Spearman, 1987). Spearman's correlation is preferable to Pearson's correlation in this scenario, as it removes the assumption of linearity (Spearman, 1987). Most stock-recruitment relationships are assumed to be monotonically increasing but not linear over the estimated range of spawning biomasses (Szuwalski et al., 2015). Unlike Spearman's correlation, DCCA can account for autocorrelation and nonstationarity in time series (Podobnik and Stanley, 2008). DCCA has been used to examine the cross correlation between nonstationary time series in climatology, physiology, and finance (Podobnik and Stanley, 2008). DCCA involves first dividing each time-series of length N into k overlapping windows of size s, where $k = N - s + 1$. Then, the windows of each time series are independently de-trended locally by fitting a linear model so the coefficients of the linear model are unique to each window of an individual time-series. Finally, the standard deviation and covariance is calculated for each window and averaged over all windows. The detrended cross-correlation coefficient for two time series X_t and Y_t, is then:

$$\rho_{DCCA} = \frac{F^2_{DCCA}(s)}{F_{DFA,x}(s)F_{DFA,y}(s)} \tag{1.1}$$

where $F^2_{DCCA}(s)$ is the average detrended covariance between X_t and Y_t, and $F_{DFA,x}(s)$ and $F_{DFA,y}(s)$ are the average detrended standard deviations for X_t and Y_t, respectively.
The DCCA correlation coefficient for each stock-recruitment time series was calculated using the R DCCA package (Prass and Pumi, 2020). Kristoufek (2014) found that DCCA coefficients can be sensitive to the window size. We tested window sizes of 5, 10, and 15 for the 432 stocks. These window sizes were chosen because they represented a varying level of detrending across a variety of time series lengths. Out of the three window sizes, a window size of 15 was determined to be most appropriate for detrending the time series. Supplementary Appendix 2 contains the results from the window size tests.

Some stocks have a dome-shaped stock-recruitment relationship due to compensatory processes at high spawning biomass. These stock-recruitment relationships are not monotonically increasing and, as such, require further processing (66 out of 432). The procedure outlined by Szuwalski et al. (2015) was used to determine the shape of the stock-recruitment relationship (Fig. 1.2 depicts a flow-chart of the procedure). First, the Ricker (1954) and Beverton and Holt (1957) stock-recruitment relationships were fit to the data for each stock. Then the relative likelihood (RL) of each model was calculated using:

$$RL(Mod_{min}) = \frac{1}{1 + e^{(\frac{AIC_c(MOD_{min}) - AIC_c(MOD_{max})}{2})}} \tag{1.2}$$

$$RL(Mod_{max}) = \frac{e^{(\frac{AIC_c(MOD_{min}) - AIC_c(MOD_{max})}{2})}}{1 + e^{(\frac{AIC_c(MOD_{min}) - AIC_c(MOD_{max})}{2})}} \tag{1.3}$$

where MOD_{min} refers to the model (Ricker or Beverton-Holt) with the lower AIC_c value and Modmax to the model with the larger AIC_c value. A relative likelihood for the Ricker model

0.75 was assumed to indicate a dome-shaped stock-recruitment curve following Szuwalski et al. (2015). Stocks below the cut-off were considered to be monotonically increasing. Stock-recruitment pairs for which spawning biomass was greater than 150% of the peak of the fitted curve were then removed before calculating Spearman's correlation for the dome-shaped stocks. The peak of the stock-recruitment curve, S_{max}, was calculated using:

$$S_{max} = \frac{1.5}{b_{ricker}} \tag{1.4}$$

where b_{ricker} represents the density-dependence parameter of the Ricker stock-recruitment relationship. The resulting data were monotonically increasing and no longer violated the assumption needed for Spearman's correlation. The entire stock-recruitment time series was used to calculate the DCCA correlation coefficient because DCCA does not require time series to be monotonically increasing, therefore the dome-shaped stocks did not require special processing.

1.3.3 Categorizing stocks

Recruitment in the RAM database is lagged to the year of fertilization. Therefore, the zero-lagged Spearman's correlation coefficient is indicative of the effect of spawning biomass on recruitment over the observed range of stock sizes. In contrast, correlations at negative lags are indicative of the influence of recruitment on spawning biomass. sizes the zero-lagged correlation coefficient had to be statistically significant and the correlation coefficients for the negative lags had to be less than the zero-lagged coefficient to conclude that such stocks exhibited some influence of spawning biomass over the observed range of stock (Szuwalski et al., 2015). Stocks without a statistically significant zero lag were considered environmentally-driven regardless of the magnitude of lagged correlations. 'Edge cases' had both a significant zero-lag coefficient and negative lag coefficients greater than the zero-lag coefficient. For these stocks, the primary driver of recruitment over the observed biomass could not be determined and choosing the largest correlation value *post hoc* could increase the chance of spurious results. However, Szuwalski et al. (2015) found that 95% of the significant negative lags in the stocks they analyzed were within two years of the age-at-50%-maturity for the stock. As such, significant negative lags were considered a sufficient indicator of the effect of recruitment on spawning biomass.

1.3.4 Regime shifts

The stocks with evidence for environmentally-driven recruitment were examined for regime changes in recruitment. Stocks with recruitment primarily influenced by spawning biomass were not included for regime shift detection because we were unable to disentangle the influence of changes in biomass from recruitment regime changes. Recruitment regimes were determined using the pruned exact linear time (PELT) algorithm in the R *changepoint* package (Killick and Eckley, 2014). PELT is based on the Optimal Partitioning approach (Jackson et al., 2005), but includes a pruning step. First, the optimal change point in the data is calculated by minimizing a cost function. A commonly used cost function in the changepoint literature is the negative log likelihood (Killick et al., 2012):

$$C(y_{(t+1):s}) = -max_{\theta} \sum_{i=t+1}^{k} logf(y_i|\theta) \tag{1.5}$$

where θ represents the parameters of interest for the observations y_i (i.e., recruitment observations) in a given segment (t + *1*, k) of the time series. Time-series segments are defined in the changepoint literature as sections of the time-series with different statistical properties, such as means or variances (Killick et al., 2012). In this case, θ would be the vector of mean recruitments. A change in the mean recruitment over the time-series is used to indicate a regime change. The *changepoint* package in R assumes that the natural logarithms of the observations are normally distributed. The start of subsequent change points is detected by conditioning the cost function on the data from the start of previous change point to the end of the series. By working backwards, the PELT algorithm “prunes” potential points from the dataset that are after the previous change point, reducing the computational time (Killick et al., 2012). The minimum regime duration was set at six years, to capture sub-decadal variation. The PELT algorithm can detect changepoints using a variety of statistical criteria such as the Akaike information criterion (AIC) and the Bayesian Information Criterion (BIC) (Killick and Eckley, 2014). We used the AIC_C given its widespread use in the statistical literature. The full function to minimize is:

$$\sum_{i=1}^{m+1}[C(y_{(\tau_{i-1}+1):\tau_i})] + 2p(\frac{n}{n-p-1}) \tag{1.6}$$

where m is the number of changepoints, which splits the data into $m + 1$ segments, τ_i are the locations of the change points in the time series, $C(y_{(\tau_{i-1}+1):\tau_i})$ is the cost function in Eqn 4, and $2p(\frac{n}{n-p-1})$ is the AIC_C penalty function, where p is the number of parameters introduced to the likelihood by the addition of a new changepoint and n is the total number of recruitment observations. Without a penalty function, the algorithm would consider any stochasticity in the data as a regime change. The *changepoint* package does not include AIC_C as a default penalty function but allows for it to be added manually. The *changepoint* package documentation includes instructions on how to create user-defined penalty functions.

We applied two additional changepoint detection methods to evaluate the robustness of the identified recruitment regimes, Bayesian change point detection and multivariate autoregressive state-space (MARSS) models . Unlike frequentist procedures for change point detection, Bayesian approaches produce a posterior probability distribution for the likelihood of a change point occurring at each location in the time series (Erdman and Emerson, 2008). We used the *bcp* package in R (Erdman and Emerson, 2008), which follows the method proposed by Barry and Hartigan (1993). For each MCMC sample, Barry and Hartigan's algorithm randomly partitions the time series into blocks using a product partition distribution (Barry and Hartigan, 1993) and calculates transition probability, p, at a given position (Erdman and Emerson, 2008). Within each partition, the data are assumed to be *i.i.d.* normal with mean, μ_{ij} (Barry and Hartigan, 1993). The posterior distributions of the location of change points and the means μ_{ij} are updated using MCMC sampling. Barry and Hartigan (1993) provide a more complete description of the algorithm. A drawback of Bayesian change point detection is that the method only provides a posterior probability. It is necessary to select a minimum posterior probability to designate a regime shift, and we chose to designate a regime shift when the posterior probability of a regime change was 0.75 by visually comparing the posterior probabilities with regime shift locations determined using the other two methods.

For the MARSS model, we used the method outlined in Holmes et al. (2021) to detect regime changes in the state process using standardized residuals from the mean recruitment. First, the recruitment observations were modelled as an autoregressive process. The MARSS model is:

$$x_i = x_{i-1} + w_i, w_i \sim N(0, q) \tag{1.7}$$

$$y_i = xy_{i-1} + v_i, v_i \sim N(0, r) \tag{1.8}$$

$$x_0 = \pi \tag{1.9}$$

where y_i is the "observed" recruitment during year i, x_i is the "true" recruitment during year i, and x_0 is the true initial recruitment at the start of the time series. x_0 is assumed to be a constant (π) and is estimated by the model (Holmes et al., 2021). Both w_i and v_i are error terms and assumed to be normally distributed with mean 0 and variances q and r, respectively. Years during which the standardized state residuals were > 2 or $<$ -2 were considered regime changes.

1.3.5 Time-series characteristics

We used simulations to determine the characteristics of a stock-recruitment time series that best leads to the detection of a stock-recruitment relationship using Spearman's correlation coefficient and detrended cross-correlation analysis (DCCA). Spawning biomass time series were simulated in two ways. In the first method, and following Szuwalski et al. (2015), each spawning biomass observation was chosen from a uniform distribution over the range

(dS_0,S_0) where S_0 represents the unfished spawning biomass and d represents the maximum depletion level of the stock. In the second method, the first observation was sampled from a uniform distribution using the same range as above. All subsequent points in the time series were then generated using an AR(1) process with an autocorrelation coefficient equal to 0.7 (roughly the average autocorrelation for spawning biomass in the RAM database), i.e.:

$$S_t = c + 0.7S_{t-1} + e_t, e_t \sim N(0, \sigma_s) \tag{1.10}$$

where c is a constant. S_t and e_t respectively refer to the spawning biomass and observation error at time t. e_t was normally distributed with variance σ_s.We used two depletion (d) levels, 0.1 and 0.5, to mimic the different exploitation histories for the stocks in the database where a depletion level of 0.1 indicates a high (past) level of exploitation. Recruitment time series were generated using the Beverton-Holt (R_{bh}) and Ricker (R_{ricker}) stock-recruitment relationships with varying levels of steepness (h = 0.3, 0.5, 0.7, 0.9), i.e.,

$$R_{bh} = \frac{4hSR_0e^{\epsilon_t - \sigma_r^2/2}}{SPR_0R_0(1-h) + (5h-1)S} \tag{1.11}$$

$$R_{ricker} = \frac{S}{SPR_0}(5h)^{1.25(1-\frac{S}{R_0SPR_0})}e^{\epsilon_t - \sigma_r^2/2} \tag{1.12}$$

$$\epsilon_t \sim N(0, \sigma_r) \tag{1.13}$$

where R_0 represents the recruitment at the unfished spawning biomass S_0. SPR_0 is the unfished spawners-per-recruit and ϵ_t represents process error at time t. 10,000 simulations were run for each combination of the spawning biomass generation process, stock-recruitment relationship, and steepness value. For each simulation, the recruitment error for the time series was sampled randomly for σ_r = 0.1, 0.3, 0.5, 0.7, and 0.9. Then, the Spearman's correlation and DCCA correlation coefficients were calculated for each simulated data set.

The results of the simulation experiment were analyzed relative to the characteristics of the stocks in each recruitment driver category (primarily influenced by spawning biomass, environment, or edge case). Specifically, we calculated σ_r and the historical depletion for each stock included in the analysis to determine whether differences existed between each of the categories.

1.4 Results

1.4.1 Categorizing stocks

The additional data (median 9 years) since Szuwalski et al. (2015) appear to have informed some of the dynamics influencing recruitment. For example, the whiting *Merlangius merlangus* stock in ICES area VIa was originally an edge case. In this analysis, it had a significant positive influence of spawning biomass on recruitment (Fig. 1.3). Conversely, additional data allowed the stock of arrowtooth flounder *Atheresthes stomias* in the Bering Sea and Aleutian Islands to be classified as environmentally-driven instead of an edge case (Fig. 1.3). Seventeen of the 39 “unmodified” stocks that were unclassified in the original analysis were reclassified in this analysis (for the full list see Supplementary Appendix 1, Table 2).

Only 14% of the stocks analyzed had evidence of some influence of spawning biomass on recruitment and 57% did not over the range of observed spawning biomasses (Table 1.1), similar to the results of Szuwalski et al. (2015); 16% and 61% respectively). More stocks than before had evidence of recruitment influencing spawning biomass more than spawning biomass influencing recruitment (29% versus 23%). This increase in ‘edge case’ stocks is partially due to the addition of new stocks to the RAM database. Out of the 127 edge case stocks, 86 were not included in Szuwalski et al. (2015)’s original analysis. Supplementary Appendix 4 contains plots for each of the 432 stocks.

The Spearman’s correlation coefficient differed markedly from the DCCA correlation coefficients (Fig. 1.4), particularly for stocks with dome-shaped stock-recruitment relationships. This difference is likely due to the pre-processing necessary to make the time series monotonically increasing for Spearman’s correlation. The simulation experiment also indicated that the DCCA coefficient has larger interquartile ranges than Spearman’s over a variety of scenarios. This may account for the lack of relationship between the two methods. Supplementary Appendix 5 contains dumbbell plots for difference in correlation coefficients among stocks.

1.4.2 Recruitment regimes

Of the 371 stocks with some influence of environment on recruitment, 44% had at least one change in average recruitment, as detected by the PELT algorithm. This decrease is likely related to the new stocks included in the analysis and the change in detection algorithm. PELT appears to be less likely to designate a regime shift than STARS.

There were major differences in the number and location of regime changes among the three methods for regime detection (PELT, Bayesian change point (BCP) detection, and MARSS). The BCP algorithm detected the fewest shifts of the three methods, with a median number of regimes detected of 0. The MARSS model detected more regime shifts than both BCP and PELT, with a median of 2 regimes detected. The median number of regimes detected in recruitment time series using the PELT algorithm was 1. BCP and PELT were able to be used on all stocks, regardless of the time series variation, but 11% of the stocks did not have enough variation in the recruitment time series to estimate both process and observation error variances for the MARSS model.

Across stocks, more years were identified as regime shifts using the PELT algorithm than using BCP. This difference is in part due to the minimum posterior probability selected to 'define' a regime shift. A lower minimum posterior probability would likely increase the number of years identified as regime shifts, while a higher minimum posterior probability would likely decrease the number of years identified. For several stocks, all methods identified a recruitment regime change within the same time frame (e.g., Fig. 1.5). Supplementary Appendix 3 compares the three methods for each stock with at least one regime change.

1.4.3 Time series characteristics

We found Spearman's correlation and DCCA performed best on stocks in our simulation experiment that were more depleted and had low steepness and recruitment error. For all the scenarios, the DCCA coefficient had a larger interquartile range than Spearman's correlation across simulations. The generation method for the spawning biomass time series and choice of stock-recruitment curve did not appear to affect whether a stock-recruitment relationship was detected or not (Supplementary Appendix 6). Fig. 1.6 depicts simulation results for one of the Beverton-Holt scenarios, with high historical depletion (0.1) and low steepness (0.3). This is a “best case” scenario for identifying a correlation between spawning biomass and recruitment. Both methods identify a correlation even when the recruitment error is relatively high.

The historical depletion for all stocks included in this analysis is given in Fig.1.7a. As expected from the simulation results, historical depletion was lowest for spawning-biomass influenced and edge case stocks, and the estimated recruitment variation was lowest for stocks with spawning biomass primarily influencing recruitment (Fig. 1.7b).

1.5 Discussion

The purpose of this study was to determine the robustness of the results of Szuwalski et al. (2015) to the addition of new data and methods. Our results suggest that the recruitment of most stocks in the RAM Legacy Stock Assessment database is primarily influenced by environmental conditions, rather than spawning biomass. Only 14% stocks analyzed had a significant positive relationship between spawning biomass and recruitment over the observed biomass range. This finding is consistent with previous meta-analyses (Gilbert, 1997; e.g., Szuwalski et al., 2015). Like Szuwalski et al. (2015), we found that a large proportion of the environmentally-influenced stocks experienced a regime change when using a change point detection algorithm. However, this is not the only method for regime detection. Our results suggest the choice of method can have a large impact on inferences regarding the number and location of regime changes in recruitment. BCP detected the fewest regime shifts of the three methods while the MARSS models identified the greatest number. More research is needed to determine which of the methods is the most appropriate for detecting regime shifts in fishery time series, which requires an agreed upon and operational definition of ‘regime shift’. Evaluating the performances of each of the methods using simulated data with ‘known’ regime shifts could provide some insight.

Our simulation results highlight that the ability to detect a stock-recruitment relationship is highly dependent on the contrast in the spawning biomass time series and the level of recruitment variation. This is consistent with our analysis of the RAM database stocks, where those with a detectable influence of spawning biomass showed greater contrast in spawning biomass and lower recruitment error (Fig. 1.7). This poses a problem for fishery managers, as many stocks are unlikely to have the necessary characteristics to have a well-defined stock-recruitment relationship, which complicates the derivation of management reference points. More data may help illuminate the influence of spawning biomass on recruitment for stocks with high recruitment variability.

Several of the edge case stocks analyzed in Szuwalski et al. (2015) were reclassified in this analysis. Continuing to update the RAM database may lead to more reclassified stocks in the future. Until longer time series can be developed, research into decreasing recruitment uncertainty could be valuable. Some studies have examined the use of environmental covariates to forecast recruitment (e.g., Basson, 1999; Olsen et al., 2011; Swain and Sinclair, 2000). However, results have been mixed. Another approach is using a non-parametric stock-recruitment model, such as in Munch et al. (2005) and Cadigan (2013). Cadigan (2013) found that using the bootstrapped confidence intervals derived from a shape-constrained additive stock-recruitment model better reflected the uncertainty in the stock-recruitment relationship than parametric models. Many stock assessments use stock-recruitment relationships to calculate reference points despite the absence of a detectable influence of spawning biomass on recruitment (Kell et al., 2016). This assumption may have ramifications on the status of the stock, especially if assessments are not regular.

We compared two correlation methods to quantify the influence of spawning biomass on recruitment. Although DCCA has advantages over Spearman’s correlation, our simulation results indicate that its performance is marginally worse when attempting to identify a stock-recruitment relationship. Across our simulations, the DCCA correlation coefficient had a larger interquartile range than Spearman’s. We suggest that DCCA is not the ideal method for stock-recruitment data. The performance issues with DCCA on the time series we analyzed may be due to a combination of the number of observations and the magnitude of observation error. DCCA was developed using longer time series with less observation error, which might impact its performance when applied to the typical spawning biomass time series. Longer time-series or pre-processing (limiting DCCA to the monotonically increasing observations, as done for Spearman's correlation) may reconcile these differences. Also, the DCCA coefficient has been shown to be sensitive to the window size. In this analysis, we chose a standard window size for all the stocks analyzed. A more appropriate method may have been to choose a window size based on the length of each stock's time series. Commonly tested window sizes range from 8 to $N/5$ or $N/4$, where N is the length of the time series (Barunik and Kristoufek, 2010); Mukli et al. (2015); Kaposzta et al. (2022)].

Three hundred and seventy-one of the 432 stocks we analyzed had a stronger influence of environment than spawning biomass on recruitment. Because all stocks likely have some influence of environment on recruitment, identifying how this relationship can change over time is important. The prevalence of recruitment regime changes is important from a management perspective (Collie et al., 2004), and several studies have been conducted to identify regime shifts in fish stocks (Collie et al., 2004; e.g., Hollowed et al., 1987; Perälä et al., 2017; Szuwalski et al., 2015). In our study, 46% of the 371 stocks with some influence of environment on recruitment had at least one regime shift detected using the PELT change point detection algorithm. This is a decrease from Szuwalski et al. (2015), which is likely due to the change from the STARS to PELT algorithm. The STARS algorithm identified on average one more regime than PELT, despite these being similar algorithms. Our results suggest that the number and location of regime shifts is dependent on the method. Further research is needed to determine the most appropriate method for identifying regime shifts naively (i.e., without prior suspected years for shifts). Each of the methods explored in this paper has advantages and disadvantages. To our knowledge, there is no peer-reviewed study directly comparing change point detection methods. We believe a simulation study could provide insight into when any one method is appropriate. Until then, we cannot comment on the best method and suggest that researchers use multiple change point detection methods to analyze their data. It is often necessary to use naïve changepoint detection methods in systems or for stocks that have not been extensively studied. Using multiple methods could allow for a more robust analysis in scenarios where methods that require *a priori* hypotheses of regime shifts are not feasible. BCP also appears promising and has been used in previous studies because it quantifies the relative uncertainty around the change point location. It has also been adapted for online forecasting, which can estimate the likelihood of a regime shift, given the time since the previous regime shift(Adams and MacKay, 2007). Perälä et al. (2017) and Munch et al. (2018) explored various methods for modeling a changing stock-recruitment relationship through time using Bayesian methods. The prevalence of stocks with some evidence of a regime shift in our analysis indicates that time-varying mean recruitment may be a worthwhile addition to assessment models.

There are several limitations to using recruitment time-series for analysis. Recruitment time series are often treated as "data" despite being model outputs (Dickey-Collas et al., 2015). The stocks in the RAM Legacy database also do not include errors for recruitment or spawning biomass estimates, which can lead misinterpretation of the stock-recruitment relationship Walters and Ludwig (1981). @Dickey-Collas et al. (2015) found recruitment time series were sensitive to the assumptions of the assessment model on which they are based. Even with recruitment indexes, such as egg production data, determining the stock-recruitment relationship is challenging (Kell et al., 2016; Marshall, 2009). The RAM Legacy database is biased towards species in the north Atlantic and Pacific (Fig. 1.1). We have comparably fewer assessments for tropical species. Furthermore, the assessments from these areas are likely to be data limited, precluding long time-series of spawning biomass and recruitment time series.

In summary, our analysis indicates that some aspects of Szuwalski et al. (2015) are robust to the addition of new data. The proportion of stocks with detectable evidence of the influence of spawning biomass on recruitment is similar to previous findings. Environmental conditions appear to play a large role in recruitment variation over time, as 86% of stocks in this analysis had a detectable influence of environment larger than the influence of spawning biomass on recruitment. We found regime detection is sensitive to the change point detection method, but further research is needed. Simulation studies including time series with known regime shifts could lead to a better understanding of the performance of the various estimation methods. However, a large proportion of stocks were found to have at least one regime shift in the time series. Current approaches to stock-recruitment modeling often assume no large, sudden temporal variation, as would be observed under regime shifts (Perälä et al., 2017), which could result in biased projections of stock biomass. Inaccurate projections can have large impacts on the management of a stock, especially when harvest control rules cannot account for abrupt shifts in biomass (see A'mar et al., 2009; Szuwalski and Punt, 2013). For example, a shift into a lower-recruitment regime could lead to weaker cohorts in the following years. If the stock is not assessed regularly, the biomass projections produced in a high recruitment regime may be inaccurate in a low recruitment regime. This may increase the likelihood that a stock is overfished. Regime shifts may also impact the biological reference points used in management. In the United States, the status of a stock is usually determined by comparing the biomass estimate relative to a reference point, such as unfished biomass (B_0) or the biomass corresponding to MSY (B_{MSY}). Estimating B_0 often requires extrapolating to time periods without much catch or abundance data (Hilborn 2002). The B_0 value used in a stock assessment may not be representative of the current environmental regime, which impacts our perception of the stock's status. Stocks are vulnerable to overfishing or collapse without reliable estimates of their status. Time-varying stock-recruitment relationships or harvest controls rules robust to regime shifts will be essential in the face of a changing environment, but these can have unintended consequences (Szuwalski et al., 2023). Identifying forecasting techniques that perform well within these systems could also allow for better management advice and more sustainable harvests. Regime shifts should be a consideration for fishery scientists developing the next generation of stock assessment models.

TABLES

Table 1.1: Summary of total number of stocks assigned to each recruitment driver based on the shape of the stock-recruitment curve.

Dome-shaped stocks		Monotonic stocks		
Environment	Spawning biomass	Environment	Spawning biomass	Environment and/or spawning biomass
50	16	194	45	127

FIGURES

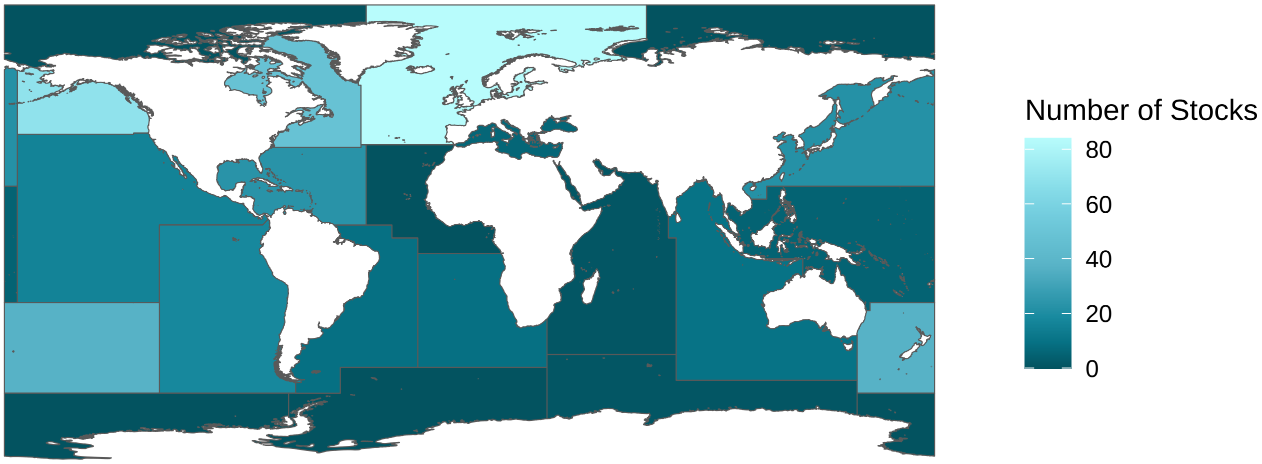

Figure 1.1: Map of stocks included in analysis by FAO region.

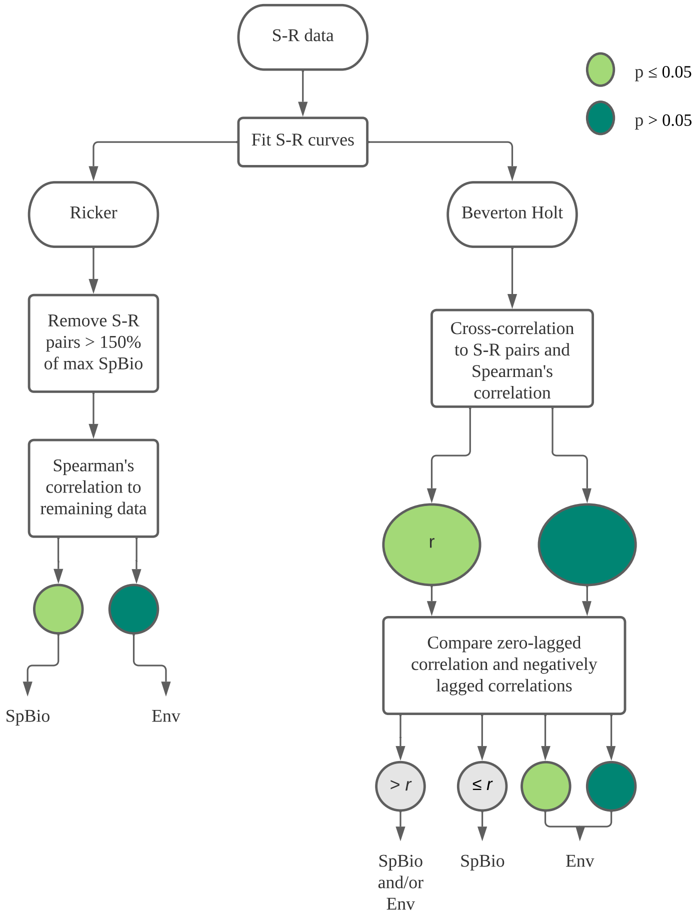

Figure 1.2: Decision process for assigning drivers to each stock's recruitment. Boxes represent tests performed; ovals are the results of the tests. r represents the coefficient for the zero-lagged correlation. If r is significant and the negatively lagged correlations are less than or equal to r, the stock is designated as driven primarily by spawning biomass. Otherwise, no primary driver can be identified.

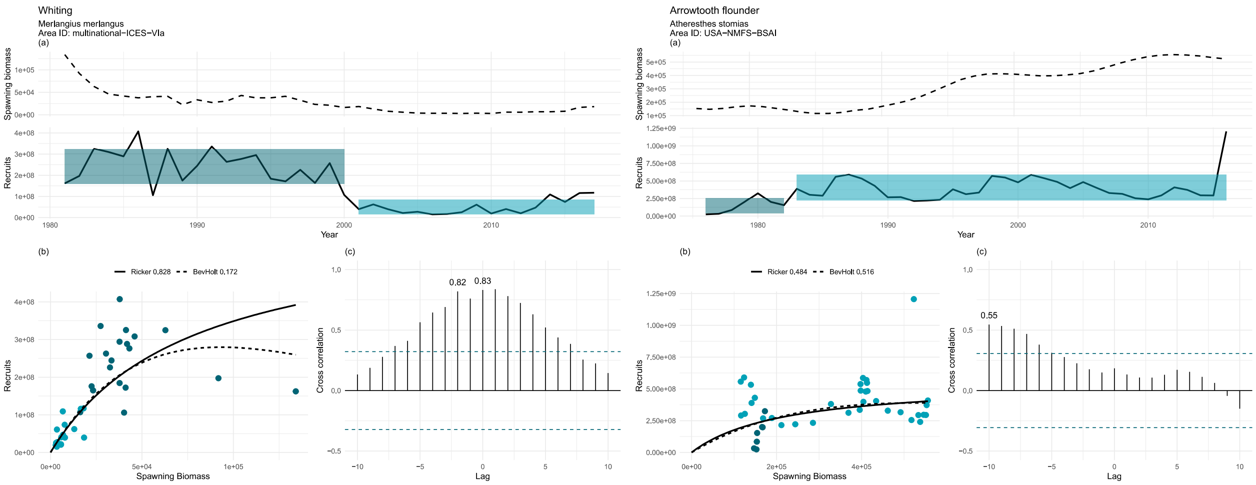

Figure 1.3: The ICES VIa stock of whiting (*Merlangius merlangus*; upper panels) was one of the stocks reclassified from edge case to spawning biomass influenced, while the Bering Sea and Aleutian Islands stock of arrowtooth flounder (*Atheresthes stomias*; lower panels) was reclassified from an edge case to environmentally-influenced. Panel (a) is the recruitment and spawning biomass time series. Boxes highlight the start and end of the recruitment regimes as identified using the PELT algorithm. The center of each box is aligned on the mean regime recruitment. The widths of the boxes represent the standard deviations during each regime. (b) depicts the stock-recruitment curve for the Ricker and Beverton-Holt models. Points are color-coded by recruitment regime. (c) is the cross-correlation values for the spawning biomass and recruitment time series. The values for significant zero-lagged coefficients and the maximum negative-lagged coefficients are listed.

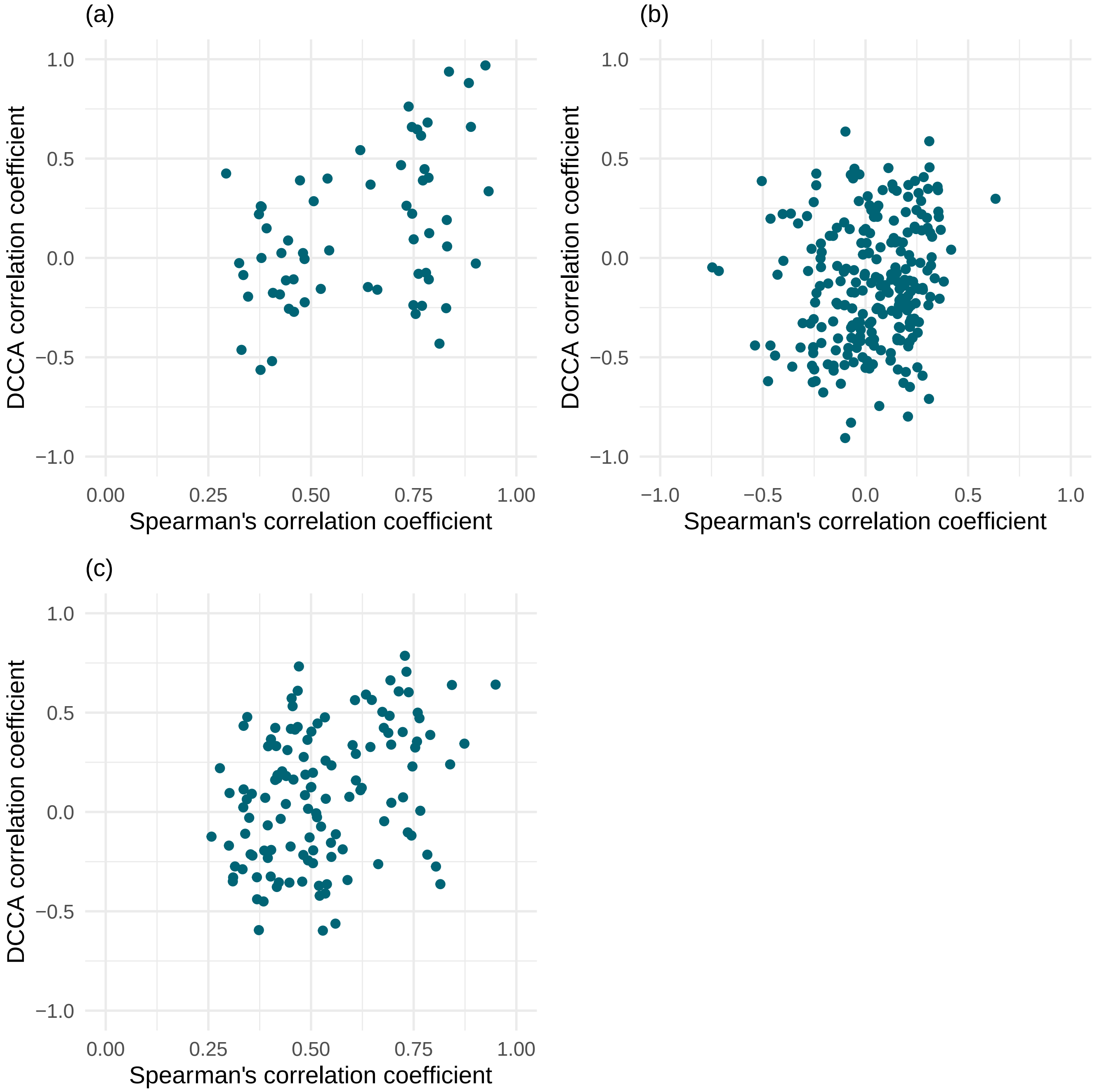

Figure 1.4: Comparison of DCCA and Spearman's correlation coefficients for each stock with the primary recruitment influence categorized as spawning biomass (a), environment (b), and edge cases (c).

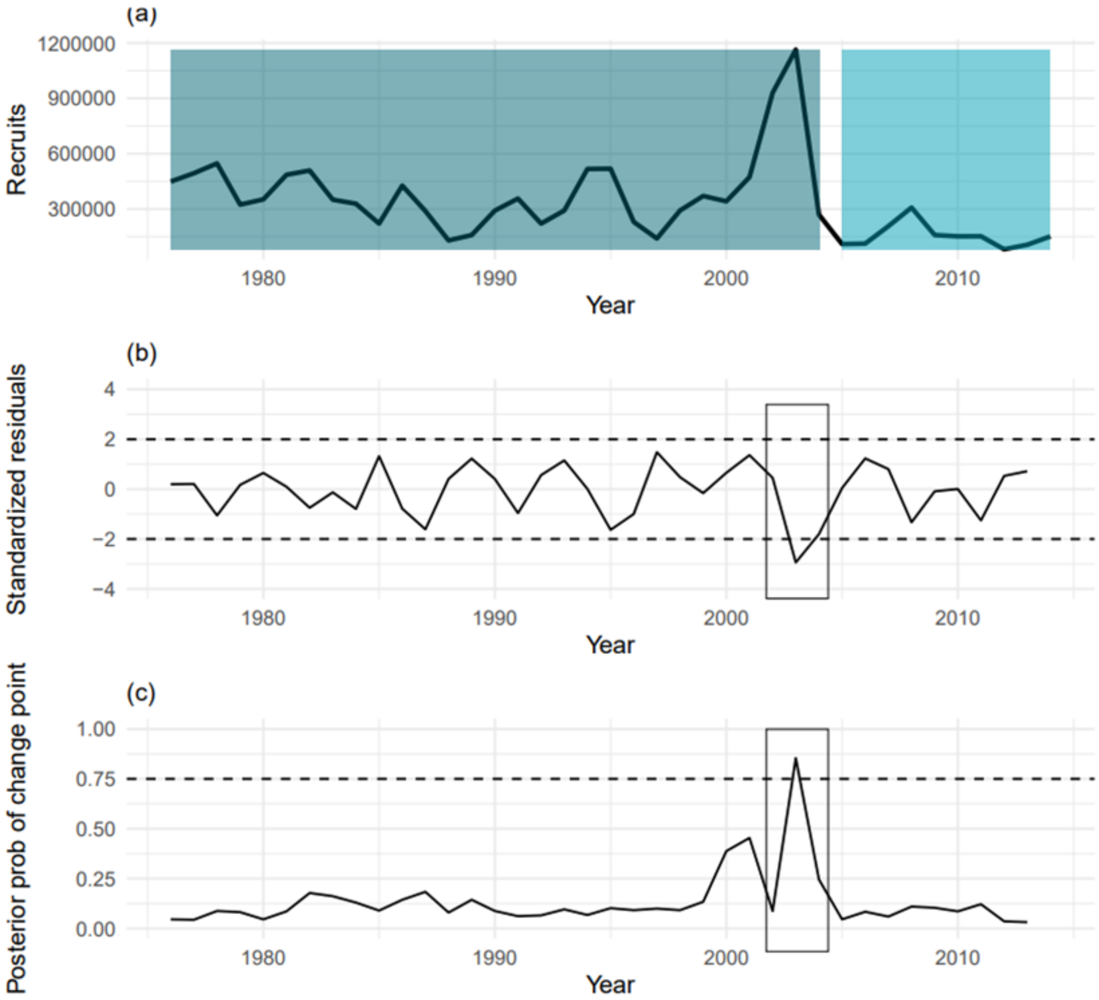

Figure 1.5: Comparison of regime shift locations for the three methods for the U.S. Southeast and Gulf of Mexico stock of red grouper, *Epinehelus morio*. All methods show evidence of a recruitment regime shift around 2003. Panels (a), (b), and (c) depict the PELT algorithim, MARSS model, and BCP results, respectively. Recruitment regime shifts for the MARSS model and BCP are highlighted in boxes.

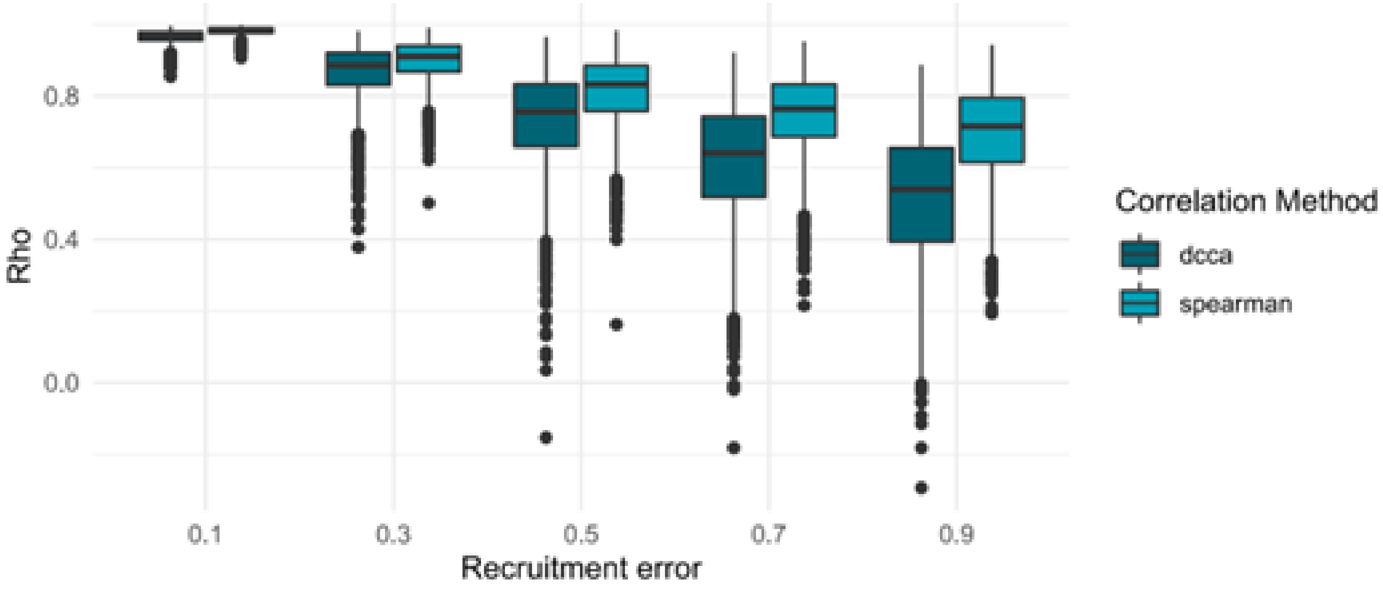

Figure 1.6: Range of DCCA and Spearman's correlation coefficients for one of the simulated scenarios. The spawning biomass time-series was generated using a uniform distribution and the recruitment time series using a Beverton-Holt function. This scenario had a minimum depletion (d) of 0.1 and a steepness parameter (h) of 0.3.

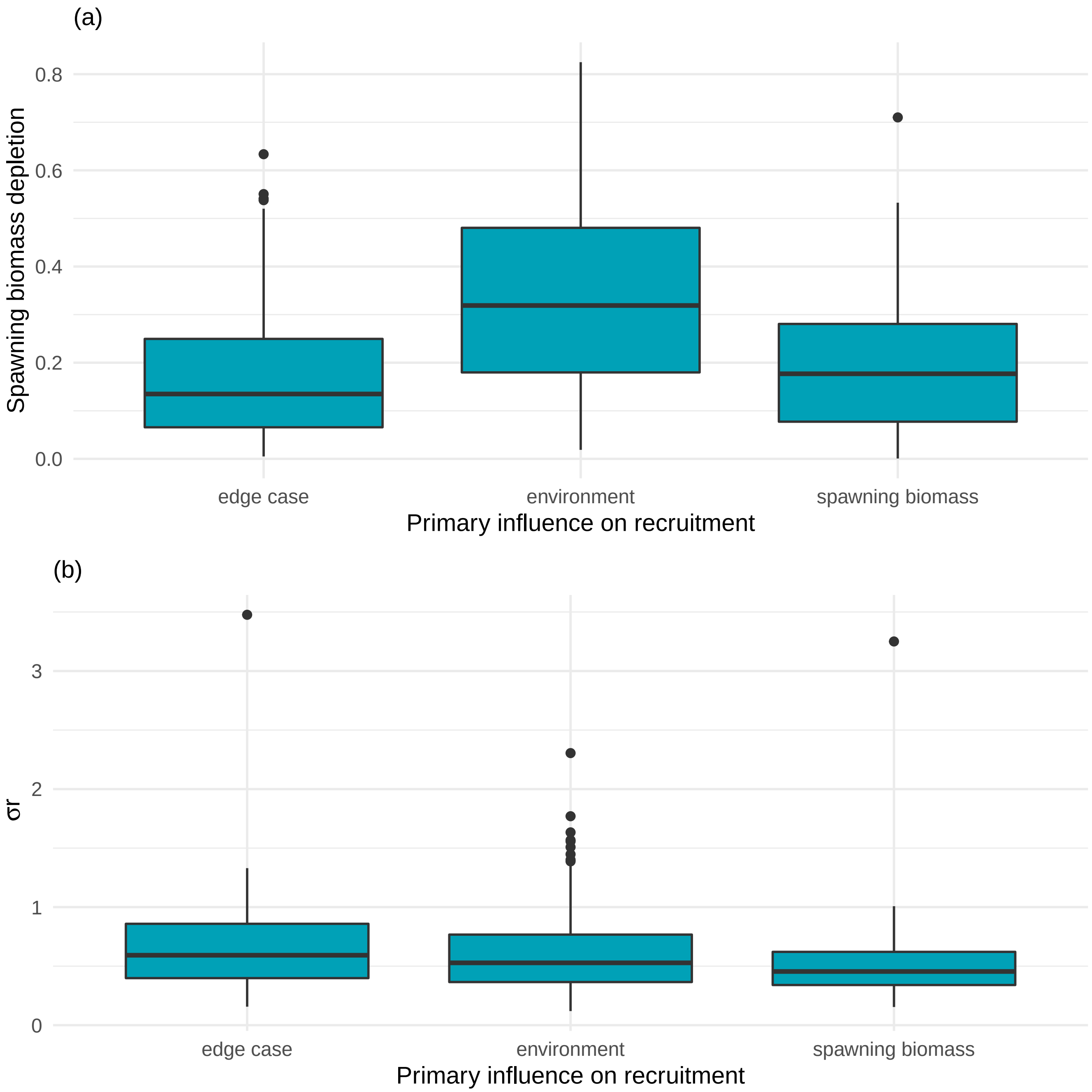

Figure 1.7: (a) Historical depletion for each of the 432 stocks in this analysis by primary influence on recruitment. Depletion was calculated as the proportion of the 5th to the 95th quantiles of spawning biomass (b) Estimated variation of the recruitment time series for each of the stocks included in this analysis.

Chapter 2

EVALUATING THE SHORT- AND MID-TERM PERFORMANCE OF SIX FORECASTING METHODS ON WEST COAST AND ALASKAN GROUNDFISH RECRUITMENT

2.1 Abstract

Predicting future recruitment has been an area of concentrated research by fishery scientists for over a hundred years. The variety of biological and environmental factors that influence recruitment have made providing consistently accurate forecasts challenging. Despite the challenges associated with forecasting recruitment, improved recruitment forecasts are for fishery scientists and managers, especially with changing ocean conditions due to climate change. We evaluated short- and mid-term performance of six forecasting methods: 1) Beverton-Holt stock-recruitment relationship, 2) an autoregressive-1 function, 3) mean recruitment, 4) a hidden Markov sampling procedure, and 5) simplex projection, 6) a PELT changepoint sampling procedure for data-rich groundfish stocks managed by the Pacific Fishery Management Council (PFMC) and the North Pacific Fishery Management Council (NPFMC). The stocks differed in their time series length, exploitation history, and the age of recruitment to the fishery. We determined that there was not one single best method. Forecast performance was related to the length of the forecast period, the performance metric used, and the characteristics of the recruitment and spawning biomass time series. An autoregressive-1 function generally performed well for short-term predictions across all the stocks we evaluated. For mid-term forecasts, PELT changepoint sampling and simplex projection performed best. We found that recruitment autocorrelation was a key characteristic to consider when choosing a forecast method. Only simplex projection per-

formed well across a wide range of autocorrelation values. Our analysis suggests that fishery scientists and managers should consider their time frame of interest and the properties of the time series when deciding on a forecast method.

2.2 Introduction

Recruitment, the numbers of young fish entering the population, can be influenced by a variety of biological and environmental factors. Due to the number of factors that can influence recruitment, providing consistently accurate forecasts has proved elusive (Subbey et al., 2014). Several methods for forecasting recruitment have been tested. Methods can be classified as parametric, semi-parametric, or nonparametric (Subbey et al., 2014). Parametric methods assume a functional form for the relationship between recruitment and predictors (Subbey et al., 2014). Classical parametric methods are stock-recruitment functions such as the Beverton and Holt (1957) and Ricker (1954) models. These models assume that mean recruitment is a function of the spawning biomass. However, the primary driver of the stock-recruitment relationship has been a source of debate. (Myers and Barrowman, 1996) suggested reduced spawning biomass was associated with reduced recruitment while Gilbert (1997) argued that instead of spawning biomass primarily driving recruitment, environmental factors played a larger role than spawning biomass in determining recruitment. For example, environmental conditions could lead to lower recruitment, which results in lower spawning biomasses. More recent analyses of stocks in the RAM Legacy Stock Assessment database (Ricard et al., 2012) have found spawning biomass often has little detectable impact on recruitment, in part due to the limited ranges of observed spawning biomass (Sellinger et al., 2024; Szuwalski et al., 2015).

Semi- and nonparametric methods for forecasting recruitment are based on a variety of modeling approaches. For example, Evans and Rice (1988) used a probability distribution to predict recruitment in place of a functional relationship with spawning biomass, while Cardinale and Arrhenius (2000) used generalized additive models (GAMs) to analyze the recruitment of Baltic cod, and Sun et al. (2009) evaluated artificial neural networks for Norwegian spring-spawning herring. Empirical dynamic modeling (EDM), described by Sugihara and May (1990), is also a popular non-parametric prediction technique. EDM accommodates nonequilibrium and nonlinear dynamics (Ye et al., 2015). Simplex projection and S-map, two EDM techniques were used to forecast recruitment of sockeye salmon in the Fraser River (Ye et al., 2015). The EDM forecasts outperformed a Ricker model for eight of the nine salmon stocks examined. In addition, an analysis conducted by Munch et al. (2018) on 185 stocks in the RAM Legacy Stock Assessment database found that EDM forecasts had lower prediction error than classic stock-recruitment relationships. The EDM method used by Munch et al. (2018) was able to take advantage of multiple recent observations, as opposed to the single spawning biomass input required for classic stock-recruitment functions. Parametric and nonparametric methods can also be used together in ensemble models, as demonstrated by (Ovando et al., 2022) on Bristol Bay sockeye salmon.

Choosing an appropriate method to forecast recruitment can be challenging. Methods differ in the amount and quality of data needed, the assumptions made by the models, and the ease of implementation. Furthermore, different methods have variable performance in short- and mid-term predictions. For example, Van Beveren et al. (2021) simulated forecasts for 31 scenarios with varying life histories. There was no single best performing method among the 16 techniques tested (Van Beveren et al., 2021). Performance was also linked to the age at maturity for the stock, recruitment variability, and autocorrelation in the recruitment time series (Van Beveren et al., 2021).

Despite the challenges associated with forecasting recruitment, it continues to be an integral part of stock assessment for certain stocks (Johnson et al., 2016). Methods that are robust to changing environments, which are expected under climate change, will be important for fishery scientists and managers in the near future (Subbey et al., 2014). Identifying forecasting techniques that perform well within systems that undergo regime shifts could also allow for more sustainable harvest.

Our goal was to test a suite of forecast methods that vary in operational use across regions and the level of parameterization. Short- and mid-term recruitment forecast performance of six forecasting methods: 1) Beverton-Holt stock-recruitment relationship, 2) an autoregressive-1 function, 3) mean recruitment, 4) a hidden Markov sampling procedure, and 5) simplex projection, 6) a PELT changepoint sampling procedure were evaluated for data-rich groundfish stocks managed by the Pacific Fishery Management Council (PFMC) and the North Pacific Fishery Management Council (NPFMC). Short-term forecasts are defined as 1 year and mid-term forecasts as 5 years.

2.3 Methods

2.3.1 Data

The groundfish stocks used in the analyses include data-rich rockfish, flatfish, and roundfish from three management areas, the Pacific West Coast, the Bering Sea and Aleutian Islands, and the Gulf of Alaska (Table 2.1). The stocks included were restricted to those with at least 20 years of recruitment estimates. Any stock that derived their recruitment time series directly from a stock-recruitment relationship was eliminated. Time series for Pacific West Coast stocks were trimmed to the main recruitment deviation period for the stock assessment. These stocks differ in their time series length, exploitation history and the age of recruitment to the fishery (Fig. 2.1).

2.3.2 Forecasting methods

Six forecasting methods are fitted to a training data set:

1. 1. Average recruitment plus process noise:

$$log(R_t) = \frac{1}{t-1}\sum_{1}^{t-1} log(R_i) + e_t, e_t \sim N(0, \sigma_r) \tag{2.1}$$

2. 2. An autoregressive-1 (AR(1)) time-series model. The log of the recruitment at time t is a function of the log of recruitment at time $t-1$, i.e.:

$$log(R_t) = \phi log(R_{t-1}) + e_t, e_t \sim N(0, \sigma_r) \tag{2.2}$$

where $|\phi| < 1$.

3. 3. The Beverton-Holt (Beverton and Holt, 1957) stock-recruitment relationship with process noise, i.e.:

$$R_t = \frac{aS_t}{1 + bS_t} exp(e_t), e_t \sim N(0, \sigma_r) \tag{2.3}$$

S_t is assumed to be equal to the spawning biomass at the final year of the training set.

4. Simplex projection (Sugihara and May, 1990). Simplex projection is based on Takens' theorem (Takens, 1981), which uses time lags of an observed variable to recreate the dynamics of the generating system. It is a nonlinear and nonparametric technique, which makes it useful for ecological time series (Munch et al., 2018). The number of time lags is referred to as the embedding dimension (E). Once the embedding dimension is chosen, the time series is transformed into a series of E-dimensional points $X^E = \{xt, xt-1, ..., xt-(E-1)\}$. For example, an embedding dimension of 3 would result in series of 3-dimensional points $\{xt, xt-1, xt-2\}$ (e.g., Fig. 2.2), which represent segments of the original time series. Prediction for a given point, x_t, follows a k-nearest-neighbors approach. Each E-dimensional point is compared to the other E-dimensional points in the time series. The "nearest neighbor" points are segments of the time series that mostly closely match the preceding dynamics of the given point $\{xt-1, xt-2\}$. The target prediction xt is calculated from the weighted average of the one-step forward projections from each of the identified nearest neighbor segments @(Sugihara and May, 1990) with the weights calculated by taking the exponential of the distance between the target E-dimensional point and the identified nearest neighbors. The number of neighbor points used by the algorithm is often set to E+1 (Sugihara and May, 1990). Figure 2.3 walks through the steps of simplex projection in 2 dimensions (adapted from Petchey, 2016). For our analysis, we included process error into the simplex forecasts,

making it semi-parametric. The recruitment at time t, R_t, is calculated by taking the average of the nearest E+1 neighbors, $(\overline{R_{E+1\,neighbors}})$ multiplied by the exponential of the process error. The process error is assumed to be normally distributed with mean 0 and standard deviation σ_r.

$$R_t = \overline{R_{E+1\,neighbors}} exp(e_t), e_t \sim N(0, \sigma_r) \tag{2.4}$$

σ_r is estimated from the training dataset by standardizing the recruitment data, R_t, using the mean recruitment, $\overline{R}$.

$$r_t = log(R_t) - log(\overline{R}) \tag{2.5}$$

and the estimate for σ_r is calculated using the *mle* function in the *stats4* package in R. The simplex projection function in the *rEDM* package (Pack et al. 2023) provides deterministic confidence intervals for predictions. Adding process error to these predictions may overestimate the variation in the data. However, without adding process error, we would be unable to perform simulations.

5. Hidden Markov model sampling procedure. The sampling procedure is based on the protocol developed by Brodziak (2022). Brodziak's sampling procedure begins by using a k-means clustering algorithm to classify recruitment and spawning biomass observations into terciles (high, medium, and low groups). Then, the number of recruitment values within each tercile are counted for each spawning biomass tercile (high, medium, or low). These values are used to calculate the transition probabilities. Recruitment predictions are generated by determining the tercile into which the spawning biomass observation falls and sampling from a multinomial distribution with the tercile-specific transition probabilities. Our sampling algorithm differs from Brodziak's. Instead of using a clustering algorithm in the first step, we fit a 2-state hidden Markov model to the training data using the *depmixS4* package in R (Visser and Speekenbrink 2010). The response in this model is the log ratio of recruitment to spawning biomass, and the predictor is spawning biomass.

$$log(\frac{R}{S} = \beta_{0,j} + \beta_{1,j}S) \tag{2.6}$$

where $\beta_{0,j}$ is the intercept for state j, and $j = 1, 2$. $\beta_{1,j}$ is the coefficient for spawning biomass in state $j = 1, 2$. The errors for the response are assumed to be gaussian. The depmixS4 output provides estimates for the transition matrix from state 1 and state 2, the βs for each state, and the standard deviation of the recruitment error in each state. The posterior state classifications at each time step in the training set can be recovered for the data and model.

The forecasting portion of the sampling protocol uses the estimated transition matrix and the posterior state classifications for the training data. First, the state at each of the forecast years is generated using the final state classification in the training data and the estimated transition matrix. For example, if the predicted state at the end of the training data was 2, the estimated transition probabilities for state 2 are used to predict the state at forecast year 1. This process can be repeated to determine the state for multiple forecast years. Once the state is determined for a forecast year, the log recruitment-spawning biomass ratio for that year is sampled from a normal distribution $N(\mu_j, \sigma_j)$ where j represents state 1 or 2. μ_j and σ_j are calculated by separating recruitment and spawning biomass observations in the training set by posterior state classification, then determining the mean and standard deviation for each state. Recruitment predictions for forecasts are calculated by multiplying the sampled recruitment-spawning biomass ratio by the spawning biomass for the year. Like the Beverton-Holt method, the spawning biomass is assumed to be equal to the spawning biomass at the final year of the training set.

$$R_t = (\frac{R}{S}) \tag{2.7}$$

6. PELT sampling procedure. The The PELT sampling procedure is similar to the Hidden Markov model sampling method but does not assume a default of recruitment regimes. The number of recruitment regimes, if any are detected, is determined by the pruned exact linear time (PELT) algorithm. PELT is based on the Optimal Partitioning approach (Jackson et al., 2005). PELT has been shown to be faster and more accurate

than other multiple changepoint detection algorithms, such as binary segmentation (Killick et al., 2012). First, the optimal change point in the data is calculated by minimizing a cost function. A commonly used cost function in changepoint literature is the negative log likelihood (Killick et al., 2012):

$$C(y_{(t+1):s}) = -max_{\theta} \sum_{i=t+1}^{k} logf(y_i|\theta) \tag{2.8}$$

where θ represents the parameters of interest for the observations y_i (i.e., recruitment observations) in a given segment (t + *1*, k) of the time series. Time-series segments are defined in the changepoint literature as sections of the time-series with different statistical properties, such as means or variances (Killick et al., 2012). In this case, θ would be the the vector of mean recruitments. The *changepoint* package in R assumes that the natural logarithms of the observations are normally distributed. The start of subsequent change points is detected by conditioning the cost function on the data from the start of previous change point to the end of the series. By working backwards, the PELT algorithm "prunes" potential points from the dataset that are after the previous change point, reducing the computational time (Killick et al., 2012). The minimum regime duration was set at seven years, to capture subdecadal variation. The PELT algorithm can detect changepoints using a variety of statistical criteria such as Akaike information criterion (AIC) and the Bayesian Information Criterion (BIC) (Killick and Eckley, 2014). We used the AIC_C given its widespread use in statistical literature. The full function to minimize is:

$$\sum_{i=1}^{m+1} [C(y_{(\tau_{i-1}+1):\tau_i})] + 2p(\frac{n}{n-p-1}) \tag{2.9}$$

where m is the number of changepoints, which splits the data into m + *1* segments, τ_i are the locations of the change points in the time series, $C(y_{(\tau_{i-1}+1):\tau_i})$ is the cost function in Eqn 4, and $2p(\frac{n}{n-p-1})$ is the AIC_C penalty function, where p is the number of parameters introduced to the likelihood by the addition of a new changepoint and n is the total number of recruitment observations. Without a penalty function, the algorithm would consider any stochasticity in the data as a regime change. The *changepoint* package does not include AICC as a default penalty function, so it was calculated manually.

The forecasting procedure works as follows. First, the PELT algorithm is run on the training data. If a regime shift is detected, the training data is split at the location of the regime change. The mean and standard deviation are calculated for the recruitment observations in the most recent regime shift. The predicted recruitment is sampled then sampled from a normal distribution $N(\mu_{i:j}, \sigma_{i:j})$, where *i:j* indicates the years that fall into the most recent regime shift. If no regime shift is detected in the training data, the sampling procedure is equivalent to the average recruitment with some process noise.

2.3.3 Evaluation approach

The time-series for each stock is divided into a training and test set. The initial training set for each stock was the first 19 years of the time series, which was sufficiently long enough to fit the forecast methods and leave enough years to forecast. The training set is used to fit each of the forecast methods. Predictions are made once the forecast model is fit, and these are compared to the recruitment in the target years in the test data set. The data for target years from the test set are then included in the training set, and the forecast models are refit. Recruitment predictions are made in this way until the end of the test data set is reached. Figure 2.4 depicts this 'expanding window' approach to forecasts. Each model run generates a series of short (1-year) and mid-term (5-year) predictions for the test data. The forecast procedure is repeated 1,000 times starting at the initial year of test data to generate confidence intervals for the short- and mid-term forecasts by year. The percentage of times the observed value of recruitment fell within the inter-quartile range was used as a measure of skill for each forecast method.

$$coverage\, probability = \frac{R_t\, in\, 95\%\, CI}{number\, of\, years\, in\, forecast} \tag{2.10}$$

We also calculated the mean absolute scaled error (MASE) of the forecasts. MASE is the ratio of the mean absolute error of the forecasted recruitment to the lag-1 difference in the recruitment training data (Equation 2.11). Unlike other common accuracy measures, such as root mean squared error and mean absolute error, MASE is scale independent. Scale independence allows which us to compare forecast accuracy across stocks (Hyndman and Koehler, 2006). MASE is also easily interpreted and insensitive to outliers (DeFilippo et al., 2021; Hyndman and Koehler, 2006; Ward et al., 2014).

$$MASE_{i,y} = \frac{\frac{1}{n}\sum_{1}^{n}|P_{i,y} - R_{i,y}|}{\frac{1}{m-1}\sum_{1}^{m}|R_{i,t} - R_{i,t-1}|} \tag{2.11}$$

The MASE for stock i in year y is calculated by first calculating the mean absolute error between the forecasted recruitment in year y, $P_{i,y}$, and the "observed" recruitment in year y, $R_{i,y}$, across the simulations, n. The denominator is the mean absolute difference ($R_{i,t} - R_{i,t-1}$) between the recruitment in the years (t) of training period of length m years. With a sufficiently long enough time series, a random walk model without drift would converge to one (Ward et al., 2014). Ideally, the MASE values for a given method would be less than one, indicating the method performs better than a random walk. However, the recruitment time series are short and a MASE value above one would not be unusual.

We generated a MASE value for each year of a stock's test data and forecast method. The distribution of MASE values by method were compared to determine accuracy of each method among stocks. To compare MASE values more easily across stocks and time, we divided MASE values into three periods. The first 10 years of forecasts for each stock were designated as the 'early' forecast period. The 'mid' forecast period represented forecast years 11 to 20, and the 'late' forecast period represented any forecasts beyond 20 years. Ideally, with more training data, we would see MASE values decrease over time. This would indicate that forecast performance improved with additional data. We also expect MASE values to be higher for longer-term forecasts, as predicting recruitment accurately farther into the future is more difficult than predicting recruitment in the short term.

2.3.4 Stock characteristics

The performance of each of the forecast methods was examined in relation to several characteristics of the stock time series. The characteristics considered were (1) the length of the time series, (2) the spawning biomass depletion ratio, (3) the recruitment autocorrelation and (4) standard deviation across the total time series, (5) the presence of a detectable influence of spawning biomass on recruitment (a stock-recruitment relationship), and (6) the presence of regime shifts. We used the procedure developed in Szuwalski et al. (2015), to test for the presence of a stock-recruitment relationship. In this procedure, a stock is considered to have a detectable relationship of spawning biomass on recruitment if (1) the zero-lagged Spearman's correlation coefficient (Spearman, 1987) is significant and (2) the coefficients for the negative lags are less than the zero-lagged coefficient. The PELT algorithm was used to determine if any of the stocks underwent a regime shift in their recruitment time series.

2.4 Results

2.4.1 Example species

For illustration, we will walk through the forecast procedure for a single Alaskan stock, the Gulf of Alaska (GOA) Pacific ocean perch, the recruitment time series for which extends from 1977 to 2020. The initial training set for short-term forecasts was 1977 to 1996 and the initial training set for mid-term forecasts was 1977 to 1993. The training set for the mid-term forecasts is shorter to have the first and last year of predictions for both short- and mid-term forecasts be 1997 and 2020, respectively. The six forecast methods were fit to the training set and 1,000 recruitment predictions were generated for each method. Using the expanding window approach, the initial test year was added to the training set and the forecast procedure was repeated. The 1,000 short- and mid-term recruitment predictions for each year were used to determine the 80% and 95% simulation intervals (Figs. 2.5 and 2.6). The coverage probability was then used to evaluate the performance of each forecast method. The coverage probability was calculated by dividing the number of times the observed recruitment for year t fell within the 95% simulation intervals for the forecasted recruitment for year t by the total number of test years. Ideally, coverage probabilities would be 0.95, but any value between above or equal to 0.8 would be considered adequate for the purposes of stock assessments. Coverage probabilities lower than 0.8 indicated that the forecast method was not capturing the recruitment dynamics, while coverage probabilities in excess of 0.957 indicate that the uncertainty around the recruitment forecasts is too high. All the methods except for simplex projection fell within the ideal range for short-term recruitment forecasts (Fig. 2.7a). However, only the mean recruitment, Beverton-Holt, and PELT sampling methods fell within the ideal range for mid-term forecasts (Fig. 2.7b). MASE values were highest for HMM sampling in both short- and mid-term forecasts (Fig. 2.8a-b). The distribution of MASE values for the other five methods were similar for both forecast lengths, and all were larger than 1, indicating that none of the methods performed better than better than a naïve forecast using the training data.

2.4.2 Overall

There was no clear best recruitment forecast method across all the stocks as different methods performed better depending on the forecast period (i.e., short or mid-term), management region, and stock characteristics. Figure 2.9 shows the frequency each of the methods fell into the ideal coverage probability range across all stocks for both short- and mid-term forecasts. AR(1) had the highest frequency (0.583) of producing short-term forecasts that fell within the “ideal” coverage probability range. The PELT sampling method and simplex projection had the highest frequency for the mid-term forecasts (0.556 and 0.472, respectively). There were differences in performance among management regions. AR(1) was the best performing method for stocks in the Bering Sea and Aleutian Islands (BSAI) region (Fig. 2.9b), but the frequency of the coverage probability for mid-term AR(1) forecasts being within the ideal range was 0 for the Gulf of Alaska region (GOA) (Fig. 2.9c). The PELT sampling method performed particularly well for short- and mid-term forecasts for the West Coast stocks (frequencies of 0.631 and 0.526, respectively; Fig. 2.9d).

The MASE values differ by stock, forecast length, and method. We found that median MASE value across stocks and method decreased as the forecast period increased, although the difference was slight - the median 1-step MASE values for each forecast period were 1.22, 1.18, and 1.14 for the early, mid, and late periods, respectively. The median 5-step MASE values showed a similar trend (1.28, 1.21, 1.19 for early, mid, and late, respectively). The median MASE values were more variable within each forecast method (Fig. 2.10). For all stocks, the frequency of outliers increased during the late forecast period. We calculated the frequency of MASE values less than 1 by region and method to evaluate how well each forecast method performed against the in-sample naïve forecast (Fig. 2.11). AR(1) and PELT sampling had some of the highest frequencies of MASE values less than 1 across each of the regions (Fig. 2.11).

Although there is no best performing method, there is one method that performed badly across the majority of stocks. Hidden Markov sampling had the lowest frequency of coverage probabilities falling within the "ideal" range for short- and mid-term forecasts for the Gulf of Alaska and the West Coast (Fig. 2.9c-d). HMM sampling also had the widest distribution of median MASE values across stocks regardless of the forecast period (Fig. 2.10) and as such, had the lowest frequency of MASE values less than 1 (Fig. 2.11).

2.4.3 Stock characteristics

Time-series length had a slight negative correlation with coverage probability for most of the methods (Fig. 2.12). This trend was more pronounced for mid-term forecasts than short-term forecasts. There was a slight improvement in the coverage probability of HMM for short-term forecasts for stocks that have longer time series. A longer time series may increase the likelihood that a stock experiences a regime shift, improving the performance of HMM sampling compared to the other methods. Like time-series length, there was a negative correlation between depletion and coverage probability (Fig. 2.13). Short- and mid-term coverage probabilities were better for stocks with high levels of depletion and worse for stocks with low levels of depletion (Fig. 2.13). The one exception was PELT sampling, which performed better in short- and mid-term forecasts when depletion was low.

Recruitment standard deviation had a slight positive correlation with coverage probability for a majority of methods (Fig. 2.14). AR(1) mid-term forecasts in particular performed better for stocks with a higher recruitment standard deviation. Recruitment autocorrelation appeared to have the largest impact on coverage probability. Low coverage probabilities were correlated with high recruitment autocorrelation for most of the methods (Fig. 2.15), especially for mid-term forecasts. Simplex projection was the exception and performed well across a range of recruitment autocorrelation values for mid-term forecasts.

The performance of the methods was also evaluated in the context of detectable stock-recruitment relationships and recruitment regime shifts. Only two methods, the Beverton-Holt equation and HMM sampling assumed a relationship between spawning biomass and recruitment. For stocks without a detectable stock-recruitment relationship (27 stocks out of 36), the Beverton-Holt stock-recruitment relationship performed similarly to other methods in short- and mid-term forecasts (Fig. 2.16). Six of the thirty-six stocks had some evidence for the influence of spawning biomass on recruitment, but not enough to determine this for certain. For these stocks, the Beverton-Holt method performed well (Fig. 2.16). For the three stocks with a detectable stock-recruitment relationship, the Beverton-Holt method performed especially well for mid-term forecasts (Fig. 2.16b), but similarly to others for short-term forecasts. This may be due to the smaller sample size compared to the stocks without a detectable stock-recruitment relationship. HMM sampling only performed well for stocks with some evidence of a stock-recruitment relationship when predicting one year ahead (Figure 2.16a). Of the thirty-six stocks, twelve had evidence of a recruitment regime shift. Two of the forecast methods tested incorporated changes in average recruitment across the time series: HMM sampling and PELT sampling. However, HMM sampling was the worst performing method, even for stocks with evidence of a regime change (Fig. 2.17). PELT sampling had a lower median coverage probability for stocks with a regime change than the other methods (Fig. 2.17). This poorer performance is particularly evident for short-term forecasts (Fig. 2.17a). This discrepancy in performance may be due to the location of the regime change. Eight of the twelve stocks with at least one regime shift detected had a regime shift within the first 19 years of the recruitment time series. The PELT sampling procedure appeared to perform better for those eight stocks than for the four stocks with later regime shifts (Fig. 2.18).

2.5 Discussion

Our goal was to evaluate a suite of forecast methods for recruitment that vary in operational use across regions and complexity. We evaluated the performances of the methods in terms of short-term (1-step) and mid-term (5-step) predictions. Our results suggest that there is no best method among the stocks evaluated, which is consistent with previous research (Van Beveren et al., 2021). Of the six methods we evaluated, the AR(1) method and PELT sampling generally performed well on the two metrics (Figs. 2.9 and 2.11) and simplex projection had consistently high coverage probabilities (Fig. 2.9). However, forecast accuracy depended on forecast length, stock characteristics, and region. Of the stock characteristics we assessed, recruitment autocorrelation had the largest impact on coverage probabilities (Fig. 2.15).

HMM sampling was the worst performing method across the Gulf of Alaska and the West Coast for both metrics. AR(1) performed best when used for short-term forecasts for all of the regions (Figs. 2.9 and 2.11). The frequency with which AR(1) forecasts fell within the ideal coverage probability range decreased for mid-term forecasts, especially for stocks in the Gulf of Alaska (Fig. 2.9c). The coverage probabilities for the AR(1) method were robust to recruitment autocorrelation in short-term forecasts (Fig. 2.15a). However, we advise against using AR(1) for mid-term forecasts when autocorrelation is high, as there is a marked decrease in coverage probability(Fig. 2.15b). Increasing autocorrelation can lead to higher prediction error (Johnson et al., 2016; Van Beveren et al., 2021), even when recruitment autocorrelation is correctly specified (Johnson et al., 2016). Time series methods, including AR(1), have conflicting performance among previous studies, especially when environmental effects are included (Subbey et al., 2014). The discrepancy between our results and those of previous studies may be due to the life history of the stocks we evaluated because differing reproductive strategies between short- and long-lived fish influences the effectiveness of some recruitment forecasting methods (Haltuch et al., 2019).

PELT sampling and simplex projection are more appropriate than AR(1) for mid-term recruitment forecasts. These methods had the highest frequency of coverage probabilities falling within the ideal range for mid-term forecasts and had smaller discrepancies in coverage probability performance for short- and mid-term forecasts when compared to AR(1) (Fig. 2.9). PELT sampling and simplex projection differed in their performance given varying levels of spawning biomass depletion and recruitment autocorrelation. Specifically, coverage probabilities for the PELT sampling method increased with lower stock depletion rates while simplex projection had a slight negative trend (Fig. 2.13). Simplex projection was the method most robust to high recruitment autocorrelation, especially for mid-term forecasts (Fig. 2.15). In contrast, PELT sampling did not perform well for stocks with high autocorrelation. Both methods had larger median MASE values than the AR(1) method for the early and middle forecast periods, but all three methods performed similarly during late forecast periods (Fig. 2.10). Larger prediction errors are common with sampling procedures and EDM techniques (Munch et al., 2018; Van Beveren et al., 2021). EDM prediction error has been shown to increase for long-lived stocks (Munch et al., 2018), which our study is limited to. PELT may be a useful method for stocks with evidence of a recruitment regime shift early in their time series. However, PELT sampling coverage probabilities were lower when the first identified regime shift is late in the time series (Fig. 2.18). This decrease in performance is likely due to the smaller sample size of recruitment points in the new regime from which to sample. Simplex projection also performed well for stocks with evidence of a recruitment regime shift. Previous research found EDM techniques had advantages over traditional stock-recruitment models when stocks underwent an abrupt change in productivity (Sguotti et al., 2020). What might appear to be regimes could also be the result of non-linear dynamics, which EDM techniques can account for in forecasts (Munch et al., 2018).

There are several limitations to this study. The metrics used to evaluate forecasts vary across the literature. Coverage probability and MASE have been used in fisheries before (DeFilippo et al., 2021; Johnson et al., 2016), but neither are used universally, and determining a suite of performance metrics would aid in comparing forecast methods among studies. Also, we evaluated forecast performance only in the context of predicting the recruitment time series. Other studies have tested the performance of recruitment forecasts within age-structured models (Van Beveren et al., 2021) and management strategy evaluation (MSE) (A'mar et al., 2009; Haltuch and Punt, 2011; Szuwalski and Punt, 2013; Wayte, 2013), which may provide more insight into the operational effectiveness of each forecast method. As recruitment time series are model outputs and can inherit quirks from model assumptions (Dickey-Collas et al., 2015), further research would be necessary to determine if the methods we tested are robust in an operational setting.

While there is no silver bullet for recruitment forecasting, we have found that evaluating stock characteristics and the desired forecast period can provide some guidance on an appropriate method. Ultimately, the best method will provide "valuable" recruitment forecasts which aid in the decision making of managers and stakeholders (Kiaer et al., 2021). Ideally, these methods will be robust to changes in stock productivity, as regime shifts are likely to increase with climate change.

TABLES

Table 2.1: Stocks included in this analysis. The stocks represent two regional management councils, the North Pacific Fishery Management Council (NPFMC) and the Pacific Fishery Management Council (PFMC). The stock regions are the Bering Sea and Aleutian Islands (BSAI), and the Gulf of Alaska (GOA).

Common name	Species name	Management council	Region	Data category	Years included
Arrowtooth flounder	*Atheresthes stomias*	NPFMC	BSAI	Tier 3	1976-2019
Atka mackerel	*Pleurogrammus monopterygius*	NPFMC	BSAI	Tier 3	1977-2020
Blackspotted and Rougheye rockfish complex	*Sebastes melanostictus, Sebastes aleutianus*	NPFMC	GOA	Tier 3	1977-2021
Dusky rockfish	*Sebastes variabilis*	NPFMC	GOA	Tier 3	1979-2020
Greenland turbot	*Reinhardtius hippoglossoides*	NPFMC	BSAI	Tier 3	1980-2020
Kamchatka flounder	*Atheresthes evermanni*	NPFMC	BSAI	Tier 3	1991-2020
Northern rock sole	*Lepidopsetta polyxystra*	NPFMC	BSAI	Tier 1	1975-2020

Table 2.1: Stocks included in this analysis. The stocks represent two regional management councils, the North Pacific Fishery Management Council (NPFMC) and the Pacific Fishery Management Council (PFMC). The stock regions are the Bering Sea and Aleutian Islands (BSAI), and the Gulf of Alaska (GOA). *(continued)*

Common name	Species name	Management council	Region	Data category	Years included
Northern and southern rock sole	*Lepidopsetta polyxystra, Lepidopsetta bilineata*	NPFMC	GOA	Tier 3	1977-2021
Northern rockfish	*Sebastes polyspinis*	NPFMC	GOA	Tier 3	1977-2020
Pacific cod	*Gadus macrocephalus*	NPFMC	Eastern Bering Sea	Tier 3	1977-2020
Pacific cod	*Gadus macrocephalus*	NPFMC	GOA	Tier 3	1977-2019
Walleye pollock	*Gadus chalcogrammus*	NPFMC	Eastern Bering Sea	Tier 1	1964-2020
Walleye pollock	*Gadus chalcogrammus*	NPFMC	GOA	Tier 3	1977-2020
Pacific Ocean perch	*Sebastes alutus*	NPFMC	BSAI	Tier 3	1977-2017
Pacific Ocean perch	*Sebastes alutus*	NPFMC	GOA	Tier 3	1977-2020
Sablefish	*Anoplopoma fimbria*	NPFMC	BSAI	Tier 3	1977-2018
Yellowfin sole	*Limanda aspera*	NPFMC	BSAI	Tier 1	1954-2020
Aurora rockfish	*Sebastes aurora*	PFMC	Pacific west coast	Data-rich	1963-2012

Table 2.1: Stocks included in this analysis. The stocks represent two regional management councils, the North Pacific Fishery Management Council (NPFMC) and the Pacific Fishery Management Council (PFMC). The stock regions are the Bering Sea and Aleutian Islands (BSAI), and the Gulf of Alaska (GOA). *(continued)*

Common name	Species name	Management council	Region	Data category	Years included
Black rockfish	*Sebastes melanops*	PFMC	California	Data-rich	1963-2014
Black rockfish	*Sebastes melanops*	PFMC	Washington	Data-rich	1963-2014
Boccacio	*Sebastes paucispinis*	PFMC	Pacific west coast	Data-rich	1960-2015
Cabezon	*Scorpaenichthys marmoratus*	PFMC	Northern Californian slope	Data-rich	1962-2018
Cabezon	*Scorpaenichthys marmoratus*	PFMC	Oregon	Data-rich	1980-2018
Cabezon	*Scorpaenichthys marmoratus*	PFMC	Southern California slope	Data-rich	1975-2018
Canary rockfish	*Sebastes pinniger*	PFMC	Pacific west coast	Data-rich	1960-2014
Chilipepper rockfish	*Sebastes goodei*	PFMC	Pacific west coast	Data-rich	1966-2014
Darkblotched rockfish	*Sebastes crameri*	PFMC	Pacific west coast	Data-rich	1975-2016
Dover sole	*Microstomus pacificus*	PFMC	Pacific west coast	Data-rich	1975-2020
Kelp greenling	*Hexagrammos decagrammus*	PFMC	Pacific west coast	Data-rich	1980-2014

Table 2.1: Stocks included in this analysis. The stocks represent two regional management councils, the North Pacific Fishery Management Council (NPFMC) and the Pacific Fishery Management Council (PFMC). The stock regions are the Bering Sea and Aleutian Islands (BSAI), and the Gulf of Alaska (GOA). *(continued)*

Common name	Species name	Management council	Region	Data category	Years included
Lingcod	*Ophiodon elongatus*	PFMC	Northern Pacific west coast	Data-rich	1960-2020
Lingcod	*Ophiodon elongatus*	PFMC	Southern pacific west coast	Data-rich	1970-2020
Petrale sole	*Eopsetta jordani*	PFMC	Pacific west coast	Data-rich	1950-2018
Sablefish	*Anoplopoma fimbria*	PFMC	Pacific west coast	Data-rich	1960-2020
Splitnose rockfish	*Sebastes diploproa*	PFMC	Pacific west coast	Data-rich	1975-2008
Widow rockfish	*Sebastes entomelas*	PFMC	Pacific West coast	Data-rich	1965-2018
Yelloweye rockfish	*Sebastes ruberrimus*	PFMC	Pacific west coast	Data-rich	1960-2016

FIGURES

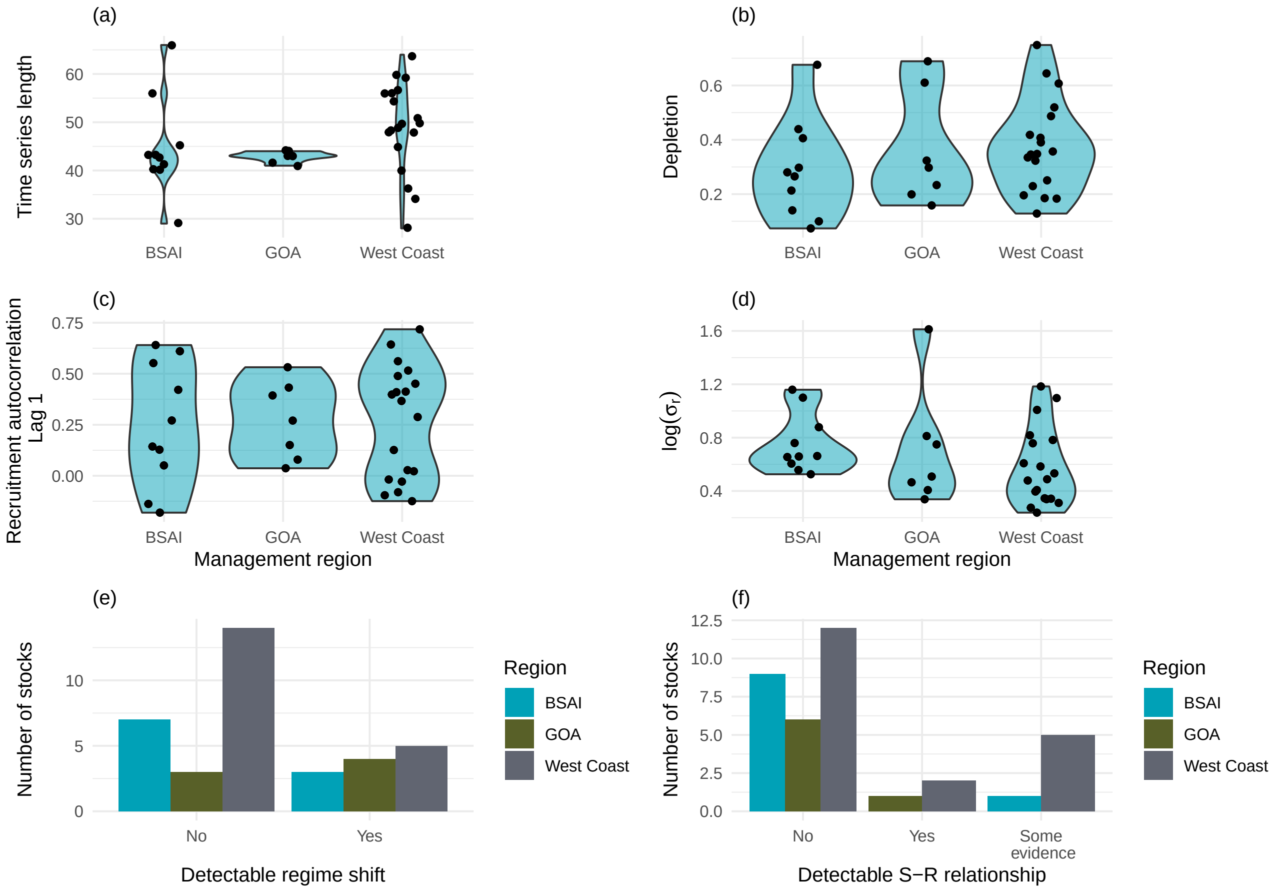

Figure 2.1: Comparison of time series characteristics for each management region. Panels depict (a) the length of each stock's recruitment time series, (b) the stock's historic spawning biomass depletion, (c) the recruitment autocorrelation at lag = 1 for each stock, (d) the log standard deviation of each stock's recruitment time series, (e) the number of stocks with a detectable regime shift, and (f) the number of stocks with a detectable stock-recruit relationship. Depletion (panel b) is defined as the 5^{th} spawning biomass quantile divided by the 95^{th} spawning biomass quantile. Regime shifts were detected using the PELT changepoint algorithm (panel e). Stock-recruitment relationships were determined using the method outlined in Szuwalski et al. (2015).

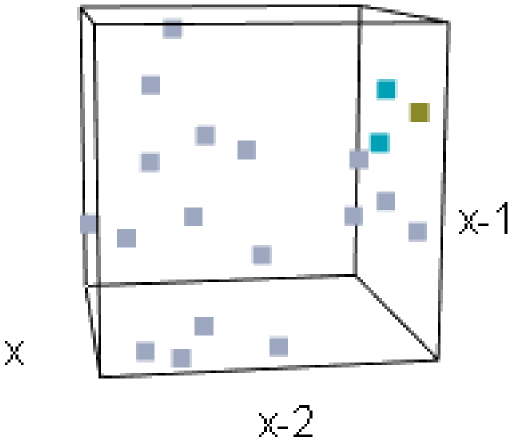

Figure 2.2: Transformation of a time series into E-dimensional points. The prediction for the target point (green) uses the nearest-neighbor E-dimensional points (blue).

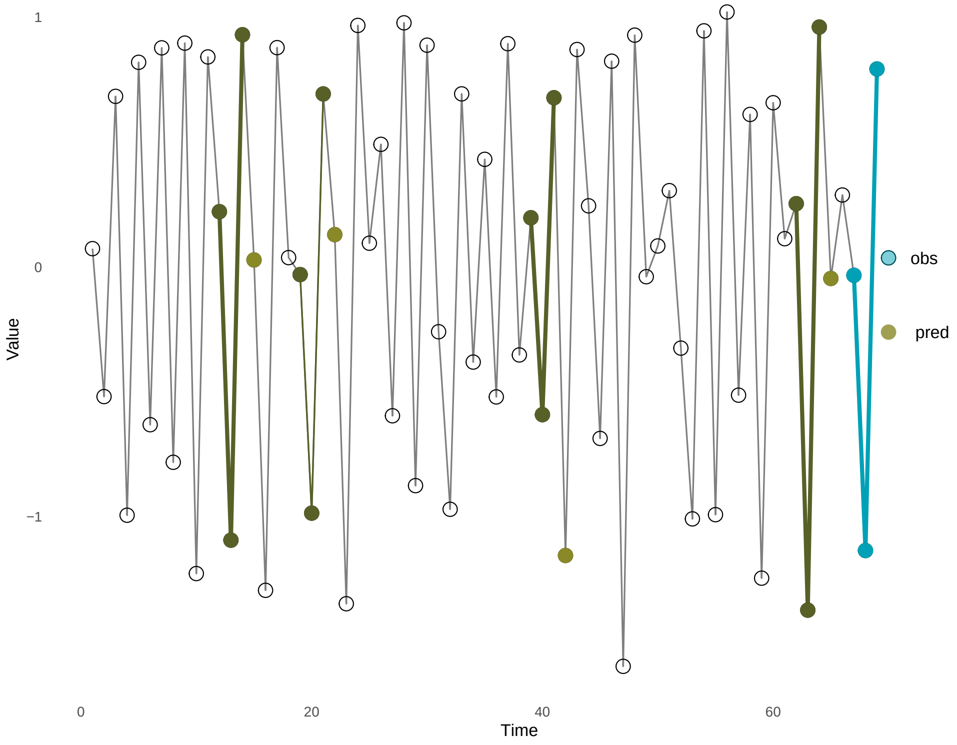

Figure 2.3: 2-dimensional representation of simplex projection. First, the embedding dimension for the time series is identified. In this figure, E = 3. To make a prediction for the observation point, labeled "obs", the data points for the previous three years, colored in blue, are considered. The most similar segments in the time series are identified by calculating the distance from the time series segment (blue points) preceding the observation point ("obs"). The "nearest-neighbor" segments are highlighted in dark green. The value for the prediction is calculated using the weighted average of the fourth points in each sequence in the identified segments, each colored green.

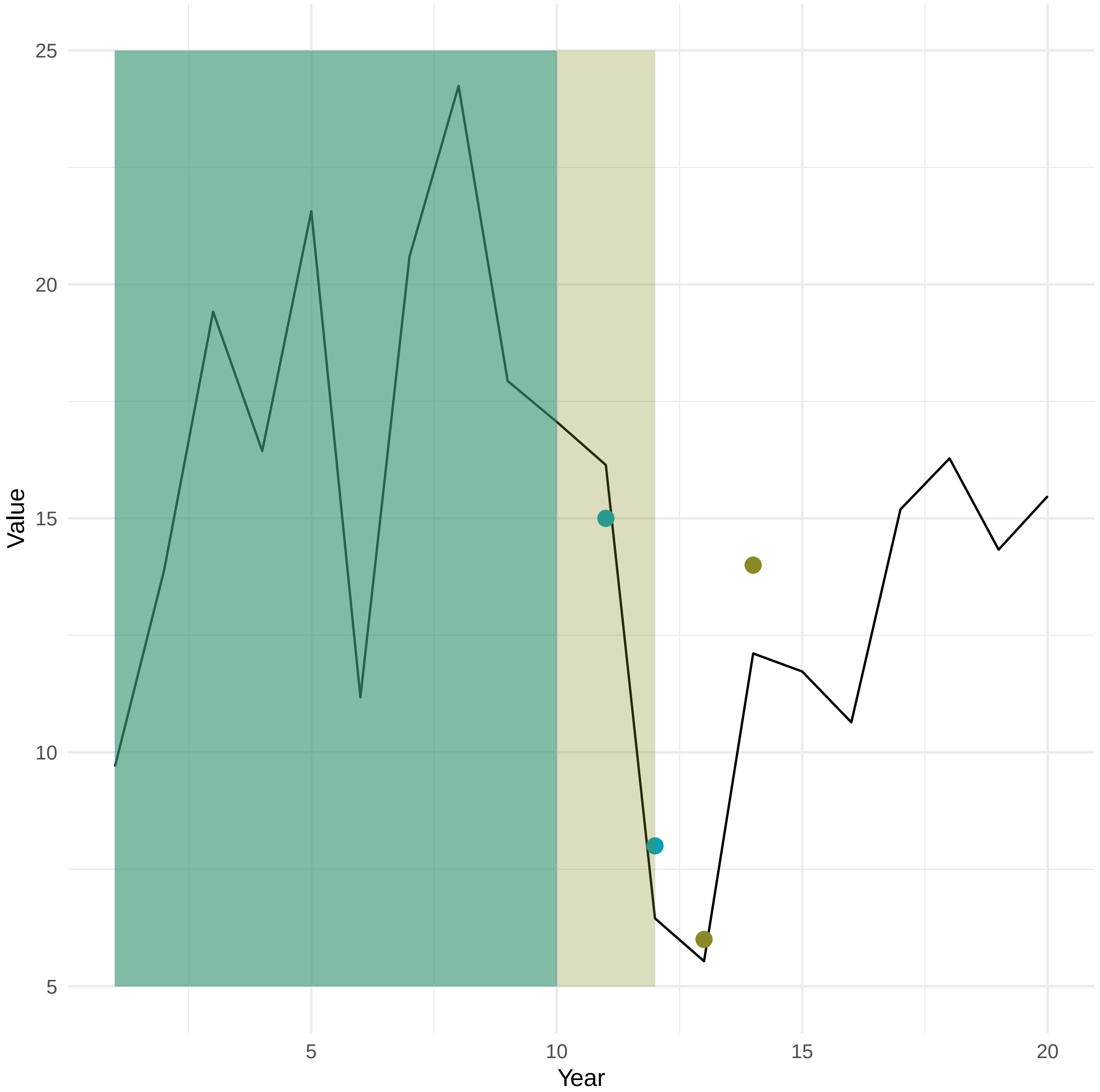

Figure 2.4: A representation of the expanding window forecast approach. The shaded dark green area indicates the initial training set and the dark green points the predictions made using the model fit using the data for this region. After making predictions, the target years are included in the training set. The shaded light green area represents the second training set. The light green points represent the predictions made using the model fit on the second training set.

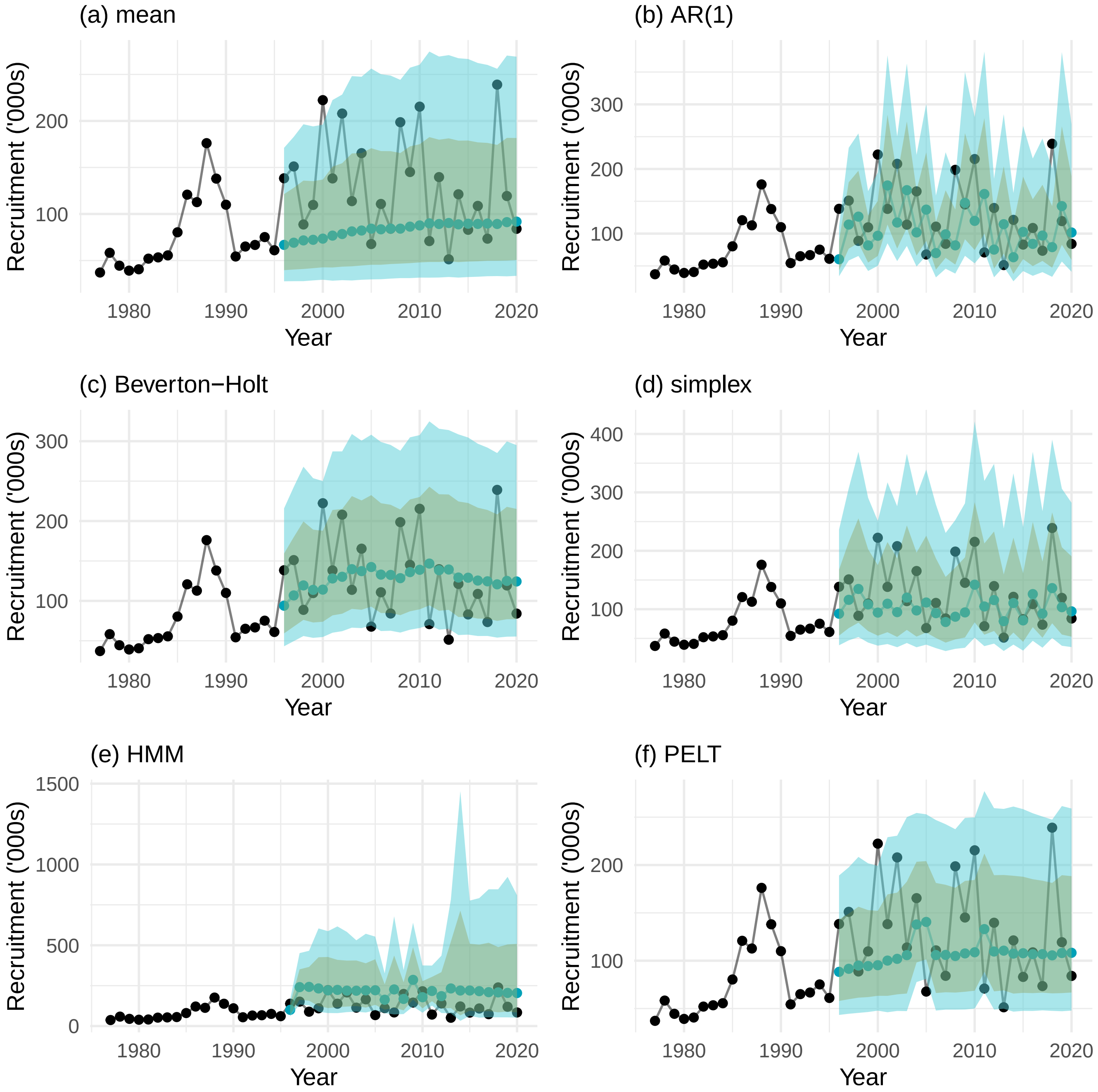

Figure 2.5: Short-term (1-year) forecasts for Gulf of Alaska Pacific ocean perch. Black points represent the recruitment time series. Blue points represent median forecasted recruitment. The blue shading represents the 95% simulation confidence intervals. The green shading represents the 80% simulation confidence intervals. Note that the y-axis differs among panels.

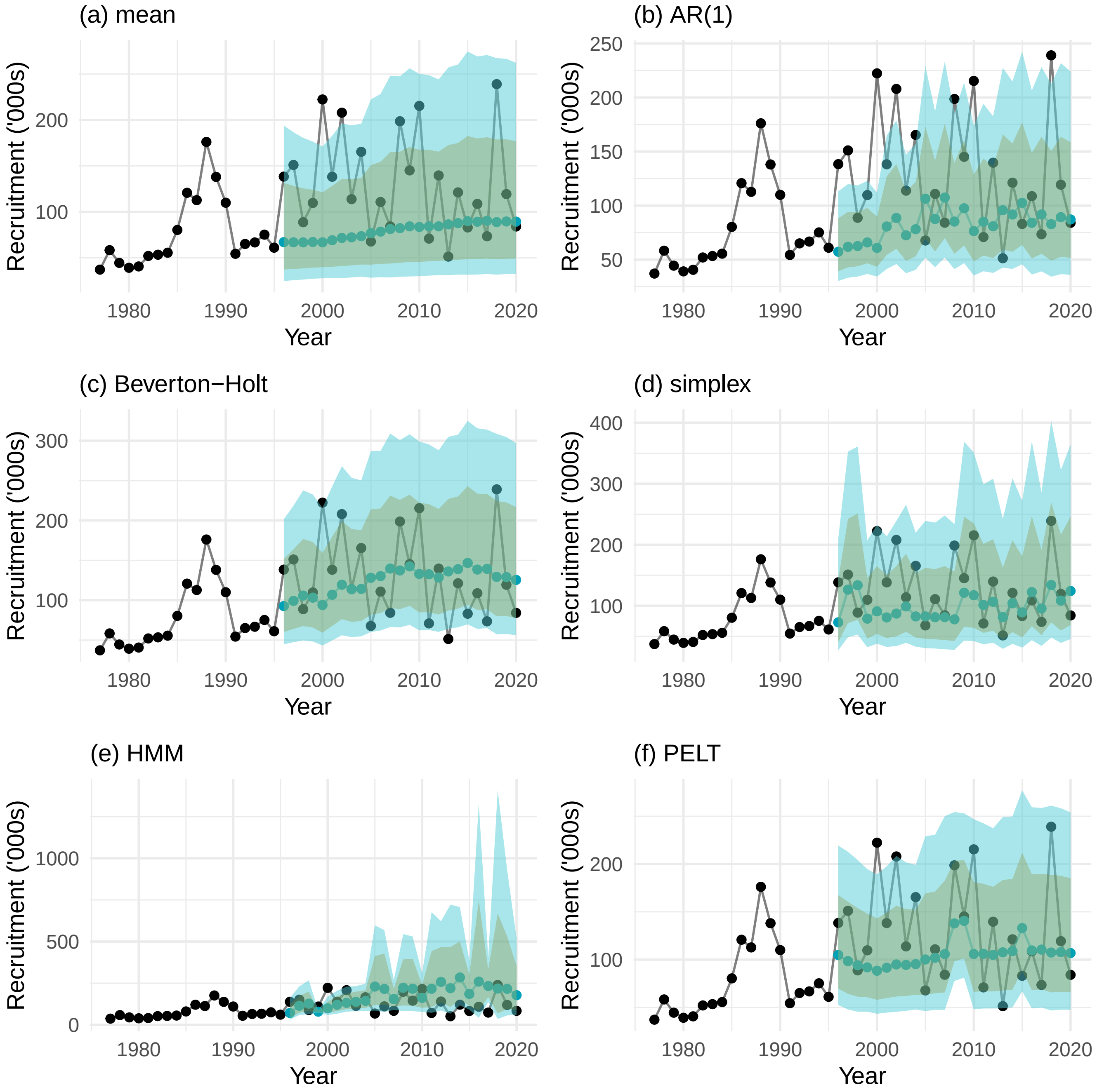

Figure 2.6: Mid-term (5-year) forecasts for Gulf of Alaska Pacific ocean perch. Black points represent the recruitment time series. Blue points represent median forecasted recruitment. The blue shading represents the 95% simulation confidence intervals. The green shading represents the 80% simulation confidence intervals. Note that the y-axis values differ among panels.

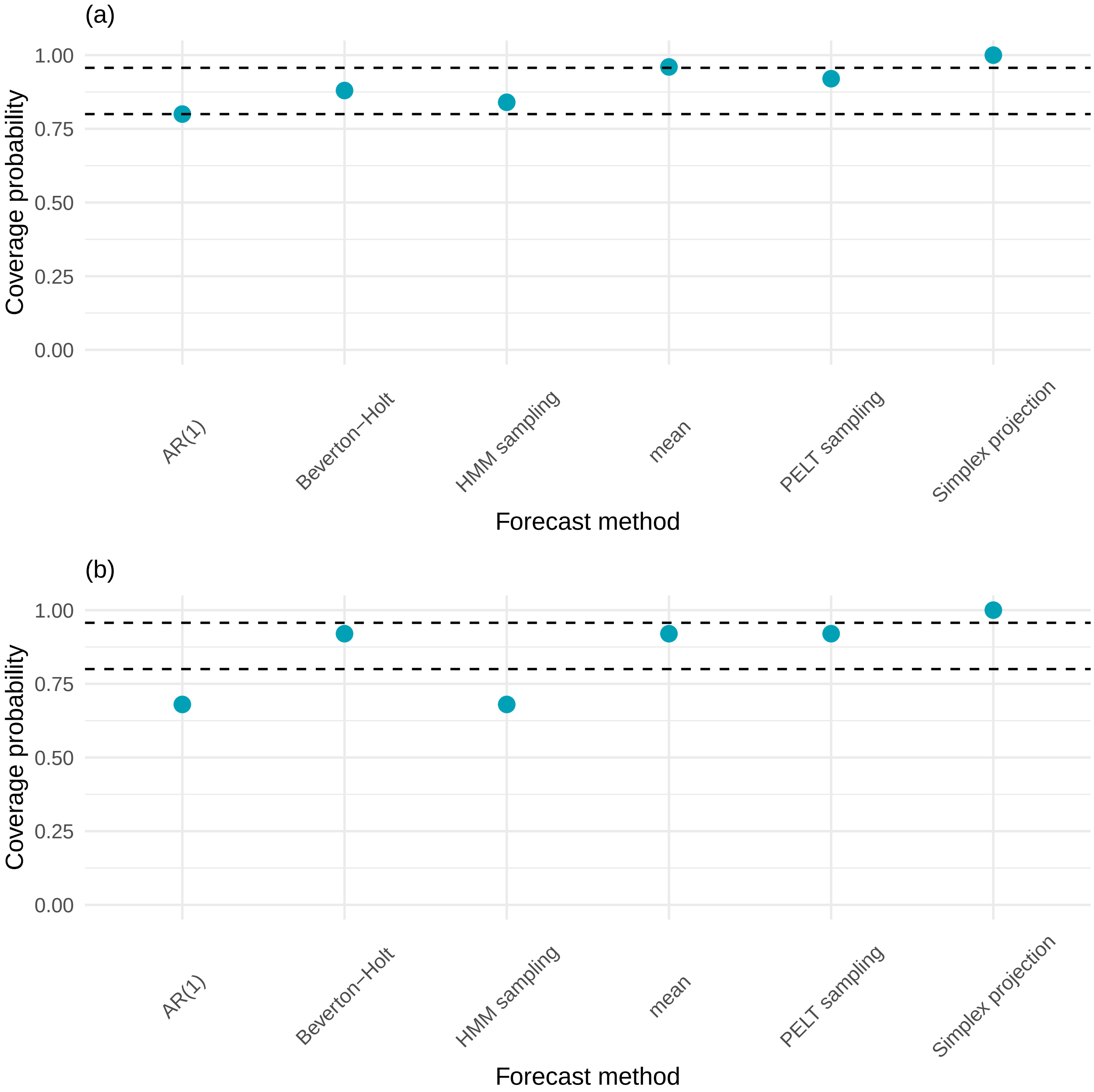

Figure 2.7: Gulf of Alaska Pacific ocean perch coverage probabilities for the six forecasting methods. (a) Short-term forecasts and (b) mid-term forecasts. Dotted lines represent the "ideal" range for coverage probabilities, between 0.8 and 0.957.

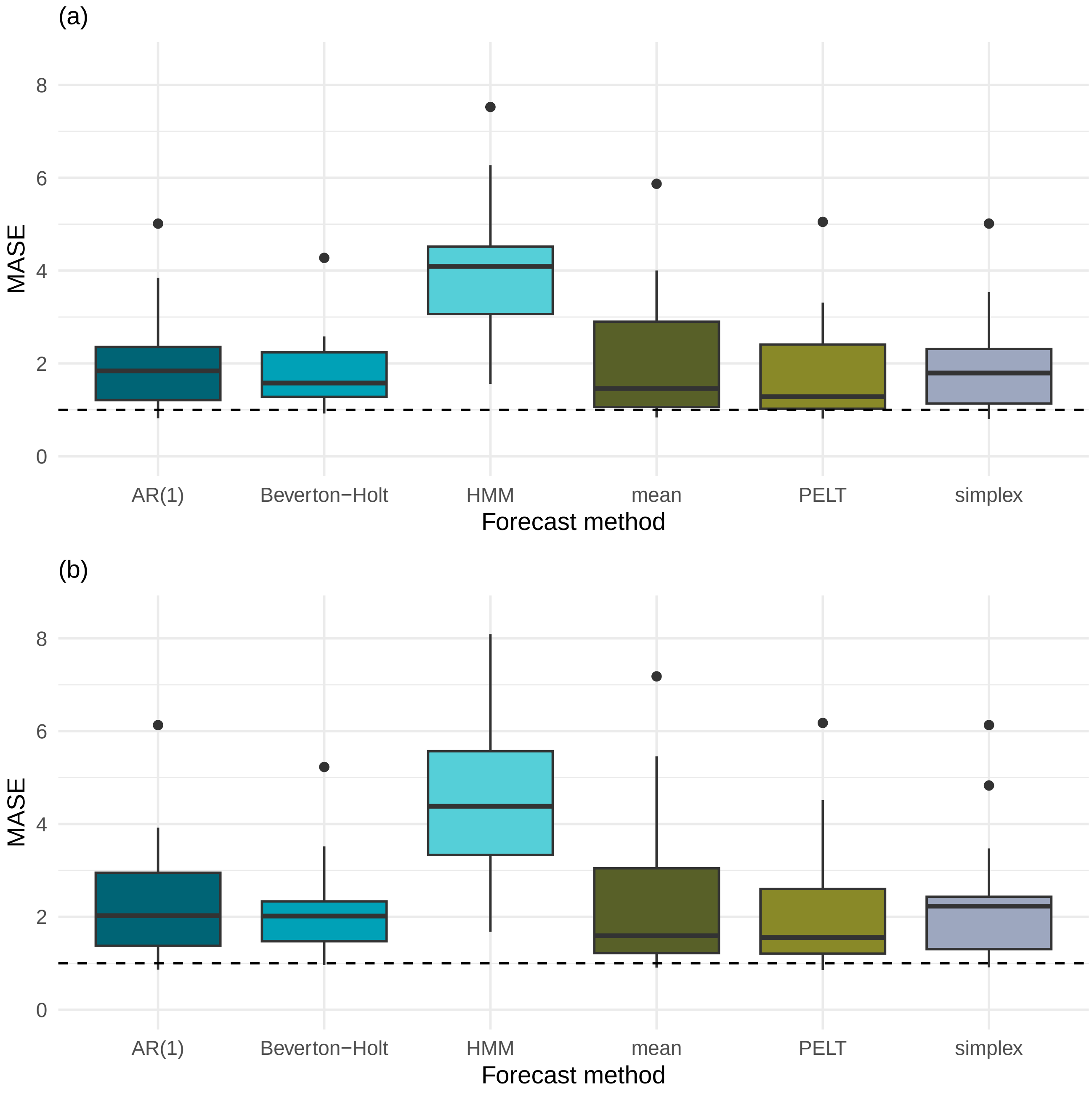

Figure 2.8: Boxplots of MASE values by method for short- (a) and mid-term (b) recruitment forecasts for Gulf of Alaska Pacific ocean perch. Dotted lines represent a MASE value of 1, where the MAE of the forecast method is equal to the in-sample naïve MAE.

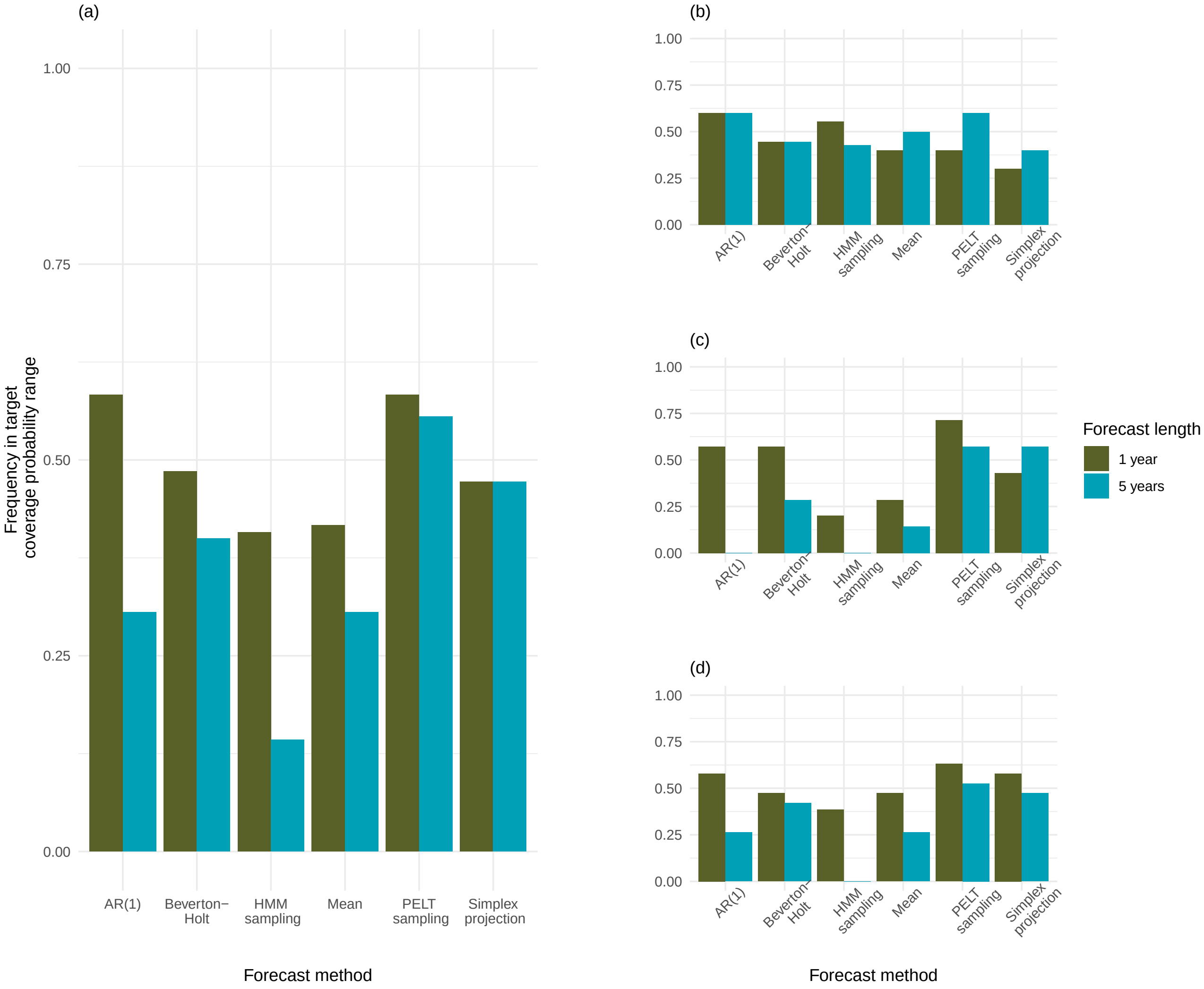

Figure 2.9: Frequency with which each method fell within the ideal coverage probability range for the short- (green) and mid-term forecasts (blue). (a) Frequency for all stocks included in analysis, (b) frequency for stocks in Bering Sea and Aleutian Islands management region, (c) frequency for stocks in Gulf of Alaska management region, and (d) frequency for stocks in the West Coast management region.

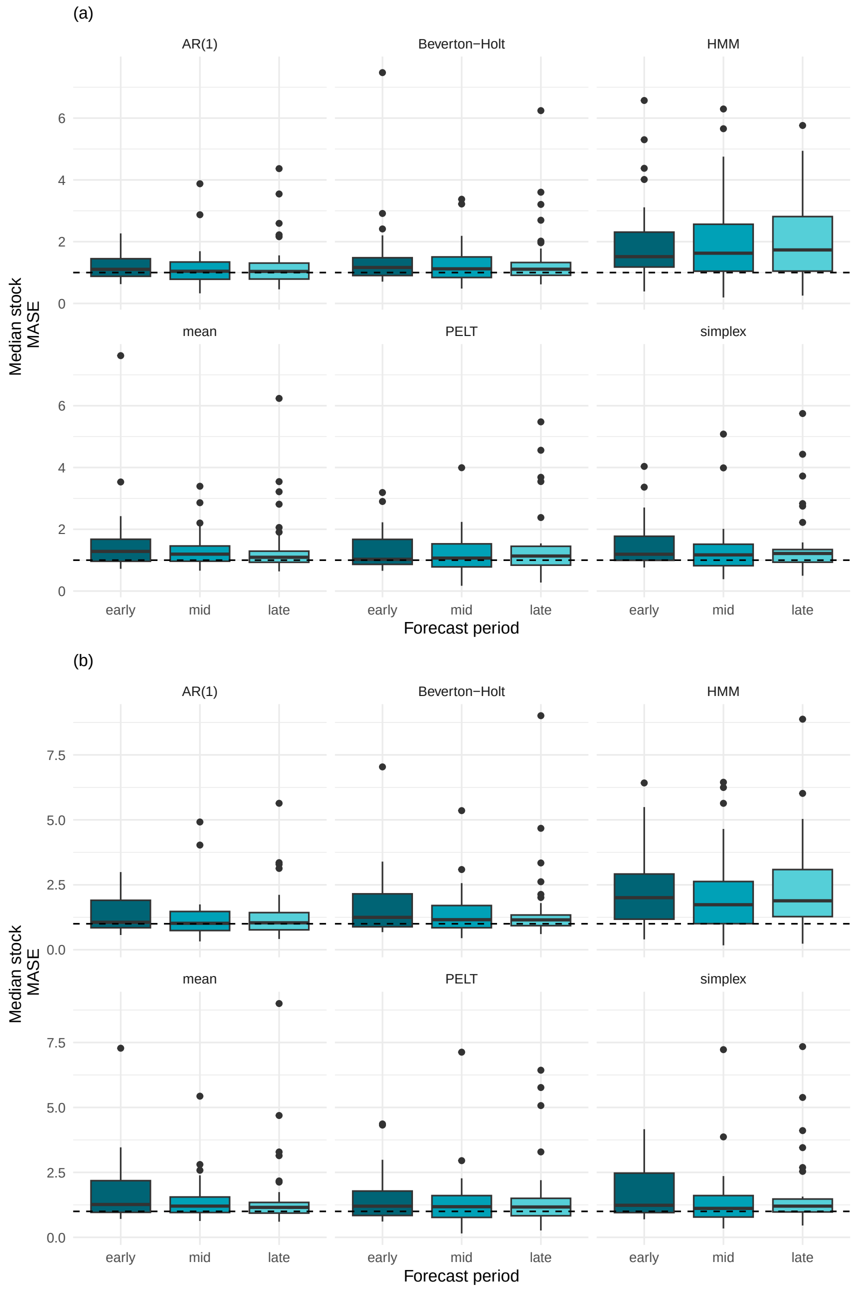

Figure 2.10: Boxplots of the median (over stocks) MASE value by forecast period and forecast method. Forecast periods are defined as early (first 10 years of the forecast period), mid (years 11 to 20 of the forecast period), and late (beyond 20 years). Panel (a) represents median MASE values for short-term (1-step) forecasts and panel (b) represents median MASE values for mid-term forecasts (5-step). Dotted lines represent a MASE value of 1, where the MAE of the forecast method is equal to the in-sample naïve MAE.

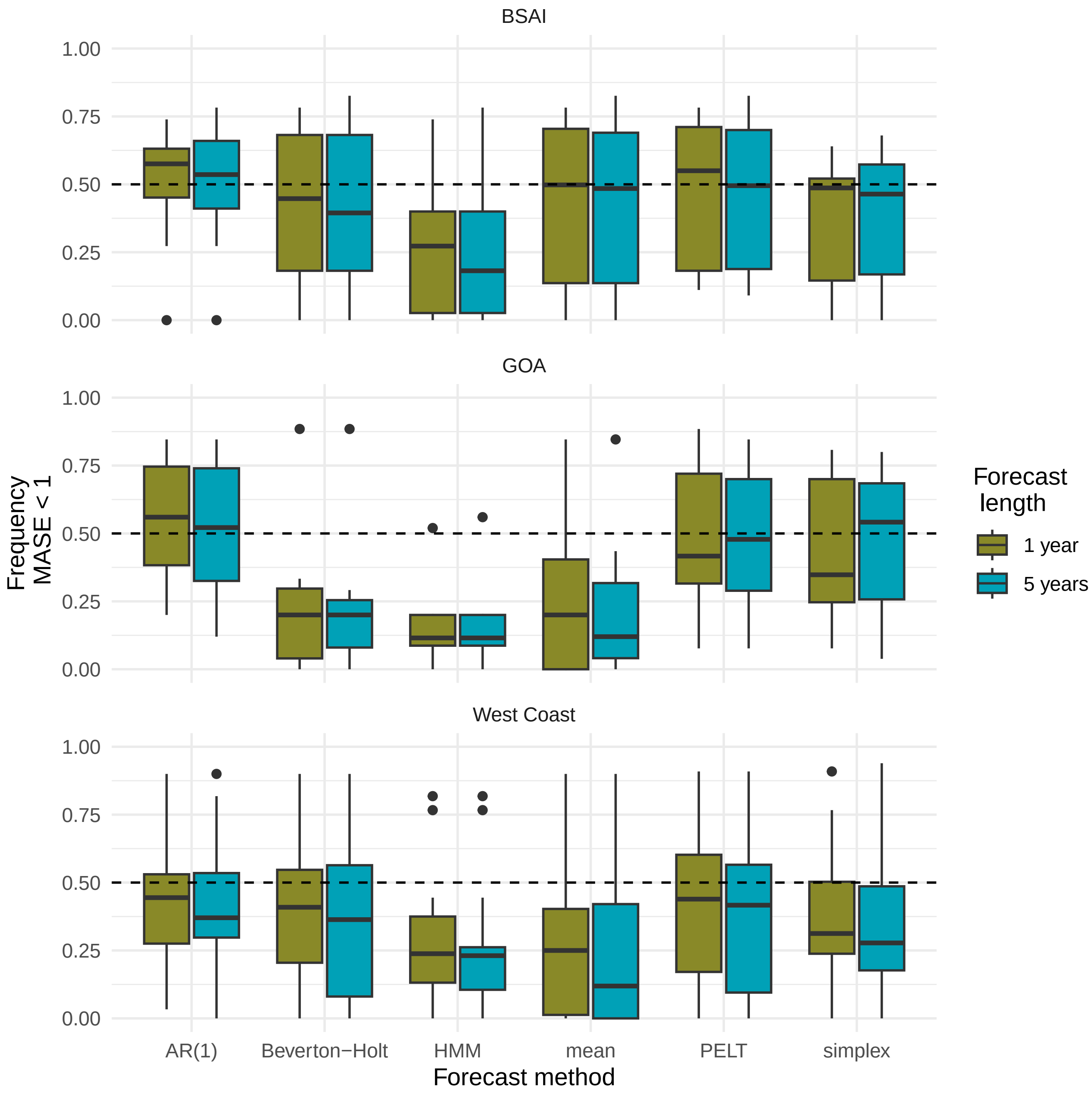

Figure 2.11: Boxplots of the frequency with which each method performed better than the in-sample naïve forecast (MASE < 1) by region and forecast length. The short forecast length is a 1-step ahead prediction, and the long forecast length is a 5-step ahead prediction.

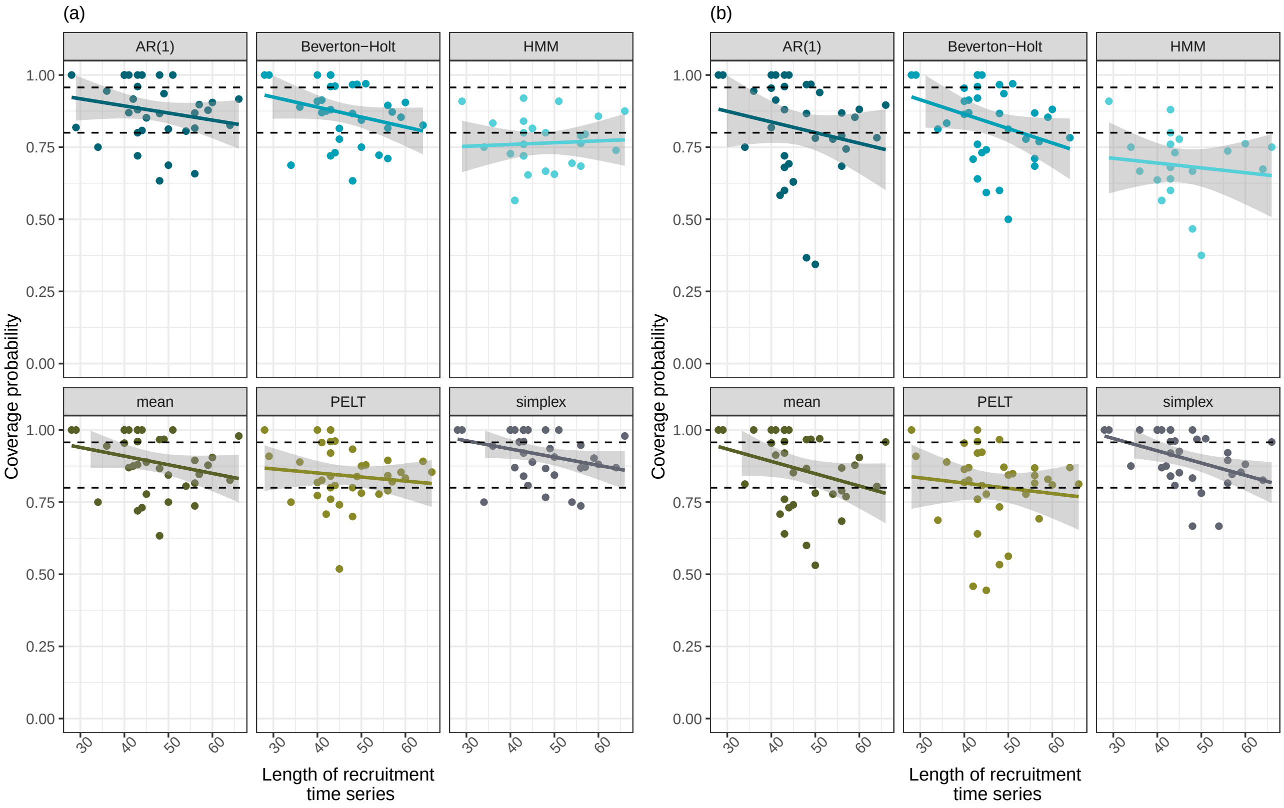

Figure 2.12: Forecast method coverage probabilities by the length of the recruitment time series for (a) short- and (b) mid-term forecasts. Points represent the coverage probability for each stock. Trend lines are shown for each method. Gray ribbons represent uncertainty around the trend lines.

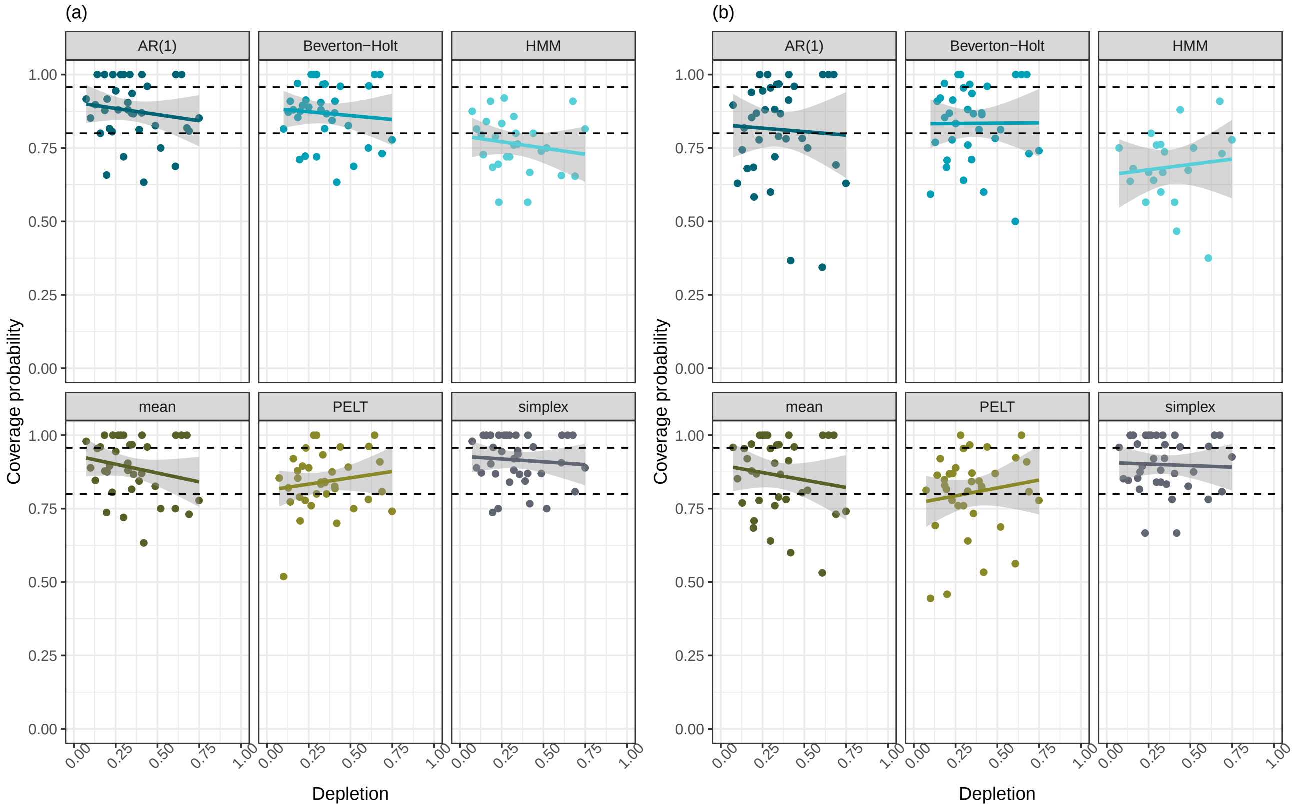

Figure 2.13: Forecast method coverage probabilities by spawning biomass depletion for (a) short- and (b) mid-term forecasts. Points represent the coverage probability for each stock. Trend lines are shown for each method. Gray ribbons represent uncertainty around the trend lines.

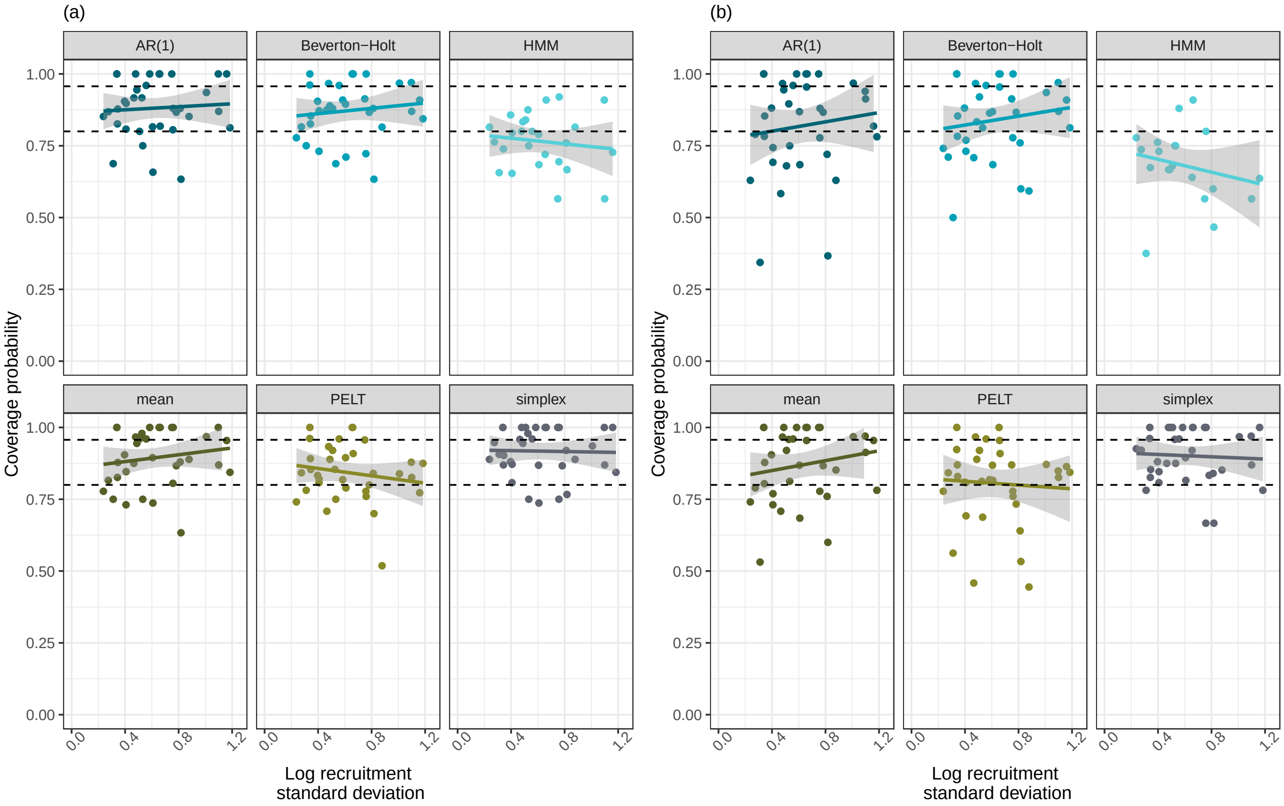

Figure 2.14: Forecast method coverage probabilities by the log recruitment standard deviation for (a) short- and (b) mid-term forecasts. Points represent the coverage probability for each stock. Trend lines are shown for each method. Gray ribbons represent uncertainty around the trend lines.

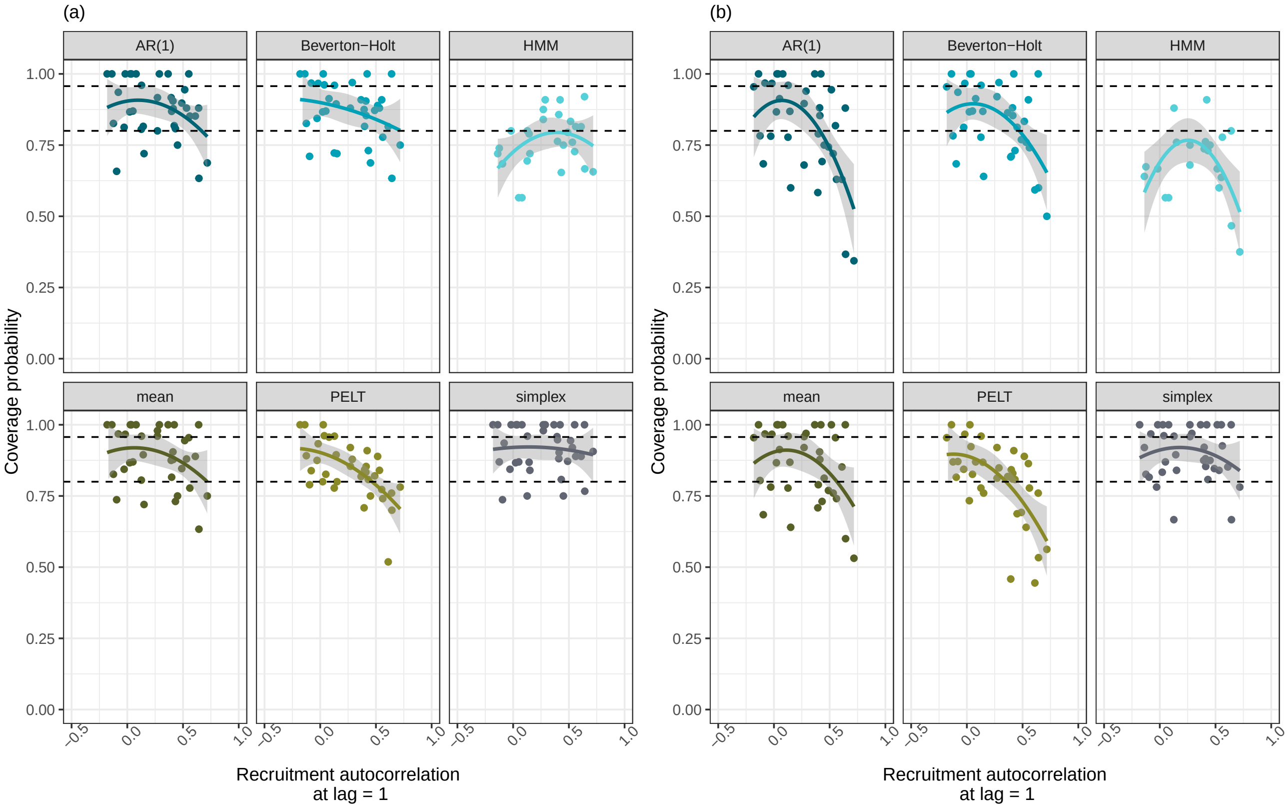

Figure 2.15: Forecast method coverage probabilities by recruitment autocorrelation for (a) short- and (b) mid-term forecasts. Points represent the coverage probability for each stock. Trend lines are shown for each method. Gray ribbons represent uncertainty around the trend lines.

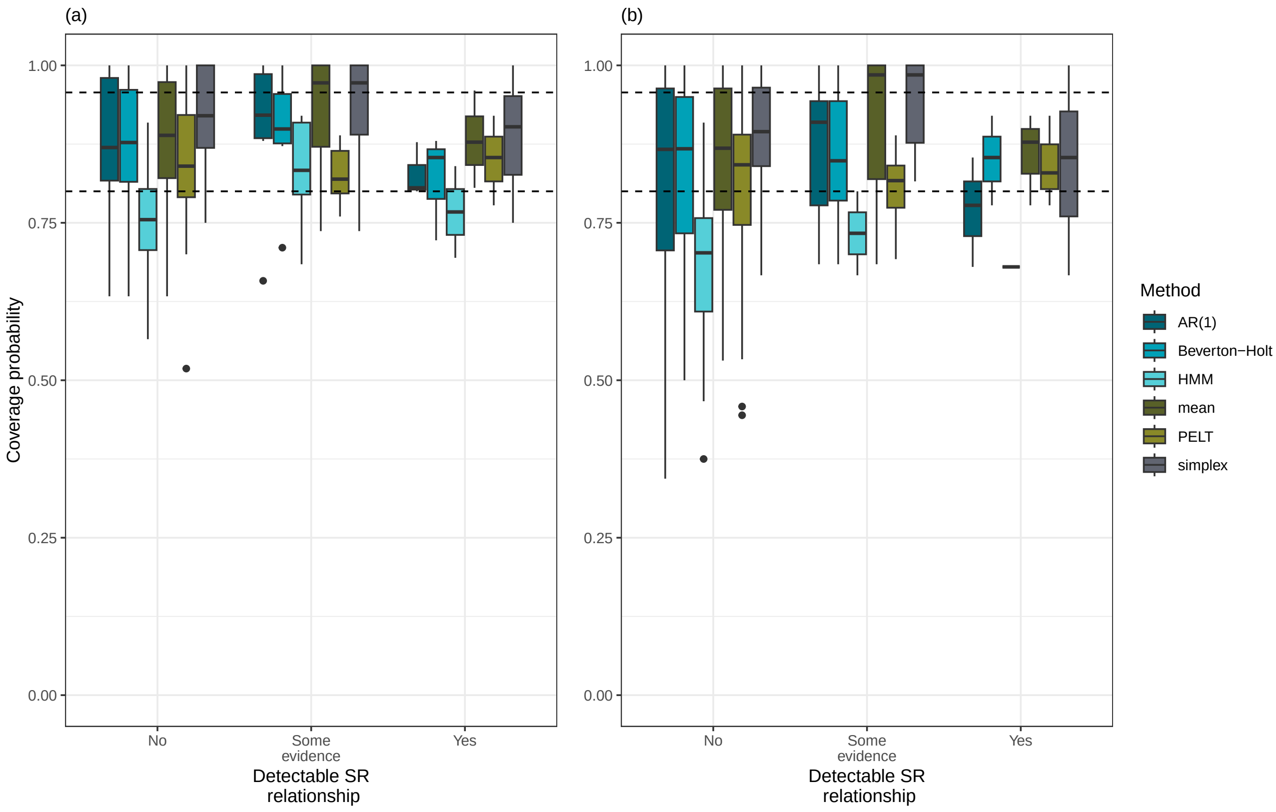

Figure 2.16: Forecast method coverage probabilities by whether a stock-recruitment relationship was detected for (a) short- and (b) mid-term forecasts.

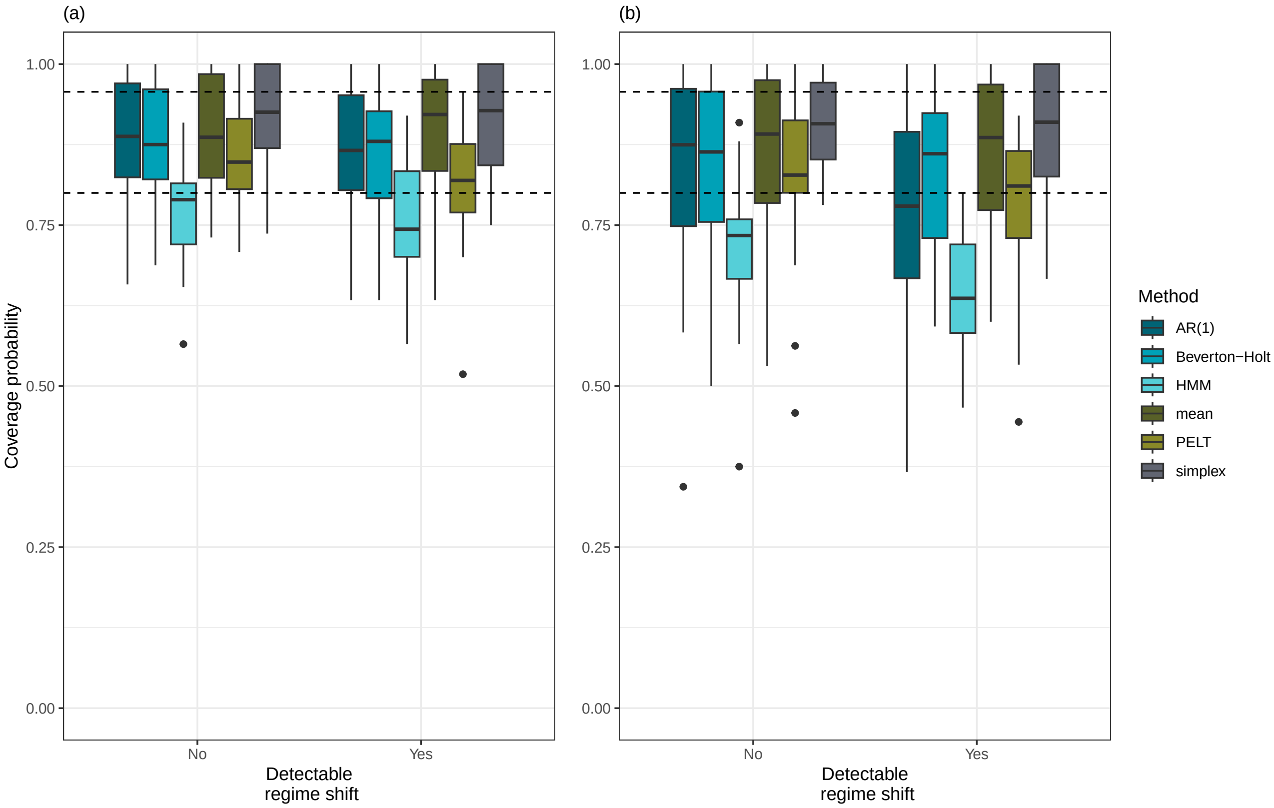

Figure 2.17: . Forecast method coverage probabilities by whether a recruitment regime shift was detected for (a) short- and (b) mid-term forecasts.

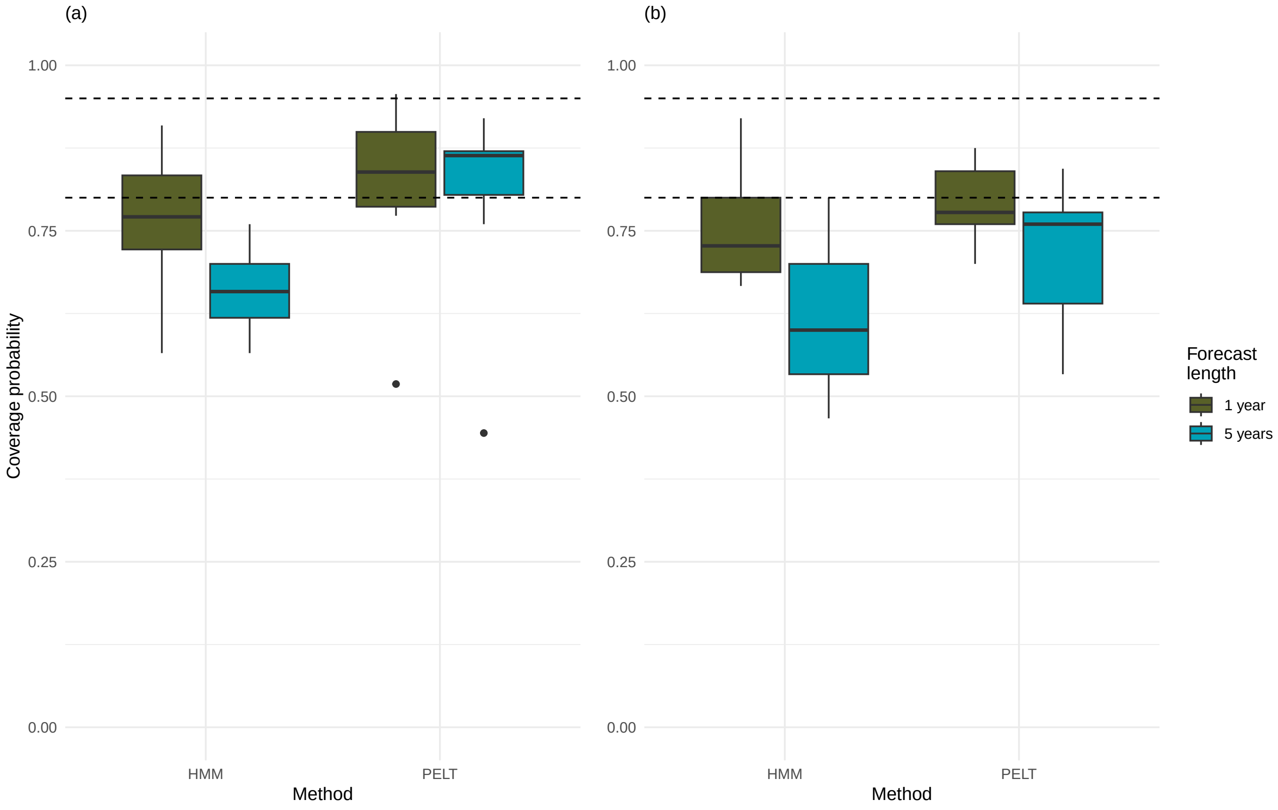

Figure 2.18: Coverage probabilities for HMM and PELT sampling methods for short- and mid-term forecasts. Panel (a) depicts coverage probabilities for stocks with a recruitment regime detected within the first 20 years of the recruitment time series. Panel (b) depicts coverage probabilities for stocks with the initial recruitment regime change after the first 20 years of the recruitment time series.

CONCLUSION

Recruitment, the number of fish entering a fish stock (often age-0 fish), is an important population process when modeling fish population dynamics. As such, research into recruitment has been extensive. Major themes in the recruitment literature include understanding the stock-recruitment relationship and improving recruitment forecasts. The stock-recruitment relationship has a direct impact on the estimation of maximum sustainable yield (MSY), which is essential for determining the sustainable harvests of fishery stocks in the United States. Forecasting recruitment is an integral part of stock biomass projections (Johnson et al., 2016). Improved projections of stock biomass help managers evaluate the status of a stock and set harvest limits (Deyle et al., 2018). The goal of this thesis was to add to the body of literature on recruitment through two research objectives, (1) re-evaluating a previous meta-analysis of stock-recruitment data in the RAM legacy database, and (2) testing a suite of recruitment forecast methods on data-rich groundfish stocks.

The first chapter of this thesis characterized robustness of the findings in Szuwalski et al. (2015), a previous meta-analysis of stock-recruitment drivers, to the addition of new data. The primary influence of spawning biomass on recruitment was determined for 432 stocks. The results indicated that 57% of stocks did not have a significant correlation between spawning biomass and recruitment over the observed biomasses. Environmental conditions therefore played a larger role in recruitment variation than spawning biomass. This chapter also examined the recruitment time series for regime shifts using three changepoint detection methods. The presence, location, and number of regime shifts in recruitment time series was highly dependent on the detection method. Despite the sensitivity to the method used to identify regime shifts, 46% of stocks without a significant correlation between spawning biomass and recruitment are estimated to have experienced at least one regime shift as determined by the PELT algorithm.

The second chapter of this thesis tested six recruitment forecast methods: 1) Beverton-Holt stock-recruitment relationship, 2) an autoregressive-1 function, 3) mean recruitment, 4) a hidden Markov sampling procedure, and 5) simplex projection, 6) a PELT changepoint sampling procedure. These methods varied in operational use across regions and the level of parameterization. Short- (1-year) and long-term (5-year) forecasts were generated using a expanding window approach. The coverage probability of the inter-quartile range and the mean absolute scaled error (MASE) of the predictions were used to evaluate the performance of each method, and it was found there was no clear best recruitment forecast method across all the stocks as different methods performed better depending on the forecast period (i.e., short vs long), management region, and stock characteristics. This is consistent with previous research (e.g., Van Beveren et al., 2021). Although it was not possible to determine a single best method, evaluating stock characteristics and the desired forecast period can provide some guidance on an appropriate method for the stocks evaluated.

This thesis demonstrates the challenges of modelling recruitment in stock assessments, especially with respect to regime shifts. Current approaches to stock-recruitment modeling often assume no large, sudden temporal variation, as would be observed under regime shifts (Perälä et al., 2017). In the first chapter, it was found that a large proportion of the stocks that were evaluated (46%) experienced at least one recruitment regime shift. The results suggest that the number and location of regime shifts depends on the method. Further research is needed to determine the most appropriate method for identifying regime shifts naively (i.e., without prior suspected years for shifts), as there is no peer-reviewed study directly comparing change point detection methods. The methods in chapter two that incorporated regime shifts were not appropriate for all stocks that were evaluated. Forecast accuracy depended on stock characteristics, particularly spawning biomass depletion and recruitment. Inaccurate recruitment projections can have large impacts on the management of a stock, especially when harvest control rules cannot account for abrupt shifts in biomass (A'mar et al., 2009; Szuwalski and Punt, 2013). Ideally, stock assessments would include recruitment forecast methods that are robust to changes in stock productivity, as regime shifts are likely to increase with climate change.

COLOPHON

This document is set in EB Garamond, Source Code Pro and Lato. The body text is set at 11pt with *lmr*.

It was written in R Markdown and LaTeX, and rendered into PDF using huskydown and bookdown.

This document was typeset using the XeTeX typesetting system, and the University of Washington Thesis class class created by Jim Fox. Under the hood, the University of Washington Thesis LaTeX template is used to ensure that documents conform precisely to submission standards. Other elements of the document formatting source code have been taken from the Latex, Knitr, and RMarkdown templates for UC Berkeley's graduate thesis, and Dissertate: a LaTeX dissertation template to support the production and typesetting of a PhD dissertation at Harvard, Princeton, and NYU

The source files for this thesis markdown, are available at `https://github.com/emilysellinger/thesis_markdown`. A hard copy of the thesis can be found in the University of Washington library.

This version of the thesis was generated on 2023-12-14 22:23:43. The repository is currently at this commit:

The computational environment that was used to generate this version is as follows:

```
- Session info ---------------------------------------------------------------
 setting  value
 version  R version 4.2.2 (2022-10-31 ucrt)
 os       Windows 10 x64 (build 22621)
```

```
 system   x86_64, mingw32
 ui       RTerm
 language (EN)
 collate  English_United States.utf8
 ctype    English_United States.utf8
 tz       America/Los_Angeles
 date     2023-12-14
 pandoc   3.1.1 @ C:/Program Files/RStudio/resources/app/bin/quarto/bin/tools/ (via rmar

- Packages -------------------------------------------------------------------
 package     * version    date (UTC) lib source
 bookdown      0.32       2023-01-17 [1] CRAN (R 4.2.2)
 cachem        1.0.6      2021-08-19 [1] CRAN (R 4.2.0)
 callr         3.7.3      2022-11-02 [1] CRAN (R 4.2.2)
 cli           3.6.0      2023-01-09 [1] CRAN (R 4.2.2)
 colorspace    2.1-0      2023-01-23 [1] CRAN (R 4.2.2)
 crayon        1.5.2      2022-09-29 [1] CRAN (R 4.2.2)
 devtools    * 2.4.5      2022-10-11 [1] CRAN (R 4.2.2)
 digest        0.6.31     2022-12-11 [1] CRAN (R 4.2.2)
 dplyr       * 1.1.0      2023-01-29 [1] CRAN (R 4.2.2)
 ellipsis      0.3.2      2021-04-29 [1] CRAN (R 4.2.0)
 evaluate      0.20       2023-01-17 [1] CRAN (R 4.2.2)
 fansi         1.0.4      2023-01-22 [1] CRAN (R 4.2.2)
 fastmap       1.1.0      2021-01-25 [1] CRAN (R 4.2.0)
 fs            1.6.1      2023-02-06 [1] CRAN (R 4.2.2)
 generics      0.1.3      2022-07-05 [1] CRAN (R 4.2.2)
 ggplot2     * 3.4.1      2023-02-10 [1] CRAN (R 4.2.2)
 git2r         0.31.0     2023-01-26 [1] CRAN (R 4.2.2)
```

```
glue          1.6.2      2022-02-24 [1] CRAN (R 4.2.0)
gtable        0.3.1      2022-09-01 [1] CRAN (R 4.2.2)
here        * 1.0.1      2020-12-13 [1] CRAN (R 4.2.0)
htmltools     0.5.4      2022-12-07 [1] CRAN (R 4.2.2)
htmlwidgets   1.6.1      2023-01-07 [1] CRAN (R 4.2.2)
httpuv        1.6.8      2023-01-12 [1] CRAN (R 4.2.2)
httr          1.4.4      2022-08-17 [1] CRAN (R 4.2.2)
huskydown   * 0.0.5      2022-11-16 [1] Github (benmarwick/huskydown@addb48e)
kableExtra  * 1.3.4.9000 2023-03-31 [1] Github (haozhu233/kableExtra@a6af5c0)
knitr       * 1.42       2023-01-25 [1] CRAN (R 4.2.2)
later         1.3.0      2021-08-18 [1] CRAN (R 4.2.1)
lifecycle     1.0.3      2022-10-07 [1] CRAN (R 4.2.2)
magrittr      2.0.3      2022-03-30 [1] CRAN (R 4.2.0)
memoise       2.0.1      2021-11-26 [1] CRAN (R 4.2.0)
mime          0.12       2021-09-28 [1] CRAN (R 4.2.0)
miniUI        0.1.1.1    2018-05-18 [1] CRAN (R 4.2.2)
munsell       0.5.0      2018-06-12 [1] CRAN (R 4.2.0)
pillar        1.8.1      2022-08-19 [1] CRAN (R 4.2.2)
pkgbuild      1.4.0      2022-11-27 [1] CRAN (R 4.2.2)
pkgconfig     2.0.3      2019-09-22 [1] CRAN (R 4.2.0)
pkgload       1.3.2      2022-11-16 [1] CRAN (R 4.2.0)
png           0.1-8      2022-11-29 [1] CRAN (R 4.2.2)
prettyunits   1.1.1      2020-01-24 [1] CRAN (R 4.2.0)
processx      3.8.0      2022-10-26 [1] CRAN (R 4.2.2)
profvis       0.3.7      2020-11-02 [1] CRAN (R 4.2.2)
promises      1.2.0.1    2021-02-11 [1] CRAN (R 4.2.1)
ps            1.7.2      2022-10-26 [1] CRAN (R 4.2.2)
purrr         1.0.1      2023-01-10 [1] CRAN (R 4.2.2)
```

```
R6              2.5.1      2021-08-19 [1] CRAN (R 4.2.0)
Rcpp            1.0.9      2022-07-08 [1] CRAN (R 4.2.2)
remotes         2.4.2      2021-11-30 [1] CRAN (R 4.2.2)
rlang           1.0.6      2022-09-24 [1] CRAN (R 4.2.2)
rmarkdown       2.20       2023-01-19 [1] CRAN (R 4.2.2)
rprojroot       2.0.3      2022-04-02 [1] CRAN (R 4.2.0)
rstudioapi      0.14       2022-08-22 [1] CRAN (R 4.2.2)
rvest           1.0.3      2022-08-19 [1] CRAN (R 4.2.2)
scales          1.2.1      2022-08-20 [1] CRAN (R 4.2.2)
sessioninfo     1.2.2      2021-12-06 [1] CRAN (R 4.2.0)
shiny           1.7.4      2022-12-15 [1] CRAN (R 4.2.2)
stringi         1.7.12     2023-01-11 [1] CRAN (R 4.2.2)
stringr         1.5.0      2022-12-02 [1] CRAN (R 4.2.2)
svglite         2.1.1      2023-01-10 [1] CRAN (R 4.2.2)
systemfonts     1.0.4      2022-02-11 [1] CRAN (R 4.2.2)
tibble          3.1.8      2022-07-22 [1] CRAN (R 4.2.2)
tidyselect      1.2.0      2022-10-10 [1] CRAN (R 4.2.2)
urlchecker      1.0.1      2021-11-30 [1] CRAN (R 4.2.2)
usethis       * 2.1.6      2022-05-25 [1] CRAN (R 4.2.0)
utf8            1.2.3      2023-01-31 [1] CRAN (R 4.2.2)
vctrs           0.5.2      2023-01-23 [1] CRAN (R 4.2.2)
viridisLite     0.4.1      2022-08-22 [1] CRAN (R 4.2.2)
webshot         0.5.4      2022-09-26 [1] CRAN (R 4.2.2)
withr           2.5.0      2022-03-03 [1] CRAN (R 4.2.0)
xfun            0.37       2023-01-31 [1] CRAN (R 4.2.2)
xml2            1.3.3      2021-11-30 [1] CRAN (R 4.2.0)
xtable          1.8-4      2019-04-21 [1] CRAN (R 4.2.0)
yaml            2.3.7      2023-01-23 [1] CRAN (R 4.2.2)
```

```
 [1] C:/Users/Emily Sellinger/AppData/Local/R/win-library/4.2
 [2] C:/Program Files/R/R-4.2.2/library

──────────────────────────────────────────────────────────────────────────────
```

REFERENCES

10 A'mar, Z.T., Punt, A.E., Dorn, M.W., 2009. The impact of regime shifts on the performance of management strategies for the Gulf of Alaska walleye pollock (Theragra chalcogramma) fishery. Canadian Journal of Fisheries and Aquatic Sciences 66, 2222–2242. https://doi.org/10.1139/F09-142

Adams, R.P., MacKay, D.J.C., 2007. Bayesian Online Changepoint Detection. https://doi.org/10.48550/arXiv.0710.3742

Andersen, T., Carstensen, J., Hernández-García, E., Duarte, C.M., 2009. Ecological thresholds and regime shifts: Approaches to identification. Trends in Ecology & Evolution 24, 49–57. https://doi.org/10.1016/j.tree.2008.07.014

Barry, D., Hartigan, J.A., 1993. A Bayesian Analysis for Change Point Problems. Journal of the American Statistical Association 88, 309–319. https://doi.org/10.2307/2290726

Barunik, J., Kristoufek, L., 2010. On Hurst exponent estimation under heavy-tailed distributions. Physica A: Statistical Mechanics and its Applications 389, 3844–3855. https://doi.org/10.1016/j.physa.2010.05.025

Basson, M., 1999. The importance of environmental factors in the design of management procedures. ICES Journal of Marine Science 56, 933–942. https://doi.org/10.1006/jmsc.1999.0541

Beverton, R.J.H., Holt, S.J., 1957. On the dynamics of exploited fish populations, 1st ed, Fish & Fisheries Series. Springer.

Cadigan, N.G., 2013. Fitting a non-parametric stock–recruitment model in R that is useful for deriving MSY reference points and accounting for model uncertainty. ICES Journal of Marine Science 70, 56–67. https://doi.org/10.1093/icesjms/fss183

Cardinale, M., Arrhenius, F., 2000. The influence of stock structure and environmental conditions on the recruitment process of Baltic cod estimated using a generalized additive model. Canadian Journal of Fisheries and Aquatic Sciences 57, 2402–2409. https://doi.org/10.1139/f00-221

Collie, J.S., Richardson, K., Steele, J.H., 2004. Regime shifts: Can ecological theory illuminate the mechanisms? Progress in Oceanography, Regime shifts in the ocean. Reconciling observations and theory 60, 281–302. https://doi.org/10.1016/j.pocean.2004.02.013

Crépin, A.-S., Biggs, R., Polasky, S., Troell, M., Zeeuw, A. de, 2012. Regime shifts and management. Ecological Economics, The Economics of Degrowth 84, 15–22. https://doi.org/10.1016/j.ecolecon.2012.09.003

Cury, P., Fromentin, J.-M., Figuet, S., Bonhommeau, S., 2014. Resolving Hjort's Dilemma How Is Recruitment Related to Spawning Stock Biomass in Marins Fish? Oceanography 27, 42–47. https://doi.org/10.5670/oceanog.2014.85

De Oliveira, J., Butterworth, D.S., 2005. Limits to the use of environmental indices to reduce risk and/or increase yield in the South African anchovy fishery. African Journal of Marine Science 27, 191–203. https://doi.org/10.2989/18142320509504078

DeFilippo, L.B., Buehrens, T.W., Scheuerell, M., Kendall, N.W., Schindler, D.E., 2021. Improving short-term recruitment forecasts for coho salmon using a spatiotemporal integrated population model. Fisheries Research 242, 106014. https://doi.org/10.1016/j.fishres.2021.106014

Deyle, E., Schueller, A.M., Ye, H., Pao, G.M., Sugihara, G., 2018. Ecosystem-based forecasts of recruitment in two menhaden species. Fish and Fisheries 19, 769–781. `https://doi.org/10.1111/faf.12287`

Dickey-Collas, M., Hintzen, N.T., Nash, R.D.M., Schön, P.-J., Payne, M.R., 2015. Quirky patterns in time-series of estimates of recruitment could be artefacts. ICES Journal of Marine Science 72, 111–116. `https://doi.org/10.1093/icesjms/fsu022`

Erdman, C., Emerson, J.W., 2008. Bcp: An R Package for Performing a Bayesian Analysis of Change Point Problems. Journal of Statistical Software 23, 1–13. `https://doi.org/10.18637/jss.v023.i03`

Evans, G.T., Rice, J.C., 1988. Predicting recruitment from stock size without the mediation of a functional relation. ICES Journal of Marine Science 44, 111–122. `https://doi.org/10.1093/icesjms/44.2.111`

Gilbert, D.J., 1997. Towards a new recruitment paradigm for fish stocks. Canadian Journal of Fisheries and Aquatic Sciences 54, 969–977. `https://doi.org/10.1139/f96-272`

Haltuch, M.A., Brooks, E.N., Brodziak, J., Devine, J.A., Johnson, K.F., Klibansky, N., Nash, R.D.M., Payne, M.R., Shertzer, K.W., Subbey, S., Wells, B.K., 2019. Unraveling the recruitment problem: A review of environmentally-informed forecasting and management strategy evaluation. Fisheries Research, Recruitment: Theory, Estimation, and Application in Fishery Stock Assessment Models 217, 198–216. `https://doi.org/10.1016/j.fishres.2018.12.016`

Haltuch, M.A., Punt, A.E., 2011. The promises and pitfalls of including decadal-scale climate forcing of recruitment in groundfish stock assessment. Canadian Journal of Fisheries and Aquatic Sciences 68, 912–926. `https://doi.org/10.1139/f2011-030`

Hollowed, A.B., Bailey, K.M., Wooster, W.S., 1987. Patterns in Recruitment of Marine Fishes in the Northeast Pacific Ocean. Biological Oceanography 5, 99–131. `https://doi.org/10.1080/01965581.1987.10749508`

Holmes, E.E., Ward, E.J., Scheuerell, M.D., 2021. Analysis of multivariate time series using the MARSS package. Version 3.11.4. `https://doi.org/10.5281/ZENODO.5781847`

Hughes, T.P., Carpenter, S., Rockström, J., Scheffer, M., Walker, B., 2013. Multiscale regime shifts and planetary boundaries. Trends in Ecology & Evolution 28, 389–395. `https://doi.org/10.1016/j.tree.2013.05.019`

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22, 679–688. `https://doi.org/10.1016/j.ijforecast.2006.03.001`

Jackson, B., Scargle, J.D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L., Tsai, T.T., 2005. An algorithm for optimal partitioning of data on an interval. IEEE Signal Processing Letters 12, 105–108. `https://doi.org/10.1109/LSP.2001.838216`

Jiao, Y., 2009. Regime shift in marine ecosystems and implications for fisheries management, a review. Reviews in Fish Biology and Fisheries 19, 177–191. `https://doi.org/10.1007/s11160-008-9096-8`

Johnson, K.F., Councill, E., Thorson, J.T., Brooks, E., Methot, R.D., Punt, A.E., 2016. Can autocorrelated recruitment be estimated using integrated assessment models and how does it affect population forecasts? Fisheries Research 183, 222–232. `https://doi.org/10.1016/j.fishres.2016.06.004`

Kaposzta, Z., Czoch, A., Stylianou, O., Kim, K., Mukli, P., Eke, A., Racz, F.S., 2022. Real-Time Algorithm for Detrended Cross-Correlation Analysis of Long-Range Coupled Processes. Frontiers in Physiology 13.

Kell, L.T., Nash, R.D.M., Dickey-Collas, M., Mosqueira, I., Szuwalski, C., 2016. Is spawning stock biomass a robust proxy for reproductive potential? Fish and Fisheries 17, 596–616. `https://doi.org/10.1111/faf.12131`

Kiaer, C., Neuenfeldt, S., Payne, M.R., 2021. A framework for assessing the skill and value of operational recruitment forecasts. ICES Journal of Marine Science 78, 3581–3591. `https://doi.org/10.1093/icesjms/fsab202`

Killick, R., Eckley, I.A., 2014. Changepoint: An R Package for Changepoint Analysis. Journal of Statistical Software 58, 1–19. `https://doi.org/10.18637/jss.v058.i03`

Killick, R., Fearnhead, P., Eckley, I.A., 2012. Optimal Detection of Changepoints With a Linear Computational Cost. Journal of the American Statistical Association 107, 1590–1598. `https://doi.org/10.1080/01621459.2012.737745`

Kristoufek, L., 2014. Measuring correlations between non-stationary series with DCCA coefficient. Physica A: Statistical Mechanics and its Applications 402, 291–298. `https://doi.org/10.1016/j.physa.2014.01.058`

Marshall, C.T., 2009. Implementing Information on Stock Reproductive Potential in Fisheries Management: The Motivation, Challenges and Opportunities, in: Fish Reproductive Biology. John Wiley & Sons, Ltd, pp. 395–420. `https://doi.org/10.1002/9781444312133.ch11`

Mukli, P., Nagy, Z., Eke, A., 2015. Multifractal formalism by enforcing the universal behavior of scaling functions. Physica A: Statistical Mechanics and its Applications 417, 150–167. `https://doi.org/10.1016/j.physa.2014.09.002`

Munch, S.B., Giron-Nava, A., Sugihara, G., 2018. Nonlinear dynamics and noise in fisheries recruitment: A global meta-analysis. Fish and Fisheries 19, 964–973. `https://doi.org/10.1111/faf.12304`

Munch, S.B., Kottas, A., 2009. A Bayesian Modeling Approach for Determining Productivity Regimes and Their Characteristics. Ecological Applications 19, 527–537.

Munch, S.B., Kottas, A., Mangel, M., 2005. Bayesian nonparametric analysis of stock-recruitment relationships. Canadian Journal of Fisheries and Aquatic Sciences 62, 1808–1821. https://doi.org/10.1139/f05-073

Myers, R.A., Barrowman, N.J., 1996. Is fish recruitment related to spawner abundance? Fishery Bulletin 94, 707–724.

Olsen, E.M., Ottersen, G., Llope, M., Chan, K.-S., Beaugrand, G., Stenseth, N.Chr., 2011. Spawning stock and recruitment in North Sea cod shaped by food and climate. Proceedings of the Royal Society B: Biological Sciences 278, 504–510. https://doi.org/10.1098/rspb.2010.1465

Ovando, D., Cunningham, C., Kuriyama, P., Boatright, C., Hilborn, R., 2022. Improving forecasts of sockeye salmon (Oncorhynchus nerka) with parametric and nonparametric models. Canadian Journal of Fisheries and Aquatic Sciences 79, 1198–1210. https://doi.org/10.1139/cjfas-2021-0287

Perälä, T.A., Swain, D.P., Kuparinen, A., 2017. Examining nonstationarity in the recruitment dynamics of fishes using Bayesian change point analysis. Canadian Journal of Fisheries and Aquatic Sciences 74, 751–765. https://doi.org/10.1139/cjfas-2016-0177

Petchey, O.L., 2016. Simplex Projection Walkthrough. https://doi.org/10.5281/ZENODO.57081

Podobnik, B., Stanley, H.E., 2008. Detrended Cross-Correlation Analysis: A New Method for Analyzing Two Nonstationary Time Series. Physical Review Letters 100, 084102. https://doi.org/10.1103/PhysRevLett.100.084102

Prass, T.S., Pumi, G., 2020. DCCA: Detrended Fluctuation and Detrended Cross-Correlation Analysis.

Punt, A.E., 2023. Those who fail to learn from history are condemned to repeat it: A perspective on current stock assessment good practices and the consequences of not following them. Fisheries Research 261, 106642. `https://doi.org/10.1016/j.fishres.2023.106642`

Ricard, D., Minto, C., Jensen, O.P., Baum, J.K., 2012. Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. Fish and Fisheries 13, 380–398. `https://doi.org/10.1111/j.1467-2979.2011.00435.x`

Ricker, W.E., 1954. Stock and Recruitment. Journal of the Fisheries Research Board of Canada 11, 559–623. `https://doi.org/10.1139/f54-039`

Scheffer, M., Carpenter, S., Foley, J.A., Folke, C., Walker, B., 2001. Catastrophic shifts in ecosystems. Nature 413, 591–596. `https://doi.org/10.1038/35098000`

Sellinger, E.L., Szuwalski, C., Punt, A.E., 2024. The robustness of our assumptions about recruitment: A re-examination of marine recruitment dynamics with additional data and novel methods. Fisheries Research 269, 106862. `https://doi.org/10.1016/j.fishres.2023.106862`

Sguotti, C., Otto, S.A., Cormon, X., Werner, K.M., Deyle, E., Sugihara, G., Möllmann, C., 2020. Non-linearity in stock–recruitment relationships of Atlantic cod: Insights from a multi-model approach. ICES Journal of Marine Science 77, 1492–1502. `https://doi.org/10.1093/icesjms/fsz113`

Spearman, C., 1987. The Proof and Measurement of Association between Two Things. The American Journal of Psychology 100, 441–471. `https://doi.org/10.2307/1422689`

Subbey, S., Devine, J.A., Schaarschmidt, U., Nash, R.D.M., 2014. Modelling and forecasting stock–recruitment: Current and future perspectives. ICES Journal of Marine Science 71, 2307–2322. https://doi.org/10.1093/icesjms/fsu148

Sugihara, G., May, R., 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error m time series. Nature 344, 734. https://doi.org/10.1038/344734a0

Sun, L., Xiao, H., Li, S., Yang, D., 2009. Forecasting Fish Stock Recruitment and Planning Optimal Harvesting Strategies by Using Neural Network. Journal of Computers 4, 1075–1082. https://doi.org/10.4304/jcp.4.11.1075-1082

Swain, D.P., Sinclair, A.F., 2000. Pelagic fishes and the cod recruitment dilemma in the Northwest Atlantic. Canadian Journal of Fisheries and Aquatic Sciences 57, 1321–1325. https://doi.org/10.1139/f00-104

Szuwalski, C.S., Hollowed, A.B., Holsman, K.K., Ianelli, J.N., Legault, C.M., Melnychuk, M.C., Ovando, D., Punt, A.E., 2023. Unintended consequences of climate-adaptive fisheries management targets. Fish and Fisheries 24, 439–453. https://doi.org/10.1111/faf.12737

Szuwalski, C.S., Punt, A.E., 2013. Fisheries management for regime-based ecosystems: A management strategy evaluation for the snow crab fishery in the eastern Bering Sea. ICES Journal of Marine Science 70, 955–967. https://doi.org/10.1093/icesjms/fss182

Szuwalski, C.S., Vert-Pre, K.A., Punt, A.E., Branch, T.A., Hilborn, R., 2015. Examining common assumptions about recruitment: A meta-analysis of recruitment dynamics for worldwide marine fisheries. Fish and Fisheries 16, 633–648. https://doi.org/10.1111/faf.12083

Takens, F., 1981. Detecting strange attractors in turbulence, in: Rand, D., Young, L.-S. (Eds.), Dynamical Systems and Turbulence, Warwick 1980, Lecture Notes in Mathematics. Springer, Berlin, Heidelberg, pp. 366–381. https://doi.org/10.1007/BFb0091924

Tang, X., Zheng, N., Rideout, R.M., Wang, S., Zhang, F., 2021. Identification of recruitment regime shifts with a hidden Markov stock-recruitment model. ICES Journal of Marine Science 78, 2591–2602. https://doi.org/10.1093/icesjms/fsab141

Van Beveren, E., Benoît, H.P., Duplisea, D.E., 2021. Forecasting fish recruitment in age-structured population models. Fish and Fisheries 22, 941–954. https://doi.org/10.1111/faf.12562

Walters, C.J., Ludwig, D., 1981. Effects of Measurement Errors on the Assessment of Stock–Recruitment Relationships. Canadian Journal of Fisheries and Aquatic Sciences 38, 704–710. https://doi.org/10.1139/f81-093

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

Wayte, S.E., 2013. Management implications of including a climate-induced recruitment shift in the stock assessment for jackass morwong (Nemadactylus macropterus) in south-eastern Australia. Fisheries Research 142, 47–55. https://doi.org/10.1016/j.fishres.2012.07.009

Xu, H., Miller, T.J., Hameed, S., Alade, L.A., Nye, J.A., 2018. Evaluating the utility of the Gulf Stream Index for predicting recruitment of Southern New England-Mid Atlantic yellowtail flounder. Fisheries Oceanography 27, 85–95. https://doi.org/10.1111/fog.12236

Ye, H., Beamish, R.J., Glaser, S.M., Grant, S.C.H., Hsieh, C., Richards, L.J., Schnute, J.T., Sugihara, G., 2015. Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. Proceedings of the National Academy of Sciences 112, E1569–E1576. `https://doi.org/10.1073/pnas.1417063112`