2021

A BAYESIAN STATE-SPACE APPROACH TO IMPROVE BIOMASS PROJECTIONS FOR MANAGING NEW ENGLAND GROUNDFISH

Joseph A. Langan
University of Rhode Island, joseph_langan@uri.edu

A BAYESIAN STATE-SPACE APPROACH TO IMPROVE BIOMASS PROJECTIONS FOR MANAGING NEW ENGLAND GROUNDFISH

BY

JOSEPH A. LANGAN

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

STATSTICS

UNIVERSITY OF RHODE ISLAND

2021

MASTER OF SCIENCE THESIS

OF

JOSEPH A. LANGAN

APPROVED:

Thesis Committee:

Major Professor Gavino Puggioni

Guangyu Zhu

Jeremy S. Collie

Christopher M. Legault

Brenton DeBoef
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2021

ABSTRACT

Fisheries management requires regularly assessing stock status and setting catch levels for the coming years. Although data-rich stock assessment models incorporating demographic and biological information are generally preferred, such approaches are often prohibited by either insufficient data or a history of poor performance. In such cases, simpler Index-Based Methods (IBMs) are often used to generate catch advice for a fishery. However, these approaches do not typically forecast future abundance levels or quantify scientific uncertainty, making it difficult to assess the performance of different candidate methods prior to implementation. As a result, there is not a consensus as to which IBMs may be best suited to a particular situation and available biological information is often under-utilized in the management process.

To address this shortcoming, this work developed a novel Index-Based Method framework using dynamic linear models (DLMs), a flexible Bayesian state-space approach. Using simulated population data mimicking member species of the Northeast Multispecies groundfish complex, the predictive performance of candidate DLM structures were evaluated via retrospective forecasting. In both an Index-Based (age-aggregated) and Age-Based formulation constructed to demonstrate how the modular nature of these models can make fuller use of available data, the tested DLMs displayed promising predictive ability. While further testing is needed, this preliminary evaluation suggests that DLMs may become a valuable approach in the management of fisheries for which a data-rich stock assessment approach is not possible.

The research in this thesis aims to provide scientists and fisheries managers with an additional tool to be developed for use in fisheries management. Because the majority of world fish stocks lack sufficient data for conventional stock assessment methods, the modeling approach and insights developed here are meant to contribute knowledge to the global pursuit of productive and sustainable marine fisheries.

ACKNOWLEDGMENTS

This has been an amazing journey. I would first like to thank my major advisor, Gavino Puggioni. You have always challenged me to take one more step and learn one more lesson, believing in my abilities even when my own confidence faltered. Your support and contributions to all my research projects and ideas have helped me to grow so much. You have taught me to approach and think about problems in a new way that will be invaluable throughout my career. I look forward to maintaining our research collaborations and friendship in the years to come. I would next like to thank my PhD advisor, Jeremy Collie. Thank you for your belief in me, your unending patience, and for pushing me to become a better scientist. Your expertise and support throughout all aspects of my graduate education have meant a lot to me and provided lessons that will benefit me for years to come. I am and will forever be tremendously grateful for your and Gavino's help and support as I took on two degree programs. Working so closely with and learning from both of you over these past six years has been an honor.

I would next like to thank my fellowship mentor, Chris Legault. Thank you for agreeing to work with me on my fellowship project. Your enthusiasm and boundless support of me and my research has meant the world to me. I have learned so much from you about science, modeling, and fisheries management and I feel very fortunate to have had the opportunity to work with you. I sincerely hope we are able to continue to work together in the future.

I would also like to thank my remaining committee member, Guangyu Zhu. Thank you for agreeing to serve as my inside committee member and helping to improve my thesis. I must also thank my other co-author on this research, Jason McNamee. Thank you for always being a supportive sounding board when I needed help or to talk out ideas. In this research and other projects, I have learned a lot from you and have really enjoyed getting to know you during my graduate education.

This thesis was funded with support from NOAA Fisheries, the National Sea Grant College Program, and the New England Fishery Management Council, for which I am grateful. Further, this research could not have been completed without additional funding and support of my various projects from the University of Rhode Island, the Rhode Island Division of Marine Fisheries, the Bureau of Ocean Energy Management, and Rhode Island Sea Grant. Additionally, the work within this thesis benefited significantly from the thoughtful input of the members of the Index Based Methods Working Group and the staff and committees of the New England Fishery Management Council.

I would next like to thank the URI community. I believe that it takes a village to train a scientist, and I could not have had a better village to be a part of during graduate school. So many of you were so generous with your time, your expertise, and your support. I can only hope to pay some of that forward in the future. I would also like to thank my many collaborators across my different research projects and endeavors. It has been a joy to work with all of you and I will forever cherish having had the opportunity to conduct research alongside each of you. I hope that our collaborations will extend far into the future.

To my fellow graduate students and friends- thank you for your tireless support and for helping me to make such fond memories during graduate school. Having the

opportunity to watch so many of you grow into amazing scientists and people has been a privilege. To my family- your unending encouragement and support has kept me motivated and grounded on this journey. I will always be grateful for the role you played in my success.

Finally, I could not have achieved any of this without the love, support, patience, and understanding of my partner Lauren. From listening to me rant about statistical models to joining me on the fish trawl, you have been my biggest cheerleader and my rock through all of it. You have picked me up when I felt discouraged and knocked me back into my place when I got too high. You have always been understanding and supportive when I had late nights coding or woke you up early in the morning as I left for field work. During this last, challenging year under a global pandemic, you have taken on so much to help and encourage me as I finished my degrees. You are my hero and I cannot express how grateful I am for your support. I would not be half the scientist or a person I am today with you and I cannot wait to continue to grow with you in the next chapter of our lives.

PREFACE

This thesis is formatted in accordance with the manuscript format guidelines established by the Graduate School of the University of Rhode Island. The manuscript is formatted for submission to the journal *Canadian Journal of Fisheries and Aquatic Sciences*, with co-authors Christopher Legault, Jeremy Collie, Jason McNamee, and Gavino Puggioni.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Since 1976, the Magnuson-Stevens Act has provided the regulatory framework to preserve living marine resources and the economic activities that they support in US waters through its requirement that fisheries management bodies prevent overfishing, rebuild depleted stocks, and ensure future sustainability of the seafood supply (National Oceanic and Atmospheric Administration 2006). To pursue these objectives, fisheries scientists generally assess the status of managed fish and invertebrate stocks every one to four years and evaluate what, if any, management actions are required. Referred to as stock assessments, these efforts involve combining fishery-independent survey data with catch information in population models that estimate current stock abundance. These estimates can then be used to set Acceptable Biological Catch (ABC) levels for the following years that will alter population abundance in the desired direction (National Marine Fisheries Service 2001; ASMFC 2009; Wiedenmann and Jensen 2018). Conducted one or more times each year, these scientific surveys consist of systematic sampling throughout the spatial range of managed species using appropriate gear types to gather abundance and other biological information over time (Politis et al. 2014). Although such survey and harvest data are usually gathered annually (ASMFC 2009; Politis et al. 2014), the ABCs set during the prior stock assessment are typically maintained until there are sufficient resources to conduct a new assessment (National Marine Fisheries Service 2001). In the interim, any short-term projection of biomass and subsequent adjustment of catch levels is constrained to a less formal consultation of recent survey and catch data. Without a designated model in place beyond the stock assessment for updating abundance estimates, such intermediate forecasts are often challenging and have high prediction error that can contribute to regulatory volatility, ratcheting down of catch quotas, increased risk of overfishing, and degraded stakeholder confidence in the assessment process (Glaser et al. 2014; Brooks and Legault 2016; Szuwalski et al. 2018).

When sufficient data are available, age-structured stock assessment approaches are preferred in the management of US stocks (e.g. Legault and Restrepo 1998). In general, these age-structured models combine survey data, catch information, age-composition data, and biological information, such as growth rates and maturation schedules, to track individual cohorts through the life cycle and ultimately estimate the true total population size. With an estimate of the true population size and potential productivity, fisheries managers can identify biological reference points to describe stock status and develop regulatory strategies to maintain fishery productivity (National Marine Fisheries Service 2001; ASMFC 2009; Punt et al. 2020). When these data-rich models perform poorly, however, they may be rejected in favor of a simpler approach (Punt et al. 2020). Furthermore, the majority of global fish stocks, including many in US waters, lack adequate data to implement an age-structured approach and thus require data-limited methods (Costello et al. 2012).

If an age-structured assessment cannot be implemented for a particular stock due either to data limitations or poor performance, one of many simple, data-limited methods may be implemented. Often known as Index-Based Methods (IBMs), these approaches typically involve the use of survey data and/or catch information to execute ad hoc procedures to develop catch advice without forecasting future stock abundance or quantifying scientific uncertainty (Wiedenmann et al. 2019; Legault et al. 2021). These procedures vary in both data requirements and technical form, which may involve calculating recent abundance trend estimates or assessing changes in the ratio between catch and survey abundances. Because IBMs do not produce forecasts of future abundance or estimate the true stock abundance, it can be difficult to determine which methods are most effective in a given situation and to develop decision rules with which to manage the fishery (Legault et al. 2021). Furthermore, the absence of an estimate of the uncertainty surrounding developed catch advice makes it difficult to assess the risk associated with candidate management strategies (Berkson et al. 2011; Maunder and Piner 2015; Kokkalis et al. 2017). There is therefore a need to develop an IBM that is able to forecast future stock abundance and quantify scientific uncertainty to aid in the management process when an age-structured approach is not possible for a target stock.

Due to their modular, flexible nature (Prado and West 2010), dynamic linear models (DLMs) are well-suited to tracking time series of observed relative abundance and develop probabilistic forecasts of managed fish stocks. As a Bayesian state-space approach, past research suggests that DLMs will likely exhibit strong performance in assessing scientific uncertainty (Fronczyk et al. 2012; Magnusson et al. 2013). Therefore, the objectives of this thesis are to 1) develop a DLM approach for forecasting relative stock abundance using only aggregate survey and catch information, 2) investigate the incorporation of additional data, like demographic information, into the DLM approach, and 3) test the predictive performance of the developed DLMs on simulated fish population data. In doing so, we hope to provide a preliminary understanding of the promise of DLMs as a tool in fisheries management and contribute to the sustainable management of data-limited marine fisheries.

REFERENCES

ASMFC, A.S.M.F.C. 2009. Guide to Fisheries Science and Stock Assessments. Report pursuant to National Oceanic and Atmospheric Administration Grant No. NA05NMF4741025.

Berkson, J., Barbieri, L., Cadrin, S.X., Cass-Calay, S., Crone, P., Dorn, M., Friess, C., Kobayashi, D., Miller, T.J., Patrick, W.S., Pautzke, S., Ralston, S., and Trianni, M. 2011. Calculating Acceptable Biological Catch for Stocks That Have Reliable Catch Data Only (Only Reliable Catch Stocks – ORCS).

Brooks, E.N., and Legault, C.M. 2016. Retrospective forecasting — evaluating performance of stock projections for New England groundfish stocks. Can. J. Fish. Aquat. Sci. **73**(6): 935–950. doi:10.1139/cjfas-2015-0163.

Costello, C., Ovando, D., Hilborn, R., Gaines, S.D., Deschenes, O., and Lester, S.E. 2012. Status and Solutions for the World's Unassessed Fisheries. Science (80-.). **338**(6106): 517–520. doi:10.1126/science.1223389.

Fronczyk, K., Kottas, A., and Munch, S. 2012. Flexible modeling for stock-recruitment relationships using Bayesian nonparametric mixtures. Environ. Ecol. Stat. **19**(2): 183–204. doi:10.1007/s10651-011-0181-0.

Glaser, S.M., Fogarty, M.J., Liu, H., Altman, I., Hsieh, C.-H., Kaufman, L., MacCall, A.D., Rosenberg, A.A., Ye, H., and Sugihara, G. 2014. Complex dynamics may limit prediction in marine fisheries. Fish Fish. **15**(4): 616–633. doi:10.1111/faf.12037.

Kokkalis, A., Eikeset, A.M., Thygesen, U.H., Steingrund, P., and Andersen, K.H. 2017. Estimating uncertainty of data limited stock assessments. ICES J. Mar. Sci. **74**(1): 69–77. doi:10.1093/icesjms/fsw145.

Legault, C.M., Bell, R.J., Cournane, J., Deroba, J.J., Fay, G., Jones, A., Miller, T.J., Muffley, B., and Wiedenmann, J. 2021. Final Report of the Index Based Methods Working Group.

Legault, C.M., and Restrepo, V.R. 1998. A Flexible Forward Age-Structured Assessment Program.

Magnusson, A., Punt, A.E., and Hilborn, R. 2013. Measuring uncertainty in fisheries stock assessment: the delta method, bootstrap, and MCMC. Fish Fish. **14**(3):

325–342. doi:10.1111/j.1467-2979.2012.00473.x.

Maunder, M.N., and Piner, K.R. 2015. Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. **72**(1): 7–18. doi:10.1093/icesjms/fsu015.

National Marine Fisheries Service. 2001. Marine Fisheries Stock Assessment Improvement Plan. Report of the National Marine Fisheries Service National Task Force for Improving Fish Stock Assessments. U.S. Dept. Commerce, NOAA Tech. Memo. NMFS-F/SPO-56. 1315 East-West Highway Silver Spring, MD 20910.

National Oceanic and Atmospheric Administration. 2006. Magnuson-Stevens Fishery Conservation and Management Act : MSA. *In* U.S. Government (2007) Public Law 94-265.

Politis, P.J., Galbraith, J.K., Kostovick, P., and Brown, R.W. 2014. Northeast Fisheries Science Center bottom trawl survey protocols for the NOAA Ship Henry B. Bigelow. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 14-06. 166 Water Street, Woods Hole, MA 02543-1026.

Prado, R., and West, M. 2010. Time series: Modeling, computation, and inference. *In* Time Series: Modeling, Computation, and Inference. doi:10.1201/9781439882757.

Punt, A.E., Tuck, G.N., Day, J., Canales, C.M., Cope, J.M., de Moor, C.L., De Oliveira, J.A.A., Dickey-Collas, M., Elvarsson, B.Þ., Haltuch, M.A., Hamel, O.S., Hicks, A.C., Legault, C.M., Lynch, P.D., and Wilberg, M.J. 2020. When are model-based stock assessments rejected for use in management and what happens then? Fish. Res. **224**: 105465. doi:10.1016/j.fishres.2019.105465.

Szuwalski, C.S., Ianelli, J.N., and Punt, A.E. 2018. Reducing retrospective patterns in stock assessment and impacts on management performance. ICES J. Mar. Sci. **75**(2): 596–609. doi:10.1093/icesjms/fsx159.

Wiedenmann, J., Free, C.M., and Jensen, O.P. 2019. Evaluating the performance of data-limited methods for setting catch targets through application to data-rich stocks: A case study using Northeast U.S. fish stocks. Fish. Res. **209**: 129–142. doi:10.1016/j.fishres.2018.09.018.

Wiedenmann, J., and Jensen, O.P. 2018. Uncertainty in stock assessment estimates for New England groundfish and its impact on achieving target harvest rates. Can. J. Fish. Aquat. Sci. **75**(3): 342–356. doi:10.1139/cjfas-2016-0484.

MANUSCRIPT

A Bayesian State-Space Approach to Improve Biomass Projections for Managing New England Groundfish

Joseph A. Langan[1], Christopher M. Legault[2], Jeremy S. Collie[1], Jason E. McNamee[3], and Gavino Puggioni[4]

[1]University of Rhode Island, Graduate School of Oceanography, Narragansett, RI 02882, USA.
[2]Northeast Fisheries Science Center, Population Dynamics Branch, Woods Hole, MA 02543, USA.
[3]Rhode Island Department of Environmental Management, Providence, RI 02908, USA.
[4]University of Rhode Island, Department of Computer Science and Statistics, Kingston, RI 02881, USA.

Corresponding Author: J.A. Langan; joseph_langan@uri.edu

Publication status: in preparation for submission to *Canadian Journal of Fisheries and Aquatic Sciences*

ABSTRACT

Fisheries management requires regularly assessing stock status and setting catch levels for the coming years. Although data-rich stock assessment models incorporating demographic and biological information are generally preferred, such approaches are often prohibited by either insufficient data or a history of poor performance. In such cases, simpler Index-Based Methods are often used to generate catch advice for a fishery without forecasting future abundance levels or quantifying scientific uncertainty, making it difficult to assess the performance of different approaches prior to implementation. To address this shortcoming, this work developed a novel Index-Based Method framework using dynamic linear models (DLMs), a flexible Bayesian state-space approach. Using simulated population data mimicking member species of the Northeast Multispecies groundfish complex, the predictive performance of candidate DLM structures were evaluated via retrospective forecasting. In both an Index-Based (age-aggregated) and Age-Based formulation, the tested DLMs displayed promising predictive ability. While further testing is needed, this preliminary evaluation suggests that DLMs may become a valuable tool in the management of fisheries for which a data-rich stock assessment approach is not possible.

INTRODUCTION

Since its first passage in 1976 and through subsequent revisions, the Magnuson-Stevens Act has provided the regulatory framework in the United States to preserve living marine resources and the economic activities that they support through its mandate that management bodies prevent overfishing, rebuild overfished stocks, and ensure a sustainable seafood supply into the future (National Oceanic and Atmospheric Administration 2006). In pursuit of these goals, fisheries scientists conduct stock assessments, generally every one to four years, in which fisheries-independent survey data are combined with catch information to assess the status of harvested fish stocks and recommend Acceptable Biological Catch levels for the following years (National Marine Fisheries Service 2001; ASMFC 2009; Wiedenmann and Jensen 2018). Although such survey and harvest data are generally compiled annually (ASMFC 2009; Politis et al. 2014), the Acceptable Biological Catch levels set during the last assessment are typically maintained until resources allow for an assessment update (National Marine Fisheries Service 2001).

Stock assessment approaches generally fall into two categories. Length- or age-structured assessments rely on models that explicitly consider the survey and catch records of a species at each stage or age throughout its post-larval life cycle in order to estimate biological reference points and develop catch advice at the scale of the true stock biomass (e.g. Legault and Restrepo 1998). While this approach is generally preferred for its greater use of biological data and desirable outputs for the management process, age-structured assessment models have for some stocks been observed to produce consistent, retrospective patterns of projection errors that lead to poor performance in maintaining stock abundance (Brooks and Legault 2016; Szuwalski et al. 2018). Defined by Mohn (1999) as "systematic inconsistencies among a series of estimates of population size based on increasing periods of data", retrospective errors are typically the result of data errors, including biased catch information, or misspecification of life history parameters, such as recruitment or natural (non-fishing) mortality, for the target fish stock (Wiedenmann and Jensen 2018; Szuwalski et al. 2018). When such patterns develop in an age-structured assessment model of a fish stock, the model may be rejected from use in the management process (Punt et al. 2020).

Strong retrospective patterns have appeared in the assessments of many stocks on the Northeast US shelf, where positive errors have led to an overestimation of stock biomass and an underestimation of fishing mortality that combine to result in Acceptable Biological Catch levels being set too high (Brooks and Legault 2016; Wiedenmann and Jensen 2018). This issue has been particularly pervasive in member stocks of the New England Multispecies groundfish complex (Deroba et al. 2010; Northeast Fisheries Science Center 2017a). For example, assessment models used to manage the Gulf of Maine Atlantic cod (*Gadus morhua*) stock suggested that the population was increasing during the late 2000s and led to higher harvest levels. Subsequent model updates, however, revealed that the stock was in fact being overfished and declining. Catch quotas were immediately and significantly cut in an effort to reverse this trend, resulting in consequential economic impacts for the regional fishing industry (Northeast Fisheries Science Center 2015; National Marine Fisheries Service 2018; Wiedenmann and Jensen 2018) While not limited to the

Northeast US, the abundance of retrospective patterns in assessment models in this region presents significant challenges to fisheries managers tasked with tracking stock status and setting catch advice.

When an age-structured assessment model is rejected due to poor performance, like a persistent, strong retrospective pattern, or cannot be fit due to data limitations, simpler approaches are employed. Known as index-based methods (IBMs), these approaches use catch information and indices of relative abundance from fisheries-independent surveys, each often aggregated across (st)ages, to develop catch advice generally without forecasting future stock abundance or estimating the involved uncertainty (Wiedenmann et al. 2019; Legault et al. 2021). IBMs vary in the calculations used to generate catch advice. Some make determinations based upon the recent trend of a smoother fit to an index of abundance, while others compare survey indices to catch records to estimate relative exploitation rates that are then adjusted toward a desired level. Several also rely upon assumptions regarding the life history parameters that may lead to poor performance in age-structured models (Wiedenmann and Jensen 2018; Szuwalski et al. 2018; Wiedenmann et al. 2019; Legault et al. 2021). Due in part to this diversity, there is not a clear consensus as to which IBM(s) are optimal in a given management situation. An additional challenge lies in the determination of reference abundance levels to target with management actions. Because IBMs do not estimate the true population abundance, there is not a clear way to determine what is a “healthy” population level and estimates often rely upon expert knowledge on a stock by stock basis. Furthermore, many authors have stressed the importance of developing full, quantified accountings of scientific uncertainty in stock assessments (Berkson et al. 2011; Punt et al. 2011; Deroba et al. 2015; Maunder and Piner 2015; Chrysafi and Kuparinen 2016) in order meet management goals. Although recent research has sought to provide guidance into the relative strengths and weaknesses of different approaches, the current situation is one in which a large amount of biological information collected for a rejected age-structured assessment goes unused, fishery managers are unable to perform risk assessment in developing catch advice, and the diversity of IBMs available may produce divergent catch advice in different situations (Wiedenmann et al. 2019; Legault et al. 2021).

There is therefore a need for an IBM that can take advantage of all available biological information to produce probabilistic forecasts with which fishery managers can make regulatory decisions and assess risk. In an effort to fill this need, this work seeks to meet the following objectives: 1) develop an IBM that produces actionable, probabilistic forecasts, 2) evaluate the performance and potential weaknesses of the developed IBM in simulations mimicking member stocks of the New England Multispecies groundfish complex, and 3) investigate the incorporation of demographic information into the IBM and the resulting changes in performance. In doing so, the results of this work will provide a new set of model options to fishery managers tasked with assessing stock abundance and developing harvest regulations in situations where an age-structured assessment model cannot be used.

METHODS

Dynamic Linear Models

This work focuses on the use of dynamic linear models (DLMs) (West and Harrison 1997) in a Bayesian framework. As state-space models, DLMs comprise two equations: (1) an observation equation defining the relationship between a latent system state and the measured variable of interest, and (2) an evolution equation describing changes in the latent system state through time.

$$y_t = \boldsymbol{F}_t \boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim N(0, V_t) \tag{1}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(0, \boldsymbol{W}_t) \tag{2}$$

$$\boldsymbol{\theta}_0 | D_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0)$$

$$V_0 \sim IG(a_v, b_v), \boldsymbol{W}_0 \sim IW(\boldsymbol{a}_w, b_w)$$

where, at time t in $1{:}T$, y_t is the observed variable of interest, $\boldsymbol{F}_t$ is the observation matrix containing covariate information, $\boldsymbol{\theta}_t$ is a vector of state variables defining the latent state, $\boldsymbol{G}_t$ is the evolution matrix describing changes in the latent state through time, ν_t is an observation error with variance V_t, and $\boldsymbol{\omega}_t$ is a vector of evolutions with variance $\boldsymbol{W}_t$. The gaussian prior on the latent state variables, given prior information D_0, has moments $\boldsymbol{m}_0$ and $\boldsymbol{C}_0$. The observation and evolution error variances follow an inverse-gamma and inverse-Wishart distribution, respectively.

The latent state(s) of the DLM model are assumed to follow a Markovian structure and can therefore be sampled iteratively, and forecasted, using a forward-filtering backward-sampling (FFBS) algorithm (Frühwirth-Schnatter 1994; Carter and Kohn 1994) derived from the Kalman Filter and Kalman Smoother (Kalman 1960). Specifically, the FFBS algorithm is used to sample the state variables $\boldsymbol{\theta}_t$, conditioned on values of $\boldsymbol{F}_t$, $\boldsymbol{G}_t$, V_t, and $\boldsymbol{W}_t$, at each time point (Appendix I). Because the annual survey data used to assess fish stocks rarely are available for more than 50 years, the evolution matrix ($\boldsymbol{G}$) and both the observation error (V) and evolution ($\boldsymbol{W}$) variances are assumed to be constant in this application. The form of the observation ($\boldsymbol{F}_t$) and evolution ($\boldsymbol{G}$) matrices are determined by the employed model structure. Sampling of $\boldsymbol{\theta}_t$, V, and $\boldsymbol{W}$ to fit each DLM is conducted using a Gibbs sampler.

Index-Based DLM for a single abundance index

We define a DLM of a single survey index of log relative abundance (S_t) to comprise two components, a trend (τ) and a regression (β_t) on log catch (H_t). In a generalized DLM form, this structure can be written as:

$$S_t = [1 \quad H_t] \begin{bmatrix} \tau_t \\ \beta_t \end{bmatrix} + \nu_t \tag{3}$$

$$\begin{bmatrix} \tau_t \\ \beta_t \end{bmatrix} = \boldsymbol{G} \begin{bmatrix} \tau_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} \omega_{\tau,t} \\ \omega_{\beta,t} \end{bmatrix} \tag{4}$$

The trend component of this model is intended to capture the broad temporal pattern in the data and could take on different forms, including a static linear trend, a random walk, a dynamic linear trend (random walk added to an otherwise static intercept term in each time step), or a first order autoregressive process. One could also define the trend as the sum of two or more of these forms.

The catch regression component requires a priori data treatment. In general, the observed relative stock abundance recorded by a fisheries-independent survey and the harvest in a given year reflect the true underlying stock abundance. As a result, there would be significant collinearity between the trend and catch regression components of the DLM if the data are not first transformed. Because abundance indices from surveys are generally considered to be more reliable than the catch information available to fisheries managers (Branch et al. 2011), the catch data are converted into catch anomalies. Specifically, a log-log linear regression of the catch history on the smoothed survey index is used to difference out the average relative exploitation rate.

The survey index is smoothed using a loess smoother with a span of 0.3, as in the PlanBSmooth IBM (Northeast Fisheries Science Center 2015) currently in use, to mitigate the inclusion of observation errors in the catch covariate. The residuals of this regression are the catch anomalies used in the catch regression component of the DLM. Despite this transformation, it is possible that the effect-per-unit catch (anomaly) will remain a function of the true stock biomass. In this case, the evolutions of the regression coefficient ($\omega_{\beta,t}$) are allowed to be correlated to the evolutions of the trend component ($\omega_{\tau,t}$) if it is assumed to be time-varying in model fitting.

The prior distributions for this and all subsequent DLM structures are defined using the same approach. The constant observation error variance (V) is characterized with an inverse-gamma distribution with prior hyperparameters $a_v = 1$ and $b_v = \frac{var(y)}{2}$, based upon the assumption that the total data variance could be allocated equally between observation error and evolution variance. Similarly, the constant evolution variance ($\boldsymbol{W}$) is characterized with an inverse-Wishart distribution where the prior hyperparameter $\boldsymbol{a}_\omega$ is defined such that $diag(\boldsymbol{a}_\omega^{-1}) = \frac{var(y)}{2p}$, where p is the number of state variables. For state variables assumed to be correlated, the corresponding off-diagonal elements of $\boldsymbol{a}_\omega$ defining the covariances are set to one half the variance terms on the main diagonal. The prior hyperparameter b_ω is set to $p + 2$. The state variables are characterized with flat gaussian prior distributions except for the regression coefficient on catch anomalies. Because the effect of catch on the population is expected to be strictly negative and have a coefficient less than one, a

$N(-0.5, 0.0625)$ prior distribution is used. Additional details are provided in Appendix II.

Multivariate extension: multiple indices of abundance

When multiple survey abundance indices are available for a fish stock of interest, IBMs currently in use often average them to create a single index (Maunder and Punt 2013; Legault et al. 2021). However, this averaging represents a loss of information. DLMs are easily adapted to model multivariate observations (Prado and West 2010) and thus a seemingly unrelated equations approach is taken here. For example, equations (3) and (4) can be rewritten as follows in a two survey case with the catch anomalies denoted by H^*:

$$\begin{bmatrix} S_{1,t} \\ S_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & H^*_{1,t} & 0 \\ 0 & 1 & 0 & H^*_{2,t} \end{bmatrix} \begin{bmatrix} \tau_{1,t} \\ \tau_{2,t} \\ \beta_{1,t} \\ \beta_{2,t} \end{bmatrix} + \boldsymbol{\nu}_t \tag{5}$$

$$\begin{bmatrix} \tau_{1,t} \\ \tau_{2,t} \\ \beta_{1,t} \\ \beta_{2,t} \end{bmatrix} = \boldsymbol{G} \begin{bmatrix} \tau_{1,t-1} \\ \tau_{2,t-1} \\ \beta_{1,t-1} \\ \beta_{2,t-1} \end{bmatrix} + \boldsymbol{\omega}_t \tag{6}$$

In this framework, estimation of the correlation within and among the trend and catch regression components of the fits to each survey allows the fitting procedure to "borrow strength" in estimating the time series of the state variables. Therefore, it can be expected that jointly modeling multiple survey indices will reduce estimation uncertainty given that they are correlated.

Hierarchical extension: demographic structure in a Stage- or Age-Based DLM

While currently employed IBMs generally do not use demographic data for a stock in setting catch advice, it is very often available in some form. If, for example, length information is available for both the survey abundance and landed catch, a stock could be split into an unfished pre-recruit (PR) stage and a harvested recruit (R) stage. These stages can then be modeled using a hierarchical structure of DLMs, in which the estimated abundance of the pre-recruit stage acts as a covariate in the model of the recruit stage with regression coefficient γ. In a case where one survey index of abundance was available, a general hierarchical model can be written as follows:

Pre-recruits

$$S_{PR,t} = \tau_{PR,t} + \nu_{PR,t} \tag{7}$$

$$\tau_{PR,t} = G\tau_{PR,t} + \omega_{PR,t} \tag{8}$$

Recruits

$$S_{R,t} = \begin{bmatrix} 1 & H_t^* & \tau_{PR,t-1} \end{bmatrix} \begin{bmatrix} \tau_{R,t} \\ \beta_t \\ \gamma_t \end{bmatrix} + \nu_{R,t} \tag{9}$$

$$\begin{bmatrix} \tau_{R,t} \\ \beta_t \\ \gamma_t \end{bmatrix} = \boldsymbol{G} \begin{bmatrix} \tau_{R,t-1} \\ \beta_{t-1} \\ \gamma_{t-1} \end{bmatrix} + \boldsymbol{\omega}_{R,t} \tag{10}$$

If age information is available for both the survey and the fishery, this stage-based model can be extended to an age-based framework where a DLM fit to each age class in the population informs the fit of the proceeding age class. Similarly, if sexual maturity information is available, an age-based model could incorporate a stock-recruit relationship in the form of a regression on the lagged abundance of older age classes in the age-1 DLM following the methods of Tableau et al. (2019). In doing so, this Age-Based DLM representation can capture the demographic structure of a population in a manner similar to a traditional age-structured assessment model without requiring strict assumptions about life history parameters or catch information.

In the case that multiple age-explicit survey indices of abundance are available, this age-based hierarchical framework can bet fit with a common set of state variables. With known or estimated ages assigned to the abundance data, differences in survey selectivity by age do not present a challenge as they would in modeling abundance aggregated across ages. Because the DLMs are fit on the log scale, the survey indices should differ by an estimable intercept term, which may be static or dynamic if there is evidence to suggest divergent trends in catchability. Therefore, the multivariate DLM for each age can be fit with a single set of state variables describing the trend, catch regression, and regression on the prior age plus additional static or dynamic intercept terms. In a case where two survey indices of abundance are available, an Age-Based model with common state variables can be written as follows:

Age-1

$$\begin{bmatrix} S_{1,1,t} \\ S_{2,1,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \tau_{1,t} \\ \mu_{2,1} \end{bmatrix} + \boldsymbol{\nu}_{1,t} \tag{11}$$

$$\begin{bmatrix} \tau_{1,t} \\ \mu_{2,1} \end{bmatrix} = \boldsymbol{G} \begin{bmatrix} \tau_{1,t-1} \\ \mu_{12,1} \end{bmatrix} + \boldsymbol{\omega}_{1,t} \tag{12}$$

Age-2 or older

$$\begin{bmatrix} S_{1,A,t} \\ S_{2,A,t} \end{bmatrix} = \begin{bmatrix} 1 & H_t^* & B_{a-1,t} & 0 \\ 1 & H_t^* & B_{a-1,t} & 1 \end{bmatrix} \begin{bmatrix} \tau_{a,t} \\ \beta_{a,t} \\ \gamma_{a,t} \\ \mu_{12,a} \end{bmatrix} + \boldsymbol{\nu}_{a,t} \tag{13}$$

$$\begin{bmatrix} \tau_{a,t} \\ \beta_{a,t} \\ \gamma_{a,t} \\ \mu_{12,a} \end{bmatrix} = \boldsymbol{G} \begin{bmatrix} \tau_{a,t-1} \\ \beta_{a,t-1} \\ \gamma_{a,t-1} \\ \mu_{12,a} \end{bmatrix} + \boldsymbol{\omega}_{a,t} \tag{14}$$

Where $S_{S,a,t}$ is an observation of age a for survey S at time t, $\tau_{a,t}$ is a common trend for age a shared across surveys, $\beta_{a,t}$ and $\gamma_{a,t}$ are the regression coefficients on the catch anomalies and the estimated abundance of the prior age class ($B_{a-1,t}$), respectively, shared across surveys, and $\mu_{S_1 S_2,a}$ is a constant intercept term for age a to account for the catchability difference between survey S_1 and survey S_2.

In fitting a single set of state variables for each modeled age class, this representation of an age-based model can identify common abundance trends across surveys if differences in catchability are assumed constant. Because each age is modeled individually, overall uncertainty will be reduced while differences in survey selectivity will not impact the fits. The additional state variable $\gamma_{a,t}$ in this model compared to the Index-Based approach is characterized with a $N(1, 0.25)$ prior distribution, corresponding to the belief that the correlations among sequential age classes should be strictly positive and have a coefficient less than or slightly larger than one. Additional details are provided in Appendix II.

Simulated case studies

To evaluate the performance of the Index-Based and Age-Based DLMs in forecasting relative abundance of fish stocks, realistic, simulated data series were generated using the PopSim tools provided in the NOAA Fisheries Toolbox made available by NOAA Fisheries' Office of Science and Technology (https://nmfs-fish-tools.github.io/). Specifically, three stocks from the New England Multispecies groundfish complex that display varying life history characteristics and have produced retrospective patterns in age-structured assessment models were chosen for simulation: 1) Georges Bank Atlantic cod, 2) Georges Bank yellowtail flounder (*Limanda ferruginea*), and 3) witch flounder (*Glyptocephalus cynoglossus*). In each case, the simulations were initialized using the values of life history parameters, recruitment, and catch history from the recent assessments for each stock (Legault et al. 2013, 2014; Northeast Fisheries Science Center 2013, 2015, 2017a, 2017b) to create scenarios similar to what would be encountered in fisheries management. The simulated stocks were each sampled by three surveys with characteristics (catchability, selectivity, and variability) set to be like those of important surveys used to track abundance for their real-world counterparts. For the Georges Bank stocks, these surveys consisted of the Northeast Fisheries Science Center Spring and Fall bottom trawl surveys (Politis et al. 2014) and the Canadian Department of Fisheries and Oceans annual trawl survey (Chadwick et al. 2007). For witch flounder, the Northeast Fisheries Science Center surveys were also used in addition to the Maine-New Hampshire Inshore survey conducted by the Maine Department of Marine Resources (https://www.maine.gov/dmr/science-research/projects/trawlsurvey/index.html). The features of, and notable differences among, the simulated stocks are summarized in Table 1 and illustrated in Figure 1. An example of the age-specific data available for modeling is shown for the Georges Bank yellowtail flounder stock in Figure 2

Additional, modified simulated scenarios were generated for the Georges Bank yellowtail flounder stock in order to study different challenging data patterns that may be encountered in management. Sharp changes in recruitment and natural mortality, which have both been implicated in causing errors in stock assessment, are present in the base yellowtail flounder scenario detailed above. To separate and study the effects of these processes, a new “Stable Recruitment” (SR) scenario was created for yellowtail flounder in which recruitment was stablized at the end of the time series by repeating the 1993-2003 observations from 2004-2014. Very large measurement error variance can also pose difficulties in the fitting of assessment models by masking signals of population dynamics and harvest effects within the data. To investigate such effects in the DLM framework, the base yellowtail flounder scenario was repeated with the survey coefficients of variation each multiplied by three to create a “High Measurement Error” (HME) case.

Evaluation of model performance

Index-Based and Age-Based DLMs were fit to 50 simulations of each scenario (Figure 1). Each model fit was performed using a Gibbs sampler with 30,000 iterations, where 20,000 iterations were discarded as the burn-in period and a thinning interval of 10 was applied to yield 1,000 posterior samples of each parameter. Model convergence and sampling of the estimated posterior distributions were evaluated using the Geweke diagonal and the effective sample size for each parameter with the R package “coda” (Plummer et al. 2006). It is important to note that each scenario produced a single stock trajectory common to all simulations, where simulations differed in the survey observations. Therefore, the results must be interpreted as performance in a particular scenario that may be faced by fisheries managers under variations in patterns of observation errors in the data used to fit the model. Predictive performance was tested using a retrospective forecasting approach similar to Brooks and Legault (2016). For each simulation, a DLM was fit to all but the last three years of the time series and a three-year forecast was generated to compare to the true future values. These forecasts used catch anomalies calculated with the true future catch and abundance values such that predictive performance could be tested when future harvest rates were known. The specific model structures within each DLM class used for a given scenario were determined by evaluating predictive performance in a subset of simulations and verifying that the fitted parameters agreed with a priori hypotheses (e.g. a negative coefficient in the catch regression and a positive coefficient in regressions on the prior age class).

Several performance metrics were calculated for each DLM class and scenario based upon the simulated survey observations and the true relative stock biomass scaled to the levels of the surveys using their relative catchabilities. Forecast performance was summarized using the mean percent prediction error and median absolute percent prediction error compared to the true relative abundance averaged across surveys for each year of the forecast. Here, the average prediction error between the median of the posterior predictive distributions and the simulated data was calculated across the three surveys for each year of the forecast period. The mean and median absolute errors across all 50 simulations were then calculated for each year using these survey-averaged values. To determine if the estimated forecast uncertainty was well-calibrated, the proportion of all instances in which the simulated survey observations fell within the 95% posterior predictive highest density interval was also calculated using the R package “HDInterval” (Meredith and Kruschke 2020). Finally, the median ratio across surveys of the root mean squared model residuals to the root mean squared measurement errors (RMSE) was calculated for both the training data and the forecast period to measure the relative magnitude of the model residuals to the observation errors of the data. The focus of this work was an evaluation of forecast performance of candidate DLM model structures and not on methods to generate catch advice from those forecasts. However, potential approaches to generate catch advice for use in management are outlined in the discussion.

RESULTS

Two Index-Based structures and one Age-Based DLM structure were selected as the best performing across the five scenarios based on testing using a subset of the simulated datasets. In all cases, fitting a dynamic coefficient in the catch regression component of the model either did not appear to improve performance in prediction or resulted in an estimated coefficient time series that did not follow expectations given the observed abundance trend. Similarly, a static relationship between age classes was selected in the Age-Based model for the same reasons. The observation error variance was also assumed to be constant in all models. In the Index-Based fits, the trend component was characterized with a random walk for Georges Bank cod and with a dynamic linear trend in all other cases. A random walk for age-1 and a first order autoregressive process for all the older age classes was used in in all cases for the Age-Based model. Full model specifications are provided in Appendix II.

Index-Based DLM Performance

The Index-Based DLMs generally exhibited strong predictive performance for the base simulations of Georges Bank cod, Georges Bank yellowtail flounder, and witch flounder (Table 2, Figures 3 and S1). The mean and median absolute prediction errors were all less than 75% in magnitude, with the largest errors observed in the DLM fit of yellowtail flounder. Unsurprisingly, prediction error increased with the length of the forecast, where mean errors across surveys were generally below 25% in the first year of the forecast period (Table 2, Figure 4). Because the data were fit on the log scale and the errors were assumed to be normally distributed, the largest prediction errors were consistently positive. The average magnitude of forecast errors was also consistently and strongly correlated to the accuracy of the model-estimated abundance in the terminal year of the data used in fitting (Figure 5). When large forecast errors occurred, they generally resulted from simulations in which multiple consecutive observation errors of the same sign were recorded at the end of the time series (ex. top two rows of Figure S1d). The median RMSE ratios between the model residuals and the observation errors were between 0.48 and 0.63 in the training data interval, indicating that the Index-Based DLM fit was able to separate signal from noise. During the forecast period, the median RMSE ratios were between 0.92 and 1.65. For the cod and yellowtail flounder scenarios, the comparison of the forecast uncertainty to the observations that occurred during the forecast window suggested that the estimated uncertainty was slightly too large.

The modified Georges Bank yellowtail flounder scenarios exhibited similar error patterns to the base scenario. In the Stable Recruitment case, the prediction performance metrics were slightly better than the results for the base yellowtail flounder simulation and particularly during the first two years of the forecast (Table 2). This limited improvement suggests that the Index-Based DLM may not be dramatically impacted by sharp changes in recruitment in proximity to the end of the training data. Predictive performance was less strong in the High Measurement Error scenario than for the base or Stable Recruitment yellowtail flounder simulations (Table 2). While the median absolute prediction errors were not a lot larger than in the base scenario, the larger differences occurring during the first two years of the forecast period indicates that decreased accuracy does not only occur in longer forecasts. The mean prediction errors, meanwhile, were nearly twice as large as in the base scenario. These results suggest that the Index-Based DLM will exhibit declining prediction performance and increasing frequency of large errors as measurement error increases. However, the median RMSE ratio during the forecast period was 0.81. Therefore, the posterior predictive median forecast still often provided a better estimate of relative stock biomass than the highly variable future observations.

Age-Based DLM Performance

The Age-Based DLM approach also exhibited strong predictive performance across the three base stock simulation scenarios (Table 2, Figures 3, 6, and S1). The mean and median absolute prediction errors were less than 30% throughout the forecast interval in all cases. As in the Index-Based approach, prediction error magnitude increased with the length of the forecast (Table 2). In general, the error patterns for the DLM fit to each age class and their relationship to estimation errors in the terminal year of the training data time series were similar to the results obtained for the Index-Based case. During the data time series used to fit the Age-Based DLMs, the median RMSE ratios between the model residuals and the observation errors were 0.51 or 0.52 for the three base scenarios. The median RMSE ratios in the forecast interval, meanwhile, were between 0.68 and 1.13. For all three stock scenarios, the estimated forecast uncertainty appeared to be too small (Table 2). However, the coverage of the posterior predictive distributions relative to the observations for the DLMs fit to each individual age class were similar to the results obtained in the Index-Based case (Figure 6). Only when the age-specific DLM fits were summed into an aggregate model estimate was the forecast uncertainty too small, likely due to correlation among the fits to each age.

In contrast to the Index-Based case, the predictive performance of the Age-Based DLM was worse in the Stable Recruitment than the base yellowtail flounder simulations. In the High Measurement Error scenario, the mean and median prediction errors were much larger than the base case and particularly so beginning in the second year of the forecast. While the median RMSE ratio for the training data suggested that the median model fit was close to the true simulated relative yellowtail flounder abundance, the ratio during the forecast period was over three due in part to patterns of observation errors at the end of the training data that resulted in large terminal year estimation errors (Figure S1d). The large prediction errors observed in the High Measurement Error scenario therefore indicate that the performance of the Age-Based approach may be sensitive to large observation error variance among the modeled age classes.

Model Comparison

As could be expected of a model based upon greater biological information, the Age-Based DLM approach performed better and with less uncertainty (Figures 3 and S1) than the Index-Based approach across the three base simulation scenarios (Table 2). In general, the mean and median absolute prediction errors were smaller for Age-

Based DLM and particularly after the first year of the forecast. However, the Age-Based DLM was not necessarily superior in the fits of all simulations. For the Georges Bank cod scenario, for example, the Age-Based DLM had a larger range of survey-averaged prediction errors despite scoring better in the mean and median absolute error metrics (Figure 5). Inspection of the model fits suggested these large errors may have stemmed from sharp increasing trends in abundance observed for individual age classes that did not manifest during the forecast period. However, the correlation between the forecast errors of the Index-Based and Age-Based approaches in this scenario was 0.65, indicating that both model structures tend to produce large errors in similar situations. Interestingly, the median RMSE ratios for the training data period were very similar between the DLM approaches, but the Age-Based models had lower values for the forecast period. This suggests that both approaches performed similarly in identifying observation errors during fitting, but that the Age-Based DLM produced better median forecasts.

This comparative performance trend was matched in the Stable Recruitment yellowtail flounder scenario. The differences in mean and median absolute prediction error and in the median RMSE ratios were smaller between the Age-Based and Index-Based approaches than the base yellowtail flounder scenario, but the performance pattern was similar. In the High Measurement Error scenario, however, the Age-Based DLM exhibited much larger prediction errors and a larger median RMSE ratio during the forecast period. This result suggests that a simpler, smoother model may perform better when the observed data exhibit a very large observation error variance that could mask relationships among age classes.

DISCUSSION

The results of the simulation testing performed in this work indicate that DLMs hold promise as a tool for managing fish stocks for which a conventional stock assessment model cannot be reliably used. Across the generated simulation scenarios mimicking population dynamics observed in the Northeast Multispecies groundfish stock complex, the tested models exhibited strong predictive performance and realistic quantifications of forecast uncertainty. Due to the flexible nature of DLMs, the model structure employed for a given stock can be drawn from a spectrum of complexity depending upon the available data. The Index-Based and Age-Based structures evaluated here represent two interior points on the possible spectrum. Although the inclusion of additional biological information allowed the Age-Based DLMs to generally perform better than their Index-Based counterparts, the results indicate that using a simpler model may be better under certain circumstances. Further development of the DLM approach could include alternative model structures or different approaches for incorporating catch information to maximize the potential of this tool in fisheries management.

Despite making use of limited biological information, the Index-Based DLMs performed well in prediction across simulation scenarios. Similar to the results of Brooks and Legault (2016) and Wiedenmann and Jensen (2018), the magnitude of the observed prediction errors were strongly correlated with the error in the terminal abundance estimate in the training data. These errors also tended to grow with the length of the forecast, suggesting that these models should be updated and used to generate new forecasts as often as data availability and resources allow in a management setting. Among the base scenarios, the prediction errors appeared to be the smallest for witch flounder and the largest for Georges Bank yellowtail flounder. Given that changes in natural mortality and recruitment have both been implicated as sources of forecast error in stock assessments (Maunder and Piner 2015; Brooks and Legault 2016; Wiedenmann and Jensen 2018), these characteristics in the yellowtail flounder simulation may be responsible for the larger errors. However, the Index-Based DLM fit of the Stable Recruitment modified yellowtail flounder scenario did not perform much better than the base scenario and the change in natural mortality occurred in the middle of the time series. Because it is unclear if changes in natural mortality and recruitment or another characteristic were responsible for the larger prediction errors observed for yellowtail flounder, additional testing would be required to better measure the impacts of changes in individual population parameters on forecast accuracy.

In general, the Age-Based DLM structure appeared to perform as well or better than the Index-Based approach. An improvement in the magnitude of the prediction errors and the median RMSE ratio during the forecast period was observed for all of the base simulation scenarios. It should be noted, however, that the Age-Based approach did produce a larger range of errors for the Georges Bank Atlantic cod scenario. This indicates that the relative strengths of the two approaches tested here may not always be clear cut. Unlike the Index-Based approach, the estimated forecast uncertainty was too small in the Age-Based model fits. However, the coverage of the 95% highest density forecast intervals for each individual age class was similar to that of the Index-Based DLM. Therefore, the small uncertainty estimate recorded for the aggregated abundance forecast is likely an artefact of the correlation among the summed age-specific model fits. The one scenario in which the Age-Based approach performed worse than its Index-Based counterpart was in the High Measurement Error yellowtail flounder simulation. Given that consecutive, large measurement errors of the same sign at the end of the fitted time series were observed to result in larger measurement errors, it is possible that such a pattern is more likely to occur across age classes when the measurement error variance is high. If so, these errors may be compounded through the summation of the individual model fits into an aggregate whole. Additionally, large errors may also mask the relationships between age classes that help provide strength to the Age-Based approach. Therefore, these results indicate that a simpler, smoother model like the Index-Based DLM may be more desirable in cases where measurement error is very high.

The relative strengths of the Index-Based and Age-Based DLM approaches highlight the benefit of the modular, additive nature of this model framework. If a single survey index of abundance is available for a fish stock with no reliable catch information, a simple, univariate DLM consisting of only a trend component could be fit and used in a similar manner to the PlanBSmooth IBM (Northeast Fisheries Science Center 2015). As demonstrated in this work, additional surveys of abundance, catch information, or length or age data can be built into an Index-Based, Stage-Based, or Age-Based approach without needing to average or summarize data and risk a detrimental loss of information (Maunder and Punt 2013). There is also additional capacity in this modeling framework beyond the scope of this work. Several previous studies have noted the importance of including the effects of environmental conditions on population processes in stock assessment (Szuwalski et al. 2015; Szuwalski and Hollowed 2016; Xu et al. 2018). Within the DLM framework, such environmental conditions can simply be added as a covariate whose effect may be allowed to change over time if such a dynamic coefficient is supported by existing knowledge of a stock. Additionally, significant work in recent years has been devoted to developing multispecies models, both in order to capture the effects of trophic interactions within a biological community or to inform the assessments of data poor stocks (Punt et al. 2011; Curti et al. 2013; Gaichas et al. 2017). The multivariate DLM structure used here to model multiple, correlated survey indices of abundance could be employed to model multiple species whose abundances tend to covary. Similarly, the hierarchical DLM structure used to model demographic information could be used to capture the direct trophic relationships among predator and prey species. In this manner, the developed DLM model framework does not represent a set of unrelated models that fisheries managers must choose between, but rather a spectrum of internally consistent models that can be adapted to make the most of available biological information.

Beyond this modular capability, the DLM framework has other characteristics that past research has identified as desirable in assessing fish stocks. Perhaps most significantly, the importance of assessing and quantifying uncertainty in stock assessment has been repeatedly stressed (Berkson et al. 2011; Magnusson et al. 2013; Maunder and Piner 2015; Kokkalis et al. 2017). Unlike many of the IBMs currently in use (Wiedenmann et al. 2019; Legault et al. 2021), fitted DLMs produce true forecasts of future stock abundance with quantified uncertainty that can be used to assess risk among potential management actions.

Furthermore, the DLMs produced here are fit in a Bayesian framework, which has been identified as well-suited to making use of available biological information and expert input and producing accurate accountings of scientific uncertainty in stock assessment (Fronczyk et al. 2012; Magnusson et al. 2013; Chrysafi and Kuparinen 2016). The use of the Bayesian approach also allows for a convenient imputation of missing data (Chrysafi and Kuparinen 2016), which could occur if, for example, a survey was not conducted in a year due to mechanical issues or as a result of the recent restrictions associated with the COVID-19 pandemic (Appendix III). This ability to produce probabilistic forecasts of abundance also has benefits in optimizing model structure and testing performance. Due in part to their reliance on decision rules to generate catch advice as opposed to predicting future stock abundance, there is not a consensus on which IBMs perform best under what circumstances and testing different approaches may involve a large simulation experiment (Wiedenmann et al. 2019; Legault et al. 2021). Using the retrospective forecasting approach developed by Brooks and Legault (2016), one could repeatedly fit candidate DLM structures to progressively longer time series of available data and compare the forecasts to the known future observations (Appendix III). While imperfect and prone to errors if underlying stock dynamics change, such testing could allow fisheries managers to identify the structure best-suited to modeling a particular fish stock and to gain a preliminary estimate of potential prediction performance. Finally, the ability of DLMs to capture changing population processes (e.g. natural mortality) may boost prediction accuracy while avoiding explicit assumptions about those processes that have resulted in significant errors in past stock assessments (Szuwalski et al. 2018).

It is difficult to directly compare the performance of the DLM approaches evaluated in this work to other IBMs used in management. Because the IBMs currently in use generally do not make forecasts of abundance or estimate the associated scientific uncertainty (Legault et al. 2021), the only way to draw comparative conclusions is through a detailed simulation experiment (Wiedenmann et al. 2019) beyond the scope of this work. However, such an experiment was recently conducted to test a variety of IBMs on simulated stocks that exhibited features that would have likely led to the rejection of age-structured assessments in a "real world" management process (Legault et al. 2021). This work included a limited, preliminary version of the Index-Based DLM and found that it performed as well or better than the other IBMs in recovering and maintaining stock abundance over the long-term. While the tested DLM model was found to be overly conservative in limiting harvest, this may have been a function of the chosen method to generate catch advice. Additionally, the DLM performed similarly well both in cases where natural mortality or catch were misspecified in the simulation experiment. While further testing is required to fully interpret the DLM performance in the IBMWG experiment, these results suggest that the different classes of DLMs described here hold great promise in providing fisheries managers with tools to evaluate potential harvest regulations and assess the associated uncertainty.

To provide this utility in management, however, the forecast produced by a DLM model for a given stock must be converted to catch advice. While this work focused on developing model structures and assessing performance in prediction, it would be straightforward to produce catch advice using these DLMs. If reliable catch information is not available for a given stock, an Index-Based DLM consisting of only a trend component could be fit and used to generate catch advice in a similar manner to that of the PlanBSmooth IBM (Northeast Fisheries Science Center 2015). This approach could be extended to an Age-Based DLM in which the fitted state variables only describe abundance trends-at-age and the relationships among successive ages. When catch information is incorporated into the model, one can use an optimization algorithm based on the estimated latent state in the terminal year of the training data and the filtering equations (above) to solve for a harvest level that will result in the population reaching a given target level by the end of the forecast interval with a desired level of certainty. For example, the use of log-transformed data in model-fitting results in larger overestimation errors on average. Given that overoptimistic forecasts that lead to higher harvest levels are more problematic in maintaining stock abundance over the long-term (Glaser et al. 2014; Brooks and Legault 2016; Szuwalski et al. 2018), users could consider requiring higher forecast certainty when providing catch advice that supports raising harvest levels. This could take the form of developing catch advice such that the 40% quantile of the posterior predictive distribution meets a target abundance at the end of the forecast window in place of the mean or median. Determining the quantile that best mitigates these risks while allowing sufficient harvest, however, will be a challenge that requires significant investigation.

These potential approaches to developing catch advice, however, rely upon a target abundance level being identified. Like other IBMs (Wiedenmann et al. 2019; Legault et al. 2021), the developed DLM structures model data on the relative (survey) scale and do not attempt to estimate the true population abundance. Therefore, users will not have clear biological reference points with which to assess current stock status and determine a desirable future population trajectory as they would when using an age-structured stock assessment model. Determining target abundance levels will therefore have to rely upon expert knowledge of each individual stock being modeled. Although imperfect, the ability of the DLM framework to generate forecasts of abundance and estimate uncertainty still represents a step forward in capability compared to current IBMs that will prove valuable in management.

Another consideration in the use of the developed DLM framework to develop catch advice for fisheries management is the differences among the Index-Based and Age-Based approaches. Because the Index-Based DLM only assumes correlation among survey indices of abundance, simulation cases emerged in which the forecasts for different surveys had divergent (positive and negative) slopes. If the surveys were representing different parts of the population that are indeed exhibiting divergent trajectories, this result would be desirable. However, it could also occur if differing patterns of observation errors among indices impacted the model estimates of abundance trends. In the case where a user is implementing the Index-Based DLM due to a lack of available size, life stage, or age information, there would not be a way to determine the root cause the divergent pattern in the forecasts and thus additional caution is warranted. When demographic structure is included in the model, however, the ability to represent multiple survey indices with a single set of state variables results in forecasts with a consistent trajectory for generating catch advice with less involved uncertainty. While testing different approaches or decision rules for using

DLM forecasts to create catch advice was outside the scope of this work, this prospective ability to directly evaluate potential harvest levels against probabilistic forecasts of abundance will be very valuable in management.

In summary, the developed DLM approach appears promising as a tool to forecast future abundance and manage index-based stocks. In the demonstration cases tested here using simulated data, strong predictive performance was observed for both the Index-Based and Age-Based approaches. Despite this performance, it is recommended that additional model development and evaluations using realistic management scenarios are undertaken. It is likely that additional model structures not tested here or alternative methods for incorporating catch information may produce better results. Nevertheless, the results demonstrate that the DLM framework addresses many of the needs and shortcomings identified by past research of IBMs currently in use. The modular ability of the model structures and development of priors can allow for the use of all available biological information and expert input to manage a stock. The majority of global fish stocks lack the data necessary for conventional stock assessment methods (Costello et al. 2012). With additional testing and validation, the DLM framework may be able to become a valuable tool for fisheries managers to use the information at hand to be more successful in maintaining productive marine fisheries around the globe.

ACKNOWLEDGEMENTS

This research was funded by an award (Number NA180AR4170321SG) from the National Marine Fisheries Service-Sea Grant Joint Fellowship Program and by a New England Fishery Management Council Contract for Service (Contract Number 18-24).

We would like to thank the staff and Scientific and Statistical Committee of the New England Fishery Management Council and the members if the Index-Based Methods Working Group for providing thoughtful feedback and suggestions that improved this work.

REFERENCES

ASMFC, A.S.M.F.C. 2009. Guide to Fisheries Science and Stock Assessments. Report pursuant to National Oceanic and Atmospheric Administration Grant No. NA05NMF4741025.

Berkson, J., Barbieri, L., Cadrin, S.X., Cass-Calay, S., Crone, P., Dorn, M., Friess, C., Kobayashi, D., Miller, T.J., Patrick, W.S., Pautzke, S., Ralston, S., and Trianni, M. 2011. Calculating Acceptable Biological Catch for Stocks That Have Reliable Catch Data Only (Only Reliable Catch Stocks – ORCS).

Branch, T.A., Jensen, O.P., Ricard, D., Ye, Yi., and Hilborn Ray. 2011. Contrasting Global Trends in Marine Fishery Status Obtained from Catches and from Stock Assessments. Conserv. Biol. **25**(4): 777–786. doi:10.1111/j.1523-1739.2011.01687.x.

Brooks, E.N., and Legault, C.M. 2016. Retrospective forecasting — evaluating performance of stock projections for New England groundfish stocks. Can. J. Fish. Aquat. Sci. **73**(6): 935–950. doi:10.1139/cjfas-2015-0163.

Carter, C.K., and Kohn, R. 1994. On Gibbs Sampling for State Space Models. Biometrika **81**(3): 541. doi:10.2307/2337125.

Chadwick, E., Brodie, W., Colbourne, E., Clark, D., Gascon, D., and Hurlbut, T. 2007. History of annual multi-species trawl surveys on the Atlantic Coast of Canada. AZMP Bull. **6**: 25–42.

Chrysafi, A., and Kuparinen, A. 2016. Assessing abundance of populations with limited data: Lessons learned from data-poor fisheries stock assessment. Environ. Rev. **24**(1): 25–38. doi:10.1139/er-2015-0044.

Costello, C., Ovando, D., Hilborn, R., Gaines, S.D., Deschenes, O., and Lester, S.E. 2012. Status and Solutions for the World's Unassessed Fisheries. Science (80-.). **338**(6106): 517–520. doi:10.1126/science.1223389.

Curti, K.L., Collie, J.S., Legault, C.M., and Link, J.S. 2013. Evaluating the performance of a multispecies statistical catch-at-age model. Can. J. Fish. Aquat.

Sci. **70**(3): 470–484. doi:10.1139/cjfas-2012-0229.

Deroba, J.J., Butterworth, D.S., Methot, R.D., De Oliveira, J.A.A., Fernandez, C., Nielsen, A., Cadrin, S.X., Dickey-Collas, M., Legault, C.M., Ianelli, J., Valero, J.L., Needle, C.L., O'Malley, J.M., Chang, Y.-J., Thompson, G.G., Canales, C., Swain, D.P., Miller, D.C.M., Hintzen, N.T., Bertignac, M., Ibaibarriaga, L., Silva, A., Murta, A., Kell, L.T., de Moor, C.L., Parma, A.M., Dichmont, C.M., Restrepo, V.R., Ye, Y., Jardim, E., Spencer, P.D., Hanselman, D.H., Blaylock, J., Mood, M., and Hulson, P.-J.F. 2015. Simulation testing the robustness of stock assessment models to error: some results from the ICES strategic initiative on stock assessment methods. ICES J. Mar. Sci. **72**(1): 19–30. doi:10.1093/icesjms/fst237.

Deroba, J.J., Shepherd, G., Gregoire, F., Nieland, J., and Rago, P. 2010. Stock assessment of Atlantic mackerel in the Northwest Atlantic for 2010. Transboundary Resources Assessment Committee, Reference Document 2010/01.

Fronczyk, K., Kottas, A., and Munch, S. 2012. Flexible modeling for stock-recruitment relationships using Bayesian nonparametric mixtures. Environ. Ecol. Stat. **19**(2): 183–204. doi:10.1007/s10651-011-0181-0.

Frühwirth-Schnatter, S. 1994. Data Augmentation and Dynamic Linear Models. J. Time Ser. Anal. **15**(2): 183–202. doi:10.1111/j.1467-9892.1994.tb00184.x.

Gaichas, S.K., Fogarty, M., Fay, G., Gamble, R., Lucey, S., and Smith, L. 2017. Combining stock, multispecies, and ecosystem level fishery objectives within an operational management procedure: simulations to start the conversation. ICES J. Mar. Sci. **74**(2): 552–565. doi:10.1093/icesjms/fsw119.

Glaser, S.M., Fogarty, M.J., Liu, H., Altman, I., Hsieh, C.-H., Kaufman, L., MacCall, A.D., Rosenberg, A.A., Ye, H., and Sugihara, G. 2014. Complex dynamics may limit prediction in marine fisheries. Fish Fish. **15**(4): 616–633. doi:10.1111/faf.12037.

Kalman, R.E. 1960. A New Approach to Linear Filtering and Prediction Problems. J. Basic Eng. **82**(1): 35–45. doi:10.1115/1.3662552.

Kokkalis, A., Eikeset, A.M., Thygesen, U.H., Steingrund, P., and Andersen, K.H. 2017. Estimating uncertainty of data limited stock assessments. ICES J. Mar. Sci. **74**(1): 69–77. doi:10.1093/icesjms/fsw145.

Legault, C.M., Alade, L.A., Gross, W.E., and Stone, H.H. 2013. Stock Assessment of Georges Bank Yellowtail Flounder for 2013. TRAC Ref. Doc. 2013/01. Available from http://www.nefsc.noaa.gov/saw/trac/.

Legault, C.M., Alade, L.A., Gross, W.E., and Stone, H.H. 2014. Stock Assessment of Georges Bank Yellowtail Flounder for 2014. TRAC Ref. Doc. 2014/01.

Legault, C.M., Bell, R.J., Cournane, J., Deroba, J.J., Fay, G., Jones, A., Miller, T.J., Muffley, B., and Wiedenmann, J. 2021. Final Report of the Index Based Methods Working Group.

Legault, C.M., and Restrepo, V.R. 1998. A Flexible Forward Age-Structured Assessment Program.

Magnusson, A., Punt, A.E., and Hilborn, R. 2013. Measuring uncertainty in fisheries stock assessment: the delta method, bootstrap, and MCMC. Fish Fish. **14**(3): 325–342. doi:10.1111/j.1467-2979.2012.00473.x.

Maunder, M.N., and Piner, K.R. 2015. Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. **72**(1): 7–18. doi:10.1093/icesjms/fsu015.

Maunder, M.N., and Punt, A.E. 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. **142**: 61–74. doi:10.1016/j.fishres.2012.07.025.

Meredith, M., and Kruschke, J. 2020. HDInterval: Highest (Posterior) Density Intervals. Available from https://cran.r-project.org/package=HDInterval.

Mohn, R. 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES J. Mar. Sci. **56**(4): 473–488. doi:10.1006/jmsc.1999.0481.

National Marine Fisheries Service. 2001. Marine Fisheries Stock Assessment Improvement Plan. Report of the National Marine Fisheries Service National Task Force for Improving Fish Stock Assessments. U.S. Dept. Commerce, NOAA Tech. Memo. NMFS-F/SPO-56. 1315 East-West Highway Silver Spring, MD 20910.

National Marine Fisheries Service. 2018. Fisheries Economics of the United States, 2016. U.S. Dept. Commer. NOAA Tech. Memo. **NMFS-F/SPO**: 243p.

National Oceanic and Atmospheric Administration. 2006. Magnuson-Stevens Fishery Conservation and Management Act : MSA. *In* U.S. Government (2007) Public Law 94-265.

Northeast Fisheries Science Center. 2013. 55th Northeast Regional Stock Assessment Workshop (55th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 13-11. 166 Water Street, Woods Hole, MA 02543-1026.

Northeast Fisheries Science Center. 2015. Operational Assessment of 20 Northeast Groundfish Stocks, Updated Through 2014. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 15-24. 166 Water Street, Woods Hole, MA 02543-1026.

Northeast Fisheries Science Center. 2017a. Operational Assessment of 19 Northeast Groundfish Stocks, Updated Through 2016. NEFSC Ref Doc. 17-17. *In* Northeast Fisheries Science Center Reference Document. Woods Hole, MA. Available from http://www.nefsc.noaa.gov/publications/.

Northeast Fisheries Science Center. 2017b. 62nd Northeast Regional Stock Assessment Workshop (62nd SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 17-03. 166 Water Street, MA 02543-1026.

Plummer, M., Best, N., Cowles, K., and Vines, K. 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News **6**: 7–11.

Politis, P.J., Galbraith, J.K., Kostovick, P., and Brown, R.W. 2014. Northeast Fisheries Science Center bottom trawl survey protocols for the NOAA Ship Henry B. Bigelow. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 14-06. 166 Water Street, Woods Hole, MA 02543-1026.

Prado, R., and West, M. 2010. Time series: Modeling, computation, and inference. *In* Time Series: Modeling, Computation, and Inference. doi:10.1201/9781439882757.

Punt, A.E., Smith, D.C., and Smith, A.D.M. 2011. Among-stock comparisons for improving stock assessments of data-poor stocks: the “Robin Hood” approach. ICES J. Mar. Sci. **68**(5): 972–981. doi:10.1093/icesjms/fsr039.

Punt, A.E., Tuck, G.N., Day, J., Canales, C.M., Cope, J.M., de Moor, C.L., De Oliveira, J.A.A., Dickey-Collas, M., Elvarsson, B.Þ., Haltuch, M.A., Hamel, O.S., Hicks, A.C., Legault, C.M., Lynch, P.D., and Wilberg, M.J. 2020. When are model-based stock assessments rejected for use in management and what happens then? Fish. Res. **224**: 105465. doi:10.1016/j.fishres.2019.105465.

Szuwalski, C.S., and Hollowed, A.B. 2016. Climate change and non-stationary population processes in fisheries management. ICES J. Mar. Sci. **73**(5): 1297–1305. doi:10.1093/icesjms/fsv229.

Szuwalski, C.S., Ianelli, J.N., and Punt, A.E. 2018. Reducing retrospective patterns in stock assessment and impacts on management performance. ICES J. Mar. Sci. **75**(2): 596–609. doi:10.1093/icesjms/fsx159.

Szuwalski, C.S., Vert-Pre, K.A., Punt, A.E., Branch, T.A., and Hilborn, R. 2015.

Examining common assumptions about recruitment: a meta-analysis of recruitment dynamics for worldwide marine fisheries. Fish Fish. **16**(4): 633–648. doi:10.1111/faf.12083.

Tableau, A., Collie, J.S., Bell, R.J., and Minto, C. 2019. Decadal changes in the productivity of New England fish populations. Can. J. Fish. Aquat. Sci. **76**(9): 1528–1540. doi:10.1139/cjfas-2018-0255.

West, M., and Harrison, J. 1997. Bayesian Forecasting and Dynamic Models. *In* Springer Series in Statistics. doi:10.1016/S1353-8020(02)00127-X.

Wiedenmann, J., Free, C.M., and Jensen, O.P. 2019. Evaluating the performance of data-limited methods for setting catch targets through application to data-rich stocks: A case study using Northeast U.S. fish stocks. Fish. Res. **209**: 129–142. doi:10.1016/j.fishres.2018.09.018.

Wiedenmann, J., and Jensen, O.P. 2018. Uncertainty in stock assessment estimates for New England groundfish and its impact on achieving target harvest rates. Can. J. Fish. Aquat. Sci. **75**(3): 342–356. doi:10.1139/cjfas-2016-0484.

Xu, H., Miller, T.J., Hameed, S., Alade, L.A., and Nye, J.A. 2018. Evaluating the utility of the Gulf Stream Index for predicting recruitment of Southern New England-Mid Atlantic yellowtail flounder. Fish. Oceanogr. **27**(1): 85–95. doi:10.1111/fog.12236.

TABLES

Table 1. Summary of the characteristics of the five simulated stock scenarios, including the time series range and the age classes evaluated for each population. CV: coefficient of variation.

Scenario	Years	Age Range	Natural Mortality Rate (M)	Fishing Mortality Rate (F)	Recent Recruitment Trend	Survey Selectivity	Survey CVs
GB Atlantic cod	1978-2012	1-10+	0.20	0.23-1.10	Stable	Similar to fishery	0.22-0.30
GB yellow-tail flounder	1973-2014	1-6+	0.20-0.40	0.32-1.94	Declining	Similar to fishery	0.30-0.45
Witch flounder	1982-2016	1-11+	0.15	0.23-1.10	Stable	One survey targets juveniles, all surveys have greater selectivity for smaller sizes than fishery	0.27-0.35
GB yellow-tail flounder-SR	1973-2014	1-6+	0.20-0.40	0.32-1.94	Stable	Similar to fishery	0.30-0.45
GB yellow-tail flounder- HME	1973-2014	1-6+	0.20-0.40	0.32-1.94	Declining	Similar to fishery	0.90-1.35

Table 2. Summary of the predictive performance of each model class (IB: Index-Based, AB: Age-Based) in the five tested simulation scenarios. Where three values are shown, they represent the metrics by year of the forecast period (Year 1, Year 2, Year 3). The change in the true relative (survey scale) abundance relative to the terminal year used in model fitting is provided for each year of the forecast period. The median root mean squared error ratio (RMSE) is provided both for the training data (T) and forecast interval (F). SR: Stable Recruitment scenario, HME: High Measurement Error scenario.

Scenario [Model Class]	True Change (%)	Mean Prediction Error (%)	Median Absolute Prediction Error (%)	Forecast Interval Coverage (%)	Median RMSE Ratio (T, F)
GB Atlantic cod [IB]	32.3, 74.4, 120.2	-17.7, -29.2, -38.3	20.5, 31.3, 40.2	96.7	(0.55, 1.65)
GB Atlantic cod [AB]	32.3, 74.4, 120.2	2.5, 7.9, 22.4	13.8, 21.9, 29.0	94.7	(0.51, 1.13)
GB yellowtail flounder [IB]	-14.1 -19.4, -40.1	18.6, 52.7, 72.8	24.6, 44.4, 57.7	99.8	(0.63, 1.38)
GB yellowtail flounder [AB]	-14.1 -19.4, -40.1	-10.6, -6.2, -7.6	14.9, 19.8, 23.2	90.7	(0.52, 0.68)
Witch flounder [IB]	0.0, 9.8, 27.1	0.1, -6.1, -13.8	15.6, 20.4, 25.0	94.7	(0.48, 0.92)
Witch flounder [AB]	0.0, 9.8, 27.1	11.1, 13.9, 19.5	16.5, 20.1, 23.6	78.7	(0.51, 0.77)
GB yellowtail flounder-SR [IB]	21.7, 45.3, 32.0	-0.6, 31.0, 70.0	18.6, 25.1, 49.7	99.3	(0.62, 1.58)
GB yellowtail flounder-SR [AB]	21.7, 45.3, 32.0	11.4, 36.5, 60.6	14.7, 29.1, 46.1	86.2	(0.53, 1.33)
GB yellowtail flounder-HME [IB]	-14.1, -19.4, -40.1	34.6, 95.3, 115.3	33.2, 62.9, 61.1	98.7	(0.35, 0.81)
GB yellowtail flounder-HME [AB]	-14.1, -19.4, -40.1	66.6, 227.7, 698.3	47.2, 164.5, 507.4	91.3	(0.47, 3.20)

FIGURES

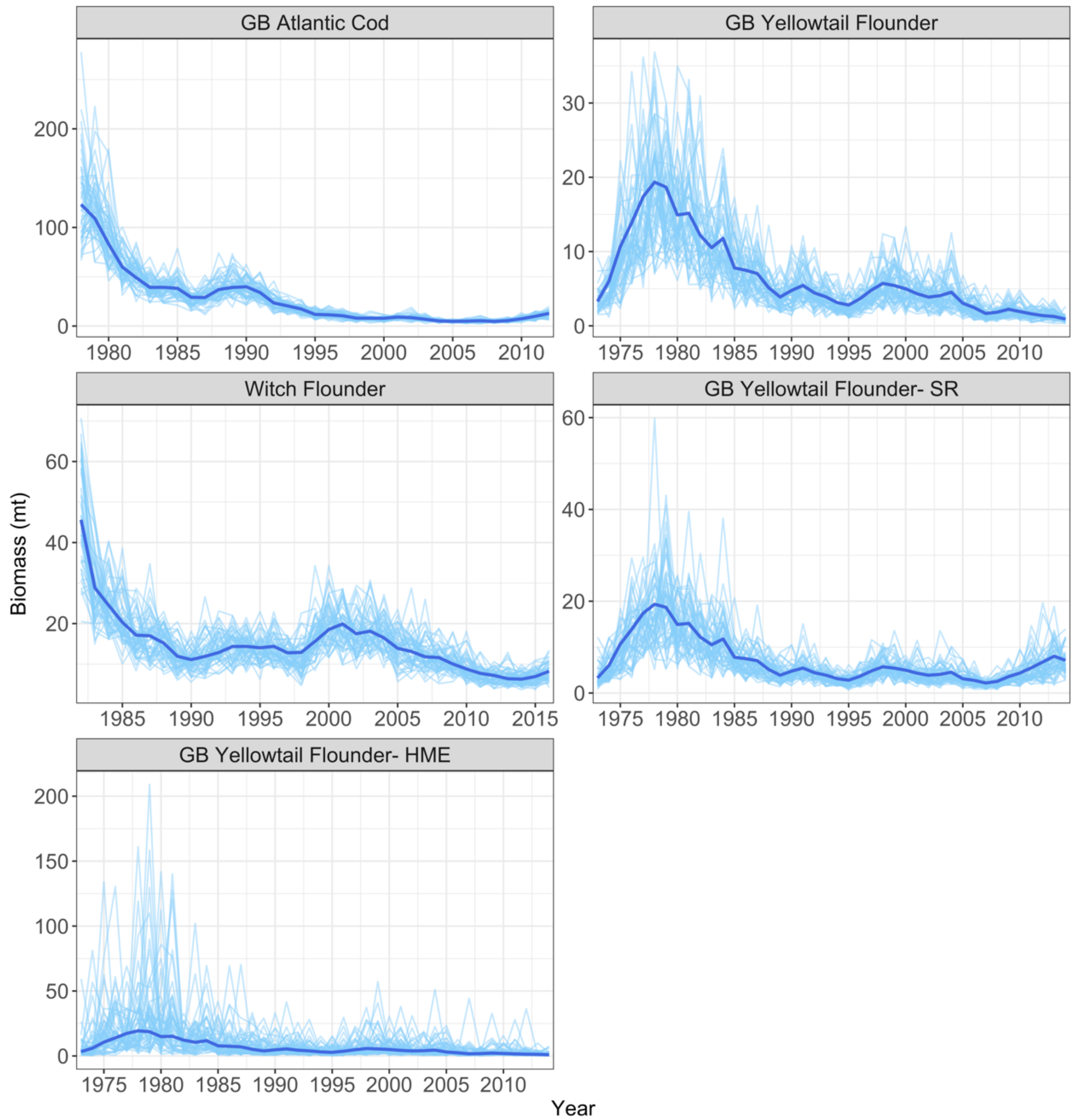


Figure 1. The true relative (survey scale) abundance (dark blue) and 50 simulated survey time series (light blue) for a single survey for each of the five stock simulation scenarios. SR: Stable Recruitment scenario, HME: High Measurement Error scenario.

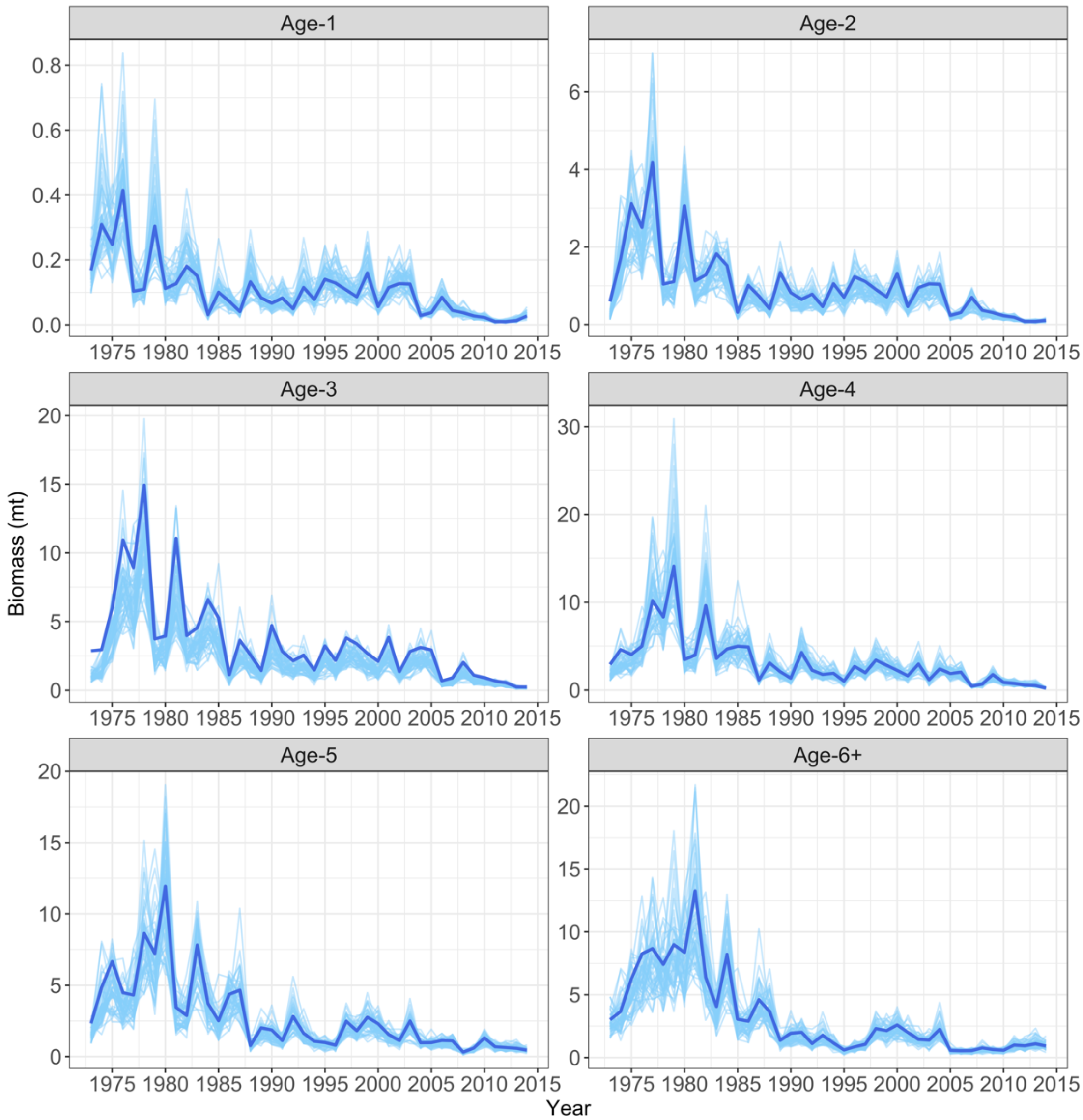


Figure 2. The true relative (survey scale) abundance (dark blue) and the 50 simulated survey time series (light blue) for the modeled age classes (1-6+) in the base yellowtail flounder scenario.

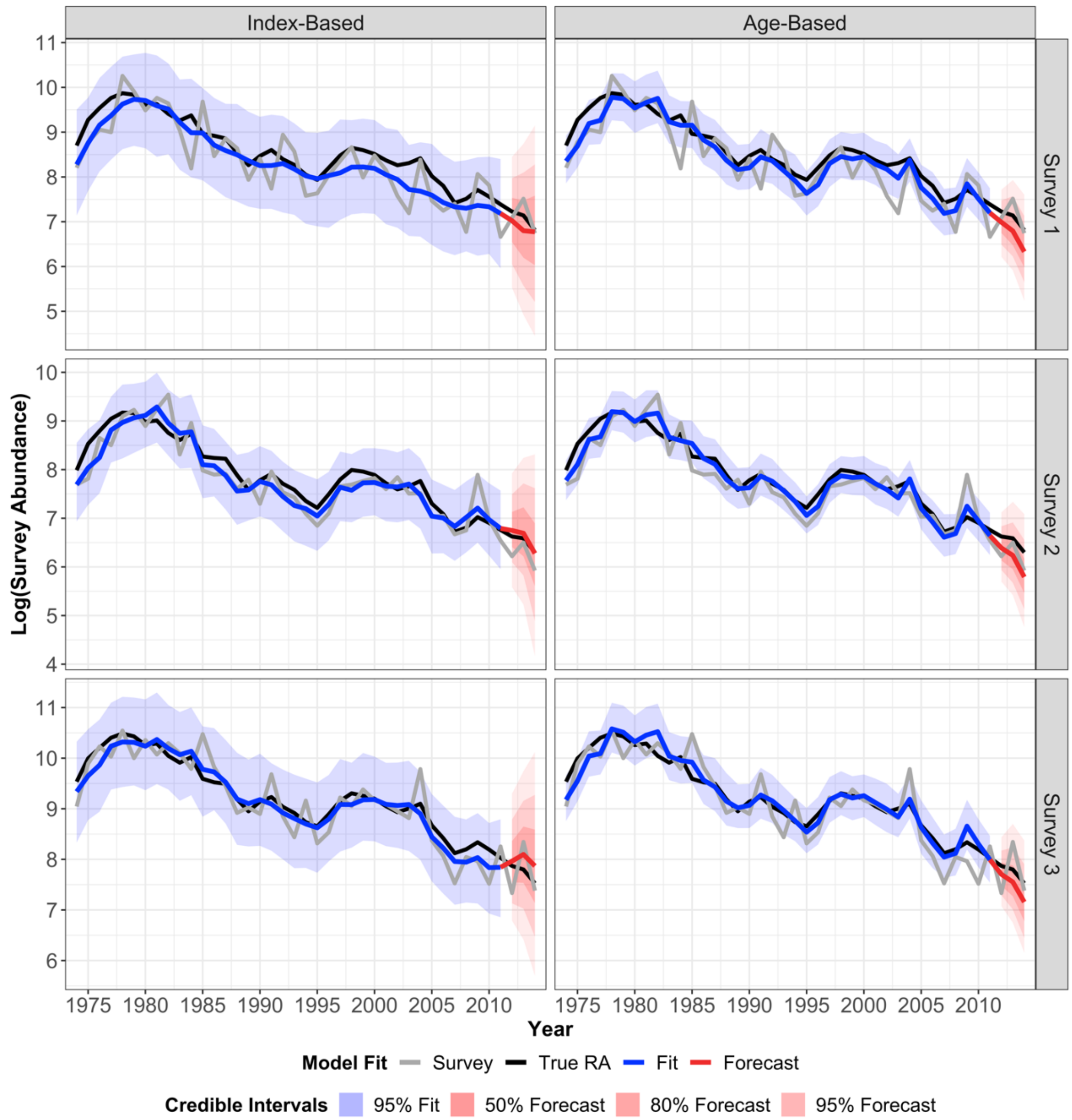


Figure 3. The Index-Based (left) and Age-Based (right) DLM fits for three surveys (rows) in a single simulation of the base yellowtail flounder scenario. The true relative (survey scale) abundance (RA, black) and observed survey abundance used in fitting (gray) is compared to the median model fit (blue) and forecast (red). The credible intervals for the posterior predictive distributions estimated for each year are represented by shading.

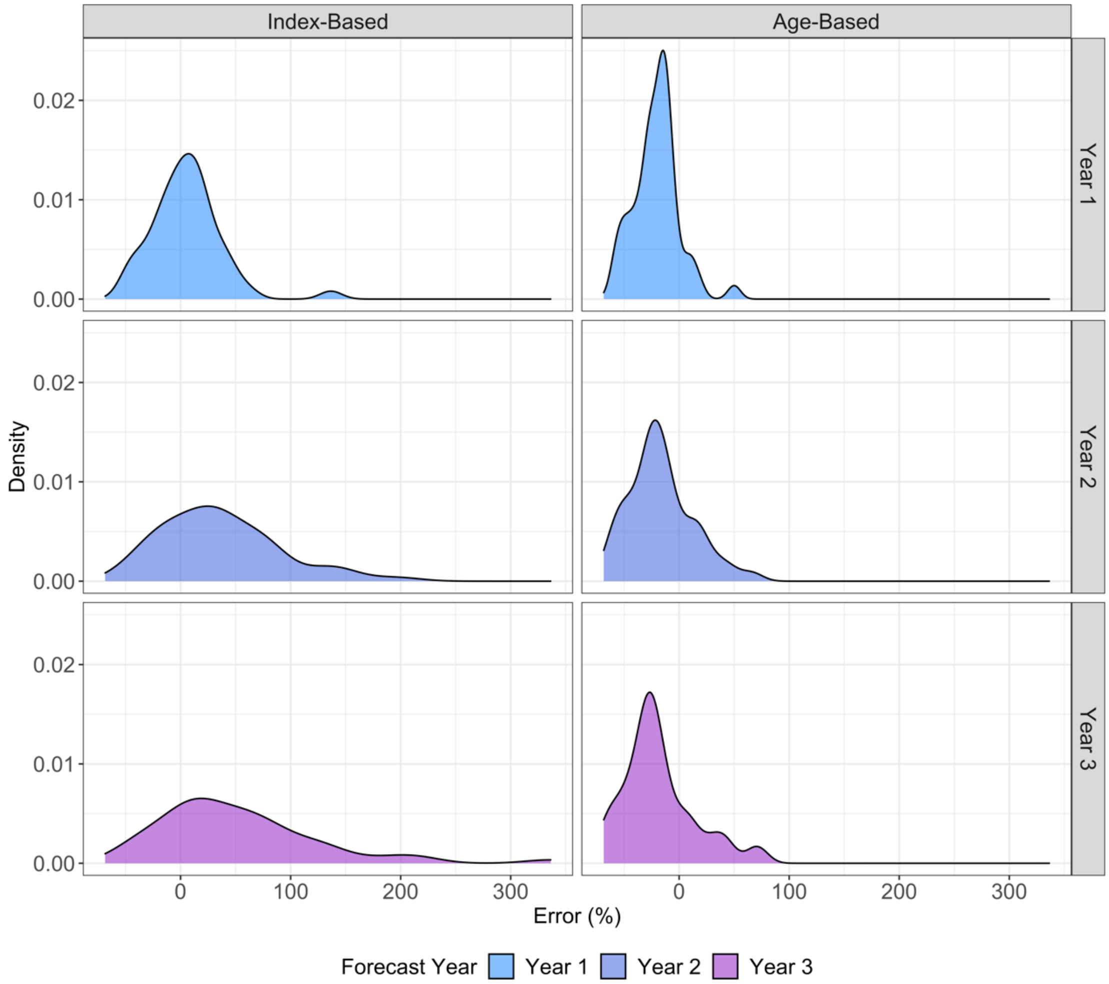


Figure 4. The distributions of survey-averaged prediction errors by year (rows) of the forecast period for the Index-Based (left) and Age-Based (right) DLM fits of the base yellowtail flounder scenario.

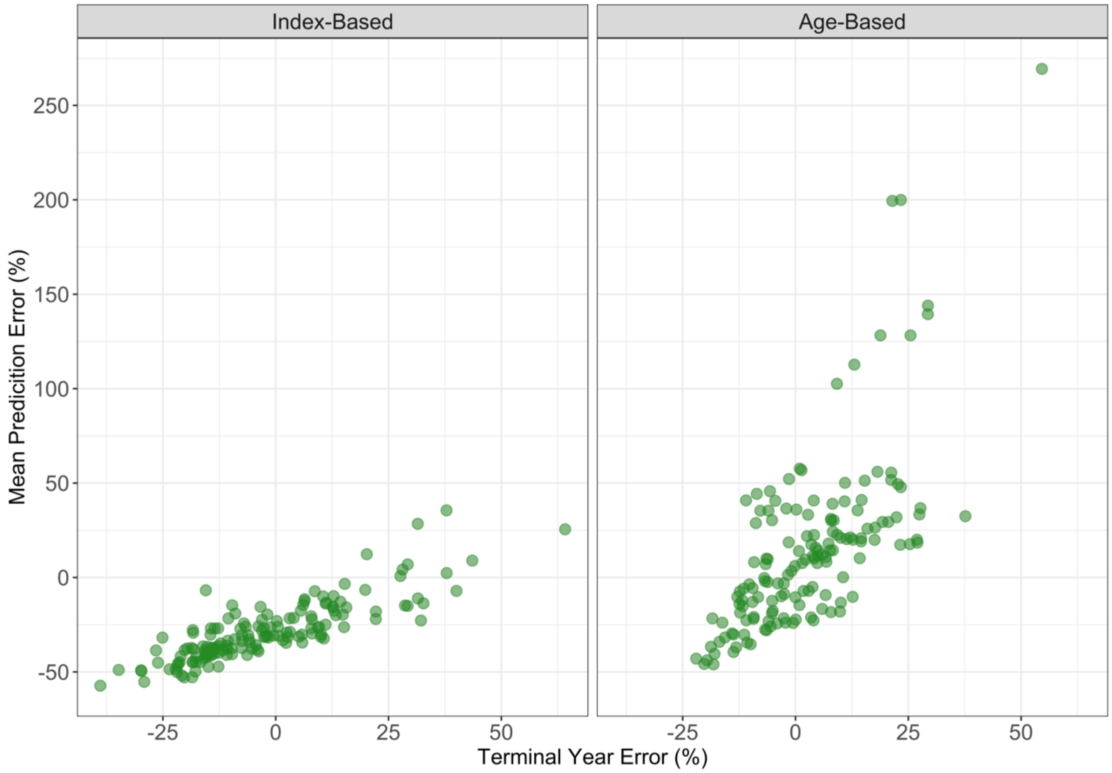


Figure 5. The relationship between the estimation error, based on the median of the posterior predictive distribution, in the terminal year of the training data and the average error observed during the forecast period for each survey for the Index-Based (left) and Age-Based (right) DLM fits to each of the 50 simulations of the base Georges Bank Atlantic cod scenario simulations.

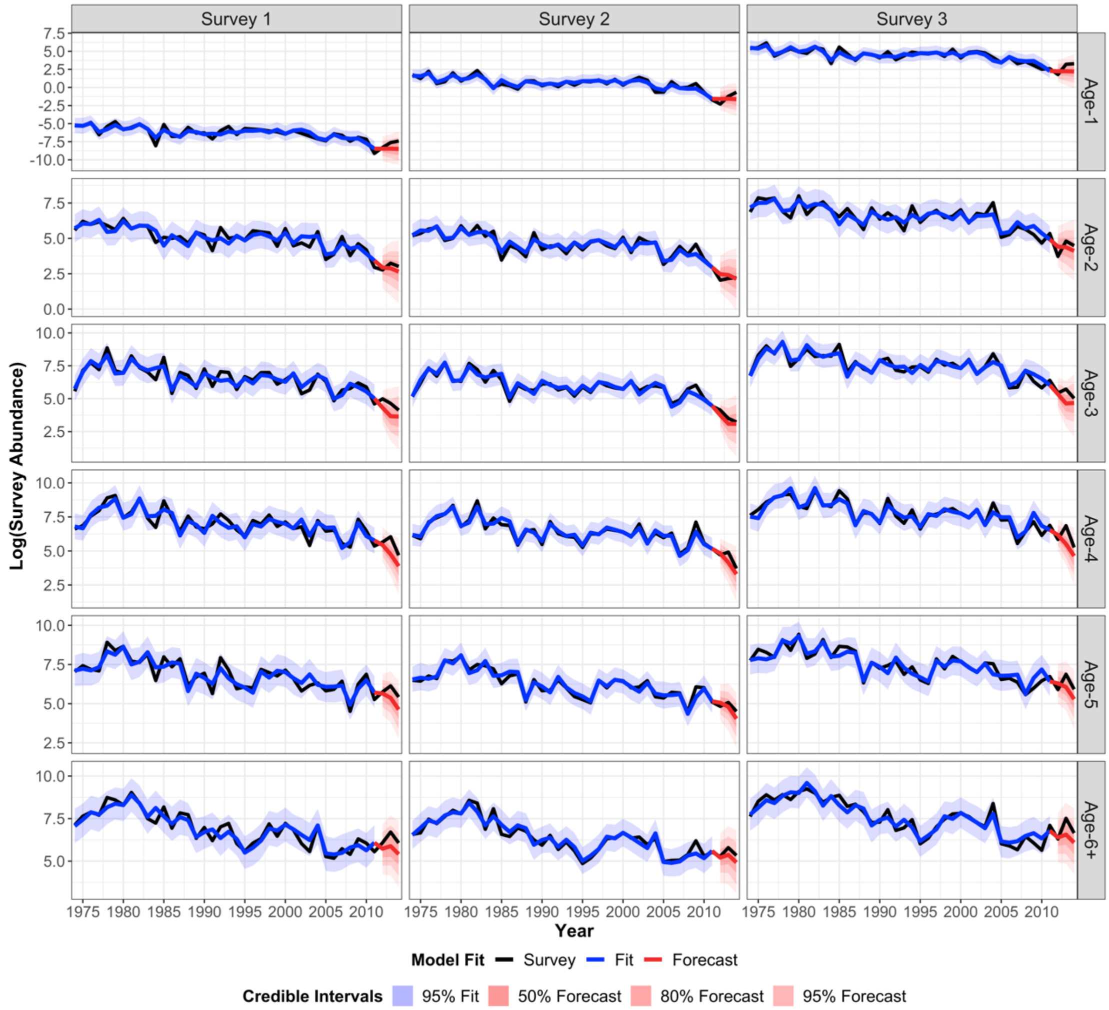
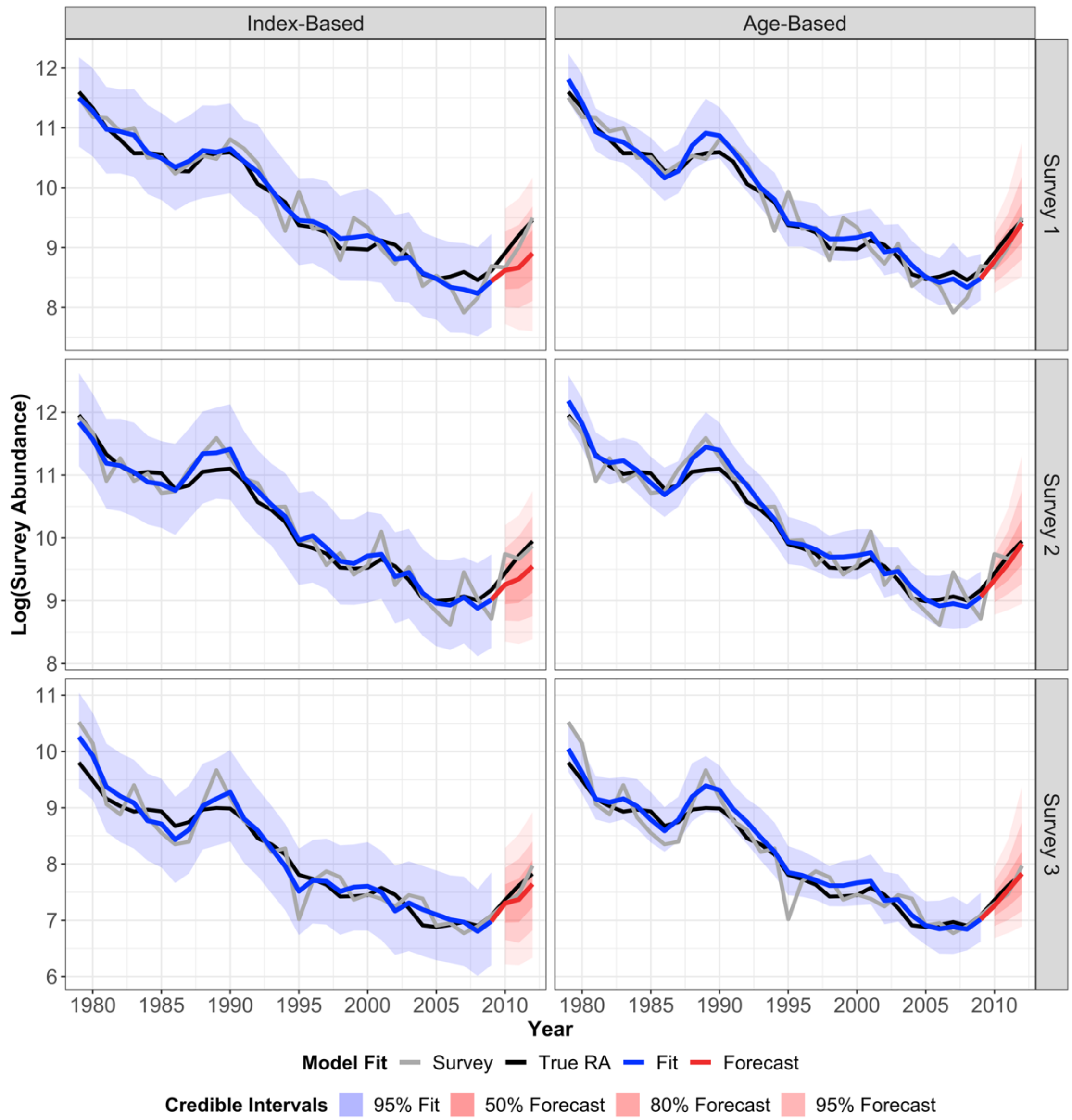


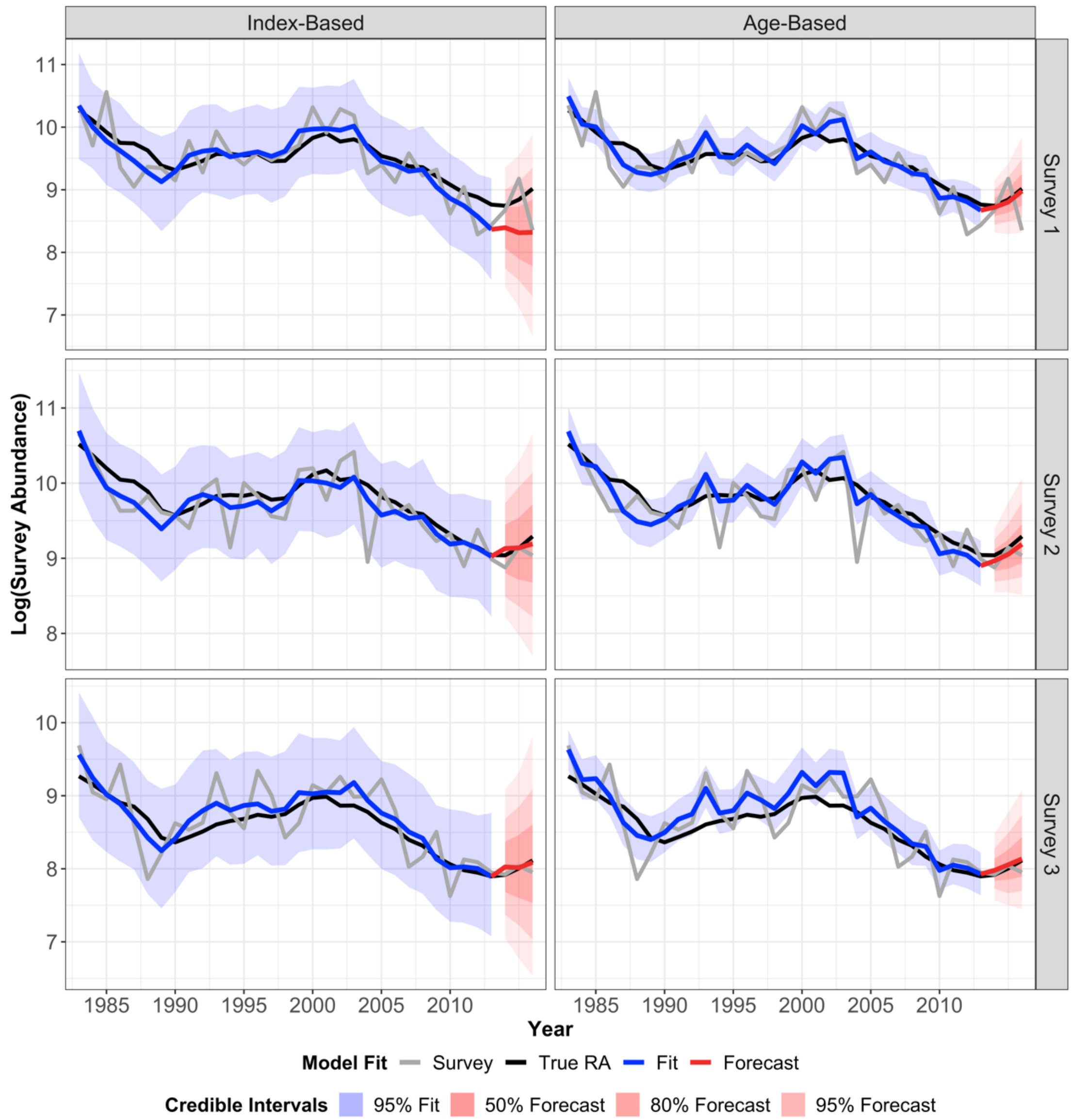
Figure 6. The Age-Based DLM fit for all surveys (columns) and age classes (rows) of a single simulation of the base yellowtail flounder scenario. The simulated survey observations used in model fitting are depicted by a black line and the median model fit and forecast are depicted by blue and red lines, respectively. The credible intervals for the model fit and forecast are shown as shaded regions.

SUPPLEMENTS

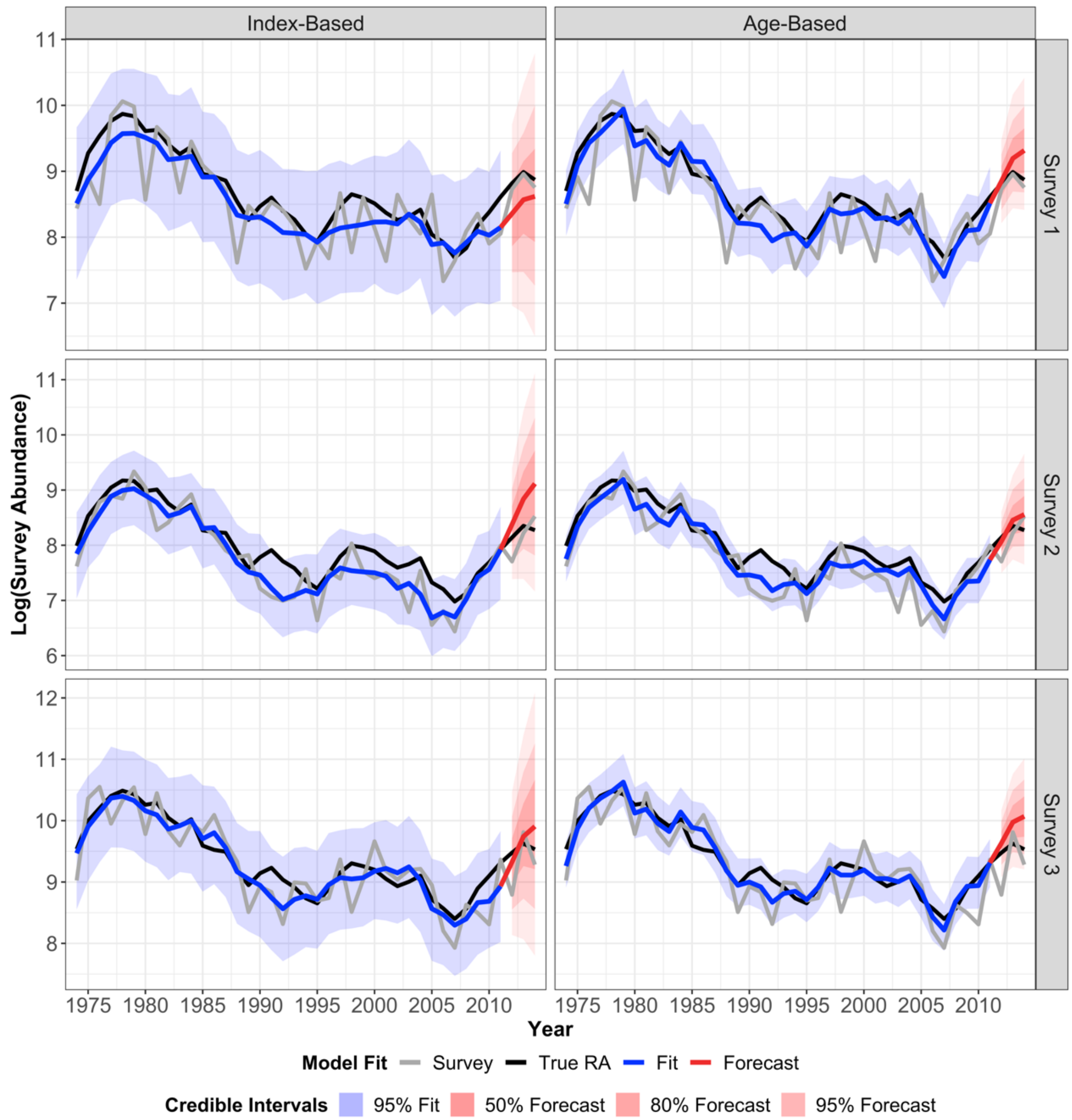
Figure S1. The Index-Based (left) and Age-Based (right) DLM fits for three surveys (rows) in a single simulation of the simulation scenarios not shown in Figure 3. The true relative (survey scale) abundance (RA, black) and observed survey abundance used in fitting (gray) is compared to the median model fit (blue) and forecast (red). The credible intervals for the posterior predictive distributions estimated for each year are represented by shading.



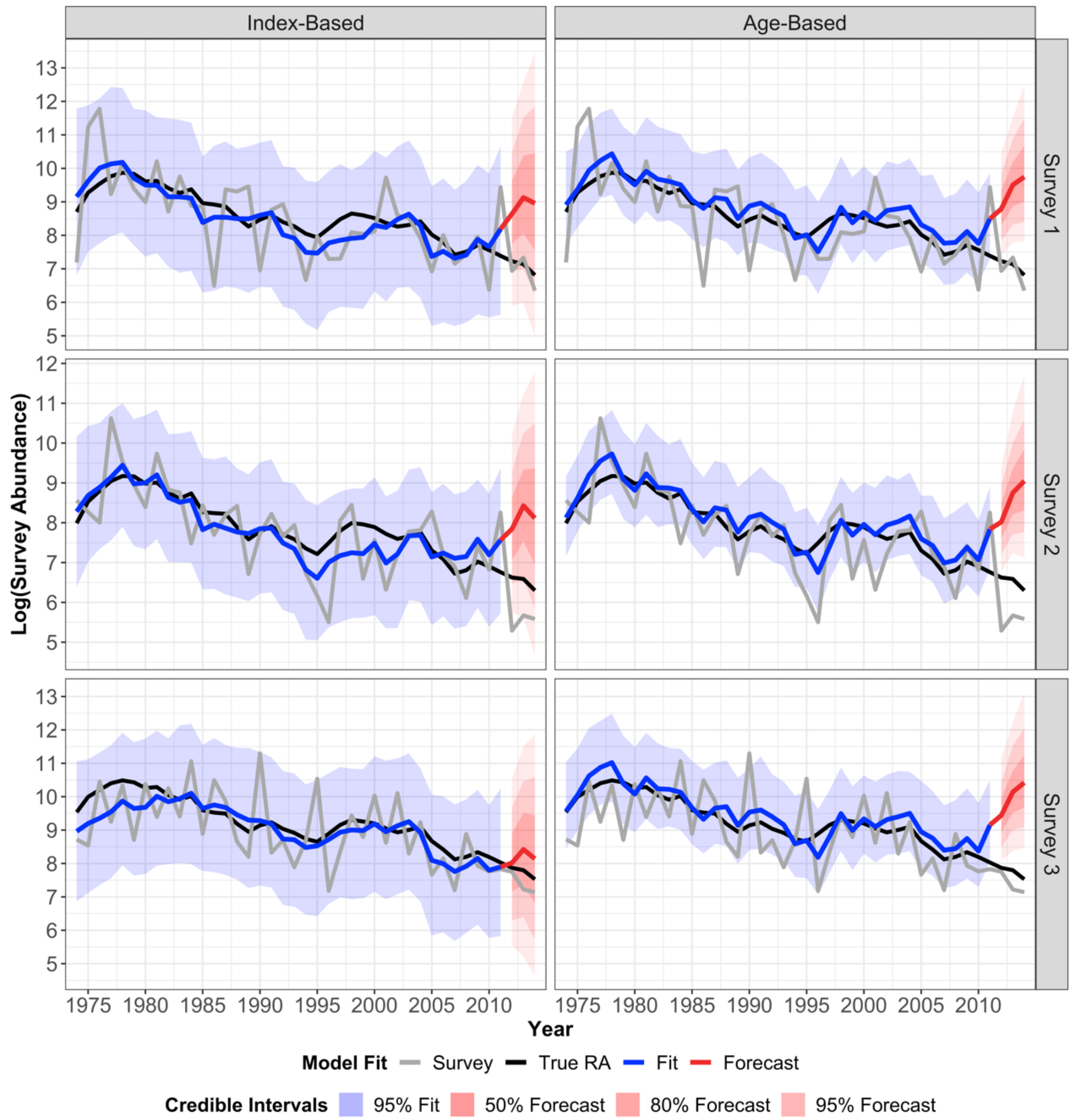
a. Georges Bank Atlantic cod base scenario



b. Witch flounder base scenario



c. Georges Bank yellowtail flounder- Stable Recruitment scenario



d. Georges Bank yellowtail flounder- High Measurement Error scenario

Appendix I: Forward-Filtering Backward-Sampling (FFBS) Algorithm

The set of state variables $\boldsymbol{\theta}_t$ to be estimated in a DLM fit to observed data y_t for time points t in $1{:}T$ are sampled using the following steps. For a given time t, the FFBS algorithm first calculates the one-step-ahead predictive distribution of the latent state, $f(\boldsymbol{\theta}_t|y_{1:t-1}) = N(\boldsymbol{a}_t, \boldsymbol{R}_t)$, where:

$$\boldsymbol{a}_t = E(\boldsymbol{\theta}_t|y_{1:t-1}) = \boldsymbol{G}_t\boldsymbol{m}_{t-1}$$

$$\boldsymbol{R}_t = Var(\boldsymbol{\theta}_t|y_{1:t-1}) = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t' + \boldsymbol{W}_t$$

The one-step-ahead predictive distribution of the observation, $f(y_t|y_{1:t-1}) = N(f_t, Q_t)$, is then defined as:

$$f_t = E(y_t|y_{1:t-1}) = \boldsymbol{F}_t\boldsymbol{a}_t$$

$$Q_t = Var(y_t|y_{1:t-1}) = \boldsymbol{F}_t\boldsymbol{R}_t\boldsymbol{F}_t' + V_t$$

The forward filtered distribution of the latent state, $f(\boldsymbol{\theta}_t|y_{1:t}) = N(\boldsymbol{m}_t, \boldsymbol{C}_t)$, has moments:

$$\boldsymbol{m}_t = E(\boldsymbol{\theta}_t|y_{1:t}) = \boldsymbol{a}_t + \boldsymbol{A}_t e_t$$

$$\boldsymbol{C}_t = Var(\boldsymbol{\theta}_t|y_{1:t}) = \boldsymbol{R}_t - \boldsymbol{A}_t\boldsymbol{A}_t' Q_t$$

$$e_t = y_t - f_t$$

$$\boldsymbol{A}_t = \boldsymbol{R}_t\boldsymbol{F}_t'/Q_t$$

Finally, the smoothed (backward sampled) distribution of the latent state, $f(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, y_{1:T}) = N(\boldsymbol{m}_t^*, \boldsymbol{C}_t^*)$, has moments:

$$\boldsymbol{m}_t^* = E(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, y_{1:T}) = \boldsymbol{m}_t + \boldsymbol{B}_t(\boldsymbol{\theta}_{t+1} - \boldsymbol{a}_{t+1})$$

$$\boldsymbol{C}_t^* = Var(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, y_{1:T}) = \boldsymbol{C}_t - \boldsymbol{B}_t\boldsymbol{R}_{t+1}\boldsymbol{B}_t'$$

$$\boldsymbol{B}_t = \boldsymbol{C}_t\boldsymbol{G}_{t+1}'\boldsymbol{R}_{t+1}^{-1}$$

The moments $\boldsymbol{m}_t^*$ and $\boldsymbol{C}_t^*$ are then recorded for all time points and returned to the Gibbs sampler for use in sampling the observation error variance V_t and the evolution error variance $\boldsymbol{W}_t$ and to make forecasts of future values of the observed data y_t.

Appendix II. DLM Specifications

Common Prior Distributions

A similar strategy was employed in setting priors for both the Index-Based and Age-Based DLMs. The state variables defining the trend component of the models were assigned flat gaussian priors (variance = 1 x 10^7). The means of these priors were set to the mean of the observed data for intercept terms and zero for the slope terms of dynamic trends. The priors on the regression coefficients on catch anomalies, which were expected to be between 0 and -1, were in all cases assigned the gaussian prior $N(-0.5, 0.0625)$.

The prior distributions on the observation and evolution error variances were informed by the data. Specifically, it was assumed a priori that half of the total data variance could be explained by observation error and the other half by evolution error. Therefore, the observation error variances for each survey were assigned the inverse-gamma prior $IG\left(1, \frac{Var(Y)}{2}\right)$. The evolution error variances for the dynamic state variables were assigned inverse-Wishart priors, where it was assumed that the state variables explained equal proportions of the total evolution variance. The covariances among the dynamic state variables were set to one half that variances a priori. The prior degrees of freedom was set to the number of dynamic state variables plus two. Thus, in an example model with two dynamic state variables, the parameters of the inverse-Wishart prior distribution would be:

$$\boldsymbol{a}_w = \begin{bmatrix} \frac{Var(Y)}{4} & \frac{Var(Y)}{8} \\ \frac{Var(Y)}{8} & \frac{Var(Y)}{4} \end{bmatrix}^{-1}, \quad b_w = 4$$

Index-Based DLMs

Georges Bank Atlantic cod

Trend component: random walk

Catch regression: static

$$\boldsymbol{F}_t = \begin{bmatrix} 1 & 0 & 0 & H^*_{1,t-1} & 0 & 0 \\ 0 & 1 & 0 & 0 & H^*_{2,t-1} & 0 \\ 0 & 0 & 1 & 0 & 0 & H^*_{3,t-1} \end{bmatrix} \qquad \boldsymbol{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{V} = \begin{bmatrix} V_1 & 0 & 0 \\ 0 & V_2 & 0 \\ 0 & 0 & V_3 \end{bmatrix} \qquad \boldsymbol{W} = \begin{bmatrix} W_{\tau_1} & W_{\tau_1\tau_2} & W_{\tau_1\tau_3} & 0 & 0 & 0 \\ W_{\tau_2\tau_1} & W_{\tau_2} & W_{\tau_2\tau_3} & 0 & 0 & 0 \\ W_{\tau_3\tau_1} & W_{\tau_3\tau_2} & W_{\tau_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{\theta}'_t = [\tau_{1,t} \quad \tau_{2,t} \quad \tau_{3,t} \quad \beta_{1,t} \quad \beta_{2,t} \quad \beta_{3,t}]$$

Where $H^*_{S,t-1}$ denotes the catch anomalies calculated for survey S lagged by one year, V_S is the measurement error variance for survey S, W_{τ_S} is the evolution error variance for a random walk (τ) fit to survey S, $W_{\tau_{S_1}\tau_{S_2}}$ is the covariance between random walks fit to two surveys, $\boldsymbol{\theta}_t$ is the vector of state variables at time t, and $\beta_{S,t}$ is the regression coefficient for the catch anomalies calculated using survey S.

Georges Bank yellowtail flounder, witch flounder, Georges Bank yellowtail flounder-Stable Recruitment, Georges Bank yellowtail flounder- High Measurement Error

Trend component: dynamic linear trend

Catch regression: static

$$\boldsymbol{F}_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & H^*_{1,t-1} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & H^*_{2,t-1} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & H^*_{3,t-1} \end{bmatrix}$$

$$\boldsymbol{G} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{V} = \begin{bmatrix} V_1 & 0 & 0 \\ 0 & V_2 & 0 \\ 0 & 0 & V_3 \end{bmatrix} \qquad \boldsymbol{W} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & W_{\tau_1} & W_{\tau_1\tau_2} & W_{\tau_1\tau_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & W_{\tau_2\tau_1} & W_{\tau_2} & W_{\tau_2\tau_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & W_{\tau_3\tau_1} & W_{\tau_3\tau_2} & W_{\tau_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{\theta}'_t = [I_{1,t} \quad I_{2,t} \quad I_{2,t} \quad \tau_{1,t} \quad \tau_{2,t} \quad \tau_{3,t} \quad \beta_1 \quad \beta_2 \quad \beta_3]$$

Where $H^{*}_{S,t-1}$ denotes the catch anomalies calculated for survey S lagged by one year, V_S is the measurement error variance for survey S, W_{τ_S} is the evolution error variance for the slope term of a dynamic linear trend (τ) fit to survey S, $W_{\tau_{S_1}\tau_{S_2}}$ is the covariance between the dynamic linear trend slopes fit to two surveys, θ_t is the vector of state variables at time t, $I_{S,t}$ is the intercept term of the dynamic linear trend for survey S at time t, and β_S is the static regression coefficient for the catch anomalies calculated using survey S.

Age-Based DLMs

All scenarios

Trend component: random walk (age-1), first order autoregressive process (age-2+)

Catch regressions: static

Regressions on prior age: static

State variables: common across surveys

Priors

The regression coefficients on the prior age class in the Age-Based DLMs were expected to be positive, but generally not much larger than 1. Therefore, these terms were all assigned the gaussian prior $N(1, 0.25)$. Flat gaussian priors, $N(0,1 \times 10^{7})$, were used for the intercept terms accounting for differences in catchability and selectivity among surveys

Age-1

$$\boldsymbol{F}_t = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \qquad \boldsymbol{G} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{V} = \begin{bmatrix} V_1 & 0 & 0 \\ 0 & V_2 & 0 \\ 0 & 0 & V_3 \end{bmatrix} \qquad \boldsymbol{W} = \begin{bmatrix} W_\tau & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{\theta}_t' = [\tau_t \quad \mu_{12} \quad \mu_{13}]$$

Where V_S is the measurement error variance for survey S and W_τ is the evolution error variance for the common random walk (τ) fit to all surveys, and $\mu_{S_1S_2}$ is the constant relative difference in catchability and selectivity between surveys 1 and 2.

Age-2 or older

$$\boldsymbol{F}_t = \begin{bmatrix} B_{a-1} & H_{t-1}^* & 1 & 0 & 0 \\ B_{a-1} & H_{t-1}^* & 1 & 1 & 0 \\ B_{a-1} & H_{t-1}^* & 1 & 0 & 1 \end{bmatrix} \qquad \boldsymbol{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \phi & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{V} = \begin{bmatrix} V_1 & 0 & 0 \\ 0 & V_2 & 0 \\ 0 & 0 & V_3 \end{bmatrix} \qquad \boldsymbol{W} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & W_\tau & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{\theta}'_t = [\gamma_a \quad \beta_a \quad \tau_{a,t} \quad \mu_{12,a} \quad \mu_{13,a}]$$

Where V_S is the measurement error variance for survey S, W_τ is the evolution error variance for the common first order autoregressive process (τ) for age a with coefficient ϕ fit to all surveys, B_{a-1} is the estimated relative abundance of the prior age class, H^*_{t-1} is the mean catch anomaly, calculated using the age-aggregated survey and catch data, across surveys lagged by one year, γ_a is the static regression coefficient for age a on the relative abundance of the prior age class, β_a is the static regression coefficient on the lagged catch anomalies, and $\mu_{S_1S_2}$ is the constant relative difference in catchability and selectivity between surveys 1 and 2.

Appendix III. DLM test with real world Georges Bank yellowtail flounder data

To perform a preliminary comparison of the retrospective forecasting performance of the Index-Based and Age-Based DLM approaches between the simulated data sets and real-world observations, survey data were gathered for the Georges Bank yellowtail flounder stock. Specifically, age-specific indices of abundance (mean number/tow) were obtained for the Department of Fisheries and Oceans Canada (DFO) trawl survey (1987-2019) and the Spring and Fall bottom trawl surveys conducted by the Northeast Fisheries Science Center of the National Marine Fisheries Service (NMFS, 1973-2019). The three survey time series were then fit with an Index-Based and an Age-Based DLM as in the main text and using model specifications detailed in Appendix II. The Index-Based model used the same model structure as for the Georges Bank Atlantic cod scenario. Because the DFO survey did not begin until 1987, missing values of the catch anomalies for 1973-1986 were estimated and imputed using a linear regression of the observed DFO catch anomalies on the anomalies calculated for the two NMFS surveys. The DLMs were then fit to subsets of the survey time series such that the generated forecasts could be compared to true observations recorded after the training data period.

Comparison of the retrospective forecasts to the true observations produced mixed results that were broadly similar to the performance observed when using simulated data. When the models were fit to the 1973-2016 survey data, the Index-Based DLM produced forecasts much closer to the observed data for 2017-2019 than the Age-Based approach (Figure A3.1). However, models fit to different subsets of the data (minimum: 1973-2005) suggested that the Index-Based DLM was not necessarily exhibiting better performance (Figure A3.2). Prior to the steep drop in abundance beginning in the 2010s, the forecasts generated by the Age-Based approach were generally closer to the future observed data. The Age-Based approach also provided a more consistent imputation of the missing (pre-1987) values of the DFO survey as additional data were added to model fitting. During the abundance decline, however, the Age-Based DLM predicted population growth that did not occur based upon low levels of harvest. As a result, the forecasts produced by the Index-Based DLM during the 2010s were closer to the real data. Taken together, these results suggest that retrospective forecasting using different subsets of available observed data can provide useful insights into prospective model performance. However, a change in the underlying population dynamics may alter the performance of a DLM structure that appeared well-suited to modeling a target stock based upon historic data. The poor forecast performance observed during the 2010s indicates that additional Age-Based DLM structures would need to be explored for use on the Georges Bank yellowtail flounder stock.

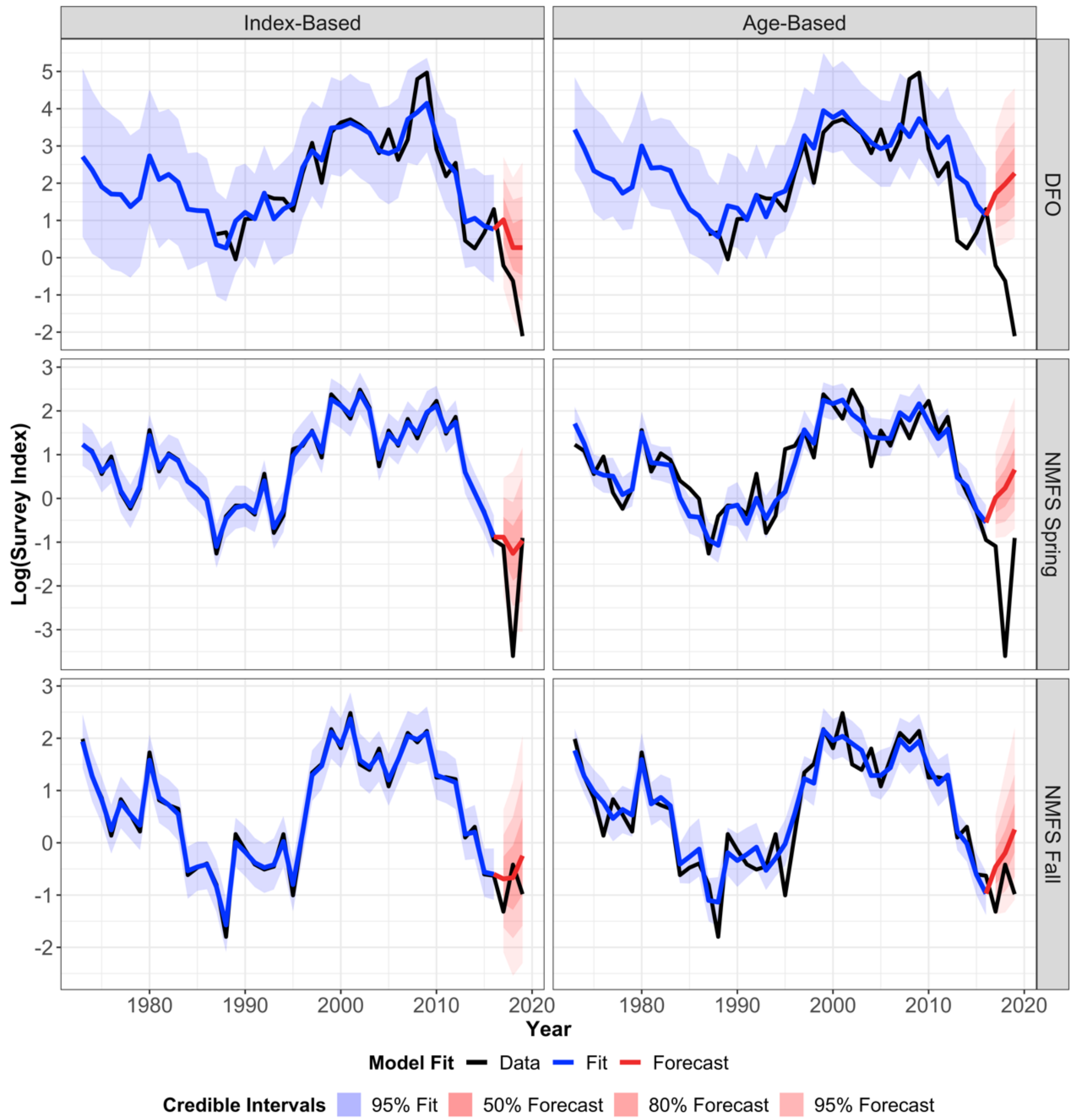


Figure A3.1. The Index-Based (left) and Age-Based (right) DLM fits of three surveys (rows) used to monitor the Georges Bank yellowtail flounder stock. The observed survey abundance used in fitting (black) is compared to the median model fit (blue) and forecast (red). The credible intervals for the posterior predictive distributions estimated for each year are represented by shading.

Figure A3.2. The Index-Based (left) and Age-Based (right) DLM fits of sequentially longer time series (minimum: 1973-2005, each year from 2005-2019 was used as an endpoint) of three surveys (rows) used to monitor the Georges Bank yellowtail flounder stock. The observed survey abundance used in fitting (black) is compared to the median model fits and three year forecasts (colors).

CONCLUSIONS

The results of this thesis support further exploration of the use of dynamic linear models (DLMs) in the assessment of fisheries for which age-structured stock assessment approaches are not possible. While the Index-Based Methods (IBMs) currently used to manage stocks in such situations do not make forecasts of future abundance or quantify scientific uncertainty (Wiedenmann et al. 2019; Legault et al. 2021), the developed DLM models exhibited strong predictive performance across the simulated scenarios. Similar to previous research into the performance of different stock assessment approaches (Brooks and Legault 2016; Wiedenmann and Jensen 2018), prediction errors in this work were strongly correlated with the estimation error magnitude in the terminal year of the observed data and tended to grow with the length of the forecast. Prediction error magnitude also appeared to increase with the observation error variance of the fitted data. Because large terminal estimation errors appeared to manifest in cases where consecutive observation errors of the same sign were observed at the end of the available time series, these results suggest that the DLM approach may perform best if the forecasts are updated frequently (annually) before prediction errors compound.

In general, the Age-Based DLM appeared to perform better than the Index-Based approach. This is not unexpected given that additional demographic information is being supplied to the model. However, the results of this thesis do not suggest that the Age-Based approach will be more desirable in every situation. For example, the Age-Based DLM produced smaller mean and median absolute prediction errors than the

Index-Based case in the Georges Bank Atlantic cod scenario, but also exhibited a larger range of errors. Apparently due to the compounding of forecast errors for individual age classes, this suggests that the preferred DLM structure in this case may depend on the acceptable risk level among fishery managers. Additionally, the High Measurement Error scenario for the Georges Bank yellowtail flounder simulation suggested that the simpler, smoother Index-Based DLM would perform better than a more complex model when observation error variance was very high. Taken together, these examples highlight the need for additional research to characterize the strengths and weaknesses of different levels of DLM complexity when applied to a variety of fisheries management situations.

Although development of specific strategies was beyond the scope of this thesis, it will be important to evaluate different methods of using the forecast produced by DLM models in generating catch advice for the management of specific fisheries. In the case that catch information is not included in a particular DLM, one could employ a similar approach for recommending harvest levels as is conducted in the PlanBSmooth IBM (Northeast Fisheries Science Center 2015). When catch information is included, the estimated DLM equations could be used to solve for the catch levels that could produce a target abundance with a desired level of uncertainty. Like other IBMs currently employed (Wiedenmann et al. 2019; Legault et al. 2021), however, the DLMs described in this work do not estimate the true population size. As a consequence of this characteristic, fisheries managers will face ambiguity in determining biomass targets that represent a healthy state for a given stock. In the absence of the ability to calculate reference levels relative to, for example, the population size at maximum sustainable yield, the development of catch advice from employed DLMs will rely upon the knowledge of experts in assessing fish stocks. However, the ability to produce true forecasts of future abundance and fulfill the well-recognized need to account for scientific uncertainty in fisheries management (Berkson et al. 2011; Maunder and Piner 2015; Kokkalis et al. 2017) gives users of the DLM framework more information with which to develop catch advice than in currently employed IBMs (Wiedenmann et al. 2019; Legault et al. 2021).

Finally, the DLM structures developed and evaluated here are not meant to be exhaustive, but rather a first exploration into the capabilities and utility of this approach in fisheries management. We strongly encourage additional testing and further development of DLMs in order to maximize their potential to improve fisheries management outcomes in the many fisheries for which a conventional stock assessment approach is not possible (Costello et al. 2012). To this end, we hope that this thesis provides scientists and fisheries managers, present and future, with inspiration and a jumping off point to continue to strive toward a more productive and sustainable future for global fisheries.

REFERENCES

Berkson, J., Barbieri, L., Cadrin, S.X., Cass-Calay, S., Crone, P., Dorn, M., Friess, C., Kobayashi, D., Miller, T.J., Patrick, W.S., Pautzke, S., Ralston, S., and Trianni, M. 2011. Calculating Acceptable Biological Catch for Stocks That Have Reliable Catch Data Only (Only Reliable Catch Stocks – ORCS).

Brooks, E.N., and Legault, C.M. 2016. Retrospective forecasting — evaluating performance of stock projections for New England groundfish stocks. Can. J. Fish. Aquat. Sci. **73**(6): 935–950. doi:10.1139/cjfas-2015-0163.

Costello, C., Ovando, D., Hilborn, R., Gaines, S.D., Deschenes, O., and Lester, S.E. 2012. Status and Solutions for the World's Unassessed Fisheries. Science (80-.).

338(6106): 517–520. doi:10.1126/science.1223389.

Kokkalis, A., Eikeset, A.M., Thygesen, U.H., Steingrund, P., and Andersen, K.H. 2017. Estimating uncertainty of data limited stock assessments. ICES J. Mar. Sci. **74**(1): 69–77. doi:10.1093/icesjms/fsw145.

Legault, C.M., Bell, R.J., Cournane, J., Deroba, J.J., Fay, G., Jones, A., Miller, T.J., Muffley, B., and Wiedenmann, J. 2021. Final Report of the Index Based Methods Working Group.

Maunder, M.N., and Piner, K.R. 2015. Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. **72**(1): 7–18. doi:10.1093/icesjms/fsu015.

Northeast Fisheries Science Center. 2015. Operational Assessment of 20 Northeast Groundfish Stocks, Updated Through 2014. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 15-24. 166 Water Street, Woods Hole, MA 02543-1026.

Wiedenmann, J., Free, C.M., and Jensen, O.P. 2019. Evaluating the performance of data-limited methods for setting catch targets through application to data-rich stocks: A case study using Northeast U.S. fish stocks. Fish. Res. **209**: 129–142. doi:10.1016/j.fishres.2018.09.018.

Wiedenmann, J., and Jensen, O.P. 2018. Uncertainty in stock assessment estimates for New England groundfish and its impact on achieving target harvest rates. Can. J. Fish. Aquat. Sci. **75**(3): 342–356. doi:10.1139/cjfas-2016-0484.