

EVALUATING THE PERFORMANCE AND APPLICABILITY OF STATE-SPACE STOCK  
ASSESSMENT MODELS

By

Emily Morgan Liljestränd

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Fisheries and Wildlife – Doctor of Philosophy  
Ecology, Evolutionary Biology and Behavior – Dual Major

2023

## ABSTRACT

Fisheries stock assessment models are used to estimate population demographics and dynamics such as abundance, biomass, and fishing mortality from input fishery data including total catch, composition of catch, and fishing effort. A goal of stock assessment is to accurately quantify population and fishery dynamics so stocks can be managed to achieve fishery objectives and long-term sustainability. Accurate and precise model estimates can be attained by using models and techniques that account for ecological complexity like variability in quantities across ages, years, or regions without overparameterization. The state-space framework is one such statistical technique that may allow for incorporating more stochasticity such that the model can better reflect reality. The state-space modeling framework assumes that unobserved “states” develop over time due to process error modeled as a random effect and that observed data have expected values based on these states but differ from expectations due to observation error.

State-space stock assessment models (SSSAM) have experienced an outpouring of research and application in the past decade as computation processing power and novel software has facilitated the approximation of the high-level integrals necessary for SSSAM. SSSAM allows for several time-varying processes in recruitment, numbers at age, mortality, selectivity, and catchability, and has become an essential part of the contemporary fisheries modeling toolbox. With their swift advancement, it is important to understand best practices of applying state-space stock assessment models, and how data availability, the variability of process or observation error, and model structure may influence model results and accuracy.

In Chapter 1 I built an age-based state-space stock assessment model that used fisheries dependent data, rather than fisheries independent surveys, as an index of abundance and was applied to Lake Michigan lake whitefish (*Coregonus clupeaformis*). The model predicted greater

abundance and lower mortality compared to the non-state-space model and domed rather than asymptotic selectivity. The state-space model also had reduced retrospective patterns in recruitment. Chapters 2 and 3 each used a simulation-estimation framework to generate catch and index data using a state-space stock assessment that assumed process variability in recruitment, expected survival (abundance), and selectivity. Simulations were based on a Gulf of Maine haddock (*Melanogrammus aeglefinus*) model and assumed different degrees of process and observation variance (Chapter 2) or assumed observation error likelihood distributions for the proportions at age (Chapter 3) to generate data. Simulated data were input into several estimation models with alternative assumptions about contributing sources of process variability and/or observation error distributions. The results show that state-space models which assume several sources of process variability can produce unbiased estimates even when processes are constant over time. The state-space models were able to estimate process variance in several stochastic processes under a broad range of true values. However, assuming variability in expected survival when it is deterministic can lead to the model not converging. Unbiased results are achieved when the observation likelihood is structured to account for inter-age correlation and overdispersion though such a framework may have difficulty allocating variance between process and observation sub-models.

The concluding chapter places this work in the context of other age-based stock assessment models and argues for the inclusion of state-space in the modeling toolbox, as they can account for multiple time-varying processes and be used in a broad range of data contexts. This work provides a blueprint for where and how SSSAM may be best utilized in the future, particularly with data limited or data poor stocks and in cases where the process variance is unknown and should be estimated within the model.

## ACKNOWLEDGMENTS

I am deeply grateful to everyone who mentored me, guided me, taught me, or provided me advice and feedback during my graduate program at Michigan State University. From those who instructed me in hierarchical modeling in the classroom, to those who led me through the administrative difficulties of being a graduate student, to even those who showed me the best way to coax ketchup from a stubborn bottle, you know who you are and that you are in my thanks. I want to acknowledge my fellow members of the Quantitative Fisheries Center, graduate students and post-docs in the Fisheries and Wildlife and Ecology, Evolution, and Behavior programs at Michigan State University, and other NOAA/Sea Grant Population Dynamics Fellows (i.e., my “fellow fellows”). In particular, Grant Adams, Matt Damiano, Nick Fisch, and Amanda Hart have been consistent supporters and have given me extensive feedback and advice. Andie Perreault, Matt Cheng, and Andres Beita are stellar and have been generous with their time and expertise. Sara Hugentobler, Zac Johnston, and Miranda Wade thank you for providing my Michigan homes away from home and thanks Zac especially for teaching me to use the HPCC. Thank you to Ben Kline for being a constant source of optimism. Liz Stebbins, you are brilliant and the greatest roommate, coworker, and friend a person could ever ask for. Beth Gerstner, thank you for all the coffee dates.

I could not have completed my work or shared it with quite as many people if not for the years of funding and support from all levels of the University. Thank you to my PhD committee, Travis Brenden, Mary Bremigan, and Elise Zipkin for your guidance and advice. Thank you to my many teachers and mentors in the Fisheries and Wildlife department who gave me the educational foundation to complete this work including Rique Campa, JP Steibel, Brian Roth, Mike Wagner, and Erin Dreelin. Members of the Modeling Sub-Committee of the 1836 Treaty



Waters offered feedback and guidance. Ted Treska and Charles Perretti provided data. Stuart Ludsin at Ohio State University was an endless wellspring of wisdom and life advice. My work was made possible by the QFC Ricker Fellowship, MSU Graduate School Research Enhancement Award, NMFS-Sea Grant Joint Fellowship, The College of Agriculture and Natural Resources (CANR) Alumni Fellowship, the IAGLR David M. Dolan Scholarship, and the Fisheries and Wildlife Robert C. Ball and Betty A. Ball Fisheries and Wildlife Fellowship, as well as many travel grants, too many to name.

Thank you to Jim Bence and Jon Deroba who have been patient and compassionate teachers, and my undeniable role models in mentorship. Your wisdom and passion for fisheries science is contagious. Everyone who works with Jon at the Northeast Fisheries Science Center during my “sabbatical” there were incredibly supporting, and I cannot thank Julie Nieland, Liz Brooks, Tim Miller, and Dan Hennen enough for creating a welcoming environment.

And lastly, I could not have done this without the love and encouragement of my family. Avery Baker graciously offered me her home, her standing desk, and her dog Louie to take on much needed afternoon walks. My parents Howard Liljestrand and Blinda McClelland set me on my scientific path and gave me substantial financial and technical support. My brother Frasier prepared me for the realities and challenges of being a graduate student. And most importantly, thank you to my best friend and partner Laura de Castro Quaglia. Every word of encouragement, every time you restarted my stalled-out R code for me, and every cup of coffee you brewed made it possible for me to reach this goal. I owe everything to you.

## TABLE OF CONTENTS

LIST OF ABBREVIATIONS .....	vii
INTRODUCTION .....	1
REFERENCES .....	6
CHAPTER 1: APPLYING A NOVEL STATE-SPACE STOCK ASSESSMENT FRAMEWORK USING A FISHERIES-DEPENDENT INDEX OF FISHING MORTALITY .....	
1.1 Abstract .....	10
1.2 Introduction.....	11
1.3 Methods .....	14
1.4 Results.....	26
1.5 Discussion .....	35
REFERENCES .....	42
APPENDIX 1A: DATA COLLECTION AND PROCESSING .....	49
APPENDIX 1B: CURRENT STATISTICAL CATCH AT AGE MODEL (SCA) .....	52
APPENDIX 1C: MODEL DIAGNOSTICS FOR THE STATE-SPACE MODEL .....	66
CHAPTER 2: THE EFFECT OF PROCESS VARIABILITY AND DATA QUALITY ON PERFORMANCE OF A STATE-SPACE STOCK ASSESSMENT MODEL .....	
2.1 Abstract .....	72
2.2 Introduction.....	73
2.3 Methods .....	76
2.4 Results.....	89
2.5 Discussion .....	102
REFERENCES .....	108
APPENDIX 2A: INPUT OBSERVATION VARIANCE VALUES .....	113
APPENDIX 2B: SUPPLEMENTARY FIGURES .....	115
CHAPTER 3: THE PERFORMANCE OF ALTERNATIVE AGE COMPOSITION LIKELIHOODS IN STATE-SPACE MODELS WITH MULTIPLE SOURCES OF PROCESS VARIATION .....	
3.1 Abstract .....	132
3.2 Introduction.....	133
3.3 Methods .....	136
3.4 Results.....	144
3.5 Discussion .....	164
REFERENCES .....	169
APPENDIX 3A: PREDICTED VALUES AND SUPPLEMENTARY DATA .....	172
CONCLUSIONS AND RECOMMENDATIONS .....	176

## LIST OF ABBREVIATIONS

SSSAM or SAM	State-space Stock Assessment Model
SCAA or SCA	Statistical catch at age
NEFSC	Northeast Fisheries Science Center
WHAM	Woods Hole Assessment Model
ASAP	Age-structured Assessment Program
SAM	State-space Assessment Model
CORA	Chippewa Ottawa Resource Authority
NOAA	National Oceanic and Atmospheric Administration
NMFS	National Marine Fisheries Service
ICES	International Council for the Exploration of the Sea
SSB	Spawning Stock Biomass
MN	Multinomial Distribution
DM	Dirichlet-Multinomial Distribution
LG	Logistic Normal Distribution
MVTW	Multivariate Tweedie Distribution
MVN	Multivariate Lognormal Distribution
kg	Kilograms

## INTRODUCTION

The aphorism “all models are wrong, some are useful” (attributed to statistician George Box) acknowledges that describing real-world systems as a collection of neat mathematical equations will inherently contain inaccuracies. Fisheries stock assessment models attempt to estimate the unobservable - abundance, natural mortality, exploitation rate - using observed catch and survey data by making explicit assumptions about how those observations are mathematically related to life history and fishery parameters. Though the resultant output will be “wrong” in some way, by designing models that account for as much of the natural system as is statistically and mathematically possible, we can strive to make the results more “useful” towards achieving long term fisheries management goals.

The history of stock assessment modeling has been typified by progressively adding complexity and realism to the models as technology and data availability have permitted. The 1980s saw the development of the statistical catch at age model (SCAA), which assumed that observed data contained error as the result of observation variability or “observation error” which could be quantified with a statistical distribution. Models could then be fit by maximizing the likelihood of the data conditional on a set of parameter values (Fournier and Archibald, 1982). When such integrated models, (those that use several data sources to fit the model), became standard, scientists had to contend with data weighting (i.e., how closely to fit some data at the expense of poorly fitting others). Research focused on how to specify or calculate variance parameters of the likelihood function that would correctly assign variability (i.e., “weight”) among catch, survey indices of abundance, and/or proportions at age (Deriso et al., 2007; Maunder and Punt, 2013). Often, data were weighted *a priori* through iterative reweighting

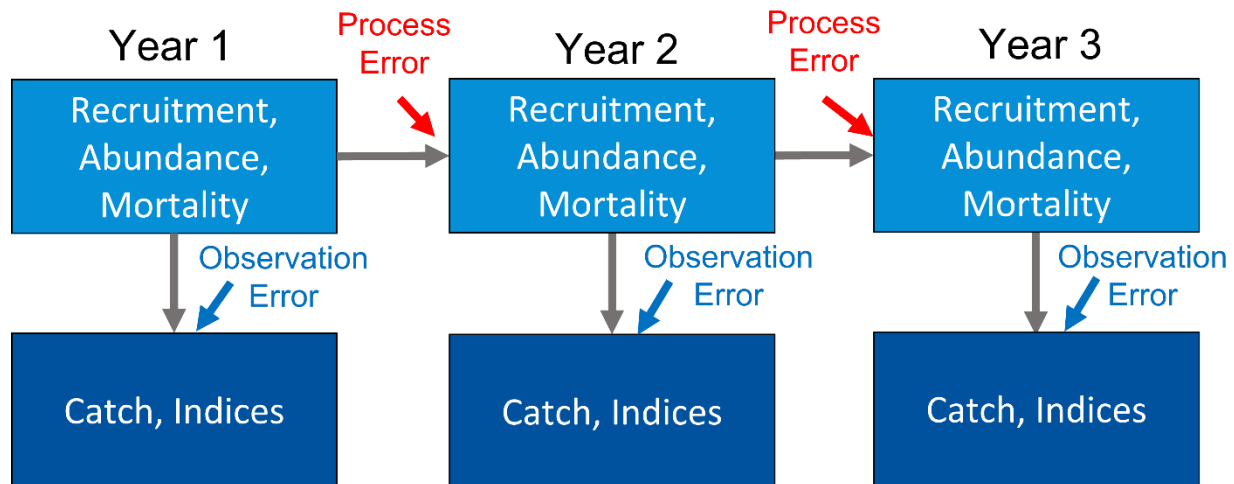
algorithms, as a function of sampling trips, or based on expert judgement (Francis, 2011; Truesdell et al., 2017).

Another important branch of stock assessment science aimed to introduce more stochasticity into previously constant or deterministic parameters like recruitment, catchability, natural mortality, and selectivity in SCAA (Deroba and Schueller, 2013; Linton and Bence, 2011; Maunder and Deriso, 2003; Wilberg and Bence, 2006). This was typically accomplished by constraining deviations from a mean or predicted value through a penalty term in a joint or “penalized” likelihood (in addition to the likelihood of the data). The variance of these deviations was usually specified using fixed ratios between the observation and process variation or assumed known *a priori*.

The development of state-space stock assessment models (SSSAM) that use an integrated likelihood, rather than the penalized likelihood, have been able to address both data weighting and process variance concerns. Such models have been able to capitalize on advanced mathematical programs that can approximate complex high-level mathematical integrals of the marginal likelihood (Kristensen *et al.*, 2016). This change has led to the estimability of process variance parameters and, under certain assumed likelihood functions for the observed catch, the observation variance parameters (Cadigan, 2015; Miller et al., 2016; Nielsen and Berg, 2014). Now the model, rather than the modeler, could determine the degree of process and observation error, and by doing so, reduce the risk of bias and model misspecification (de Valpine and Hilborn, 2005). Knowing the variability of ecological processes, especially recruitment, can guide model projections which is vital given that forecasting the future abundance of a stock is key to sustainability (Thorson et al., 2014).

A state-space stock assessment model (SSSAM) is a hierarchical model with two sub-model components - a “process” model that describes the population dynamics (birth, growth, death of the fish) and “observation” model that describes the fishery (how harvest or other data are collected) (Figure 0.1). The process model quantifies how underlying unobserved “states” like abundance, recruitment, or mortality progress over time as a function of the state in the previous year and the process error (Aeberhard et al., 2018; Auger-Méthé et al., 2016). The observation model describes how the observed data like catch or indices of abundance are a function of the state in each year and observation error. In a frequentist framework, the process and observation error are quantified with probability distributions and the former is integrated from a marginal likelihood that integrates across all possible state values.

Figure 0.1. Simplified structure of a state-space stock assessment model.



SSSAMs are now employed as a supplement to traditional stock assessment methods in the United States, Canada, and Europe (Cadigan, 2015; Miller et al., 2016; Nielsen and Berg, 2014). Two R modeling packages have been developed- WHAM in the Northeastern United States and SAM in Northern Europe- to facilitate fitting state-space stock assessment models (Nielsen and Berg, 2014; Stock and Miller, 2021). SSSAM have been used for multispecies models, multifleet models, and to flexibly estimate process variance where once it was fixed or

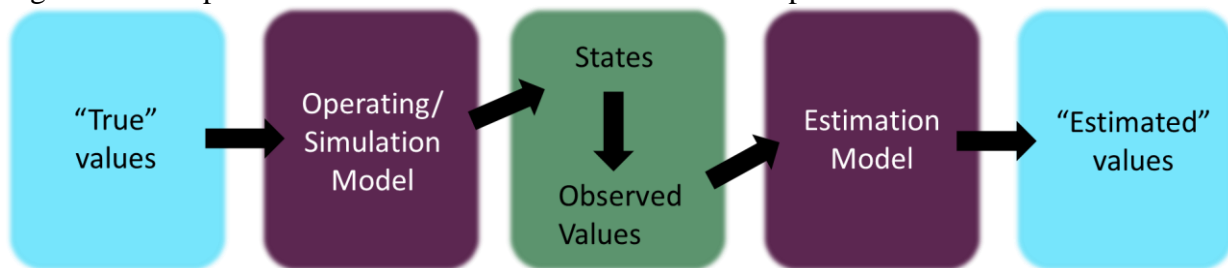
estimated with constraints (Albertsen et al., 2018; Nielsen et al., 2021; Perreault and Cadigan, 2021; Stock et al., 2021).

Several SSSAMs have been able to estimate observation error variance and thus self-weight the data within the model (Berg and Nielsen, 2016; Cadigan, 2015). However, this ability is contingent on how the observation model is parameterized, usually by fitting catch-at-age data (rather than total catch and proportions at age) using a multivariate lognormal distribution. As research and application of SSSAM was flourishing in the past decade, so was research into alternative likelihood distributions for the proportion at age data, other than the conventional multinomial, that better reflect the procedures used to collect age information and adjust the information content of such data (Francis, 2011, 2014; Maunder, 2011). Alternatives such as the Dirichlet-multinomial, logistic normal, or multivariate Tweedie, may better quantify overdispersion and inter-age correlation (Fisch et al., 2021; Thorson et al., 2022, 2017). These two avenues of inquiry - into SSSAM and alternative likelihoods - have recently begun to intersect and there is considerable interest in how to parameterize both the process and observation sub-models to correctly apportion variability between the two and among multiple time-varying processes or multiple data sources within the process and observation models, respectively (Albertsen *et al.*, 2017; Cronin-Fine and Punt, 2021; Xu *et al.*, 2020).

As the interest in applying SSSAM and alternative age composition likelihoods continues to expand, it is more important than ever to understand where and how this framework is best applied. This can and has involved 1) applying stock assessment models to real data in novel contexts and comparing the output against non-state-space models (e.g., Perreault *et al.*, 2020) or 2) fitting state-space models with simulated data using a simulation-estimation procedure (e.g., Miller and Hyun, 2018) (Figure 0.2). Self- and cross- tests are especially elucidating and can

quantify model robustness and the illustrated consequences of model misspecification (Deroba *et al.*, 2015). Simulation-estimation testing asks not only how well models perform under ideal (correctly specified) circumstances, but also what is the preferred model to choose when the true underlying dynamics are *unknown*. When faced with substantial uncertainty or poor data quality, which model configuration will “get it right” most of the time? Among many wrong models, which is the most useful?

Figure 0.2. Simplified structure of a simulation-estimation experiment.



This research employs both approaches- fitting both real and simulated data- to understand how SSSAM can be applied when fisheries independent indices of abundance are not available (Chapter 1), when the underlying recruitment, expected survival, and selectivity processes are highly variable or constant or the data are “noisy” (Chapter 2), and when the observation error in proportions at age has substantial overdispersion or inter-age correlation (i.e., non-multinomial distribution) (Chapter 3). The findings reveal some of the current upper limits of model complexity for SSSAM and how these models can be designed in the future to accurately and precisely estimate recruitment, abundance, spawning stock biomass, and exploitation rate, and process or observation error variances.



## REFERENCES

- Aeberhard, W.H., Flemming, J.M., Nielsen, A., 2018. Review of state-space models for fisheries science. *Annu. Rev. Stat. Appl.* 5, 215-35. <https://doi.org/10.1146/annurev-statistics-031017-100427>.
- Albertsen, C.M., Nielsen, A., Thygesen, U.H., 2018. Connecting single-stock assessment models through correlated survival. *ICES J. Mar. Sci.* 75, 235-244. <https://doi.org/10.1093/icesjms/fsx114>.
- Albertsen, C.M., Nielsen, A., Thygesen, U.H., 2017. Choosing the observational likelihood in state-space stock assessment models. *Can. J. Fish. Aquat. Sci.* 74, 779-789. <https://doi.org/10.1139/cjfas-2015-0532>.
- Auger-Méthé, M., Field, C., Albertsen, C.M., Derocher, A.E., Lewis, M.A., Jonsen, I.D., Flemming, J.M., 2016. State-space models' dirty little secrets: Even simple linear Gaussian models can have estimation problems. *Sci. Rep.* 6, 1-10. <https://doi.org/10.1038/srep26677>.
- Berg, C.W., Nielsen, A., 2016. Accounting for correlated observations in an age-based state-space stock assessment model. *ICES J. Mar. Sci.* 73, 1788-1797. <https://doi.org/10.1093/icesjms/fsw046>.
- Cadigan, N.G., 2015. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. *Can. J. Fish. Aquat. Sci.* 308, 1-13. <https://doi.org/10.1139/cjfas-2015-0047>.
- Cronin-Fine, L., Punt, A.E., 2021. Modeling time-varying selectivity in size-structured assessment models. *Fish. Res.* 239, 105927. <https://doi.org/10.1016/j.fishres.2021.105927>.
- de Valpine, P., Hilborn, R., 2005. State-space likelihoods for nonlinear fisheries time-series. *Can. J. Fish. Aquat. Sci.* 62, 1937-1952. <https://doi.org/10.1139/f05-116>.
- Deriso, R.B., Maunder, M.N., Skalski, J.R., 2007. Variance estimation in integrated assessment models and its importance for hypothesis testing. *Can. J. Fish. Aquat. Sci.* 64, 187-197. <https://doi.org/10.1139/f06-178>.
- Deroba, J.J., Butterworth, D.S., Methot, R.D., DeOliveria, J.A.A., Fernandez, C., Nielsen, A., Cadrin, S.X., Dickey-Collas, M., Legault, C.M., Ianelli, J.N., Valero, J.L., Needle, C.L., Malley, J.M.O., Chang, Y., 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., 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, 19-30. <https://doi.org/10.1093/icesjms/fst237>.
- Deroba, J.J., Schueller, A.M., 2013. Performance of stock assessments with misspecified age- and time-varying natural mortality. *Fish. Res.* 146, 27-40. <https://doi.org/10.1016/j.fishres.2013.03.015>.
- Fisch, N., Camp, E., Shertzer, K., Ahrens, R., 2021. Assessing likelihoods for fitting composition data within stock assessments, with emphasis on different degrees of process and observation error. *Fish. Res.* 243, 106069. <https://doi.org/10.1016/j.fishres.2021.106069>.
- Fournier, D.A., Archibald, C.P., 1982. A general theory for analyzing catch at age data. *Can. J. Fish. Aquat. Sci.* 39, 1195-1207. <https://doi.org/10.1139/f82-157>.
- Francis, R.I.C.C., 2011. Corrigendum: Data weighting in statistical fisheries stock assessment models. *Can. J. Fish. Aquat. Sci.* 68, 2228-2228. <https://doi.org/10.1139/f2011-165>.
- Francis, R.I.C.C., 2014. Replacing the multinomial in stock assessment models: A first step. *Fish. Res.* 151, 70-84. <https://doi.org/10.1016/j.fishres.2013.12.015>.
- Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., Bell, B.M., 2016. TMB: Automatic differentiation and laplace approximation. *J. Stat. Softw.* 70, 1-21. <https://doi.org/10.18637/jss.v070.i05>.
- Linton, B.C., Bence, J.R., 2011. Catch-at-age assessment in the face of time-varying selectivity. *ICES J. Mar. Sci.* 68, 611-625. <https://doi.org/10.1093/icesjms/fsq173>.
- Maunder, M.N., 2011. Review and evaluation of likelihood functions for composition data in stock-assessment models: Estimating the effective sample size. *Fish. Res.* 109, 311-319. <https://doi.org/10.1016/j.fishres.2011.02.018>.
- Maunder, M.N., Deriso, R.B., 2003. Estimation of recruitment in catch-at-age models. *Can. J. Fish. Aquat. Sci.* 60, 1204-1216. <https://doi.org/10.1139/f03-104>.
- Maunder, M.N., Punt, A.E., 2013. A review of integrated analysis in fisheries stock assessment. *Fish. Res.* 142, 61-74. <https://doi.org/10.1016/j.fishres.2012.07.025>.
- Miller, T.J., Hare, J.A., Alade, L.A., 2016. A state-space approach to incorporating environmental effects on recruitment in an age-structured assessment model with an application to southern new england yellowtail flounder. *Can. J. Fish. Aquat. Sci.* 73, 1261-1270. <https://doi.org/10.1139/cjfas-2015-0339>.
- Miller, T.J., Hyun, S., 2018. Evaluating evidence for alternative natural mortality and process error assumptions using a state-space, age-structured assessment model. *Can. J. Fish. Aquat. Sci.* 75, 691-703. <https://doi.org/10.1139/cjfas-2017-0035>.

- Nielsen, A., Berg, C.W., 2014. Estimation of time-varying selectivity in stock assessments using state-space models. *Fish. Res.* 158, 96-101. <https://doi.org/10.1016/j.fishres.2014.01.014>.
- Nielsen, A., Berg, C.W., Hintzen, N.T., Mosegaard, H., Trijoulet, V., 2021. Multi-fleet state-space assessment model strengthens confidence in single-fleet SAM and provides fleet-specific forecast options. *ICES J. Mar. Sci.* 78, 2043-2052. <https://doi.org/10.1093/icesjms/fsab078>.
- Perreault, A.M.J., Cadigan, N.G., 2021. Natural mortality diagnostics for state-space stock assessment models. *Fish. Res.* 243, 106062. <https://doi.org/10.1016/j.fishres.2021.106062>.
- Perreault, A.M.J., Wheeland, L.J., Joanne Morgan, M., Cadigan, N.G., 2020. A state-space stock assessment model for American plaice on the Grand Bank of Newfoundland. *J. Northwest Atl. Fish. Sci.* 51, 45-104. <https://doi.org/10.2960/J.v51.m727>.
- Stock, B.C., Miller, T.J., 2021. The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time- and age- varying processes via random effects and links to environmental covariates. *Fish. Res.* 240, 105967. <https://doi.org/10.1016/j.fishres.2021.105967>.
- Stock, B.C., Xu, H., Miller, T.J., Thorson, J.T., Nye, J.A., 2021. Implementing two-dimensional autocorrelation in either survival or natural mortality improves a state-space assessment model for Southern New England-Mid Atlantic yellowtail flounder. *Fish. Res.* 237, 105873. <https://doi.org/10.1016/j.fishres.2021.105873>.
- Thorson, J.T., Jensen, O.P., Zipkin, E.F., 2014. How variable is recruitment for exploited marine fishes? A hierarchical model for testing life history theory. *Can. J. Fish. Aquat. Sci.* 71, 973-983. <https://doi.org/10.1139/cjfas-2013-0645>.
- Thorson, J.T., Johnson, K.F., Methot, R.D., Taylor, I.G., 2017. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. *Fish. Res.* 192, 84-93. <https://doi.org/10.1016/j.fishres.2016.06.005>.
- Thorson, J.T., Miller, T.J., Stock, B.C., 2022. The multivariate-Tweedie: a self-weighting likelihood for age and length composition data arising from hierarchical sampling designs. *ICES J. Mar. Sci.* fsac159. <https://doi.org/10.1093/icesjms/fsac159>.
- Truesdell, S.B., Bence, J.R., Syslo, J.M., Ebener, M.P., 2017. Estimating multinomial effective sample size in catch-at-age and catch-at-size models. *Fish. Res.* 192, 66-83. <https://doi.org/10.1016/j.fishres.2016.11.003>.
- Wilberg, M.J., Bence, J.R., 2006. Performance of time-varying catchability estimators in statistical catch-at-age analysis. *Can. J. Fish. Aquat. Sci.* 63, 2275-2285. <https://doi.org/10.1139/f06-111>.

Xu, H., Thorson, J.T., Methot, R.D., 2020. Comparing the performance of three data-weighting methods when allowing for time-varying selectivity. *Can. J. Fish. Aquat. Sci.* 77, 247-263. <https://doi.org/10.1139/cjfas-2019-0107>.

CHAPTER 1:  
APPLYING A NOVEL STATE-SPACE STOCK ASSESSMENT FRAMEWORK USING A  
FISHERIES-DEPENDENT INDEX OF FISHING MORTALITY

1.1 Abstract

State-space models have a hierarchical framework that assumes the observed data are derived from a time series of unobserved latent states. State-space stock assessment models have emerged as an alternative framework to conduct stock assessment in Canada, the east coast of the United States of America, and in Europe though little research has investigated where and how they are optimally used and if they would be appropriate in other management contexts. I built a novel state-space stock assessment model with process variation in recruitment and time- and age- specific catchability. I fit the model to fisheries dependent commercial trap net and gill net catch and effort data of Lake Michigan lake whitefish (*Coregonus clupeaformis*) using a marginal likelihood that integrated over latent states. Compared to the previously employed statistical catch at age model, the state-space model estimated dome-shaped, rather than asymptotic selectivity for both fisheries, 15% lower average total instantaneous mortality, and 20% higher average recruitment. To my knowledge this is the first application of a state-space stock assessment model fit by maximum likelihood in the Laurentian Great Lakes and the first such model to exclude a fisheries-independent survey. These results demonstrate the feasibility of employing a maximum likelihood state-space framework in fisheries that lack such fishery independent indices of abundance and instead use catch per unit effort as an index of abundance. This work presents a novel approach to applying state-space stock assessment modeling and offers insights and suggestions for future in the Great Lakes and in similar circumstances of data availability.

## 1.2 Introduction

Fisheries stock assessment must contend with multiple sources of model variability, both from the process sub-model (i.e., process variation) and from the observation sub-model (i.e., observation error) which introduce uncertainty and obscure the relationship between underlying model dynamics and the observed data (Fournier and Archibald, 1982). Until recently, age-structured fisheries stock assessment models have generally assumed the degree of process variation and/or observation error was known. Deriso *et al.* (2007) emphasized that it is possible to estimate the variances, even from multiple sources, when all variation is due to observation error. If process variation is present and deviations from the mean are treated like another source of data, the variance of those deviations cannot be independently estimated by maximizing the joint likelihood of the deviations and data, a procedure sometimes called penalized likelihood (see Schnute 1994). The penalized likelihood, also called “errors-in-variables” (EV) method, is still used in many stock assessment packages such as ASAP, MULTIFAN-CL, and stock synthesis (SS) (Fournier *et al.*, 1998; Legault and Restrepo, 1998; Methot and Wetzel, 2013). However, because the estimates can be asymptotically biased, alternative approaches have been adopted in newer assessment platforms (de Valpine and Hilborn, 2005; Stock and Miller, 2021). State-space models (SSM) have been able to partition process variation and observation error and estimate both with high precision (Aeberhard *et al.*, 2018b). SSM accomplish this by assuming the data are observations (with error) derived from unobserved states like population abundance and mortality, that are progressing through time following a stepwise Markovian process. The states, or their deviations from a shared mean, are specified as random effects and in a maximum likelihood framework are integrated over to obtain the marginal likelihood, which is maximized.

The theorized advantages of state-space stock assessment models have been realized, with models applied in several different geographical regions, using different model configurations, able to simultaneously estimate the magnitude of several different sources of process variation, including recruitment, survival, fishing mortality, selectivity, and natural mortality (Aeberhard *et al.*, 2018; Cadigan, 2015; Stock *et al.*, 2021; Stock and Miller, 2021; Yin *et al.*, 2019). State-space stock assessment models (herein defined as those that use a marginal likelihood) have estimated mortality and abundance greater accuracy and precision than non-state-space models (herein defined as those that use a penalized likelihood), and exhibit less retrospective patterns (Gunnlaugsson, 2012; Perreault *et al.*, 2020; Stock *et al.*, 2021). A state-space modeling research track is currently exploring the feasibility of applying such models to NOAA-managed stocks in New England and the Mid-Atlantic and state-space stock assessment models are used for a large portion of stocks managed by the International Council for the Exploration of the Sea (ICES) (Aanes *et al.*, 2020). Nevertheless, there remains considerable interest and uncertainty in what kinds of process variation can be considered, and under what circumstances and with what data the SSM approach can be practically implemented (Aanes *et al.*, 2020).

This work contributes to the ongoing inquiry by using a state-space modeling framework to fit total catch and age composition data for two fisheries (with two distinct gears) without a fisheries-independent index of abundance. Indices quantify trends in stock abundance over time and are often assumed to be proportional to abundance therefore allowing an estimation of stock size. In the absence of a fisheries independent survey, an index can be calculated from data collected during the fishing process, such as catch per unit effort (CPUE). However, CPUE is only a reliable index of abundance if the proportionality constant, catchability, is correctly

specified (Wilberg *et al.*, 2009). Catchability is often assumed constant or varying around a constant mean, though can vary systematically due to density-dependent effects, gear changes, preferential sampling, or angler behavior (Ducharme-Barth *et al.*, 2022; Hilborn and Walters, 1992; Quirijns *et al.*, 2008). In such cases, potential causes of changes in catchability might be unknown, but can still be partially accounted for by allowing catchability to vary through a stochastic process, an approach I adopt herein (Wilberg *et al.*, 2009).

This novel state-space model was applied to lake whitefish (*Coregonus clupeaformis*) in the northmost region of Lake Michigan. This stock experiences no recreational fishing and two commercial fisheries, a trap net fishery and a gill net fishery, have yielded on average 222 and 159 thousand kilograms per year, respectively, since 1986, the first year included in the assessment. Fishing effort is reported in numbers of lifts of the trap nets and feet of gill net deployed. These effort data were used to inform fishing mortality within the model and catchability was a state variable estimated for each age and year that scaled effort to fishing mortality. This parameterization distinguishes this state-space model from previous applications and existing state-space platforms like SAM or WHAM where instantaneous fishing mortality is a state variable (Nielsen and Berg, 2014; Perreault *et al.*, 2020).

To my knowledge, this is the first state-space stock assessment model application in the Laurentian Great Lakes (fit by marginal maximum likelihood) and this work confirms the applicability of SSM without a fisheries independent survey, which is customary in previous applications (Cadigan, 2015; Nielsen *et al.*, 2021; Perreault *et al.*, 2020; Stock and Miller, 2021). My objective was to apply a novel state-space stock assessment model and to provide time- and age- varying estimates of recruitment and catchability and their associated variances. Ultimately,



I developed an assessment tool applicable in situations without fisheries independent surveys and from this tool I make general recommendations about the use of SSM models in such situations.

### 1.3 Methods

Table 1.1. Symbols used in the text, appendices, tables, and figures. Whether the symbol is used in the statistical catch at age model (SCA), the state-space model (SSM), or both, is indicated. The value is provided for indices and time and age invariant data.

Symbol	Description	Usage	Value(s)
<b>Index Variables</b>			
$y$	Year	Both	1986-2017
$a, \tilde{a}$	Age	Both	4-15+
$G$	Fishery	Both	g- gill net, t-trap net
<b>Data and Priors</b>			
$E_{y,G}$	Observed effort by year and fishery	Both	See Figure 2
$\tilde{C}_{y,G}$	Observed harvest by year and fishery	Both	
$\tilde{P}_{a,y,G}$	Observed proportions by age, year, and fishery	Both	
$\tilde{M}$	Observed natural mortality	SSM	0.2
$\hat{M}$	Median of prior on natural mortality	SCA	0.2
$\sigma_M$	Standard deviation of observed or prior natural mortality	Both	0.1
$m_{a,y}$	Maturity by age and year	Both	
$W_{a,y}^{(spawn)}$	Weight at the time of spawning by age and year	Both	
$e$	Average number of eggs per kilogram	SCA	19937
$f$	Average proportion of females in spawning population	SCA	0.4845
$\varphi_G$	Standard deviation multiplier for fishery-specific catchability, $q_{y,G}$ , process variability	SCA	1.5
$\varphi_s$	Standard deviation multiplier for trap net selectivity parameter $p_{1,y,t}$ process variability	SCA	1.0
$\varphi_R$	Standard deviation multiplier for recruitment process variability	SCA	15
$\varphi_{CG}$	Standard deviation multiplier for total catch observation error	SCA	4
$\sigma_{CG}$	Standard deviation of likelihood of total catch	SSM	0.067
<b>Parameters and State Variables</b>			
$M$	Instantaneous natural mortality	Both	
$\sigma_G$	Standard deviation of process variability of fishery-specific catchability	SSM	
$\rho_G$	Correlation coefficient (among ages) of process variability of fishery-specific catchability	SSM	
$\sigma_R$	Standard deviation of process variability of recruitment	SSM	
$\varepsilon_y^{(R)}$	Process variability of recruitment	SSM	
$\varepsilon_y^{(G)}$	Vector of process variability of fishery-specific catchability	SSM	
$\varepsilon_y^{(s)}$	Process variability of trap net selectivity parameter	SCA	
$q_{y,G}$	Catchability by year and fishery	SCA	

Table 1.1 (cont'd)

$\sigma$	Baseline standard deviation for process variability and observation error	SCA	
$p_{1,y,t}$	Trap net selectivity inflection point parameter by year	SCA	
$p_{1,g}$	Gill net selectivity parameter	SCA	
$p_{2,g}$	Trap and gill net selectivity parameters	SCA	
$\bar{R}$	Average recruitment and abundance in 1986	SCA	
$D_y$	Deviations by year from average recruitment	SCA	
$d_a$	Deviations by age from average abundance in 1986	SCA	
$\alpha$	Ricker stock recruitment parameter	SCA	
$\beta$	Ricker stock recruitment parameter	SCA	
<b>Derived Quantities</b>			
$N_{a,y}$	Abundance by age and year	Both	
$\hat{N}_{4,y}$	Predicted abundance of age 4 individuals by year (recruitment)	SCA	
$F_{a,y,G}$	Instantaneous fishing mortality by age, year, and fishery	Both	
$\mathbf{q}_{y,G}$	Vector of age-specific catchability by year and fishery	SSM	
$s_{a,y,G}$	Selectivity by age, year, and fishery	SCA	
$Z_{a,y}$	Instantaneous total mortality by age and year	Both	
$\bar{Z}_a$	Average instantaneous total mortality by age from 1986-1988	SCA	
$C_{a,y,G}$	Expected harvest by age, year, and fishery	Both	
$P_{a,y,G}$	Expected proportions by age, year, and fishery	Both	
$B_y^{(spawn)}$	Spawning stock biomass by year	Both	
$\epsilon_y$	Spawning stock size in eggs by year	SCA	
$T$	Total objective function value	Both	
$\sigma_a$	Standard deviation of multivariate normal random walk deviations of fishery-specific catchability	SSM	Set to single value, $\sigma_G$ , for each fishery
$n_G$	Effective sample size of multinomial likelihood of observing given proportion at age of catch	Both	
$\Sigma_G$	Covariance matrix of process variability of fishery-specific catchability	SSM	
$\sigma_s$	Standard deviation of process variability of trap net selectivity parameter	SCA	

### 1.3.1 Data and Model Inputs

The assessment model was fit using annual fishery-specific values of fishing effort ( $E_{y,G}$ ), observed fishery harvest ( $\tilde{C}_{y,G}$ ), and observed age compositions in units of proportions at age ( $\tilde{P}_{a,y,G}$ ) (variables are described in Table 1.1). Annual reported yield in weight from the catch reporting system was converted into estimates of annual catch in numbers using the annual

average weight of fish harvested by each fishery and estimates of underreporting (Appendix 1A). In addition, the model used input values of average weight at age, average length at age, and maturity schedules in each year, which were collected from harvest subsampling, to calculate selectivity and spawning stock biomass.

### 1.3.2 State-space Assessment Model

A state-space stock assessment model (SSM) was developed in Template Model Builder (TMB) that assumed two state variables, recruitment and catchability at age, varied across years (Kristensen *et al.*, 2016). The SSM was structurally similar to the existing statistical catch at age (SCA) models used to assess lake whitefish and lake trout in the Laurentian Great Lakes which fit data using a penalized likelihood (described in Appendix 1B). The SSM had structural elements that mimicked the state-space modeling software “SAM” including the state variables, process variance structure, and a marginal likelihood (Berg and Nielsen, 2016; Nielsen and Berg, 2014). The SSM had two sub-models: a process or “population” model, and an observation or “fishery” model. The process model described how the underlying states, recruitment and catchability, changed over time as a function of the state in the previous year and process variability. The observation model described how the predicted total catch and catch compositions derived from those true unobserved states owing to observation error.

Table 1.2. Equations used in the statistical catch at age model (SCA), the state-space model (SSM), or both. Note that though equations 3.5a and 3.5b appear identical, each contributes to a different part of the objective function total, which is an important distinction in the integrated likelihood framework.

Index	Description	Equation	Usage
<b>Process Model</b>			
1.1	Recruitment random walk	$\log N_{4,y} = \log N_{4,y-1} + \varepsilon_y^{(R)}; \varepsilon_y^{(R)} \sim N(0, \sigma_R^2)$	SSM
1.2	Abundance at age 4-9 in initial year	$\log N_{a,1986} = \log N_{4,1986-(a-4)} - \sum_4^{a-1} \log \bar{Z}_a, 4 < a \leq 9$	SSM
1.3	Abundance at age 9+ in initial year	$\log N_{a,1986} = 0, a > 9$	Both

Table 1.2 (cont'd)

1.4	Abundance at age exponential decay with plus group	$\log N_{a,y} = \log N_{a-1,y-1} - Z_{a-1,y-1}, 4 \leq a < A$ $\log N_{A,y} = \log(N_{A-1,y-1} e^{-Z_{A-1,y-1}} + N_{A,y-1} e^{-Z_{A,y-1}})$	Both
1.5	Total instantaneous mortality	$Z_{a,y} = M + \sum_{G=g,t} F_{a,y,G}$	Both
1.6a	Fishing mortality	$F_{a,y,G} = q_{y,G} E_{y,G} S_{a,y,G}, G = g, t$	SCA
1.6b	Fishing Mortality	$F_{a,y,G} = q_{a,y,G} E_{y,G}, G = g, t$	SSM
1.7a	Year-specific catchability random walk	$\log q_{y+1,G} = \log q_{y,G} + \varepsilon_y^{(G)}; \varepsilon_y^{(G)} \sim N(0, \sigma_G^2), G = g, t$	SCA
1.7bi	Year- and age- specific vector of catchability random walk	$\log \mathbf{q}_{y,G} = \log \mathbf{q}_{y-1,G} + \boldsymbol{\varepsilon}_y^{(G)}; \boldsymbol{\varepsilon}_y^{(G)} \sim N(0, \boldsymbol{\Sigma}_G), G = g, t$	SSM
1.7bii	Correlated error of catchability random walk	$\boldsymbol{\Sigma}_{a,\tilde{a}} = \rho^{ \tilde{a}-a } \sigma_a \sigma_{\tilde{a}}, 4 < a < A, 4 < \tilde{a} < A$	SSM
1.8	Spawning stock biomass	$B_y^{(spawn)} = \sum_{a=4}^A m_{a,y} W_{a,y}^{(spawn)} \log N_{a,y}$	Both
<b>Observation Model</b>			
2.1	Baranov catch equation	$C_{a,y,G} = \frac{F_{a,y,G}}{Z_{a,y}} (1 - \exp(-Z_{a,y})) N_{a,y}, G = g, t$	Both
2.2	Yearly catch proportions at age	$P_{a,y,G} = \frac{C_{a,y,G}}{\sum_{a=1}^A C_{a,y,G}}, G = g, t$	Both
<b>Negative Log Likelihoods</b>			
3.1	Total objective function on the log-scale is the sum of prior/penalty and data likelihood components	$T = \sum NLP + \sum NLL$	Both
3.2a	Recruitment deviations from those expected by Ricker curve (see Appendix B)	$NLP^{(R)} = \sum_{y=1991}^{2017} \frac{1}{2\sigma_R^2} \left( \log \frac{\hat{N}_{4,y}}{N_{4,y}} \right)^2 + \log \sigma_R$	SCA
3.2b	Recruitment deviations from those expected around 0	$NLP^{(R)} = \sum_{y=1981}^{2017} \frac{1}{2\sigma_R^2} \left( \varepsilon_y^{(R)} \right)^2 + \log \sigma_R$	SSM
3.3a	Fishery-specific deviations in catchability following a random walk	$NLP^{(G)} = \sum_{y=1986}^{2017} \frac{1}{2\sigma_G^2} \left( \varepsilon_y^{(G)} \right)^2 + \log \sigma_G, G = g, t$	SCA
3.3b	Fishery-specific deviations in vector of catchability at age following a random walk	$NLP^{(G)} = \sum_{y=1986}^{2017} \frac{1}{2} \boldsymbol{\varepsilon}_y^{(G)'} \boldsymbol{\Sigma}_T^{-1} \boldsymbol{\varepsilon}_y^{(G)} + \log \sqrt{ \boldsymbol{\Sigma}_G }, G = g, t$	SSM
3.4	Deviations for random walk for trap net selectivity parameter	$NLP^{(s)} = \sum_{y=1986}^{2017} \frac{1}{2\sigma_s^2} \left( \varepsilon_y^{(s)} \right)^2 + \log \sigma_s$	SCA
3.5a	Penalty on natural mortality from deviating from median of the prior	$NLP^{(M)} = \frac{1}{2\sigma_M^2} \left( \log \tilde{M} - \log M \right)^2 + \log \sigma_M$	SCA
3.5b	Likelihood of observing a given natural mortality value based on assumed true value	$NLL^{(M)} = \frac{1}{2\sigma_M^2} \left( \log \tilde{M} - \log M \right)^2 + \log \sigma_M$	SSM
3.6	Likelihood of observing the given total catch based on an assumed true value	$NLL^{(CG)} = \sum_{y=1986}^{2017} \frac{1}{2\sigma_{CG}^2} \left( \log \frac{C_{y,G}}{\tilde{C}_{y,G}} \right)^2 + \log \sigma_{CG}, G = g, t$	Both
3.7	Likelihood of observing the given proportions at age of catch based on assumed true values	$NLL^{(PG)} = - \sum_{y=1986}^{2017} \sum_{a=4}^A n_{y,G} \tilde{P}_{a,y,G} \log P_{a,y,G}, G = g, t$	Both

Recruitment was modeled as a random walk which effectively penalizes recruitment estimates that change greatly from year to year with the extent of the penalty depending on the random walk variance. Though recruitment can differ greatly between adjacent years due to several abiotic and biotic processes, treating these changes as arising from random walk is a recommended approach in a state-space framework especially when estimating the variance of the process variability (Maunder and Thorson, 2018). Recruitment,  $N_{4,y}$ , was calculated on the log-scale by adding normally distributed error to the log-scale recruitment in the previous year, starting 5 years prior to the first year being modeled (Table 1.2, Eq. 1.1). Abundances of individuals ages 5 to 9 in the initial year, 1986, were calculated using the recruitment from years preceding the time series, 1981-1985, which were estimated by extending the random walk process backwards in time (Eq. 1.2). This method assumed that recruitment prior to 1986 came from the same stochastic process as recruitment throughout the time series and the initial abundances are a function of prior recruitment, adjusted downward using average total mortality. Given that there was no information on mortality rates prior to 1986, the average estimated age-specific instantaneous mortality rates for 1986-1988,  $\bar{Z}_a$ , were used as reasonable substitutes. The abundance of older ages in the initial year were set to 0, consistent with the very low or zero harvest for cohorts that were age-9 or older in 1986 (Eq. 1.3).

Abundances in subsequent years and ages were a function of an exponential mortality model (Eqs. 1.4-1.5). Instantaneous fishing mortality was the product of age-, year-, and fishery-specific catchability,  $q_{a,y,G}$  and year- and fishery- specific effort,  $E_{y,G}$  (Eq. 1.6a), a common approach in stock assessment models (Fournier and Archibald, 1982). Effectively,  $q_{a,y,G}$  combines what is conventionally modeled as two separate processes, year-specific catchability, and age-specific selectivity, into a single age- and year- specific value. Time-varying selectivity

is commonly used in stock assessment to account for changes in size at age or changes in gear technology that may preferentially target certain ages, among other reasons (Martell and Stewart, 2014). The vector of log-scale catchability at age in a given year, for a given fishery, i.e.  $\mathbf{q}_{y,G} = (q_{4,y,G}, q_{5,y,G}, \dots, q_{A,y,G})$ , varied over time according to a random walk with multivariate normal error, analogous to how SAM models age specific instantaneous fishing mortality (Eq. 1.7bi) (Nielsen and Berg, 2014). The covariance matrix,  $\mathbf{\Sigma}$ , is a flexible matrix that quantifies the variability and correlation among ages of the year- and fishery- specific process variability vectors,  $\boldsymbol{\epsilon}_y^{(G)}$  (Eq. 1.7bii). Though the covariance matrix can be parameterized in many ways depending on assumed similarity of age-specific process variability, in this model the correlation structure was a first order autoregressive process (AR(1)), where the degree of correlation was a function of the absolute difference in ages. This correlation structure, when applied to fishing mortality or selectivity, has a demonstrably better fit to data in other stocks and is now the default structure in SAM models and a popular option in WHAM models (Berg and Nielsen, 2016; Nielsen et al., 2021; Nielsen and Berg, 2014; Stock and Miller, 2021). The assumption of stronger correlation in catchability variations between adjacent ages, implicit in the AR(1) structure, is sensible given that many factors (such as size or depth distributions of fish) will tend to be more similar for fish closer in age (Nielsen and Berg, 2014). Catchability in this case is expressed as a proportionality between fishing effort and fishing mortality. This is also the proportionality between annual catch and average annual abundance (Ricker, 1975). Thus, my assumption that fishing mortality is proportional to fishing effort is equivalent to treating annual catch divided by annual effort as an abundance index.

Fishery-specific correlation coefficients  $\rho_G$  and standard deviation,  $\sigma_G$ , parameters were estimated for each fishery and were invariant among ages. The correlation coefficients were

estimated on the logit scale, and thus constrained between 0 and 1. These determined how the random walk errors of age-specific catchabilities were related among ages. I did not allow for negative correlations as I could not envision a mechanism that would cause age-specific catchability to differ most for adjacent ages. At one extreme of the allowed range, log-scale catchabilities would change by the same amount for every age from year to year, and their trajectories over time would be parallel ( $\rho = 1$ ). At the other extreme age-specific catchabilities would develop over time completely independently ( $\rho = 0$ ). The estimated value determined the model dynamics between these extremes.

In the observation sub-model, the predicted age-, year-, and fishery- specific catch was calculated using the Baranov catch equation (Eq. 2.1). The predicted proportions at age were derived from those values (Eq. 2.2).

The SSM assessment model was fit by maximizing a marginal (integrated) likelihood. The random effects,  $\psi$ , were integrated out of the likelihood, so the objective function was conditional only on the fixed effect parameters,  $\lambda$ . The marginal likelihood contained two components, the likelihood of the data given the random effects and the parameters,  $L(\lambda|X, \psi)$ , and the probability density function of the random effects,  $f(\psi)$ ,

$$L(\lambda|X) = \int_{-\infty}^{\infty} L(\lambda|X, \psi)f(\psi)d\psi.$$

The likelihood of observing the total catch and catch composition were modeled as a lognormal distribution and a multinomial distribution, respectively (Eqs. 3.6-3.7). Estimated age- and year- invariant  $M$  was estimated using a normally distributed likelihood with mean  $\tilde{M} = 0.2$  and fixed standard deviation  $\sigma_M^2$  (Eq. 3.5b). Because of this distribution's inclusion in the likelihood portion of the marginal likelihood,  $\tilde{M}$  is effectively a single observed data point which the model fits by predicting  $M$ . Error from the random walk of log-scale recruitment,  $\log N_{4,y}$ ,

was assumed to come from a 0-mean normal distribution (Eq. 3.2b). Error in the age- and fishery- specific random walk of catchability was assumed to arise from a 0-mean multivariate normal distribution (Eq. 3.3b). The standard deviations for the observed log-scale total catch by fishery,  $\sigma_{CG}$ , were fixed at 0.067, the values used in the SCA model. These were found in the SCA model by following an iterative procedure during fitting that adjusted the ratios between process and observation error variance, until the target value of  $\sigma_{CG}$  was achieved, as described by Richards *et al.* (1997). This effectively means variance was set based on expert judgment at a level that assumes that two-thirds of observed catches would be within about 6.7% of the true values of catch. I fixed  $\sigma_{CG}$  at the same value in the SSM model because the model did not converge when I attempted to estimate it along with the other parameters. I conducted a sensitivity analysis that set  $\sigma_{CG}$  at values 50% above or below 0.067.

The point estimates of fixed effect parameters (including the standard deviation parameters for the random effects) were estimated by maximum likelihood. The predicted random effects and derived quantities (i.e., abundance, spawning stock biomass, mortality) were calculated using the “epsilon” method bias correction algorithm which is now standard in TMB (Thorson and Kristensen, 2016).

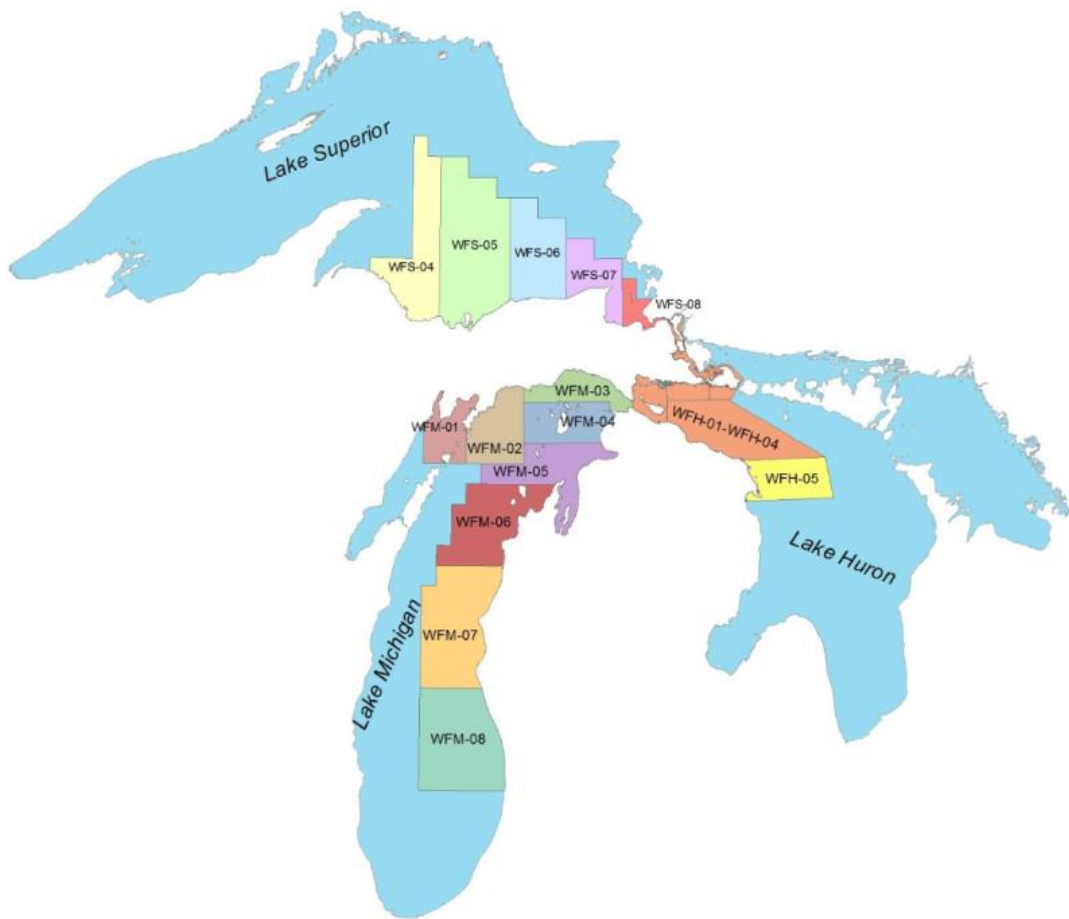
### 1.3.3 Case Study- Lake Michigan Lake Whitefish

The lake whitefish stock used as a case study for the application of the novel state-space stock assessment is the northmost management region of Lake Michigan, from the Straits of Mackinac to Seul Choix Point, WFM-03 (Figure 1.1). Lake whitefish in the Laurentian Great Lakes have experienced a declining recruitment since the 1990s owing to changes in ice cover and food web dynamics (Ebener *et al.*, 2021). Estimated recruitment in WFM-03 has ranged from 1.5 million individuals during peaks in the 90s and early 00s, to less than half a million in



the past five years (Caroffino and Seider, 2020). Individuals recruit to the fishery at age 4, and on average adults are equally susceptible to environmental and biological drivers of natural mortality over time. Changes in environmental conditions that are not fully understood, notably declines in an important food source, *Diporeia*, have driven declines in lake whitefish growth and condition which are reflected in the age, length, and weight data (Fera *et al.*, 2015).

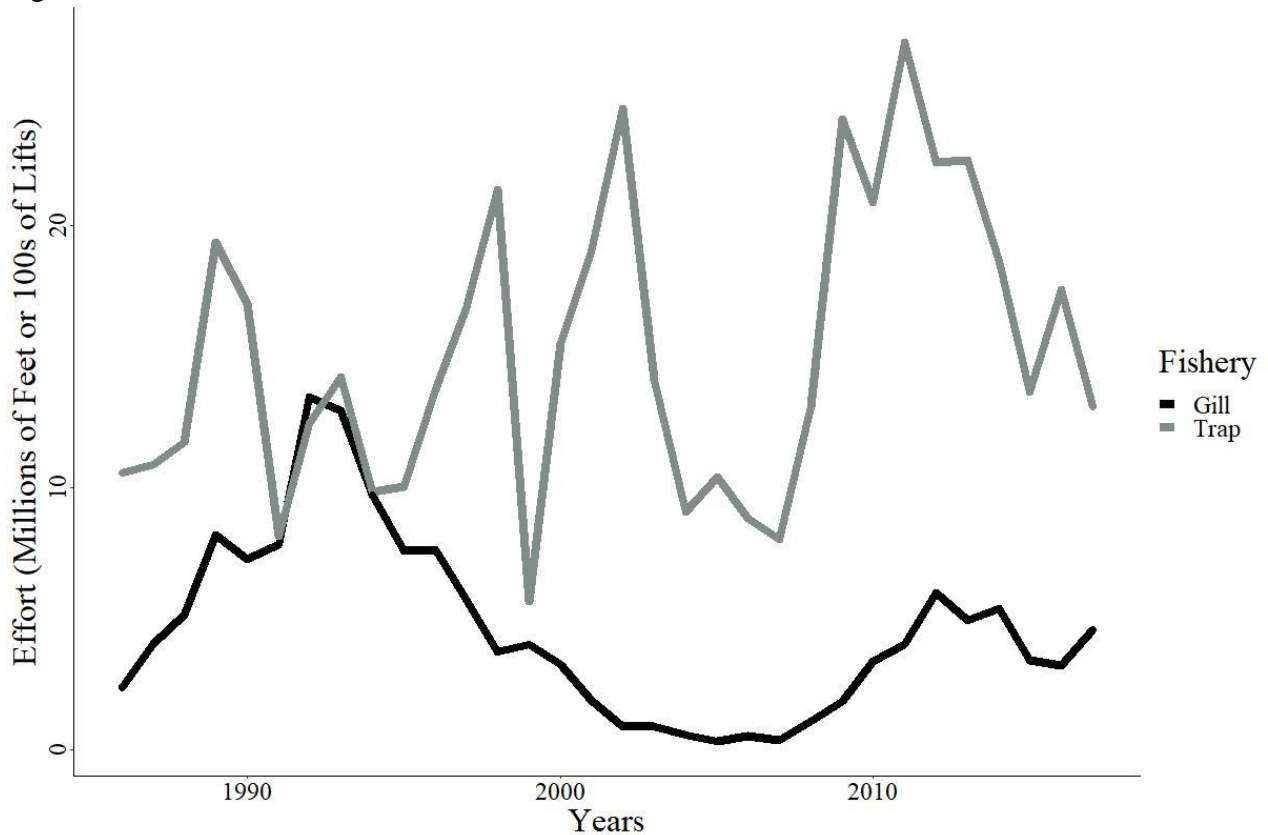
Figure 1.1. Lake whitefish management units of the 1836 Treaty-Ceded Waters of Lakes Superior, Huron, and Michigan, including the management region of interest, WFM-03, in northern Lake Michigan. Reproduced from Caroffino and Barton (2019).



Emigration or immigration were not included in the model, because sub-populations in each management unit tend to reproduce in the same region in which they were born, which has been observed from tagging and genetic studies (Ebener *et al.*, 2010; VanDeHey *et al.*, 2009).

However, some mixing does occur during harvest, so recruitment should be interpreted as both the reproduction rate and the net movement from other stock areas. Two commercial fisheries occur on lake whitefish in WFM-03, a trap net and gill net fishery, both of which are exclusively fished by members of tribes represented by the Chippewa Ottawa Resource Authority (CORA) (Caroffino and Lenart, 2012). Licensed fishers are obligated to report daily weight of landed fish and the amount of gear used (length of gill net or number of trap net lifts), which are aggregated to a single yearly value for the management region (Ebener *et al.*, 2005) (Figure 1.2). The harvest and effort data are not aggregated across both fisheries because each are reported in distinct units and several features make the two fisheries unique. The gears tend to exhibit different age-specific selectivity curves; gill nets are dome shaped whereas the trap nets are believed to be asymptotic- catching more older and larger fish than the gill nets (Zhao and Morbey, 2017). Additionally, the temporal patterns of relative effort of the two fisheries differed considerably over time, often due to changes in the market and management actions, including a gill net conversion program, wherein Michigan exchanged fishers' gill nets for trap nets, which was implemented with the 2000 consent decree (*United States vs. Michigan Consent Decree*, 2000). Lastly, an adjustment was necessary to define effective gill net effort because of changes to the average height of gill net gear over time (Ebener *et al.*, 2005).

Figure 1.2. Fishery effort of the gill net and trap net fisheries on Lake Michigan lake whitefish in region WFM-03 from 1986-2017.



For Lake Michigan lake whitefish, fisheries management is conducted through regulations intended to limit harvest so as to not exceed a constant annual mortality rate of 65% for any age, and to sustain a spawning potential ratio (SPR) of at least 20% (Ebener *et al.*, 2005). SPR is the ratio between the spawning stock biomass per recruit (SSBR) obtained with the current age-specific fishing mortality schedule held constant over the life of a cohort and the spawning stock biomass per recruit given no fishing mortality ( $SSBR_{F=0}$ ). This equates, conditioned on life-history, to a constant fishing mortality rule because in Lake Michigan for lake whitefish natural mortality is assumed constant. The recommended harvest level is calculated by scaling age-specific fishing mortality rates (averaged over the last three years of the assessment) up or down to meet mortality and SPR conditions in projections.

I compared the estimated abundance, recruitment, spawning stock biomass, and instantaneous mortality, between the SCA and SSM, and determined if differences in output would change management metrics. Spawning stock biomass was the product of proportion mature, average weight at spawning, and abundance at age (Eq. 1.8). The percent difference of each year-specific value, and the minimum, the maximum, and average over the whole time series were calculated. The ranges (differences between maximum and minimum values) were compared between the SCA and SSM. The patterns of age- and year- specific catchability of the SSM was compared to the catchability and selectivity in the SCA. Because the SCA model estimates annual catchability and age- and year- specific selectivity, the product of these values was compared to the age- and year- specific catchabilities from the SSM. The annual total mortality in the final three years and the SPR based on those mortality rates were compared against the established limits. Other diagnostic values, like normalized one step ahead (OSA) residuals for the age composition data in both the SCA and SSM models, ordinary least squared (OLS) or “Pearson residuals” for the total catch in the SCA model, OSA residuals for the total catch in the SSM model, and retrospective patterns were examined to compare relative model fits (Appendix B and C) (Hurtado-Ferro *et al.*, 2015; Trijoulet *et al.*, 2023). OSA residuals account for random effects and the correlation between observations in the multinomial distribution and were calculated either externally (for SCA) or internally (for SSM) as described in Trijoulet *et al.*, (2023). OLS residuals were used to evaluate the SCA model’s fit to total catch because these observations are continuous and normally distributed and this model does not have random effects. Mohn’s Rho statistic was used to quantify the degree of retrospective bias in recruitment and spawning stock biomass and compared against the standard acceptable range for short-lived species, -0.22 to 0.3 (Hurtado-Ferro *et al.*, 2015; Mohn, 1999).

## 1.4 Results

The SSM converged on a global minimum of the objective function (Eq. 3.1) and estimated the mean and standard error for the fixed effects: instantaneous natural mortality,  $M$ , the standard deviations of the process variability,  $\sigma_R$  and  $\sigma_G$ , and the degree of correlation in the process variability of the age-specific catchability for each fishery,  $\rho_G$  (Table 1.3). The output of the SSM exhibited similar trends as that of the SCA model despite the differences in model configuration (Figures 1.3-1.5). A quantitative comparison among the average, minimum, maximum, and range of derived quantities revealed a -62 to +53% difference in some output values (Table 1.4).

Table 1.3. Point estimates and asymptotic standard errors of fixed effect parameters in the State-space model (SSM) for Lake Michigan lake whitefish in WFM-03.

Parameter (Symbol)	Estimate	Standard Error
$M$	0.1896	0.0184
$\sigma_R$	0.2474	0.0428
$\sigma_t$	0.3652	0.0442
$\sigma_g$	0.3905	0.0492
$\rho_t$	0.8438	0.0502
$\rho_g$	0.9142	0.0356

Table 1.4. Estimates of derived quantities in the state-space model (SSM) and statistical catch at age (SCA) model for Lake Michigan lake whitefish in WFM-03.

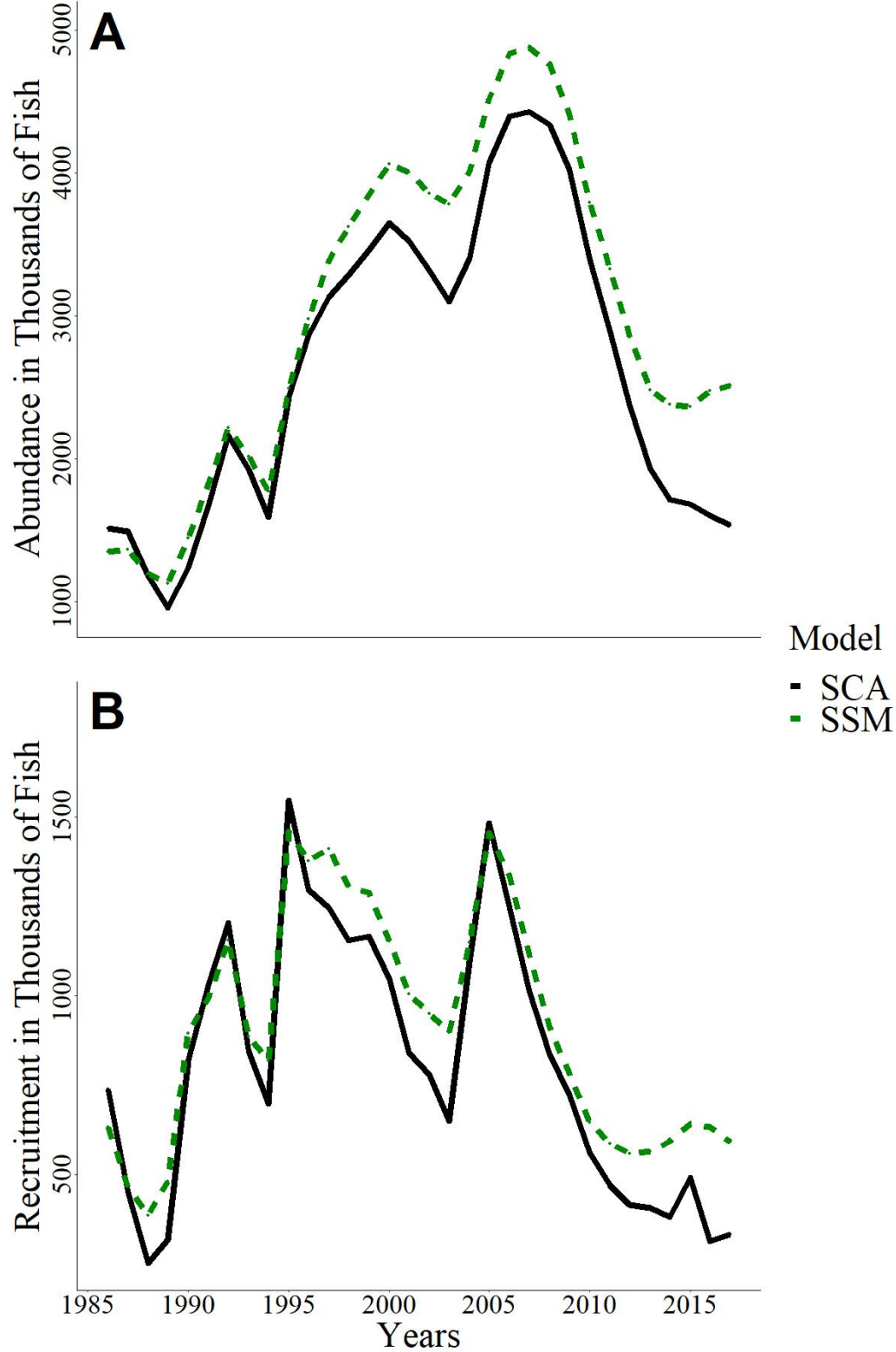
Quantity	SCA Estimate	SSM Estimate	Percent (%) difference
Average Abundance	2,633,377	3,001,214	13.97
Maximum Abundance	4,431,335	4,885,740	10.25
Minimum Abundance	953,555	1,126,510	18.14
Abundance Range	3,477,780	3,759,230	8.10
Average Recruitment	807,254	910,279	12.76
Average Recruitment in Terminal 10 years	492,468	650,837	32.16
Minimum Recruitment	252,249	385,213	52.71
Maximum Abundance	1,545,150	1,460,345	-5.49
Recruitment Range	1,292,901	1,075,132	-16.84
Average Spawning Stock Biomass (SSB) (lbs)	3,931,830	4,489,802	14.19

Table 1.4 (cont'd)

Gill Net Mortality Range	0.691	0.740	7.09
Minimum SSB (lbs)	2,182,870	2,418,414	10.79
Maximum SSB (lbs)	5,985,860	6,542,546	9.30
SSB Range (lbs)	3,802,990	4,124,132	8.44
Average Total Mortality	0.647	0.543	-16.07
Maximum Total Mortality	0.303	0.268	-11.55
Minimum Total Mortality	1.288	1.211	-5.98
Total Mortality Range	0.985	0.942	-4.37
Average Trap Net Mortality	0.251	0.159	-36.65
Maximum Trap Net Mortality	0.060	0.036	-40.00
Minimum Trap Net Mortality	0.485	0.306	-36.91
Trap Net Mortality Range	0.424	0.270	-36.32
Average Gill Net Mortality	0.202	0.195	-3.47
Maximum Gill Net Mortality	0.013	0.005	-61.54
Minimum Gill Net Mortality	0.704	0.745	5.82
Spawning Potential Ratio (SPR)	0.53	0.63	18.87

Total abundance and recruitment (abundance at age 4) followed similar trends across time in the two models (Figure 1.3). The average abundance in the SSM model was 3 million and in the SCA model it was 2.6 million. Estimates of abundance for both models peaked in 2007 and were lowest in 1989 (Figure 1.3A). The range in abundance was smaller in the SSM model. In the SSM the maximum total abundance was 4.3 times larger than the minimum abundance, and in the SCA, the maximum was 4.6 times the minimum. Neither model estimated abundance consistently higher or lower across the entire time series, though the SSM estimates ranged from 11% lower to 64% higher than of those of the SCA model and was larger for most years. In the final 5 years, the SSM estimated a positive trend and the SCA model a negative trend.

Figure 1.3. Estimates of (A) total abundance and (B) recruitment of age 4 individuals of Lake Michigan lake whitefish in WFM-03 from 1986-2017, from the statistical catch at age (SCA) model (black, solid), and the state-space model (SSM) (green, dashed).

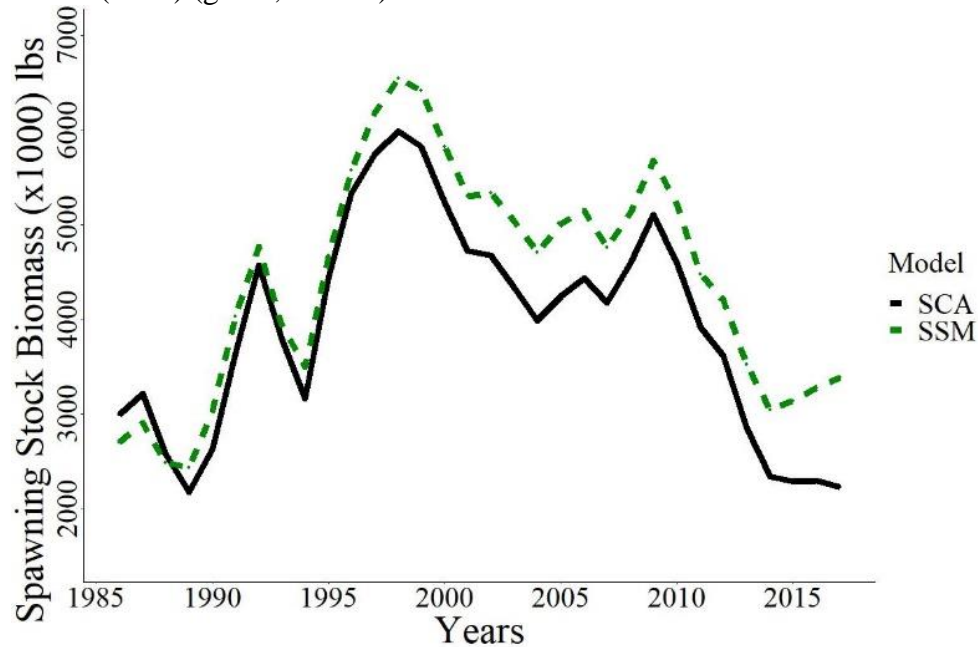


Estimates of recruitment differed more than abundance among the models, though both the SSM and SCA models estimated similar trends (Figure 1.3B). Estimated recruitment for both models were greatest in 1995 and lowest in 1988. The range in recruitment was smaller in the SSM. In the SSM the maximum estimated recruitment was 3.8 times larger than the minimum estimate and in the SCA model, the maximum was 6.1 times larger than the minimum. Neither model consistently estimated higher recruitment than the other. Recruitment estimates in the SSM were 14% lower to 103% higher than those of the SCA model (at the very beginning and penultimate point of the time series, respectively). Mean recruitment was 13% larger in the SSM (807,000 versus 889,000, for SCA and SSM respectively) and 32% larger in just the final 10 years (492,000 versus 633,000, for SCA and SSM respectively).

Both models estimated similar trends in spawning stock biomass (SSB), except for near the end of the time series, when the SSM estimates rose and the SCA estimates fell (Figure 1.4). In both models, the maximum SSB occurred in 1998 and the minimum in 1989, and in both the maximum value was approximately 2.7 times larger than the minimum. SSM estimates of SSB ranged from 10% lower to 59% higher than those of the SCA model, and the average difference was 15%. The mean SSB was 3.9 million lbs. for the SSM and 4.5 million lbs. for the SCA. For both models the trends in SSB were different from trends in abundance and recruitment due to temporal patterns in mortality and weight at age.

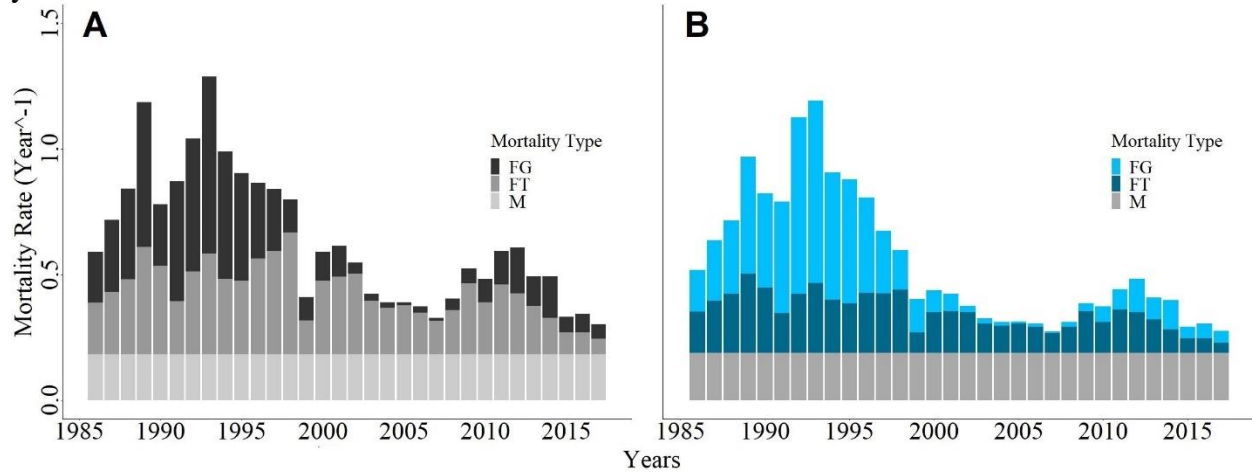


Figure 1.4. Estimates of spawning stock biomass of Lake Michigan lake whitefish in WFM-03 from 1986-2017, from the statistical catch at age (SCA) model (black, solid), and the state-space model (SSM) (green, dashed).



In both models, the total instantaneous mortality rate initially increased in the first third of the time series, then gradually declined (Figure 1.5). The SCA model estimated higher mortality than the SSM for most of the time series, reaching a maximum in 1993 at an instantaneous total mortality rate of  $1.288 \text{ yr}^{-1}$  (averaged across ages 4-15+) (Figure 1.5A). The SSM reached a maximum total mortality in the same year but at  $1.211 \text{ yr}^{-1}$  (Figure 1.5B). In only 1990 and 1992 did the SSM estimate a larger total instantaneous mortality rate than the SCA model, by 3% and 6%, respectively. The average total instantaneous mortality rate was  $0.54 \text{ yr}^{-1}$  in the SSM and  $0.64 \text{ yr}^{-1}$  in the SCA. The SSM estimated a yearly instantaneous trap net fishing mortality rate 15-54% lower than the SCA model. The yearly instantaneous gill net fishing mortality rate in the SSM ranged from 49% larger to 57% lower than the SCA model, though it was lower for most of the time series (24 out of 32 years). Time-invariant instantaneous natural mortality rate was approximately the same in both models-  $0.190 \pm 0.018 \text{ yr}^{-1}$  and  $0.184 \pm 0.017 \text{ yr}^{-1}$  in the SSM and SCA, respectively.

Figure 1.5. Estimated instantaneous fishing mortality rate of the gill net fishery (FG), the trap net fishery (FT) and the natural mortality rate (M) over time for (A) the statistical catch at age model (SCA) and (B) the state-space model (SSM). The instantaneous fishing mortality rate was averaged across ages 4-15+ for each year, and instantaneous natural mortality rate was age- and year- invariant.



In the SSM the catchabilities provide estimates of the age- and year- specific proportionality between instantaneous fishing mortality rates and fishing effort. These fishery-specific relationships are presented in two ways- as catchability over years for each age (Figure 1.6B, 1.6D) and as catchability over ages for each year, but normalized so that the maximum value is 1, as is typical for selectivity (Figure 1.7B, 1.7D). The estimated degree of correlation in the AR(1) random walk for catchability ( $\pm$  95% confidence intervals) was  $0.843 \pm 0.050$  for the trap net fishery and  $0.914 \pm 0.036$  for the gill net fishery (Table 1.3). For the SCA, the product of year-specific catchability and age- and year- specific selectivity has the same meaning as SSM catchabilities. I calculated these derived age- and year- specific catchabilities for the SCA and present them in the same way to compare how models quantify the relationship between fishing mortality and fishing effort (Figures 1.6A, 1.6C, 1.7A, 1.7C).

Figure 1.6. Estimated catchability over years for each age (on scale from red to purple from age 4 to 15+) of the (A) gill net and (C) trap net fishery in the statistical catch at age (SCA) model, and the (B) gill net and (D) trap net fishery in the state-space model (SSM).

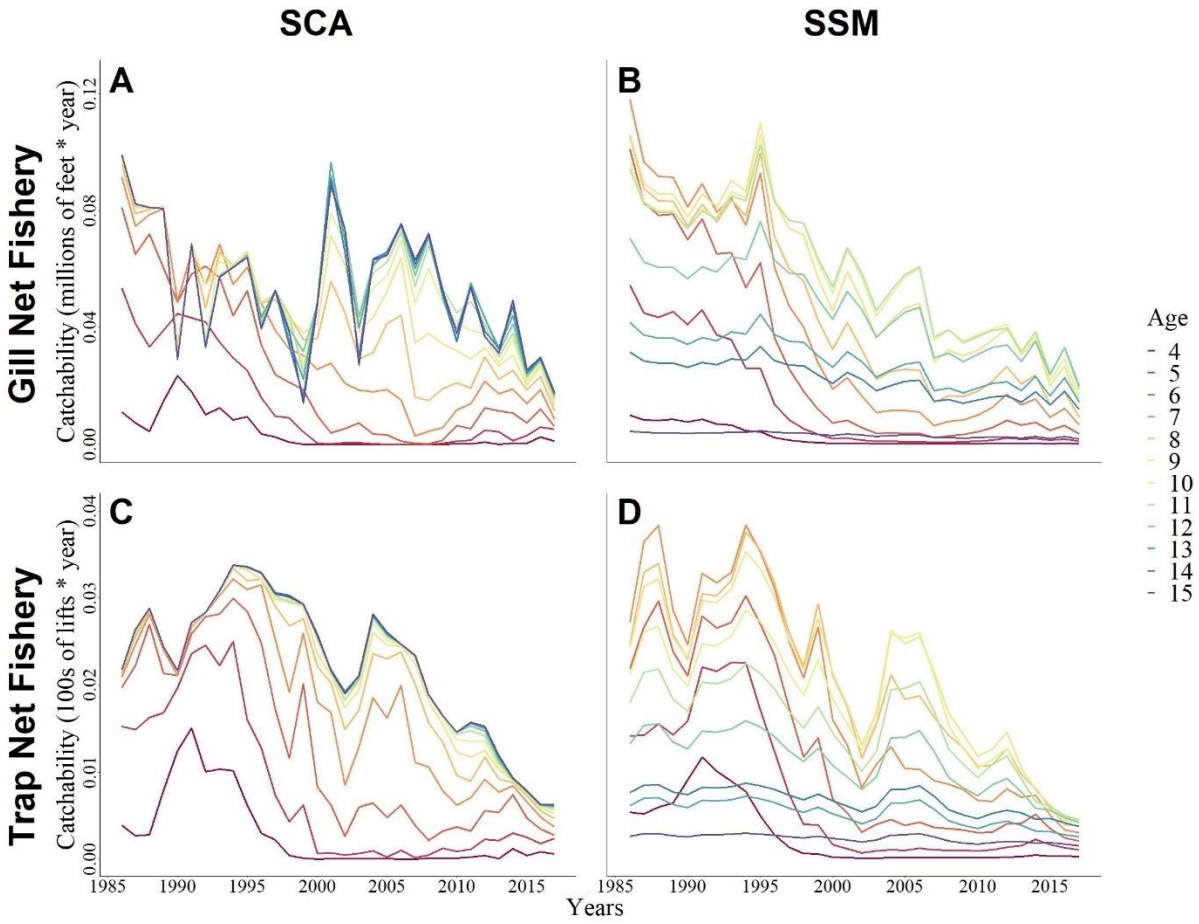
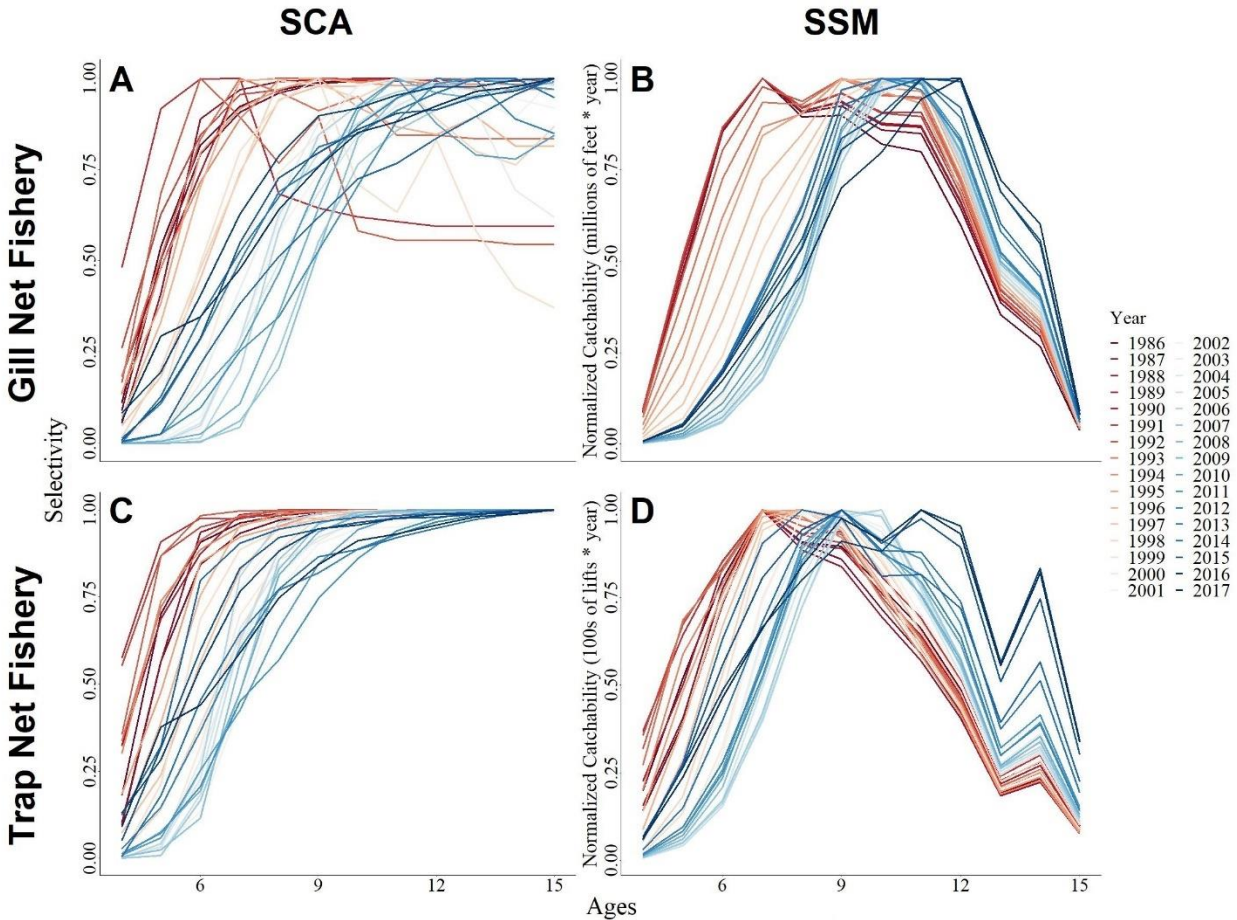


Figure 1.7. Estimated selectivity over ages for each year (on scale from red for early years to blue for later years) of the (A) gill net and (C) trap net fishery in the statistical catch at age (SCA) model, and of the (B) gill net and (D) trap net fishery in the state-space model (SSM). Selectivity was obtained by normalizing age-specific catchability such that the maximum value for each year was 1.



Catchability at age was lower later in the time series than near the beginning across both models and both fisheries, but there were some marked differences in patterns between SSM and SCA (Figures 1.6). Catchability of the gill net fishery steadily declined after 1995 in the SSM but peaked mid-way through the time series, in 2001, in the SCA (contrast Figures 1.6A and 1.6B). Catchability of the trap net fishery had similar trends for both models, with peaks occurring in 1994 and 2004-2006 for both the SCA and SSM (contrast Figures 1.6C and 1.6D). The selectivity for both fisheries was dome shaped in the SSM, and largely asymptotic in the SCA (Figure 1.7). The one notable deviation from this pattern was the gill net fishery in the SCA

model, where some dome-like patterns occurred in the initial 10 years of the time series with a modest decline in selectivity at older ages. This pattern can be explained because selectivity in the SCA model was a function of mean length at age, and because fish got smaller on average through the time series, the old fish were as small late in the time series as the young highly selected fish earlier in the time series. In general, selectivity peaked at younger ages earlier in the time series for both models and both fisheries, and the age of full selectivity became progressively older in later years.

Management reference points differed between the SCA and SSM though not to the extent that differences in recommended harvest limits would have differentially impinged on fishery operations. In the last three years of the SCA and SSM assessments, the maximum (over ages) annual total mortality was 31% and 25%, and the SPR based on these mortality rates was 0.53 and 0.63, respectively. Thus, mortality rates were estimated to be below 65% and SPR above 0.2. In short, harvest limits based on either assessment would correspond to higher than status quo levels of fishing mortality.

Residuals for the fit to log scale total catch for both fisheries and both models were centered on zero, approximately normal, and generally did not trend through time, although the residuals for catch from the SSM model did exhibit more skew than the SCA model (Figures 1B.1-1B.3, 1C.2-1C.4). Residuals for the age compositions for both fleets from the SCA model were patterned for several cohorts, especially in the first half of the time series (Figures 1B.4-1B.5). In the SSM model, the model overestimated the proportion of older ages in the trap net fishery and underestimated the proportion of individuals in the gill net fishery at the beginning of the time series, but the cohort effects were generally resolved relative to the SCA model (Figures 1C.5-1C.6). Both models exhibited retrospective patterns in recruitment, though not spawning

stock biomass (Figures 1B.6-1B.7, 1C.7-1C.8). The Mohn’s Rho for recruitment was 0.989 for the SCA and 0.448 for the SSM, which both suggest a systemic bias in recruitment, though less so in the SSM. The Mohn’s Rho for spawning stock biomass was marginally larger and in the opposite direction in the SSM model (0.092) compared to the SCA (-0.038). The retrospective bias on SSB was within the acceptable range of -0.22 to 0.3 suggested for short lived species.

Sensitivity tests demonstrated that fixing the standard deviation of the lognormal distribution of the total catch at values 50% above or below the baseline did not change most estimated fixed effect parameters by more than 1.5% (Table 1.5). The estimated standard deviation of the trap net and gill net catchability process variability decreased by 2.5% and 8% respectively, when the observation error was increased, suggesting a tradeoff between these values. For a full description of the model diagnostics of SSM, including time series of one step ahead (OSA) residuals and retrospective analyses, see Appendix 1C.

Table 1.5. Results of sensitivity analysis. The maximum likelihood estimates of fixed effect parameters are reported for each scenario as well as the percent (%) difference relative to the baseline.

Scenario	$M$	% Dif	$\sigma_R$	% Dif	$\sigma_t$	% Dif	$\sigma_g$	% Dif	$\rho_t$	% Dif	$\rho_g$	% Dif
$\sigma_{CG} = 0.067$ (Baseline)	0.1896		0.247		0.365		0.391		0.844		0.914	
$\sigma_{CG} = 0.034$	0.1899	0.18	0.248	0.09	0.370	1.2	0.403	3.26	0.847	0.32	0.920	0.58
$\sigma_{CG} = 0.100$	0.1896	-0.17	0.247	-0.2	0.360	-2.5	0.371	-7.96	0.841	-0.7	0.905	-1.55

## 1.5 Discussion

This work demonstrated the feasibility of applying a state-space stock assessment model (SSM) with age-specific catchability as a state variable akin to year- and age- specific fishing mortality in previous applications (Berg and Nielsen, 2016; Nielsen and Berg, 2014). This also confirmed the possibility of applying SSM in cases where there are no fisheries independent

indices of abundance. To capitalize on state-space modeling's capacity to estimate several sources of process variability, the existing SCA model had to be re-parameterized such that recruitment and age-specific catchability were random walk processes. Merely specifying that yearly variability in catchability, recruitment, or selectivity were random effects, without changing the parameterization of the SCA model resulted in model non-convergence.

These necessary changes to the process model parameterization, though increasing realism and model flexibility, inevitably led to changes in model output that cannot be parsed from changes due only to converting a statistical catch at age model (SCA) to an SSM. Some changes in output between the SCA and SSM, like trends in abundance and fishing mortality, were minimal. Others, like recruitment in the latter part of the time series, and the catchability and selectivity patterns, substantially altered the interpretation of population and fisheries dynamics.

Estimates of time-varying recruitment from the SSM exhibited less interannual variation and less difference between the maximum and minimum value than the SCA model, which reflects autocorrelation in the random walk process (Thorson *et al.*, 2014). During much of the assessment, recruitment can be strongly informed by how a cohort was represented in subsequent years of catch data which is why estimates at the end of the assessment period are generally the most uncertain (Brooks and Legault, 2016). In the absence of future information, the SCA model was guided by the explicit relationship between recruitment and spawning stock biomass (SSB) in past years while SSM, lacking such a function, remained largely the same in the final 5 years (Figure 3B). Ideally, the Lake Michigan lake whitefish model should have enough flexibility to account for the relationship between recruitment and SSB while also incorporating how ecosystem changes like ice cover, prey availability, and habitat degradation due to dreissenid

invasion may create additional yearly variability (Ebener *et al.*, 2021). The solution may lie in one or several alternative recruitment structures including 1) parameterizing variable parameters of the stock recruitment curve, 2) employing more dynamic autocorrelation processes with time series models that estimate how much to “remember” from previous years (e.g., AR(1)), or 3) utilizing a “mixture-distribution” of two distributions controlled via a Bernoulli distribution that accommodates occasional spikes or dips in recruitment (Johnson *et al.*, 2016; Maunder and Thorson, 2018; Thorson *et al.*, 2014).

The SSM estimated dome-shaped selectivity for both fisheries, so it predicted older (and thus larger) individuals were present but less vulnerable to capture. The SCA was constrained to fit an asymptotic selectivity curve for the trap net fishery such that the oldest fish were catchable at a comparable rate to middle-aged fish. An asymptote was estimated for most years for the gill net fishery as well. Sensitivity tests of the SSM which forced an asymptote for the trap net fishery confirmed that differences in the estimated selectivity curve drove the differences in abundance, and by extension, spawning stock biomass. Both asymptotic and dome-shaped selectivity have been previously reported for trap net gear and either relationship could be reasonably expected for Lake Michigan lake whitefish in WFM-03 (Dunlop *et al.*, 2018; Hansen *et al.*, 2008; Jeong *et al.*, 2000). The SSM had limited data to inform selectivity because the average age of lake whitefish in this region is lower than in adjacent regions and older ages were not strongly represented in the harvest (Caroffino and Seider, 2020). This flexibility to estimate time-varying selectivity without specifying an explicit and somewhat arbitrary function, is a hallmark of state-space stock assessment models (Nielsen *et al.*, 2021). However, if there is good reason to specify or restrain it, there are several future options to do so. The functional form could be implemented as a “prior” about which penalties are assessed, rather than requiring an



exact match. This approach could be implemented with the selectivity parameters fixed, changing gradually over time (perhaps by an autoregressive process), or changing at discrete time blocks (Cronin-Fine and Punt, 2021; Martell and Stewart, 2014; Stock and Miller, 2021). Future lake whitefish assessments might consider a thorough model selection and review procedure, with dome-shaped selectivity for the trap net fishery included as a possible interpretation of fishery dynamics.

The simultaneous estimation of observation and process error variances is a touted hallmark of state-space modeling but was not achieved in this model. This inability to estimate the observation error variance within the model may be caused by the estimation of variance of time-varying catchability. Sensitivity tests demonstrated a tradeoff between observation error standard deviation,  $\sigma_{CG}$ , and catchability standard deviation,  $\sigma_G$ , because when the former was fixed at a higher value, the latter was estimated at a lower value. The observation error standard deviation may have been estimable in this model if there were greater contrast in fishing mortality and catch over time, or if there were informative survey data. Observation error standard deviation can be approximated from the observed data, so estimating the standard deviation of catchability, an unobservable state variable, was prioritized in this model, a practice also used by others (Francis, 2011). The age composition data were fit using a multinomial distribution for which the effective sample sizes were specified *a priori*. Efforts to calculate effective sample sizes through established iterative reweighting techniques led to data overweighting (i.e. the effective sample sizes were orders of magnitude *larger* than observed sample size), fitting the age proportion data precisely at the expense of high process model variability (Francis, 2011; Truesdell et al., 2017). Estimability of observation error variance may be beyond the upper limits of model flexibility for this SSM given the process model

parameterization and data availability. Limitations on estimating the observation error variance have not been evident in other state-space models that included index data and assumed constant catchability (Nielsen and Berg, 2014; Perreault *et al.*, 2020). Future research could investigate the reasons for this difference. In particular, it is an open question whether estimation of observation error variance and time-varying catchability is possible with different distributions of the catch data (e.g., Dirichlet multinomial for proportions at age or a multivariate lognormal for catch at age (Nielsen and Berg, 2014; Thorson *et al.*, 2017) or if provision of external estimates of observation error variance is a fundamental requirement when allowing time-varying catchability. If it is a fundamental property, providing external estimates of observation error variance in order to allow for time-varying fishery catchability seems preferable.

Though fisheries-dependent sampling is non-random and prone to bias, its continued use may be justified by its extensive availability, inexpensiveness, and the capabilities of modern assessment models to account for violations of the typical assumption of proportionality with abundance. This state-space stock assessment model uses variability in the proportionality coefficient, catchability, to account for a stochastic relationship between CPUE and abundance across age and time, combatting the concerns of hyperstability or hyperdepletion that surround the use of effort data (Rose and Kulka, 1999). Other research has explored standardization methods using generalized linear models (GLM) or generalized additive models (GAM) to explain how factors like area, time, and gear can influence the relationship between abundance and CPUE in a non-linear fashion (Deroba and Bence, 2009; Ducharme-Barth *et al.*, 2022; Grüss *et al.*, 2019). Specific covariates such as season, sampling depth, or hook and bait characteristics can be directly included in the model to explain variation in catchability (Grüss *et al.*, 2019; Johnson *et al.*, 2019). Even with fine scale catch data and informative covariates, specifying

catchability as a time-varying random effect can allow for temporal variability not explained by the covariates, without overfitting the model.

Use of this SSM in future cases may challenge preconceived assumptions about population or fishery dynamics that were assumed impossible, unrealistic, or inestimable given the available data and limits to parameterization. The SSM structure easily lends itself to estimating process variation in natural mortality, selectivity, abundance, and many other factors previously assumed to be invariant or fixed. This research demonstrated that a lack of fisheries independent indices is not a hurdle to employing SSM. Many more fisheries may consider this approach a viable option, maintaining that there is sufficient informative effort data to relate catchability to fishing mortality. Highly migratory species like tunas, billfish, and sharks, data-poor species like elasmobranchs, deep sea fisheries, and developing fisheries are examples of those that have limited to no fisheries independent data, or the continued collection of such data may not be cost effective and may therefore benefit from this SSM approach (Costello *et al.*, 2012; Dennis *et al.*, 2015; Langley *et al.*, 2009; Lynch *et al.*, 2018; Victorero *et al.*, 2018). Several Great Lakes stocks which have been previously assessed using a penalized likelihood statistical catch-at-age model may also be candidates for fitting a state-space stock assessment model, especially one such as this which can accommodate fisheries-dependent data. For example, yellow perch in Lake Michigan are assessed using recreational and commercial fisheries effort and time-varying catchability, though these models include informative fisheries-independent data (Wilberg *et al.*, 2005). As another example, Chinook salmon stocks in Lakes Michigan and Huron lack reliable survey information and abundance is estimated entirely with catch and effort data (Brenden *et al.*, 2012; Clark *et al.*, 2016). Likewise, consideration of fishery dependent CPUE information has been suggested for use alongside fishery independent

information when stakeholder perceptions of stock status are inconsistent with a stock assessment, as in Gulf of Maine and Georges Bank Cod (NEFSC, 2022a). While a lack of fishery independent data may create new challenges, such as the inability to estimate observation error variance, they are not insurmountable, and many benefits of state-space modelling can still be achieved.

## REFERENCES

- Aanes, S., Aeberhard, W.H., Albertsen, C.M., Aldrin, M., Babyn, J., Berg, C.W., Breivik, O., Cook, R., Flemming, J.M., Fryer, R., Liljestr nd, E.M., Millar, C., Miller, T.J., Minto, C., Newman, K.B., Nielsen, A., Perreault, A., Regular, P., Skaug, H.J., Spence, M., Trijoulet, V., Vatnehol, S., 2020. Workshop on the review and future of state space stock assessment models in ICES (WKRFSAM). *ICES Sci. Rep.* 2:32. 1-23. <http://doi.org/10.17895/ices.pub.6004>.
- Aeberhard, W.H., Flemming, J.M., Nielsen, A., 2018. Review of state-space models for fisheries science. *Annu. Rev. Stat. Appl.* 5, 215-35. <https://doi.org/10.1146/annurev-statistics-031017-100427>.
- Berg, C.W., Nielsen, A., 2016. Accounting for correlated observations in an age-based state-space stock assessment model. *ICES J. Mar. Sci.* 73, 1788-1797. <https://doi.org/10.1093/icesjms/fsw046>.
- Brenden, T.O., Bence, J.R., Szalai, E.B., 2012. An age-structured integrated assessment of chinook salmon population dynamics in Lake Huron's main basin since 1968. *T. Am. Fish. Soc.* 141, 919-933. <https://doi.org/10.1080/00028487.2012.675910>.
- Brooks, E.N., Legault, C.M., 2016. Retrospective forecasting - evaluating performance of stock projections for New England groundfish stocks. *Can. J. Fish. Aquat. Sci.* 73, 935-950. <https://doi.org/10.1139/cjfas-2015-0163>.
- Cadigan, N.G., 2015. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. *Can. J. Fish. Aquat. Sci.* 308, 1-13. <https://doi.org/10.1139/cjfas-2015-0047>.
- Caroffino, D.C., Lenart, S.J., 2012. Executive summary in Caroffino, D.C. and Lenart, S.J., eds. Technical fisheries committee administrative report 2012: Status of lake trout and lake whitefish populations in the 1836 treaty-ceded waters of Lakes Superior, Huron and Michigan, with recommended yield and effort levels for 2012. <http://www.michigan.gov/greatlakesconsentdecree>.
- Caroffino, D.C., Seider, M.J., 2020. Executive Summary in Caroffino, D.C. and Seider, M.J., eds. Technical fisheries committee administrative report 2020: Status of lake trout and lake whitefish populations in the 1836 treaty-ceded waters of Lakes Superior, Huron and Michigan, with recommended yield and effort levels for 2020. <https://www.michigan.gov/greatlakesconsentdecree>.
- Clark, R.D., Bence, J.R., Claramunt, R.M., Johnson, J.E., Gonder, D., Legler, N.D., Robillard, S.R., Dickinson, B.D., 2016. A spatially explicit assessment of changes in chinook salmon fisheries in Lakes Michigan and Huron from 1986 to 2011. *N. Am. J. Fish. Manage.* 36, 1068-1083. <https://doi.org/10.1080/02755947.2016.1185060>.

- Costello, C., Ovando, D., Hilborn, R., Gaines, S.D., Deschenes, O., Lester, S.E., 2012. Status and solutions for the world's unassessed fisheries. *Science* 338, 517-520. <https://doi.org/10.1126/science.1223389>.
- Cronin-Fine, L., Punt, A.E., 2021. Modeling time-varying selectivity in size-structured assessment models. *Fish. Res.* 239, 105927. <https://doi.org/10.1016/j.fishres.2021.105927>.
- de Valpine, P., Hilborn, R., 2005. State-space likelihoods for nonlinear fisheries time-series. *Can. J. Fish. Aquat. Sci.* 62, 1937-1952. <https://doi.org/10.1139/f05-116>.
- Dennis, D., Plagányi, É., Van Putten, I., Hutton, T., Pascoe, S., 2015. Cost benefit of fishery-independent surveys: Are they worth the money? *Mar. Policy* 58, 108-115. <https://doi.org/10.1016/j.marpol.2015.04.016>.
- Deriso, R.B., Maunder, M.N., Skalski, J.R., 2007. Variance estimation in integrated assessment models and its importance for hypothesis testing. *Can. J. Fish. Aquat. Sci.* 64, 187-197. <https://doi.org/10.1139/f06-178>.
- Deroba, J.J., Bence, J.R., 2009. Developing model-based indices of Lake Whitefish abundance using commercial fishery catch and effort data in Lakes Huron, Michigan, and Superior. *N. Am. J. Fish. Manage.* 29, 50-63. <https://doi.org/10.1577/m07-205.1>.
- Ducharme-Barth, N.D., Grüss, A., Vincent, M.T., Kiyofuji, H., Aoki, Y., Pilling, G., Hampton, J., Thorson, J.T., 2022. Impacts of fisheries-dependent spatial sampling patterns on catch-per-unit-effort standardization: A simulation study and fishery application. *Fish. Res.* 246, 106169. <https://doi.org/10.1016/j.fishres.2021.106169>.
- Dunlop, E.S., Feiner, Z.S., Höök, T.O., 2018. Potential for fisheries-induced evolution in the Laurentian Great Lakes. *J. Great Lakes Res.* 44, 735-747. <https://doi.org/10.1016/j.jglr.2018.05.009>.
- Ebener, M.P., Bence, J.R., Newman, K., Schneeberger, P., 2005. Application of statistical catch-at-age models to assess lake whitefish stocks in the 1836 treaty-ceded waters of the upper Great Lakes. In: Mohr, L.C., Nalepa, T.F. (Eds.), *Proceedings of a Workshop on the Dynamics of Lake Whitefish and the Amphipod Diporeia spp. in the Great Lakes Great Lakes Fishery Commission Technical Report 66*, pp. 271-309.
- Ebener, M.P., Brenden, T.O., Wright, G.M., Jones, M.L., Faisal, M., 2010. Spatial and temporal distributions of lake whitefish spawning stocks in Northern lakes Michigan and Huron, 2003-2008. *J. Great Lakes Res.* 36, 38-51. <https://doi.org/10.1016/j.jglr.2010.02.002>.
- Ebener, M.P., Dunlop, E.S., Muir, A.M., 2021. Declining recruitment of lake whitefish recruitment to fisheries in the Laurentian Great Lakes: Management considerations and research priorities. <http://www.glfc.org/pubs/misc/2021-01.pdf>.

- Fera, S.A., Rennie, M.D., Dunlop, E.S., 2015. Cross-basin analysis of long-term trends in the growth of lake whitefish in the Laurentian Great Lakes. *J. Great Lakes R.* 41, 1138-1149. <https://doi.org/10.1016/j.jglr.2015.08.010>.
- Fournier, D.A., Archibald, C.P., 1982. A general theory for analyzing catch at age data. *Can. J. Fish. Aquat. Sci.* 39, 1195-1207. <https://doi.org/10.1139/f82-157>.
- Fournier, D.A., Hampton, J., Sibert, J.R., 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore. *Can. J. Fish. Aquat. Sci.* 55, 2105-2116. <https://doi.org/10.1139/f98-100>.
- Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J.N., Magnusson, A., Maunder, M.N., Nielsen, A., Sibert, J., 2012. AD Model Builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optim. Method Softw.* 27, 233-249. <https://doi.org/10.1080/10556788.2011.597854>.
- Francis, R.I.C.C., 2011. Corrigendum: Data weighting in statistical fisheries stock assessment models. *Can. J. Fish. Aquat. Sci.* 68, 2228-2228. <https://doi.org/10.1139/f2011-165>.
- Grüss, A., Walter, J.F., Babcock, E.A., Forrestal, F.C., Thorson, J.T., Laretta, M.V., Schirripa, M.J., 2019. Evaluation of the impacts of different treatments of spatio-temporal variation in catch-per-unit-effort standardization models. *Fish. Res.* 213, 75-93. <https://doi.org/10.1016/j.fishres.2019.01.008>.
- Gunnlaugsson, T., 2012. Some catch-at-age analysis methods and models compared on simulated data. *Open J. Mar. Sci.* 2, 16-24. <https://doi.org/10.4236/ojms.2012.21003>.
- Hansen, M.J., Horner, N.J., Liter, M., Peterson, M.P., Maiolie, M.A., 2008. Dynamics of an increasing lake trout population in Lake Pend Oreille, Idaho. *N. Am. J. Fish. Manage.* 28, 1160-1171. <https://doi.org/10.1577/M07-149.1>.
- He, J.X., Bence, J.R., 2007. Modeling annual growth variation using a hierarchical Bayesian approach and the von Bertalanffy growth function, with application to lake trout in southern Lake Huron. *T. Am. Fish. Soc.* 136, 318-330. <https://doi.org/10.1577/t06-108.1>.
- Herbst, S.J., Marsden, J.E., 2011. Comparison of precision and bias of scale, fin ray, and otolith age estimates for lake whitefish (*Coregonus clupeaformis*) in Lake Champlain. *J. Great Lakes R.* 37, 386-389. <https://doi.org/10.1016/j.jglr.2011.02.001>.
- Hilborn, R., Walters, C.J., 1992. Quantitative fisheries stock assessment: choice, dynamics, and uncertainty. Chapman and Hall, New York, NY.
- Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R.R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. *ICES J. Mar. Sci.* 72, 99-110. <https://doi.org/10.1093/icesjms/fsu198>.

- Jeong, E.C., Park, C.D., Park, S.W., Lee, J.H., Tokai, T., 2000. Size selectivity of trap for male red queen crab *Chionoecetes japonicus* with the extended SELECT model. *Fish. Res.* 66, 494-501. <https://doi.org/10.1046/j.1444-2906.2000.00079.x>.
- Johnson, K.F., Councill, E., Thorson, J.T., Brooks, E., Methot, R.D., Punt, A.E., 2016. Can autocorrelated recruitment be estimated using integrated assessment models and how does it affect population forecasts? *Fish. Res.* 183, 222-232. <https://doi.org/10.1016/j.fishres.2016.06.004>.
- Johnson, K.F., Thorson, J.T., Punt, A.E., 2019. Investigating the value of including depth during spatiotemporal index standardization. *Fish. Res.* 216, 126-137. <https://doi.org/10.1016/j.fishres.2019.04.004>.
- Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., Bell, B.M., 2016. TMB: Automatic differentiation and laplace approximation. *J. Stat. Softw.* 70, 1-21. <https://doi.org/10.18637/jss.v070.i05>.
- Langley, A., Harley, S., Hoyle, S., Davies, N., Hampton, J., Kleiber, P., 2009. Stock assessment of yellowfin tuna in the western and central Pacific Ocean. WCPFC SC5 SA WP-3, Port Vila, Vanuatu, 10–21.
- Legault, C.M., Restrepo, V.R., 1998. A flexible forward age-structured assessment program. *Collect. Vol. Sci. Pap. ICCAT*, 49, 246-253.
- Lynch, P.D., Shertzer, K.W., Cortés, E., Latour, R.J., 2018. Abundance trends of highly migratory species in the Atlantic Ocean: accounting for water temperature profiles. *ICES J. Mar. Sci.* 75, 1427-1438. <https://doi.org/10.1093/icesjms/fsy008>.
- Martell, S., Stewart, I., 2014. Towards defining good practices for modeling time-varying selectivity. *Fish. Res.* 158, 84-95. <https://doi.org/10.1016/j.fishres.2013.11.001>.
- Maunder, M.N., Thorson, J.T., 2018. Modeling temporal variation in recruitment in fisheries stock assessment: A review of theory and practice. *Fish. Res.* 217, 71-86. <https://doi.org/10.1016/j.fishres.2018.12.014>.
- Methot, R.D., Wetzel, C.R., 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fish. Res.* 142, 86-99. <https://doi.org/10.1016/j.fishres.2012.10.012>.
- Mohn, R., 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. *ICES J. Mar. Sci.* 56, 473-488. <https://doi.org/10.1006/jmsc.1999.0481>.
- [NEFSC] Northeast Fisheries Science Center. 2022. Management Track Assessments Fall 2021. US Dept Commer, Northeast Fish. Sci. Cent. Ref. Doc. 22-07, 53 p.



- Nielsen, A., Berg, C.W., 2014. Estimation of time-varying selectivity in stock assessments using state-space models. *Fish. Res.* 158, 96-101. <https://doi.org/10.1016/j.fishres.2014.01.014>.
- Nielsen, A., Berg, C.W., Hintzen, N.T., Mosegaard, H., Trijoulet, V., 2021. Multi-fleet state-space assessment model strengthens confidence in single-fleet SAM and provides fleet-specific forecast options. *ICES J. Mar. Sci.* 78, 2043-2052. <https://doi.org/10.1093/icesjms/fsab078>.
- Patriarche, M.H., 1977. Biological basis for management of lake whitefish in the Michigan waters of northern Lake Michigan. *T. Am. Fish. Soc.* 106, 295-308. [https://doi.org/10.1577/1548-8659\(1977\)106<295:bbfmol>2.0.co;2](https://doi.org/10.1577/1548-8659(1977)106<295:bbfmol>2.0.co;2)
- Perreault, A.M.J., Wheeland, L.J., Joanne Morgan, M., Cadigan, N.G., 2020. A state-space stock assessment model for American plaice on the Grand Bank of Newfoundland. *J. Northwest Atl. Fish. Sci.* 51, 45-104. <https://doi.org/10.2960/J.v51.m727>.
- Quirijns, F.J., Poos, J.J., Rijnsdorp, A.D., 2008. Standardizing commercial CPUE data in monitoring stock dynamics: Accounting for targeting behaviour in mixed fisheries. *Fish. Res.* 89, 1-8. <https://doi.org/10.1016/j.fishres.2007.08.016>
- Richards, L.J., Schnute, J.T., Olsen, N., 1997. Visualizing catch–age analysis: a case study. *Can. J. Fish. Aquat. Sci.* 54, 1636-1658. <https://doi.org/10.1139/f97-073>.
- Ricker, W.E., 1975. Computation and interpretation of biological statistics of fish populations. *Bull. Fish. Res. Board. Can.* 191, 1-382.
- Rose, G.A., Kulka, D.W., 1999. Hyperaggregation of fish and fisheries: how catch-per-unit-effort increased as the northern cod (*Gadus morhua*) declined. *Can. J. Fish. Aquat. Sci.* 56, 118-127. <https://doi.org/10.1139/f99-207>.
- Schnute, J.T., 1994. A general framework for developing sequential fisheries models. *Can. J. Fish. Aquat. Sci.* 51, 1676–1688. <https://doi.org/10.1139/f94-168>.
- Stock, B.C., Miller, T.J., 2021. The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time- and age- varying processes via random effects and links to environmental covariates. *Fish. Res.* 240, 105967. <https://doi.org/10.1016/j.fishres.2021.105967>.
- Stock, B.C., Xu, H., Miller, T.J., Thorson, J.T., Nye, J.A., 2021. Implementing two-dimensional autocorrelation in either survival or natural mortality improves a state-space assessment model for Southern New England-Mid Atlantic yellowtail flounder. *Fish. Res.* 237, 105873. <https://doi.org/10.1016/j.fishres.2021.105873>.
- Thorson, J.T., Jensen, O.P., Zipkin, E.F., 2014. How variable is recruitment for exploited marine fishes? A hierarchical model for testing life history theory. *Can. J. Fish. Aquat. Sci.* 71, 973-983. <https://doi.org/10.1139/cjfas-2013-0645>.

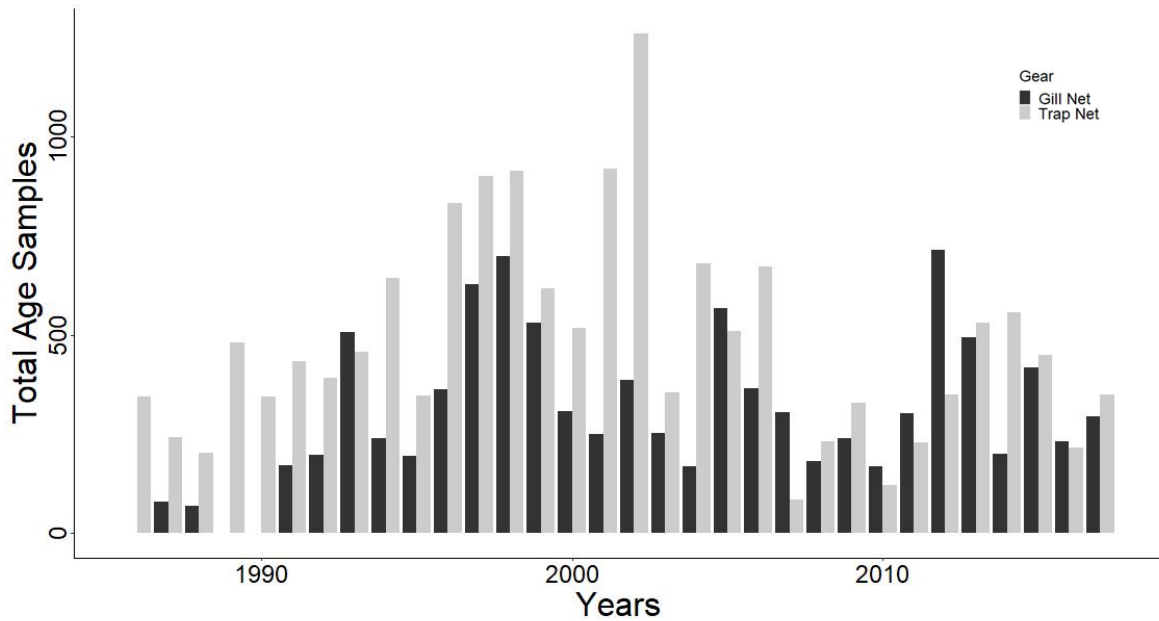
- Thorson, J.T., Kristensen, K., 2016. Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. *Fish. Res.* 175, 66-74. <https://doi.org/10.1016/j.fishres.2015.11.016>.
- Thorson, J.T., Johnson, K.F., Methot, R.D., Taylor, I.G., 2017. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. *Fish. Res.* 192, 84-93. <https://doi.org/10.1016/j.fishres.2016.06.005>.
- Trijoulet, V., Albertsen, C.M., Kristensen, K., Legault, C.M., Miller, T.J., Nielsen, A., 2023. Model validation for compositional data in stock assessment models: Calculating residuals with correct properties. *Fish. Res.* 257, 106487. <https://doi.org/10.1016/j.fishres.2022.106487>.
- Truesdell, S.B., Bence, J.R., 2016. A Review of stock assessment methods for lake trout and lake whitefish in 1836 treaty waters of Lake Huron, Lake Michigan and Lake Superior. Quantitative Fisheries Center Technical Report T2016-01. <https://doi.org/10.6084/m9.figshare.3123949>.
- Truesdell, S.B., Bence, J.R., Syslo, J.M., Ebener, M.P., 2017. Estimating multinomial effective sample size in catch-at-age and catch-at-size models. *Fish. Res.* 192, 66-83. <https://doi.org/10.1016/j.fishres.2016.11.003>.
- U.S. v. Michigan. 2000. Stipulation for Entry of Consent Decree. U.S. District Court Western District of Michigan Southern Division, Case No. 2:73 CV 26.
- VanDeHey, J.A., Sloss, B.L., Peeters, P.J., Sutton, T.M., 2009. Genetic structure of lake whitefish (*Coregonus clupeaformis*) in Lake Michigan. *Can. J. Fish. Aquat. Sci.* 66 (3), 382-393. <https://doi.org/10.1139/F08-213>.
- Victorero, L., Watling, L., Deng Palomares, M.L., Nouvian, C., 2018. Out of Sight, But Within Reach: A Global History of Bottom-Trawled Deep-Sea Fisheries From >400 m Depth. *Front. Mar. Sci.* 5, 1-17. <https://doi.org/10.3389/fmars.2018.00098>.
- Wilberg, M.J., Bence, J.R., Eggold, B.T., Maskauskas, D., Clapp, D.F., 2005. Yellow Perch dynamics in Southwestern Lake Michigan during 1986–2002. *N. Am. J. Fish. Manage.* 25, 1130-1152. <https://doi.org/10.1577/M04-193.1>.
- Wilberg, M.J., Thorson, J.T., Linton, B.C., Berkson, J., 2009. Incorporating time-varying catchability into population dynamic stock assessment models. *Rev. Fish. Sci.* 18, 7-24. <https://doi.org/10.1080/10641260903294647>.
- Yin, Y., Aeberhard, W.H., Smith, S.J., Flemming, J.M., 2019. Identifiable state-space models : A case study of the Bay of Fundy sea scallop fishery. *Can. J. Stat.* 47, 27-45. <https://doi.org/10.1002/cjs.11470>.

Zhao, Y., Morbey, Y.E., 2017. Estimating the size selectivity of trap nets using a gill-net selectivity experiment: Method development and application to lake whitefish in lake Huron. *N. Am. J. Fish. Manage.* 37, 1341-1349.  
<https://doi.org/10.1080/02755947.2017.1381206>.

## APPENDIX 1A: DATA COLLECTION AND PROCESSING

Fishery and biological data of lake trout and lake whitefish have been collected in the 1836 Treaty waters of the Great Lakes since 1985, though the time series for some regions begins later. Each year, the time series up to but *not* including that year are used to build stock assessments that are used to make harvest recommendations for the following year (i.e., the time series 1986-2017 is used in a model built in 2018 that makes recommendations for 2019). A catch reporting system provides monthly total harvest,  $H_{y,G}$ , in weight, and effort data,  $E_{y,G}$ , in lifts for the trap net fishery, or length of net for the gill net fishery from 1986 to 2017, which are pooled from daily reports mandated for each licensed commercial fisher (Ebener *et al.*, 2005). Biological data were collected opportunistically by sampling catch from commercial trap and gill net fisheries from 1986-2017. The number of individual fish sampled each year varied considerably over time and between fisheries, from 86-1261 in the trap net fishery and 0-716 in the gill net fishery (Figure 1A.1).

Figure 1A.1. Number of fish sampled and aged from the gill net and trap net fishery of Lake Michigan lake whitefish in region WFM-03 from 1986-2017.



In general, aging was done using scale samples for smaller (and presumably younger) fish and otoliths for larger (and presumably older) fish. A total of 2,484 scale samples were taken from individuals ranging from 375-669mm (averaging 464mm), and a total of 748 otolith samples were collected, from individuals ranging from 430-623mm (averaging 523mm). Otoliths are a more accurate and precise aging tool than either fin rays or scales to age lake whitefish and were assumed to be unbiased measurements (Herbst and Marsden, 2011). Because scale samples yield similar age estimates as otoliths at younger ages but are less precise at older ages, dividing the aging between the two structures should minimize aging error (Herbst and Marsden, 2011).

Weight at age,  $W_{a,y}^{(harv)}$ , and length at age,  $L_{a,y}$ , for the harvest in each year was determined using a growth model with year- and cohort- specific parameters, using data from both fisheries (He and Bence, 2007). These calculations were done prior to assessment model fitting and provided as inputs to the assessment model. Weight at age from the sampled harvest and samples from a graded-mesh gill net survey were used to calculate population weight at age at the

beginning of the year,  $W_{a,y}^{(init)}$ , and at the time of spawning,  $W_{a,y}^{(spawn)}$ , assuming that harvest occurs on June 30<sup>th</sup>, spawning occurs on October 30<sup>th</sup>, and growth from one year to the next follows an exponential model.

The total harvest in weight by year and fishery,  $H_{y,G}$ , was converted into total harvest in numbers using the average weight by year and fishery,  $\bar{W}_{y,G}$ . The observed total catch in each year by each fishery  $C_{y,G}$  was a function of total harvest, the mean weight, and a year- and fishery- specific adjustment term,  $A_{y,G}$ , that corrected for under reporting of each fishery,

$$C_{y,G} = \frac{H_{y,G}}{\bar{W}_{y,G} A_{y,G}}.$$

Underreporting rate was estimated as the ratio of reported whitefish harvested by the fishers to the reported whitefish purchased by wholesalers (Ebener *et al.*, 2005). Observed proportions at age for each fishery and year,  $P_{a,y,G}$  were calculated from the commercial sampling data, by normalizing the number sampled at age and year by the total sampled in each year. Only fish age 4 and older were included in the stock (younger fish were extremely rarely caught) and all fish age 15 and older were aggregated into a plus group. The oldest recorded age was 32, however, individuals aged 20 or older were exceedingly rare in this region (<0.1% of sampled individuals).

## APPENDIX 1B: CURRENT STATISTICAL CATCH AT AGE MODEL (SCA)

I refer to this model as the current assessment, despite some minor modifications that occurred to the model in 2021, and it is an age-based statistical catch at age model (henceforth “SCA”) fit to commercial trap net and gill net fishery data with additional data inputs described below (Caroffino and Seider, 2020; Ebener *et al.*, 2005; Truesdell and Bence, 2016). The model was built using AD Model Builder (Fournier *et al.*, 2012). The years of data used in the assessment were from 1986 to 2017 (i.e.,  $y = 1986, 1989 \dots 2017$ ), the age classes ranged from 4 to 15+ (i.e.,  $a = 4, 5, \dots A$ , where  $A=15+$ ), and there were catch and effort data for each of the two fisheries (i.e.,  $G=g, t$ ).

The model estimated the abundance of individuals ages 4-9 in 1986 and the abundance (i.e., recruitment) of age 4 individuals in each year. Log-scale recruitment was calculated as the sum of two log-scale parameters- an average recruitment,  $\bar{R}$ , and a vector of deviations,  $D_y$ ,

$$\log N_{4,y} = \log \bar{R} + \log D_y.$$

The number of individuals from age 4 to 9 in the initial year were calculated in a similar way, using the same average recruitment and another set of deviation values  $d_a$ ,

$$\log N_{a,1986} = \log \bar{R} + \log d_a \text{ where } a < 10.$$

The number of individuals in the initial year from age 10 to 15 were set to 0, because few or no fish of these ages were detected in the catch and there were insufficient data to reliably inform abundance and catchability. The effects of this simplification on model output, particularly in later years of the time series, are likely negligible. The vectors of  $D_y$  and  $d_a$  were constrained to collectively sum to 0. All other abundances at age and year, after the initial year and initial age,  $N_{a,y}$ , followed an exponential decay with all fish age  $A$  or older subject to the same year-specific instantaneous mortality (Table 1.2, Eq. 1.4). The instantaneous total mortality

at age and year,  $Z_{a,y}$ , was the sum of age-, year-, and fishery- specific instantaneous fishing mortality,  $F_{a,y,G}$ , and a single estimated instantaneous natural mortality,  $M$  (Eq. 1.5).

Instantaneous fishing mortality was the product of year- and fishery- specific catchability,  $q_{y,G}$ , year- and fishery- specific fishing effort,  $E_{y,G}$ , and age- and year- and fishery- specific selectivity,  $s_{a,y,G}$  (Eq. 1.6a). Log-scale catchability varied over time according to a random walk (Eq. 1.7a).

Selectivity at age for each fishery was a function of mean length at age and varied over time in the trap net fishery. The selectivity of the trap net fishery was modeled as a logistic function of mean length at age each year, normalized against a reference length, 455 mm (i.e., selectivity was 1.0 at this mean length at age):

$$s_{a,y,t} = \frac{\frac{1}{1 + \exp(-p_{2,t}(L_{y,a} - p_{1,y,t}))}}{1 + \exp(-p_{2,t}(455 - p_{1,y,t}))}$$

$$\log p_{1,y+1,t} = \log p_{1,y,t} + \varepsilon_y^{(s)}; \varepsilon_y^{(s)} \sim N(0, \sigma_s^2); y < 2017.$$

The parameters,  $p_{1,y,t}$  and  $p_{2,t}$ , are the inflection point (varying over years), and the determinant of the slope at the inflection point (constant over years), respectively. The inflection point for the initial year was estimated and then varied over according to a random walk, the distribution of which was included in the total objective function (Eq. 3.4),

Selectivity of the gill net fishery used a lognormal function with two estimated parameters,  $p_{1,g}$  and  $p_{2,g}$ , and was normalized to equal 1.0 for a mean length at age of  $\exp(p_{2,g})$ ,



$$S_{a,y,g} = \frac{\frac{1}{\sqrt{2\pi}p_{1,g}L_{y,a}} \exp\left(-\frac{(\log L_{y,a} - p_{2,g})^2}{2p_{1,g}^2}\right)}{\frac{1}{\sqrt{2\pi}p_{1,g} \exp p_{2,g}}}.$$

Spawning stock biomass,  $B_y^{(Spawn)}$ , in each year was calculated by multiplying the weight at age during spawning (adjusted with maturity schedule), by the number of individuals at age in each year (Eq. 1.8).

The predicted age-, year-, and fishery- specific catch,  $C_{a,y,G}$ , was calculated from the Baranov catch equation (Eq. 2.1). The annual proportions at age were calculated from the age- and year- specific total catch (Eq. 2.2). The objective function of the SCA model,  $T$ , was the sum of the negative log prior probability density of the parameters, NLP, and the negative log likelihood of the data given the parameters, NLL (Eq. 3.1). The NLP was the sum of explicit components explaining variability in recruitment,  $N_{4,y}$ , log-scale catchability (trap and gill net), and selectivity (trap net). Priors for all other parameters were from uniform distributions, specified via bounds for the parameters.

Recruitment was constrained such that estimated values did not deviate substantially from a mean predicted by a Ricker stock recruitment function,  $\hat{N}_{4,y}$  which depended on Ricker stock recruitment parameters,  $\alpha$  and  $\beta$ , and year-specific spawning stock size in eggs,  $\epsilon_y$

$$\log \hat{N}_{4,y+5} = \alpha \epsilon_y^{(-\beta \epsilon_y)}.$$

The predicted value was compared to the current estimates,  $N_{a,y}$ , using a likelihood framework during model fitting (Eq. 3.2a). Fecundity data, including age- and year- specific maturity,  $m_{a,y}$ , the average number of eggs per kilogram,  $e$ , and the average proportion of females in the spawning population,  $f$ , were input into the model. The maturity matrix was a 5-year running average that was the same across all 1836 treaty waters and calculated by applying

an age/length key to maturity at length data. Eggs per kilogram was also a universal value based on samples collected in 1983 and 1996 (Ebener *et al.*, 2005). The proportion of females was calculated from samples of the commercial harvest and was a single age- and year-invariant value for each management unit, which was 0.4845 for WFM-03 (Ebener *et al.*, 2005; Patriarche, 1977).

Spawning stock size (i.e., eggs) was calculated from maturity at age, weight at age during spawning,  $W_{y,i}^{(spawn)}$ , eggs per kilogram, percent female, numbers at age, and total instantaneous mortality for the proportion of the year that occurs before spawning (0.838),

$$\epsilon_y = \sum_{i=4}^A m_{y,i} W_{y,i}^{(spawn)} ef N_{i,y}^{-Z_{y,i} * 0.838}$$

The deviations from predicted mean recruitment were assumed to be lognormally distributed (Eq. 3.2a). Note that these terms start in 1991, which is the year class produced by the spawning biomass in 1986 (the first year being modeled). The errors for the log-scale trap net and gill net catchability random walks were assumed to come from fishery-specific normal distributions (Eq. 3.3a). Similarly, the errors for the random walk of the time-varying trap net selectivity parameter,  $p_{1,y,t}$ , were assumed to come from a normal distribution (Eq. 3.4). The NLP also included a prior expectation on instantaneous natural mortality, that it was assumed to arise from a lognormal distribution with median  $M = 0.2$  and standard deviation,  $\sigma_M = 0.1$  (Eq. 3.5a).

The NLL included four summed components, for the total catch and proportions at age of the trap net and gill net fisheries. The observed total catch by year, for each fishery was assumed to come from a lognormal distribution, with median equal to the predicted catch (Eq. 3.6). The observed catch proportions at age by year for each fishery were assumed to behave as though

they resulted from sampled numbers at age following a multinomial distribution, with fishery-specific effective sample size,  $n_G$  determined *a priori* using an iterative reweighting algorithm (Francis, 2011; Truesdell et al., 2017) (Eq. 3.7). The effective sample size in each year was the product of the number of fish sampled and the gear-specific scalar, 0.12 and 0.06 for the trap and gill net fisheries, respectively, rounded to the nearest whole number. Those effective sample sizes are reported in Table 1B.1.

Table 1B.1. Year-specific effective sample size of the multinomial distribution of proportion at age of the trap net,  $n_t$ , and gill net,  $n_g$ , fisheries.

Year	$n_t$	$n_g$
1986	40	0
1987	28	5
1988	24	4
1989	56	0
1990	40	0
1991	51	10
1992	46	12
1992	53	30
1993	75	14
1994	40	11
1995	97	21
1996	105	37
1997	107	41
1998	72	31
1999	60	18
2000	107	15
2001	147	23
2002	41	15
2003	79	10
2004	59	33
2005	78	21
2006	10	18
2007	27	11
2008	38	14
2009	14	10
2010	27	18
2011	41	42
2012	61	29
2013	65	12

Table 1B.1 (cont'd)

2014	52	24
2015	25	13
2016	41	17
2017	40	0

The standard deviations for the above distributions were calculated as the product of a single common standard deviation term,  $\sigma$ , which was estimated, and a standard deviation specific multiplier,  $\varphi$ , that was pre-specified. This constraint was implemented because in this model structure there was insufficient information to independently estimate the standard deviations of each process variability and observation error distribution. The model used a minimization algorithm to determine the point estimates for parameters that minimized the total objective value  $T$ . The estimated parameters and standard errors are reported in Table 1B.2.

Table 1B.2. Parameter estimates and standard error of SCA model.

Parameter (Symbol)	Estimate	Standard Error
$M$	0.1838	0.0171
$\sigma$	0.0445	0.0026
$q_{1986,t}$	0.0221	0.0051
$q_{1987,t}$	0.0275	0.0055
$q_{1988,t}$	0.0287	0.0047
$q_{1989,t}$	0.0241	0.0040
$q_{1990,t}$	0.0208	0.0035
$q_{1991,t}$	0.0270	0.0042
$q_{1992,t}$	0.0278	0.0040
$q_{1993,t}$	0.0302	0.0044
$q_{1994,t}$	0.0334	0.0048
$q_{1995,t}$	0.0328	0.0051
$q_{1996,t}$	0.0317	0.0048
$q_{1997,t}$	0.0303	0.0047
$q_{1998,t}$	0.0281	0.0043
$q_{1999,t}$	0.0286	0.0038
$q_{2000,t}$	0.0254	0.0037
$q_{2001,t}$	0.0220	0.0029
$q_{2002,t}$	0.0188	0.0029
$q_{2003,t}$	0.0206	0.0032

Table 1B.2 (cont'd)

$q_{2004,t}$	0.0280	0.0041
$q_{2005,t}$	0.0258	0.0039
$q_{2006,t}$	0.0247	0.0036
$q_{2007,t}$	0.0235	0.0039
$q_{2008,t}$	0.0189	0.0028
$q_{2009,t}$	0.0164	0.0023
$q_{2010,t}$	0.0142	0.0022
$q_{2011,t}$	0.0155	0.0025
$q_{2012,t}$	0.0147	0.0025
$q_{2013,t}$	0.0116	0.0023
$q_{2014,t}$	0.0093	0.0022
$q_{2015,t}$	0.0080	0.0023
$q_{2016,t}$	0.0064	0.0022
$q_{2017,t}$	0.0067	0.0025
$q_{1986,g}$	0.1015	0.0147
$q_{1987,g}$	0.0858	0.0114
$q_{1988,g}$	0.0809	0.0103
$q_{1989,g}$	0.0811	0.0112
$q_{1990,g}$	0.0486	0.0059
$q_{1991,g}$	0.0687	0.0086
$q_{1992,g}$	0.0608	0.0067
$q_{1993,g}$	0.0682	0.0084
$q_{1994,g}$	0.0614	0.0081
$q_{1995,g}$	0.0659	0.0086
$q_{1996,g}$	0.0484	0.0064
$q_{1997,g}$	0.0525	0.0075
$q_{1998,g}$	0.0454	0.0063
$q_{1999,g}$	0.0389	0.0052
$q_{2000,g}$	0.0486	0.0070
$q_{2001,g}$	0.1061	0.0167
$q_{2002,g}$	0.0734	0.0113
$q_{2003,g}$	0.0435	0.0066
$q_{2004,g}$	0.0631	0.0101
$q_{2005,g}$	0.0656	0.0104
$q_{2006,g}$	0.0754	0.0120
$q_{2007,g}$	0.0701	0.0122
$q_{2008,g}$	0.0743	0.0118
$q_{2009,g}$	0.0532	0.0073
$q_{2010,g}$	0.0448	0.0060
$q_{2011,g}$	0.0555	0.0084
$q_{2012,g}$	0.0435	0.0072
$q_{2013,g}$	0.0329	0.0066
$q_{2014,g}$	0.0509	0.0125

Table 1B.2 (cont'd)

$q_{2015,g}$	0.0257	0.0070
$q_{2016,g}$	0.0303	0.0088
$q_{2017,g}$	0.0198	0.0063
$p_{1,1986,t}$	475.38	9.5069
$p_{1,1987,t}$	483	8.735
$p_{1,1988,t}$	475.83	10.337
$p_{1,1989,t}$	471.67	7.9452
$p_{1,1990,t}$	479.03	10.072
$p_{1,1991,t}$	463.82	8.2848
$p_{1,1992,t}$	466.99	6.9187
$p_{1,1993,t}$	470.45	6.5829
$p_{1,1994,t}$	467.32	6.5738
$p_{1,1995,t}$	479.12	5.7553
$p_{1,1996,t}$	480.74	5.4075
$p_{1,1997,t}$	478.89	5.1605
$p_{1,1998,t}$	491.21	5.3963
$p_{1,1999,t}$	470.89	6.7586
$p_{1,2000,t}$	473.77	6.4606
$p_{1,2001,t}$	447.77	5.9078
$p_{1,2002,t}$	473.12	5.6466
$p_{1,2003,t}$	472.59	7.4029
$p_{1,2004,t}$	458.43	6.8059
$p_{1,2005,t}$	461	8.519
$p_{1,2006,t}$	437.27	7.8146
$p_{1,2007,t}$	435	8.3108
$p_{1,2008,t}$	435.06	7.9925
$p_{1,2009,t}$	450.69	6.734
$p_{1,2010,t}$	472.2	7.4435
$p_{1,2011,t}$	469.65	6.3439
$p_{1,2012,t}$	478.42	6.9458
$p_{1,2013,t}$	477.52	7.1065
$p_{1,2014,t}$	439.58	7.3909
$p_{1,2015,t}$	463.28	9.3819
$p_{1,2016,t}$	476.64	13.21
$p_{1,2017,t}$	478.23	11.348
$p_{2,t}$	0.0600	0.0040
$p_{1,g}$	0.0833	0.0063
$p_{2,g}$	6.29	0.0154
$\bar{R}$	504140	49815
$D_{1986}$	1.4627	0.2157
$D_{1987}$	0.8984	0.1619
$D_{1988}$	0.5004	0.1086
$D_{1989}$	0.6276	0.1224

Table 1B.2 (cont'd)

$D_{1990}$	1.624	0.2177
$D_{1991}$	2.0457	0.2577
$D_{1992}$	2.3842	0.2723
$D_{1993}$	1.6725	0.2007
$D_{1994}$	1.3842	0.1679
$D_{1995}$	3.0649	0.2843
$D_{1996}$	2.5721	0.2435
$D_{1997}$	2.4709	0.2369
$D_{1998}$	2.2904	0.2348
$D_{1999}$	2.3116	0.2563
$D_{2000}$	2.0771	0.2498
$D_{2001}$	1.6651	0.2206
$D_{2002}$	1.5444	0.2248
$D_{2003}$	1.2869	0.2152
$D_{2004}$	2.1706	0.3452
$D_{2005}$	2.9389	0.4021
$D_{2006}$	2.4684	0.3507
$D_{2007}$	2.0104	0.2881
$D_{2008}$	1.657	0.2356
$D_{2009}$	1.4356	0.2023
$D_{2010}$	1.1113	0.1674
$D_{2011}$	0.9298	0.1522
$D_{2012}$	0.8216	0.1522
$D_{2013}$	0.8075	0.1805
$D_{2014}$	0.7559	0.1974
$D_{2015}$	0.9726	0.3089
$D_{2016}$	0.6199	0.2238
$D_{2017}$	0.6573	0.3222
$d_5$	1.006	0.1666
$d_6$	0.4012	0.0992
$d_7$	0.0923	0.0462
$d_8$	0.0242	0.0235
$d_9$	0.0134	0.0175
$\alpha$	0.0003	0.0001
$\beta$	1.17E-10	3.64E-11

Several plots are included to illustrate model diagnostics. Histogram plots of ordinary least squared (OLS) residuals (observed-expected value) of (A) total trap net catch on the log-scale, (B) total gill net catch on the log-scale, and one step ahead (OSA) residuals of (C) trap net

proportions at age, and (D) gill net proportions at age are reported in Figure 1.B1. The OLS residuals of the total trap net and gill net fishery catch over time are reported in Figures 1B.2-1B.3. The OSA residuals of proportions at age are plotted as a bubble plot to highlight potential patterns across age and time (Figures 1B.4-1B.5). Retrospective plots for the recruitment and spawning stock biomass are reported in Figures 1B.6-1B.7.

Figure 1B.1. Histogram plots of ordinary least squared (OLS) residuals of (A) total trap net catch on the log-scale and (B) total gill net catch on the log-scale, and one step ahead (OSA) residuals of (C) trap net proportions at age and (D) gill net proportions at age.

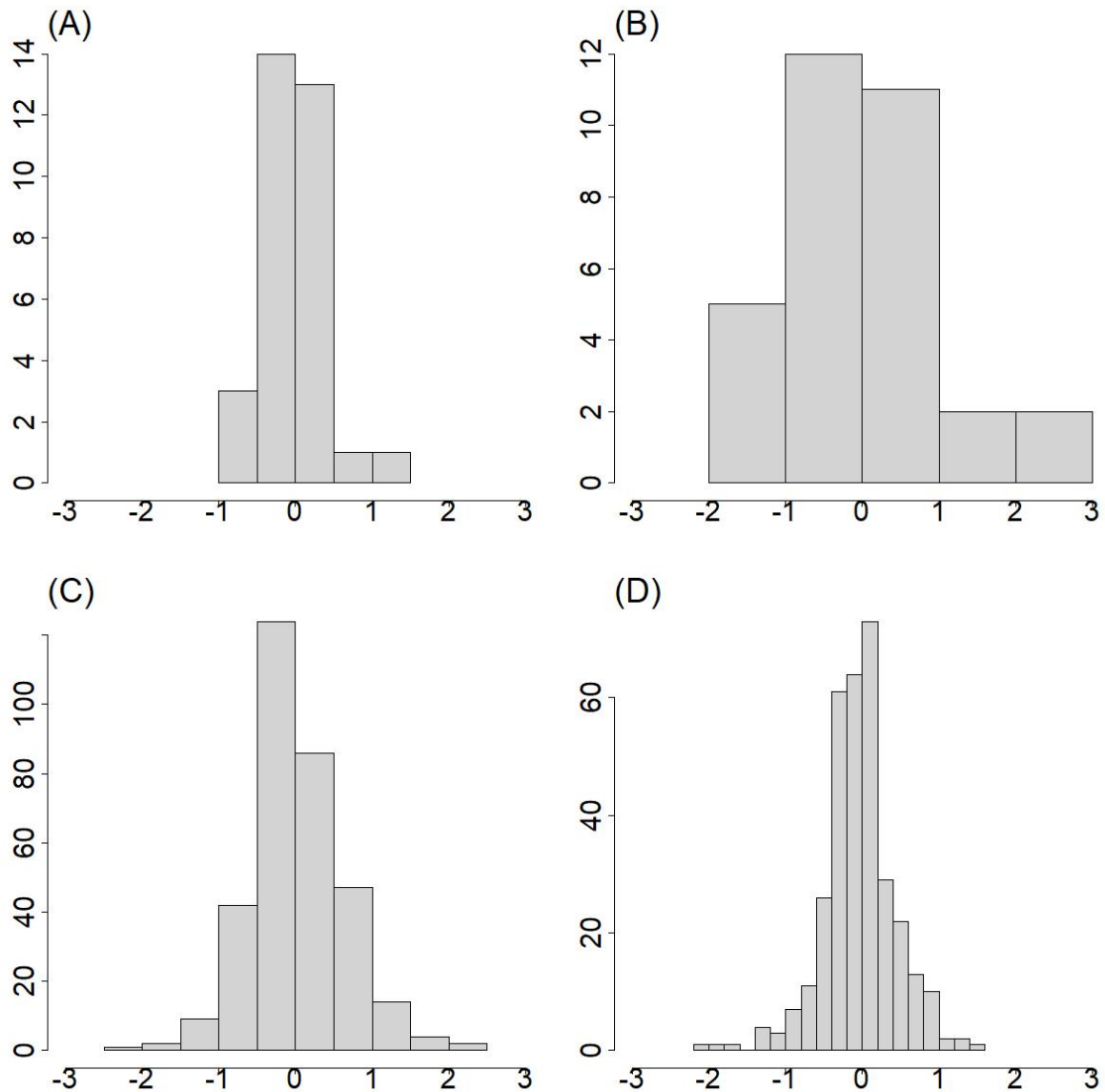




Figure 1B.2. Ordinary least squared (OLS) residuals of total log-scale catch of the trap net fishery over time. Observed catch was modeled as a lognormal distribution with mean equal to the expected catch and standard deviation 0.067.

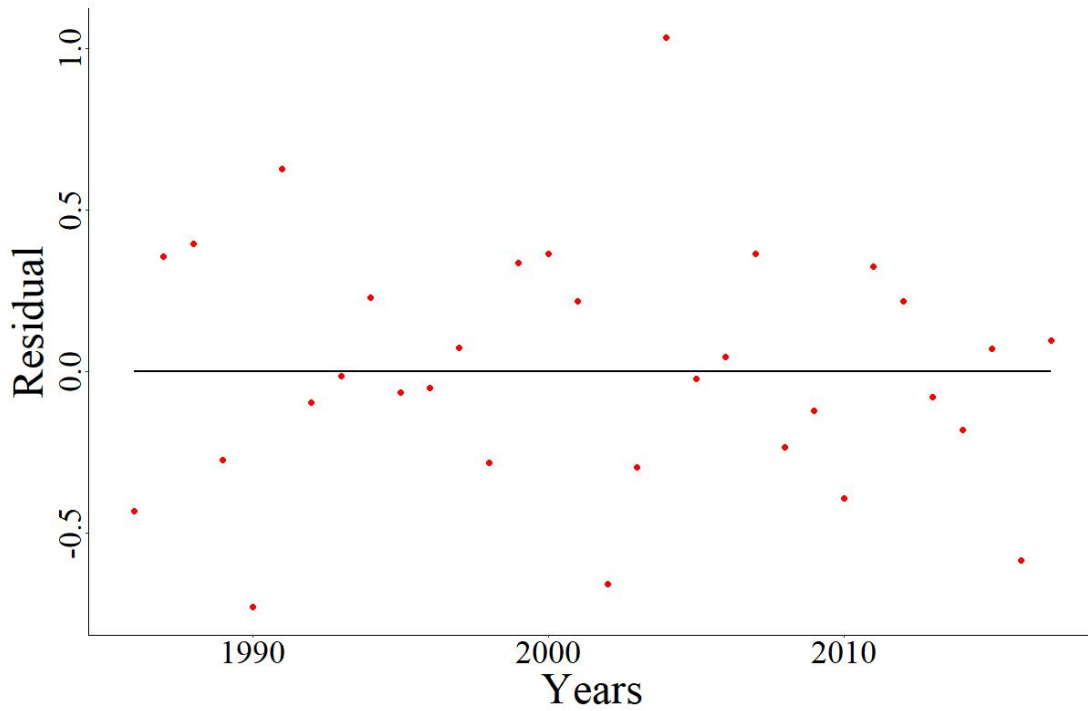


Figure 1B.3. Ordinary least squared (OLS) residuals of total log-scale catch of the gill net fishery by year. Observed catch was modeled as a lognormal distribution with mean equal to the expected catch and standard deviation 0.067.

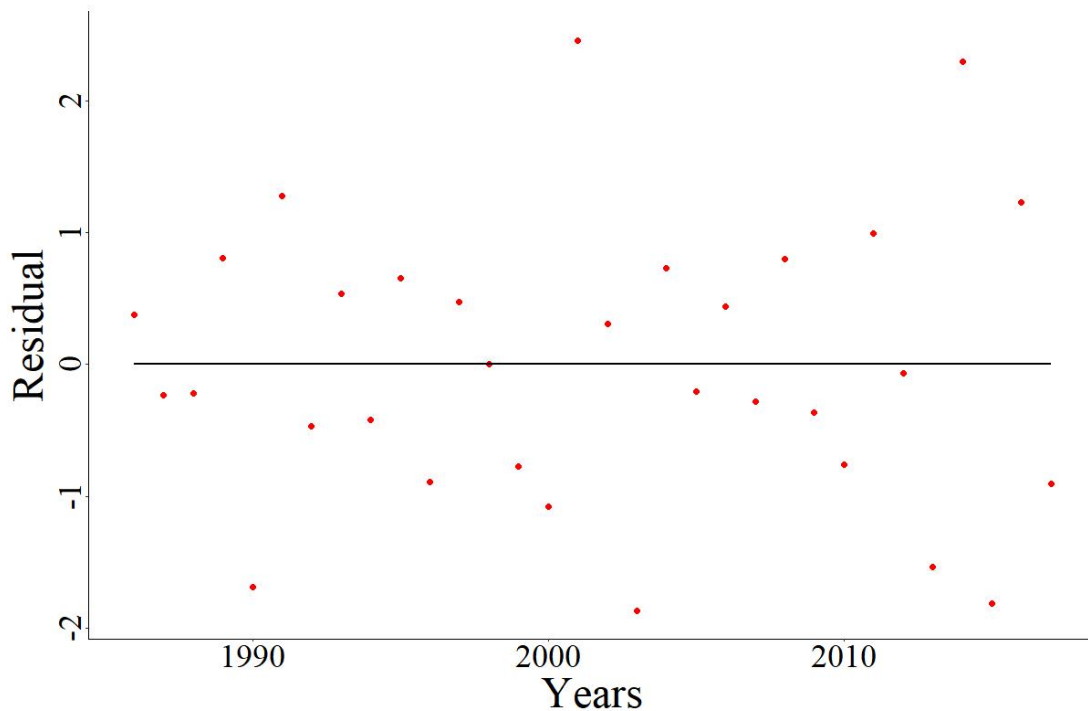


Figure 1B.4. One step ahead (OSA) residuals of proportions at age of the trap net fishery. Observed proportions were modeled as a multinomial distribution with year-specific effective sample size. No residuals were reported for year and age combination where the expected and observed value was 0 (older ages in the first 6 years) or for age 15+, because the way OSA residuals are calculated does not allow for estimates in the oldest age.

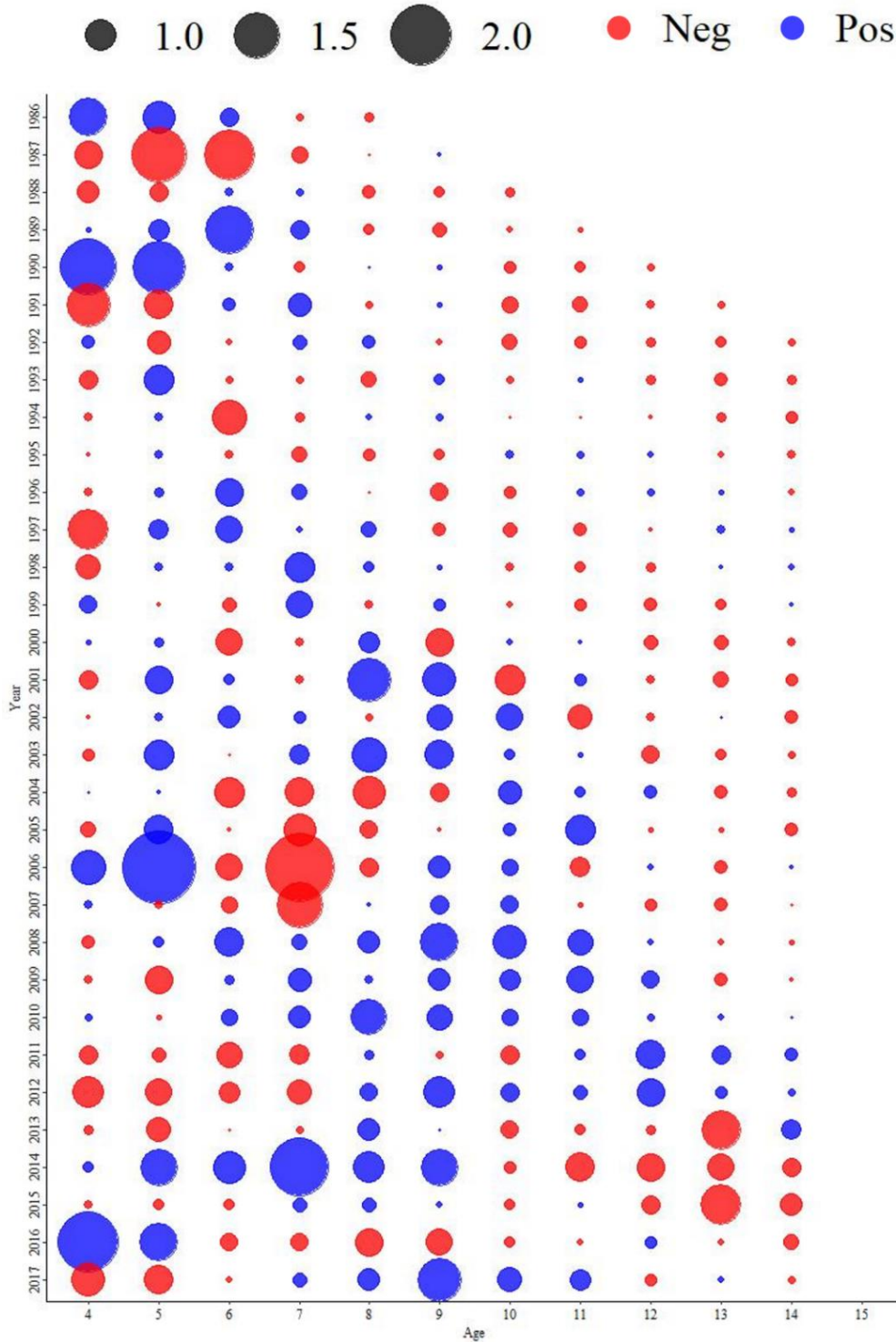


Figure 1B.5. One step ahead (OSA) residuals of proportions at age of the gill net fishery. Observed proportions were modeled as a multinomial distribution with year-specific effective sample size. No residuals were reported for year and age combination where the expected and observed value was 0 (older ages in the first 6 years) or for age 15+, because the way OSA residuals are calculated does not allow for estimates in the oldest age. Note that years 1986, 1989, and 1990 are also absent because the sample size was zero.

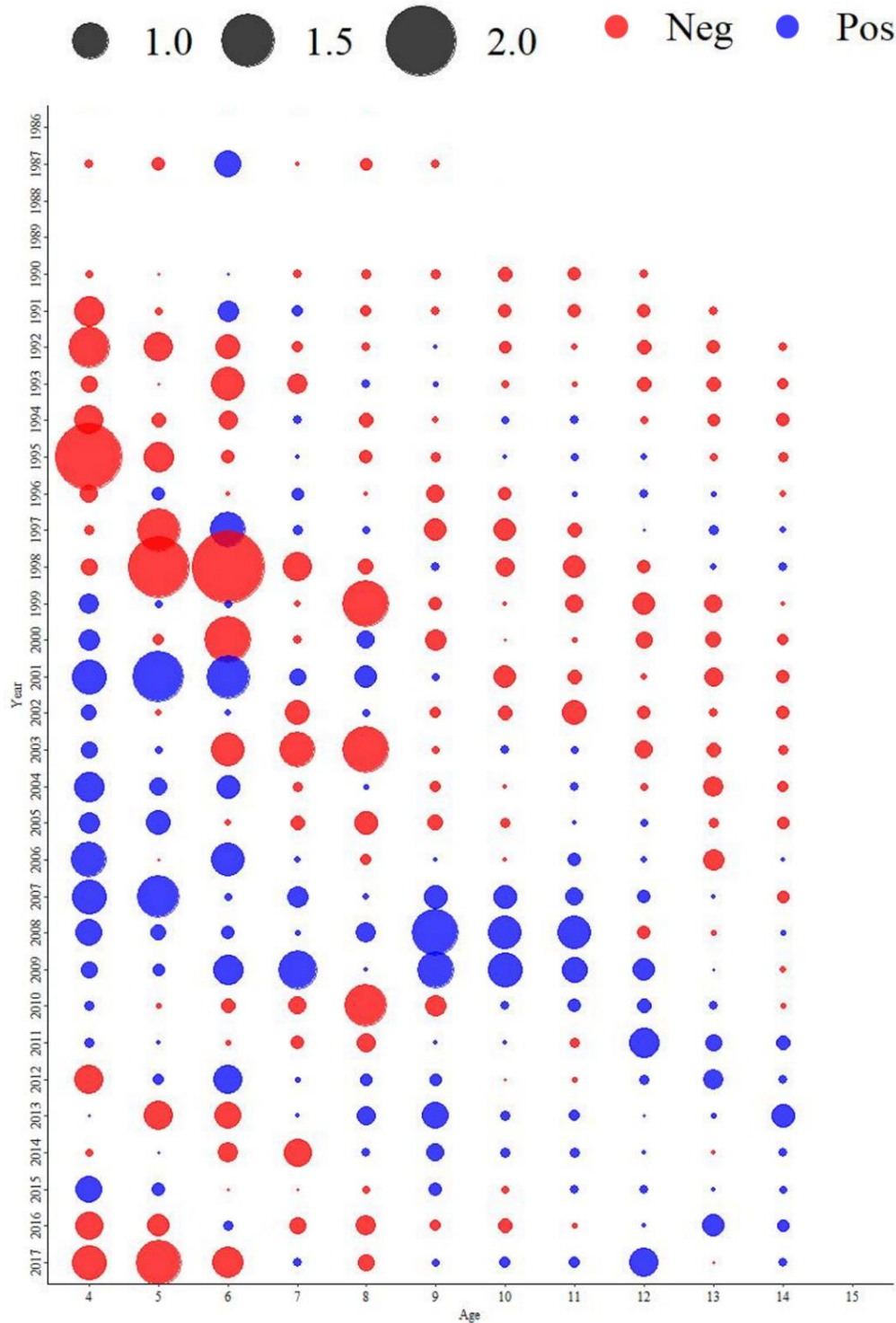


Figure 1B.6. Retrospective analysis of estimated recruitment. The SCA model was refit after removing sequential years of the most recent observations. Dark colors indicate more complete data sets and lighter colors indicate less complete data sets.

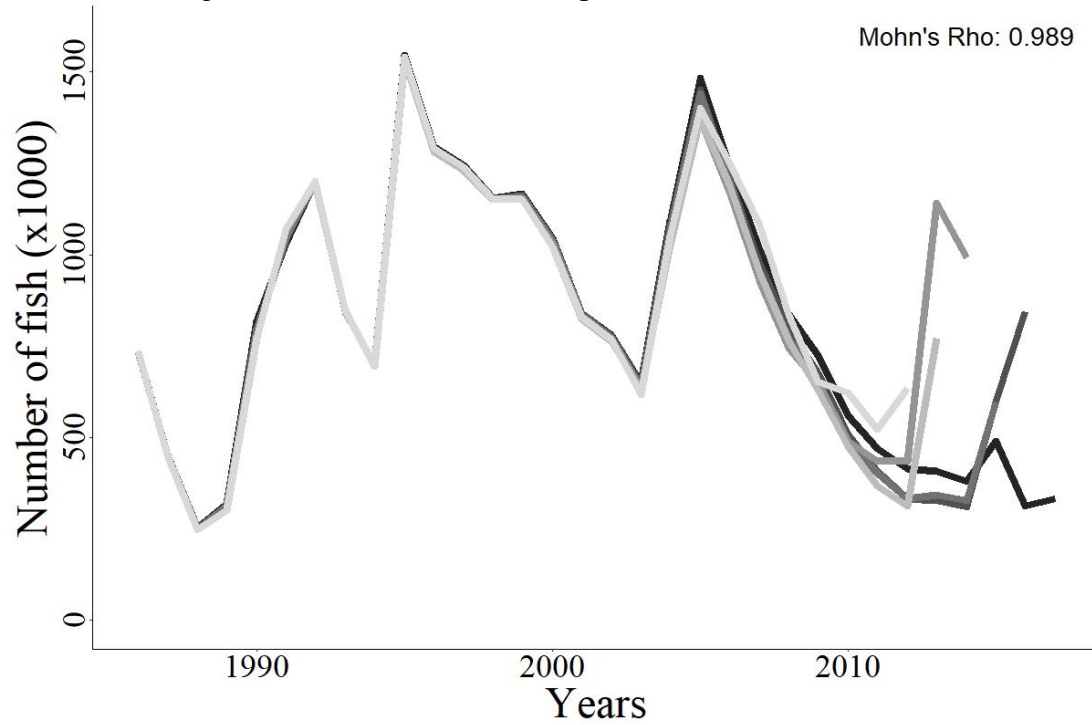
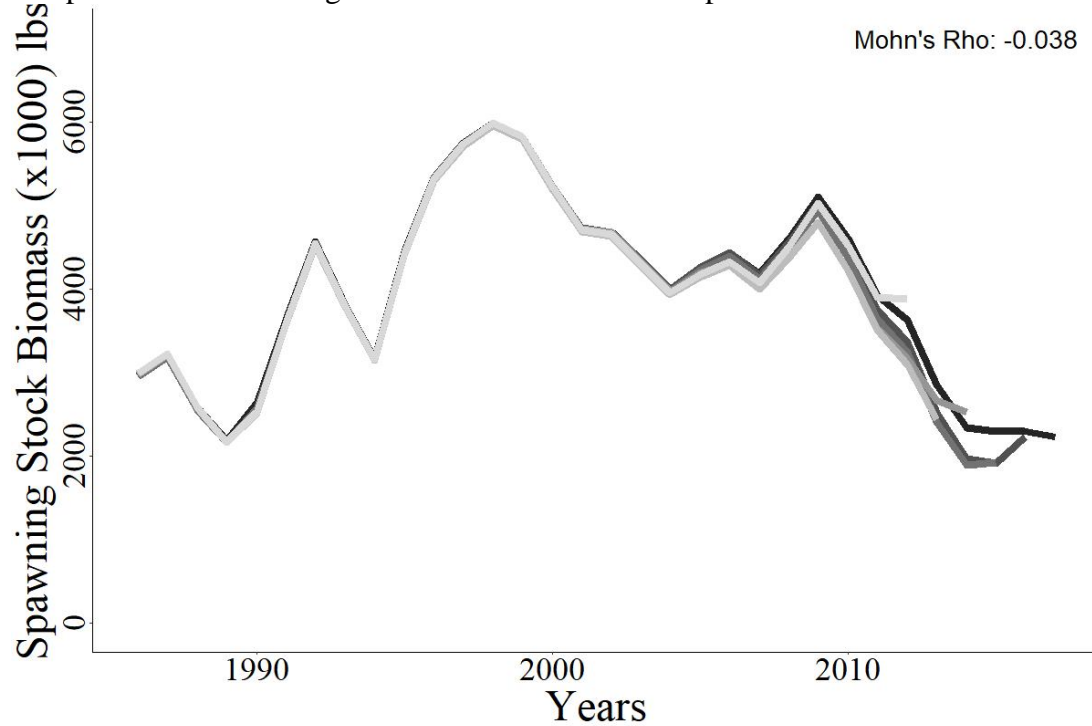


Figure 1B.7. Retrospective analysis of estimated spawning stock biomass. The SCA model was refit after removing sequential years of the most recent observations. Dark colors indicate more complete data sets and lighter colors indicate less complete data sets.



## APPENDIX 1C: MODEL DIAGNOSTICS FOR THE STATE-SPACE MODEL

The state-space stock assessment model predicted temporally variability in age structure with older ages being more represented in the population in later years (Figure 1C.1). The state-space stock assessment model (SSM) fit the catch and catch composition data well for both fisheries with plots of the predicted and observed total harvest over time entirely overlapping, with no difference distinguishable through line graphs, and therefore are not reported here. The histograms of one step ahead (OSA) residuals for the total catch on the log-scale, proportions at age, and catchability process variability are plotted in Figure 1C.2. The residuals were plotted over time in Figures 1C.3-1C.6. Retrospective plots for the recruitment and spawning stock biomass are reported in Figures 1C.7-1C.8.

Figure 1C.1. Predicted proportions at age over time in the state-space stock assessment model (SSM).

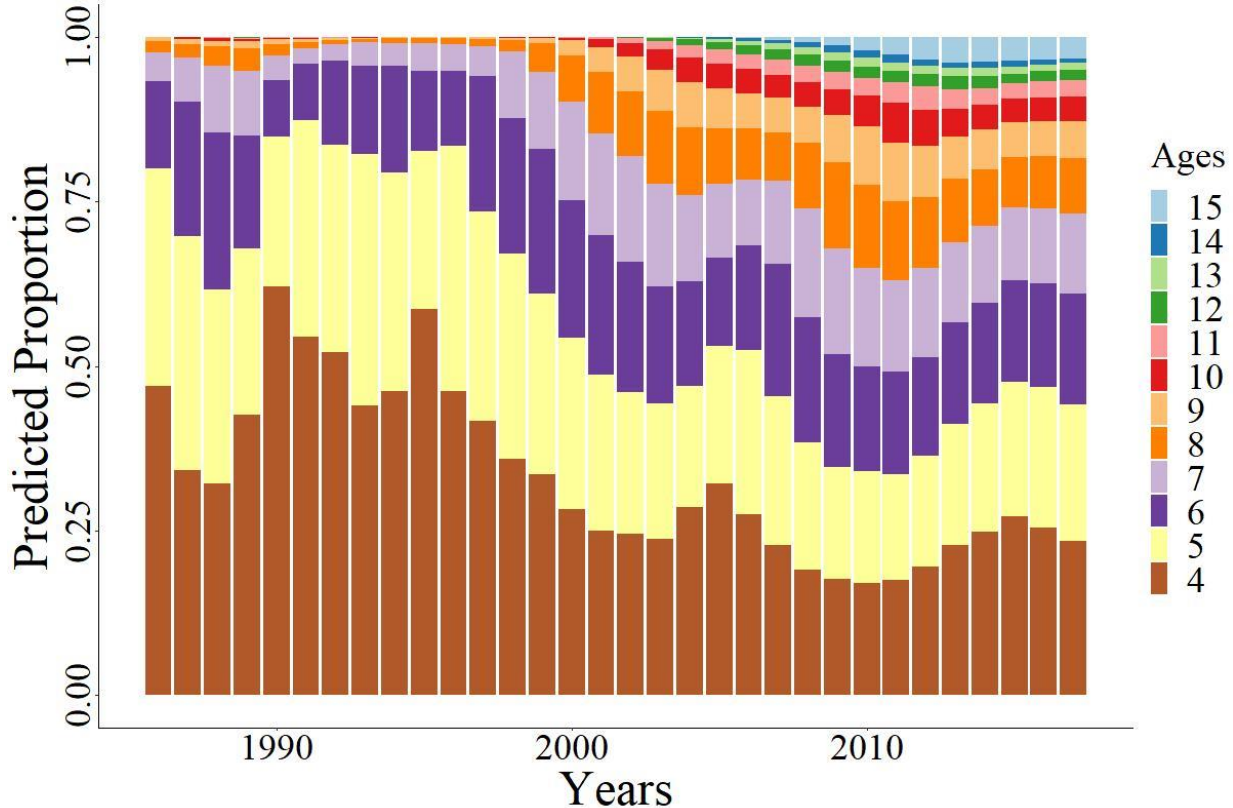


Figure 1C.2. Histogram plots of one step ahead (OSA) residuals of (A) total trap net catch on the log-scale, (B) total gill net catch on the log-scale, (C) trap net proportions at age and (D) gill net proportions at age, and the process variability at age of the (E) trap net catchability and (F) gill net catchability.

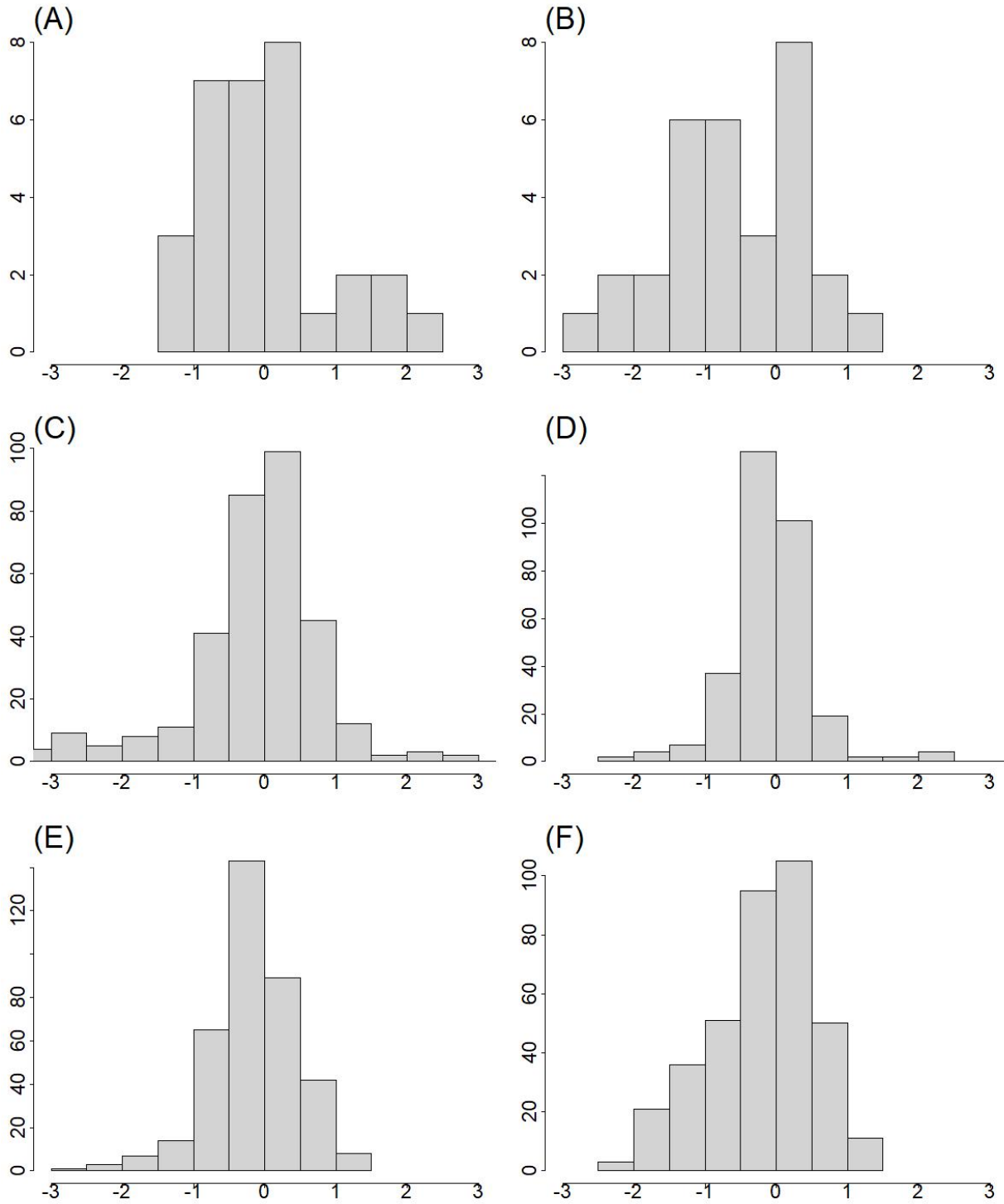


Figure 1C.3. One step ahead (OSA) residuals of total log-scale catch of the trap net fishery by year. Observed catch was modeled as a lognormal distribution with mean equal to the expected catch and standard deviation 0.067.

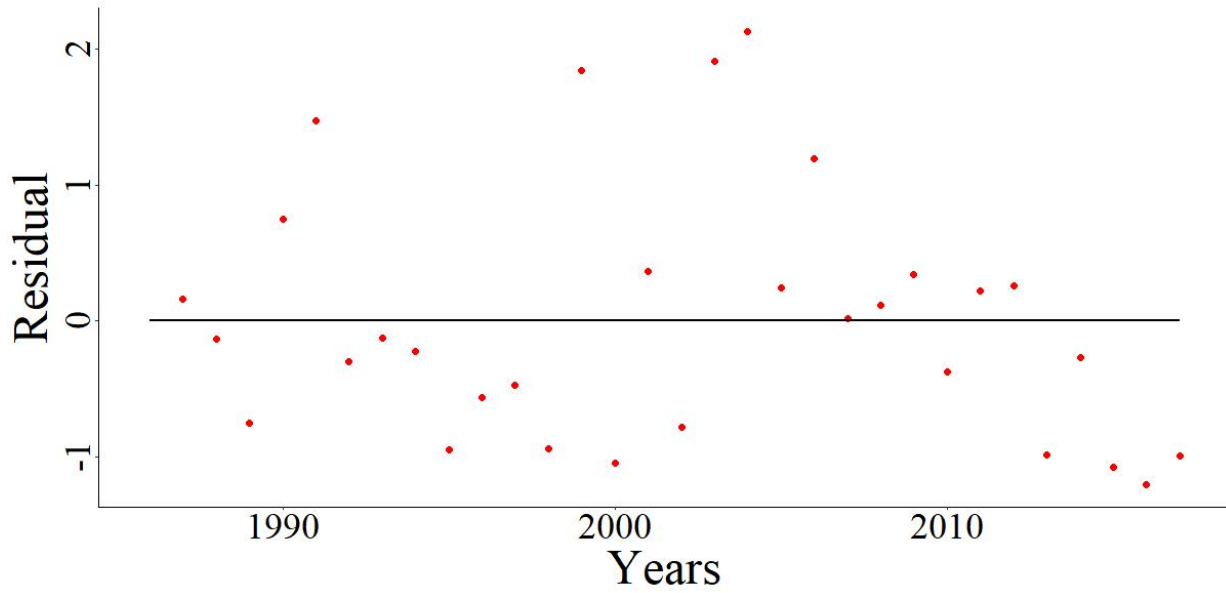


Figure 1C.4. One step ahead (OSA) residuals of total log-scale catch of the gill net fishery by year. Observed catch was modeled as a lognormal distribution with mean equal to the expected catch and standard deviation 0.067.

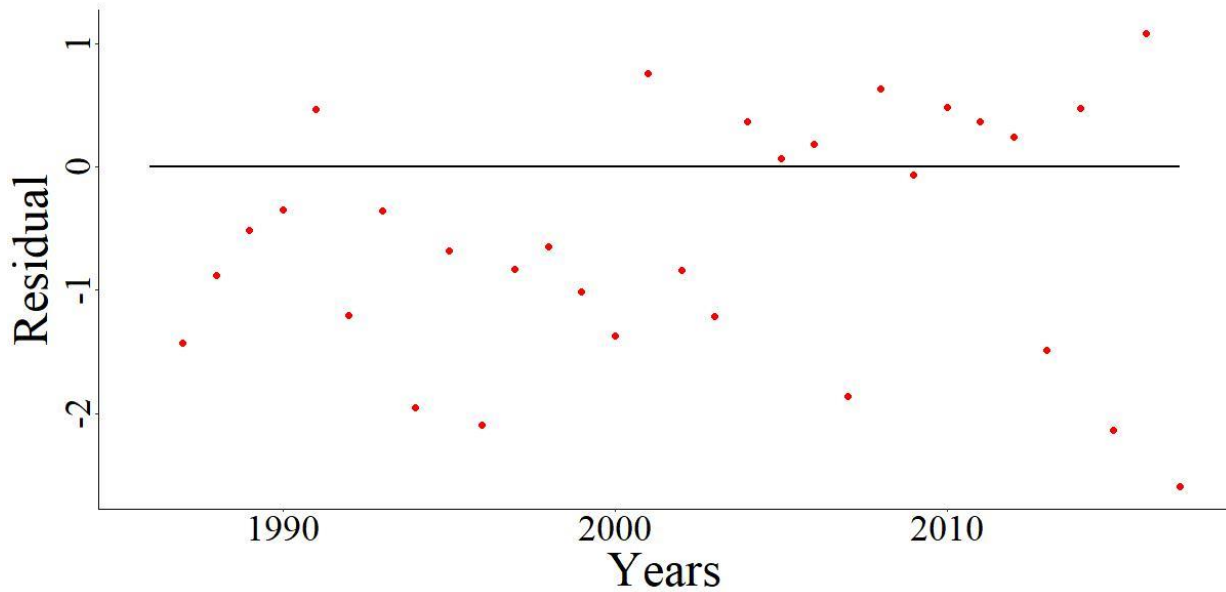




Figure 1C.5. One step ahead (OSA) residuals of proportions at age of the trap net fishery. Observed proportions were modeled as a multinomial distribution with year-specific effective sample size. No residuals were reported for year and age combination where the expected and observed value was 0 (older ages in the first 6 years) or for age 15+, because the way OSA residuals are calculated does not allow for estimates in the oldest age.

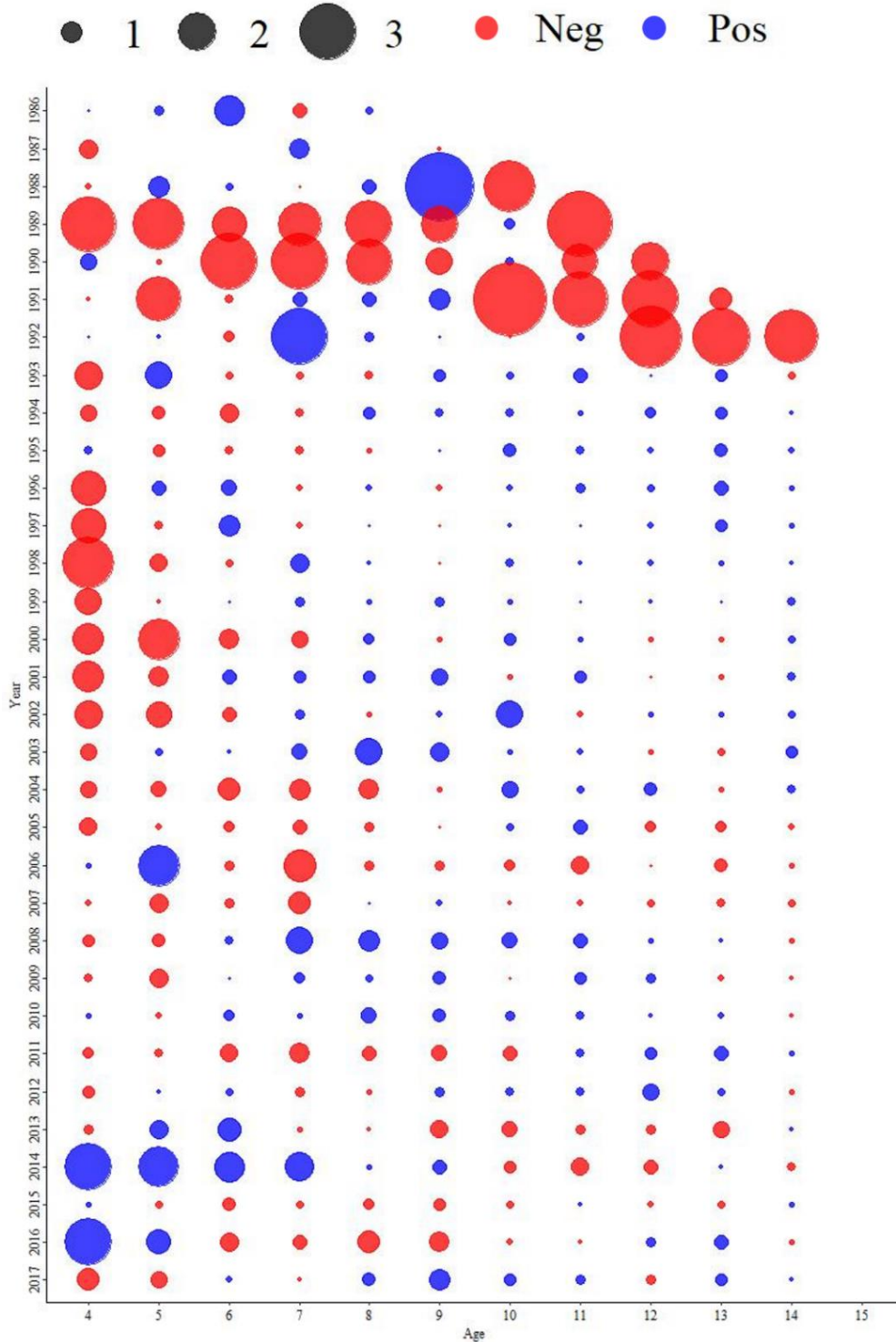




Figure 1C.6. One step ahead (OSA) residuals of proportions at age of the gill net fishery. Observed proportions were modeled as a multinomial distribution with year-specific effective sample size. No residuals were reported for year and age combination where the expected and observed value was 0 (older ages in the first 6 years) or for age 15+, because the way OSA residuals are calculated does not allow for estimates in the oldest age. Note that years 1986, 1989, and 1990 are also absent because the sample size was zero.

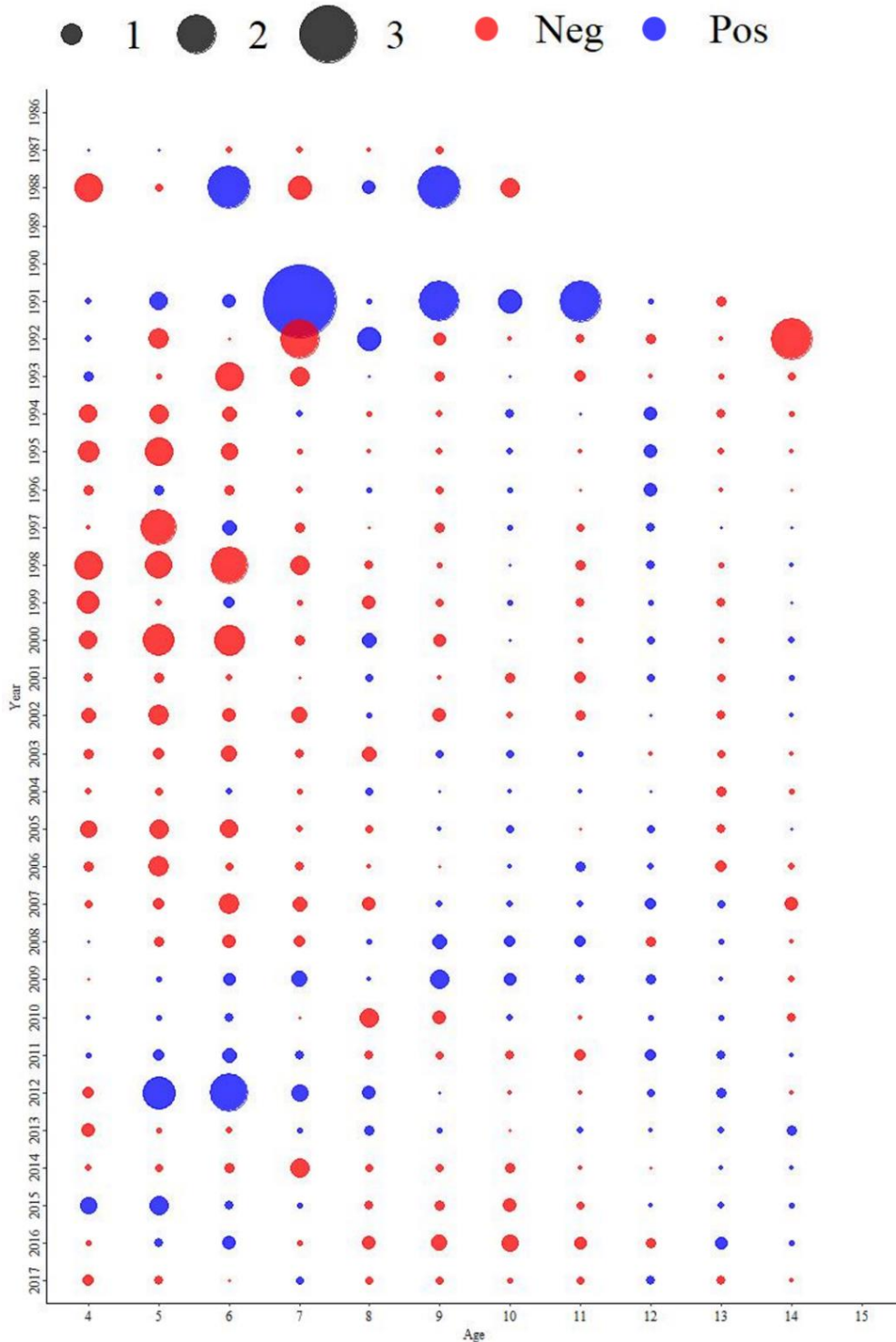


Figure 1C.7. Retrospective analysis of estimated recruitment in number of fish. The SSM model was refit after removing sequential years of the most recent observations. Dark colors indicate more complete data sets and lighter colors indicate less complete data sets.

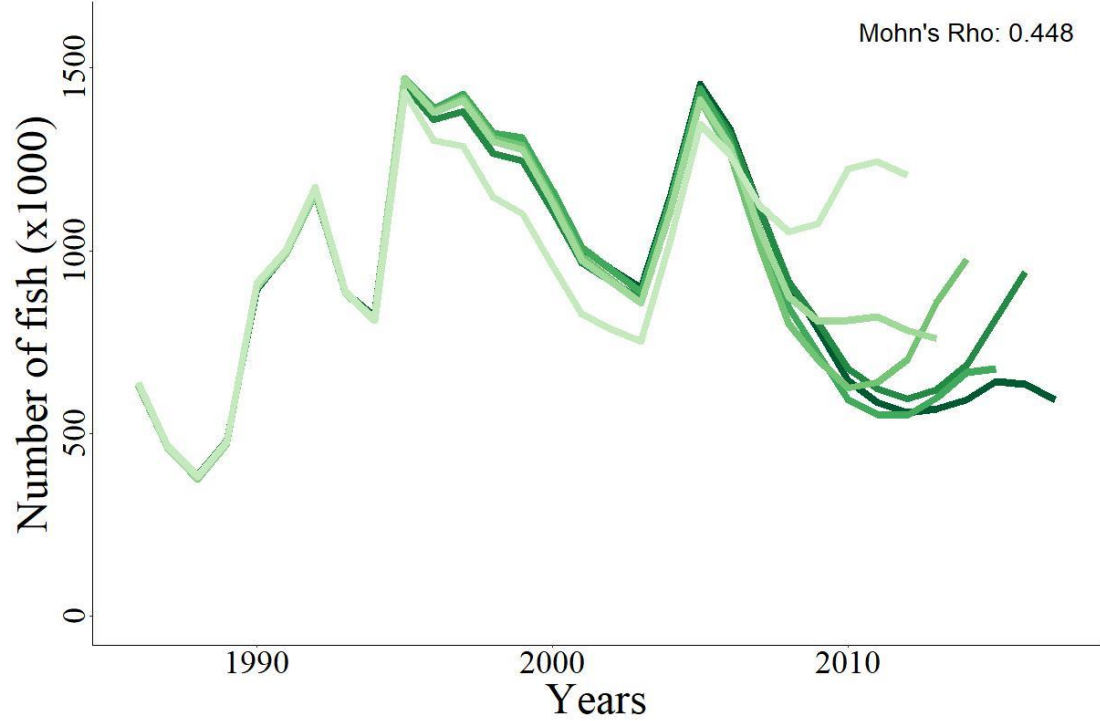
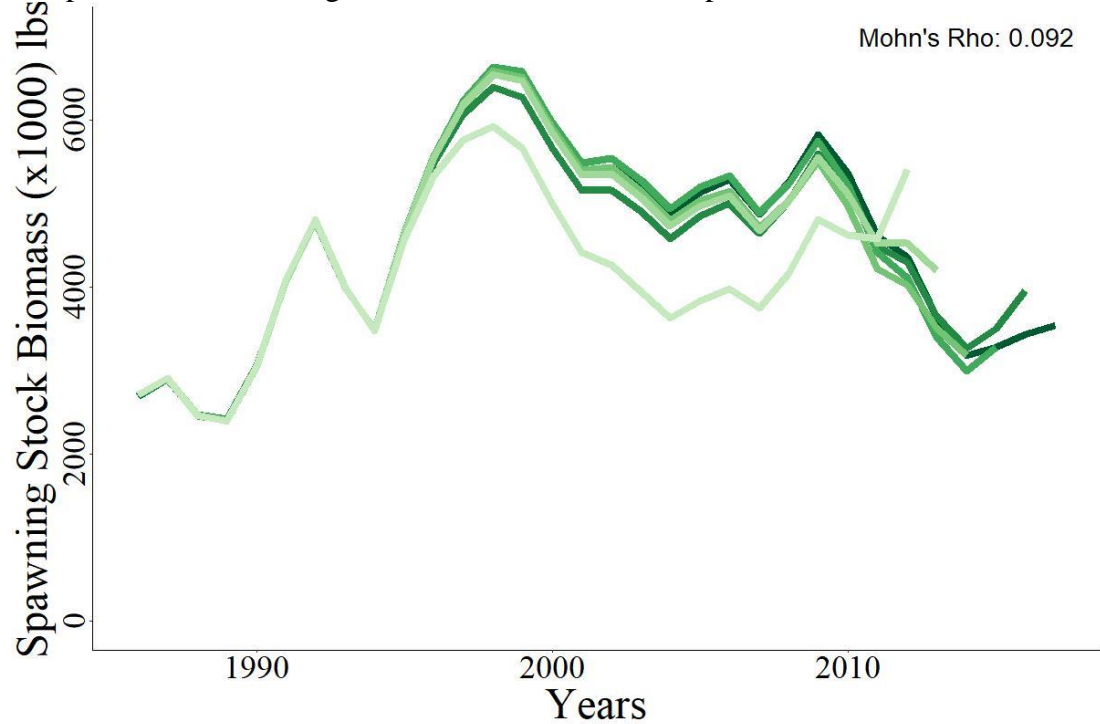


Figure 1C.8. Retrospective analysis of estimated spawning stock biomass. The SSM model was refit after removing sequential years of the most recent observations. Dark colors indicate more complete data sets and lighter colors indicate less complete data sets.



CHAPTER 2:  
THE EFFECT OF PROCESS VARIABILITY AND DATA QUALITY ON PERFORMANCE  
OF A STATE-SPACE STOCK ASSESSMENT MODEL

2.1 Abstract

State-space modeling is an emerging approach to age structured fisheries stock assessment that can accommodate multiple sources of variability in processes like recruitment, abundance, and selectivity. By maximizing the marginal likelihood by treating yearly deviations as random effects and then integrating them from the likelihood, these models can estimate multiple process variances. Several fisheries software packages have been developed that use a state-space framework with marginal likelihood, which has increased their popularity and usage across the U.S. Atlantic coast, Canada, and Europe. However, robust testing is still needed to gauge the applicability of these models and understand how they perform over a range of realistic variability in the process or observation error. Using an assessment model fit to Gulf of Maine haddock as a baseline, I used a simulation-estimation procedure to evaluate if state-space stock assessment models could produce unbiased and precise estimates over process variances that ranged from zero to well above the levels estimated in the Gulf of Maine haddock assessment, or when observations were noisier (i.e., more variable around their true value) than had been assumed in the assessment. I fit alternative estimation models that differed in which processes errors were included (of recruitment, expected survival, and fishery selectivity). State-space models that specify random effects in all three processes produced relatively unbiased and precise estimates of biomass and exploitation under most simulation scenarios and therefore are recommended except when variability in expected survival is absent (or very low), in which case the model is unlikely to converge. A conventional statistical-catch-at-age model with recruitment

estimated as a fixed effect for each year, deterministic expected survival, and constant selectivity produced estimates that were comparable to the best performing state-space model, but did not provide internal estimates of process variances and did not perform well when recruitment was highly variable. This work will facilitate the use of state-space stock assessment models and the choice of parameterization that will produce the most accurate output to inform future predictions and management.

## 2.2 Introduction

Fisheries stock assessment modelers have long worked to parse process variation from observation noise (Fournier and Archibald, 1982). Process variation (often “process error”) is the stochasticity in underlying biological state processes like recruitment, natural mortality, abundance, catchability, and selectivity. Observation noise (often “observation error”) are factors that create differences between what is real and what is measured such as deviations from random sampling. Process and observation error can be quantified using probability distributions and in a maximum likelihood estimation framework, those probability distributions are included in the likelihood, either added on the log-scale in a joint likelihood or errors-in-variables framework or integrated from the likelihood in the marginal likelihood (de Valpine and Hilborn, 2005). When maximizing marginal likelihood, it is at least theoretically possible to estimate multiple process variances and in some cases observation variances (Cadigan, 2015; Nielsen and Berg, 2014). Thanks to advances in fisheries stock assessment modeling software like Template Model Builder (TMB) that can approximate the high-level integrals, state-space stock assessment models that maximize the marginal likelihood and estimate many of the variances have emerged as a popular addition to the modeling toolbox (Aanes *et al.*, 2020; Kristensen *et al.*, 2016). In fact, state-space and the marginal likelihood have become largely synonymous in quantitative

fisheries discourse, though Bayesian methods still flourish, particular in state-space surplus production models (Han *et al.*, 2023; Soto *et al.*, 2022). Where once a stock assessment model could perhaps only account for stochasticity in one underlying process like recruitment or catchability, state-space modeling software like SAM or the Woods Hole Assessment Model (WHAM) now default to including variability in several underlying processes (Johnson *et al.*, 2014; Linton and Bence, 2008; Nielsen and Berg, 2014; Stock and Miller, 2021).

State-space stock assessment models (i.e., those that use a marginal likelihood) have several demonstrated advantages over non-state-space assessment models. They can more realistically represent dynamics by including stochasticity in recruitment, abundance, selectivity, and natural mortality, and allowing these to vary among ages, years, and region (Perreault *et al.*, 2020; Thorson, 2019; Trijoulet *et al.*, 2020). This increased realism could decrease bias caused by misspecification (Deroba and Schueller, 2013). Specifying variation as a random effect in a marginal likelihood will decrease bias even further (Thorson *et al.*, 2019). With such a framework, the standard deviation of the probability distribution of these random effects can be estimated as a parameter in the model, rather than being fixed *a priori* or by calculating the distribution of independent fixed effect parameters after the model is fit. Accurately estimating the variance of recruitment is especially important because it can inform our understanding of a species extinction probability or guide rebuilding plans, and estimating it within a statistical model can propagate uncertainty (Maunder and Deriso, 2003; Thorson *et al.*, 2014a). Since they account for more of the process variability that actually occurs, state-space models should exhibit less retrospective patterns than non-state-space models (Hurtado-Ferro *et al.*, 2015; Perreault *et al.*, 2020; Stewart and Martell, 2014; Stock *et al.*, 2021; Szuwalski *et al.*, 2018). The additional

variation may also resolve data conflicts in integrated assessments (Maunder and Piner, 2017; Stewart and Monnahan, 2017).

As stock assessment practitioners move to adopt state-space methods it is important to understand how well these models perform at producing unbiased and precise estimates of quantities like spawning stock biomass, recruitment, and exploitation, if the process variances can be estimated under a broad range of true values, and if they reduce retrospective patterns (the differences in model output when the most recent years are included in the analysis).

My objective was to evaluate the performance of state-space stock assessment models using a simulation-estimation procedure wherein data were generated with different levels of variability in recruitment, numbers at age or “expected survival” (given the fishing mortality), and fishery selectivity, ranging from no variability (i.e., constant over time) to variability that was two times the value estimated from fitting to real Gulf of Maine haddock catch and index data. In a separate procedure, I tested two different levels of observation error by generating data with higher observation error variability in total catch, the survey indices, or the proportions at age of the catch or indices set at higher levels than in the baseline Gulf of Maine haddock assessment. I fit each simulated dataset with a suite of alternative estimation models that made different assumptions about the existence and nature of the variability in the three processes of recruitment, expected survival, and selectivity.

## 2.3 Methods

### 2.3.1 Overview

My approach was to generate datasets using an operating model and fit alternative estimation models to each simulated dataset. I simulated 100 datasets for each scenario, which were defined by the variance of the process errors assumed for recruitment, expected survival, and selectivity, or the observation variance in total catch and index data and the “effective sample size” describing the observation noise in the respective proportions at age. The baseline estimation model was based on a state-space stock assessment model fit to Gulf of Maine haddock (*Melanogrammus aeglefinus*). In alternative estimation models I altered the treatment of process errors (including leaving out process errors in some or all the processes). The structure of the simulation model, as well as the baseline values of process error variances and the values of other parameters was also based on the state-space stock assessment model fit to Gulf of Maine haddock.

### 2.3.2 Determination of Baseline Estimation Model

The baseline estimation model was an age-based state-space stock assessment model, adapted from the statistical catch at age model used to assess Gulf of Maine haddock using the program ASAP (Legault and Restrepo, 1998). Description of symbols are provided in Table 2.1. The state-space stock assessment model was built using the Woods Hole Assessment Model (WHAM) package which is an R package dependent on Template Model Builder (TMB) (Kristensen et al., 2016; Stock and Miller, 2021). WHAM offers several options for parameterizing process error in recruitment, numbers at age (i.e., expected survival), selectivity, and other state variables that were not explored in the current study. The process errors can be specified as random effects and integrated from the joint likelihood of data and random effects

(i.e., the marginal likelihood). Recruitment can be parameterized as a random walk or random about a mean, and deviations from the previous year’s value or mean, respectively, can be independent and identically distributed (“iid”) or correlated among years. Age- and year- specific deviations in expected survival or mean selectivity can be iid, correlated across years or ages in a 1-dimensional autoregressive process (“AR1\_y” or “AR1\_a”) or both ages and years simultaneously in a 2-dimensional autoregressive process (“2DAR1”). Refer to Stock and Miller (2021) for a full description of WHAM and its many parameterization options.

Table 2.1. Symbols used in the text, tables, and figures. For indices, time and age invariant data, and some simple derived quantities, the value is provided.

Symbol	Description	Usage
<b>Index Variables</b>		
$y, \tilde{y}$	Year	1977-2019
$Y$	Terminal Year	2019
$a, \tilde{a}$	Age	1-8
$A$	Terminal age plus group	9+
$i$	Survey Index	F (Fall), S (Spring)
<b>Data and Inputs</b>		
$C_y$	Total observed catch by year	
$P_{a,y}$	Observed proportion at age by year in the catch	
$I_{y,i}$	Observed total index value by year and survey	
$p_{a,y,i}$	Observed proportion at age by year and survey of the index	
$M$	Instantaneous natural mortality	0.2
$m_{a,y}$	Maturity at age and year	
$W_{a,y}$	Weight at age and year at time of capture	
$W_{a,y}^{(S)}$	Weight at age and year at time of spawning	
$f$	Fraction of the year that passes before spawning occurs	0.25
$N_{7:9+,1977}$	Numbers at age (in thousands of individuals) for the oldest age groups in the first year	[5,1,20]
$S_{4:9+,F}$	Selectivity of ages 4-9+ in the fall survey	[1,1,1,1,1,1]
$S_{5:9+,S}$	Selectivity of ages 5-9+ in the spring survey	[0.9,1,1,1,1]
$\sigma_{C_y}$	Standard deviation of log-normally distributed observation error of total catch	See Appendix 2A
$\sigma_{I_{y,i}}^2$	Standard deviation of log-normally distributed observation error of index values	See Appendix 2A
$n_{y,i}^{(I)}$	Effective sample size of multinomial distribution of observation error of proportions at age of the indices	See Appendix 2A
<b>States/Random Effects</b>		
$\varepsilon_y^{(Rec)}$	Deviations from mean recruitment	
$\varepsilon_{a,y}^{(Sur)}$	Deviations from expected abundance at age, which is 0	
$\varepsilon_{a,y}^{(Sel)}$	Deviations from mean selectivity at age	



Table 2.1 (cont'd)

<b>Fixed Effect Parameters</b>	
$\bar{R}$	Mean recruitment
$N_{1:6,1977}$	Numbers at age for the youngest age groups in the first year
$\bar{s}_a$	Mean selectivity at age of the catch
$F_y$	Fishing mortality at year
$\sigma_R$	Standard deviation of process error around mean recruitment
$\sigma_{Sur}$	Standard deviation of process error around survival
$\sigma_{Sel}$	Standard deviation of process error around selectivity at age of catch
$\rho_{NAA}$	Correlation coefficient of process errors around recruitment and survival, which is correlated among years
$\rho_{Sel}$	Correlation coefficient of process errors around selectivity which is correlated among years
$q_i$	Catchability of indices
$S_{a,i}$	Selectivity of indices at age
<b>Derived Quantities</b>	
$N_{a,1978:2019}$	Abundance (“numbers at age”) in thousands
$F_y$	Fully selected instantaneous fishing mortality
$F_{a,y}$	Age-specific instantaneous fishing mortality
$s_{a,y}$	Selectivity at age and year
$\hat{C}_{a,y}$	Expected catch at age in numbers
$\hat{C}_y$	Expected total catch in weight
$\hat{P}_{a,y}$	Expected proportions at age of the catch
$\hat{I}_{a,y,i}$	Expected fisheries independent index value at age
$\hat{I}_{y,i}$	Expected total fisheries independent index value
$\hat{P}_{a,y,i}$	Expected proportions at age of the index
$B_y^{(S)}$	Spawning stock biomass
$E_y$	Exploitation rate
$T$	Average spawning stock biomass trend

To determine the structure of the baseline estimation model, the real Gulf of Maine haddock data were fit using several parameterizations of the WHAM model, each with different combinations of process error structure in recruitment, expected survival, or selectivity, from those options described above, as well as for several options for the age composition likelihood (multinomial, Dirichlet-multinomial, and logistic normal). The Gulf of Maine haddock data includes annual total biomass of landings (for commercial and recreational fisheries combined), survey indices from the spring and fall Northeast Fisheries Science Center (NEFSC) bottom trawl surveys, and the numerical proportion at age of those landings and surveys (from 1-9+)

from 1977 through 2019 (NEFSC, 2022b). Other input data included biological information like natural mortality, maturity at age, and weight at age at time of catch or spawning. The estimation model was chosen based on the minimum Akaike's Information Criteria (AIC) value (Akaike, 1998), and by robustness in generating and fitting simulated data.

In all the estimation models fit using the Gulf of Maine haddock data, observation error variances for total annual landings and annual survey indices of total abundance were pre-specified and treated as known. Similarly, effective sample sizes of the multinomial distribution for the fishery and survey proportions at age were pre-specified and not estimated. WHAM has optional parameters that can scale the input year-specific standard deviation of the lognormal distribution of total annual catch or annual survey indices. However, when I tried to estimate these parameters, the model failed to converge. The effective sample size of the multinomial distribution cannot be estimated within the model and I did not use iterative reweighting when fitting the haddock data or simulated data in WHAM because it is computationally costly and particularly challenging in state-space models (Maunder, 2011). The effective sample sizes were premised on an iterative reweighting procedure following Francis (2011) used during the haddock assessment using ASAP, but the values were not re-estimated for the WHAM model (NEFSC, 2022b). Alternative age composition likelihoods like the Dirichlet-multinomial or logistic normal do have weighting parameters, but both performed poorly at generating data that could be fit using the estimation model, likely owing to too much process and observation error variability. The values that determined observation error variance- the standard deviation of the lognormal distribution and the effective sample size of the multinomial, - were fixed at the values used in the original ASAP assessment model (NEFSC, 2022b).

In the selected estimation model, recruitment (i.e., the number of age 1 individuals) in each year,  $N_{1,y}$  varied about the mean recruitment following a lognormal distribution and AR(1) process (Table 2.2, Eqs. 2.1-2.2). Under this model, deviations in a given year would tend to be more similar to those in years immediately prior or subsequent (Eq. 2.2). The model estimated the mean recruitment,  $\bar{R}$ , standard deviation,  $\sigma_R$ , and correlation coefficient,  $\rho_{NAA}$ , as fixed effects, and the deviations from the mean,  $\varepsilon_y^{(Rec)}$ , as random effects. Log-scale abundance in the initial year,  $\log N_{a,1977}$ , was estimated individually for ages 1-6 and fixed for ages 7-9+. Fixing the abundance for the older ages in the initial year improved model stability and parameter estimability when fitting to both real and simulated data.

Table 2.2. Equations used in the estimation and simulation models.

Index	Description	Equation
<b>Process Model</b>		
2.1	Yearly recruitment is a function of mean recruitment and random process error	$\log N_{1,y} = \log \bar{R} + \varepsilon_y^{(Rec)}, y > 1977$
2.2	Recruitment deviations are randomly distributed as stationary AR(1)	$\varepsilon_y^{(Rec)} \sim N\left(\rho_{NAA}\varepsilon_{y-1}^{(Rec)} - \frac{\sigma_R^2}{2(1-\rho_{NAA}^2)}, \sigma_R^2\right)$
2.3	Fishing mortality at age and year is the product of fully selected F, and age and year specific selectivity	$F_{a,y} = F_y s_{a,y}, a < 7$
		$F_{a,y} = F_y, a \geq 7$
2.4	Selectivity at age and year are the sum of a mean parameter for each age and deviations, transformed on the logit scale	$s_{a,y} = \frac{1}{1 + \exp(-(\text{logit}(\bar{s}_a) + \varepsilon_{a,y}^{(Sel)}))}$
2.5	Selectivity at age deviations are distributed as stationary AR(1)	$\varepsilon_{a,y}^{(Sel)} \sim N\left(\rho_{Sel}\varepsilon_{a,y-1}^{(Sel)} - \frac{\sigma_{Sel}^2}{2(1-\rho_{Sel}^2)}, \sigma_{Sel}^2\right), a < 7$
2.6	Age- and year-specific abundance is a function of abundance in the previous age and year, survival, and random process error	$\log N_{a,y} = \log(N_{a-1,y-1}) - F_{a-1,y-1} - M + \varepsilon_{a,y}^{(Sur)}, 1 < a < A$ $\log N_{A,y} = \log(N_{A-1,y-1}e^{-F_{A-1,y-1}-M} + N_{A,y-1}e^{-F_{A,y-1}-M})$ $+ \varepsilon_{A,y}^{(Sur)}, a = A$

Table 2.2 (cont'd)

2.7	Numbers at age deviations are distributed as stationary AR(1)	$\varepsilon_{a,y}^{(Sur)} \sim N\left(\rho_{NAA}\varepsilon_{a,y-1}^{(Sur)} - \frac{\sigma_{Sur}^2}{2(1-\rho_{NAA}^2)}, \sigma_{Sur}^2\right), a > 1$
<b>Observation Model</b>		
2.8	Expected catch at age is a function of abundance and mortality	$\hat{C}_{a,y} = N_{a,y}(1 - e^{-M-F_{a,y}}) \frac{F_{a,y}}{F_{a,y}+M}$
2.9	Expected total catch in weight is a function of catch at age and weight at age and year	$\hat{C}_y = \sum_{a=1}^A \hat{C}_{a,y} W_{a,y}$
2.10	Expected catch proportion at age is a function of age specific values and total	$\hat{p}_{a,y} = \frac{\hat{C}_{a,y}}{\sum_{a=1}^A \hat{C}_{a,y}}$
2.11	Expected survey index in weight for each age, year, and index	$\hat{I}_{a,y,i} = q_i s_{a,i} N_{a,y} e^{(-F_{a,y}-M)f_{y,i}}$
2.12	Expected total survey index in weight is a sum of age-specific values	$\hat{I}_{y,i} = \sum_{a=1}^A \hat{I}_{a,y,i}$
2.13	Expected index proportion at age is a function of age specific values and total	$\hat{p}_{a,y,i} = \frac{\hat{I}_{a,y,i}}{\hat{I}_{y,i}}$
<b>Likelihood Components</b>		
2.14	Observed total catch is derived from the expected values and lognormal error	$\log(C_y) \sim N(\log(\hat{C}_y) - \frac{\sigma_{C_y}^2}{2}, \sigma_{C_y}^2)$
2.15	Observed catch proportion at age is derived from multinomial distribution with input effective sample size	$P_{a,y} \sim MN(\hat{p}_{a,y}, n_y^{(C)})$
2.16	Observed total survey index is derived from expected values and lognormal error	$\log(I_{y,i}) \sim N(\log(\hat{I}_{y,i}) - \frac{\sigma_{I_{y,i}}^2}{2}, \sigma_{I_{y,i}}^2)$

Table 2.2 (cont'd)

2.17	Observed index proportion at age is derived from multinomial distribution with input effective sample size	$p_{a,y,i} \sim MN(\hat{p}_{a,y,i}, n_{y,i}^{(I)})$
<b>Derived Quantities and Simulation Output</b>		
2.18	Average Recruitment in terminal 3 years	$\bar{N}_{1,(Y-3:Y)} = \frac{\sum_{y=Y-3}^Y N_{1,y}}{3}$
2.19	Spawning stock biomass	$B_y^{(S)} = \sum_{a=1}^A N_{a,y} W_{a,y}^{(S)} m_{a,y} e^{f(-M-F_{a,y})}$
2.20	Exploitation rate of stock	$E_y = \frac{\hat{C}_y}{\sum_{a=1}^A N_{a,y} W_{a,y}}$
2.21	Average spawning stock biomass trend in recent 10 years	$D = \frac{B_T^{(S)}}{B_{T-10}^{(S)}}$
2.22	Relative Error of a performance metric	$RE_x = \frac{\hat{x} - \tilde{x}}{\hat{x}}$

In the selected estimation model, fully selected fishing mortality in each year,  $F_y$ , was a fixed effect parameter. Fishing mortality at age and year,  $F_{a,y}$ , for  $a < 7$  was the product of the fully selected value,  $F_y$  and selectivity at age and year,  $s_{a,y}$  (Eq. 2.3). For age 7 and above  $F_{a,y} = F_y$ . Selectivity at age and year was a function of an age specific mean,  $\bar{s}_a$ , and age- and year-specific selectivity deviations,  $\varepsilon_{a,y}^{(Sel)}$ , and constrained to be between zero and one (Eq. 2.4). Deviations from the mean selectivity on the logit scale,  $\varepsilon_{a,y}^{(Sel)}$  were random effects correlated among years but not among ages (Eq. 2.5). The standard deviation of this distribution,  $\sigma_{Sel}$ , and the correlation coefficient,  $\rho_{Sel}$ , were fixed effects.

After the initial year, abundance at age in the selected estimation model was a function of abundance at age of the previous age in the previous year, expected survival, itself a function of instantaneous fishing mortality,  $F_{a,y}$ , and natural mortality,  $M$ , and process error on expected survival,  $\varepsilon_{a,y}^{(Sur)}$  (Eq. 2.6). Abundance in the 9+ group also included the number of 9+ individuals

that survived since the previous year. This parameterization mimics how process errors in numbers at age were incorporated in the “full state-space model” described previously in Nielsen and Berg, (2014) and Miller *et al.* (2016). The expected survival process errors had an AR(1) structure with correlation among years but not ages. The standard deviation for the expected survival process error  $\sigma_{Sur}$ , was estimated as a fixed effect, and the correlation coefficient was the same as that for the recruitment process error,  $\rho_{NAA}$  (Eq. 2.7).

The predicted catch at age in numbers,  $\hat{C}_{a,y}$ , followed the Baranov catch equation (Eq. 2.8). The predicted total catch in weight (i.e., yield),  $\hat{C}_y$ , was a sum of the products of catch at age and weight at age,  $W_{a,y}$ , across ages (Eq. 2.9) The predicted proportions at age,  $\hat{P}_{a,y}$ , were derived from the total catch in numbers and catch at age (Eq. 2.10). The predicted survey index at age and year and survey was the product of abundance at age, survey-specific catchability,  $q_i$ , age- and survey- specific selectivity,  $s_{a,i}$ , and expected survival (adjusted for the fraction of the year that has elapsed at the time of the survey,  $f_{y,i}$ ) (Eq. 2.11). Ages 4-9 were assumed to be fully susceptible to survey gear in the fall and ages 5-9 were assumed to be fully susceptible to survey gear in the spring; hence survey selectivity was fixed at 1, apart from survey selectivity for age 5 individuals in the spring which was fixed at 0.9. Unlike for the fishery, age-specific selectivity for the surveys was time-invariant. The predicted total survey index value,  $\hat{I}_{y,i}$ , and predicted survey proportions at age,  $\hat{p}_{a,y,i}$ , were calculated from the age-specific indices in the same way as the fishery total catch and proportions were from the age-specific values (Eqs. 2.12-2.13).

The observed total catch in each year,  $C_{a,y}$ , was assumed to have a lognormal distribution with mean equal to the predicted catch and the year-specific standard deviation that was input as data,  $\sigma_{C_y}$  (Eq. 2.14, Appendix 2A). The observed proportions were assumed to have a

multinomial distribution with year specific effective sample sizes  $n_y^{(C)}$ , which were also input (Eq. 2.15, Appendix 2A). The observed total index values and proportions at age of the indices followed the same distributions (lognormal and multinomial) as total catch and catch proportions at age but had unique standard deviations and effective sample sizes, respectively that determined the degree of variation (Eq. 2.16-2.17, Appendix 2A).

The estimation model was fit by maximizing the marginal likelihood in which the random effects on recruitment, expected survival, and selectivity were integrated from the joint distribution of random effects and data. Therefore, the objective function was minimized by varying the fixed effects. Mean recruitment, numbers at age in the first age, fishing mortality, and process variance parameters were estimated on the log-scale and survey catchability and selectivity and AR(1) process error correlation coefficients were estimated on the logit scale. The predicted values for the random effects and derived quantities were those that maximized the joint likelihood of data and process variability, conditioned on the estimates of the fixed effects, using the epsilon method for bias correction (Thorson and Kristensen, 2016).

### *2.2.3 Simulation/Operating Model Scenarios*

Data were simulated using the fitted WHAM model described above. The simulated data had the same number of age groups and years, the same natural mortality, maturity, weight at age at capture and spawning, numbers at age in the oldest age group in the first year, and observation error variance of the fall and spring surveys as estimated or assumed for the Gulf of Maine haddock assessment. In all simulation scenarios, data were simulated using most of the same fixed effect parameter estimates, including the same mean recruitment, yearly fully selected fishing mortality, mean selectivity at age, numbers at age for ages 1-6 in the first year, catchability and selectivity of the index data, and correlation coefficients of the process errors in

numbers at age or selectivity. The difference among simulation scenarios was the process or observation error variance or the effective sample size used in the simulation model. In the baseline simulation scenario, random effects on recruitment,  $\varepsilon_y^{(Rec)}$ , expected survival,  $\varepsilon_{a,y}^{(Sur)}$ , and selectivity,  $\varepsilon_{a,y}^{(Sel)}$ , were drawn from their respective distributions using the mean (for recruitment and selectivity), correlation coefficients, and standard deviations estimated in the model fit to Gulf of Maine haddock. The randomly generated values of deviations in yearly recruitment, survival, and selectivity then informed the realized catch and index values,  $\tilde{C}_y$ ,  $\tilde{P}_{a,y}$ ,  $\tilde{I}_y$ , and  $\tilde{p}_{a,y}$  (i.e., Eqs. 2.8-2.13 but  $\tilde{X}$  rather than  $\hat{X}$ ) and the simulated catch and index values were a function of these realized values and observation error (i.e., Eqs. 2.14-2.17, but  $\tilde{X}$  rather than  $\hat{X}$ ).

Each alternative simulation scenario assumed that one of the process or observation error variance values differed from the baseline (Table 2.3), while the others were kept at their baseline values. The first set of simulation scenarios addressed how increases or decreases in the variance of process error in recruitment, expected survival, or selectivity affected model performance (Table 2.3A). The four alternative variances used for each source of process variability are categorized as “very low,” “low,” “high,” and “very high.” The “very low” simulation scenario meant the variance was set to 0. For the recruitment or age-specific selectivity processes, this meant that the year-specific values equaled the mean,  $\bar{R}$  or  $\bar{s}_a$ , respectively. For the expected survival process, this meant that the survival was deterministic and abundance at age did not deviate from what was expected given  $F_{a,y}$  and  $M$  and abundance at age in the previous year. The “low” simulation scenario meant the standard deviation of each process error distribution was half of the baseline value. “High” and “very high” scenarios for expected survival or selectivity meant the standard deviation was double or three times the baseline value. However, because the estimated standard deviation of recruitment process was



already quite high (1.17) the “high” and “very high” treatments were constrained to 1.5 and 2, which are 1.28 and 1.7 times the baseline value, respectively. Because there were four different levels of process error standard deviation for each of the three types of process errors, plus baseline levels, there were 13 total simulation scenarios for the process error (including the baseline).

Table 2.3. Process error standard deviations in simulation scenarios (A) and multiplier for year- and index- specific observation error standard deviation (total catch and index value) and effective sample sizes (proportions at age) (B). Only increases in observation error variance were tested as these reflected increased “noisiness” in the simulated data and the consequences of having poor data.

(A)

Process Error	Base	Very Low	Low	High	Very High
$\sigma_R$	1.17	0.00	0.58	1.50	2.00
$\sigma_{Sur}$	0.19	0.00	0.10	0.40	0.60
$\sigma_{Sel}$	0.30	0.00	0.15	0.60	0.90

(B)

Observation Error	Base	High	Very High
$\sigma_{C_y}$	x1.0	x1.1	x1.5
$\sigma_{I_{y,i}}$	x1.0	x1.1	x1.5
$n_y^{(C)}$	x1.0	x0.9	x0.5
$n_{y,i}^{(I)}$	x1.0	x0.9	x0.5

The second set of simulation scenarios involved decreasing the “quality” (i.e., increasing the “noisiness”) of the simulated data by either increasing the observation error standard deviation of the total catch or total index, or decreasing the effective sample size of the proportions at age of the catch or index, both of which assume that the observed data can deviate more from the predicted values than is assumed in the baseline model (Table 2.3B). Observation error variance or effective sample size was adjusted accordingly in the estimation models, so the data quality was poorer but correctly specified. The process error standard deviations were kept at the baseline values. The standard deviation of the lognormal error around total catch or total

index value (for both the spring and fall surveys), which were year-specific values, were increased by a factor of 1.1 or 1.5 for “high” or “very high” treatments, respectively. The effective sample size of the multinomial distribution for the catch proportions at age or index proportions at age (for both the spring and fall surveys), which were also year-specific values, were decreased by a factor of 0.9 or 0.5 for a “high” or “very high” treatment, respectively. As in the process error simulation scenarios, each component was tested individually, so in total there were nine simulation scenarios for observation error (including the baseline).

#### *2.3.4 Alternative Estimation Model Scenarios*

The estimation model described in section 2.3.2 was fit to simulated data. The estimation model specified yearly deviations in recruitment, expected survival, and selectivity as correlated random effects. Seven other estimation models were also used to fit the data, each which specified one or more of these processes as a fixed effect or constant over time, rather than a random effect. Each of the eight versions of the estimation model were designated by a three-letter code that indicate whether the recruitment, expected survival, and selectivity processes were modeled as a random effect (R), fixed effect (F) or constant (C), in that order. For example, the estimation model described in section 2.3.2 was designated “RRR” and the model that assumed recruitment and expected survival were random effects but selectivity was constant was “RRC”. When recruitment was a fixed effect (F), the year-specific deviations were estimated parameters and their distribution was not constrained with a probability function. When a value was constant, the year-specific deviations were fixed at zero. When applied to expected survival, “constant” means no variation in the survival process, so survival is deterministic and abundance is only a function of abundance in the previous year and total mortality. Not every combination of random effect, fixed effect, or constant parameterization could be tested because of the

structure of WHAM. The intent with fitting these additional alternative versions of the estimation model was to examine the consequences of ignoring process error when it is present (in most cases) or trying to estimate it when it is absent (in the “very low” scenarios). The “FCC” model, which specifies yearly recruitment as a fixed effect parameter, expected survival as deterministic, and selectivity as constant, mimics the statistical catch-at-age model (SCAA) currently employed to assess many fisheries in the northwest Atlantic using program ASAP. Within each simulation scenario of process error variability or data quality, 100 data sets were simulated and fit using the eight estimation models. If the model did not converge, and/or did not result in an invertible Hessian matrix, the model was considered a failure and excluded from the analysis. Model performance was also evaluated by retrospective analysis where retrospective patterns were quantified by Mohn’s Rho calculated for spawning stock biomass, fishing mortality, and recruitment (Hurtado-Ferro *et al.*, 2015).

#### 2.4.5 Performance of Assessment Models

Several metrics were used to evaluate model performance in fitting to simulated data, each of which reflected values of interest in fisheries management. Mean recruitment in the terminal three years,  $\bar{N}_{1,(Y-3:Y)}$ , spawning stock biomass in the terminal year,  $B_Y^{(S)}$ , exploitation rate in the terminal year,  $E_Y$ , and depletion in spawning stock biomass in the terminal 10 years,  $D$ , were compared against the true values used to simulate the data (Eqs. 2.18-2.21). Depletion was defined as the ratio between spawning stock biomass in the terminal year and spawning stock biomass 10 years prior to the terminal year (Eq. 2.21). Relative error of a given model performance metric  $x$  was the difference between the estimated,  $\hat{x}$ , and realized,  $\tilde{x}$ , values relative to the realized value (Eq. 2.22).

For estimation models that assumed one or more process was a random effect the distributions of the estimated standard deviation and correlation coefficient of the AR(1) process of the random effect were compared against the true value used to generate the data. Convergence rate, the bias (systemic over- or under- estimation of a quantity), and precision (spread of relative error) were compared across treatments. For each of the thirteen simulation scenarios for process error variability, and nine simulation scenarios for observation error variability, simulated data were fit using eight estimation models. Thus, there were 104 simulation/estimation combinations for the process variability scenarios and 72 such combinations for the observation error scenarios.

## 2.4 Results

The median convergence rate across the 104 simulation/estimation combinations for process error scenarios was 90% and ranged from 7-100% (Table 2.4A). Among process variability scenarios, average convergence rate was highest when the recruitment variability was very low (94%) or low (95%) and lowest when the recruitment variability was high (73%) or expected survival variability was very low (74%) or very high (76%). Models converged most frequently when recruitment included a random effect, expected survival was deterministic, and selectivity was constant (RCC) (98%), and least frequently when recruitment was constant, survival was deterministic, and selectivity included a random effect (CCR) (76%). The convergence rate was approximately the same across observation error simulation scenarios, as were the relative convergence rates for the different estimation models (Table 2.4B).

Table 2.4. Convergence rate in process variability simulation scenarios (A) and observation error simulation scenarios (B) across 100 simulated data sets. The criteria for convergence were that the model had to complete the optimization without error and all the parameters and standard error values had to be estimated (i.e., the Hessian matrix was invertible).

(A) SIMULATION SCENARIO			ESTIMATION MODELS								Average	Median	Min	Max
Recruitment	Survival	Selectivity	RRR	RRC	RCR	RCC	FCR	FCC	CCR	CCC				
Baseline	Baseline	Baseline	94	100	93	96	81	88	82	84	89.75	90.5	81	100
Very High	Baseline	Baseline	90	100	94	97	38	46	61	65	73.88	77.5	38	100
High	Baseline	Baseline	91	100	94	98	71	77	65	77	84.13	84	65	100
Low	Baseline	Baseline	94	100	99	98	94	99	84	97	95.63	97.5	84	100
Very Low	Baseline	Baseline	91	93	94	98	90	98	97	98	94.88	95.5	90	98
Baseline	Very High	Baseline	87	98	90	90	54	59	59	73	76.25	80	54	98
Baseline	High	Baseline	86	98	89	98	74	87	77	81	86.25	86.5	74	98
Baseline	Low	Baseline	54	84	95	99	85	94	81	89	85.13	87	54	99
Baseline	Very Low	Baseline	7	48	98	99	82	96	79	86	74.38	84	7	99
Baseline	Baseline	Very High	97	98	95	95	92	90	77	85	91.13	93.5	77	98
Baseline	Baseline	High	92	100	96	100	88	92	73	82	90.38	92	73	100
Baseline	Baseline	Low	84	98	94	100	77	93	80	84	88.75	88.5	77	100
Baseline	Baseline	Very Low	85	98	86	100	64	91	72	82	84.75	85.5	64	100
<b>Average</b>			80.92	93.46	93.62	97.54	76.15	85.38	75.92	83.31				
<b>Median</b>			90	98	94	98	81	91	77	84				
<b>Min</b>			7	48	86	90	38	46	59	65				
<b>Max</b>			97	100	99	100	94	99	97	98				

Table 2.4 (cont'd)

(B) SIMULATION SCENARIO				ESTIMATION MODELS											
Total Catch	Catch	Total Index	Index Proportions	RRR	RRC	RCR	RCC	FCR	FCC	CCR	CCC	Average	Median	Min	Max
Baseline	Baseline	Baseline	Baseline	94	100	93	96	81	88	82	84	89.75	90.5	81	100
Very High	Baseline	Baseline	Baseline	97	100	94	98	84	93	87	93	93.25	93.5	84	100
High	Baseline	Baseline	Baseline	90	100	93	97	83	88	80	86	89.63	89	80	100
Baseline	Very High	Baseline	Baseline	94	99	91	98	77	88	80	86	89.13	89.5	77	99
Baseline	High	Baseline	Baseline	90	97	93	97	76	84	77	81	86.88	87	76	97
Baseline	Baseline	Very Low	Baseline	84	95	84	100	63	86	56	84	81.5	84	56	100
Baseline	Baseline	Low	Baseline	91	98	90	97	86	90	75	82	88.63	90	75	98
Baseline	Baseline	Baseline	Very Low	82	97	93	98	75	82	77	80	85.5	82	75	98
Baseline	Baseline	Baseline	Low	93	98	99	98	77	89	71	88	89.13	91	71	99
<b>Average</b>				90.56	98.22	92.22	97.67	78	87.56	76.11	84.89				
<b>Median</b>				91	98	93	98	77	88	77	84				
<b>Min</b>				82	95	84	96	63	82	56	80				
<b>Max</b>				97	100	99	100	86	93	87	93				

Spawning stock biomass in the terminal year was relatively unbiased across all estimation models in all process variance simulation scenarios (Figure 2.1). Precision was lower in estimation models that assumed recruitment was constant (C--), except when the true recruitment variance was zero (left column, bottom row). Spawning stock biomass was estimated with low precision when the variation in expected survival was high or very high (middle column, top two rows), except when expected survival was included as a random effect.

Average recruitment in the terminal three years was relatively unbiased across all estimation models in all process variance simulation scenarios but precision was lower in estimation models that assumed constant recruitment (Figure 2.2). When variability in selectivity was very high (right column, top row) and the estimation model incorrectly assumed it was constant and treated recruitment and expected survival as random effects (RRC), average recruitment was positively biased. The most precise and relatively unbiased estimates of average recruitment in the terminal three years were in cases where the true recruitment variance was zero (left column, bottom row).

Exploitation rate in the terminal year was relatively unbiased except when the model incorrectly assumed recruitment was constant in which case it was underestimated (Figure 2.3). Exploitation rate was underestimated when selectivity variability was very high (right column, top row) and the model assumed it was constant and recruitment and expected survival were random effects (RRC), suggesting a tradeoff between exploitation rate and recruitment estimates. Precision in estimated exploitation rate was low when expected survival process variance was high or very high (middle column, top two rows) and the model assumed the expected survival was deterministic. Depletion was estimated without bias across all estimation models and process variance scenarios (Figure 2.4). Depletion had low precision when recruitment was

assumed to be constant and high precision when the variance in expected survival was zero (middle column, bottom row).

Recruitment process error standard deviation was estimated without substantial bias in most estimation models and process variance simulation scenarios (Figure 2.5). When the true recruitment variance was zero or very low (left column, bottom two rows), the standard deviation was unbiased, as long as the estimation model assumed process error in expected survival, otherwise the variance was overestimated. The recruitment standard deviation was also overestimated when the true variation in expected survival was high or very high (middle column, top two rows) and the estimation model assumed the expected survival was deterministic. The survival process error standard deviation was underestimated in almost all scenarios, and only substantially overestimated when the true expected survival variance was zero and the model assumed expected survival process variation was present (Figure 2.6). The recruitment and expected survival process error correlation coefficient, which quantifies inter-year correlations was unbiased in most simulation scenarios, except when selectivity variation was very high, and the estimation model assumed selectivity was constant (Figure 2B.1).

The selectivity process error standard deviation was underestimated in estimation models that assumed all processes were random effects (RRR) but was relatively unbiased or overestimated in other estimation models across other process error simulation scenarios (Figure 2.7). In particular, the standard deviation was substantially overestimated, sometimes by a factor of 2 or 3 when selectivity was the only source of process variability in the model (CCR), and the bias became more extreme as either recruitment, expected survival, or selectivity variability increased. The selectivity process error correlation coefficient was underestimated in all



scenarios, and the bias was greater when selectivity was the only source of process variability (Figure 2B.2).

The Mohn's Rho statistics showed low retrospective bias and high precision in spawning stock biomass or average fishing mortality, except in estimation models where recruitment was incorrectly specified as constant (Figures 2B.4-2B.5). When recruitment was specified as a random effect, there were positive retrospective patterns in recruitment, and when specified as a fixed effect, the retrospective patterns were less (Figure 2B.6).

Increasing the variability in total catch or index data or the proportions at age did not noticeably change the precision or accuracy in model output, though the precision was lower in estimation models that assumed recruitment was constant across all observation error operating model scenarios (Figures 2B.7-2B.9). The retrospective patterns in spawning stock biomass, average fishing mortality, and recruitment was also equivalent across observation error scenarios (Figures 2B.10-2B.12). Recruitment and selectivity process error standard deviation, and both correlation coefficients were estimated about as well across observation error operating model scenarios (Figures 2B.13-2B.17).

Figure 2.1. Relative error in spawning stock biomass in the terminal year across process variability operating model scenarios. The black lines indicated zero relative error. Results presented in each column are for when process error variance was set at a range of values (left-recruitment, middle-survival, right-selectivity) with the standard deviations for the other processes at baseline values. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). The number above each boxplot indicates the number of models that converged and produced positive definite Hessians, out of 100 simulated data sets. The black dot is the median and the red dot is the average relative error across the converged models.

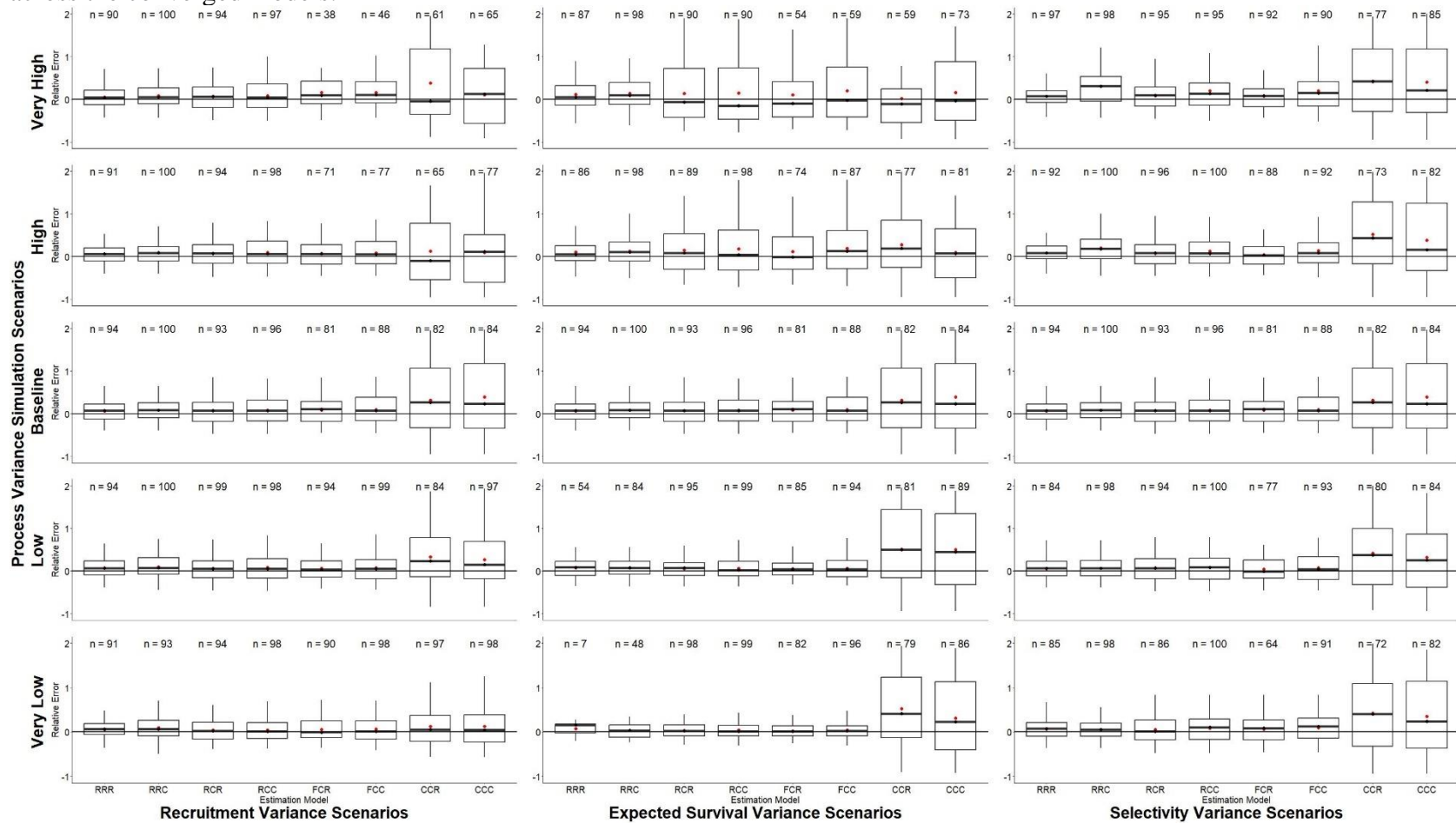


Figure 2.2. Relative error in average recruitment in the terminal 3 years across process variability operating model scenarios. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C), in that order. See caption of Figure 2.1 for description of plot elements.

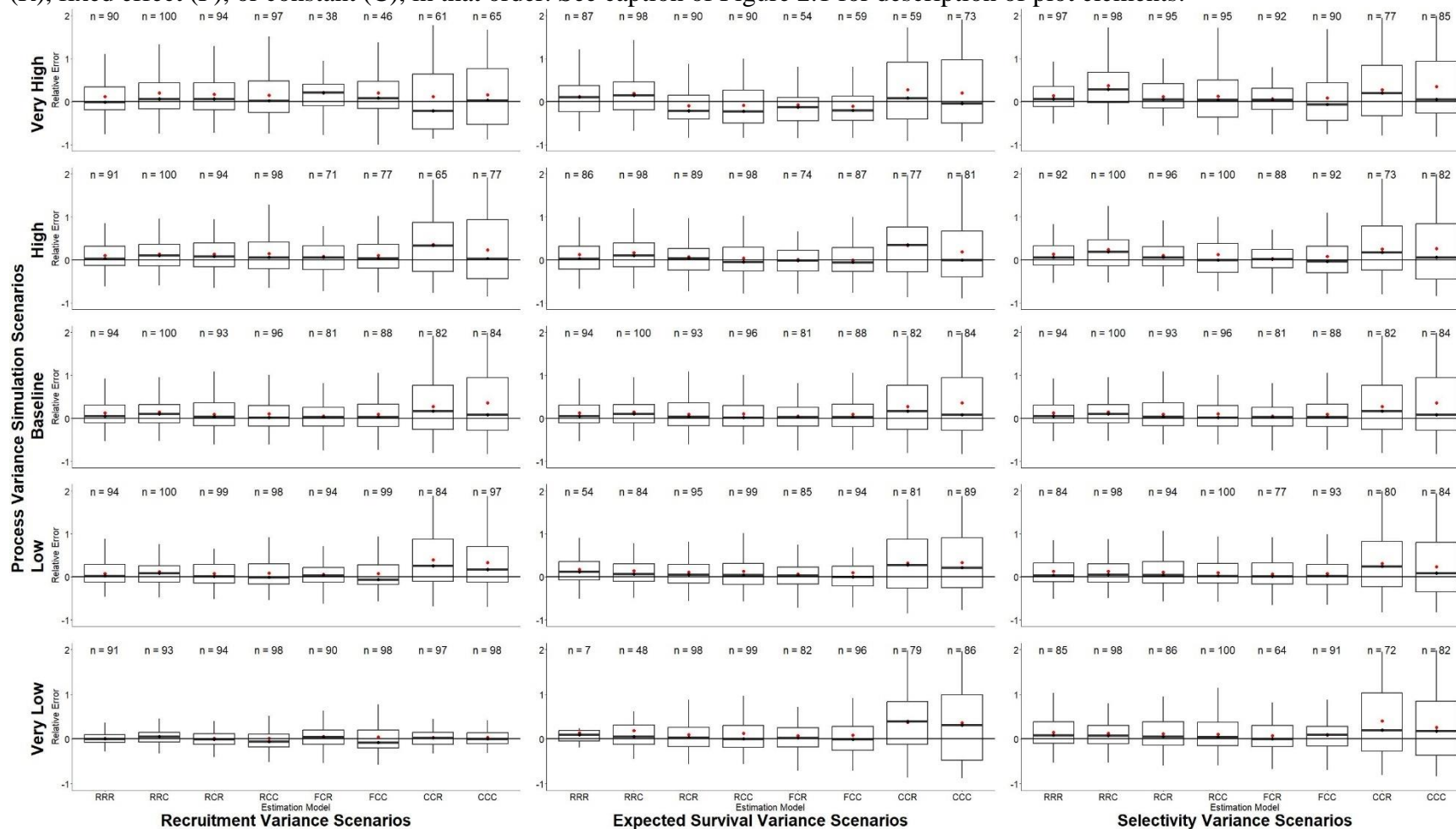


Figure 2.3. Relative error in average exploitation (predicted catch biomass/total biomass) in the terminal year across process variability operating model scenarios. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C), in that order. See caption of Figure 2.1 for description of plot elements.

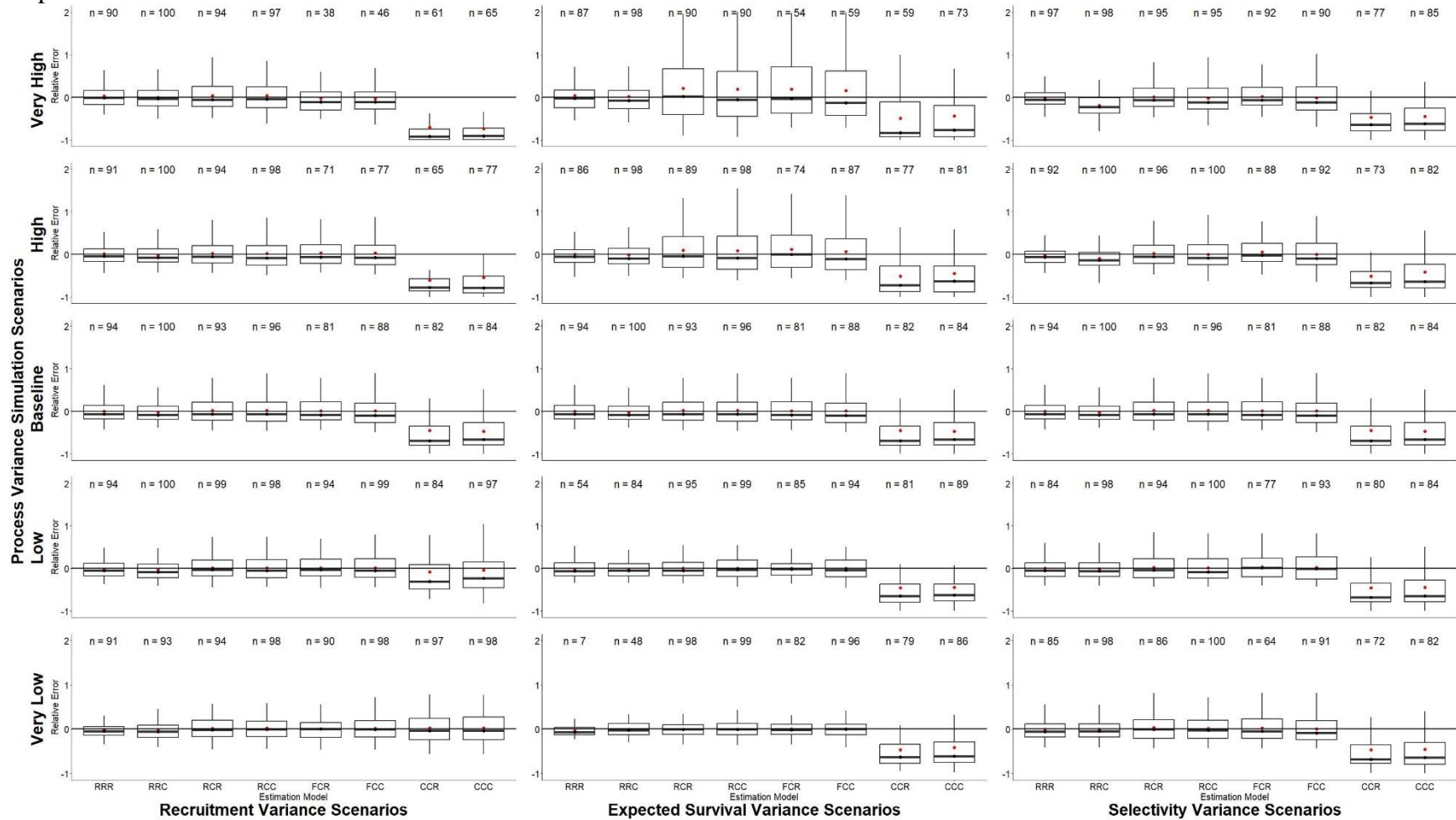


Figure 2.4. Relative error in depletion in the terminal 10 years across process variability operating model scenarios. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C), in that order. See caption of Figure 2.1 for description of plot elements.

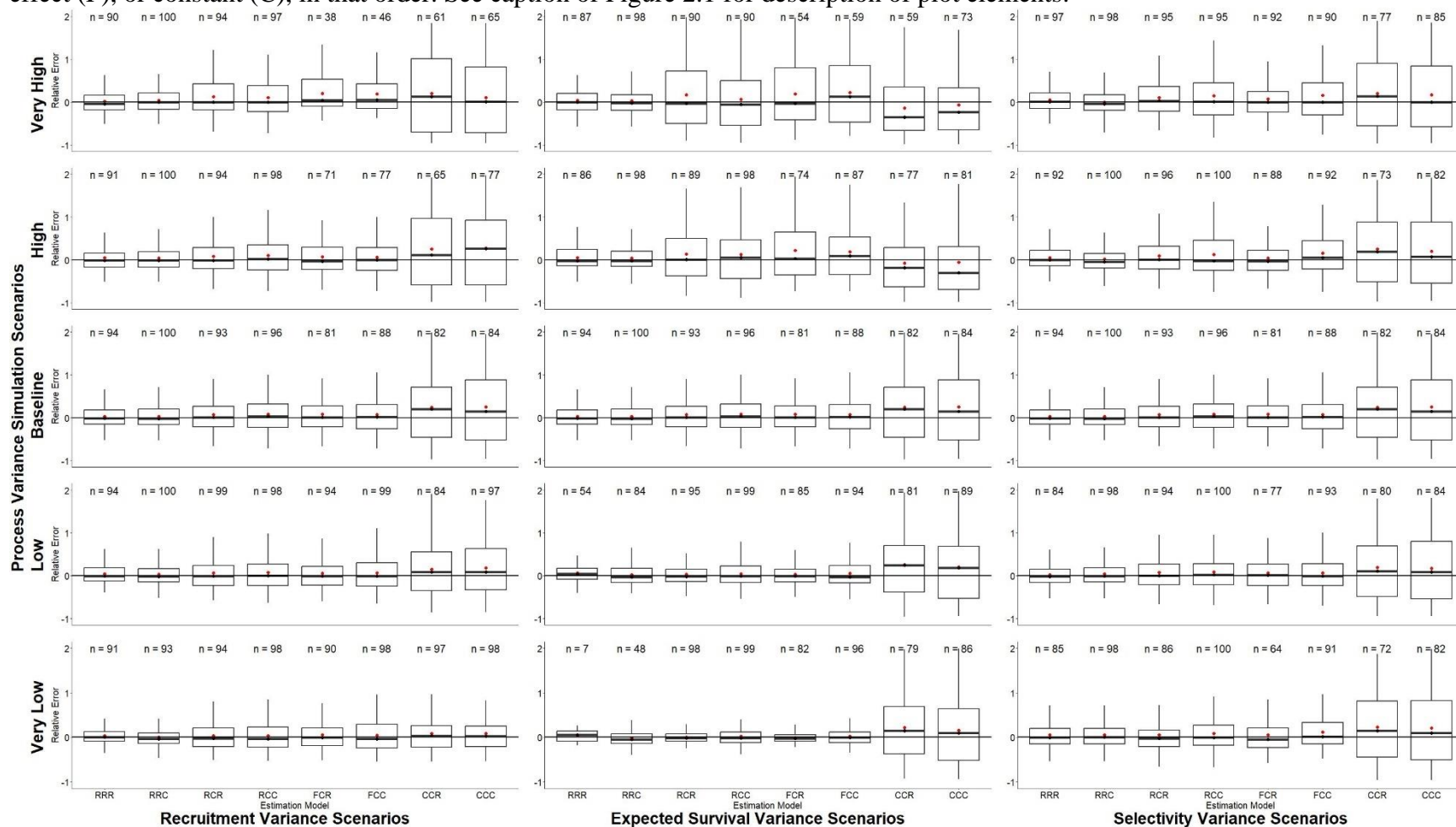


Figure 2.5. Estimated recruitment process error standard deviation across process variability operating model scenarios. Black lines indicate the true recruitment standard deviation. Results presented in each column are for when process error standard deviation was set at a range of values (left-recruitment, middle-survival, right-selectivity) with the standard deviations for the other processes at baseline values. Note that results in all three columns are for recruitment process variation. When the survival (middle column) or selectivity (right column) standard deviations are changed from their baseline value, the true recruitment standard deviation remains constant at its baseline value of 1.176 as indicated by the horizontal lines. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Only estimation models that specify recruitment as a random effect (R--) are represented. See caption of Figure 2.1 for description of plot elements. Note slight differences in the y-axis.

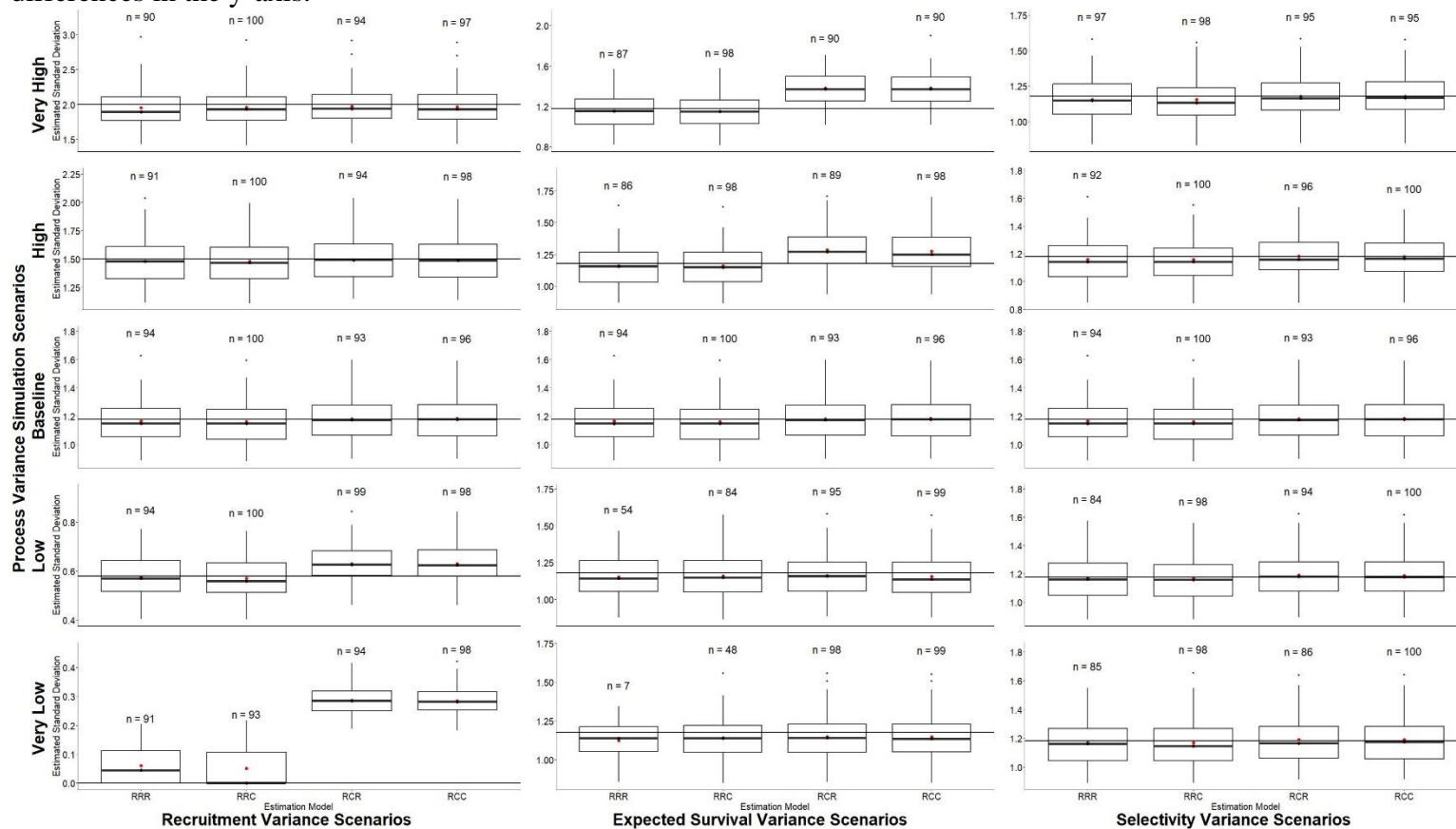


Figure 2.6. Estimated expected survival process error standard deviation across process variability operating model scenarios. Black lines indicate the true expected survival standard deviation. Results presented in each column are for when process error standard deviation was set at a range of values (left-recruitment, middle-survival, right-selectivity) with the standard deviations for the other processes at baseline values. Note that results in all three columns are for survival process variation. When the recruitment (left column) or selectivity (right column) standard deviations are changed from their baseline value, the true expected survival standard deviation remains constant at its baseline value of 0.19 as indicated by the horizontal lines. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Only estimation models that specify survival as a random effect (-R-) are represented. See caption of Figure 2.1 for description of plot elements. Note slight differences in the y-axis.

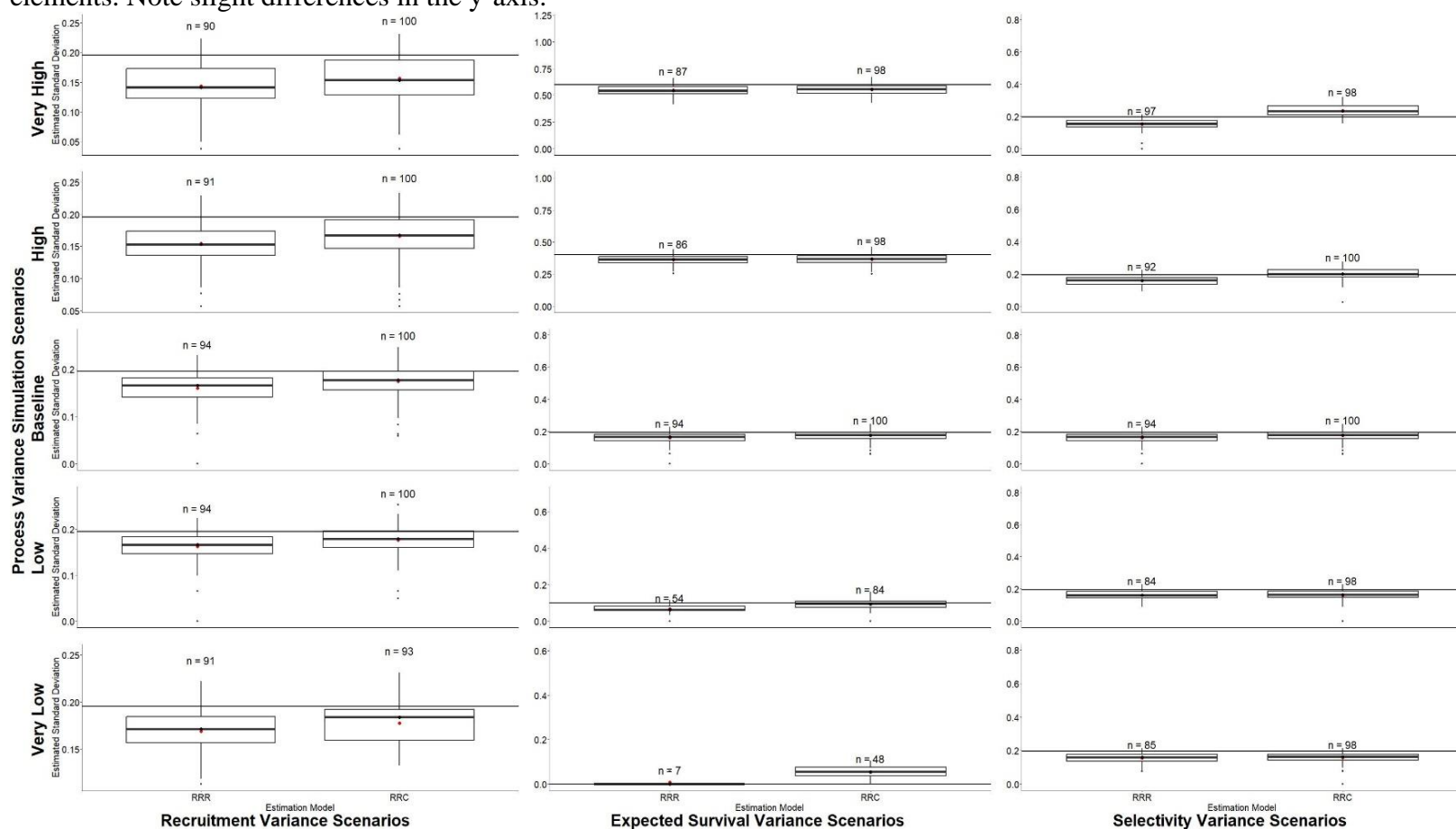
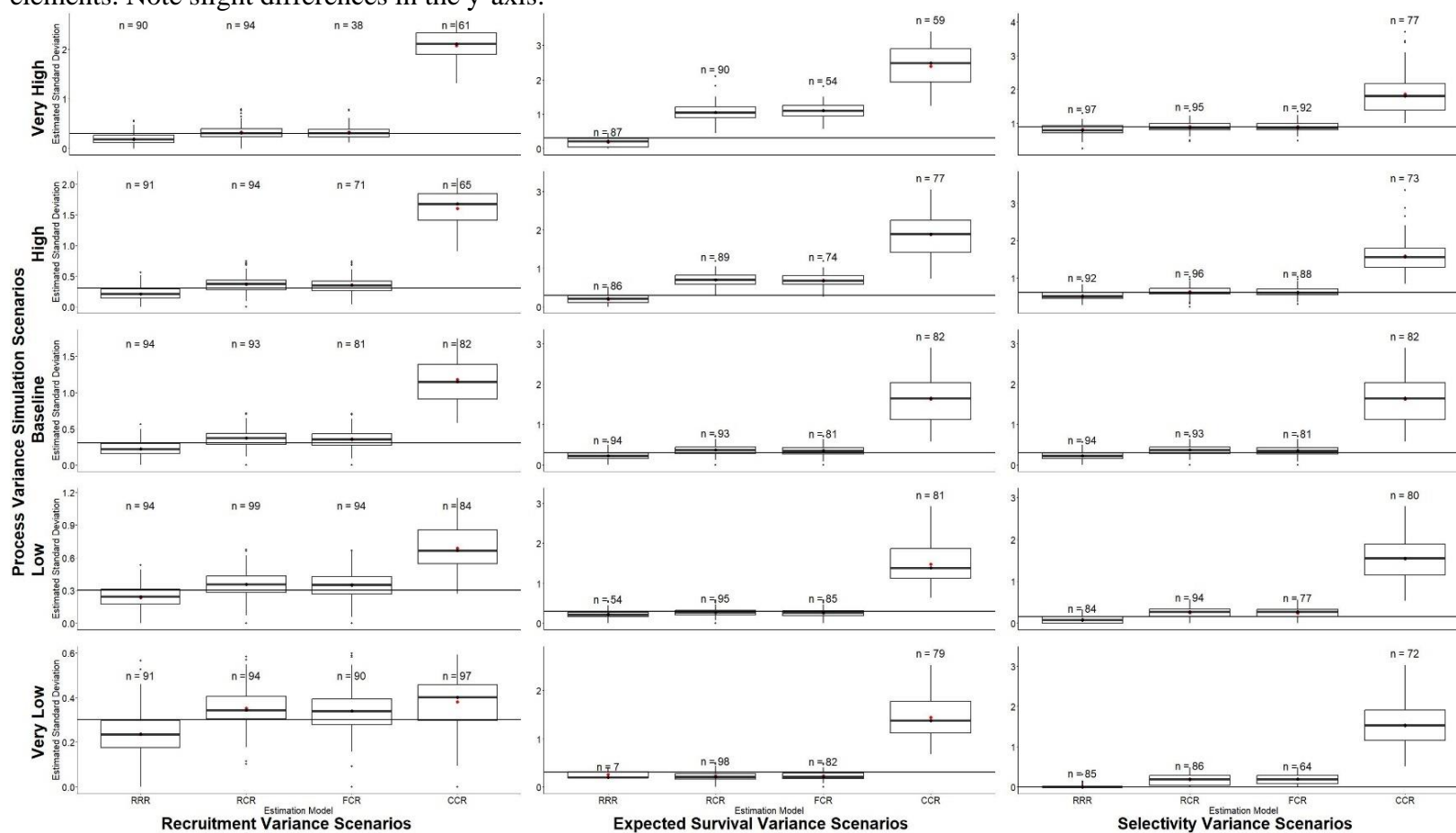




Figure 2.7. Estimated expected selectivity process error standard deviation across process variability operating model scenarios. Black lines indicate the true selectivity standard deviation. Results presented in each column are for when process error standard deviation was set at a range of values (left-recruitment, middle-survival, right-selectivity) with the standard deviations for the other processes at baseline values. Note that results in all three columns are for selectivity process variation. When the recruitment (left column) or expected survival (middle column) standard deviations are changed from their baseline value, the true selectivity standard deviation remains constant at its baseline value of 0.3 as indicated by the horizontal lines. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Only estimation models that specify selectivity as a random effect (--R) are represented. See caption of Figure 2.1 for description of plot elements. Note slight differences in the y-axis.





## 2.5 Discussion

Even under a wide range of true variability in recruitment, survival, and selectivity, estimation models that assumed recruitment varied and treated them as either random or fixed effects estimated spawning stock biomass, recruitment, exploitation, and depletion without substantial bias. Estimation models that assumed constant recruitment performed poorly under most circumstances, but recruitment is rarely, if ever, assumed constant in an integrated age-based stock assessment model because many biotic and abiotic factors will influence the production of eggs and their survival to recruitment from year to year (Maunder and Thorson, 2018b). Though several estimation models were used in this study to provide contrast and elucidate the influence of assuming random effects in each process individually or in combination, only a few resemble those that are conventionally employed in stock assessment. The FCC resembles the architecture of a program like ASAP and a traditional statistical-catch-at-age model (Legault and Restrepo, 1998). When to apply a state-space model that includes process error on everything (RRR), b) a “reduced” version of RRR that specifies only one or two of the processes as random effects, or c) a statistical-catch-at-age model for a particular system will depend on the assumed degree of process and observation error variance and whether the model converges. Ultimately, the decision to employ a state-space model depends strongly on what quantities of interest need to be estimated with high accuracy and if modelers want estimates of process variance that are estimated within the model. Internal estimates of process variance may be preferable in fisheries management because they may better reflect the range of future recruitment, abundance, and spawning stock biomass.

If the recruitment or expected survival process variation is suspected to be very high, then it is better to employ an estimation model that specifies that both processes are random effects.

Such models tended to have higher convergence rates and less biased model estimates compared to models that specify recruitment as a fixed effect and survival as deterministic. Similar research that tested the estimability of recruitment and recruitment standard deviation in a deterministic, fixed effect, penalized likelihood, or marginal likelihood framework found similar results, i.e., that having a constraint on recruitment deviations leads to less bias in estimated recruitment and biomass (Maunder and Deriso, 2003; Thorson et al., 2019). If there is reason to believe the survival is deterministic (i.e., that the abundance at age can be fully explained by the previous abundance at age and mortality), the estimation model should not include survival as a random effect, because the convergence rate of such models under these conditions was very low. When recruitment or survival process variability is low RR- models are likely trying to estimate phantom variability and misallocating variance in recruitment or selectivity to expected survival (more on this discussed below). In the present case, there is little disadvantage to assuming recruitment or selectivity process error is present when it is absent. This aligns with similar work that found that specifying process error in selectivity when there is none did not create bias, as long as data were weighted correctly (Cronin-Fine and Punt, 2021; Martell and Stewart, 2014; Stewart and Monnahan, 2017). However, this conclusion is contingent upon the parameterization of selectivity, as other work has found that sometimes estimating time-varying selectivity when it varies minimally can introduce substantial bias (Linton and Bence, 2011).

Modelers can anticipate the degree of variability in recruitment or expected survival before choosing among estimation models using information outside of the assessment model and/or by understanding what influences interannual variability. Recruitment variability has been weakly linked to fecundity, length of spawning season, and other life history characteristics, and to environmental conditions, though such relationships tend to become unstable over time

(Brosset *et al.*, 2020; Cushing, 1990; Mertz and Myers, 1996; Rose *et al.*, 2001). Thorson *et al.*, (2014) developed a hierarchical stock-recruitment model that estimated the standard deviation and inter-annual correlation coefficient of recruitment for 154 marine species summarized across taxa, which could serve as a reference, though this work quantified variation in addition to that explained by a stock-recruitment relationship. Variation in expected survival includes any change in the numbers-at-age that is not accounted for in the exponential total mortality function. Expected survival variability can reflect immigration/emigration from the system, unaccounted for process variability in natural mortality or selectivity, or catch misreporting, which may be detected from process error residuals (Stock and Miller, 2021). Therefore, if movement is not occurring and other process variability is correctly specified, such as on natural mortality, expected survival can reasonably be excluded from the random effects in the estimation model. If substantial movement or stock mixing is occurring, models can account for this through auxiliary information like tagging data, and/or restructuring management units to reflect closed populations (Berger *et al.*, 2021; Bosley *et al.*, 2022; Goethel *et al.*, 2021). Therefore state-space stock assessment models may perform best when such auxiliary information about the stock movement is routinely collected.

The estimation model that specified random effects in recruitment, expected survival, and selectivity (RRR) did not reliably partition variability across the three processes in all circumstances. When all three processes varied, the standard deviation in recruitment was accurately estimated but expected survival and selectivity variability were underestimated, a phenomenon noted between recruitment and survival in Trijoulet *et al.*, (2020). This underestimation may be because the survival and selectivity processes are confounded, regardless of the degree of process variation, or that the AR(1) process constrains the simulated

values to be closer to the mean and the full extent of variation is not realized with only 42 years of data (Linton and Bence, 2008). When recruitment and selectivity are at their baseline variability but expected survival variation is high, some will be allocated to recruitment. The problem of variance not being apportioned correctly was exacerbated when one or more processes were incorrectly specified. There is an apparent tradeoff between expected survival and selectivity process variance, because when one was assumed fixed (RCR and RRC), the standard deviation of the other process is overestimated. When *only* one source of variability is specified, either the recruitment or selectivity, in the RCC and CCR models, respectively, that process must absorb all model variance, so the standard deviation is substantially overestimated.

The decision to employ a state-space model should not depend on perceived quality of the index or catch data. Other work has investigated the estimability of observation and process error variance more closely and noted trade-offs between data weighting or data availability (increased number of surveys and auxiliary data) and correctly specifying the process model (Dickey-Collas et al., 2015; Stewart and Monnahan, 2017). Within the scope of this work, having quality catch proportion at age data slightly increased the accuracy of expected survival standard deviation estimates. Age composition likelihoods like the Dirichlet-multinomial that include data-weighting parameters have been explored as an alternative to the multinomial distribution when there is time-varying selectivity (Albertsen *et al.*, 2017; Cronin-Fine and Punt, 2021; Xu *et al.*, 2020). The use of such alternative likelihoods may allow for estimation of observation error variance, which was not realized in my model. However, it may exacerbate the issue of correctly apportioning variability and is a topic of future investigation.

Intriguingly, I saw little decline in the precision of biomass or exploitation estimates when data quality was low or process error in recruitment was substantial. This suggests that

assessments of comparable quality might be achieved with alternative data collection schemes involving fewer trips or sampled fish. Gulf of Maine haddock is a data rich case with multiple survey indices, so increasing the variance by no more than 50% and limiting the investigation to one data source at a time may not have been extensive enough to investigate how noisy data can still lead to unbiased and precise estimates. Future work may consider reductions of data quality in several data sources simultaneously, or removal of whole surveys or years of data. As research in data-limited or data-poor methods in fisheries becomes increasingly necessary to support more global fisheries, state-space models may be an important avenue of investigation (Cope et al., 2023). I did not use AIC or DIC as a model selection criterion but rather based model performance on bias and precision of relative error and retrospective patterns. Comparing the likelihood between state-space and non-state-space models can sometimes lead to choosing over-fitted models with poor predictive abilities (Trijoulet et al., 2020). Contemporary model selection is not based solely on AIC or DIC, but rather also using retrospective patterns, consistency, prediction skill, and (of particular importance in state-space models) one step ahead residuals (Carvalho et al., 2021; Trijoulet et al., 2023). Using retrospective patterns to select among models with different selectivity parameterization can lead to better performing models than if selection was done based on DIC (Linton and Bence, 2011).

A state-space model is categorically recommended instead of a SCAA model for its ability to incorporate more stochasticity and estimate process error standard deviations. Correctly quantifying the process error variability in recruitment, expected survival, and selectivity can lead to model projections that are more likely to reflect reality. Accurate estimates of recruitment process variance in particular are vital to predicting the future trajectory of a stock because it can be used to approximate extinction probability and inform recovery plans (Maunder and Deriso,

2003). Using anything except the state-space model that assumes random effects on recruitment, expected survival, and selectivity is only recommended under a few circumstances explored in this study, which the modeler may be able to determine *a priori*. Though this work offers guidelines, its findings are rooted in a particular stock and a model with the variance of observation error distributions specified *a priori*. The observation error variance was fixed at the same value used in the SCAA model which may have been too large because observation error in SCAA tends to account for process variability which the state-space model explicitly specifies. Future work should expand this simulation approach to include even greater observation error bias, greater combinations of low/high variance in processes like recruitment, expected survival, selectivity, and with alternative observation error likelihoods that allow for self-weighting and variance estimation like the Dirichlet-multinomial, logistic normal, or multivariate Tweedie.

## REFERENCES

- Aanes, S., Aeberhard, W.H., Albertsen, C.M., Aldrin, M., Babyn, J., Berg, C.W., Breivik, O., Cook, R., Flemming, J.M., Fryer, R., Liljestr and, E.M., Millar, C., Miller, T.J., Minto, C., Newman, K.B., Nielsen, A., Perreault, A., Regular, P., Skaug, H.J., Spence, M., Trijoulet, V., Vatnehol, S., 2020. Workshop on the review and future of state space stock assessment models in ICES (WKRFSAM). *ICES Sci. Rep.* 2:32. 1-23. <http://doi.org/10.17895/ices.pub.6004>.
- Akaike, H., 1998. Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pp. 199-213. Ed. By E. Parzen, K. Tanabe, and G. Kitagawa. Springer, New York, NY. [https://doi.org/10.1007/978-1-4612-1694-0\\_15](https://doi.org/10.1007/978-1-4612-1694-0_15).
- Albertsen, C.M., Nielsen, A., Thygesen, U.H., 2017. Choosing the observational likelihood in state-space stock assessment models. *Can. J. Fish. Aquat. Sci.* 74, 779-789. <https://doi.org/10.1139/cjfas-2015-0532>.
- Berg, C.W., Nielsen, A., 2016. Accounting for correlated observations in an age-based state-space stock assessment model. *ICES J. Mar. Sci.* 73, 1788-1797. <https://doi.org/10.1093/icesjms/fsw046>.
- Berger, A.M., Deroba, J.J., Bosley, K.M., Goethel, D.R., Langseth, B.J., Schueller, A.M., Hanselman, D.H., 2021. Incoherent dimensionality in fisheries management: Consequences of misaligned stock assessment and population boundaries. *ICES J. Mar. Sci.* 78, 155-171. <https://doi.org/10.1093/icesjms/fsaa203>.
- Bosley, K.M., Schueller, A.M., Goethel, D.R., Hanselman, D.H., Fenske, K.H., Berger, A.M., Deroba, J.J., Langseth, B.J., 2022. Finding the perfect mismatch: Evaluating misspecification of population structure within spatially explicit integrated population models. *Fish Fish.* 23, 294-315. <https://doi.org/10.1111/faf.12616>.
- Brosset, P., Smith, A.D., Plourde, S., Castonguay, M., Lehoux, C., and Van Beveren, E., 2020. A fine-scale multi-step approach to understand fish recruitment variability. *Sci. Rep.*, 10, 16064. <https://doi.org/10.1038/s41598-020-73025-z>.
- Cadigan, N.G., 2015. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. *Can. J. Fish. Aquat. Sci.* 308, 1-13. <https://doi.org/10.1139/cjfas-2015-0047>.
- Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., Kitakado, T., Yemane, D., Piner, K.R., Maunder, M.N., Taylor, I., Wetzel, C.R., Doering, K., Johnson, K.F., Methot, R.D., 2021. A cookbook for using model diagnostics in integrated stock assessments. *Fish. Res.* 240, 105959. <https://doi.org/10.1016/j.fishres.2021.105959>.

- Cope, J. M., Dowling, N. A., Hesp, S. A., Omori, K. L., Bessell-Browne, P., Castello, L., Chick, R., Dougherty, D., Holmes, S.J., McGarvey, R., Ovando, D., Nowlis, J., Prince, J., 2023. The stock assessment theory of relativity: deconstructing the term “data-limited” fisheries into components and guiding principles to support the science of fisheries management. *Rev. Fish. Sci.* 33, 241-263. <https://doi.org/10.1007/s11160-022-09748-1>.
- Cronin-Fine, L., Punt, A.E., 2021. Modeling time-varying selectivity in size-structured assessment models. *Fish. Res.* 239, 105927. <https://doi.org/10.1016/j.fishres.2021.105927>.
- Cushing, D. H., 1990. Plankton production and year-class strength in fish populations: an update of the match/mismatch hypothesis. *Adv. Mar. Biol.* 26, 249-293. [https://doi.org/10.1016/S0065-2881\(08\)60202-3](https://doi.org/10.1016/S0065-2881(08)60202-3).
- De Valpine, P., Hilborn, R., 2005. State-space likelihoods for nonlinear fisheries time-series. *Can. J. Fish. Aquat. Sci.* 62, 1937-1952. <https://doi.org/10.1139/f05-116>.
- Deroba, J.J., Schueller, A.M., 2013. Performance of stock assessments with misspecified age- and time-varying natural mortality. *Fish. Res.* 146, 27-40. <https://doi.org/10.1016/j.fishres.2013.03.015>.
- Dickey-Collas, M., Hintzen, N.T., Nash, R.D.M., Schon, P.-J., Payne, M.R., 2015. Quirky patterns in time-series of estimates of recruitment could be artefacts. *ICES J. Mar. Sci.* 72, 111-116. <https://doi.org/10.1093/icesjms/fsu022>.
- Francis, R.I.C.C., 2011. Corrigendum: Data weighting in statistical fisheries stock assessment models. *Can. J. Fish. Aquat. Sci.* 68, 2228-2228. <https://doi.org/10.1139/f2011-165>.
- Goethel, D.R., Bosley, K.M., Langseth, B.J., Deroba, J.J., Berger, A.M., Hanselman, D.H., Schueller, A.M., 2021. Where do you think you’re going? Accounting for ontogenetic and climate-induced movement in spatially stratified integrated population assessment models. *Fish. Fish.* 22, 141-160. <https://doi.org/10.1111/faf.12510>.
- Han, Q., Shan, X., Jin, X., Gorfine, H., 2023. Contrasting stock status trends obtained from survey and fishery CPUE, taking *Larimichthys polyactis* in Yellow Sea Large Marine Ecosystem as an example. *Ecol. Indic.* 147, 110032. <https://doi.org/10.1016/j.ecolind.2023.110032>.
- Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R.R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. *ICES J. Mar. Sci.* 72, 99-110. <https://doi.org/10.1093/icesjms/fsu198>.
- Johnson, K.F., Monnahan, C.C., McGilliard, C.R., Vert-Pre, K.A., Anderson, S.C., Cunningham, C. J., Hurtado-Ferro, F., Licandeo, R.R., Muradian, M.L., Ono, K., Szuwalski, C.S., Valero, J.L., Whitten, A.R., Punt, A.E., 2014. Time-varying natural mortality in fisheries



- stock assessment models: identifying a default approach. *ICES J. Mar. Sci.* 72, 137-150. <https://doi.org/10.1093/icesjms/fsu055>.
- Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., Bell, B.M., 2016. TMB: Automatic differentiation and laplace approximation. *J. Stat. Softw.* 70, 1-21. <https://doi.org/10.18637/jss.v070.i05>.
- Legault, C.M., Restrepo, V.R., 1998. A flexible forward age-structured assessment program. *Collect. Vol. Sci. Pap. ICCAT*, 49, 246-253.
- Linton, B.C., Bence, J.R., 2008. Evaluating methods for estimating process and observation error variances in statistical catch-at-age analysis. *Fish. Res.* 94, 26-35. <https://doi.org/10.1016/j.fishres.2008.06.014>.
- Linton, B.C., Bence, J.R., 2011. Catch-at-age assessment in the face of time-varying selectivity. *ICES J. Mar. Sci.* 68, 611-625. <https://doi.org/10.1093/icesjms/fsq173>.
- Martell, S., Stewart, I., 2014. Towards defining good practices for modeling time-varying selectivity. *Fish. Res.* 158, 84-95. <https://doi.org/10.1016/j.fishres.2013.11.001>.
- Maunder, M.N., Deriso, R.B., 2003. Estimation of recruitment in catch-at-age models. *Can. J. Fish. Aquat. Sci.* 60, 1204-1216. <https://doi.org/10.1139/f03-104>.
- Maunder, M.N., 2011. Review and evaluation of likelihood functions for composition data in stock-assessment models: Estimating the effective sample size. *Fish. Res.* 109, 311-319. <https://doi.org/10.1016/j.fishres.2011.02.018>.
- Maunder, M.N., Piner, K.R., 2017. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. *Fish. Res.* 192, 16-27. <https://doi.org/10.1016/j.fishres.2016.04.022>.
- Maunder, M.N., Thorson, J.T., 2018. Modeling temporal variation in recruitment in fisheries stock assessment: a review of theory and practice. *Fish. Res.* 217, 1-16. <https://doi.org/10.1016/j.fishres.2018.12.014>.
- Mertz, G., Myers, R.A., 1996. Influence of fecundity on recruitment variability of marine fish. *Can. J. Fish. Aquat. Sci.* 53, 1618-1625. <https://doi.org/10.1139/f96-089>.
- Miller, T.J., Hare, J.A., Alade, L.A., 2016. A state-space approach to incorporating environmental effects on recruitment in an age-structured assessment model with an application to southern new england yellowtail flounder. *Can. J. Fish. Aquat. Sci.* 73, 1261-1270. <https://doi.org/10.1139/cjfas-2015-0339>.
- [NEFSC] Northeast Fisheries Science Center. 2022. Stock Assessment Update of 14 Northeast Groundfish Stocks Through 2018. US Dept Commer, Northeast Fish. Sci. Cent. Ref. Doc. 22-06.

- Nielsen, A., Berg, C.W., 2014. Estimation of time-varying selectivity in stock assessments using state-space models. *Fish. Res.* 158, 96-101. <https://doi.org/10.1016/j.fishres.2014.01.014>.
- Perreault, A.M.J., Wheeland, L.J., Joanne Morgan, M., Cadigan, N.G., 2020. A state-space stock assessment model for American plaice on the Grand Bank of Newfoundland. *J. Northwest Atl. Fish. Sci.* 51, 45-104. <https://doi.org/10.2960/J.v51.m727>.
- Rose, K.A., Cowan, J.H., Winemiller, K.O., Myers, R.A., Hilborn, R., 2001. Compensatory density dependence in fish populations: importance, controversy, understanding and prognosis. *Fish and Fish.* 2, 293-327. <https://doi.org/10.1046/j.1467-2960.2001.00056.x>.
- Rossi, S.P., Cox, S.P., Benoît, H.P., Swain, D.P., 2019. Inferring fisheries stock status from competing hypotheses. *Fish. Res.* 216, 155-166. <https://doi.org/10.1016/j.fishres.2019.04.011>.
- Soto, M., Fernández-Peralta, L., Pennino, M.G., Kokkalis, A., Rey, J., Salmerón, F., Liébana, M., Meissa, B., Kell, L., 2022. Effects of misreporting landings, discards, and Catch Per Unit of Effort index in state-space production models: the case of black hake in northwest Africa. *ICES J. Mar. Sci. fsac188*. <https://doi.org/10.1093/icesjms/fsac188>.
- Stewart, I.J., Martell, S.J.D., 2014. A historical review of selectivity approaches and retrospective patterns in the Pacific halibut stock assessment. *Fish. Res.* 158, 40-49. <https://doi.org/10.1016/j.fishres.2013.09.012>.
- Stewart, I.J., Monnahan, C.C., 2017. Implications of process error in selectivity for approaches to weighting compositional data in fisheries stock assessments. *Fish. Res.* 192, 126-134. <https://doi.org/10.1016/j.fishres.2016.06.018>.
- Stock, B.C., Miller, T.J., 2021. The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time- and age- varying processes via random effects and links to environmental covariates. *Fish. Res.* 240, 105967. <https://doi.org/10.1016/j.fishres.2021.105967>.
- Stock, B.C., Xu, H., Miller, T.J., Thorson, J.T., Nye, J.A., 2021. Implementing two-dimensional autocorrelation in either survival or natural mortality improves a state-space assessment model for Southern New England-Mid Atlantic yellowtail flounder. *Fish. Res.* 237, 105873. <https://doi.org/10.1016/j.fishres.2021.105873>.
- Szuwalski, C.S., Ianelli, J.N., Punt, A.E., 2018. Reducing retrospective patterns in stock assessment and impacts on management performance. *ICES J. Mar. Sci.* 75, 596-609. <https://doi.org/10.1093/icesjms/fsx159>.
- Thorson, J.T., Jensen, O.P., Zipkin, E.F., 2014. How variable is recruitment for exploited marine fishes? A hierarchical model for testing life history theory. *Can. J. Fish. Aquat. Sci.* 71, 973-983. <https://doi.org/10.1139/cjfas-2013-0645>.

- Thorson, J.T., Kristensen, K., 2016. Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. *Fish. Res.* 175, 66-74. <https://doi.org/10.1016/j.fishres.2015.11.016>.
- Thorson, J.T., 2019. Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fish. Res.* 210, 143-161. <https://doi.org/10.1016/j.fishres.2018.10.013>.
- Thorson, J. T., Rudd, M. B., Winker, H., 2019. The case for estimating recruitment variation in data-moderate and data-poor age-structured models. *Fish. Res.* 217, 87-97. <https://doi.org/10.1016/j.fishres.2018.07.007>.
- Trijoulet, V., Fay, G., Miller, T.J., 2020. Performance of a state-space multispecies model: What are the consequences of ignoring predation and process errors in stock assessments? *J. Appl. Ecol.* 57, 121-135. <https://doi.org/10.1111/1365-2664.13515>
- Trijoulet, V., Albertsen, C.M., Kristensen, K., Legault, C.M., Miller, T.J., Nielsen, A., 2023. Model validation for compositional data in stock assessment models: Calculating residuals with correct properties. *Fish. Res.* 257, 106487. <https://doi.org/10.1016/j.fishres.2022.106487>.
- Xu, H., Thorson, J.T., Methot, R.D., 2020. Comparing the performance of three data-weighting methods when allowing for time-varying selectivity. *Can. J. Fish. Aquat. Sci.* 77, 247-263. <https://doi.org/10.1139/cjfas-2019-0107>.

APPENDIX 2A: INPUT OBSERVATION VARIANCE VALUES

Table 2A.1. Input standard deviation of lognormal distribution for total catch and total index, and effective sample size of multinomial distribution for catch and index proportions at age used to fit Gulf of Maine haddock.

Year	$\sigma_{C_y}$	$\sigma_{I_{y,S}}^2$	$\sigma_{I_{y,F}}^2$	$n_y^{(C)}$	$n_{y,S}^{(I)}$	$n_{y,F}^{(I)}$
1977	0.2	0.6142	0.5225	60	15	25
1978	0.2	0.7069	0.3756	60	15	25
1979	0.2	0.5174	0.402	60	15	25
1980	0.2	0.6821	0.5415	60	15	25
1981	0.2	0.5442	0.3841	60	15	25
1982	0.2	0.6061	0.5385	60	15	25
1983	0.2	0.6917	0.4717	60	15	25
1984	0.2	0.7048	0.467	60	15	7
1985	0.2	0.6785	0.5869	60	15	7
1986	0.2	0.7443	0.6136	60	15	7
1987	0.2	0.6549	0.5188	60	5	7
1988	0.2	0.8185	0.8237	20	5	7
1989	0.15	1.0139	0.5951	20	5	7
1990	0.15	0.8431	0.5812	20	5	7
1991	0.15	0.8238	0.771	20	5	7
1992	0.15	0.915	0.7289	20	5	7
1993	0.15	0.7611	0.9208	60	10	7
1994	0.15	0.635	0.6189	60	10	7
1995	0.15	0.7618	0.7218	60	10	7
1996	0.15	0.6179	0.5799	60	10	7
1997	0.15	0.698	0.5582	60	10	7
1998	0.15	0.709	0.6899	60	10	30
1999	0.15	0.6945	0.4931	60	10	30
2000	0.15	0.69	0.6627	60	10	30
2001	0.1	0.8868	0.4699	60	10	30
2002	0.1	0.8074	0.5369	60	10	30
2003	0.1	0.543	0.4162	140	10	30
2004	0.1	0.6241	0.4569	140	10	30
2005	0.1	0.6961	0.3908	140	25	30
2006	0.1	0.744	0.4546	140	25	30
2007	0.1	0.6686	0.4973	140	25	30
2008	0.1	0.7815	0.5064	140	25	30
2009	0.1	0.666	0.5245	140	25	12
2010	0.1	0.6778	0.6653	140	25	12
2011	0.1	0.6538	0.6523	140	25	12
2012	0.1	0.6848	0.7937	140	25	12

Table 2A.1 (cont'd)

2013	0.1	0.73	0.4409	140	25	12
2014	0.1	0.492	0.5957	140	25	12
2015	0.1	0.8862	0.409	140	25	12
2016	0.1	0.5058	0.3803	140	25	12
2017	0.1	0.5909	0.3692	140	25	12
2018	0.1	0.4801	0.3968	140	25	12
2019	0.1	0.4973	0.4331	140	25	12

---

## APPENDIX 2B: SUPPLEMENTARY FIGURES

Figure 2B.1. Estimated recruitment and expected survival process error correlation coefficient. Black lines indicate the true correlation coefficient. Results presented in each column are for when process error standard deviation was set at a range of values (left-recruitment, middle-survival, right-selectivity) with the standard deviations for the other processes at baseline values. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Only estimation models that specify recruitment as a random effect (R--) are represented. The number above each boxplot indicates the number of models that converged and produced positive definite Hessians, out of 100 simulated data sets. The black dot is the median and the red dot is the average relative error across the converged models. Note slight differences in the y-axis.

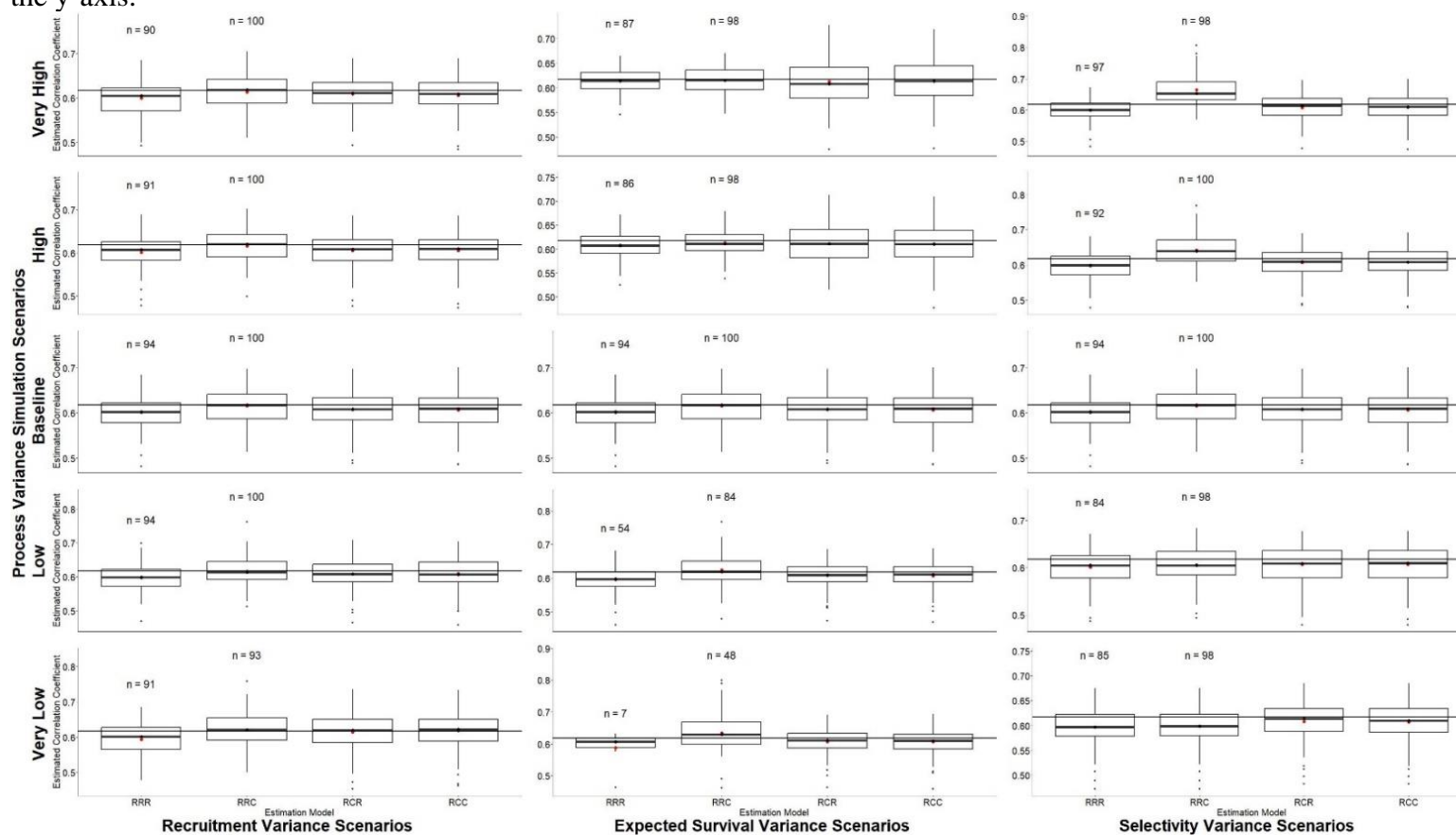


Figure 2B.2. Estimated selectivity process error correlation coefficient. Black lines indicate the true correlation coefficient. Results presented in each column are for when process error standard deviation was set at a range of values (left-recruitment, middle-survival, right-selectivity) with the standard deviations for the other processes at baseline values. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Only estimation models that specify selectivity as a random effect (--R) are represented. See caption of Figure 2B.1 for description of plot elements. Note slight differences in the y-axis.

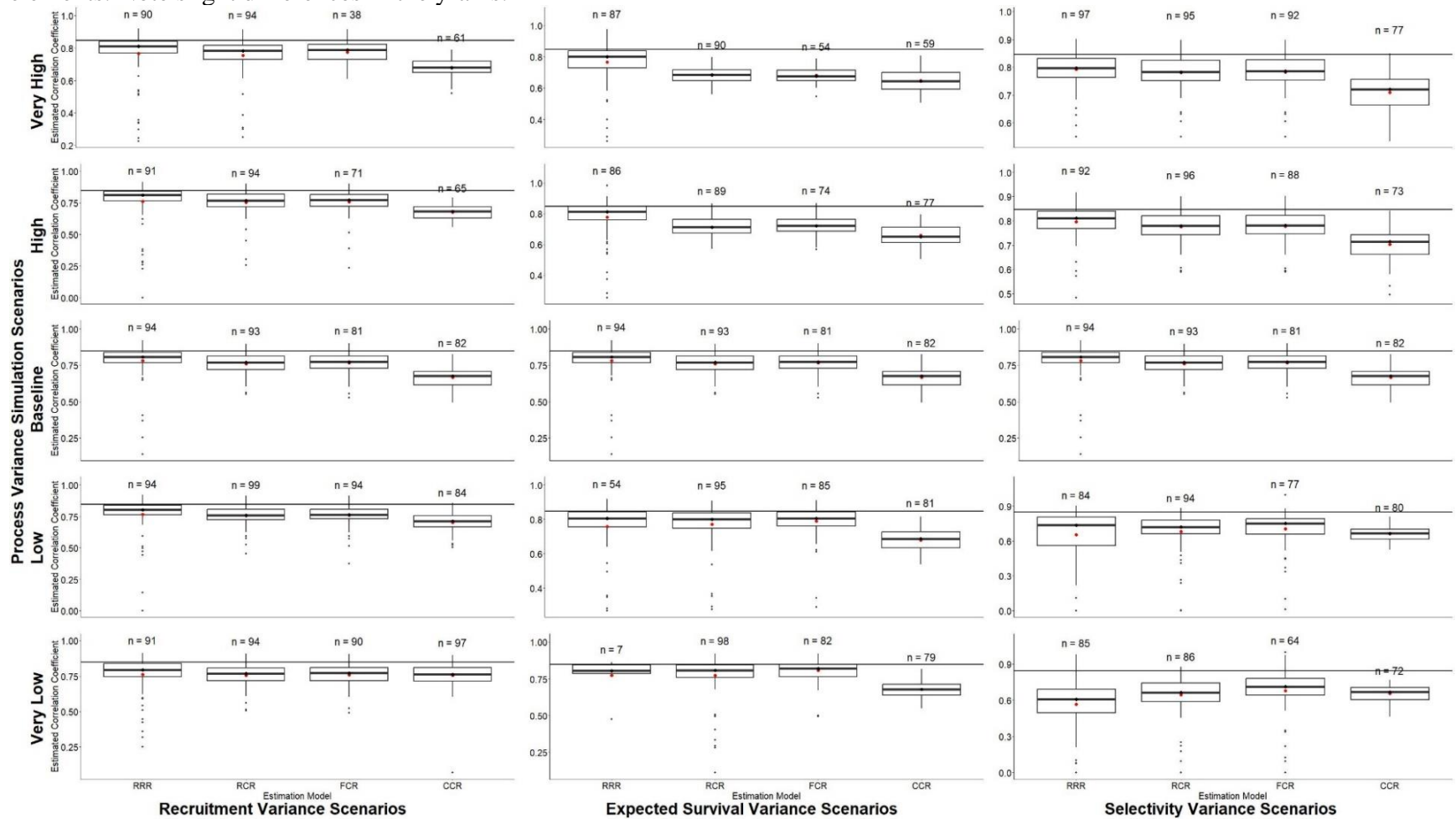


Figure 2B.3. Mohn's Rho of spawning stock biomass across process variability operating model scenarios. Black lines indicate zero retrospective bias. Results presented in each column are for when process error standard deviation was set at a range of values (left-recruitment, middle-survival, right-selectivity) with the standard deviations for the other processes at baseline values. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). See caption of Figure 2B.1 for description of plot elements.

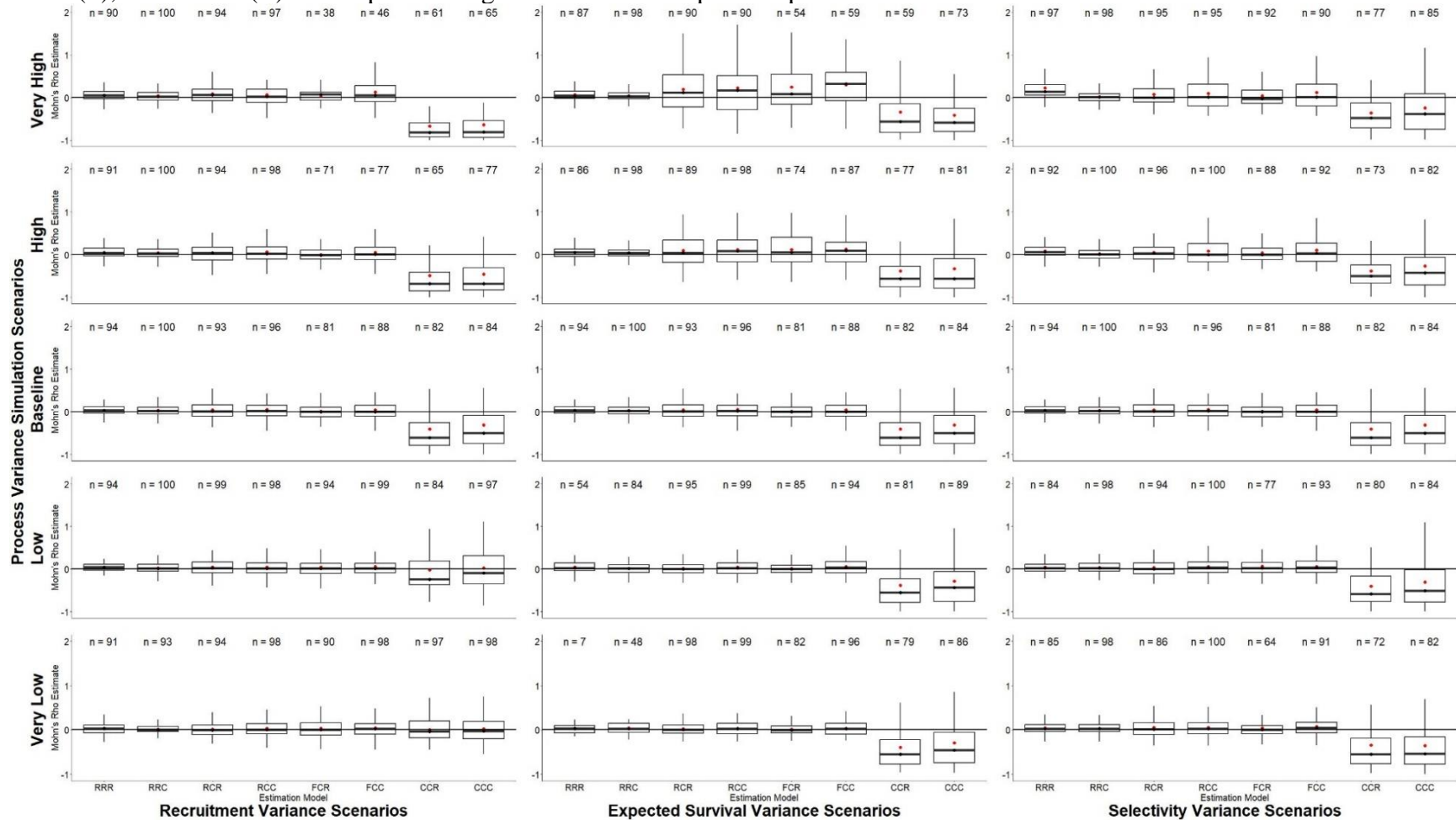




Figure 2B.4. Mohn's Rho of average fishing mortality across process variability operating model scenarios. Black lines indicate zero retrospective bias. Results presented in each column are for when process error standard deviation was set at a range of values (left-recruitment, middle-survival, right-selectivity) with the standard deviations for the other processes at baseline values. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). See caption of Figure 2B.1 for description of plot elements.

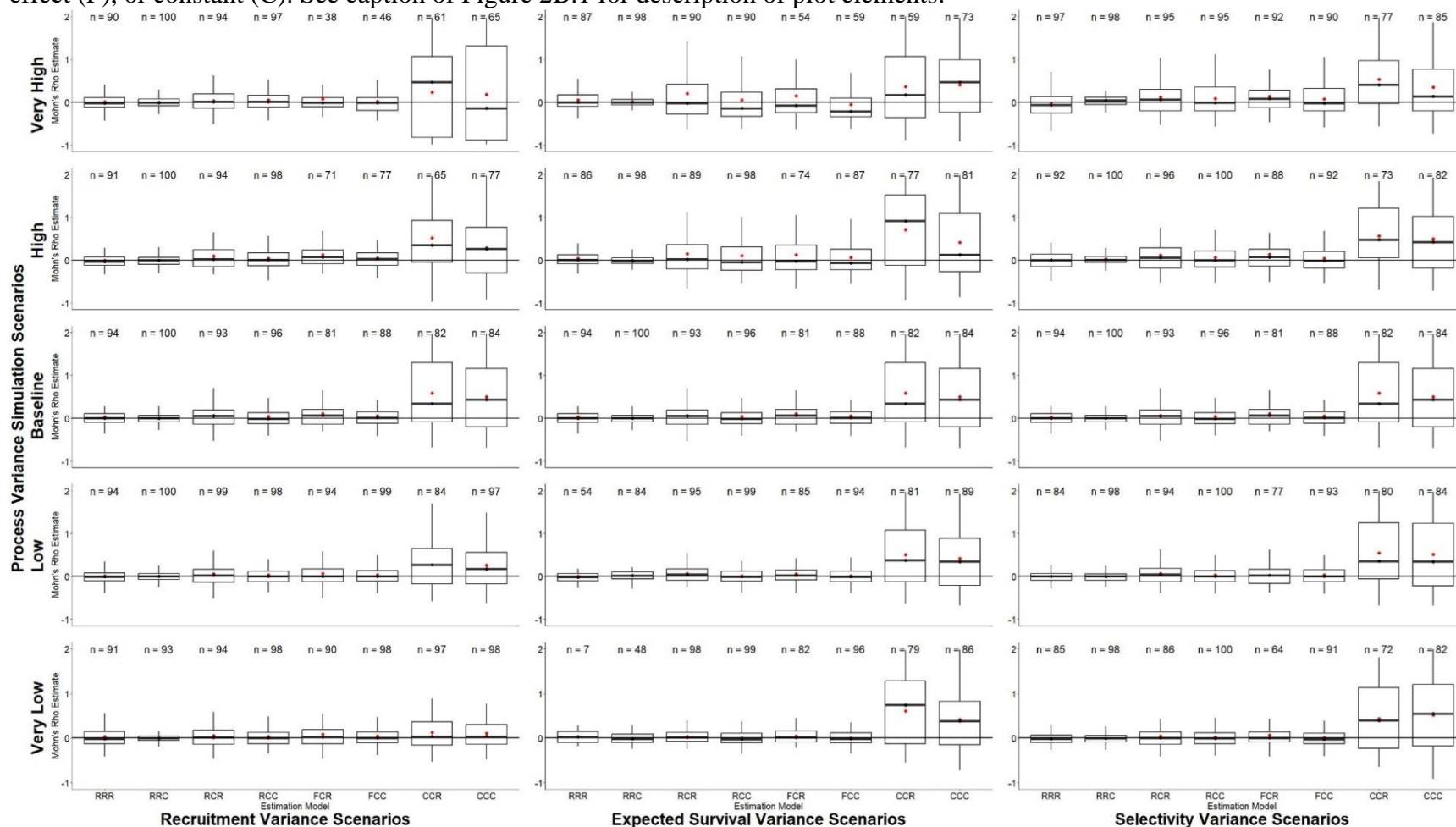


Figure 2B.5. Mohn's Rho of recruitment across process variability operating model scenarios. Black lines indicate zero retrospective bias. Results presented in each column are for when process error standard deviation was set at a range of values (left-recruitment, middle-survival, right-selectivity) with the standard deviations for the other processes at baseline values. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). See caption of Figure 2B.1 for description of plot elements.

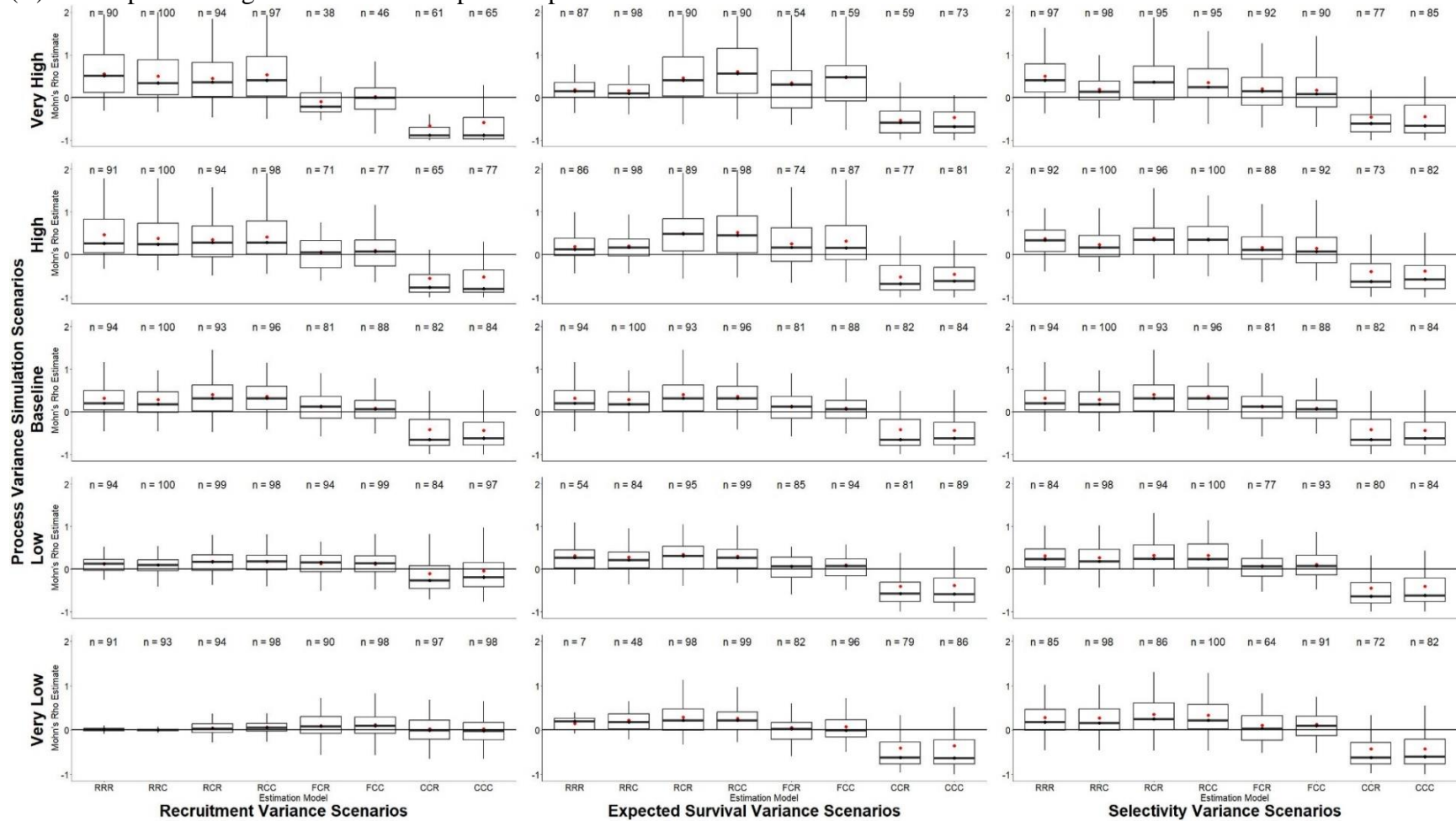


Figure 2B.6. Relative error in spawning stock biomass in the terminal year across observation error operating models scenarios where the standard deviation of the lognormal error of total catch or total indices (columns 1-2) or the effective sample size of the multinomial catch or index proportions (columns 3-4) are adjusted to make the data noisier. The black lines indicated zero relative error. Results presented in each column are for when process error variance was set at a range of values (left-recruitment, middle-survival, right-selectivity) with the standard deviations for the other processes at baseline values. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). The number above each boxplot indicates the number of models that converged and produced positive definite Hessians, out of 100 simulated data sets. The black dot is the median and the red dot is the average relative error across the converged models.

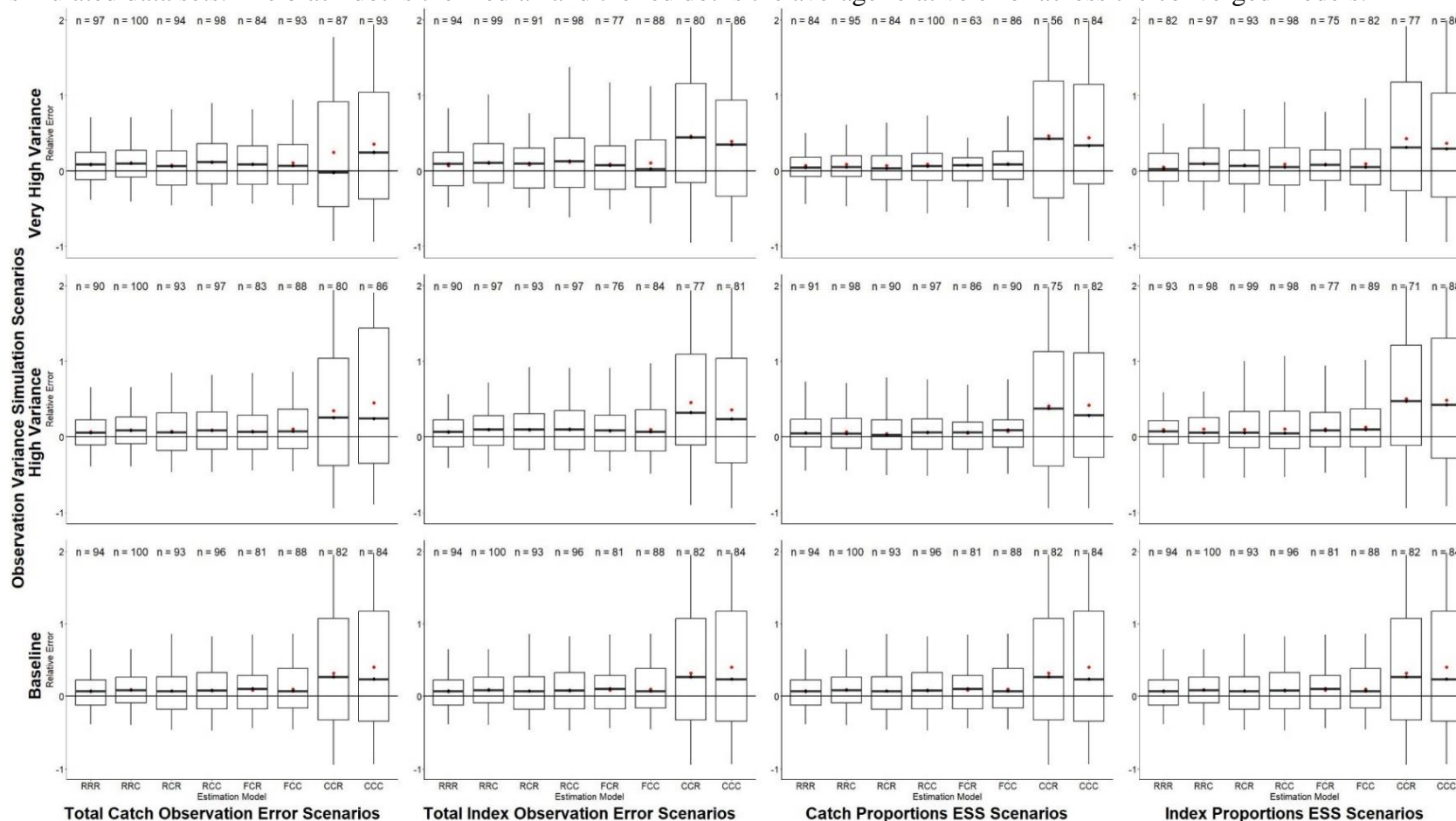


Figure 2B.7. Relative error in average recruitment in the terminal 3 years across observation error operating models scenarios. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C), in that order. See caption of Figure 2B.6 for description of plot elements.

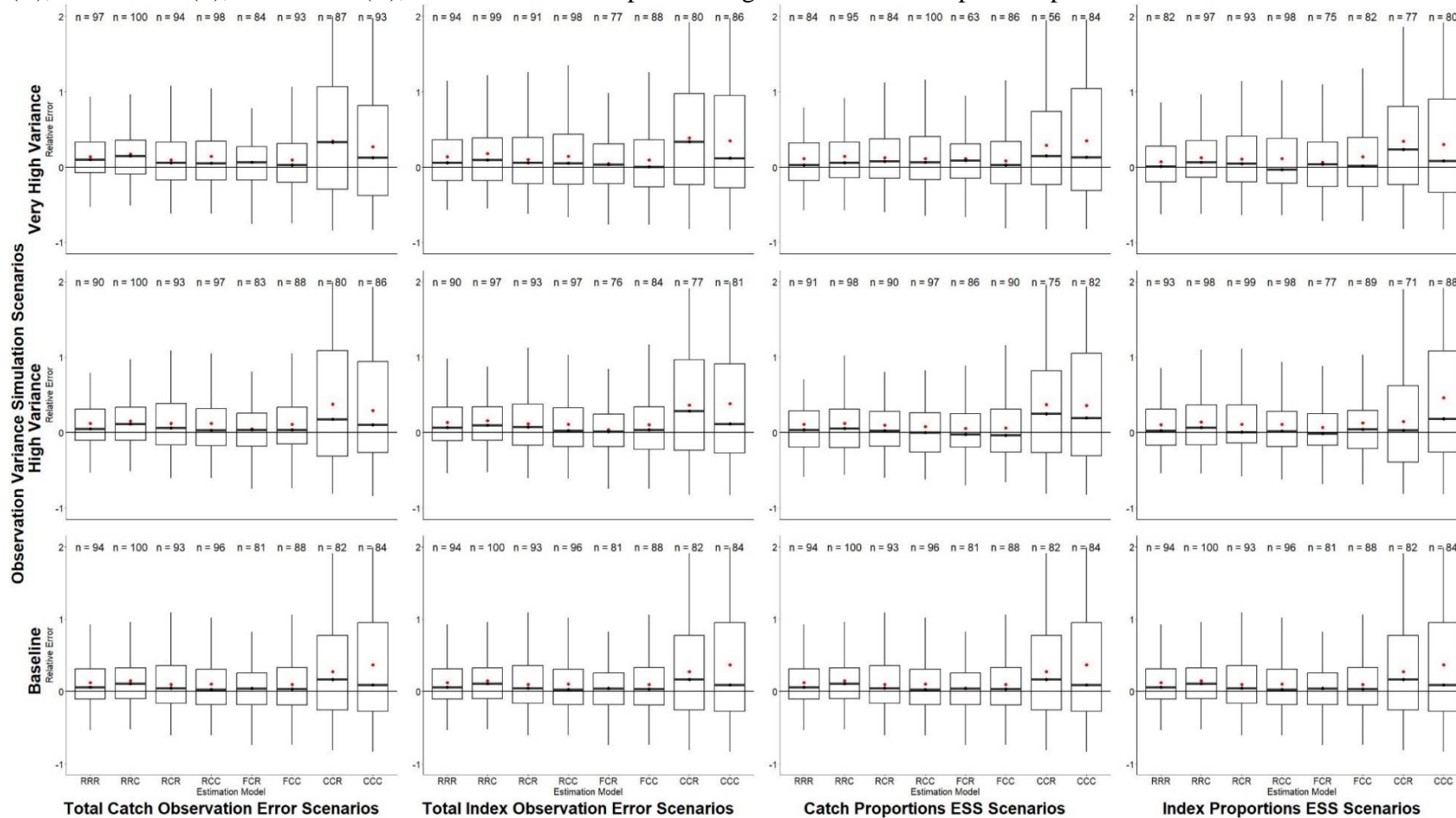


Figure 2B.8. Relative error in exploitation (predicted catch biomass/total biomass) in the terminal year across observation error operating models scenarios. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C), in that order. See caption of Figure 2B.6 for description of plot elements.

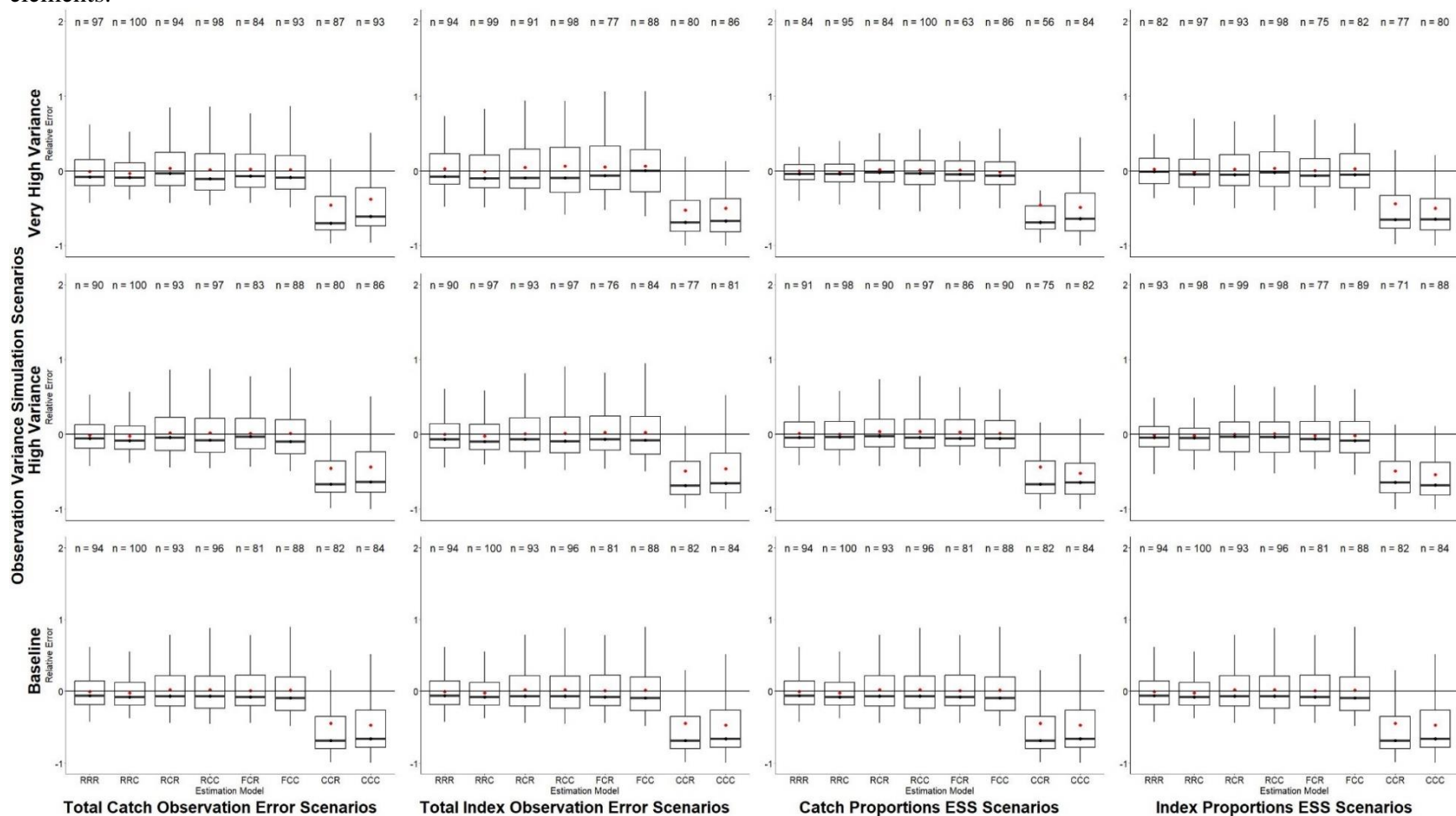


Figure 2B.9. Relative error in depletion in the terminal 10 years across observation error operating models scenarios. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C), in that order. See caption of Figure 2B.6 for description of plot elements.

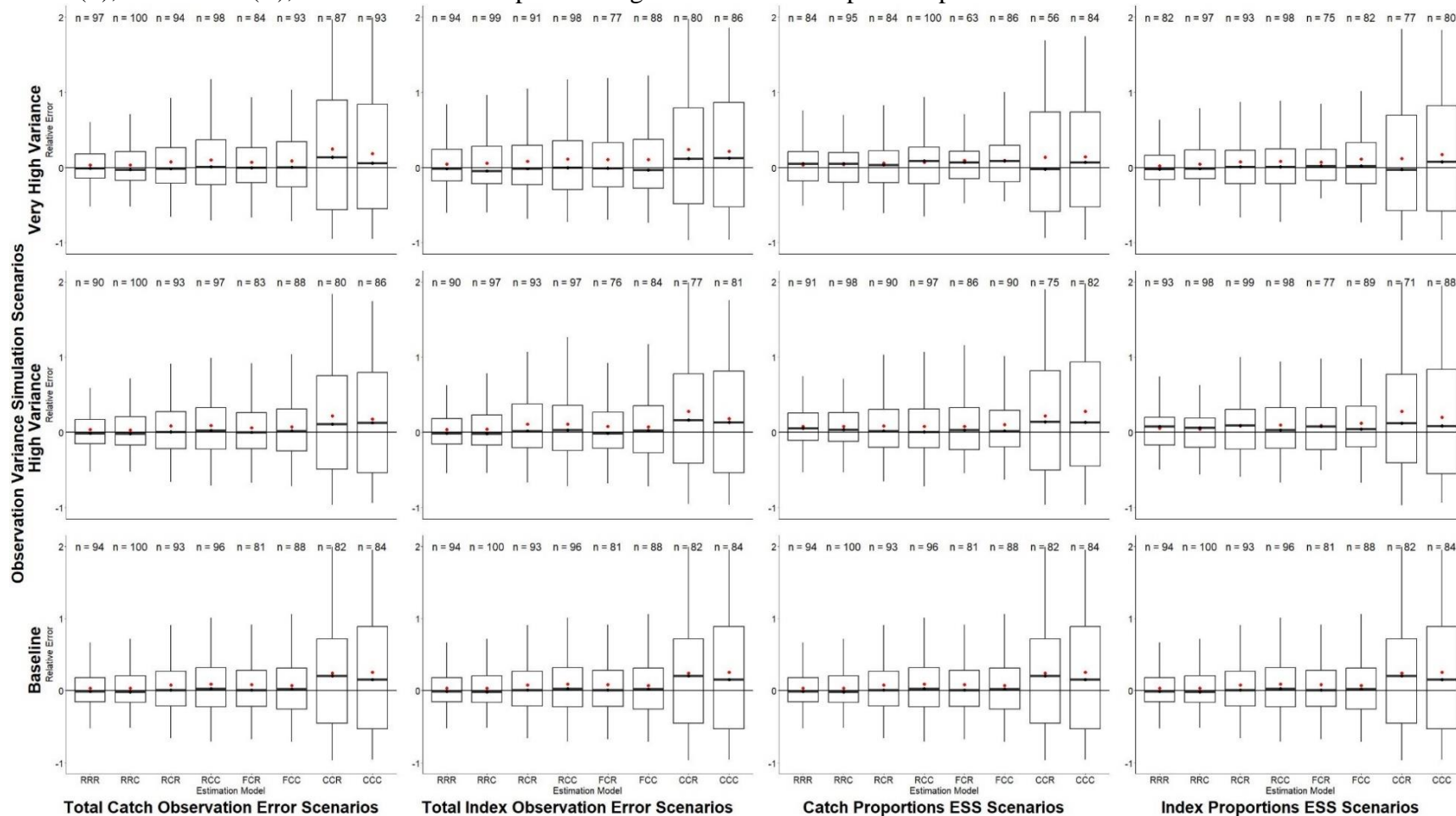




Figure 2B.10. Mohn's Rho of spawning stock biomass across observation error operating models scenarios. Black lines indicate zero retrospective bias. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). See caption of Figure 2B.6 for description of plot elements.

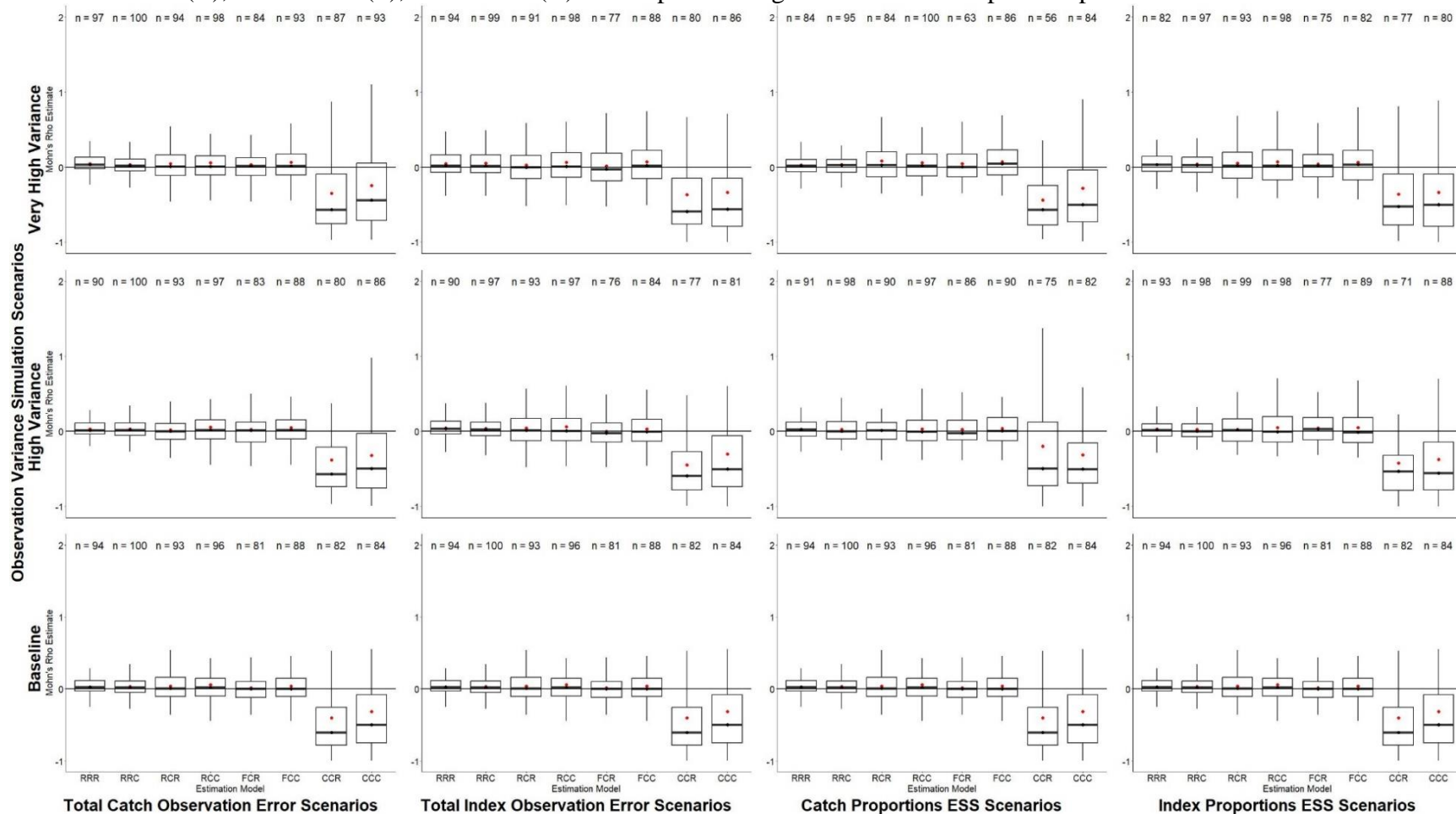


Figure 2B.11. Mohn's Rho of average fishing mortality across observation error operating models scenarios. Black lines indicate zero retrospective bias. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). See caption of Figure 2B.6 for description of plot elements.

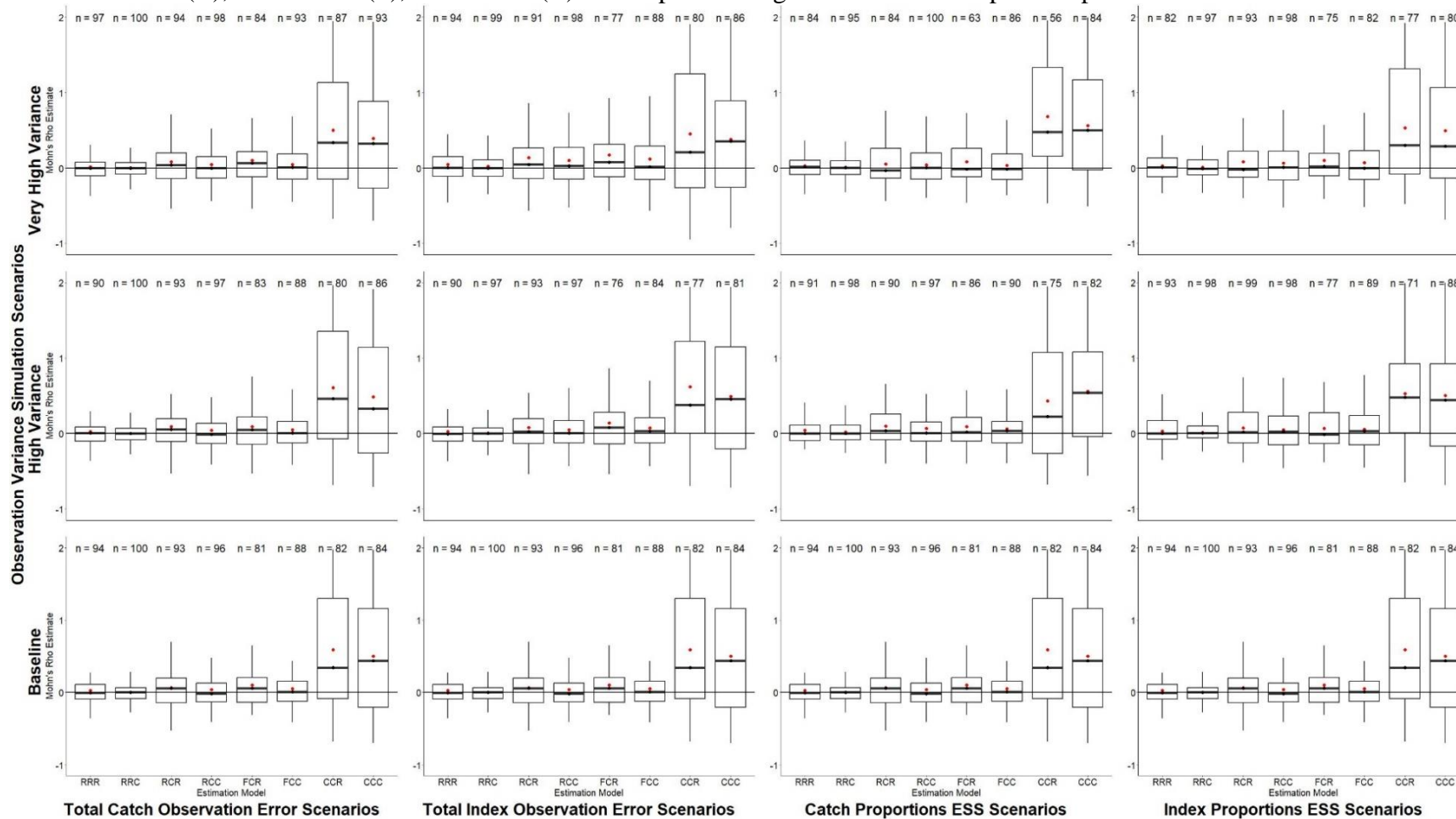




Figure 2B.12. Mohn's Rho of recruitment across observation error operating models scenarios. Black lines indicate zero retrospective bias. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). See caption of Figure 2B.6 for description of plot elements.

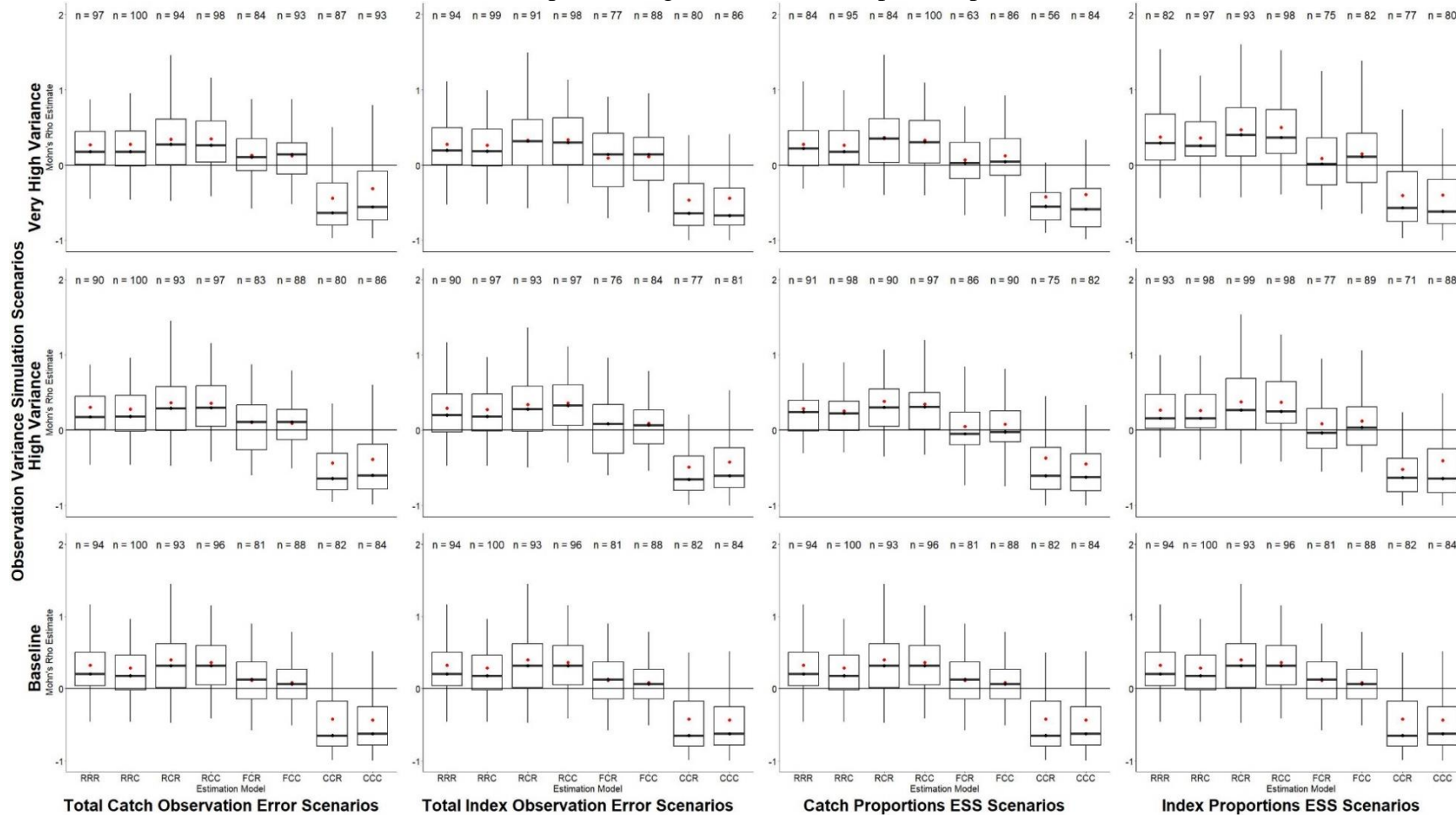


Figure 2B.13. Estimated recruitment process error standard deviation across observation error operating model scenarios. Black lines indicate the true recruitment standard deviation. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Only estimation models that specify recruitment as a random effect (R--) are represented. See caption of Figure 2B.6 for description of plot elements.

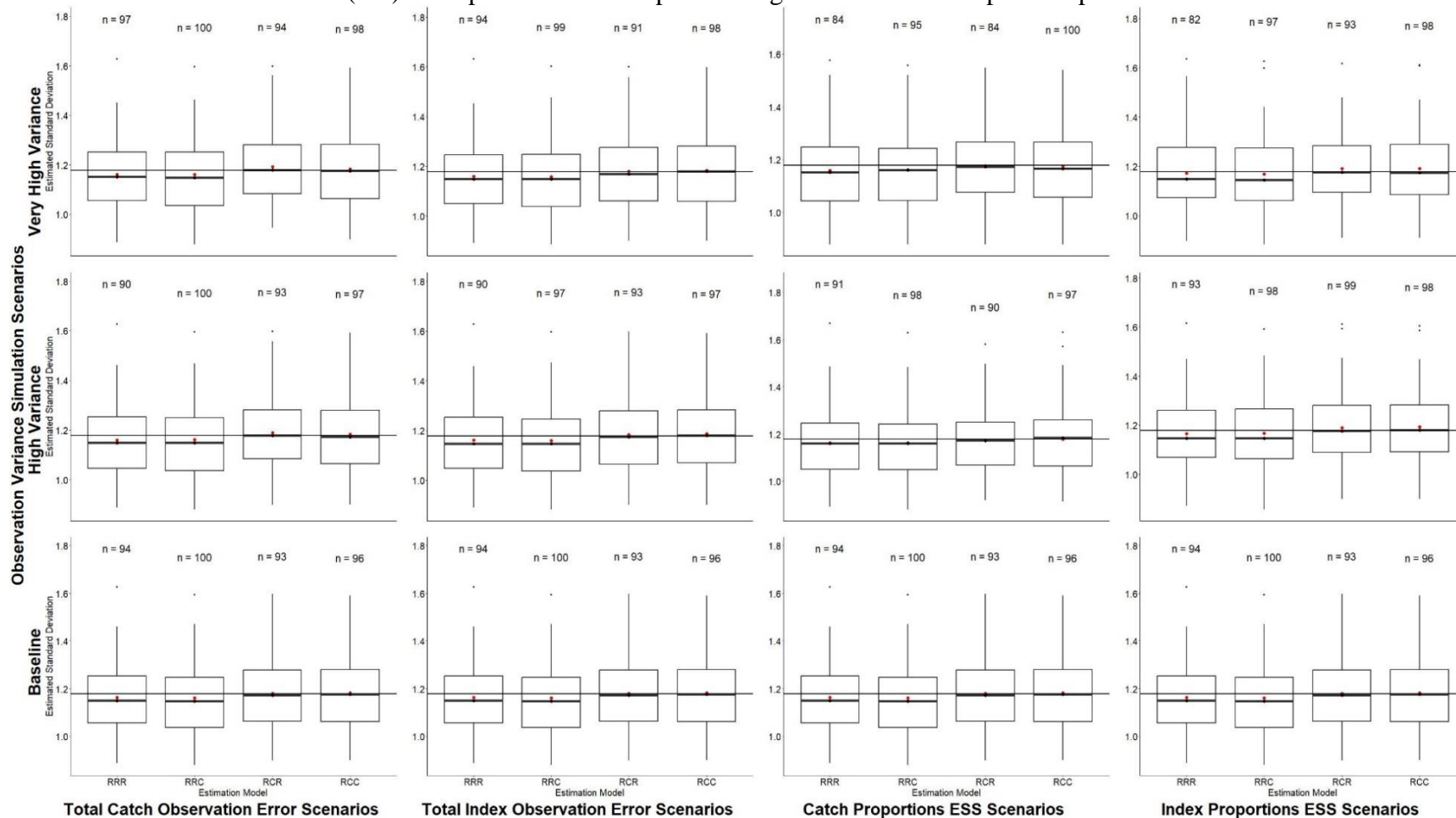


Figure 2B.14. Estimated expected survival process error standard deviation across observation error operating model scenarios. Black lines indicate the true expected survival standard deviation. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Only estimation models that specify survival as a random effect (-R-) are represented. See caption of Figure 2B.6 for description of plot elements.

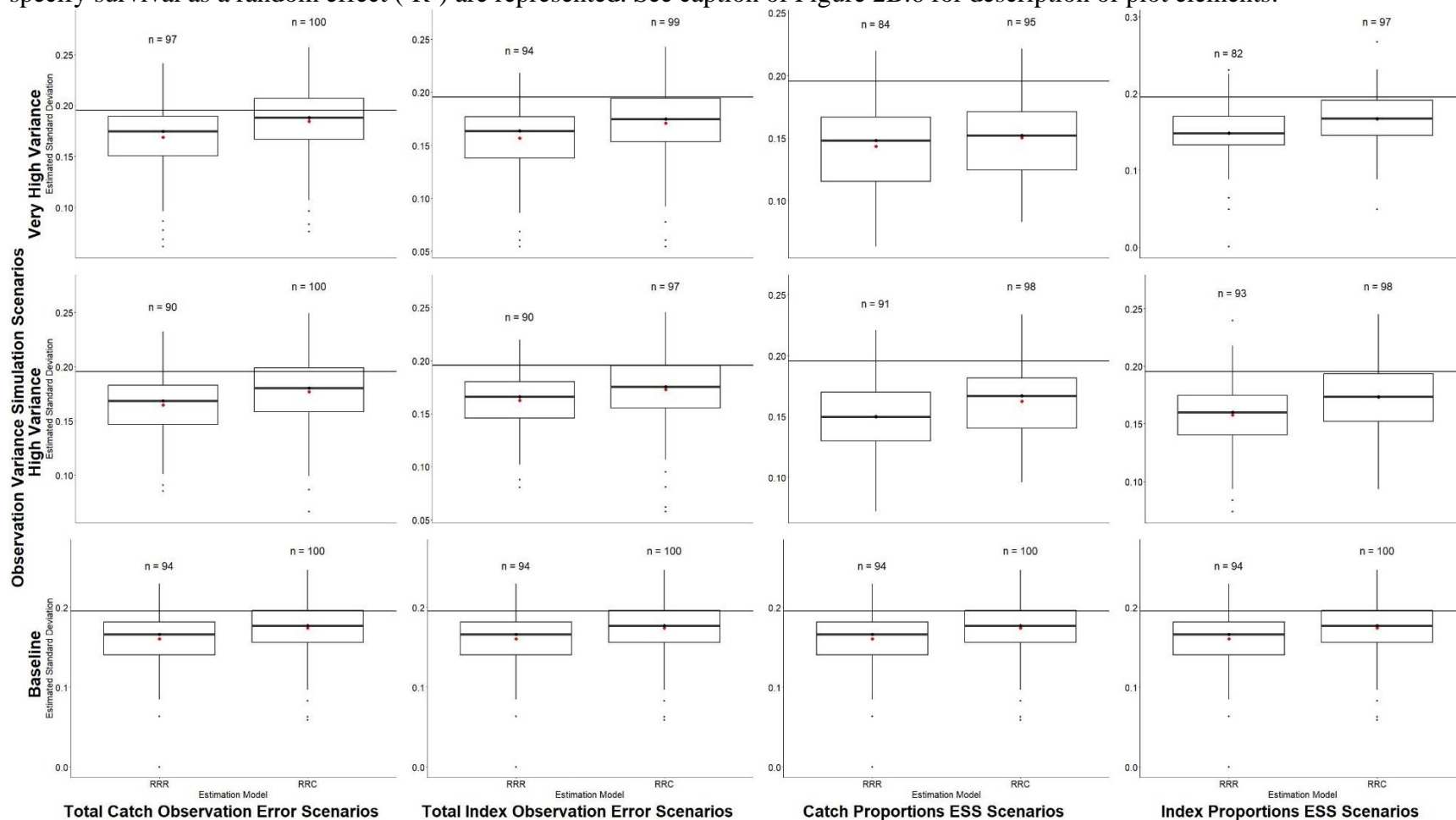


Figure 2B.15. Estimated selectivity process error standard deviation across observation error operating model scenarios. Black lines indicate the true selectivity standard deviation. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Only estimation models that specify selectivity as a random effect (R--) are represented. See caption of Figure 2B.6 for description of plot elements.

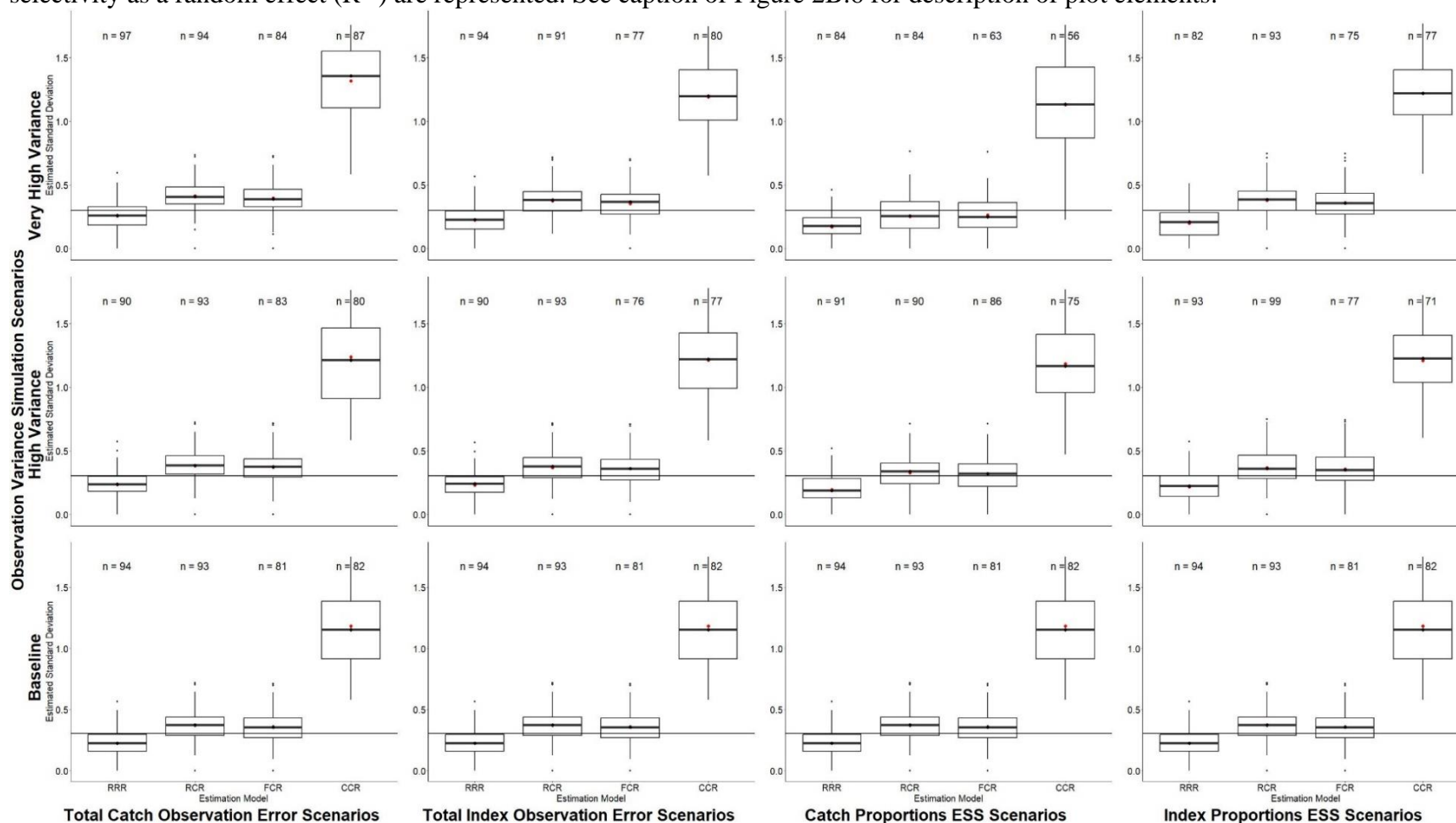


Figure 2B.16. Estimated recruitment and expected survival process error correlation coefficient across observation error operating model scenarios. Black lines indicate the true correlation coefficient. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Only estimation models that specify recruitment as a random effect (R--) are represented. See caption of Figure 2B.6 for description of plot elements.

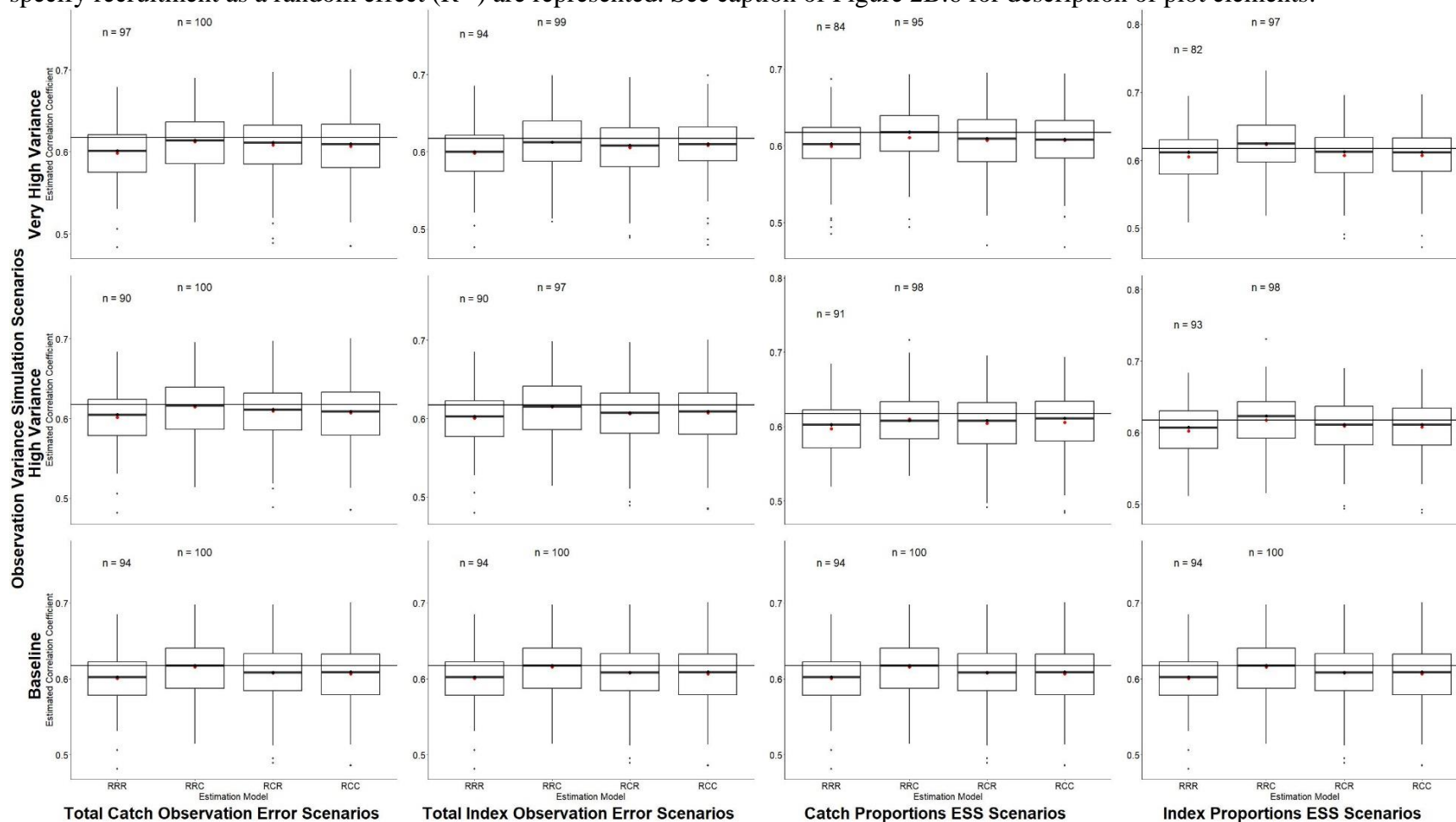
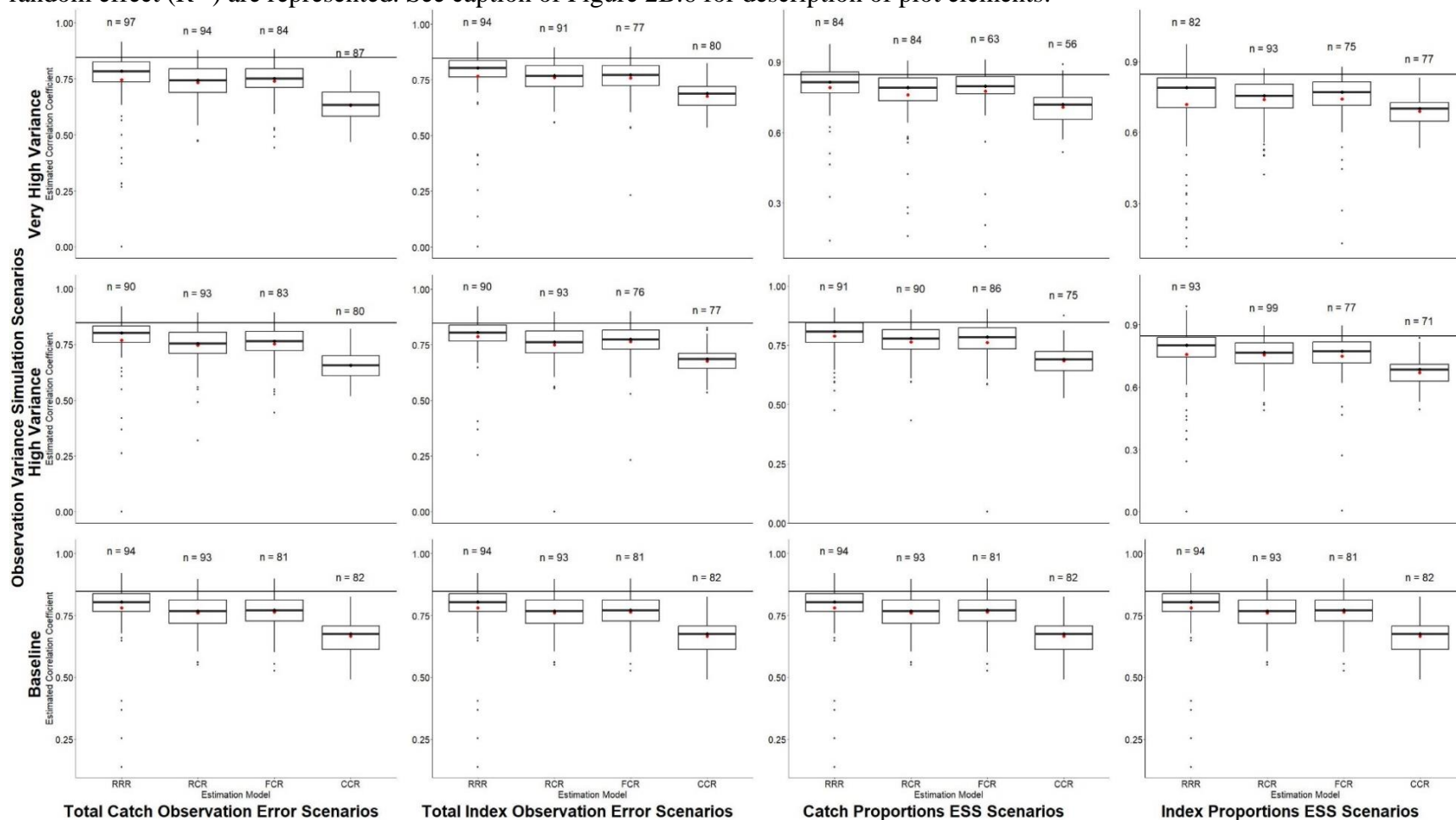


Figure 2B.17. Estimated selectivity process error correlation coefficient across observation error operating model scenarios. Black lines indicate the true correlation coefficient. Individual boxplots represent the eight estimation models that treat recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Only estimation models that specify selectivity as a random effect (R--) are represented. See caption of Figure 2B.6 for description of plot elements.



## CHAPTER 3:

### THE PERFORMANCE OF ALTERNATIVE AGE COMPOSITION LIKELIHOODS IN STATE-SPACE MODELS WITH MULTIPLE SOURCES OF PROCESS VARIATION

#### 3.1 Abstract

Integrated stock assessment models incorporate multiple sources of data like catch, indices of abundance, and proportions at age, each with different degrees of quality in either fishery or survey catch. In a maximum likelihood estimation framework, integrated models explicitly apportion observation variability through values supplied to the assumed likelihood, such as the standard deviation of a lognormal distribution (for catch) or the effective sample size of a multinomial distribution (for proportions at age), and these terms are typically supplied by the user. Several alternatives to the multinomial (MN) distribution, such as the Dirichlet-multinomial (DM), logistic normal (LG), and multivariate Tweedie (MVTW) distributions, can allow observation variance to be approximated within the model by estimating parameters that scale or replace the input effective sample size. These distributions can better account for the intra-haul correlation and overdispersion of sampled ages. However, these distributions have yet to be extensively tested in a state-space stock assessment model with estimated process error variance, which has the added challenge of apportioning total model variability between the observation and process sub-models. I used a simulation-estimation framework to generate data with process variability in recruitment, expected survival (i.e., abundance), and selectivity at age, and with observation variability in the catch proportions at age derived from one of several distributions. Simulated data were fit using several estimation models that differed in which processes (recruitment, expected survival, or selectivity) were incorporated and in the assumed distribution for observation error influencing age composition data. Estimation models had

difficulty apportioning variability among processes, especially when the observation model for age compositions did not account for realized inter-age correlation and overdispersion.

Estimation models that assumed the survival processes were stochastic produced biased model output, but only models that assumed an intermediate amount of process variability (either stochastic survival or constant selectivity at age) were able to accurately estimate the parameters determining the characteristics of observation error in age composition data. It is recommended that the logistic normal or multivariate Tweedie distributions be employed in fitting state-space stock assessment models, especially in the face of several time-varying processes.

### 3.2 Introduction

Statistical estimation models fit data using a likelihood function, which quantifies the probability of observing the data conditional on the model and the values of the parameters. Choices are made about observation error (e.g., how it is structured, the magnitude relative to the process error, and how much is distributed across multiple data sources) through specification of the likelihood function. Through the variance parameters, such as the standard deviation of the normal distribution or the effective sample size of the multinomial distribution, the model can be “weighted” such that some data are fit more closely (with less observation error such that the predicted and observed values are more similar) than others. Correctly weighting among multiple data sources in integrated stock assessment models is an ongoing challenge especially when data have conflicting trends (Maunder and Piner, 2017; Maunder and Punt, 2013; Punt *et al.*, 2013). Changing the data weighting or the likelihood function as a whole can cause substantial changes in model output and model evaluation tools like AIC as the model fits certain data points more accurately (Francis, 2014, 2011). It is therefore imperative to specify an observation error



structure that accurately reflects the data collection process and correctly weights among, for example, catch, discards, indices of abundance, and tagging data.

Age composition, or the fraction of each age represented in the catch or index, is recorded by sub-sampling individuals from the catch or survey and aging from otoliths, scales, or other bony structures. In standard statistical-catch-at-age models, the proportions at age are often assumed to have observation error that has a multinomial distribution (MN) (Fournier and Archibald, 1982; Legault and Restrepo, 1998; Methot and Wetzel, 2013). However, such a distribution assumes that individuals are sampled independently, when in reality factors like schooling or targeted fishing may lead to overdispersion (more variability than expected) and “intra-haul correlation” among age groups (Pennington and Vølstad, 1994). Therefore, alternative composition distributions with more desirable properties that can account for overdispersion and inter-age correlation have been explored instead of the multinomial (Francis, 2014; Maunder, 2011). In a state-space modeling framework, which include one or more time-varying random effects which are integrated from the marginal likelihood, the process variation is self-weighting and the applicability of alternative proportions at age likelihoods may be contingent on how well they can parse variance among data sources and between the process and observation sub-models (Besbeas and Morgan, 2017; Francis, 2017; Linton and Bence, 2008). The multinomial is still recommended over some of the alternative proportions at age distributions (listed below) when the model accounts for time-varying selectivity (Cronin-Fine and Punt, 2021).

The Dirichlet-multinomial (DM) is a hierarchical distribution that assumes that the effective sample size of a multinomial distribution follows a Dirichlet distribution (Francis, 2017; Thorson *et al.*, 2017). The DM outperforms other data weighting schemes like iteratively

reweighting the multinomial effective sample size or estimating effective sample sizes with a strong prior (Thorson *et al.*, 2017; Xu *et al.*, 2020). The logistic normal (LG) is another hierarchical distribution, but unlike DM can account for substantial correlation among age classes. LG is a transformed multivariate-normal distribution that can be parameterized to be linear, or an autoregressive first order AR(1) or second order AR(2) process, or moving average ARMA (Fisch *et al.*, 2021; Francis, 2014). Its flexibility and ability to self-weight has led the LG distribution, and others that can account for correlations, to outperform the DM especially when there's moderate process error in selectivity and the composition sample size is large (Albertsen *et al.*, 2017; Fisch *et al.*, 2021). The LG distribution is a recommended option when employing state-space stock assessment software WHAM and when there is variance in numbers at age, natural mortality, selectivity, or environmental covariates (Stock *et al.*, 2021; Stock and Miller, 2021). The multivariate Tweedie distribution (MVTW) has recently been adapted for application in fisheries stock assessment. The MVTW is an attractive alternative to the multinomial that assumes that the proportions at age are derived from a transformed gamma and arises naturally from the process of subsampling age and length information (Thorson *et al.*, 2022). Though previous research and application are limited for MVTW, it has outperformed the DM and performed comparably to the LG in a state-space model with time-varying selectivity and is already included in the state-space modeling software WHAM (Stock and Miller, 2021; Thorson *et al.*, 2022). The MVTW can upweight the data (produce effective sample sizes *larger* than the input sample size, which combats the risk of setting the input sample size too low) and estimate heteroscedasticity unlike the DM, and account for zeros in the data, unlike the LG, so it exhibits many of the stated important qualities for an catch proportions at age distribution (Francis, 2014; Thorson *et al.*, 2022).

Models may perform differently and produce biased model output when they do not account for overdispersion or inter-age correlation when these factors are present or when they do try to account for these processes when they are absent, particularly when the model is also accounting for several other stochastic processes. To address this question, catch data were simulated using alternative age composition scenarios that assumed several observation error distributions for the proportions at age (of those described above), then the simulated data were fit using estimation models that assumed different likelihoods for the proportions at age, and different sources of process variability in recruitment, expected survival, and selectivity. The additional process variability scenarios were intended to test how well the model can apportion variability when there are one or several time-varying processes.

### 3.3 Methods

#### 3.3.1 Overview

Catch data were simulated using a state-space stock assessment operating model that was based on a model fit with Gulf of Maine haddock data. The operating model generated total catch and catch proportions at age using four scenarios that assumed the observation error in total catch had a lognormal distribution and the observation errors in catch proportions at age had a multinomial, Dirichlet-multinomial, logistic-normal with AR(1) correlations, or multivariate Tweedie distribution. A fifth simulation scenario modeled observation error in catch-at-age, rather than the total catch and proportions separately, using a multivariate lognormal distribution with AR(1) error (MVN). The MVN distribution for catch-at-age is the default assumption in the state-space modeling software SAM and a common observation error distribution used to fit Canadian stocks (Berg and Nielsen, 2016; Nielsen and Berg, 2014; Perreault *et al.*, 2020). The operating model included intra-annual process variance in recruitment, expected survival, and

selectivity, with an autocorrelated AR(1) distribution. Yearly deviations in these processes were not simulated from their respective distributions, but rather fixed at the estimated values, such that the realized abundance, spawning stock biomass, and fishing mortality was the same for every generated data set. The simulated data were fit using several estimation models that differed in the assumed likelihood of the proportions at age, either MN, DM, LG or MVTW, but not MVN because no such option exists in the modeling package used for fitting. Several versions of the estimation models were also used which differed in which processes (of recruitment, expected survival or selectivity) had process errors. Model output from fitting simulated data were compared to the realized values used to generate that data to assess model performance.

### *3.3.2 Simulation/Operating Model Scenarios*

Each simulation model generated total catch, total index, and catch and index age composition data (for ages 1-9+) in each year using the predicted values from an state-space stock assessment model fit with Gulf of Maine haddock catch and index data from 1977-2019 using the WHAM state-space assessment package (Stock and Miller, 2021). The process and observation sub-model were identical to that described in Table 2.1 and summarized below. The process sub-model assumed time-varying random effects in recruitment, expected survival at age, and selectivity at age, and the yearly deviations were assumed to have an autoregressive AR(1) process with inter-annual correlation. In the observation sub-model, catch at age was predicted each year from a Baranov catch equation and the predicted total catch and proportions at age were derived from these age-specific values. Indices of abundance from a spring and fall bottom trawl survey were predicted from the abundance at age, index catchability, and index selectivity, and similarly decomposed into total index and index age compositions. The model

estimated fixed effect parameters of mean recruitment, mean selectivity at age, numbers at age for ages 1-6 in the first year, index catchability and selectivity, and the variance and correlation parameters of the random effects. The model integrated the random effects from the marginal likelihood and predicted values of the random effects and derived quantities were calculated using the epsilon method for bias correction (Thorson and Kristensen, 2016). For a full description of the process and observation sub-models see section 2.3.2.

The operating model generated catch and index data using the same estimated fixed effect and random effect values from the model fitted to real data, i.e., the recruitment, expected survival, and selectivity deviations were kept the same and not re-drawn from their respective distributions. Therefore, the process variance was the same across simulation scenarios. The total index and index proportions at age were not simulated, but rather matched that of the Gulf of Maine Haddock data. What distinguished the five simulation scenarios was the assumed probability distribution of the observation error in total catch and catch proportions at age (Table 3.1).

Table 3.1. Simulation distributions of total catch, proportion at age, or catch-at-age used in likelihood simulation scenarios and likelihood equations used in likelihood estimation scenarios.  $LN()$  denotes lognormal distribution with mean and standard deviation.  $N()$  is a normal distribution with mean and standard deviation.  $MN()$  is a multinomial distribution with effective sample size and predicted proportions.  $Gam()$  denotes gamma distribution with shape and scale parameters.  $Pois()$  is a Poisson distribution with mean/variance parameter.  $\Gamma()$  is the gamma function and  $det()$  is the determinant of a matrix. Predicted values of total catch  $\hat{C}_y$  in weight, proportions at age  $\hat{P}_{a,y}$ , catch-at-age in numbers,  $\hat{C}_{a,y}$ , and average weight at age at time of capture,  $W_{a,y}$ , are in Appendix 3A.

Index	Equation	Variance and Parameter Values
<b>Simulation Model</b>		
<i>Total Catch</i>		
Sim		$\sigma_{C_y} = 0.2,$
Lognormal		$1977 \leq y \leq 1988$
	$C_y \sim LN(\hat{C}_y, \sigma_{C_y})$	$\sigma_{C_y} = 0.15,$
		$1989 \leq y \leq 2000$
		$\sigma_{C_y} = 0.1,$
		$2001 \leq y \leq 2019$
<i>Catch Proportions at Age</i>		
SimMulti		$n_y^{(C)} = 60,$
		$1977 \leq y \leq 1987$
		$n_y^{(C)} = 20,$
		$1988 \leq y \leq 1992$
		$n_y^{(C)} = 60,$
		$1993 \leq y \leq 2002$
		$n_y^{(C)} = 140,$
		$2003 \leq y \leq 2019$
SimDirMult		$n_y^{(C)}$ see SimMulti
		$\beta = 18.504$
	$P_{a,y} \sim MN(n_y^{(C)}, \hat{P}_{a,y}^*)$	
	$\hat{P}_{a,y}^* = \frac{X_{a,y}}{\sum_{a=1}^A X_{a,y}}$	
	$X_{a,y} \sim Gam(k_{a,y}, 1)$	
	$k_{a,y} = \hat{P}_{a,y} * \beta$	
SimLogi		$n_y^{(C)}$ see SimMulti
		$\sigma = 1.95$
		$\sigma_y = \exp(\sigma +$
		$\frac{1}{2} \log(n_y^{(C)}))$
		$\rho = 0.648$
	$P_{A,y} = 1 - \sum_{a=1}^{A-1} P_{a,y}$	
	$P_{a,y} = \frac{P_{a,y}^*}{1 + \sum_{a=1}^{A-1} P_{a,y}^*}, a < A$	
	$P_{a,y}^* \sim N(\mu_{a,y}, \Sigma_y), a < A$	
	$\Sigma_{y,a,\tilde{a}} = \rho^{ a-\tilde{a} } \sigma_y^2, a < A, \tilde{a} < A$	
	$\mu_{a,y} = \log(\hat{P}_{a,y}) - \log(\hat{P}_{A,y}), a < A$	

Table 3.1 (cont'd)

SimTwe

$$\begin{aligned}
 P_{a,y} &= \frac{P_{a,y}^*}{\sum_{a=1}^A P_{a,y}^*} \\
 P_{a,y}^* &\sim \text{Gam}(k_{a,y}, \gamma_{a,y}) \\
 k_{a,y} &= -\text{Pois}(\lambda_{a,y})\alpha \\
 \gamma_{a,y} &= \phi(\psi - 1)\mu_{a,y}^{\psi-1} \\
 \alpha &= \frac{2 - \psi}{1 - \psi} \\
 \lambda_{a,y} &= \frac{\mu_{a,y}^{2-\psi}}{\phi(2 - \psi)} \\
 \mu_{a,y} &= \hat{P}_{a,y}n_y^{(C)}
 \end{aligned}$$

$$\begin{aligned}
 n_y^{(C)} &\text{ see SimMulti} \\
 \psi &= 1.294 \\
 \phi &= 0.296
 \end{aligned}$$

Catch at Age  
SimMultiV

$$\begin{aligned}
 C_{a,y} &\sim N(\hat{C}_{a,y}, \Sigma_y) \\
 \Sigma_{a,\tilde{a}} &= \rho^{|\tilde{a}-a|}\sigma_a\sigma_{\tilde{a}}, 1 < a < A, 1 < \tilde{a} < A
 \end{aligned}$$

$$\begin{aligned}
 \rho &= 0.024 \\
 \sigma_1 &= 0.599 \\
 \sigma_2 &= 0.478 \\
 \sigma_3 &= 0.424 \\
 \sigma_4 &= 0.389 \\
 \sigma_5 &= 0.412 \\
 \sigma_6 &= 0.430 \\
 \sigma_7 &= 0.444 \\
 \sigma_8 &= 0.490 \\
 \sigma_{9+} &= 0.448
 \end{aligned}$$

$$\begin{aligned}
 C_y &= \sum_{a=1}^A C_{a,y}W_{a,y} \\
 P_{a,y} &= \frac{C_{a,y}}{\sum_{a=1}^A C_{a,y}}
 \end{aligned}$$

**Estimation Model Negative Log Likelihood**

Total Catch

Est  
Lognormal

$$\sum_{y=1977}^Y \frac{1}{2\sigma_{C_y}^2} \left(\log \frac{C_y}{\hat{C}_y}\right)^2 + \log \sigma_{C_y}$$

$\sigma_{C_y}$  is same as  
SimLognormal

Catch Proportions at Age

EstMult

$$- \sum_{y=1977}^Y \sum_{a=1}^A n_y^{(C)} \hat{P}_{a,y} \log P_{a,y}$$

$n_y^{(C)}$  see SimMulti

EstDirMult

$$\begin{aligned}
 &- \sum_{y=1977}^Y \log(\Gamma(n_y^{(C)} + 1) + \log(\Gamma(\varphi_y) - \log(\Gamma(n_y^{(C)} + \varphi_y) \\
 &\quad + \sum_{a=1}^A (-\log(\Gamma(n_y^{(C)}P_{a,y} + 1) \\
 &\quad + \log(\Gamma(n_y^{(C)}P_{a,y} + \hat{P}_{a,y} * \beta) \\
 &\quad - \log(\Gamma(\hat{P}_{a,y} * \beta))] \\
 &\quad \varphi_y = \sum_{a=1}^A k_{a,y} \\
 &\quad k_{a,y} = \hat{P}_{a,y} * \beta
 \end{aligned}$$

$n_y^{(C)}$  see SimMulti  
 $\beta$  estimated on the log  
scale

Table 3.1 (cont'd)

EstLogi

$$\begin{aligned}
 & - \sum_{y=1977}^Y \left( \log \frac{1}{\sqrt{(2\pi)^{\frac{A-1}{2}} \det(\Sigma)}} \right) \\
 & \quad - \frac{1}{2} \left( (\mathbf{X}_y - \boldsymbol{\mu}_y)^T \boldsymbol{\Sigma}^{-1} (\mathbf{X}_y - \boldsymbol{\mu}_y) \right) \\
 & \quad + \sum_{a=1}^{A-1} \log(P_{a,y}) \\
 & \quad \Sigma_{y,a,\tilde{a}} = \rho^{|a-\tilde{a}|} \sigma^2, a < A, \tilde{a} < A \\
 & \quad \mu_{a,y} = \log(\hat{P}_{a,y}) - \log(\hat{P}_{A,y}), a < A \\
 & \quad X_{a,y} = \log(P_{a,y}) - \log(P_{A,y}), a < A
 \end{aligned}$$

$\sigma$  estimated on the log scale, one value for all years  
 $\rho$  estimated on the logit scale  
 $\mathbf{X}_y$  and  $\boldsymbol{\mu}_y$  are vector notation and contain all ages for a given year

EstTwe

$$\sum_{y=1977}^Y -\log \left( \prod_{a=1}^A dtweedie(P_{a,y} | \hat{P}_{a,y}, n_y^{(C)}, \psi, \phi) \right)$$

$n_y^{(C)}$  see SimMulti  
 $(\psi - 1)$  estimated on logit scale (to constrain between 1 and 2)  
 $\phi$  estimated on log scale

Note- there is no closed form version of the multivariate Tweedie distribution. See Thorson *et al.* (2022)

The first scenario mimicked the structure of the model fit to real Gulf of Maine haddock data such that total catch was lognormally distributed with the mean equal to the predicted value in each year, and the catch proportions at age had a multinomial distribution (“SimMulti”) with the probability equal to the predicted values (for predicted values, see Appendix 3A). The standard deviation and effective sample size of these distributions, respectively, were the same as in the Gulf of Maine haddock data (see table 2A.1). Other simulation scenarios assumed the total catch was drawn from a lognormal distribution, like the first model, but the proportions at age were generated using either a Dirichlet-multinomial (“SimDirMult”), logistic-normal distribution with AR(1) correlations (“SimLogi”), or multivariate Tweedie distribution (“SimTwe”). A final simulation scenario assumed that the catch-at-age, rather than total catch and proportions at age, was simulated from a multivariate lognormal distribution (“SimMultiV”) but decomposed into



total catch and proportions at age before being supplied to the estimation models. The data were simulated in R rather than using the simulation options built into the package WHAM, and the package MASS was used for multivariate lognormal distributions (a component of “SimLogi” and “SimMultiV”), and the package tweedie was used for “SimTwe” (Venables and Ripley, 2002; Dunn, 2022).

The Dirichlet-multinomial, logistic-normal, and multivariate Tweedie distributions all contain additional “age composition parameters” that determine the variance and/or correlation of the observation error distribution. To ensure that the observation error variance was approximately the same across scenarios, appropriate age composition parameters were found by fitting the Gulf of Maine haddock data assuming a DM, LG, or MVTW distribution for the proportions at age while fixing all the parameters at their estimated values except the age composition parameter. All the data inputs were kept the same as well, such that the alternative proportion at age models were fit using the same input sample size as the model fit using the multinomial distribution. Because there was no option to fit the catch at age data to a multivariate lognormal distribution in WHAM, an alternative approach was used to find appropriate standard deviation and correlation values for the multivariate log normal distribution that approximated the variance of the lognormal catch and multinomial proportions at age. 100,000 data sets of total catch and proportion at age were generated from the lognormal and multinomial distributions, converted into catch-at-age, and the average standard deviation at age and correlation of the residuals between the observed and predicted values were estimated. Each simulation scenario also necessitated an input sample size, which was assumed to equal the effective sample size in the Gulf of Maine haddock data. In total, 100 data sets were generated for each of the five simulation scenarios.

### *3.3.3 Estimation Model Scenarios*

Simulated data were fit by different estimation models that included four different assumptions about the distribution for the observed age-compositions (observation scenarios) and five different assumptions about process variability (process scenarios) estimation scenarios and five process variance scenarios. The observation scenarios assumed the catch proportions at age followed a multinomial “EstMulti,” Dirichlet-multinomial “EstDirMult,” logistic normal “EstLogi,” or multivariate Tweedie “EstTwe” distribution, each corresponding to one of the simulation scenarios (Table 3.1). The logistic normal distribution “EstLogi” cannot account for zeros in the proportions at age data and therefore zeros were pooled with adjacent age bins automatically in WHAM. Each observation scenario was crossed with process scenarios that assumed random effects in recruitment, expected survival, and selectivity which were integrated from the marginal likelihood (which matches the operating model), or that one or more process was constant or a fixed effect (which does not match the operating model). The process scenarios were designated by a three-letter code which indicated if the recruitment, expected survival, or selectivity was specified as a random effect (R), a fixed effect (F), or constant (C). See section 2.3.4 for a complete description of the process variation estimation scenarios. The intent was to investigate trade-offs in weighing total model variability between the observation and process variance. In total there were 100 combinations of simulation and estimation model.

### *3.3.4 Performance of Assessment Models*

The estimation models were assessed by how precisely and accurately they estimated spawning stock biomass in the terminal year, average recruitment in the terminal 3 years, exploitation rate in the terminal year, and the depletion, defined as the ratio of the predicted spawning stock biomass in the terminal year to that 10 years prior to the terminal year. Metrics in

the terminal year are often the basis for management reference points and management decisions, so evaluating their accuracy is of primary interest. A retrospective analysis was performed, which involved refitting the model after sequentially removing the ultimate seven years of data to investigate how recent observations influenced the estimates, and retrospective patterns were quantified by Mohn's Rho statistic for spawning stock biomass, average fishing mortality, and recruitment (Hurtado-Ferro *et al.*, 2015; Mohn, 1999). See section 2.4.5 and Table 2.2 "Derived Quantities and Simulation Output" for a complete description of these metrics and how relative error was calculated.

Additional performance metrics were evaluated if the estimation model correctly specified the likelihood of the proportions at age i.e., a "self-test." When the simulation and estimation models matched in the assumed distributions for proportions at age the estimated age composition parameters, which quantify the observation error variance, were compared against the true value used to generate the data. The bias in estimated process error variance and correlation coefficients in recruitment, expected survival, and selectivity cannot be determined because these processes were not simulated with a known variance value, however these estimates were compared against that estimated from fitting to the real Gulf of Maine Haddock. "Biased" and "unbiased" were used to explain how closely the estimates from simulated data matched those from the original estimation model on which the simulations are based.

### 3.4 Results

The estimation models successfully converged and produced an invertible Hessian matrix for most replicates in most combinations of simulation and estimation scenario (Table 3.2). However, when the likelihood simulation scenario was multinomial and the estimation scenario was multivariate Tweedie ("SimMulti" + "EstTwe"), no replicates ran successfully for any

assumed process variance model. Similarly, when the data were generated from a multivariate Tweedie and fit assuming a Dirichlet-multinomial (“SimTwe” + “EstDirMult”), only 19 out of 500 simulated data sets successfully converged. Convergence rate was zero for several cases where the estimation model assumed that age compositions followed a -multinomial, (“EstDirMult”), especially for the (“RRR”) process scenario. Convergence rate was lowest when the simulated data had a multinomial distribution (64.7%), and highest when the simulated data had a Dirichlet-multinomial distribution (95.75%). Convergence rate was lowest when the model assumed process variance in all three processes (72.85%) and highest when the survival was deterministic and selectivity was constant (90.8% for RCC and 90.05% for FCC).

Table 3.2. Convergence rate across likelihood simulation scenarios, likelihood estimation scenarios, and process variance estimation scenarios. Convergence criteria included a successful model run without errors and production of an invertible Hessian matrix.

Likelihood Simulation Scenario	Likelihood Estimation Scenario	Process Variance Scenarios					Average	Median	Min	Max
		RRR	RRC	RCR	RCC	FCC				
SimMulti	EstMulti	100	100	100	100	100	100	100	100	100
SimMulti	EstDirMult	2	15	91	100	100	61.6	91	2	100
SimMulti	EstLogi	93	100	93	100	100	97.2	100	93	100
SimMulti	EstTwe	0	0	0	0	0	0	0	0	0
SimDirMult	EstMulti	96	100	99	100	100	99	100	96	100
SimDirMult	EstDirMult	96	100	94	100	100	98	100	94	100
SimDirMult	EstLogi	83	100	59	100	100	88.4	100	59	100
SimDirMult	EstTwe	98	100	90	100	100	97.6	100	90	100
SimLogi	EstMulti	100	100	100	100	100	100	100	100	100
SimLogi	EstDirMult	0	62	96	100	100	71.6	96	0	100
SimLogi	EstLogi	96	100	95	100	100	98.2	100	95	100
SimLogi	EstTwe	100	100	100	100	100	100	100	100	100
SimTwe	EstMulti	100	100	100	100	100	100	100	100	100
SimTwe	EstDirMult	0	0	0	17	2	3.8	0	0	17
SimTwe	EstLogi	98	100	100	100	99	99.4	100	98	100
SimTwe	EstTwe	100	100	100	100	100	100	100	100	100
SimMultiV	EstMulti	100	100	100	100	100	100	100	100	100
SimMultiV	EstDirMult	0	6	92	100	100	59.6	92	0	100
SimMultiV	EstLogi	99	100	100	99	100	99.6	100	99	100
SimMultiV	EstTwe	96	100	99	100	100	99	100	96	100

Table 3.2 (cont'd)

<b>Average</b>	72.85	79.15	85.4	90.8	90.05
<b>Median</b>	96	100	97.5	100	100
<b>Min</b>	0	0	0	0	0
<b>Max</b>	100	100	100	100	100

Estimation models that assumed no random effects (“FCC”), only process variance in recruitment (“RCC”), or only process variance in recruitment and selectivity (“RCR”) had similar distributions of relative errors (i.e., similar precision and bias) across all model evaluation criteria. Collectively, these estimation models can be referred to as the “deterministic survival” scenarios because they all assume that the expected numbers at age do not deviate from that predicted by the abundance in the previous year and the instantaneous mortality, and the other scenarios assume a random effect on survival.

Spawning stock biomass (SSB) in the terminal year was underestimated when the estimation model assumed deterministic survival (Figure 3.1). Conversely, SSB was overestimated when the estimation model assumed process variance in only recruitment and expected survival (“RRC”), except when the Dirichlet-multinomial or logistic normal was correctly specified (“SimDirMult” + “EstDirMult” or “SimLogi” + “EstLogi”). When the process scenario “RRR” was correctly specified in the estimation model, the model over-, under- or accurately estimated SSB, depending upon the simulated and assumed distribution for the age composition data and was least biased when the simulated data were logistic-normal (“SimLogi”).

Average recruitment in the terminal three years was underestimated when the estimation model assumed deterministic survival (Figure 3.2). When the expected survival was a random effect, the average recruitment could be over-, under- or accurately estimated depending on the

scenario. The average recruitment was most precise and bias was lowest when the estimation model assumed a logistic normal distribution (“EstLogi”).

Exploitation rate in the terminal year was overestimated when the estimation model assumed deterministic survival, except when the catch at age data were simulated from a multivariate lognormal distribution (“MultiV”) (Figure 3.3). Estimation models that assumed process variability in recruitment and expected survival (“RRC”) underestimated or accurately estimated exploitation rate and was the least biased when the Dirichlet-multinomial or logistic normal distributions were correctly specified. When the process model correctly assumed variability in all three processes, the exploitation rate was not substantially biased, especially when the estimation model assumed a Logistic normal or multivariate Tweedie distribution.

Depletion was underestimated when the estimation model assumed deterministic survival (Figure 3.4). Depletion was slightly negatively biased even when the model did assume process variability in expected survival and the bias was often less when the selectivity was assumed to be constant.

The retrospective patterns in spawning stock biomass (SSB) was negative and progressively larger in magnitude in estimation models that specified more of the processes as constant or fixed effects rather than as random effects (Figure 3.5). The same pattern was seen for the magnitude of retrospective patterns in average fishing mortality, though in this case the retrospective bias was positive (Figure 3.6). The retrospective patterns in recruitment was also consistently negative across all simulation and estimation scenarios (like SSB) but estimation models that assumed constant selectivity exhibited less retrospective patterns than models that specified selectivity was a random effect (Figure 3.7).

Recruitment process error variance was estimated with high precision and low bias, except when age composition data were simulated from a Dirichlet-multinomial in which case it was overestimated (Figure 3.8). The expected survival process variance tended to be overestimated, and the bias was greater when the simulated data were Dirichlet-multinomial (Figure 3.9). The selectivity process was underestimated or greatly overestimated, though the bias was less when the simulated data were multivariate lognormal (Figure 3.10). The correlation coefficient for the AR(1) process error in recruitment and expected survival was estimated with the least bias when the model assumed constant selectivity (--C) and overestimated when the expected survival was deterministic (-C-) (Figure 3.11). The correlation coefficient for the AR(1) process error in selectivity was strongly underestimated or slightly overestimated, and the bias was worse when the simulated data were logistic normal (Figure 3.12).

When the simulation and estimation model both assumed a Dirichlet-multinomial distribution for the age compositions, the scaling parameter,  $\beta$  was estimated with low bias when the estimation model assumed two of the processes were varying and one was constant (RRC or RCR), but not when all three were varying (Figure 3.13A).  $\beta$  was overestimated (so the effective sample size was higher and the variance was lower) when the process variance scenario was RRR and was underestimated (corresponding to higher estimated variance) when the survival was deterministic and selectivity was constant (-CC). A similar pattern of bias was present in the logistic normal distribution which estimated variance parameter,  $\sigma$ . The observation error variance was accurately estimated when only two processes were assumed to be random effects but predicted lower variance when all processes are varying and higher variance when survival was deterministic and selectivity was constant. (Figure 3.13B). The inter-age correlation of the age compositions,  $\rho$ , was estimated with little bias and high precision when selectivity was a

random effect but overestimated when selectivity was constant. The pattern was also present in the estimation of multivariate Tweedie parameters  $\phi$  and  $\psi$  (Figure 3.13C). Larger  $\phi$  or  $\psi$  parameters correspond to a smaller effective sample size and therefore greater observation variance. Like the Dirichlet-multinomial or the logistic-normal distributions, the multivariate Tweedie distribution predicted more observation variability when the process model assumed less process variability and less observation variability with the process model assumed substantial process variability.



Figure 3.1. Relative error of spawning stock biomass in the terminal year across alternative age composition simulation (rows) and estimation (columns) scenarios and process variance estimation scenarios (individual boxes). The number above each box indicates the number of the 100 generated data sets that were successfully fit. The catch proportions at age were either simulated (Sim) or estimated (Est) as a multinomial distribution (Multi), Dirichlet-multinomial distribution (DirMulti), logistic normal distribution (Logi), multivariate Tweedie distribution (Twe), or multivariate normal distribution (MultiV) (catch at age rather than proportions). Individual boxplots represent the five estimation models that assume recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Black lines are drawn through zero, red dots indicate the mean, and black dots indicate the median. When no plot is present, no replicates converged for that combination of simulation and estimation scenarios.

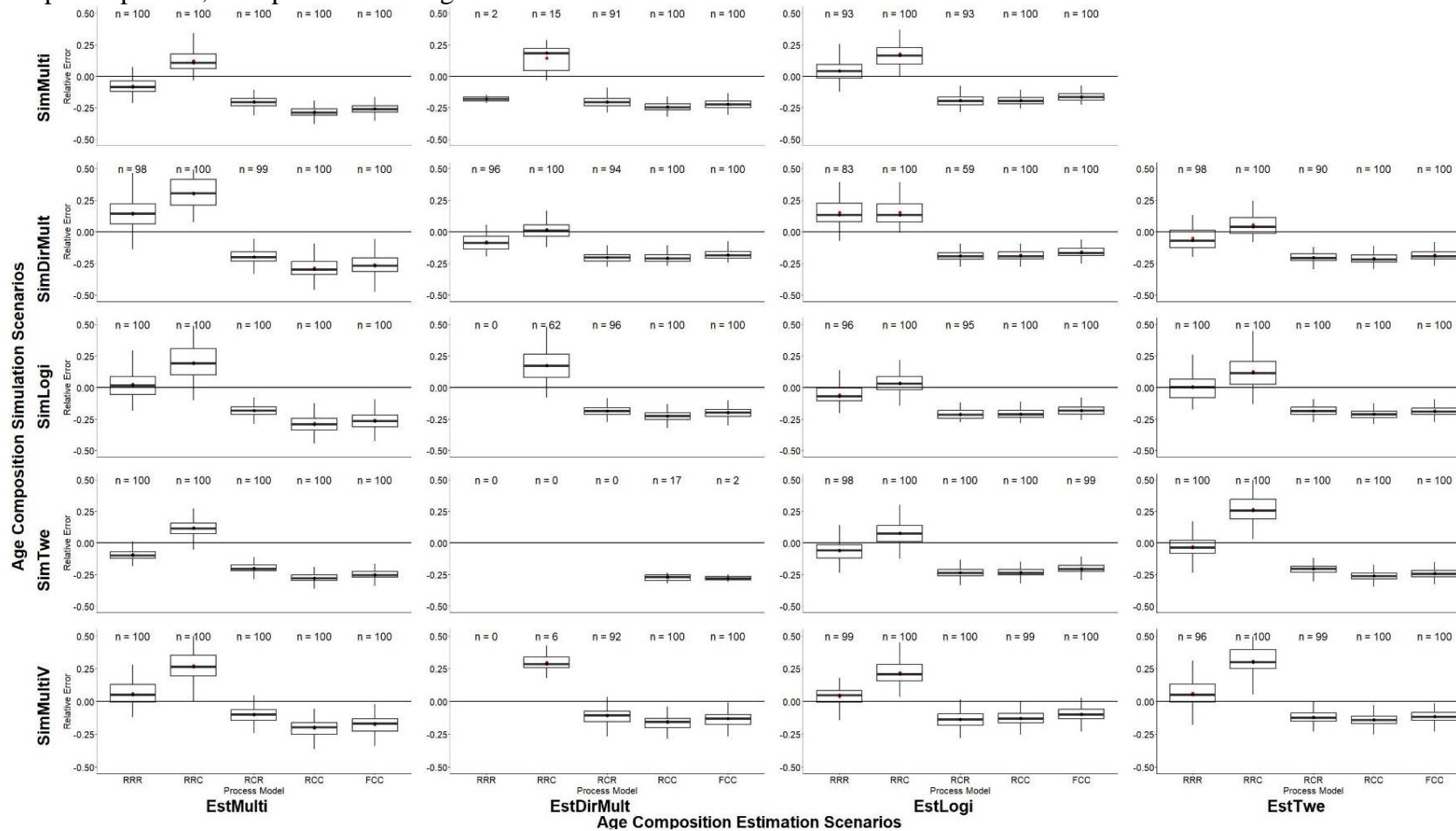


Figure 3.2. Relative error of average recruitment in the terminal three years across alternative age composition simulation (rows) and estimation (columns) scenarios and process variance estimation scenarios (individual boxes). The number above each box indicates the number of the 100 generated data sets that were successfully fit. The catch proportions at age were either simulated (Sim) or estimated (Est) as a multinomial distribution (Mult), Dirichlet-multinomial distribution (DirMult), logistic normal distribution (Logi), multivariate Tweedie distribution (Twe), or multivariate normal distribution (MultiV) (catch at age rather than proportions). Individual boxplots represent the five estimation models that assume recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Black lines are drawn through zero, red dots indicate the mean, and black dots indicate the median. When no plot is present, no replicates converged for that combination of simulation and estimation scenarios.

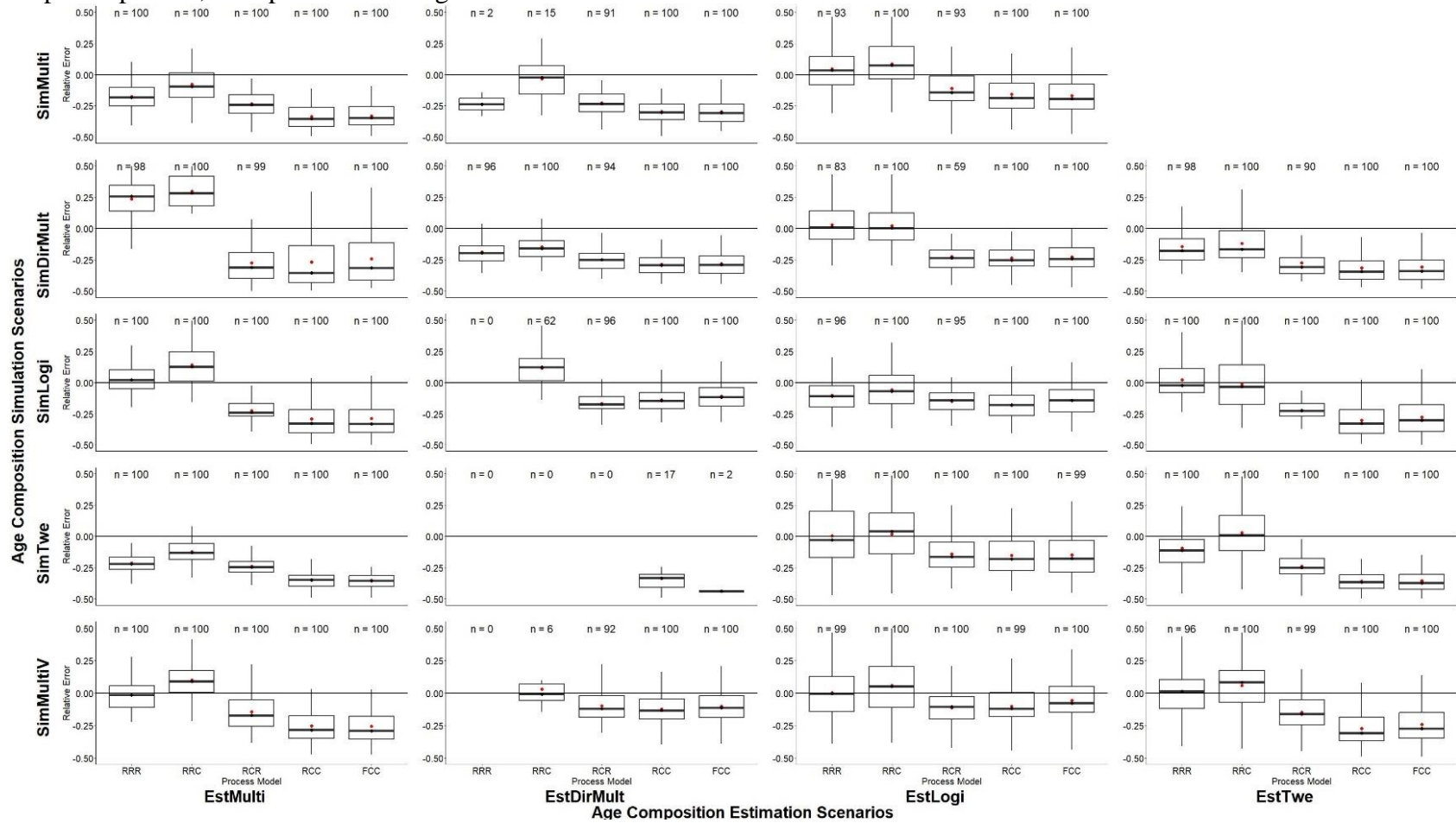


Figure 3.3. Relative error of exploitation rate in the terminal year across alternative age composition simulation (rows) and estimation (columns) scenarios and process variance estimation scenarios (individual boxes). The number above each box indicates the number of the 100 generated data sets that were successfully fit. The catch proportions at age were either simulated (Sim) or estimated (Est) as a multinomial distribution (Multi), Dirichlet-multinomial distribution (DirMulti), logistic normal distribution (Logi), multivariate Tweedie distribution (Twe), or multivariate normal distribution (MultiV) (catch at age rather than proportions). Individual boxplots represent the five estimation models that assume recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Black lines are drawn through zero, red dots indicate the mean, and black dots indicate the median. When no plot is present, no replicates converged for that combination of simulation and estimation scenarios.

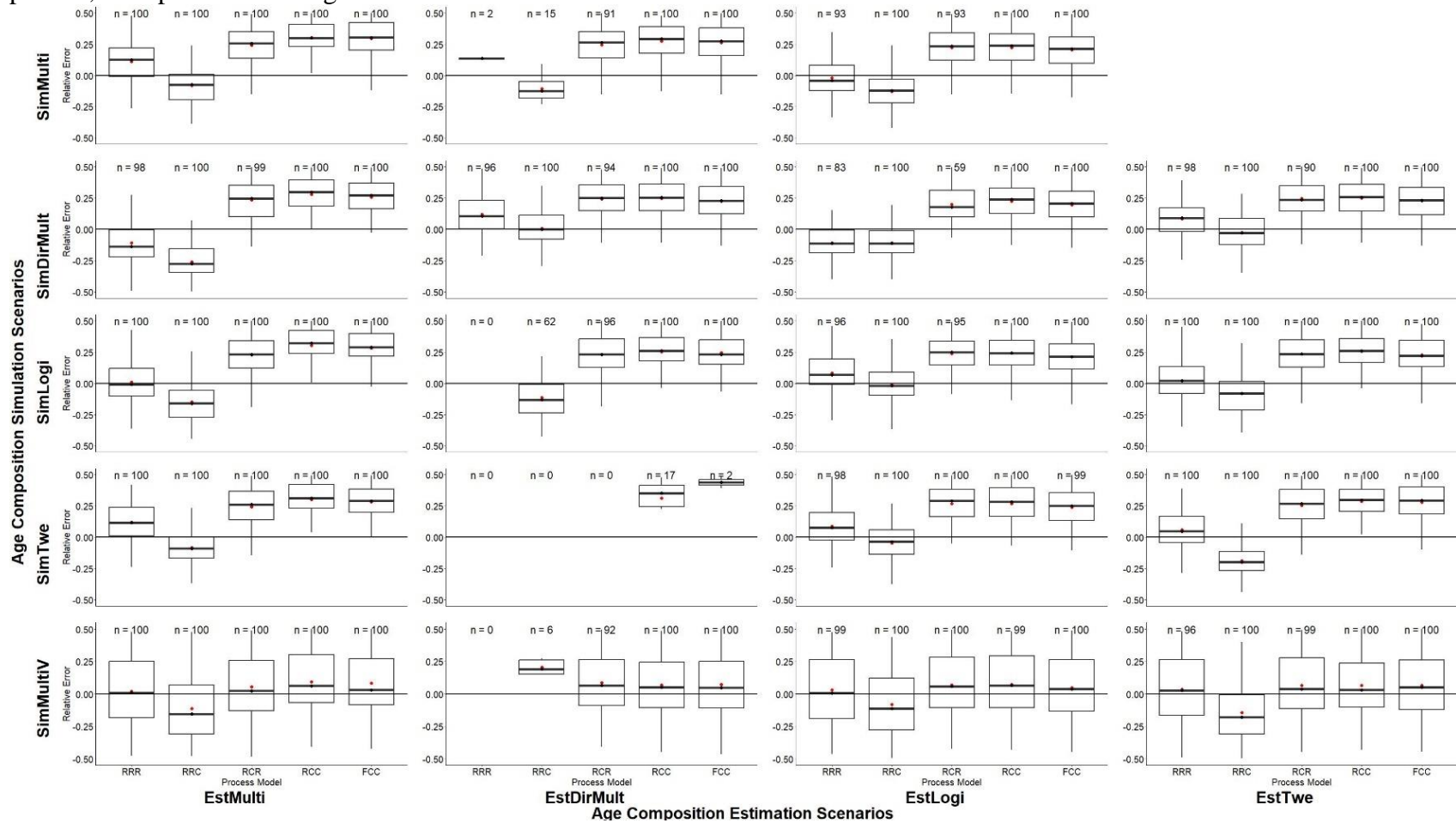


Figure 3.4. Relative error of depletion across alternative age composition simulation (rows) and estimation (columns) scenarios and process variance estimation scenarios (individual boxes). The number above each box indicates the number of the 100 generated data sets that were successfully fit. The catch proportions at age were either simulated (Sim) or estimated (Est) as a multinomial distribution (Multi), Dirichlet-multinomial distribution (DirMulti), logistic normal distribution (Logi), multivariate Tweedie distribution (Twe), or multivariate normal distribution (MultiV) (catch at age rather than proportions). Individual boxplots represent the five estimation models that assume recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Black lines are drawn through zero, red dots indicate the mean, and black dots indicate the median. When no plot is present, no replicates converged for that combination of simulation and estimation scenarios.

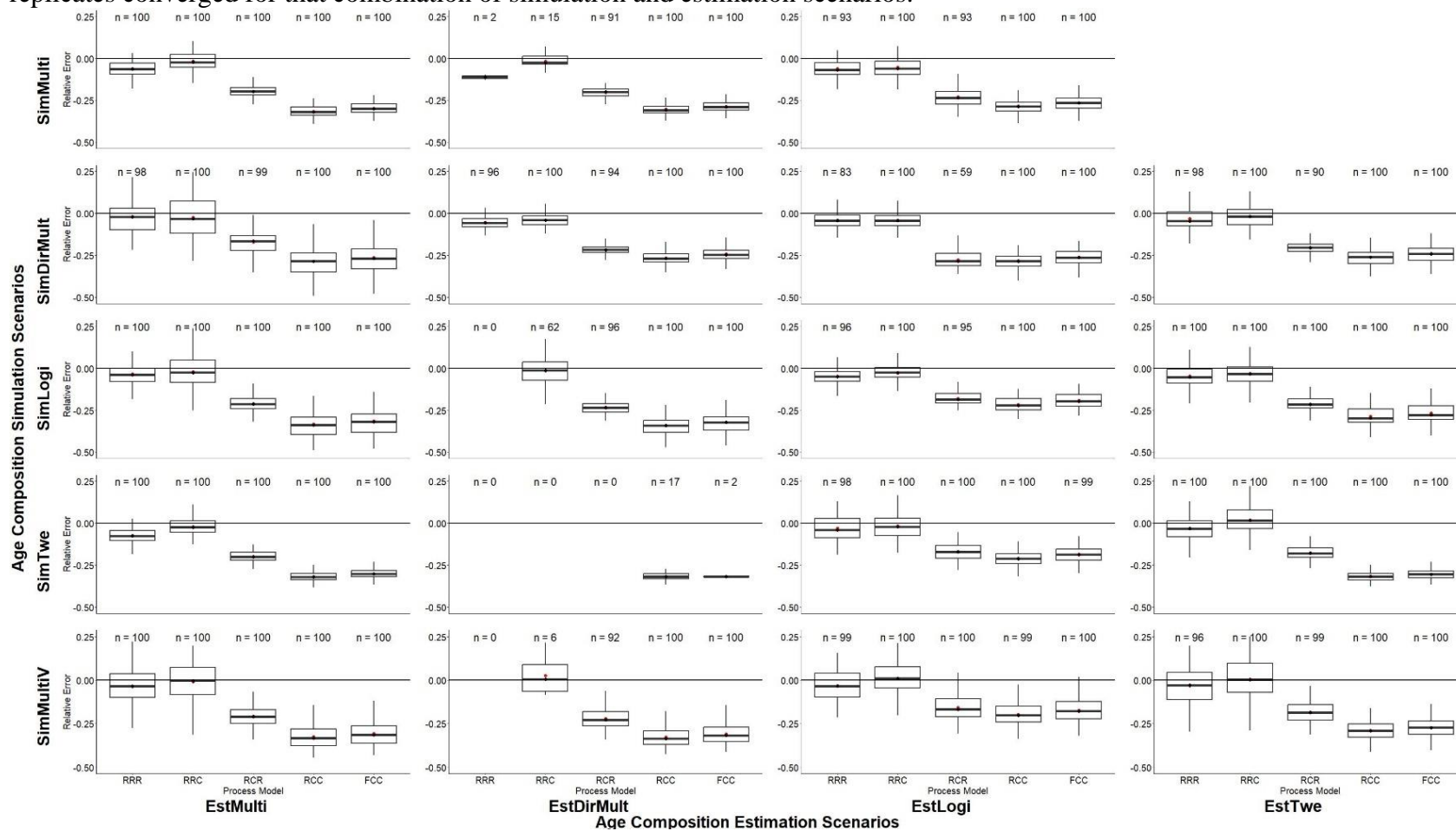


Figure 3.5. Mohn's Rho in spawning stock biomass across alternative age composition simulation (rows) and estimation (columns) scenarios and process variance estimation scenarios (individual boxes). The number above each box indicates the number of the 100 generated data sets that were successfully fit. The catch proportions at age were either simulated (Sim) or estimated (Est) as a multinomial distribution (Multi), Dirichlet-multinomial distribution (DirMulti), logistic normal distribution (Logi), multivariate Tweedie distribution (Twe), or multivariate normal distribution (MultiV) (catch at age rather than proportions). Individual boxplots represent the five estimation models that assume recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Black lines are drawn through zero, red dots indicate the mean, and black dots indicate the median. When no plot is present, no replicates converged for that combination of simulation and estimation scenarios.

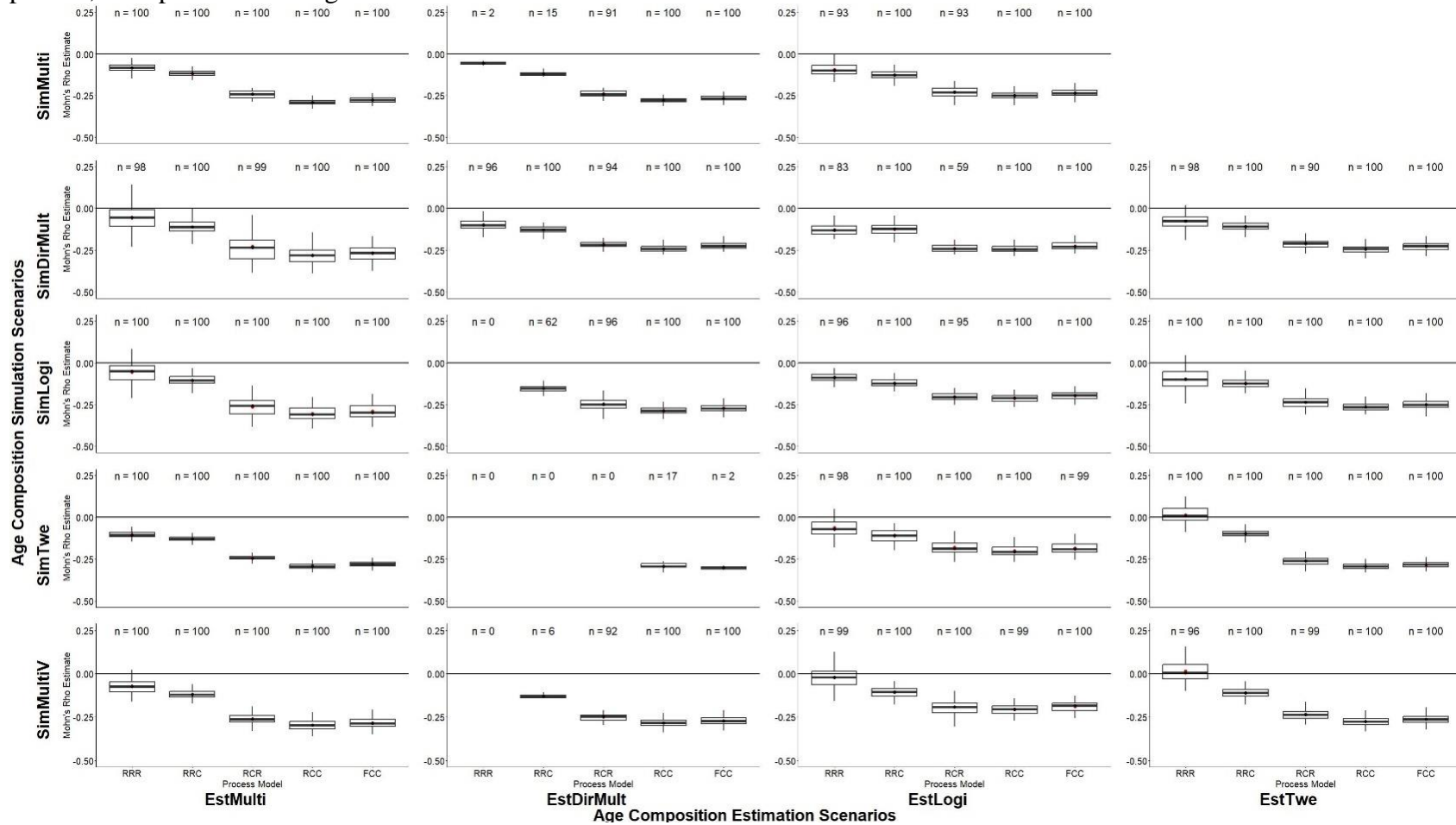




Figure 3.6. Mohn's Rho in average fishing mortality across alternative age composition simulation (rows) and estimation (columns) scenarios and process variance estimation scenarios (individual boxes). The number above each box indicates the number of the 100 generated data sets that were successfully fit. The catch proportions at age were either simulated (Sim) or estimated (Est) as a multinomial distribution (Multi), Dirichlet-multinomial distribution (DirMulti), logistic normal distribution (Logi), multivariate Tweedie distribution (Twe), or multivariate normal distribution (MultiV) (catch at age rather than proportions). Individual boxplots represent the five estimation models that assume recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Black lines are drawn through zero, red dots indicate the mean, and black dots indicate the median. When no plot is present, no replicates converged for that combination of simulation and estimation scenarios.

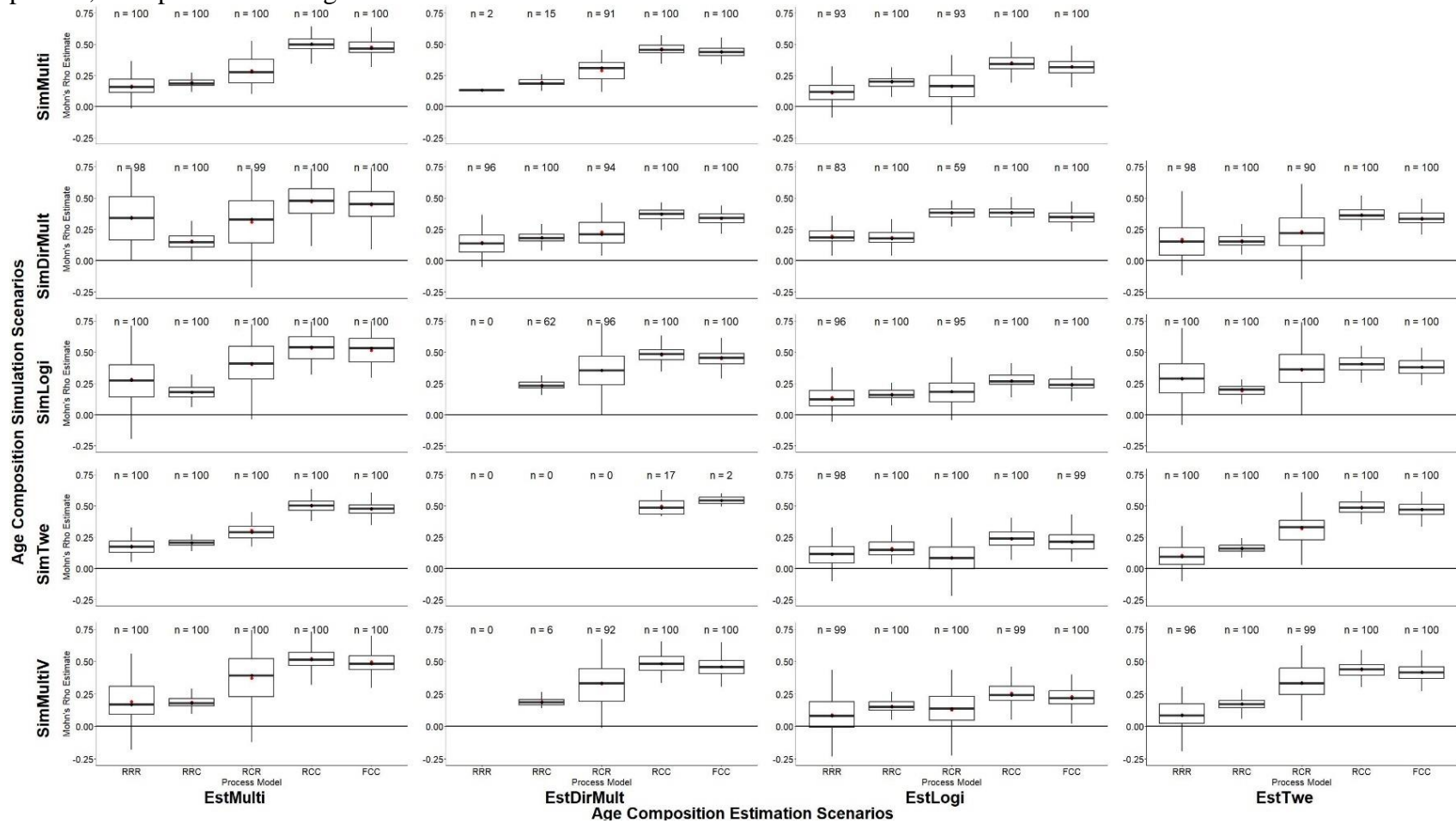


Figure 3.7. Mohn's Rho in recruitment across alternative age composition simulation (rows) and estimation (columns) scenarios and process variance estimation scenarios (individual boxes). The number above each box indicates the number of the 100 generated data sets that were successfully fit. The catch proportions at age were either simulated (Sim) or estimated (Est) as a multinomial distribution (Multi), Dirichlet-multinomial distribution (DirMulti), logistic normal distribution (Logi), multivariate Tweedie distribution (Twe), or multivariate normal distribution (MultiV) (catch at age rather than proportions). Individual boxplots represent the five estimation models that assume recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Black lines are drawn through zero, red dots indicate the mean, and black dots indicate the median. When no plot is present, no replicates converged for that combination of simulation and estimation scenarios.

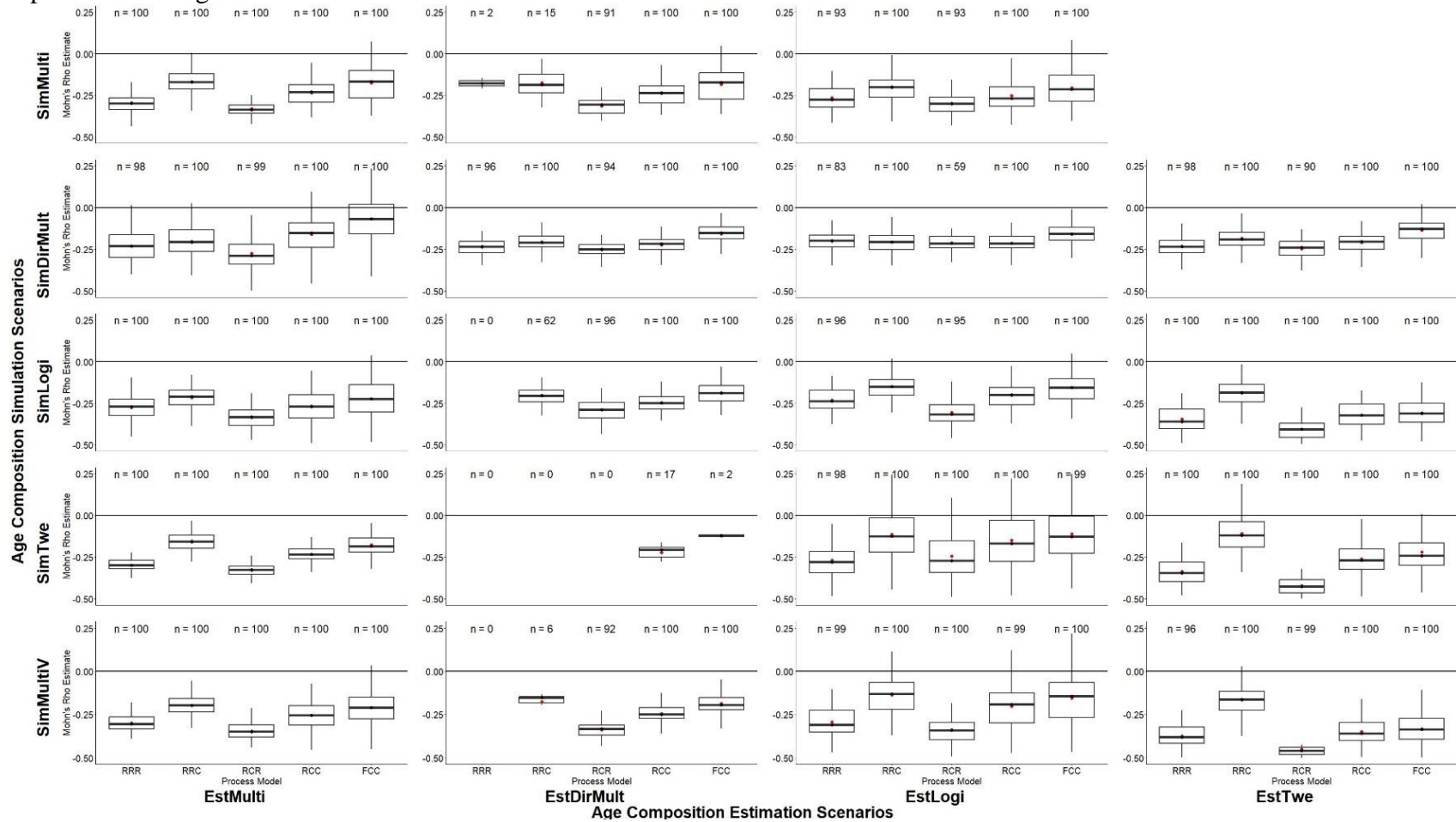


Figure 3.8. Estimated recruitment process error variance across alternative age composition simulation (rows) and estimation (columns) scenarios and process variance estimation scenarios (individual boxes). The number above each box indicates the number of the 100 generated data sets that were successfully fit. The catch proportions at age were either simulated (Sim) or estimated (Est) as a multinomial distribution (Multi), Dirichlet-multinomial distribution (DirMulti), logistic normal distribution (Logi), multivariate Tweedie distribution (Twe), or multivariate normal distribution (MultiV) (catch at age rather than proportions). Individual boxplots represent the five estimation models that assume recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Black lines indicate the original estimate of the variance from the model on which the simulations are based, red dots indicate the mean, and black dots indicate the median. When no plot is present, no replicates converged for that combination of simulation and estimation scenarios.

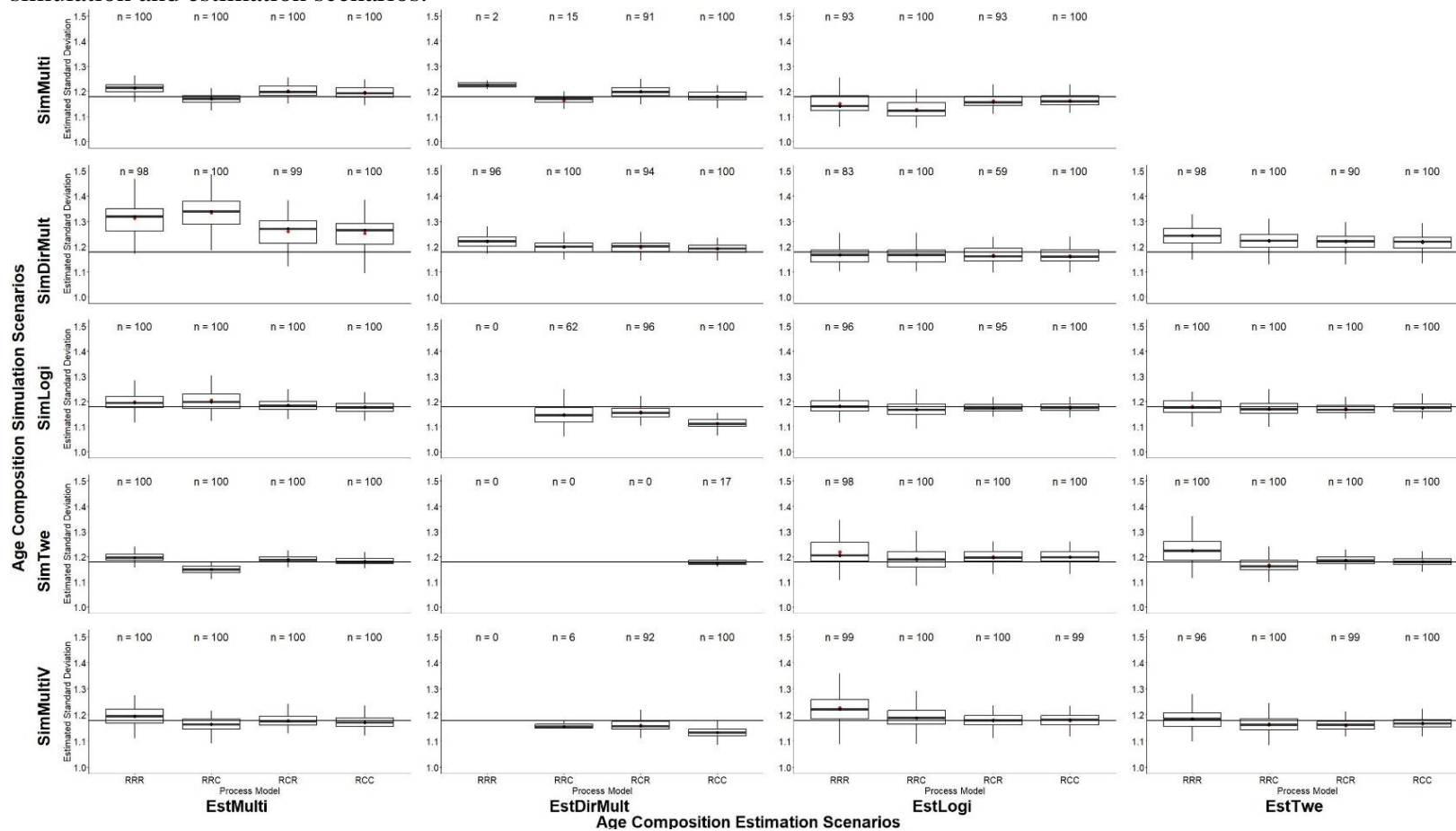




Figure 3.9. Estimated expected survival process error variance across alternative age composition simulation (rows) and estimation (columns) scenarios and process variance estimation scenarios (individual boxes). The number above each box indicates the number of the 100 generated data sets that were successfully fit. The catch proportions at age were either simulated (Sim) or estimated (Est) as a multinomial distribution (Mult), Dirichlet-multinomial distribution (DirMult), logistic normal distribution (Logi), multivariate Tweedie distribution (Twe), or multivariate normal distribution (MultiV) (catch at age rather than proportions). Individual boxplots represent the five estimation models that assume recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Black lines indicate the original estimate of the variance from the model on which the simulations are based, red dots indicate the mean, and black dots indicate the median. Two plots are removed because no models converged for those estimation models.

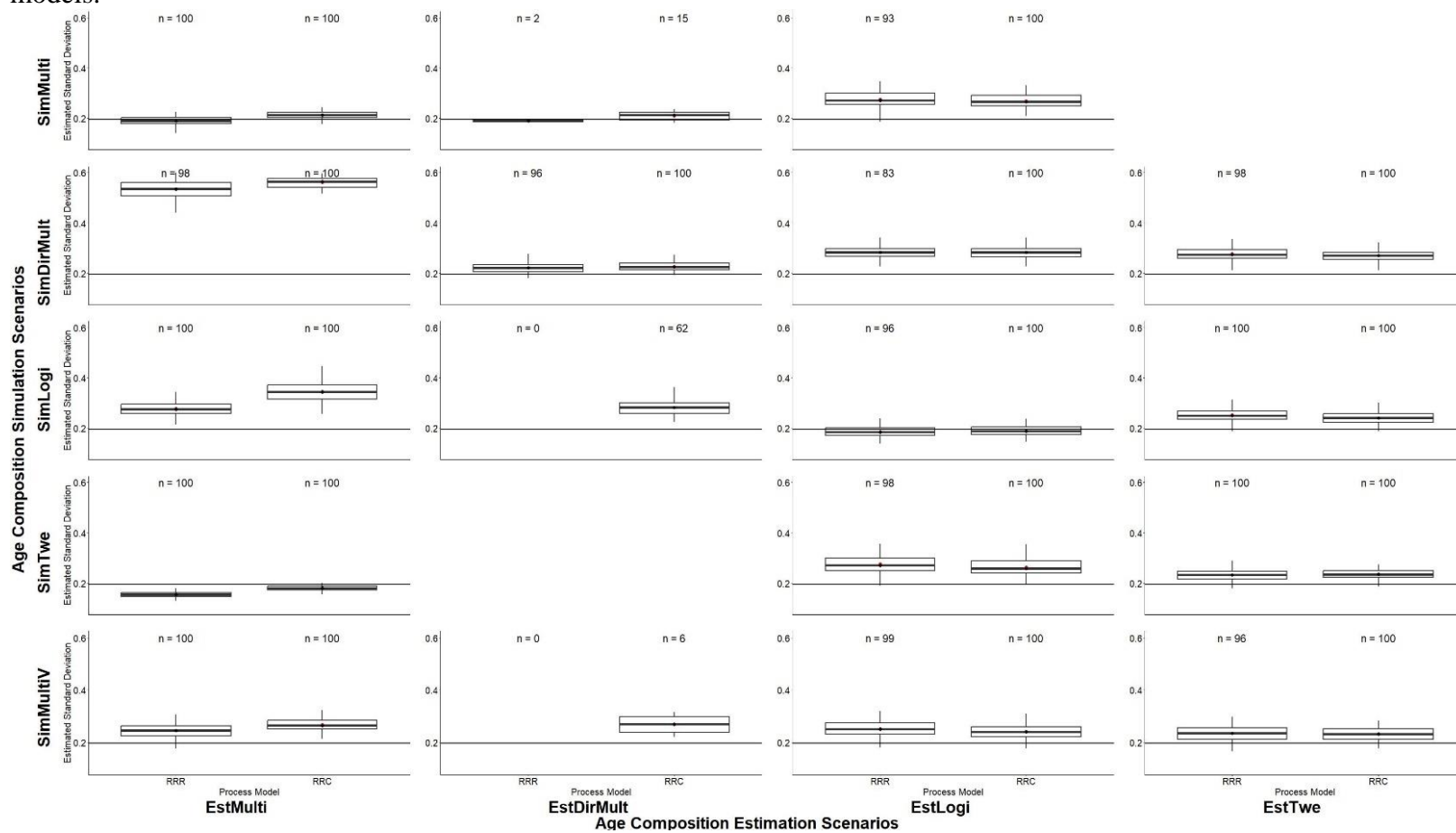


Figure 3.10. Estimated selectivity process error variance across alternative age composition simulation (rows) and estimation (columns) scenarios and process variance estimation scenarios (individual boxes). The number above each box indicates the number of the 100 generated data sets that were successfully fit. The catch proportions at age were either simulated (Sim) or estimated (Est) as a multinomial distribution (Mult), Dirichlet-multinomial distribution (DirMult), logistic normal distribution (Logi), multivariate Tweedie distribution (Twe), or multivariate normal distribution (MultiV) (catch at age rather than proportions). Individual boxplots represent the five estimation models that assume recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Black lines are drawn through the predicted selectivity variance, red dots indicate the mean, and black dots indicate the median. Two plots are removed because no models converged for those estimation models.

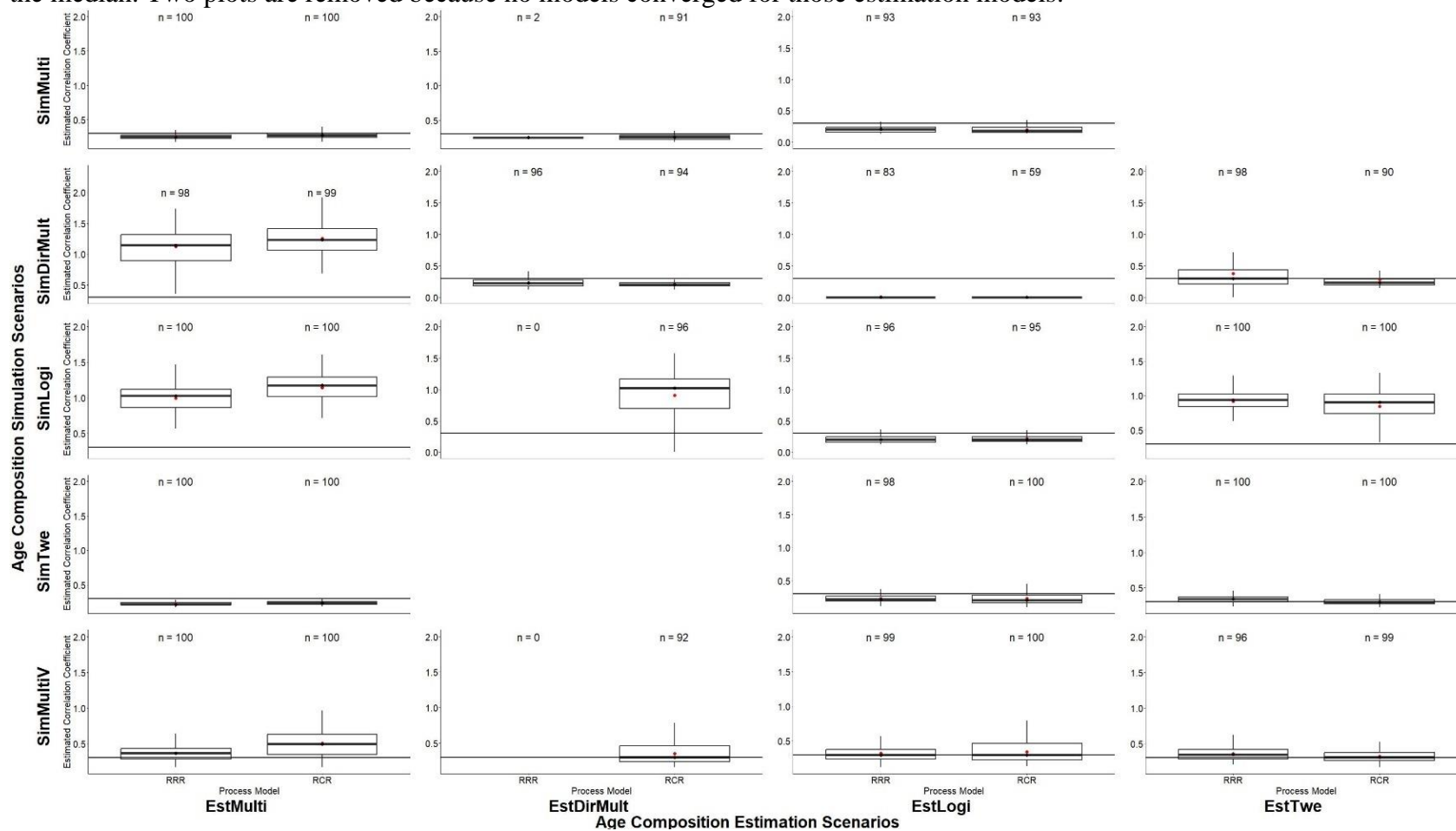


Figure 3.11. Estimated recruitment and expected survival AR(1) correlation coefficient across alternative age composition simulation (rows) and estimation (columns) scenarios and process variance estimation scenarios (individual boxes). The number above each box indicates the number of the 100 generated data sets that were successfully fit. The catch proportions at age were either simulated (Sim) or estimated (Est) as a multinomial distribution (Mult), Dirichlet-multinomial distribution (DirMult), logistic normal distribution (Logi), multivariate Tweedie distribution (Twe), or multivariate normal distribution (MultiV) (catch at age rather than proportions). Individual boxplots represent the five estimation models that assume recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Black lines indicate the original estimate of the correlation coefficient from the model on which the simulations are based, red dots indicate the mean, and black dots indicate the median. Plots are removed if no model converged.

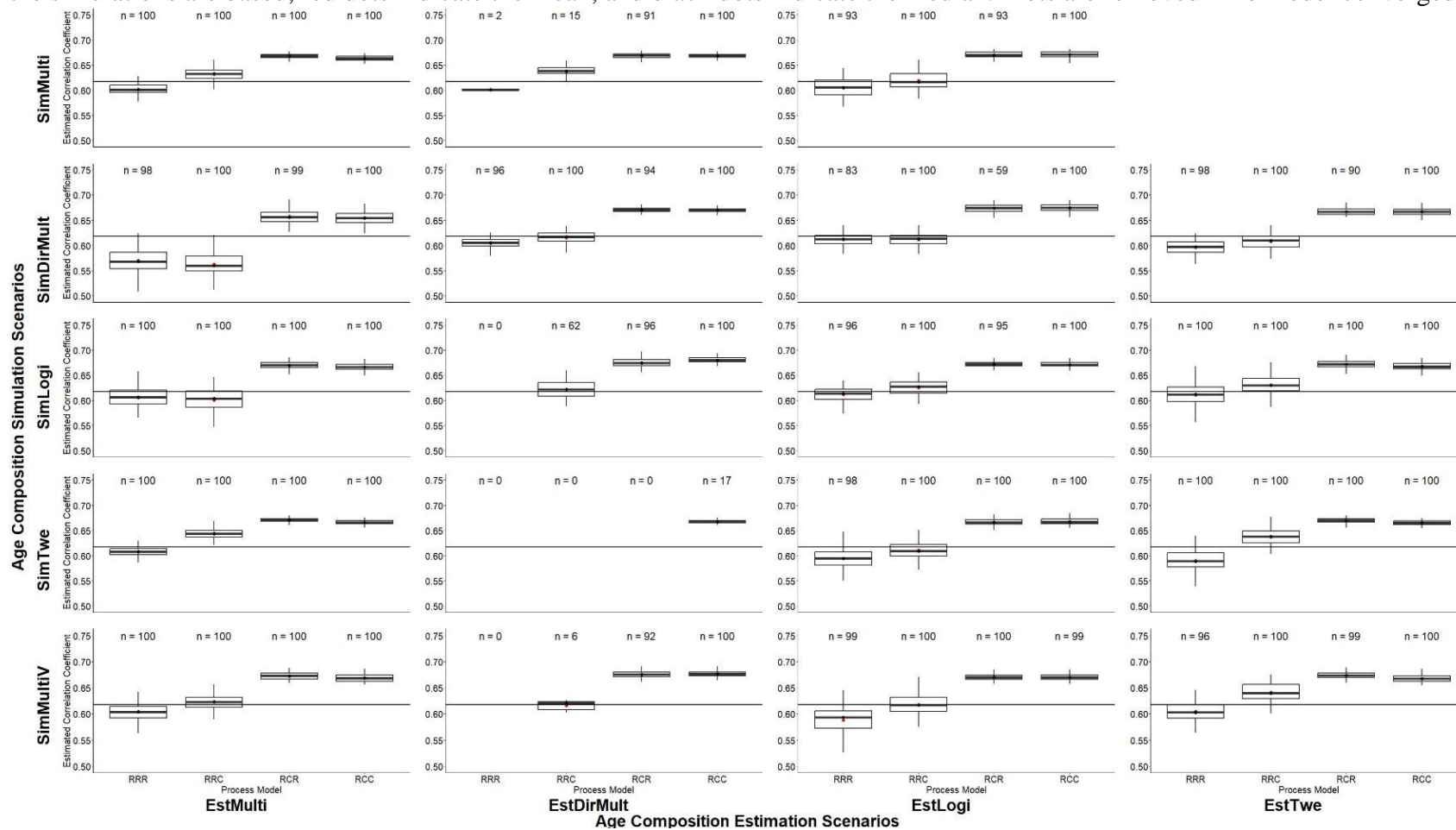


Figure 3.12. Estimated selectivity correlation coefficient across alternative age composition simulation (rows) and estimation (columns) scenarios and process variance estimation scenarios (individual boxes). The number above each box indicates the number of the 100 generated data sets that were successfully fit. The catch proportions at age were either simulated (Sim) or estimated (Est) as a multinomial distribution (Multi), Dirichlet-multinomial distribution (DirMulti), logistic normal distribution (Logi), multivariate Tweedie distribution (Twe), or multivariate normal distribution (MultiV) (catch at age rather than proportions). Individual boxplots represent the five estimation models that assume recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Black lines indicate the original estimate of the correlation coefficient from the model on which the simulations are based, red dots indicate the mean, and black dots indicate the median. Two plots are removed because no models converged for those estimation models.

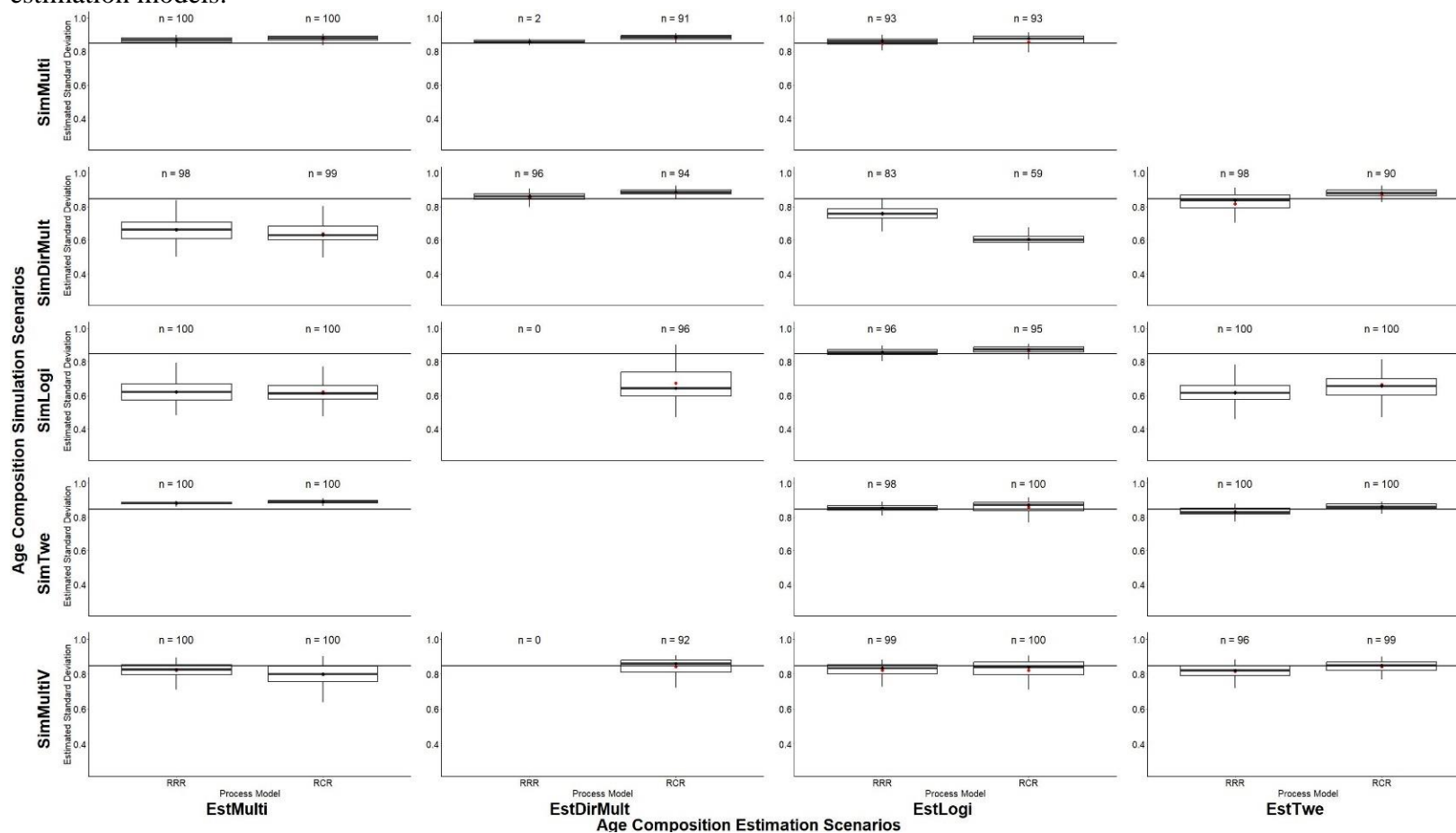


Figure 3.13. Estimated observation error parameters in self-tests that generate and estimate age compositions assuming a (A) Dirichlet-multinomial, (B) logistic normal, and (C) multivariate Tweedie distribution. The number above each box indicates the number of the 100 generated data sets that were successfully fit. Individual boxplots represent the five estimation models that assume recruitment, expected survival, or selectivity as a random effect (R), fixed effect (F), or constant (C). Black lines are drawn through the predicted value, red dots indicate the mean, and black dots indicate the median.

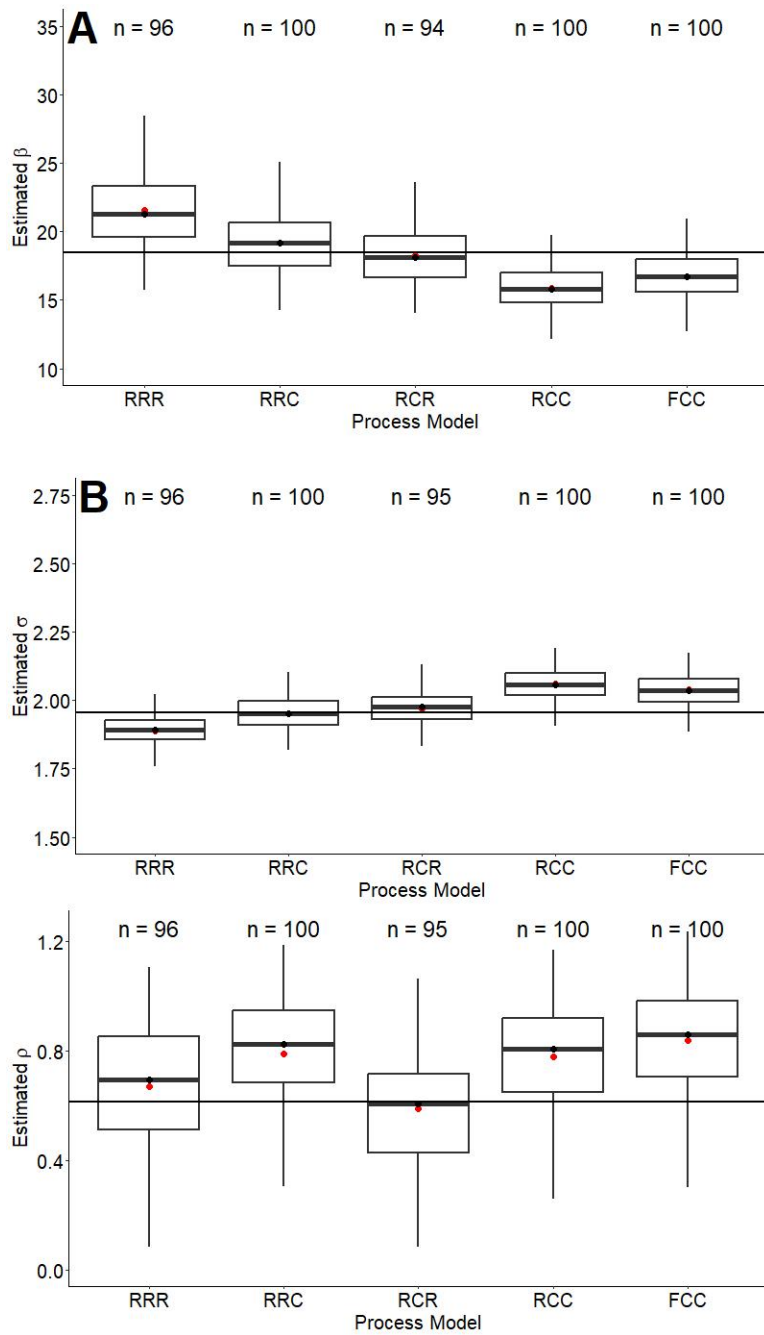
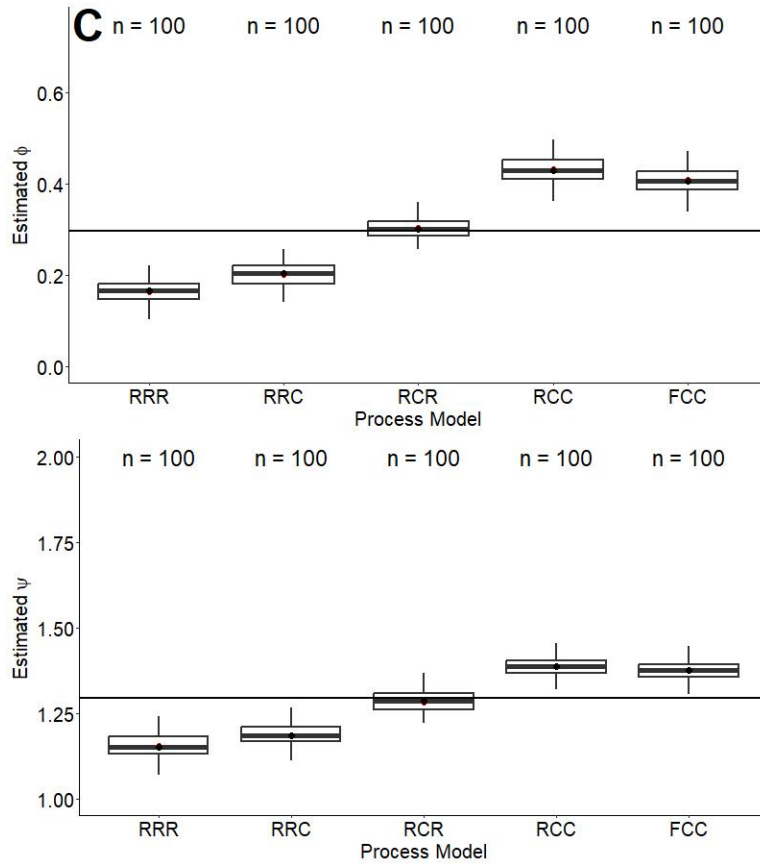


Figure 3.13 (cont'd)



### 3.5 Discussion

The estimation models that specified random effects in recruitment, expected survival, and selectivity, were able to produce relatively unbiased and precise estimates of spawning stock biomass, recruitment, exploitation, and depletion under a wide range of real observation error distributions for catch proportions at age, with the important exception of when the Dirichlet-multinomial was assumed in the estimation model. In such cases, the model rarely converged unless the proportion at age data was actually from a Dirichlet-multinomial. This shortcoming was likely because the Dirichlet-multinomial could not account for inter-age correlation, only overdispersion. Likewise, using the multivariate Tweedie distribution should be avoided if the catch proportions at age contain observation error from a multinomial distribution. However there is substantial evidence that real age composition data is not sampled randomly and has inter-age correlation and overdispersion, so the data generated using the multinomial distribution is unlikely to resemble real conditions (Francis, 2014; Pennington and Vølstad, 1994). The logistic normal and the multivariate Tweedie distribution outperformed other estimation models when the process variation was correctly specified. This aligns with previous work that has recommended the logistic normal distribution or the multivariate Tweedie over the Dirichlet-multinomial in simulated cases when process error was substantial (Fisch *et al.*, 2021) and when applied to real catch data (Albertsen *et al.*, 2017; Francis, 2014; Thorson *et al.*, 2022). If deciding between the logistic normal and multivariate Tweedie, the latter might be preferred for two reasons: its ability to accommodate zeros in the data and its flexibility to predict an effective sample size greater than the input sample size, which can correct for an input sample size set too low (Thorson *et al.*, 2022).

In cases where the RRR model did produce biased outputs, model performance could not be reliably improved by assuming one or more processes was constant or a fixed effect rather than random effect. In some cases, specifying selectivity as constant, rather than time-varying, could increase the convergence rate and reduce bias, however, assuming survival was deterministic consistently led to under- or over- estimation of model outputs. Care should be taken to correctly specify the variability in recruitment and expected survival, regardless of whether the likelihood for the proportions at age includes weighting parameters. It is recommended to include selectivity as a random effect because there is little consequence to trying to estimate it even when it is constant (Cronin-Fine and Punt, 2021; Martell and Stewart, 2014; Stewart and Monnahan, 2017). And if the Dirichlet-multinomial is employed, the number of assumed processes will likely need to be reduced for the model to successfully run. A similar upper limit on the complexity of the process model was noted in Stock *et al.*, (2021) who found that a model that assumed 2D autoregressive AR(1) process in expected survival and natural mortality and a logistic normal distribution for the catch proportions at age failed to converge.

The RRR model had difficulty correctly allocating variability among the three processes across all age composition scenarios and had additional trouble when age composition quality was estimated (i.e., the Dirichlet-multinomial, logistic normal, or multivariate Tweedie). On average, the recruitment and expected survival process error variance tended to be overestimated and the selectivity process error variance was underestimated. When the data generating and estimation model both assume a multinomial distribution, the expected survival variance is usually underestimated under a wide arrange of true expected survival variances (see Chapter 2, Figure 2.6). In the current study, when the age composition likelihood is incorrectly specified, expected survival, more than recruitment and selectivity, may be absorbing the error that results



from model misspecification. The model could be accounting for differences between the expected and observed proportions at age by deviating the numbers at age much more than would be expected by the previous numbers at age and instantaneous mortality alone. This could also explain why models that assume that survival is deterministic will routinely underestimate recruitment, spawning stock biomass, and depletion and overestimate fishing mortality, and why those that assume variability are more likely to be unbiased.

The principal disadvantage to assuming process error in recruitment, expected survival, and selectivity is that if the assumed observation error in the proportions at age is not multinomial, the observation error parameters-  $\beta$  for the Dirichlet-multinomial,  $\sigma$  and  $\rho$  for the logistic normal and  $\phi$  and  $\psi$  for the multivariate Tweedie, could be biased. Even when the observation model is correctly specified, the process model estimated variance in recruitment, expected survival, and selectivity at higher or lower values than the fit to real data. The logistic normal distribution produces estimates of process variance most similar to the fit to real data, though still underestimates the observation variance. Getting accurate estimates of the observation error variance may depend on reducing the amount of assumed variability in the process model. A two-step approach may be appropriate wherein the model is first fit assuming the survival is deterministic, then a second time assuming random effects on expected survival but the observation error parameter is fixed at its estimated value from the first step.

Previous work has generated catch proportions at age data with overdispersion and inter-age correlation, either 1) by simulating from a distribution other than the multinomial as in Cronin-Fine and Punt (2021) or 2) by designing an operating model that accounts for clustering/schooling, aging error and targeted fishing, all factors that can lead to overdispersion and inter-age correlation as in Maunder (2011) or Fisch *et al.* (2021). My approach was the

former which is computationally simpler and allows for explicit self- tests that match the simulation and estimation models. However, I did not allow for different degrees of overdispersion and inter-age correlation to be evaluated and future work may focus on generating catch data with low, medium, or high inter-haul correlation to test what effect these have on the performance of age composition likelihoods. Additionally, it would be prudent to expand this work to several more parameterizations and options of the Dirichlet-multinomial and logistic normal including the “linear parameterization” of the former and other correlation structures of the variance-covariance like a second order autoregressive AR(2) or autoregressive moving average (ARMA) parameterization of the latter, either of which may be more flexible in its ability to adjust observation error variance (Fisch *et al.*, 2021; Thorson *et al.*, 2017). Because state-space stock assessment models with time-varying random effects have been shown to perform differently when the variability is large or small, this work could be further expanded to include cases where, for example, selectivity is not time-varying but the model assumes it is (see chapter 2). Future work should also consider generating new time series of recruitment, expected survival, and selectivity rather than limiting to those predicted from fitting the model to Gulf of Maine haddock. However, this may require use of new information or approaches, given that in Chapter 2 I found that when data were generated from a new set of states this often led to convergence failure for the self-weighting approaches, which is why I used the multinomial distribution for all simulations in Chapter 2.

This study uniquely contends with the question of how to parameterize process and observation stochasticity using a simulation-estimation framework. My recommendations to use a likelihood function other than the multinomial to fit catch proportions at age data is consistent with research that used simulation-estimation methods with fewer varying processes (Cronin-

Fine and Punt, 2021; Fisch *et al.*, 2021; Thorson *et al.*, 2022; Xu *et al.*, 2020) or applied models to real data with several time-varying processes (Albertsen *et al.*, 2017). This work contributes to the ongoing inquiry in how to parameterize state-space stock assessment models in the face of substantial stochasticity and uncertainty.

## REFERENCES

- Albertsen, C.M., Nielsen, A., Thygesen, U.H., 2017. Choosing the observational likelihood in state-space stock assessment models. *Can. J. Fish. Aquat. Sci.* 74, 779-789. <https://doi.org/10.1139/cjfas-2015-0532>.
- Berg, C.W., Nielsen, A., 2016. Accounting for correlated observations in an age-based state-space stock assessment model. *ICES J. Mar. Sci.* 73, 1788-1797. <https://doi.org/10.1093/icesjms/fsw046>.
- Besbeas, P., Morgan, B.J.T., 2017. Variance estimation for integrated population models. *Adv. Stat. Anal.* 101, 439-460. <https://doi.org/10.1007/s10182-017-0304-5>
- Cronin-Fine, L., Punt, A.E., 2021. Modeling time-varying selectivity in size-structured assessment models. *Fish. Res.* 239, 105927. <https://doi.org/10.1016/j.fishres.2021.105927>.
- Dunn PK. 2022. Tweedie: Evaluation of Tweedie Exponential Family Models. R package version 2.3.5
- Fisch, N., Camp, E., Shertzer, K., Ahrens, R., 2021. Assessing likelihoods for fitting composition data within stock assessments, with emphasis on different degrees of process and observation error. *Fish. Res.* 243, 106069. <https://doi.org/10.1016/j.fishres.2021.106069>.
- Fournier, D.A., Archibald, C.P., 1982. A general theory for analyzing catch at age data. *Can. J. Fish. Aquat. Sci.* 39, 1195-1207. <https://doi.org/10.1139/f82-157>.
- Francis, R.I.C.C., 2011. Corrigendum: Data weighting in statistical fisheries stock assessment models. *Can. J. Fish. Aquat. Sci.* 68, 2228-2228. <https://doi.org/10.1139/f2011-165>.
- Francis, R.I.C.C., 2014. Replacing the multinomial in stock assessment models: A first step. *Fish. Res.* 151, 70-84. <https://doi.org/10.1016/j.fishres.2013.12.015>.
- Francis, R.I.C.C., 2017. Revisiting data weighting in fisheries stock assessment models. *Fish. Res.* 151, 5-15. <https://doi.org/10.1016/j.fishres.2016.06.006>.
- Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R.R., Mcgilliard, C.R., Monnahan, C.C., Muradian, M.L., 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. *ICES J. Mar. Sci.* 72, 99-110. <https://doi.org/10.1093/icesjms/fsu198>.
- Legault, C.M., Restrepo, V.R., 1998. A flexible forward age-structured assessment program. *Collect. Vol. Sci. Pap. ICCAT*, 49, 246-253.

- Linton, B.C., Bence, J.R., 2008. Evaluating methods for estimating process and observation error variances in statistical catch-at-age analysis. *Fish. Res.* 94, 26-35. <https://doi.org/10.1016/j.fishres.2008.06.014>.
- Martell, S., Stewart, I., 2014. Towards defining good practices for modeling time-varying selectivity. *Fish. Res.* 158, 84-95. <https://doi.org/10.1016/j.fishres.2013.11.001>.
- Maunder, M. N. 2011. Review and evaluation of likelihood functions for composition data in stock-assessment models: Estimating the effective sample size. *Fisheries Research*, 109: 311–319. Elsevier B.V.
- Maunder, M.N., Punt, A.E., 2013. A review of integrated analysis in fisheries stock assessment. *Fish. Res.* 142, 61-74. <https://doi.org/10.1016/j.fishres.2012.07.025>.
- Maunder, M.N., Piner, K.R., 2017. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. *Fish. Res.* 192, 16-27. <https://doi.org/10.1016/j.fishres.2016.04.022>.
- Method, R.D., Wetzel, C.R., 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fish. Res.* 142, 86-99. <https://doi.org/10.1016/j.fishres.2012.10.012>.
- Mohn, R., 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. *ICES J. Mar. Sci.* 56, 473-488. <https://doi.org/10.1006/jmsc.1999.0481>.
- Nielsen, A., Berg, C.W., 2014. Estimation of time-varying selectivity in stock assessments using state-space models. *Fish. Res.* 158, 96-101. <https://doi.org/10.1016/j.fishres.2014.01.014>.
- Pennington, M., Vølstad, J.H., 1994. Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. *Biometrics* 50, 725-732. <https://doi.org/10.2307/2532786>.
- Perreault, A.M.J., Wheeland, L.J., Joanne Morgan, M., Cadigan, N.G., 2020. A state-space stock assessment model for American plaice on the Grand Bank of Newfoundland. *J. Northwest Atl. Fish. Sci.* 51, 45-104. <https://doi.org/10.2960/J.v51.m727>.
- Punt, A.E., Huang, T., Maunder, M.N., 2013. Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. *ICES J. Mar. Sci.* 70, 16-33. <https://doi.org/10.1093/icesjms/fss185>.
- Stewart, I.J., Monnahan, C.C., 2017. Implications of process error in selectivity for approaches to weighting compositional data in fisheries stock assessments. *Fish. Res.* 192, 126-134. <https://doi.org/10.1016/j.fishres.2016.06.018>.

- Stock, B.C., Miller, T.J., 2021. The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time- and age- varying processes via random effects and links to environmental covariates. *Fish. Res.* 240, 105967. <https://doi.org/10.1016/j.fishres.2021.105967>.
- Stock, B.C., Xu, H., Miller, T.J., Thorson, J.T., Nye, J.A., 2021. Implementing two-dimensional autocorrelation in either survival or natural mortality improves a state-space assessment model for Southern New England-Mid Atlantic yellowtail flounder. *Fish. Res.* 237, 105873. <https://doi.org/10.1016/j.fishres.2021.105873>.
- Thorson, J.T., Kristensen, K., 2016. Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. *Fish. Res.* 175, 66-74. <https://doi.org/10.1016/j.fishres.2015.11.016>.
- Thorson, J.T., Johnson, K.F., Methot, R.D., Taylor, I.G., 2017. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. *Fish. Res.* 192, 84-93. <https://doi.org/10.1016/j.fishres.2016.06.005>.
- Thorson, J.T., Miller, T.J., Stock, B.C., 2022. The multivariate-Tweedie: a self-weighting likelihood for age and length composition data arising from hierarchical sampling designs. *ICES J. Mar. Sci.* fsac159. <https://doi.org/10.1093/icesjms/fsac159>.
- Venables WN, Ripley BD. 2002. *Modern Applied Statistics with S*, Fourth edition. Springer, New York. ISBN 0-387-95457-0, <https://www.stats.ox.ac.uk/pub/MASS4/>.
- Xu, H., Thorson, J.T., Methot, R.D., 2020. Comparing the performance of three data-weighting methods when allowing for time-varying selectivity. *Can. J. Fish. Aquat. Sci.* 77, 247-263. <https://doi.org/10.1139/cjfas-2019-0107>.

APPENDIX 3A: PREDICTED VALUES AND SUPPLEMENTARY DATA

Note that the catch at age is not the product of total catch and proportions because total catch is in weight and catch at age is in numbers of fish.

Table 3A.1. Predicted total catch in weight,  $\hat{C}_y$ , in metric tons (mt) and proportions at age,  $\hat{P}_{a,y}$ , for ages 1-9+.

Year	$\hat{P}_{1,y}$	$\hat{P}_{2,y}$	$\hat{P}_{3,y}$	$\hat{P}_{4,y}$	$\hat{P}_{5,y}$	$\hat{P}_{6,y}$	$\hat{P}_{7,y}$	$\hat{P}_{8,y}$	$\hat{P}_{9+,y}$	$\hat{C}_y$
1977	0.02485	0.58862	0.07487	0.17968	0.08162	0.04801	0.00045	9.00E-05	0.00181	3306.993
1978	0.00279	0.1628	0.61264	0.05695	0.10153	0.04222	0.01996	0.00021	0.00091	5039.263
1979	0.01114	0.03139	0.24924	0.55267	0.04931	0.06854	0.02471	0.0123	0.00072	4667.412
1980	0.03503	0.16492	0.06502	0.23823	0.38412	0.03931	0.04692	0.01731	0.00915	6827.306
1981	0.01573	0.28606	0.25863	0.05399	0.13209	0.19393	0.0206	0.02567	0.01329	6078.218
1982	0.00286	0.13785	0.36099	0.21853	0.04023	0.08418	0.11801	0.01373	0.02362	6887.568
1983	0.00899	0.02676	0.26407	0.28601	0.20623	0.03496	0.06158	0.08651	0.0249	7162.477
1984	0.0028	0.10825	0.06465	0.30634	0.20092	0.18593	0.02612	0.03962	0.06538	4209.778
1985	0.00092	0.04324	0.30163	0.08128	0.23333	0.14017	0.13471	0.01534	0.04938	3309.395
1986	0.00189	0.02002	0.17546	0.36877	0.07214	0.16318	0.09312	0.07656	0.02886	2268.542
1987	0.00156	0.0418	0.08844	0.28995	0.26971	0.06605	0.12712	0.06071	0.05465	960.6168
1988	0.00714	0.02627	0.13866	0.13345	0.28472	0.2206	0.04714	0.0829	0.05912	493.2147
1989	0.00461	0.12793	0.08612	0.19807	0.13259	0.2184	0.13636	0.02739	0.06855	320.0106
1990	0.0076	0.06396	0.37221	0.09601	0.15385	0.08441	0.11544	0.0651	0.04141	481.9544
1991	0.01105	0.1002	0.19423	0.38361	0.07487	0.09713	0.04293	0.05112	0.04486	443.6066
1992	0.031	0.12773	0.28023	0.20016	0.22649	0.041	0.04201	0.01609	0.03529	311.7735
1993	0.03811	0.24328	0.25114	0.19286	0.10479	0.11576	0.01846	0.01602	0.01958	203.4478
1994	0.03622	0.24449	0.37519	0.14535	0.07797	0.05061	0.0499	0.00672	0.01354	186.6776
1995	0.01662	0.26103	0.35847	0.22601	0.05739	0.03433	0.02085	0.01792	0.00737	448.9902
1996	0.00417	0.11105	0.46693	0.26256	0.09567	0.02672	0.01519	0.00818	0.00955	340.4701
1997	0.00818	0.03312	0.24951	0.45465	0.16408	0.05755	0.01588	0.00796	0.00906	1058.044
1998	0.00531	0.06458	0.07915	0.31495	0.35244	0.12178	0.04109	0.01012	0.01057	975.1465
1999	0.02128	0.04706	0.14231	0.12113	0.28738	0.25265	0.08948	0.02566	0.01304	601.3936
2000	0.00305	0.17644	0.10578	0.17526	0.10612	0.21514	0.14825	0.04831	0.02163	1113.83
2001	0.00076	0.02783	0.35275	0.1401	0.15285	0.08534	0.12652	0.07679	0.03706	1206.591
2002	0.0009	0.0069	0.0546	0.47061	0.13761	0.13102	0.05239	0.07688	0.0691	1281.283
2003	0.00012	0.00907	0.01786	0.07707	0.60629	0.11505	0.07319	0.02612	0.07522	1385.141
2004	0.00171	0.00173	0.02746	0.0349	0.10919	0.65622	0.0712	0.04121	0.05637	1426.343
2005	0.00026	0.04256	0.00753	0.05709	0.06636	0.15168	0.54991	0.05463	0.06998	1755.596
2006	0.00114	0.00437	0.1818	0.01276	0.07291	0.07543	0.11914	0.4486	0.08385	1125.342
2007	0.00229	0.02097	0.01711	0.35806	0.01561	0.07277	0.05865	0.08623	0.36831	1885.741
2008	0.00041	0.03435	0.06598	0.02628	0.4774	0.01343	0.05577	0.03854	0.28785	1537.247
2009	0.0006	0.0063	0.0905	0.0909	0.03233	0.49476	0.01071	0.03986	0.23404	1341.602
2010	0.00202	0.01013	0.01803	0.11091	0.10726	0.03464	0.48652	0.00771	0.2228	1294.561
2011	0.01612	0.04239	0.03068	0.02508	0.13736	0.11964	0.03343	0.39921	0.19609	931.1966

Table 3A.1 (cont'd)

2012	0.00745	0.29621	0.09671	0.03835	0.02228	0.11626	0.07804	0.02063	0.32407	1066.45
2013	0.02434	0.07007	0.54144	0.08655	0.02602	0.01334	0.06311	0.03717	0.13796	1274.27
2014	0.04052	0.16908	0.11163	0.53616	0.0448	0.01266	0.00535	0.02362	0.05617	1413.875
2015	0.00178	0.3748	0.23443	0.07934	0.26332	0.01868	0.0043	0.00166	0.02169	1520.039
2016	0.001	0.0129	0.62262	0.1553	0.0422	0.14815	0.00786	0.00168	0.00832	3417.433
2017	0.00098	0.00984	0.02767	0.7187	0.11606	0.02976	0.08788	0.0041	0.005	4919.816
2018	0.00032	0.01227	0.02628	0.0379	0.77052	0.08417	0.01764	0.04601	0.0049	3735.021
2019	0.00019	0.00442	0.03543	0.04084	0.04193	0.77665	0.05632	0.01176	0.03247	4572.009

Table 3A.2. Predicted catch at age,  $\hat{C}_{a,y}$ , in thousands of fish, for ages 1-9+.

Year	$\hat{C}_{1,y}$	$\hat{C}_{2,y}$	$\hat{C}_{3,y}$	$\hat{C}_{4,y}$	$\hat{C}_{5,y}$	$\hat{C}_{6,y}$	$\hat{C}_{7,y}$	$\hat{C}_{8,y}$	$\hat{C}_{9+,y}$
1977	64.36553	1524.642	193.9273	465.4	211.4082	124.3626	1.170178	0.234036	4.68071
1978	9.647603	563.2727	2119.684	197.0473	351.2694	146.074	69.06422	0.719318	3.132671
1979	30.11564	84.88222	674.0665	1494.711	133.3532	185.3607	66.8179	33.27062	1.947584
1980	119.8756	564.4389	222.5368	815.3669	1314.654	134.5246	160.574	59.23828	31.33018
1981	51.02478	927.7554	838.7944	175.1043	428.3966	628.936	66.7942	83.26101	43.11055
1982	10.54797	508.4561	1331.498	806.0282	148.3758	310.5089	435.2775	50.62773	87.10347
1983	31.29291	93.09956	918.8288	995.1624	717.5684	121.6312	214.2858	301.0209	86.62564
1984	5.073111	196.3108	117.2492	555.5566	364.3804	337.2003	47.36384	71.85188	118.5626
1985	1.400902	65.49444	456.8645	123.1039	353.4084	212.3017	204.0388	23.2416	74.80125
1986	2.119151	22.43891	196.6509	413.3133	80.85484	182.8851	104.364	85.81045	32.3497
1987	0.590107	15.80274	33.43479	109.6133	101.9615	24.97019	48.05652	22.95174	20.66113
1988	1.375507	5.062218	26.71892	25.71581	54.8635	42.50886	9.083762	15.97367	11.39138
1989	0.5895	16.37471	11.02297	25.35299	16.97125	27.95547	17.45419	3.505387	8.774151
1990	1.512147	12.71739	74.01403	19.09191	30.5929	16.78417	22.95563	12.94558	8.234291
1991	1.889746	17.14003	33.22292	65.61805	12.80715	16.61451	7.343159	8.743497	7.673595
1992	4.711277	19.41443	42.59333	30.42376	34.42487	6.231598	6.386077	2.445837	5.364227
1993	4.263278	27.21651	28.09623	21.57666	11.7232	12.95011	2.065694	1.79212	2.190663
1994	3.860754	26.06329	39.99627	15.49485	8.311407	5.395068	5.319272	0.716742	1.443894
1995	4.926308	77.36357	106.2421	66.98581	17.01032	10.17556	6.178856	5.311216	2.185125
1996	0.99144	26.38946	110.9612	62.39491	22.73529	6.348823	3.608618	1.943576	2.268914
1997	4.603832	18.64612	140.4697	255.9595	92.37351	32.3997	8.942162	4.482889	5.099105
1998	2.487198	30.23084	37.05621	147.4437	164.9968	57.00897	19.23646	4.738927	4.950635
1999	7.380731	16.32361	49.36437	42.01725	99.68484	87.63504	31.0393	8.900266	4.52376
2000	2.058675	118.9442	71.31067	118.146	71.54197	145.0352	99.94023	32.56987	14.5832
2001	0.535843	19.55248	247.8588	98.4428	107.4029	59.96165	88.89631	53.95432	26.04177
2002	0.684518	5.233093	41.39876	356.8458	104.3426	99.34635	39.72212	58.29197	52.39453
2003	0.105072	7.780729	15.31173	66.07663	519.8392	98.64247	62.75669	22.39426	64.49716
2004	1.487501	1.498638	23.85292	30.30837	94.83503	569.9329	61.83989	35.79431	48.95461
2005	0.289912	46.86504	8.292384	62.86154	73.06885	167.0184	605.5313	60.15509	77.05889



Table 3A.2 (cont'd)

2006	0.859622	3.296312	137.0759	9.620302	54.97207	56.87209	89.82911	338.2413	63.22189
2007	3.031117	27.77332	22.65559	474.1867	20.6707	96.37616	77.6682	114.2025	487.7589
2008	0.438864	37.03156	71.14057	28.32969	514.7272	14.4844	60.12627	41.55791	310.3514
2009	0.52068	5.456948	78.32872	78.67509	27.98236	428.2262	9.271401	34.49954	202.5658
2010	1.594541	8.000589	14.24218	87.63078	84.74826	27.36814	384.4188	6.091257	176.0378
2011	8.851628	23.28364	16.85143	13.7746	75.44232	65.70769	18.35991	219.2538	107.6999
2012	6.12203	243.4982	79.50022	31.52638	18.31593	95.5677	64.15483	16.95681	266.4025
2013	27.57266	79.38815	613.4074	98.05162	29.48127	15.11717	71.4959	42.10628	156.3008
2014	54.39156	226.9524	149.8382	719.6802	60.13748	16.99709	7.186617	31.70809	75.39438
2015	3.161245	663.892	415.255	140.5441	466.4308	33.08125	7.615274	2.934991	38.41228
2016	4.130328	53.50235	2582.956	644.2627	175.06	614.5949	32.59656	6.951739	34.50351
2017	5.118253	51.4821	144.7045	3758.474	606.9205	155.6328	459.5756	21.46536	26.13838
2018	1.190751	46.1663	98.88559	142.5834	2899.036	316.6825	66.37148	173.0962	18.44742
2019	0.868842	20.23003	162.1355	186.9135	191.9033	3554.357	257.7378	53.81914	148.5856

Table 3A.3. Weight at age at time of harvest,  $W_{a,y}$ , in kilograms (kg).

Year	$W_{1,y}$	$W_{2,y}$	$W_{3,y}$	$W_{4,y}$	$W_{5,y}$	$W_{6,y}$	$W_{7,y}$	$W_{8,y}$	$W_{9+,y}$
1977	0.113	0.757	1.163	2.008	2.558	3.358	3.709	3.587	4.686
1978	0.113	0.777	1.234	1.684	2.438	3.108	4.642	4.075	6.088
1979	0.56	0.774	1.155	1.805	2.261	2.659	2.775	3.587	4.725
1980	0.468	0.76	1.168	1.852	2.389	3.354	3.602	4.562	4.204
1981	0.56	0.685	1.515	1.978	2.641	3.027	3.658	4.185	3.921
1982	0.376	0.62	0.995	2.138	2.598	3.107	3.646	4.129	4.293
1983	0.181	0.667	1.2	1.727	2.376	2.969	3.373	3.719	4.215
1984	0.313	0.803	1.23	1.801	2.324	3.166	3.917	4.498	4.071
1985	0.315	0.981	1.068	1.86	2.34	2.652	3.588	4.09	4.153
1986	0.503	0.507	1.193	1.457	2.264	2.494	3.063	3.636	4.589
1987	0.349	0.856	1.592	2.008	2.402	2.609	3.272	4.236	5.277
1988	0.331	0.412	1.116	1.626	2.562	2.583	3.872	4.664	5.18
1989	0.247	1.146	1.818	1.827	2.375	2.863	3.537	4.4	4.244
1990	0.296	0.831	1.541	3.331	2.454	3.043	3.738	3.618	4.189
1991	0.347	1.46	1.882	2.657	3.027	2.958	3.35	4.433	3.881
1992	0.448	1.192	1.764	1.973	2.654	3.067	2.079	3.721	3.45
1993	0.364	0.885	1.593	2.041	2.436	3.035	3.393	3.422	3.657
1994	0.361	0.776	1.589	2.187	3.065	2.788	3.623	3.41	3.73
1995	0.278	0.821	1.16	1.747	2.515	3.512	4.091	5.209	5.665
1996	0.336	0.674	1.074	1.803	2.197	3.149	2.472	2.388	3.163
1997	0.353	0.835	1.766	1.65	2.329	2.975	2.991	3.073	3.608
1998	0.25	0.976	1.447	1.828	2.213	2.843	3.375	3.152	2.989
1999	0.265	0.609	1.311	1.609	1.764	1.926	2.279	3.025	3.29
2000	0.261	0.607	1.005	1.54	1.782	2.009	2.369	2.669	3.081

Table 3A.3 (cont'd)

2001	0.242	0.899	1.271	1.495	1.813	2.213	2.245	2.481	2.528
2002	0.121	0.475	1.022	1.34	1.631	2.141	2.596	2.299	2.638
2003	0.318	0.583	0.89	1.228	1.466	1.766	2.13	2.418	2.507
2004	0.189	0.582	0.782	1.382	1.341	1.659	1.813	2.01	2.204
2005	0.173	0.55	0.752	1.134	1.458	1.436	1.674	1.946	2.273
2006	0.255	0.426	0.841	0.798	1.345	1.649	1.508	1.682	2.035
2007	0.243	0.589	0.799	1.138	1.216	1.512	1.635	1.576	1.708
2008	0.289	0.55	0.986	1.188	1.322	1.302	1.658	1.734	1.746
2009	0.304	0.714	0.781	1.118	1.277	1.608	1.533	1.849	1.906
2010	0.37	0.591	0.776	1.086	1.396	1.533	1.705	2.099	2.014
2011	0.365	0.593	1.03	1.177	1.371	1.57	1.773	1.896	2.096
2012	0.247	0.54	0.903	1.056	1.41	1.444	1.736	1.798	1.962
2013	0.292	0.572	0.842	1.207	1.341	1.566	1.657	1.851	2.088
2014	0.301	0.59	0.759	1.096	1.556	1.656	1.961	1.924	2.179
2015	0.187	0.441	0.799	0.976	1.285	1.672	1.575	1.991	2.218
2016	0.185	0.366	0.603	1.012	1.175	1.372	1.706	2.004	2.003
2017	0.182	0.399	0.598	0.832	1.137	1.29	1.503	1.608	2.627
2018	0.174	0.408	0.564	0.747	0.943	1.298	1.37	1.644	1.808
2019	0.234	0.391	0.575	0.719	0.851	0.984	1.279	1.495	1.786

## CONCLUSIONS AND RECOMMENDATIONS

The overarching intent of this work was to answer one central question – how can state-space stock assessment models (SSSAM) be designed and employed most effectively across a range of modeling contexts now that computing software has eased their use? In what contexts and with what data can this modeling framework be used to minimize bias in crucial management estimates like spawning stock biomass, exploitation, and recruitment, or the process and observation error variances. I sought to test the limits of SSSAM's purported abilities to estimate process and observation variance, reduce retrospective patterns, and allow for more stochasticity to be included in the population model than was previously possible in age based statistical catch at age models.

Chapter 1 demonstrated that existing SSSAM structures can be adapted to accommodate fisheries dependent data such as catch per unit effort in the absence of a fisheries independent index. Several fisheries across the world rely on fisheries dependent data either exclusively or alongside fisheries independent surveys. I recommend that such fisheries at least consider the use of a state-space stock assessment model in their management practices, alongside conventional approaches. This chapter also demonstrated that yearly recruitment, catchability-at-age, and their associated variances could be predicted in a stock assessment model without additional functions (a stock-recruitment curve or a selectivity curve, respectively) to inform them, though the observation error variance would need to be specified. This added flexibility contributed to observed differences between the state-space model and the non-state-space model in output recruitment and spawning stock biomass. There is some evidence that the estimated recruitment in the SSSAM may have been positively biased - the model had negative residuals in the proportions at age for the youngest age classes and there was positive bias in the recruitment

retrospective analysis. I recommend that careful consideration be taken when parameterizing and interpreting recruitment in SSSAMs. Without enough structure or data to inform it, the model may estimate constant recruitment or very little yearly variability. However, with a specific and restrictive structure such as a stock-recruitment curve, which constricts the expected recruitment to be a function of predicted stock size in a previous year, or a random walk, which constricts the recruitment to be similar to recruitment in previous and subsequent years, the model may be biased. I recommend further research, particularly simulation-estimation models or model evaluation by likelihood or AIC, to investigate the performance of SSSAM under several parameterizations of the recruitment.

Chapter 2 demonstrated how well SSSAM can achieve unbiased estimates of recruitment, spawning stock biomass, exploitation, and depletion when the recruitment, abundance (or “expected survival”) and selectivity are highly varying, slightly variable, or constant. By using SSSAM, I was able to include more process variance than was previously possible for Gulf of Maine haddock. I used a simulation-estimation experiment to generate data with known underlying states like abundance and selectivity, such that I could compare the estimated values to the true values and thus quantify bias and precision. I found that a SSSAM that assumes a process is varying when it is constant is preferable to assuming a process is constant when it is varying, with one exception. If the survival is deterministic (i.e., the numbers at age are a function of numbers at age in the previous year and age and instantaneous mortality without an additional random error), the estimation model should be specified as such, or else the model is unlikely to converge. I recommend that process errors be added to natural mortality, rather than survival, if variability is expected in numbers at age due to stochasticity in natural mortality, and that models explicitly account for movement if the variability in numbers at age are due to

immigration or emigration. Using these approaches, the model may better reflect the true population dynamics and the output is likely to be more accurate. This work also revealed that when SSSAMs incorporate error in several different processes and estimate their variances, there is risk of misapportioning the variability across those sources. I would recommend that the expected survival and selectivity process variances especially be interpreted with caution, particularly if there is AR(1) autoregression in the process errors, as it might lead to underestimation of the variance.

Chapter 3 compared the performance of several existing age composition likelihood options in a popular state-space stock assessment software, WHAM, to determine which of these choices may be preferable under the widest array of circumstances, and when the data have overdispersion, inter-age correlation, and zeros. If catch proportion at age data is to be used in SSSAM, alternatives to the multinomial must be employed in order to estimate the observation variance. When the estimation model accounts for several time-varying processes, age composition likelihoods that can account for overdispersion and inter-age correlation outperformed those that could not. I recommend employing a logistic normal or multivariate Tweedie distribution to fit catch proportions at age data in SSSAM, especially when the model accounts for two or more time-varying processes. Though this resulted in the least biased outputs, I caution that the estimated observation error variance may be underestimated and the process error variance biased in such cases. The findings from this work support my previous recommendation to account for multiple sources of process variation in the process model, as such models reduced relative error and retrospective patterns compared to those that assumed fixed effects and constant parameters regardless of how the observation model was specified. Data weighting and alternative likelihoods in SSSAM are ongoing areas of investigation, and I

recommend continued research into the logistic normal and multivariate Tweedie distributions using more simulation-estimation experiments to understand how much overdispersion and/or inter-age correlation can be accounted for by their parameters.

This work encountered several reoccurring issues, none more omnipresent than difficulties surrounding model fitting and convergence. Model failure, more than anything, revealed the limits of SSSAM, such as trying to estimate standard deviation parameters of the total catch for Lake Michigan lake whitefish (Chapter 1), trying to estimate survival process variance when it was absent (Chapter 2), or trying to scale the effective sample size while failing to account for overdispersion and correlation (Chapter 3). When George Box revisited his famous aphorism, he included the clause: “the practical question is how wrong do [models] have to be to not be useful.” There is no more useless model than a failed model, and no stronger sign throughout this work that a model was inappropriate in a given context, either because processes were specified incorrectly, or the model tried to estimate too many processes with not enough data. A continued challenge in quantitative fishery science in general and SSSAM in particular is to find synergy among system complexity, data availability, and model structure.

Another consistent theme throughout this work was questioning how to collect catch and index data. If there are fisheries dependent indices of abundance like catch per unit effort, a survey is not required to fit a SSSAM. And if a survey is used, fewer samples may be required than was previously assumed to achieve the same level of model performance. When the standard deviation of the realized total index values was increased (by 10% or 50%) or the effective sample size of the proportions at age were lowered (also by 10 or 50%) the model did just as well in estimating performance metrics like recruitment and spawning stock biomass or the process error variances, suggesting fewer sampling trips may be taken or fewer individuals

might be aged to achieve the same accuracy and precision. The multivariate Tweedie distribution can increase or decrease the effective sample size relative to the input sample size, suggesting that even if the total number of sampled individuals in the catch is low, the model may be able to right-weight the information internally. In the future, models may be able to do more with less, and resources can be allocated towards collecting as much *informative* data as possible without being redundant.

My intent was to investigate and offer a supplement, rather than a replacement, to the quantitative fisheries modeling toolbox. SSSAMs present an exciting avenue to introduce additional stochasticity and estimate more varying processes but they are not without their limitations. It should never be assumed that one model or class of models will always be right for every fishery or circumstance, and it is wise to consider a suite of model parameterizations and structures, if nothing else to determine how models are sensitive to particular assumptions. I hope that the findings of this research will lend confidence to the continued inclusion of SSSAM in management plans of several fisheries stocks across the globe.