Exploring drivers of capelin (*Mallotus villosus*) and Atlantic cod (*Gadus morhua*) population dynamics using Empirical Dynamic Modelling (EDM)

by

Reid Steele

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirement for the degree of Master of Science in Fisheries Science (Stock Assessment)

Center for Fisheries Ecosystem Research (CFER)

Marine Institute of Memorial University of Newfoundland

September 2023

St. John's, Newfoundland

Abstract

Capelin (*Mallotus villosus*) populations on the Newfoundland shelf collapsed in the early 1990s, coinciding with an ecosystem regime shift and greatly reduced capelin biomass which both persist to this day. The dual-regime nature of this stock's history suggests it may experience nonlinear dynamics, which are difficult to predict using linear models. This thesis explores the application of nonlinear Empirical Dynamic Modelling (EDM) forecasting tools to capelin biomass data, seeking to determine if capelin dynamics are nonlinear, if nonlinear predictive models of capelin population dynamics outperform linear models, what climatic and ecological factors drive nonlinear changes in capelin biomass, and if these driving forces can be measured and compared. In my first chapter, I found capelin dynamics were nonlinear, and EDM predictive models returned equal or improved model diagnostics to linear models in most situations. In my second chapter, I identified a strong positive association between capelin and Atlantic cod dynamics, with both species being driven by long term climatic change and likely to benefit from mild warming. This thesis clearly identifies the utilities of EDM as a tool for use in stock assessment in detecting and forecasting nonlinear stock dynamics, and identifying and characterizing factors driving population dynamics.

General Summary

Capelin is a key species in the Newfoundland shelf ecosystem which provides the primary link in the food web between zooplankton and larger consumers, such as groundfish, seabirds, and marine mammals. Capelin populations in Newfoundland collapsed in 1991, possibly due to an abnormally cold period, and have not yet recovered. The resulting shift in capelin dynamics before and after the collapse makes it difficult to predict this population's trajectory using parametric linear models. This thesis explores the use of nonlinear, nonparametric forecasting methods, known as Empirical Dynamic Modelling (EDM), to predict capelin population trajectories, identify climate and ecosystem factors which drive these trajectories, and assess how these factors have influenced the capelin population and will likely continue to in the future. The latter two objectives are also explored using Atlantic cod as an additional study species to demonstrate the use of EDM for species with different life history strategies.

Acknowledgements

I would like to thank my primary supervisor Dr. Jin Gao for providing the main ideas behind this project, providing feedback, and guiding me through the process resulting in this thesis. I would also like to thank the remainder of my committee, Dr. Jonathan Fisher and Dr. Paul Regular, for their comments on my work and for providing a different perspective to help ground me in the ecological foundations of my work. Lastly, I would like to thank my collaborator Dr. Mariano Koen-Alonso for his insights into capelin and Atlantic cod dynamics and my modelling of them, and for providing much of the data used in this study.

Thanks to the National Science and Engineering Research Council of Canada (NSERC) and to the Fisheries and Oceans Canada Technical Expertise in Stock Assessment (TESA) program for providing the funding for this project.

Co-authorship Statement

I completed the majority of the work presented in this thesis, including literature review, data analysis, and writing and editing of the manuscripts within. My supervisor Dr. Jin Gao assisted with the conceptualization of the studies within this thesis, in addition to providing general supervisory support and manuscript editing. My co-supervisor Dr. Jonathan Fisher and committee member Dr. Paul Regular also assisted in manuscript editing and provided helpful support and ideas about my work during committee meetings. Dr. Mariano Koen-Alonso similarly helped provide research ideas during committee meetings.

I was not involved in any data collection for this study. All of the data used in this study were initially collected by Fisheries and Oceans Canada. The capelin acoustic index dataset was provided by Dr. Mariano Koen-Alonso, the Atlantic cod and Greenland halibut bottom trawl index datasets and ice timing dataset were provided by Dr. Paul Regular, and the Newfoundland Climate Index dataset and its components and capelin and Atlantic cod catch datasets are publicly available online (Cyr and Galbraith, 2021; NAFO, 2021).

Both chapters of this thesis are intended to be submitted as independent peer-reviewed scientific journal articles, with all collaborators mentioned above as co-authors. The overview and summary sections are designed to link the ideas behind the two chapters together into a single thesis.

Table of Contents

Overview 1

Chapter 1: Exploring Capelin Population Dynamics using EDM 7

1.1 Introduction 7

1.2 Methods 9

1.3 Results 14

1.4 Discussion 16

1.5 Tables 24

1.6 Figures 25

Chapter 2: Comparing Capelin and Atlantic Cod Biomass Drivers using Scenario Exploration 36

2.1 Introduction 36

2.2 Methods 38

2.3 Results 40

2.4 Discussion 43

2.5 Tables 46

2.6 Figures 47

Summary and Conclusion 61

References 65

List of Tables

Table 1.2 *Model selection metrics for MVE, ARIMA AR1, and GLS predictive models for both model fits and the forecast experiment.*

Table 2.1 *p-values for CCM significance testing against all covariates for both capelin and Atlantic cod. * indicates significant result ($p < 0.05$).*

List of Figures

Figure 1.1 *Capelin acoustic index, standardized separately for all available years (1982-2019) and for years following the capelin collapse (1991-2019). Open points are missing data filled in using Gaussian Process regression*

Figure 1.2 *Raw standardized covariate time series used in this study.*

Figure 1.3 *Prediction skill (ρ) of univariate (A) simplex projections using embedding dimensions from 1 to 5 and (B) S-Map projections using nonlinear tuning parameter (θ) values ranging from 0.01 to 9 for the standardized capelin acoustic index for both all available years and only years following the capelin collapse.*

Figure 1.4 *Attractor manifold of the standardized capelin acoustic index for all available years, constructed using 1 and 2 year lags.*

Figure 1.5 *Convergent cross-mapping prediction skill (ρ) as a function of library size (time series length) between the normalized capelin acoustic index and ecological and climactic covariates, using all available years of data and E=1. Maximum cross-correlation with 1-year lag and p-values calculated by comparison to cross maps using randomized surrogate time series are included for validation.*

Figure 1.6 *Interaction strengths calculated via the S-Map method between capelin and capelin catch, Greenland halibut biomass, and Atlantic cod biomass for both the all years and post-collapse datasets.*

Figure 1.7 *Interaction strengths calculated via the S-Map method between capelin and capelin catch, Greenland halibut biomass, and Atlantic cod biomass as a function of the cumulative Newfoundland Climate Index (NLCI) and sea surface temperature (SST) for both the all years and post-collapse datasets. Dashed lines are linear regression lines. The all years dataset includes both red and blue points.*

Figure 1.8 *Model fits for MVE, ARIMA AR1, and GLS predictive models for the standardized capelin acoustic index using post-collapse data.*

Figure 1.9 *Forecast experiment for MVE, ARIMA AR1, and GLS predictive models for the standardized capelin acoustic index using post-collapse data.*

Figure 1.10 *Model fits for MVE, ARIMA AR1, and GLS predictive models for the standardized capelin acoustic index using all available years.*

Figure 1.11 *Forecast experiment for MVE, ARIMA AR1, and GLS predictive models for the standardized capelin acoustic index using all available years.*

Figure 2.1 *Boxplots of predicted change in normalized capelin acoustic index per change in covariate (ΔB/ΔCov) calculated from EDM scenario exploration using ±0.5 standard deviation (top, blue) and ±1 standard deviation (bottom, red) perturbation scenarios.*

Figure 2.2 *Boxplots of predicted change in normalized cod fall bottom trawl survey biomass index per change in covariate (ΔB/Δcov) calculated from EDM scenario exploration using ±0.5 standard deviation (top, blue) and ±1 standard deviation (bottom, red) perturbation scenarios.*

Figure 2.3 *Predicted changes in the capelin acoustic index using S-Map scenario exploration with SST perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against normalized SST.*

Figure 2.4 *Predicted changes in the capelin acoustic index using S-Map scenario exploration with the cumulative NLCI perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against the normalized cumulative NLCI.*

Figure 2.5 *Predicted changes in the capelin acoustic index using S-Map scenario exploration with air temperature perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against normalized air temperature.*

Figure 2.6 *Predicted changes in the capelin acoustic index using S-Map scenario exploration with ice timing perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against normalized ice timing.*

Figure 2.7 *Predicted changes in the capelin acoustic index using S-Map scenario exploration with capelin catch perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against normalized capelin catch.*

Figure 2.8 *Predicted changes in the capelin acoustic index using S-Map scenario exploration with the Atlantic cod bottom trawl survey biomass index perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against the normalized cod biomass index.*

Figure 2.9 *Predicted changes in the capelin acoustic index using S-Map scenario exploration with the Greenland halibut bottom trawl survey biomass index perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against the normalized Greenland halibut biomass index.*

Figure 2.10 *Predicted changes in the Atlantic cod bottom trawl survey biomass index using S-Map scenario exploration with SST perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against normalized SST.*

Figure 2.11 *Predicted changes in the Atlantic cod bottom trawl survey biomass index using S-Map scenario exploration with the cumulative NLCI perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against the normalized cumulative NLCI.*

Figure 2.12 *Predicted changes in the Atlantic cod bottom trawl survey biomass index using S-Map scenario exploration with cod catch perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against normalized cod catch.*

Figure 2.13 *Predicted changes in the Atlantic cod bottom trawl survey biomass index using S-Map scenario exploration with the capelin acoustic index perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against the normalized capelin acoustic index.*

Figure 2.14 *Predicted changes in the Atlantic cod bottom trawl survey biomass index using S-Map scenario exploration with the Greenland halibut bottom trawl survey biomass index perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against the normalized Greenland halibut biomass index.*

Glossary of Empirical Dynamic Modelling (EDM) Terms

Attractor Manifold: A representation of the dynamics of a system and its states. May be constructed from multiple related time series, or from a time series and its lagged coordinates.

Convergent Cross Mapping: EDM technique designed to identify causality between variables in a system by using the attractor manifolds of each variable to predict the other (Sugihara et al., 2012).

Dimensionality: Refers to the value of the embedding dimension. A high embedding dimension corresponds to high dimensionality, and vice versa.

Embedding Dimension I: The number of successive time steps used to reconstruct a time series and generate predictions. Measured using Simplex forecasting.

Empirical Dynamic Modelling: Refers to a group of nonlinear, nonparametric forecasting tools predicated on Taken's theorem that the dynamics of a system (e.g. an ecosystem) can be reconstructed from one of its time series using lagged versions of that time series (Takens, 1981). Includes Simplex, S-Map, Convergent Cross Mapping, and Multiview Embedding.

Multiview Embedding: An EDM forecaster which leverages system complexity to generate forecasts by combining many valid variable and attractor manifold combinations into a single model. Described in Ye and Sugihara (2016).

Nonlinearity: The degree to which dynamics are nonlinear. Low nonlinearity corresponds to linear dynamics, whiles high nonlinearity responds to very nonlinear (chaotic) dynamics. Also called state-dependence (e.g. Clark and Luis, 2020). Measured by S-Map θ value – higher θ indicates more nonlinearity.

Nonlinear Dynamics: Dynamics where change in input is not proportional to change in output (Clark and Luis, 2020). May be the result of differing interactions between variables depending on the state of the system (i.e., phase changes), multiplicative interactions between multiple variables, stochastic noise, etc.

Scenario Exploration: A method of assessing nonlinear relationships between a target variable and its covariate(s) by fitting a multivariate EDM model to the target time series, then forecasting on the same time series after artificially increasing and/or decreasing the values of the covariate time series. The scenario exploration performed in this thesis uses methodology from Deyle et al. (2016a).

Simplex Projection: A simple nonlinear EDM forecaster, which uses E lagged coordinates to predict future values. It is also used to determine the best value of E for a time series, and thus quantify the dimensionality of the time series, by comparing prediction skill across simplex forecasts with different E values. Described in Sugihara and May (1990).

S-Map: A nonlinear EDM forecaster, which uses E lagged coordinates and weighting parameter θ to predict future values. θ represents how heavily the model weights nearby points when making projections. When $\theta = 0$, the S-Map is equivalent to a linear AR1 model, and nonlinearity increases with θ. Similar to how simplex projection determines the best value of E, S-Maps are used to determine the best value of θ for a time series, and thus quantify the nonlinearity of the time series, by comparing prediction skill across simplex forecasts with different θ values. Described in Sugihara (1994).

Overview

Capelin (*Mallotus villosus*) is a subarctic forage fish with a circumpolar distribution throughout the Arctic, North Pacific, and North Atlantic oceans. It is considered a keystone species on the Newfoundland Shelf, where it serves as the primary trophic link between plankton and larger consumers, including seabirds, marine mammals, and many commercially valuable groundfish species (Buren et al., 2014). Forage fishes, including sardines, anchovies (Chavez et al., 2003; Schwartzlose et al., 1999), herring (Skagseth et al., 2015; Toresen and Østvedt, 2000), and Barents Sea capelin (ICES, 2021), are typically subject to cyclical boom-bust cycles, characterized by alternating periods of very high and very low biomass separated by population booms and collapses. Uniquely on the Newfoundland Shelf, capelin biomass collapsed in the early 1990s, hypothesized as the result of a period of abnormally cold climate, and despite a period of partial recovery in the mid 2010s, have yet to recover to pre-collapse levels (Bourne et al., 2021; Buren et al., 2019). This collapse is also associated with changes to capelin life history, including earlier maturation (Bourne et al., 2021) and spawning later in the year (Murphy et al., 2021), and a general regime shift in the ecosystem, punctuated in declines in groundfish due to overfishing and the cold period (Gomes et al., 1995; NAFO, 2010), and increases in shrimp (Lilly et al., 2000; Worm and Myers, 2003) and seal (DFO, 2019) populations.

Predictive models are often used in fisheries science and stock assessment to explain past changes in fish stock biomass with the objective of using that knowledge to predict how biomass may change in the future under different conditions. Traditional stock assessment models seek to parameterize population dynamics and predict how populations would be affected by differing levels of fishing pressure. As ecosystems change and models improve, multi-species models and models designed for ecosystem-based fisheries management have become more common,

seeking to improve these predictions by identifying and accounting for the influence of changes in climate and/or other species in the ecosystem. For example, the Barents sea assessment model for capelin accounts for cod and herring populations (Bogstad and Gjøsæter, 2001; Gjøsæter, 2002; Gjøsæter et al., 2015), and the current Newfoundland shelf capelin forecast model includes sea ice metrics (Lewis et al., 2019). However, linear parametric models may not be sufficient to properly predict chaotic dynamics which are often present in marine systems (Clark and Luis, 2020). Nonlinear dynamics – where changes in predictors are not proportional to the changes they cause in their targets – are extremely common in animal populations, including bony fishes (Clark and Luis, 2020). The incidence of nonlinearity is compounded by fisheries, which increase the likelihood of their targets exhibiting nonlinear dynamics (Glaser et al., 2014a). To properly predict the nonlinear dynamics prevalent in fisheries, nonlinear models are required.

Empirical Dynamic Modelling (EDM) is a suite of nonlinear, nonparametric forecasting tools which use lagged versions of a time series (or related time series in multivariate cases) to reconstruct its dynamics and make future predictions. It includes a basic nonlinear simplex forecaster used to select the optimal maximum number of lags (Sugihara and May, 1990), the weighted nonlinear forecaster S-Map used to measure nonlinearity and strength and sign of species interactions (Sugihara, 1994), Convergent Cross Mapping (CCM), which is used to identify causal relationships (Sugihara et al., 2012), and Multiview Embedding (MVE), which creates multivariate predictions by identifying the best possible combinations of time series lags from multiple time series and averaging them. EDM models can also identify, rank, and visualize nonlinear relationships between a target time series and its driving covariates by using EDM scenario exploration, which compares how EDM model predictions differ between simulations with an increase or decrease to the driving covariate.

In the context of stock assessment, EDM has been used successfully on many occasions to forecast fisheries and fish population dynamics, including forage fishes similar to capelin. Linear models used for stock assessment are often bound to strict assumptions and do not account for the existence of multiple states within ecosystems, which may cause the driving forces of population dynamics to change based on the biomass of the target species and their environments. EDM is not bound by these restrictions, which can make it a better tool for predicting the dynamics of complex, dynamic ecosystems. For example, EDM models have been shown to outperform linear models at predicting recruitment on many occasions, when considered both past stock size, past recruitment, and environmental conditions (Deyle et al., 2018; Munch et al., 2018; Ye et al., 2015a). EDM can also be used as part of an ensemble approach with linear models – for example, Sguotti et al. (2020) found that Atlantic cod stock-recruitment relationships were better modelled by linear models in stable populations, and by nonlinear multivariate simplex projections in chaotic populations which exhibited abrupt changes in recruitment and stock size.

EDM can also provide additional utility to managers beyond improving predictions by identifying the environmental, ecological, and anthropogenic conditions that drive species biomass, and predicting how such driving effects may change depending on the states of an ecosystem. CCM has been used to show that Pacific sardine and northeastern anchovy populations off California are both driven by temperature rather than by competition and, as such, temperature state-dependent rules are required for proper management of these stocks (Sugihara et al., 2012). Similarly, CCM has been used to compare the influence of anthropogenic and environmental factors on Bohai Sea and Yellow Sea fish stocks (Zhang et al., 2022), and to discern the effects of life history strategies, environmental conditions, and fishing on North Sea

fish stocks (Wang et al., 2020). EDM has been used to model climate effects on sardines, and disentangle the effects of climate and fisheries on sardines using EDM scenario exploration (Deyle et al., 2013; Giron-Nava et al., 2021). EDM has also been used to identify potential for regime shifts as compared to continuing stable states by quantifying nonlinearity (Dakos et al., 2017; Hsieh et al., 2005). Glaser et al. (2014b) found that nonlinear S-Map spatial catch per unit effort time series predictions outperformed comparable linear forecasts. In addition to modelling, S-Maps can also be used to quantify the magnitude and direction of species interactions, such as interactions between temperature, fish, and zooplankton (Deyle et al., 2016b). The abilities of EDM to identify, assess, and compare causative relationships between variables, model chaotic population dynamics, and work in the context of regime shifts make it an ideal tool for modelling capelin on the Newfoundland shelf.

Predictive models on capelin dynamics by Fisheries and Oceans Canada (DFO) identify timing of sea ice retreat, sea ice area, capelin condition, larval capelin abundance, and the regime shift itself as useful predictors of capelin biomass (Buren et al., 2014; Lewis et al., 2019). Sea ice retreat is hypothesised to indirectly affect capelin biomass via prey availability, but it is not clear if sea ice retreat is simply a correlate of the actual physical forcer affecting the spring bloom and prey availability. Moreover, the environmental factors and/or ecological factors that contributed to the regime shift and changes in capelin condition and both adult and larval capelin abundance persistent since the collapse remain unknown. Identifying drivers of capelin biomass is particularly difficult due to the chaotic nature of its near instantaneous collapse and partial recovery in the mid 2010s, especially using parametric linear models which are not designed to work with such dynamics. The main goal of the first chapter of my thesis is to address these issues using Empirical Dynamic Modelling (EDM). In this thesis I assess the influence of

predator species, capelin catch, or the newly available Newfoundland and Labrador Climate Index (Henceforth referred to as NLCI, Cyr and Galbraith, 2021) and its components on capelin dynamics, which has been shown to be a useful predictor of capelin spawning timing and year class strength (Murphy et al., 2021).

The objectives for the first chapter of my thesis explores the application of a wide range of EDM techniques to capelin biomass data, with the objectives of using EDM to (1) determine if capelin dynamics are nonlinear, (2) identify and measure the influence of ecological (predator biomass and capelin catch) and climatic (the NLCI and it's components, as well as timing of sea ice retreat) drivers of capelin population dynamics, and (3) create a multivariate forecast model for capelin population dynamics using these drivers and compare it to linear forecast models using the same data. To assess the influence of the regime shift, I have compared results using all available years of data (referred to as the all years dataset) to results using only data from after the capelin collapse (referred to as the post-collapse dataset) where possible.

In the second chapter of my thesis, I expand on the first chapter's objective of using EDM to assess relationships between population dynamics and the climatic and ecological covariates which drive them by applying EDM scenario exploration to biomass data for capelin. This chapter complements the first by directly exploring specific relationships between capelin and their drivers, and how EDM uses those relationships to generate biomass predictions. The second chapter also adds Atlantic cod (*Gadus morhua*) as a second study species. Atlantic cod is a larger-bodied species, which as a demersal predator, has different ecological functions and a longer lifespan and generation time than capelin, suggesting it is less likely to exhibit nonlinear dynamics. However, cod also shares a similar biomass and life history evolution trajectory with capelin over the study period (Bourne et al., 2021; DFO, 2012; NAFO, 2010; Olsen et al., 2004).

This makes it an ideal species to test whether EDM analyses of capelin are similarly applicable to other fish species, particularly commercially valuable groundfish. The objectives of my second chapter are to (1) identify drivers of capelin and cod biomass using CCM, (2) use scenario exploration to model the potential magnitude and direction of each driver's effect on the system state of capelin and cod biomass and (3) predict how the magnitude and direction change with the relative value of the driver and/or biomass.

Chapter 1: Exploring Capelin Population Dynamics using EDM

2.14 Introduction

Capelin (*Mallotus villosus*) are a keystone forage species in the subarctic Northwest Atlantic, which form the primary energy linkage between zooplankton and high trophic level predators, especially commercial fishery species (Buren et al., 2019). The capelin populations of the Newfoundland shelf collapsed in the early 1990s during a period of abnormally cold climate and has not recovered since (Bourne et al., 2021). This collapse was accompanied by an overall regime shift in the region (DFO, 2012), punctuated by declines in groundfish (Gomes et al., 1995; NAFO, 2010) and increases in shrimp (Lilly et al., 2000; Worm and Myers, 2003) and seal populations (DFO, 2019). Capelin life history has also changed since this collapse – on average, capelin after the collapse mature a year earlier (Bourne et al., 2021) and spawn 3 weeks later in the year (Murphy et al., 2021). Despite reductions in fishing and a return to the normal climate in following years, capelin continue to exhibit early maturation, delayed spawning, and low productivity (Bourne et al., 2021; Buren et al., 2014; Murphy et al., 2021).

Past research into the drivers of capelin dynamics identify sea ice timing, capelin condition, larval capelin abundance, and the regime shift itself as predictors of capelin biomass (Buren et al., 2014; Lewis et al., 2019). The effects of fisheries landings, predator biomass, and other climactic variables remain unclear. These knowledge gaps are hard to fill using parametric linear models as it is difficult to separate correlation from causation. The current forecast model used for capelin stock assessment 68% of the variation in capelin biomass from 2003-2017 with the previously mentioned drivers (Lewis et al., 2019, Bourne et al., 2021). However, forage fishes often exhibit chaotic, boom-bust dynamics (Buren et al., 2019; Deyle et al., 2013; Giron-Nava et

al., 2021) which are difficult to predict using typical parametric linear models. This is particularly relevant to capelin forecast modelling in Newfoundland due to the population's sudden and unpredicted collapse in 1991 and lack of recovery since. Empirical dynamic modelling (EDM) offers an alternate approach well suited to such problems as this method can capture non-linear dynamics and identify causal relationships (Sugihara, 1994; Sugihara and May, 1990).

Specifically, EDM is a method for reconstructing time series dynamics using lagged versions of the time series. These lags may come from the original time series itself, or in multivariate cases, they may be combined with lags from related time series (Ye and Sugihara, 2016). By comparing the ability of one time series to reconstruct another, EDM can also identify and differentiate between correlation and causation between time series using convergent cross-mapping (Sugihara et al., 2012). In many cases, EDM has been successfully used to model fish population dynamics (Glaser et al., 2014b; Kuriyama et al., 2020; Wasserman et al., 2022), including chaotic systems such as the boom-bust cycles of pacific sardine (Deyle et al., 2013; Giron-Nava et al., 2021) and fisheries recruitment dynamics (Munch et al., 2018). EDM can also be used to identify potential for regime shifts and other ecological phase changes by quantifying nonlinearity (Dakos et al., 2017; Hsieh et al., 2005). The flexibility of EDM as a nonlinear multivariate modelling method, and its ability to identify correlation and causation between time series makes it ideal for modelling the dynamics of a short-lived keystone forage fish like capelin, which exhibits chaotic population dynamics.

The three objectives of this study are to: (1) use EDM to test whether capelin dynamics are nonlinear; (2) identify environmental and ecological drivers of capelin biomass; and (3) combine them to forecast capelin biomass using both EDM and traditional parametric linear models, and

(4) to determine if and how nonlinearity and model prediction ability have changed as a result of the regime shift.

1.2 Methods

Data

Capelin dynamics in NAFO divisions 2J3KL were measured using the capelin stock biomass index from Fisheries and Oceans Canada's spring acoustic survey (henceforth referred to as the capelin acoustic index), the methodology of which is described in detail in Mowbray (2012). This index was chosen due to its superior temporal coverage, standardization, and accuracy for capelin specifically as compared to the depth stratified bottom trawl survey index used for groundfish. Two different datasets were analyzed in this study – one using all available data (1983-2019), and one using only data from after the capelin collapse (1991-2019). This was done because 1) the capelin collapse and regime shift are highly nonlinear events, and separating the datasets is of interest to test for changes in the presence and/or degree of nonlinearity in capelin dynamics, and 2) because there is no recovery to the pre-collapse state after the collapse, it is unclear if including pre-collapse data is helpful, detrimental, or irrelevant for predicting post-collapse population dynamics.. Both datasets contained some missing data years, which were filled using predicted values from a Gaussian process regression fit using maximum likelihood estimation (Figure 1.1). To prevent pre-collapse values from influencing filled years after capelin is known to have collapsed, post-collapse missing data were filled using regression on only post-collapse data for both datasets. Both datasets were then normalized by subtracting each data by the mean biomass and dividing by the standard deviation.

Capelin catch and biomass indices for Atlantic cod (*Gadus morhua*) and Greenland halibut (*Reinhardtius hippoglossoides*) were also included in this study to represent potential top-down pressures on capelin. Capelin catch data for 2J3KL were gathered from the publicly available NAFO STATLANT 21a database (NAFO, 2021). Atlantic cod and Greenland halibut biomass indices were derived from Fisheries and Oceans Canada's Fall random stratified bottom trawl surveys comprising NAFO divisions 2J3KL. Methodological details on these surveys can be found in Doubleday (1981). Biomass indices were calculated using standard stratified analyses described in Smith and Somerton (1981). These datasets were also converted to normalized anomalies for use in EDM analyses.

Climate dynamics were gathered from the Newfoundland Climate Index, and included winter North Atlantic Oscillation (NAO), air temperature, sea ice duration and area, iceberg count below 48°N, sea surface temperature (SST), vertically averaged temperature and salinity at Station 27, cold intermediate layer (CIL) core temperature at Station 27, Newfoundland shelf CIL area and bottom temperature, and the Newfoundland and Labrador Climate Index (NLCI), which is the sign-adjusted sum of all the previous climate metrics. The climate data used in this study, as well as detailed methodology regarding these climate metrics is publicly available and can be found in Cyr and Galbraith (2021). All NLCI climate data used in this study were gathered as normalized anomalies from their original source and were not transformed further. To act as a long-term index of climate phase, the cumulative sum of the NLCI over its history (1951-2019) was calculated and used in this study (henceforth referred to as cumulative NLCI). Lastly, timing of sea ice retreat (henceforth referred to as ice timing) in day of year was provided by Fisheries and Oceans Canada and converted to a normalized anomaly using the years 1969 to 2021. All covariate time series are plotted in Figure 1.2, and all datasets are listed in Table 1.1.

Analyses

EDM is predicated on Takens' theorem that the state space of a dynamic system, such as the capelin acoustic index, can be reconstructed using its lagged coordinates (Takens, 1981). This can be visualized as a 3D attractor manifold by plotting a time series against its 1- and 2-coordinate lags. The first step of EDM analysis is to determine the embedding dimension I – the number of lagged coordinates required to best reconstruct the system's dynamics (Chang et al., 2017). This can be done by comparing prediction skill at different values of E in the form of correlation coefficient ρ for simplex projections, a nonparametric method using $E + 1$ neighboring points in state space to generate time series forecasts (Sugihara and May, 1990). The embedding dimension (E) for both capelin datasets was determined by carrying out univariate simplex projections over the full dataset using leave one out cross-validation with E values over the range of 1-5 and selecting the value of E which maximized predictive skill as measured by the correlation coefficient (ρ). Five was chosen as the maximum value to prevent excessive loss of temporal coverage due to missing lags. Though dimensionality above $E = 5$ is possible, it is uncommon in bony fishes (Clark and Luis, 2020). These selected E values were used in all subsequent analyses.

Once E has been determined, nonlinearity can be assessed similarly using forecasts via S-Map projections. Rather than using only neighboring points, S-Map uses the entire time series with a weighting parameter (θ) controlling the degree to which nearby points are prioritized – larger values of θ indicate greater influence of nearby points on projections and thus greater nonlinearity, with θ=0 indicating a fully linear model (Sugihara, 1994). I tested the nonlinearity of both capelin acoustic index datasets with S-Map projections using leave one out cross validation with a range of values from 0.01-9 for θ.

To identify drivers of capelin population dynamics, the previously discussed climatological and ecological covariates were analyzed with both capelin datasets using Convergent Cross Mapping (CCM). CCM compares points on attractor manifolds of two time series variables to predict each other – if a causative relationship is present, evidence of the manifold of the causative agent should be present in the manifold of the affected variable, which is evidenced by increased prediction skill of the affected variable cross-mapping the causative variable (Sugihara et al., 2012). In a true causative relationship, this increase in prediction skill should increase with time series length (Sugihara et al., 2012). Ecological covariates were cross-mapped in both directions, while climatological covariates were cross-mapped by capelin as capelin cannot drive climate. To assess convergence and validate my CCM results, CCMs were compared to maximum cross-correlation at the same maximum lag (E), and p-values were calculated by comparing CCM results to a distribution of 1000 CCMs (Chang et al., 2017) using phase-randomized surrogate datasets (Ebisuzaki, 1997).

EDM can also be used to measure the direction and strength of species interactions using the S-Map method, which uses S-Maps to calculate partial derivatives in multivariate state space for each point of a time series (Chang et al., 2017). These partial derivatives are a useful proxy for interspecific interactions, and can be used to discern the strength and direction of interspecific interactions, and how these factors change in different ecosystem states (Chang et al., 2017; Deyle et al., 2016b; Ushio et al., 2018; Wasserman et al., 2022). To assess interspecific interactions of capelin, I used the S-Map method to calculate interaction strength with capelin catch, Atlantic cod biomass, and Greenland halibut biomass throughout the time series for both datasets. I also regressed these interactions against the climatic variables which returned

significant CCM results to see if these interactions were mediated by climate (e.g. Deyle et al., 2016b).

Lastly, I used Multiview Embedding (MVE) with the aforementioned climatological and ecological covariates to generate a multivariate forecast model for the capelin acoustic index. Where most models use a single set of covariates, MVE uses many different combinations of lags and covariates to reconstruct the original target time series (in this case, the capelin acoustic index), making it particularly useful in time series that are short (~25 points) or highly chaotic (Ye and Sugihara, 2016), both of which apply to the capelin acoustic index. To compare the predictive ability of MVE to linear models with best performances, I generated an autoregressive integrated moving average (ARIMA) model with autoregressive order 1 autocorrelated errors and a generalized least squares (GLS) model with no correlation structure to compare to the MVE model. To meet the variance assumptions of ARIMA and GLS models, these models were fitted using logged raw data for the capelin acoustic index and other ecological covariates, and the results were exponentiated and converted to the same standardized scale as the MVE model. To balance model fit and computational limitations, all post-collapse models were allowed to use a maximum of 4 parameters, selected from the climatological and ecological covariates and their lags up to E years. This restriction was not imposed on the all years dataset due to its lower value of E, thus requiring less lags, and removing the associated computational limitations. The best combination of parameters for the ARIMA and GLS models were chosen by testing all possible models and selecting the model returning the lowest AICc. To assess the forecast skill of the models, each model was fit to a training set of the first 15 years of fully available data (accounting for lags), then used to predict the capelin acoustic index for the next year. After each

forecast, the actual value for the forecasted year was added to the training set and the subsequent year was forecasted, continuing for the remainder of the dataset.

1.3 Results

Univariate Analyses

E=4 and E=1 were selected as the best embedding dimensions for the post-collapse and all years datasets, respectively (Figure 1.3). Prediction skill declined with increasing embedding dimension in the all years dataset, but was largely unaffected by embedding dimension in the post-collapse dataset with the exception of E=2, which returned much lower prediction skill than all other embedding dimensions tested (Figure 1.3). Both datasets exhibited an increase in prediction skill with θ, indicating nonlinear dynamics were present in the capelin time series (Figure 1.3), though this difference was small for the post-collapse dataset. This increase was greater in the all years dataset, and was especially pronounced at low θ values. Prediction skill peaked at θ=7 in the all years dataset, and continued to increase through the entire range of tested θ values in the post-collapse dataset (Figure 1.3). The capelin acoustic index attractor manifold reveals 3 different phases of capelin biomass magnitude and dynamics by visual inspection – the pre-collapse period from 1985-1992, followed by the collapsed population from 1991-2012, and a partial recovery period distinct from the collapsed population but still well below pre-collapse levels from 2013-2017, after which the population returned to the collapsed state in 2018-2019 (Figure 1.4).

Convergent Cross-Mapping

CCM analyses revealed capelin catch, Greenland halibut biomass, Atlantic cod biomass, the cumulative NLCI, ice timing, and SST to be the strongest correlates of capelin biomass (Figure

1.5). The cumulative NLCI, catch, and cod biomass exhibited very high cross-correlation and achieved very high CCM predictive skill at relatively low library sizes (Figure 1.5), suggesting these covariates were in synchrony with the capelin acoustic index. This means strong forcing was present between these covariates and the capelin acoustic index, preventing CCM from reliably discerning the direction (or bidirectionality) of causation. Conversely, the capelin cross-map of Greenland halibut was stronger than its opposite and continually increased in library size (Figure 1.5). This indicates convergence, meaning a causative relationship of Greenland halibut on the capelin acoustic index can be inferred (Sugihara et al., 2012). Similar patterns indicative of a causative relationship were observed in cross-map patterns with SST and ice timing (Figure 1.5). Winter NAO, air temperature, bottom temperature, Station 27 temperature, Station 27 salinity, and Station 27 CIL core temperature were all weakly coupled with the capelin acoustic index, though only winter NAO appeared convergent and none of these effects approached statistical significance (Figure 1.5). No relationship was detected between the capelin acoustic index and the remaining climatic covariates (Figure 1.5).

Interaction Sign and Strength

No clear temporal patterns were present by visual inspection in species interaction sign and strength except for the post-collapse period for the all years dataset, where they were near 0 (Figure 1.6). During the pre-collapse period, capelin interactions with cod were primarily positive, and interactions with Greenland halibut and catch switched between positive and negative, with the negative interactions being stronger (Figure 1.6). Post-collapse, cod interactions were stronger than Greenland halibut and catch interactions, with variable signs (Figure 1.6). Capelin interactions with Greenland halibut and catch over this period were

primarily positive, though they were also highly variable (Figure 1.6). No significant correlations were detectable between interaction strength and any cross mapped climatic drivers (Figure 1.7).

Predictive Modelling

The MVE predictive model outperformed both linear test models by ρ and R^2 in all modelling tests except for the all years forecast, where it performed similarly to the ARIMA AR1 model and was outperformed by the GLS model (Table 1.2). In all cases, the GLS model outperformed the ARIMA model (Table 1.2). As expected, prediction skill was reduced in the forecast experiment when compared to model fits (Table 1.2). Error metrics were not directly comparable between model fits and forecasts due to differing time series lengths. Models using the post-collapse dataset returned higher ρ and R^2 values than those using the all-years dataset (Table 1.2).

Time series plots reveal that all 3 models struggled to predict phase changes in both the capelin collapse in 1991 and subsequent partial recovery in 2013. In both model fits and the forecast experiment, ARIMA and GLS predicted values returned similar patterns (Figures 1.8-1.11). In model fits, the linear test models predicted the capelin collapse too early (Figure 1.10), and better matched the magnitude of the recovery in 2013, but failed to predict the continued recovery in 2014 (Figure 1.8). In contrast, the MVE model was generally better able to fit year-to-year dynamics but tended to underpredict phase changes (Figure 1.8, Figure 1.10). Similar patterns held in the forecast experiment with the exception of the all years GLS model, which predicted the 2013-2015 phase change period well, but predicted collapsed years somewhat erratically (Figure 1.9, Figure 1.11).

1.4 Discussion

Embedding Dimension and Nonlinearity

We found the best embedding dimension for the post-collapse period was larger than that for the all years period, and that dynamics over both periods were nonlinear, with the all years dataset exhibiting stronger nonlinearity. The presence of nonlinearity and relatively high dimensionality of the post-collapse period was expected, as animals with shorter generation times are more likely to exhibit nonlinear dynamics and higher values of *E* (Clark and Luis, 2020). The difference in predictability between θ at 0 (linear) and its maximum is smaller in the post-collapse dataset than in the all-years dataset, indicating greater nonlinearity in the all-years dataset (Dakos et al., 2017). This agrees with previous literature studying regime shifts using EDM, which find increasing nonlinearity is typical of instability and impending collapse in marine ecosystems (Dakos et al., 2017; Hsieh et al., 2005), and that nonlinearity is reduced after population collapse as the population settles into a more stable post-collapse state (Dakos et al., 2017). Though nonlinearity does decrease, nonlinear dynamics are maintained after the collapse (Figure 1.3) suggesting that small-scale capelin dynamics are still nonlinear, even during "stable" periods. The dynamics around nonlinearity and the regime shift in this stock could theoretically be better explored by independent nonlinearity testing of each state (e.g. pre-collapse, post-collapse before partial recovery), but such experiments cannot be performed reliably with the data used in this study due to lack of temporal coverage, particularly in the pre-collapse years.

Convergent Cross-Mapping

Our CCM results revealed the cumulative NLCI, SST, and ice timing as potential climatic drivers of capelin biomass. I can infer from this result that capelin dynamics are strongly forced by climatic change, and respond to such changes quickly. My CCM results suggest that SST and ice timing are the primary drivers of these climatic changes, which is in line with previous literature that capelin are extremely sensitive to changing temperature and react to such changes

very quickly (Rose, 2005), and that sea ice retreat timing is a useful predictor of capelin dynamics (Buren et al., 2014; Lewis et al., 2019). As capelin cannot drive climate, the synchrony between capelin and the cumulative NLCI suggest capelin is strongly forced by long-term climatic change, with multiple subsequent years of cold or warm climate having large effects on capelin biomass. With the context of CCM alone, it is difficult to tell how exactly changing climate will affect capelin, though it seems likely the effect will be pronounced. I detected no significant influence of sea ice area, icebergs, NAO, CIL metrics on capelin dynamics. CCM final correlation values are more sensitive to noise than the rate of convergence (Mønster et al., 2017), suggesting that weak, insignificant CCM results with converging patterns such as that of Winter NAO in my study may represent a true causative pattern that is obscured by time series noise as climatic effects work their way up the food chain to capelin. Further research using more advanced CCM and/or CCM-adjacent algorithms such as pairwise asymmetric inference, cross map smoothness, multispatial CCM, and continuity scaling may help to clarify these patterns (Clark et al., 2015; Ma et al., 2014; McCracken and Weigel, 2014; Ying et al., 2022).

CCM also revealed capelin catch, Atlantic cod biomass, and Greenland halibut biomass as potential drivers of capelin dynamics. Unlike climate, ecological mechanisms can be logically applied to clarify directionality of potential capelin drivers identified by CCM, complicating interpretation of the synchrony between capelin and both capelin catch and Atlantic cod biomass. Previous research suggests that capelin dynamics are driven by bottom-up environmental forcing, and not top-down forces (Buren et al., 2014). When present, capelin makes up the majority of cod diets, and the presence of capelin has been shown to drive cod condition, growth, distribution, biomass and fecundity on both the Newfoundland shelf, and in the Barents sea (Bogstad and Gjøsæter, 2001; Rose, 2002, 2005). Thus, scientific literature on the relationship

between capelin and cod suggest capelin is far more likely to be driving cod than the reverse. This is also supported by data used in this study, as the collapse of capelin occurred one year before the collapse of cod (Figure 1.2). Similarly, catch often mirrors biomass of the target species and is sometimes used as an abundance index, suggesting it is more likely that capelin catch is driven by capelin biomass than the opposite case, especially considering the positive relationship between capelin catch and capelin biomass in Figure 1.6.

Conversely, the CCM results in this chapter suggest Greenland halibut biomass drives capelin biomass, with no evidence for the opposite case. Unlike cod and capelin, Greenland halibut biomass was relatively unaffected by the collapse and regime shift of the 1990s (Bowering and Lilly, 1992; Dawe et al., 2012; Dwyer et al., 2010). Greenland halibut is known to be an important predator of capelin, but they are also opportunistic and wide-ranging predators and will eat whatever prey species are most available in the ecosystem (Bowering and Lilly, 1992; Dawe et al., 2012; Dwyer et al., 2010), which likely explains why CCM found no evidence of capelin driving Greenland halibut. It is possible that the reduced population of capelin coupled with a similar population of Greenland halibut after the collapse has led to increased relative predation pressure on capelin, producing a top-down effect partially responsible for the non-recovery of capelin, but the effect of Greenland halibut predation on capelin biomass is largely unstudied. It is also possible that this result represents indirect causality stemming from a different original source (Ye et al., 2015b), or that it is spurious, as it is on the edge of significance and CCM is known to sometimes identify causation where there is none for oscillating data patterns (Bartsev et al., 2021) such as that of Greenland halibut (Figure 1.2), and it runs counter to the convention of past literature that capelin dynamics are primarily bottom-up

driven (Buren et al., 2014). Regardless, this result suggests more research into the influence of Greenland halibut on capelin population dynamics is needed.

Interaction Strength

My interaction strength results found no consistent negative interactions between capelin and any of the predators I tested, further supporting the idea that capelin dynamics are primarily bottom-up driven. I also found that capelin species interactions did not change with changing climate, suggesting that climatic effects likely drive capelin directly, or indirectly through bottom-up processes rather than through top-down processes. Surprisingly, most of the interactions between capelin and its predators were positive, particularly over the post-collapse period, which is not consistent with expected patterns from predation (Chang et al., 2017; Deyle et al., 2016b). Potential explanations could be that these patterns are proxies for other dynamics (such as the regime shift), are statistical artifacts due to some combination of noise, autocorrelation, and cross-correlation, or that the presence of these predators assists capelin indirectly through some other mechanism that I did not test. The positive relationship between capelin and it's predators suggests that top-down influences on capelin dynamics are weak, which agrees with previous literature on the Newfoundland Shelf suggesting capelin dynamics are primarily bottom-up driven (Buren et al., 2014). Exploring capelin interactions with potential prey and competitors is a potential avenue for future research using these analyses should sufficient data become available.

Predictive modelling

MVE predictive models explained more variation in capelin dynamics for both fits and forecasts with the exception of the all-years forecast, for which the GLS model provided the best forecasts

(Table 1.2). The all years forecast result is surprising, as the regime shift is a highly nonlinear event, and hypothetically should be more easily modelled by nonlinear models. The failure of the MVE model at all years forecasting in comparison to the GLS model, as well as the MVE model's success during the post-collapse period, suggest that pre-collapse dynamics may be detrimental to making post-collapse predictions in nonlinear state-space. More specifically, it is also possible that the long-term dynamics of capelin are linear, particularly with regards to the cumulative NLCI, and the short term, year-to-year dynamics are nonlinear and may be driven by different factors. Methodological explanations for the all years forecast results include methodological differences in variable selection (the GLS model covariates were selected from the fit test and were unchanged for the forecast test and are thus informed by the full dataset, while MVE model selection is built into the model and must be run independently each year), MVE model averaging causing underprediction of phase changes, and lack of sufficient information in the training set to properly calibrate linear dynamics, as each chaotic event in the all years dataset only occurs once. Regardless, the success of the MVE model in all other cases indicates that it is capable of modelling capelin dynamics, and that MVE and other nonlinear models may be useful forecasting and modelling tools for capelin management. The fact that all 3 models struggled to predict the phase shifts in 1991 and 2013 suggests either that the covariate or combination of covariates responsible for these phase shifts was not present in my model, or more phase shifts are required in the data for the models to identify patterns responsible for causing them. Potential covariates that could improve my models include indices of phytoplankton and zooplankton, or more detailed information on the capelin population, such as larval abundance and capelin condition (Buren et al., 2014; Lewis et al., 2019).

Limitations of EDM

Though my results show EDM is a useful tool for modelling fisheries population dynamics, it is bound by some notable limitations. The first of these is its data requirements – by default, EDM requires a time series at least 35-40 points long (Ye et al., 2015a), with no holes. In this study, I was unable to retrieve sensible results from CCM using the post-collapse dataset due to this limitation. These issues can be somewhat mitigated by using MVE, which recommends a lower minimum of ~25 points due to its ability to synthesize dynamics from multiple sources (Ye and Sugihara, 2016), by using modified forms of EDM designed to work with missing data (e.g. Johnson and Munch, 2022), or by filling missing data, as I did in this study. However, in the stock assessment context, a consistent time series of at least 35-40 years is required for abundance indices, which limits the number of stocks that can be EDM can be used on and the covariates they can be modelled and/or cross mapped with, and precludes EDM from being useful on data poor stocks. Even if the time series is long enough, my results also show that it is difficult for EDM to predict chaotic dynamics associated with phase changes, such as the capelin collapse and partial recovery in 1991 and 2013 respectively. Standard EDM is also not designed to incorporate uncertainties in input data, which can call EDM results into question when working on point estimates with a high degree of uncertainty. However, these limitations are not unique to EDM, and it is possible they can be overcome as time series lengthen, population estimates improve, and more chaotic events occur, allowing EDM to better identify patterns in other covariates associated with them. Similarly, using more advanced EDM tools designed to work on survey components rather than point estimates, and/or to work on shorter time series such as the CCM algorithms mentioned previously could help to resolve these issues in the future.

Another limitation of EDM is that its nonparametric nature makes it difficult or even impossible to understand the exact nature of how covariates affect each other, and to perform simulation studies to observe the influence of changing covariates, such as harvest rate. Recent research has addressed the latter issue through the development of fisheries management strategies using Empirical Dynamic Programming. which combines EDM models with a temporal difference learning algorithm to simulate different fisheries catch scenarios and enable the use of EDM in risk-based management (Brias and Munch, 2021; Giron-Nava et al., 2021). It is also difficult to calculate uncertainties for MVE models because they average a multitude of simpler models, though this is not an issue for other EDM models. As a result, EDM's most useful application to fisheries management is as part of an ensemble approach, which can improve predictive modelling and identify covariates driving population dynamics, while more standard stock assessment models or extensions of EDM such as Empirical Dynamic Programming are used for such simulation studies, setting reference points, and quantifying uncertainties.

1.5 Tables

Table 1.1 List of datasets used in this chapter, their classifications, and their sources.

Variable	Class	Source
Capelin Acoustic Index	Target	Received from Fisheries and Oceans Canada
Capelin Catch	Ecological	NAFO STATLANT 21a Database
Atlantic Cod Bottom Trawl Index		Received from Fisheries and Oceans Canada
Greenland Halibut Bottom Trawl Index		Received from Fisheries and Oceans Canada
Timing of Sea Ice Retreat	Climatic	Received from Fisheries and Oceans Canada
Cumulative Sum of the Newfoundland and Labrador Climate Index (NLCI)		Cyr and Galbraith (2021)
Winter North Atlantic Oscillation (NAO)		Cyr and Galbraith (2021)
Air Temperature		Cyr and Galbraith (2021)
Sea Ice Duration and Area		Cyr and Galbraith (2021)
Iceberg Count Below 48°N		Cyr and Galbraith (2021)
Sea Surface Temperature (SST)		Cyr and Galbraith (2021)
Cold Intermediate Layer (CIL) Area		Cyr and Galbraith (2021)
Bottom Temperature		Cyr and Galbraith (2021)
Station 27 Temperature		Cyr and Galbraith (2021)
Station 27 Salinity		Cyr and Galbraith (2021)
Station 27 CIL Core Temperature		Cyr and Galbraith (2021)

Table 1.2 Model selection metrics for MVE, ARIMA AR1, and GLS predictive models for both model fits and the forecast experiment

Dataset	Type	Model	ρ	R^2	MAE	MSE	RMSE
Post-Collapse	Fit	Multiview	0.916	0.84	0.409	0.484	0.695
Post-Collapse	Fit	ARIMA	0.662	0.438	0.446	0.656	0.81
Post-Collapse	Fit	GLS	0.814	0.662	0.324	0.404	0.636
Post-Collapse	Forecast	Multiview	0.82	0.673	0.858	1.786	1.336
Post-Collapse	Forecast	ARIMA	0.281	0.079	1.082	2.656	1.63
Post-Collapse	Forecast	GLS	0.763	0.583	0.917	1.759	1.326
All Years	Fit	Multiview	0.89	0.791	0.259	0.232	0.482
All Years	Fit	ARIMA	0.813	0.661	0.244	0.363	0.602
All Years	Fit	GLS	0.845	0.715	0.196	0.307	0.554
All Years	Forecast	Multiview	0.505	0.255	0.105	0.021	0.145
All Years	Forecast	ARIMA	0.51	0.26	0.114	0.021	0.146
All Years	Forecast	GLS	0.791	0.626	0.084	0.011	0.104

1.6 Figures

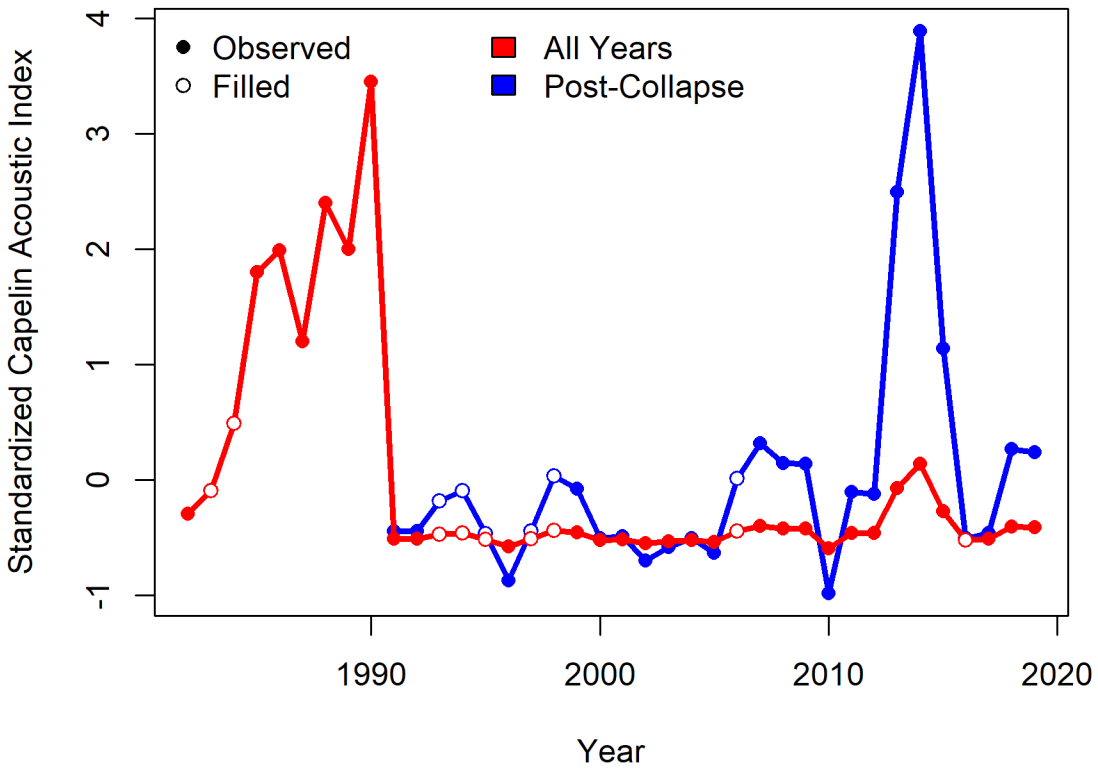

Figure 1.1 Capelin acoustic index, standardized separately for all available years (1982-2019) and for years following the capelin collapse (1991-2019). Open points are missing data filled in using Gaussian Process regression

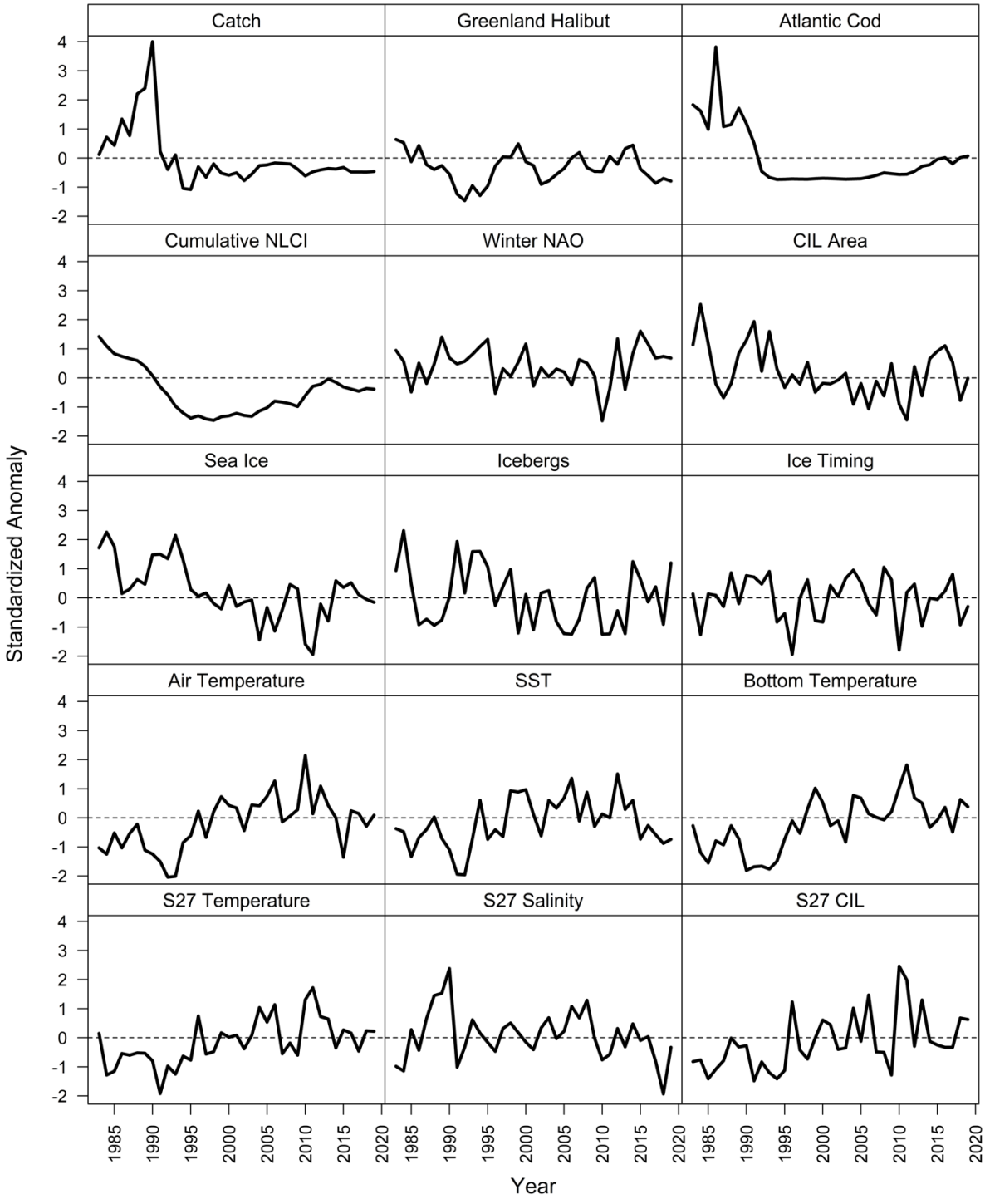

Figure 1.2 Raw standardized covariate time series used in this study.

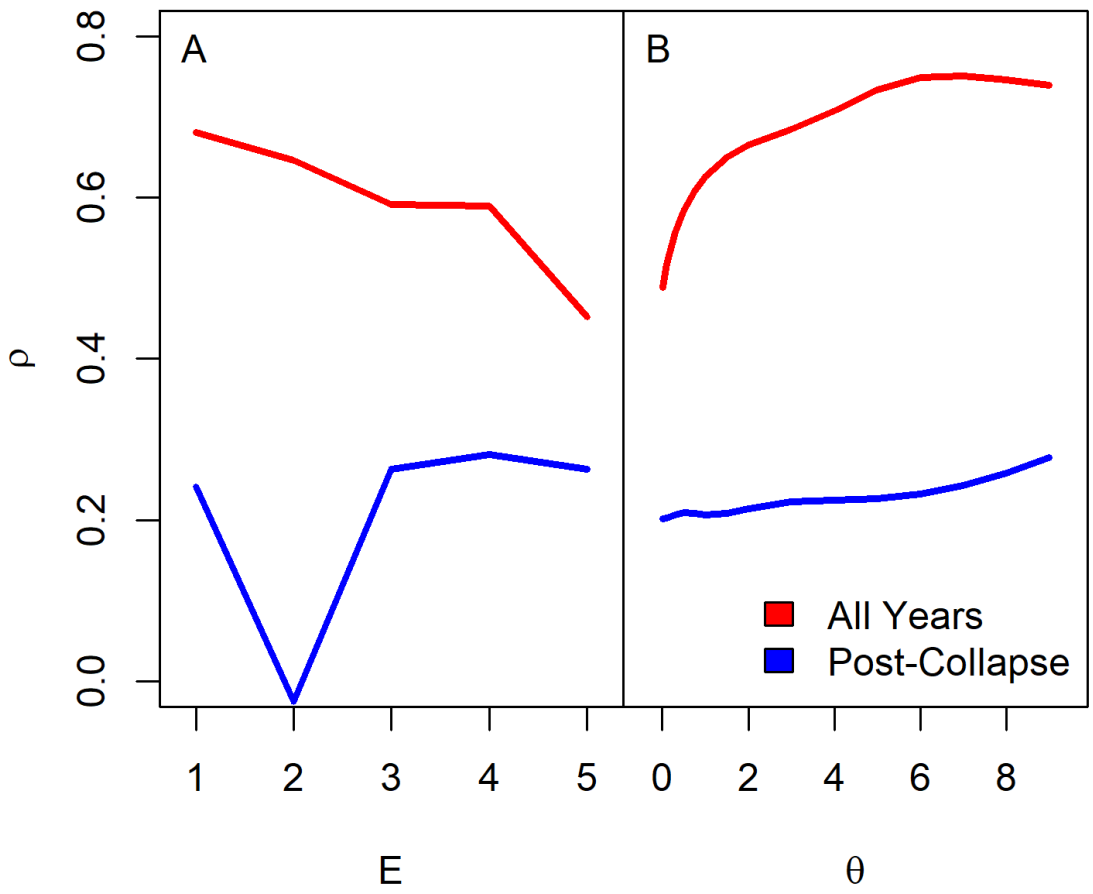

Figure 1.3 Prediction skill (ρ) of univariate (A) simplex projections using embedding dimensions from 1 to 5 and (B) S-Map projections using nonlinear tuning parameter (θ) values ranging from 0.01 to 9 for the standardized capelin acoustic index for both all available years and only years following the capelin collapse.

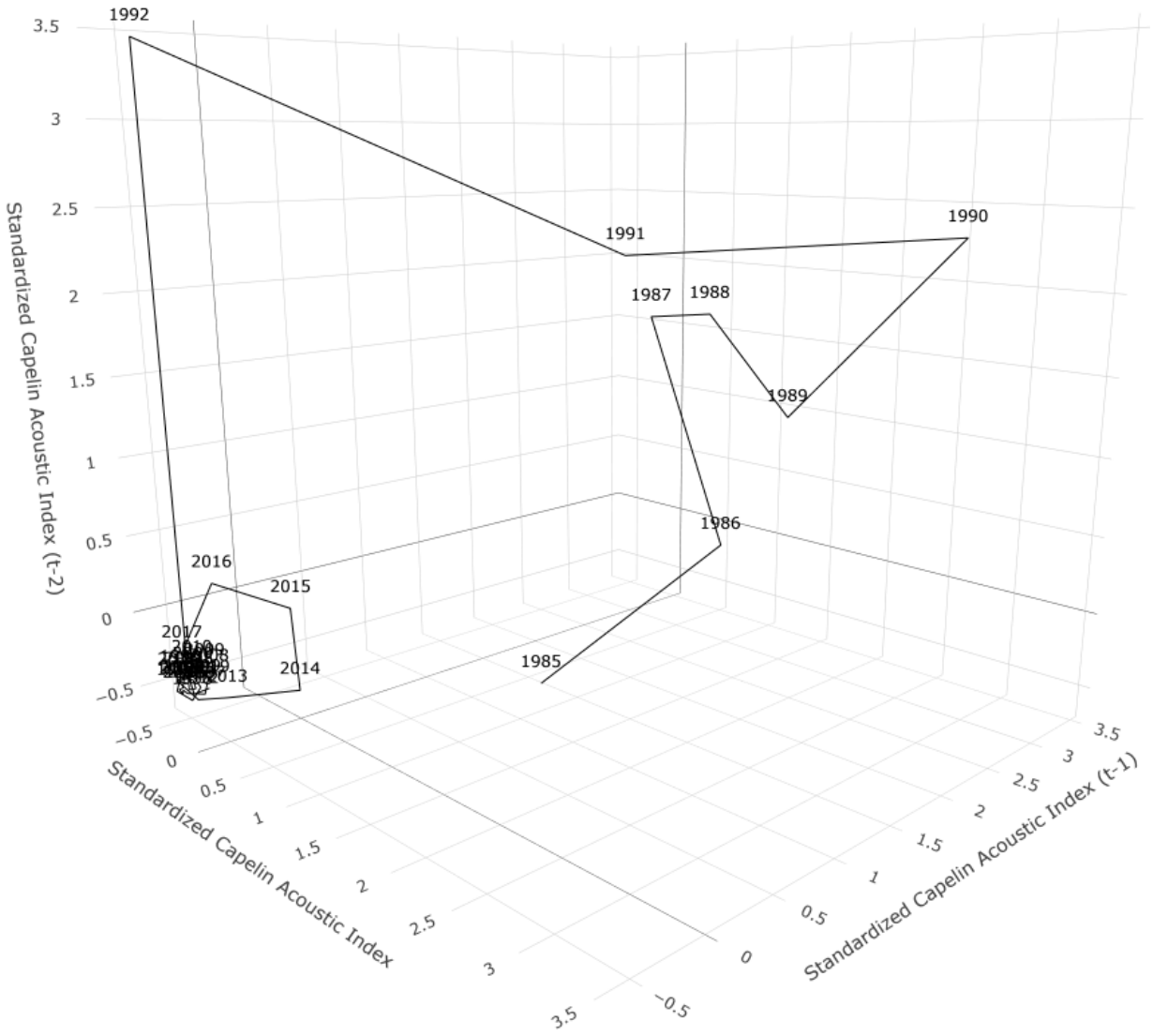

Figure 1.4 Attractor manifold of the standardized capelin acoustic index for all available years, constructed using 1 and 2 year lags.

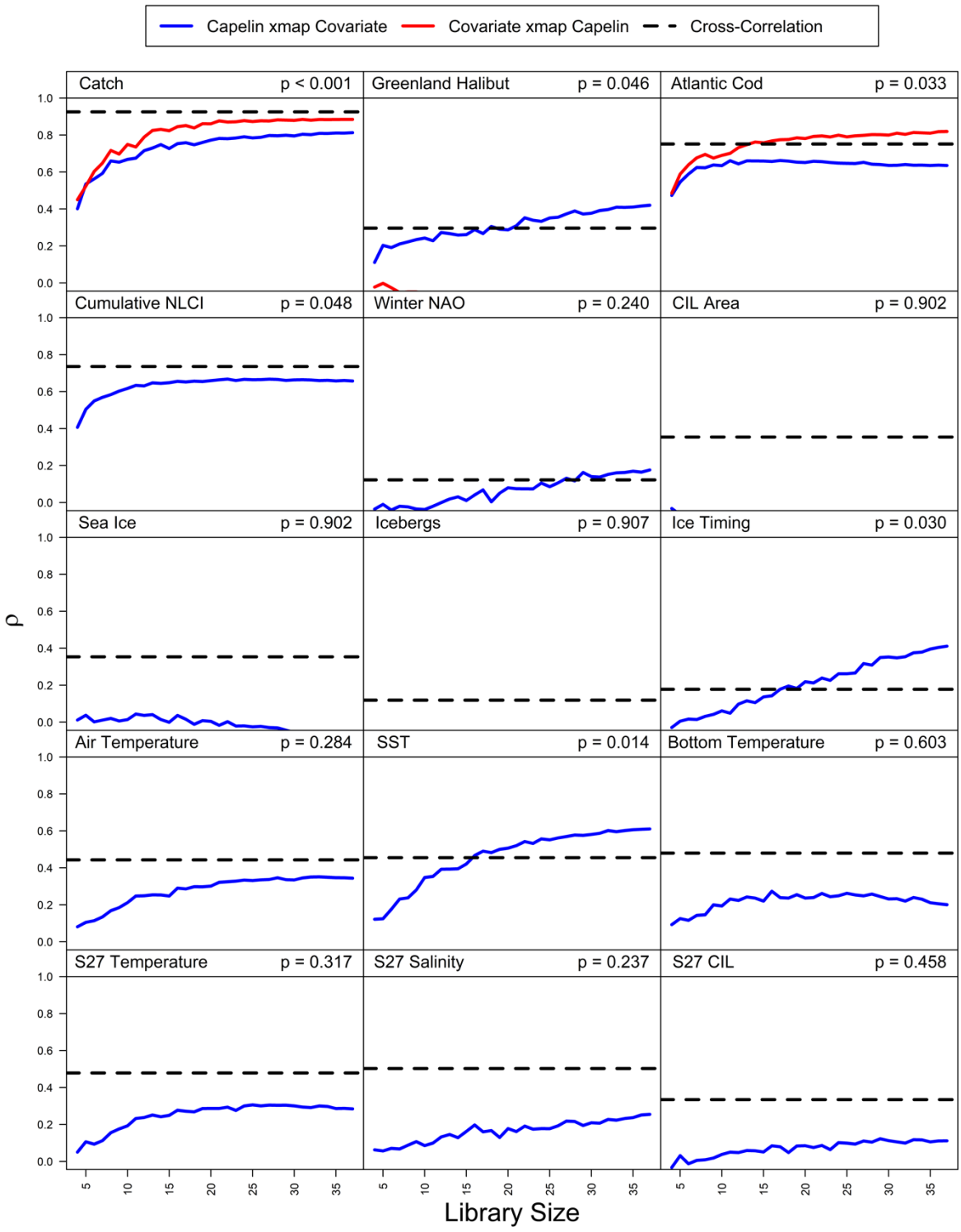

Figure 1.5 Convergent cross-mapping prediction skill (ρ) as a function of library size (time series length) between the normalized capelin acoustic index and ecological and climactic covariates, using all available years of data and E=1. Maximum cross-correlation with 1-year lag and p-values calculated by comparison to cross maps using randomized surrogate time series are included for validation.

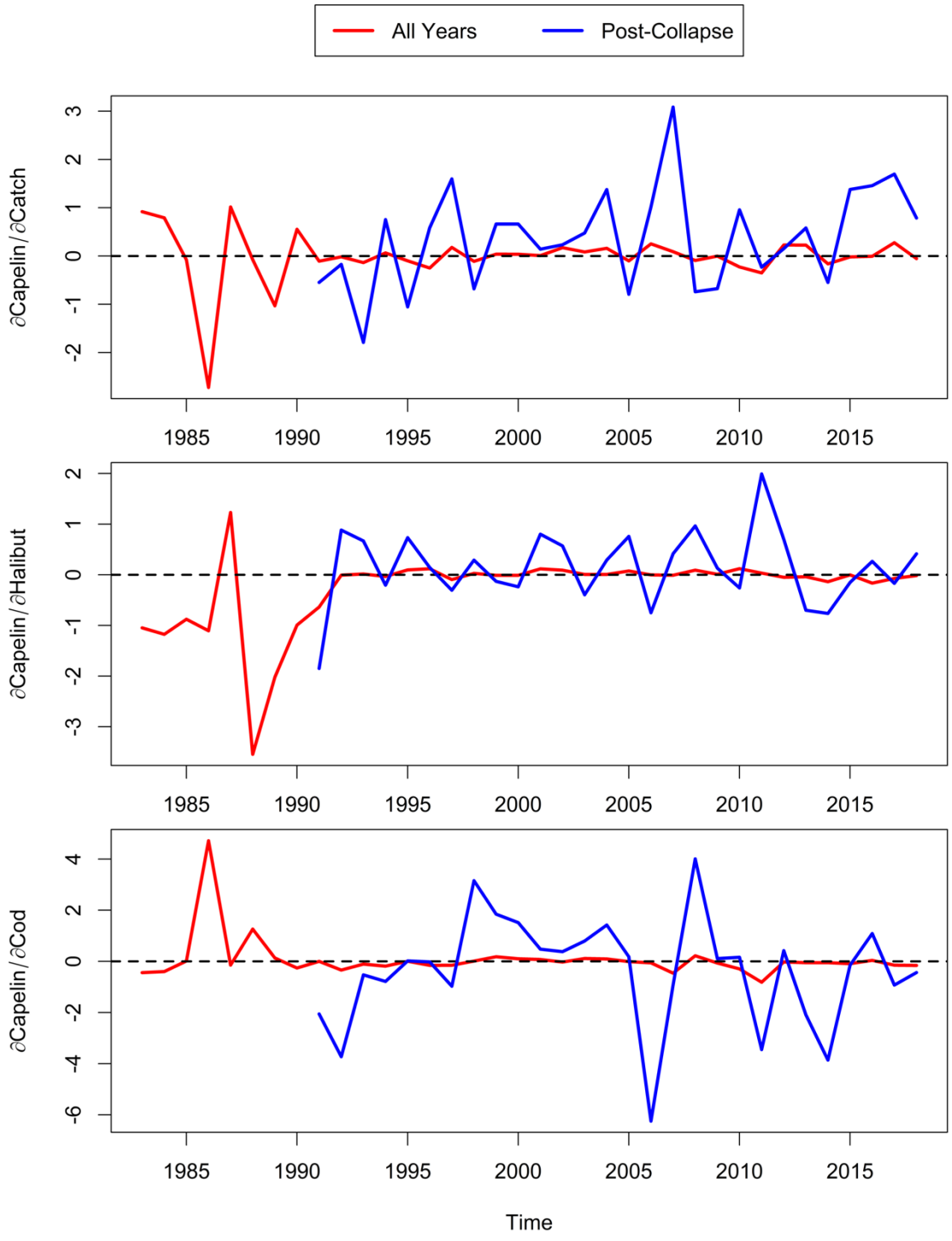

Figure 1.6 Interaction strengths calculated via the S-Map method between capelin and capelin catch, Greenland halibut biomass, and Atlantic cod biomass for both the all years and post-collapse datasets.

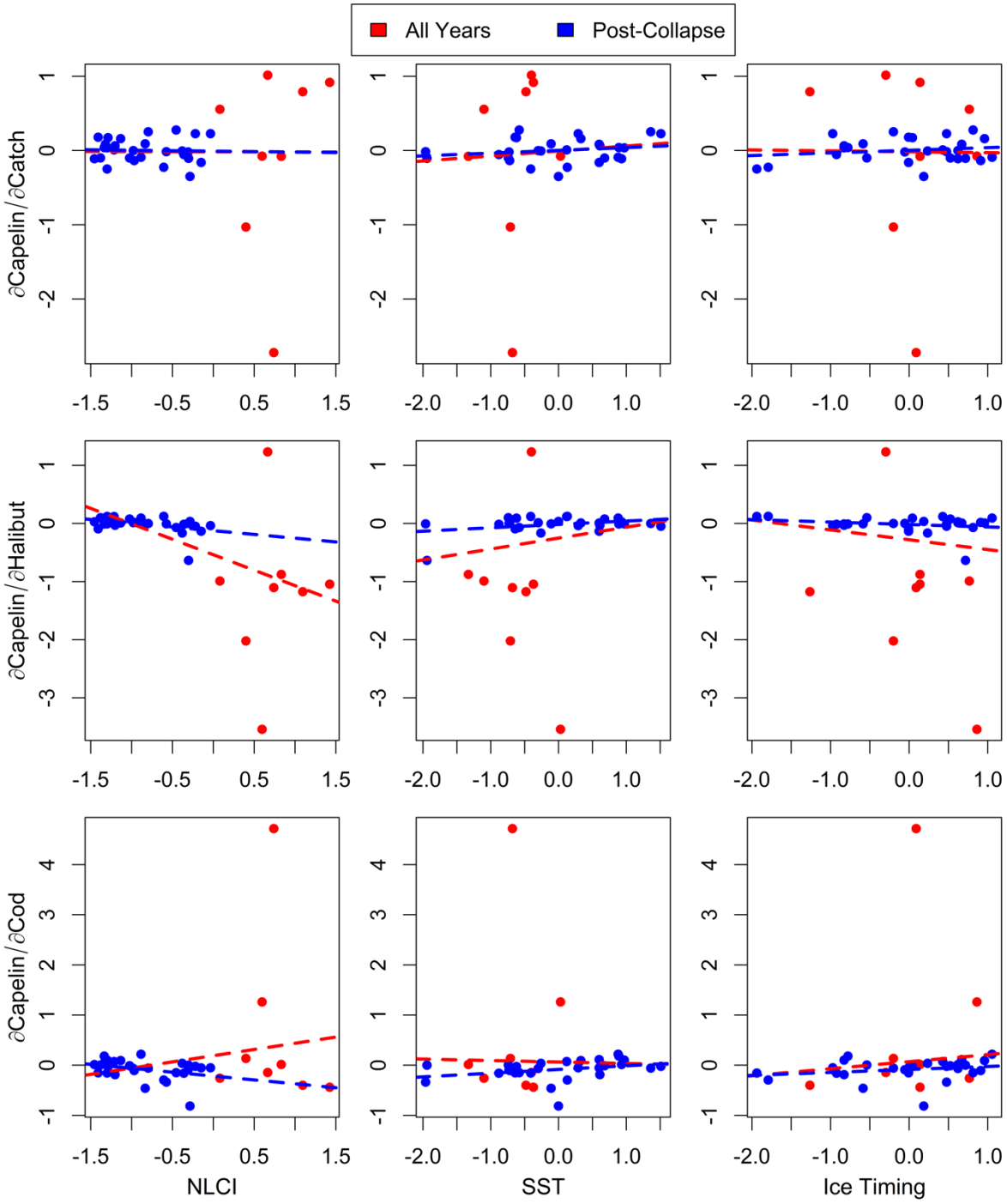

Figure 1.7 Interaction strengths calculated via the S-Map method between capelin and capelin catch, Greenland halibut biomass, and Atlantic cod biomass as a function of the cumulative Newfoundland Climate Index (NLCI) and sea surface temperature (SST) for both the all years and post-collapse datasets. Dashed lines are linear regression lines. The all years dataset includes both red and blue points.

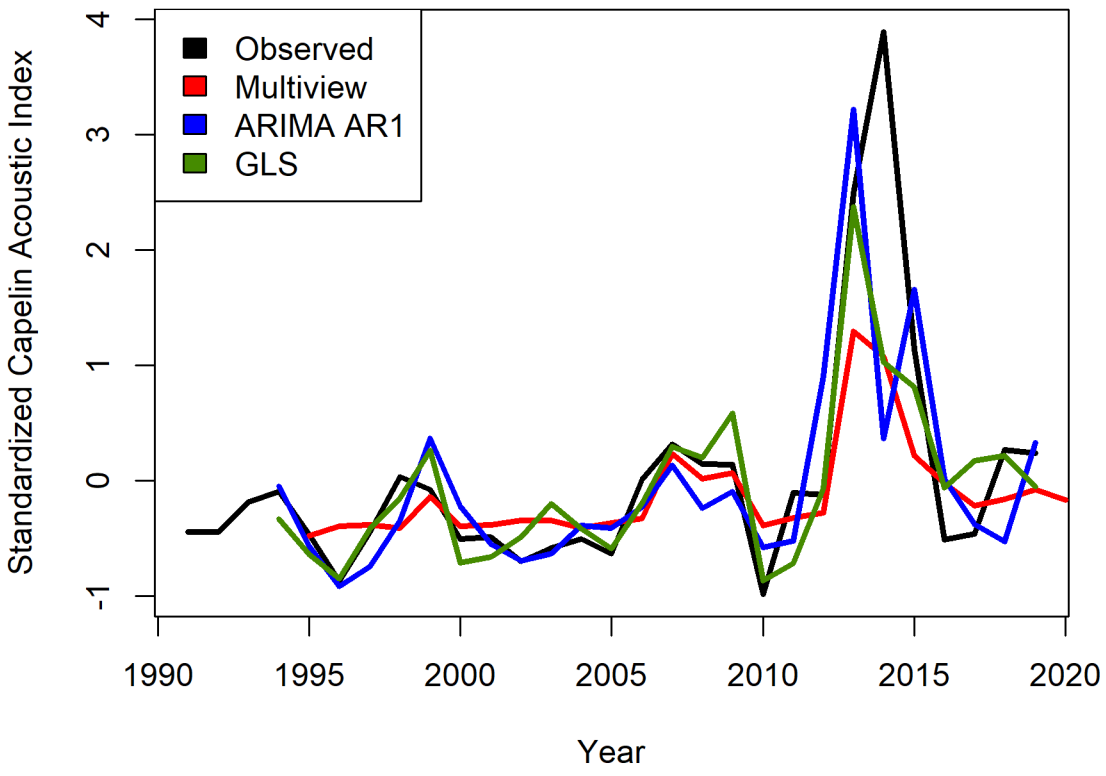

Figure 1.8 Model fits for MVE, ARIMA AR1, and GLS predictive models for the standardized capelin acoustic index using post-collapse data.

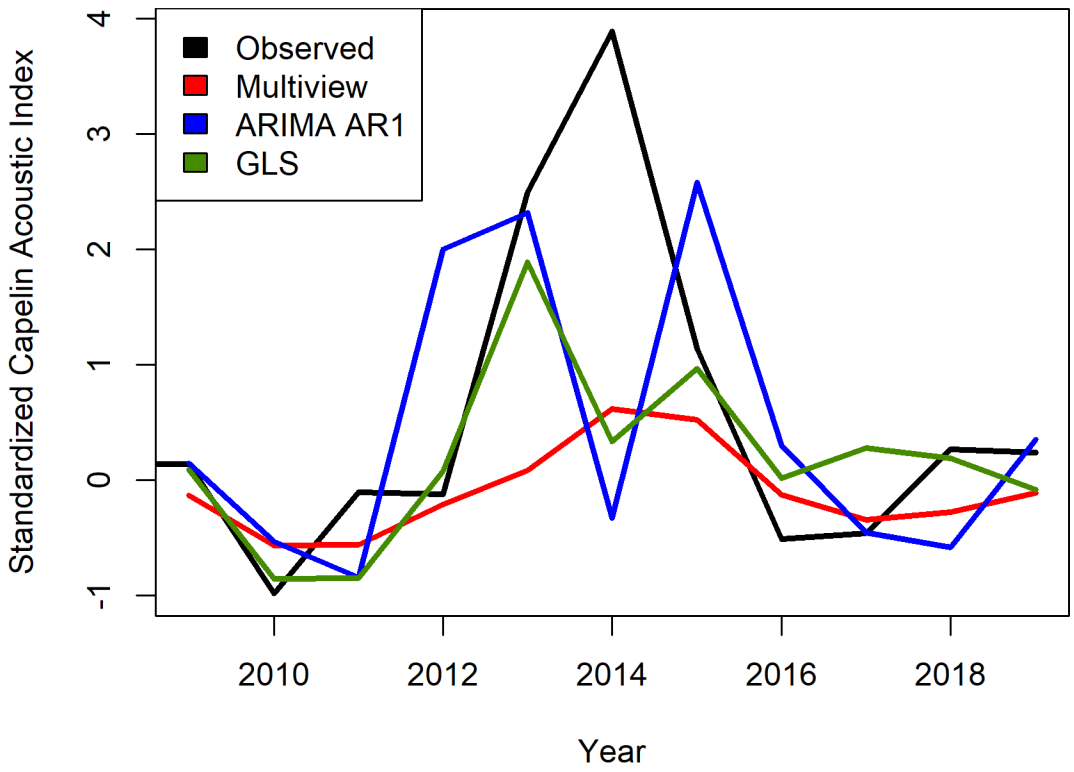

Figure 1.9 Forecast experiment for MVE, ARIMA AR1, and GLS predictive models for the standardized capelin acoustic index using post-collapse data.

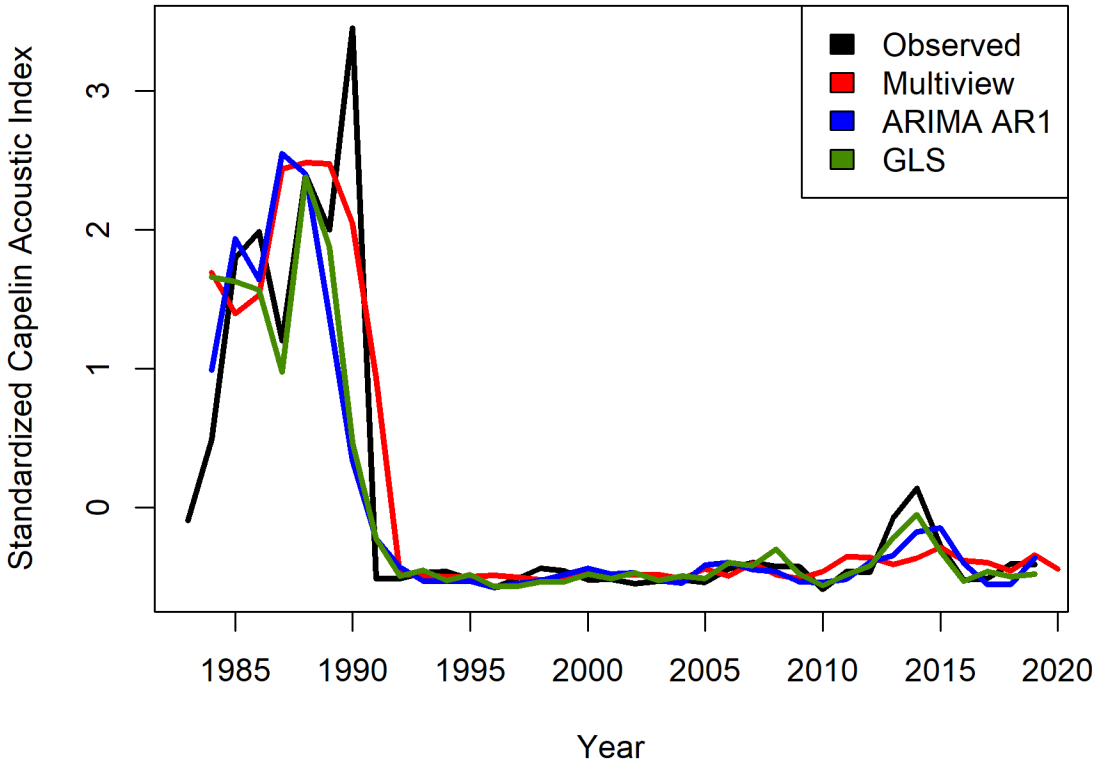

Figure 1.10 Model fits for MVE, ARIMA AR1, and GLS predictive models for the standardized capelin acoustic index using all available years.

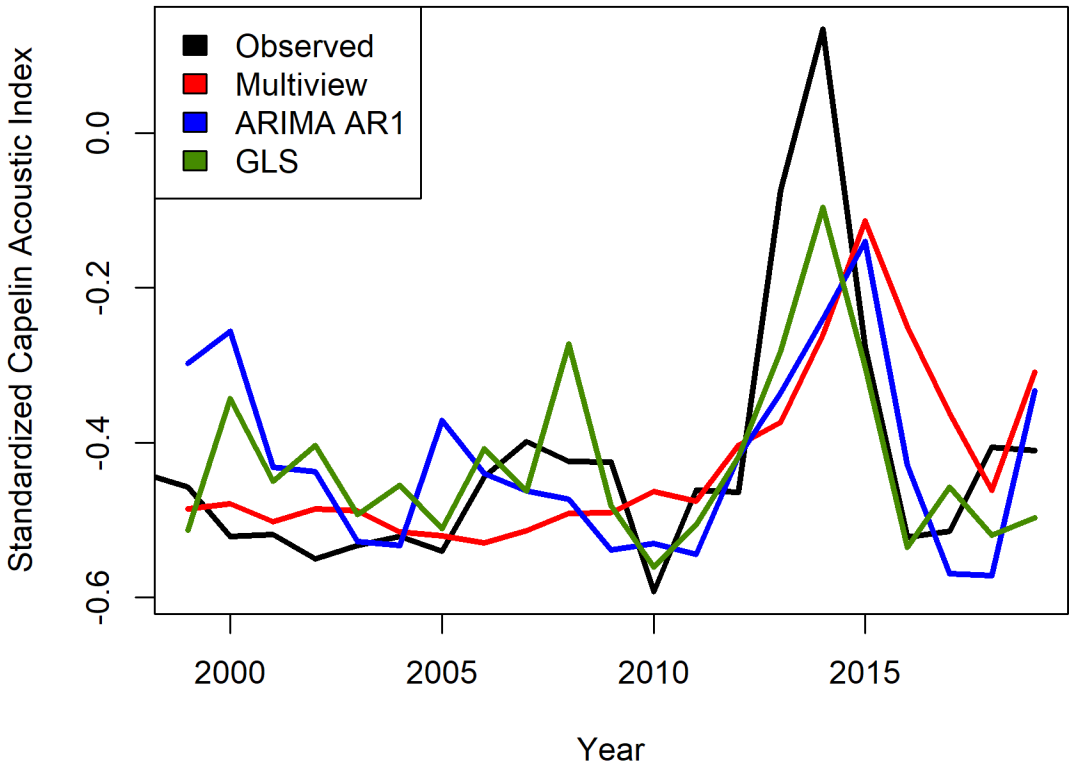

Figure 1.11 Forecast experiment for MVE, ARIMA AR1, and GLS predictive models for the standardized capelin acoustic index using all available years.

Chapter 2: Comparing Capelin and Atlantic Cod Biomass Drivers using Scenario Exploration

2.1 Introduction

As mentioned in chapter 1, the capelin (*Mallotus villosus*) stock on the Newfoundland Shelf collapsed in the 1991 during a period of abnormally cold climate , contributing to an overall regime shift in the ecosystem (Bourne et al., 2021; DFO, 2012). In addition to biomass, capelin life histories have changed since then, with capelin maturing earlier (Bourne et al., 2021), and capelin spawning later in the year (Murphy et al., 2021). Where chapter 1 focuses exclusively on capelin, Atlantic cod (*Gadus morhua*) is also analyzed in this chapter for comparison. Atlantic cod is a larger fish with a higher trophic position in the food web, a longer generation time, and a less stochastic time series (Figure 1.2); yet, cod has a similar history to capelin in terms of population dynamics, making it an ideal bridge between EDM analyses of capelin and other larger fish species. Atlantic cod experienced a similar population collapse in 1992, which is often associated with overfishing, and also have not yet recovered to pre-collapse levels (Gomes et al., 1995; NAFO, 2010). Like capelin, cod also now mature earlier (Gomes et al., 1995). Despite a brief period of partial recovery in the mid 2010s, the regime shift and collapsed populations of capelin and cod, as well as their changed life history strategies, remain to this day despite reduced fishing pressure and more typical climatic conditions (Bourne et al., 2021; Buren et al., 2014; DFO, 2021; Murphy et al., 2021). As ecosystems on the Newfoundland Shelf change due to climate change, changing fisheries, and the slow recovery of capelin, cod, and other collapsed fish species, predicting what conditions may lead to cod and capelin recovery or further decline is necessary to responsibly manage these species into the future.

One potential tool for this purpose is Empirical Dynamic Modelling (EDM) scenario exploration. As shown in chapter 1, EDM is capable of identifying causative covariates which drive fisheries time series using convergent cross-mapping (CCM, Sugihara et al., 2012), and using these drivers to generate nonlinear predictions (Deyle et al., 2013; Giron-Nava et al., 2021; Glaser et al., 2014b; Kuriyama et al., 2020; Munch et al., 2018; Wasserman et al., 2022). However, these models do not directly show what influence causative drivers have on fish biomass, or how that influence may change with changing ecosystem conditions. Scenario exploration seeks to solve this issue by comparing how EDM model predictions change when the values of driving covariates are perturbed. For example, scenario exploration has been used to show that incidence of influenza and humidity are positively correlated at high humidity and negatively correlated at low humidity (Deyle et al., 2016a).

In the context of fisheries, scenario exploration can be a useful tool in ecosystem-based management, where it can be used to predict how species react to changes in climatic conditions, fishing pressure, or biomass of other species in the ecosystem (Munch et al., 2023). As an example, scenario exploration has been used to predict how changes in temperature affect sardine abundance, and how this relationship relates to fishing pressure and management of future sardine biomass (Deyle et al., 2013). This utility suggests scenario exploration may be an ideal tool for predicting population dynamics for capelin and cod, which have experienced extreme changes in climatic conditions, fishing pressure, and ecological regimes in recent history.

The objectives of this chapter are (1) to identify drivers of capelin and cod biomass using CCM, (2) use scenario exploration to assess the magnitude and direction of each driver's effect on

capelin and cod biomass and (3) how the magnitude and direction change with the relative value of the driver and/or biomass.

2.2 Methods

Data Series

Data series used were the same as for chapter 1, except for the separation of the NLCI in to the cumulative NLCI and annual NLCI to distinguish long-term and short-term climatic change, and the addition of cod catch data for 2J3KL, which was also gathered from the NAFO STATLANT 21 database (NAFO, 2021). Capelin biomass dynamics were gathered from the Fisheries and Oceans Canada (DFO) spring acoustic survey (Mowbray, 2012) using the hole-filling methodology from chapter 1 to fill missing years. Cod biomass dynamics were derived from the DFO fall random stratified bottom trawl survey index in NAFO divisions 2J3KL (Doubleday, 1981; Smith and Somerton, 1981). Covariate data series tested against the cod and capelin biomass indices included timing of sea ice retreat, the Newfoundland Climate Index (NLCI, Cyr and Galbraith, 2021), its component parts [winter North Atlantic Oscillation (NAO), air temperature, sea ice duration and area, iceberg count below 48°N, sea surface temperature (SST), vertically averaged temperature and salinity at Station 27, cold intermediate layer (CIL) core temperature at Station 27, Newfoundland shelf CIL area and bottom temperature], the cumulated sum of the NLCI (cumulative NLCI), day of year of sea ice retreat (henceforth referred to as ice timing), capelin and cod catch (NAFO, 2021), and Greenland halibut fall random stratified bottom trawl survey index in NAFO divisions 2J3KL (Doubleday, 1981; Smith and Somerton, 1981). The same datasets were used for analyses on both capelin and cod except for catch, which was matched to the target species.

EDM Analyses

As in chapter 1, univariate simplex projections were used to calculate the optimal embedding dimension (E) in the range of 1 to 5, and univariate S-Map projections were used to calculate the optimal weighting parameter (θ) in the range of 0.01 to 9 for both the capelin acoustic index and cod bottom trawl survey index. Capelin and cod biomass indices at optimal E were tested for causative relationships against all covariates using convergent cross mapping (CCM). Significance of CCM results was tested by producing p-values by comparison to a null distribution of 1000 CCMs (Chang et al., 2017) using phase-randomized surrogate datasets (Ebisuzaki, 1997).

Scenario Exploration

Scenario exploration methodology was adapted from Deyle et al. (2016). Data were time-delay embedded (*i.e.,* separated into lags) and converted to normalized anomalies before scenario exploration. For both cod and capelin, multivariate S-Maps predicting the biomass index (B) one year in the future ($t+1$) were fit using the previously calculated optimal weighting parameter θ and E lags of the biomass index and the most recent year of each covariate (Cov) which returned significant CCM results:

$$B(t+1) = f[B(t), B(t-1), \dots \big(B\big(t-(E-1)\big), Cov(t)\big],$$

These S-Maps were then used to run predictions on the same dataset with the covariate perturbed positively and negatively by a set value (Δ):

$$B_{+}(t+1) = f[B(t), B(t-1), \dots B\big(t-(E-1)\big), Cov(t) + \Delta], \quad (1)$$

$$B_{-}(t+1) = f[B(t), B(t-1), \dots B\big(t-(E-1)\big), Cov(t) - \Delta], \quad (2)$$

Δ values of 0.5 and 1 (corresponding to half a standard deviation and one standard deviation) were used, yielding four perturbation scenarios. The magnitude and direction of each covariate's influence on the biomass indices per change in the covariate (henceforth referred to as $\Delta B/\Delta Cov$) were calculated by subtracting the positive perturbation predictions by the negative perturbation predictions, and dividing by the total perturbed difference in the covariate to get a rate, for each year in the time series:

$$\frac{\Delta B}{\Delta Cov} = \frac{B_{+}(t+1) - B_{-}(t+1)}{2\Delta}, \quad (3)$$

Where $2\Delta = [Cov(t) + \_] - [Cov(t) - \Delta]$. Boxplots of $\Delta B/\Delta Cov$ for all covariate and Δ combinations were compared to assess the relative magnitude and overall direction of the influence of each covariate on capelin and cod biomass indices. For each covariate and Δ combination, time series plots of B_{+} and B_{-}, and plots of $\Delta B/\Delta Cov$ against the covariate were produced to assess the relationship between the value of the covariate and its resulting influence on biomass, and how this influence changed over time and with the value of the covariate. All models used the full dataset, and post-collapse and pre-collapse points were distinguished in plots for ease in detecting differences in patterns before and after the regime shift.

2.3 Results

CCM testing results are in Table 2.1. CCM testing returned significant results for catch, Greenland halibut biomass, the cumulative NLCI, and SST across both capelin and cod. Ice timing and air temperature also returned significant CCM results for capelin. No covariates returned significant results for cod and not for capelin. Cod and capelin also returned significant cross-map results against each other.

SST, cumulative NLCI, Greenland halibut biomass, and catch scenario explorations revealed positive effects on both capelin and cod (Figure 2.1, Figure 2.2). Capelin and cod also had primarily positive effects on each other. Air temperature and ice timing effects were slightly negative overall but were generally weak and outlier driven. For both cod and capelin, catch and cumulative NLCI had the largest effects on *ΔB/ΔCov*, followed by their effects on each other. The positive effects of SST and Greenland halibut on *ΔB/ΔCov* were weaker and more outlier-driven (Figure 2.1, Figure 2.2).

In scenario explorations using climatic covariates, capelin generally benefitted from warming climate during the collapsed period and were affected minimally or negatively by warming climate in the pre-collapse period. Perturbing SST upwards led to predictions of several small recoveries throughout the post-collapse period, and there was a clear positive relationship between *ΔB/ΔSST* and SST at high SST in the post-collapse period (Figure 2.3). Conversely, there was a negative relationship between SST and *ΔB/ΔSST* in the pre-collapse period, which also extended into the brief recovery of capelin in the mid 2010s (Figure 2.3). Increased cumulative NLCI decreased capelin predictions before the collapse, increased them in the early 1990s and from 2005 to 2019, had no effect from 1995 to 2005 (Figure 2.4). This pattern is reflected in *ΔB/ΔCNLCI* as no effect at low cumulative NLCI, to increasing *ΔB/ΔCNLCI* with cumulative NLCI, which eventually turns downwards and becomes negative at the high cumulative NLCI values of the pre-collapse period (Figure 2.4). Air temperature and ice timing similarly negatively affected *ΔB/ΔCov* when increased during the pre-collapse period, though *ΔB/ΔCov* was slightly increased by increased air temperature and decreased by increased ice timing for most of the post-collapse period (Figure 2.5, Figure 2.6).

Capelin surprisingly tended to benefit both from increased presence of predators and increased catch in scenario explorations,. Through the entire time series, capelin predictions and *ΔB/ΔCov* consistently increased with increased catch and increased cod biomass (Figure 2.7, Figure 2.8). Increased Greenland halibut biomass increased capelin biomass predictions except for the pre-collapse period using the 0.5 standard deviation perturbation, for which this trend was flipped (Figure 2.9). *ΔB/ΔGH* trended upwards with Greenland halibut biomass at high values, though the relative lack of years contributing to this pattern suggests it could be due to chance (Figure 2.9).

Across all scenarios explored, Atlantic cod biomass predictions consistently increased with both climatic and ecological covariates. Cod biomass predictions increased with SST in the pre-collapse period and slightly during the post-collapse period and showed a clear increasing pattern of *ΔB/ΔSST* with SST over both periods (Figure 2.10). This pattern was repeated in cumulative NLCI (Figure 2.11). Increased catch perturbations resulted in increased cod predictions in the years before and immediately after the collapse, and immediately before and during the slight recovery period in the mid 2010s (Figure 2.12). Perturbing capelin biomass resulted in cod biomass predictions increasing with capelin in every year of the time series, though this increase was much larger in the post-collapse period (Figure 2.13). There was no clear pattern between capelin biomass and cod *ΔB/ΔCapelin* (Figure 2.13). Lastly, increasing Greenland halibut biomass similarly greatly increased cod predictions in the pre-collapse period and slightly increased them in most of the post-collapse period, but conversely exhibited a decreasing pattern in *ΔB/ΔGH* with increasing Greenland halibut biomass (Figure 2.14).

2.4 Discussion

In this study, EDM scenario exploration was able to identify and rank drivers of both capelin and cod biomass, and illuminate clear relationships between these biomass drivers and the direction and degree to which they drive biomass, though it should be noted that this is predicated on the assumption that the relationships between cod and capelin biomass and the other covariates are correctly characterized by the S-Map model. For both cod and capelin, I found long-term climatic change as measured by the cumulative NLCI was the strongest climatic driver of biomass. Capelin and cod biomass were the strongest ecological drivers of each other outside of catch, which was likely confounded by it's use in the model as a driver of biomass when it is more likely to be driven by biomass. I also found that warming climate was generally beneficial for both capelin and cod across all climatic drivers, though the magnitude of this benefit, and the conditions in which it manifested differed. For example, warming climate as measured by the cumulative NLCI had a negligible effect on either cod or capelin until a certain threshold was reached, and capelin exhibited a second threshold after which warming climate began to have a negative effect on biomass, where cod did not (Figure 2.4, Figure 2.11). Conversely, ecological drivers were less consistent, exhibiting a variety of different $\Delta B/\Delta Cov$ patterns.

The main weakness of EDM scenario exploration iI IhiI study is its inability to discern directionality in closely related time series. Resulting violations in the assumption that the S-Map model is suitable for predicting biomass based on the covariates can cause spurious results, which may be further exacerbated by covariate perturbations outside the range of the observed data. This weakness is most obvious in the positive relationship returned between catch and biomass for both capelin and cod. Logically, it would be expected that catch would lead to decreased biomass, especially for the 2J3KL Atlantic cod stock, which has a well-documented

history of being overfished (DFO, 2012; Gomes et al., 1995; NAFO, 2010). The positive relationship returned by scenario exploration was extremely large in comparison to actual cod dynamics and is a statistical artifact likely associated with cod catch acting as an index of cod, leading to high cross-correlation between catch and cod (Figure 1.5). Similarly, this also represents a weakness of using raw catch data to assess fisheries. It is possible that a different relationship could be detected using other metrics which better account for the proportion of fishing mortality being experienced by the stock, such as fishing effort or catch per unit effort (e.g. Giron-Nava et al., 2021). This serves as an avenue for potential future research into the utility of EDM scenario exploration in fisheries management.

Another potential weakness of scenario exploration in my study is my focus on single covariate models. In reality, many environmental and ecological covariates may be constantly changing in response to each other, which complicates predictions in comparison to the single covariate tests I perform in this study. Focusing on single covariates may result in missed synergistic and/or antagonistic interactions between covariates – for example, I show that an increase in cumulative NLCI benefits cod, but this does not include how the increase in capelin I also associate with an increase in cumulative NLCI would affect cod alongside it. This also applies to covariates which are not accounted for in my model. As an example, it remains to be seen whether a return to the pre-collapse condition in long-term climate state would result in recovery for capelin and cod, or if such a recovery would be hampered by other mechanisms, such as their altered life history strategies since the collapse. As these time series grow and experience new phases, these interactions, and the ability of EDM to account for them, will become clearer. Though careful thought would be required to perturb multiple covariates realistically, multivariate scenario

exploration remains an interesting avenue for potential research into the use of EDM scenario exploration as a tool for stock assessment.

Regardless, this study shows EDM scenario exploration can be utilized to predict how species may react to environmental and/or ecological changes in their environment. This information can be applied in the context of management to predict how species may respond to changes in the environment, management actions, or a combination of both (e.g. Deyle et al., 2013). For example, EDM scenario exploration could be combined with climate change projections (e.g. Deyle et al., 2022), fishing scenarios (e.g. Giron-Nava et al., 2021), or some combination of both to predict how species may react to different stressors, or combinations of stressors. In the context of cod and capelin, future research into scenario exploration comprising more detailed catch information, more species, and more specific potential future scenarios would be beneficial to fully explore the potential of scenario exploration as a tool for their management.

2.5 Tables

Table 2.1 p-values for CCM significance testing against all covariates for both capelin and Atlantic cod. * indicates significant result ($p < 0.05$).

Covariate	Capelin	Atlantic Cod
Catch	<0.001*	<0.001*
Capelin		0.039*
Atlantic Cod	<0.001*	
Greenland Halibut	0.015*	0.008*
Cumulative NLCI	<0.001*	<0.001*
Annual NLCI	0.220	0.107
Winter NAO	0.163	0.342
CIL Area	0.807	0.334
Sea Ice	0.594	0.162
Icebergs	0.853	0.566
Ice Timing	0.009*	0.673
Air Temperature	0.026*	0.060
SST	0.001*	0.028*
Bottom Temperature	0.149	0.089
S27 Temperature	0.072	0.060
S27 Salinity	0.085	0.640
S27 CIL	0.269	0.456

2.6 Figures

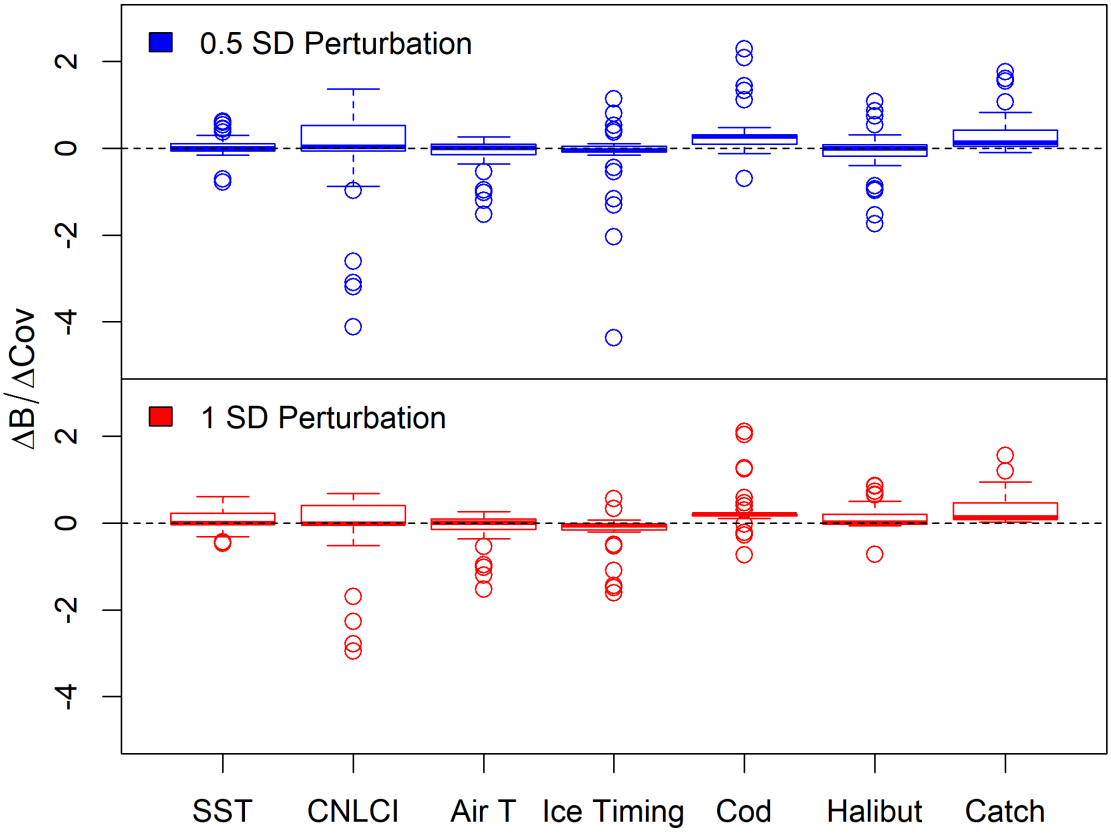

Figure 2.1 Boxplots of predicted change in normalized capelin acoustic index per change in covariate (ΔB/ΔCov) calculated from EDM scenario exploration using ±0.5 standard deviation (top, blue) and ±1 standard deviation (bottom, red) perturbation scenarios.

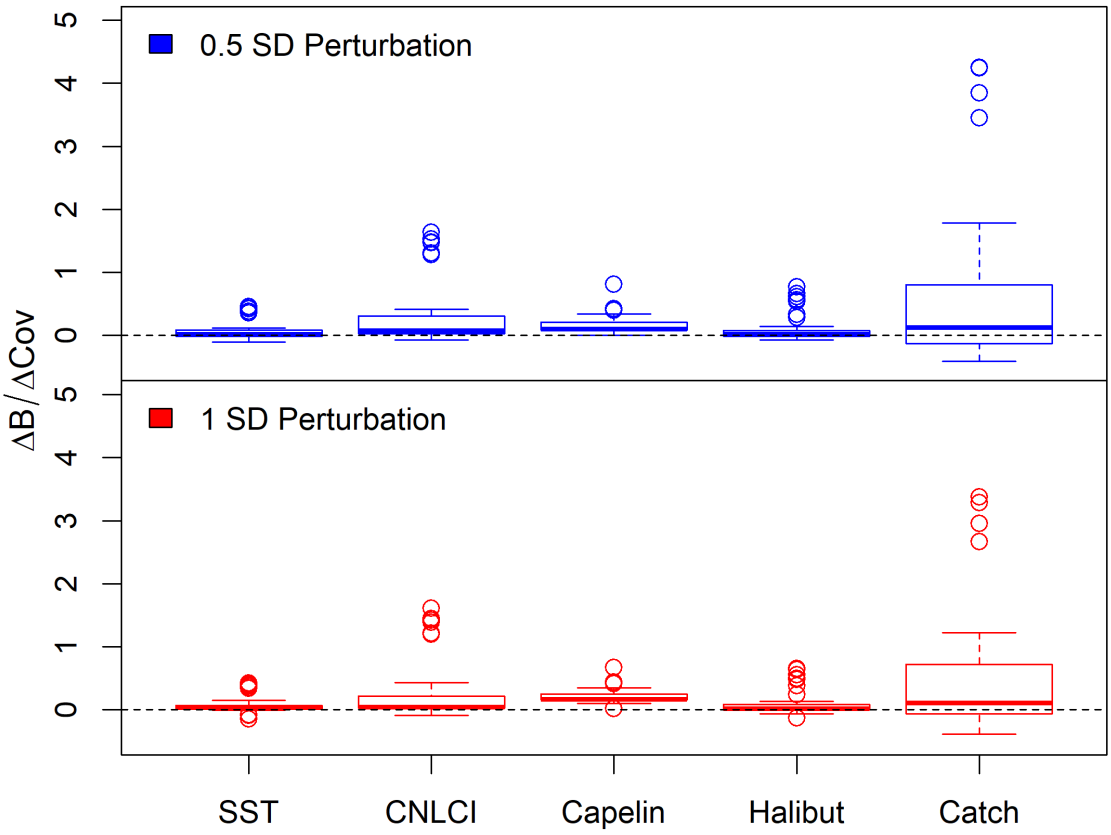

Figure 2.2 Boxplots of predicted change in normalized cod fall bottom trawl survey biomass index per change in covariate (ΔB/ΔCov) calculated from EDM scenario exploration using ±0.5 standard deviation (top, blue) and ±1 standard deviation (bottom, red) perturbation scenarios.

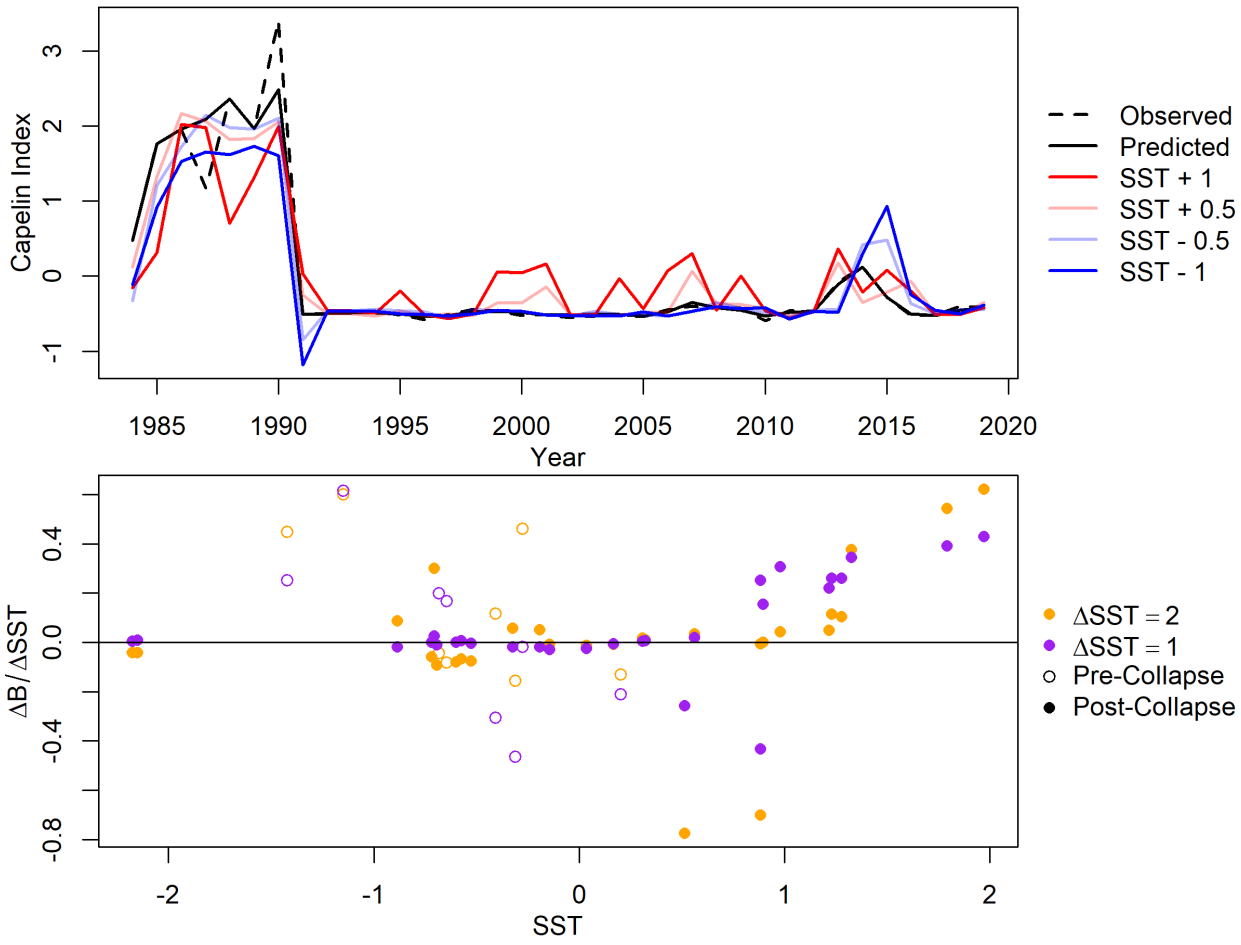

Figure 2.3 Predicted changes in the capelin acoustic index using S-Map scenario exploration with SST perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against normalized SST.

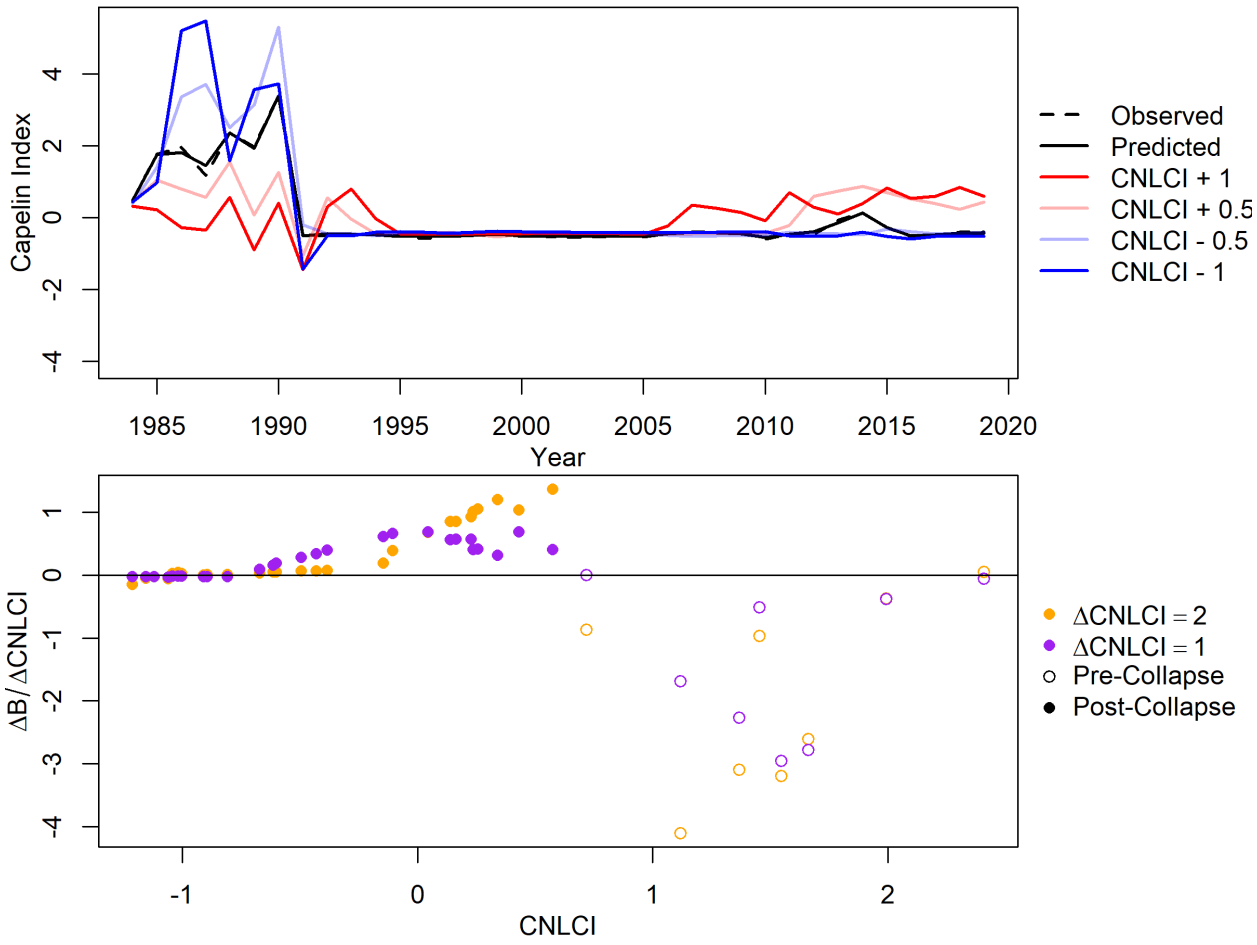

Figure 2.4 Predicted changes in the capelin acoustic index using S-Map scenario exploration with the cumulative NLCI perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against the normalized cumulative NLCI.

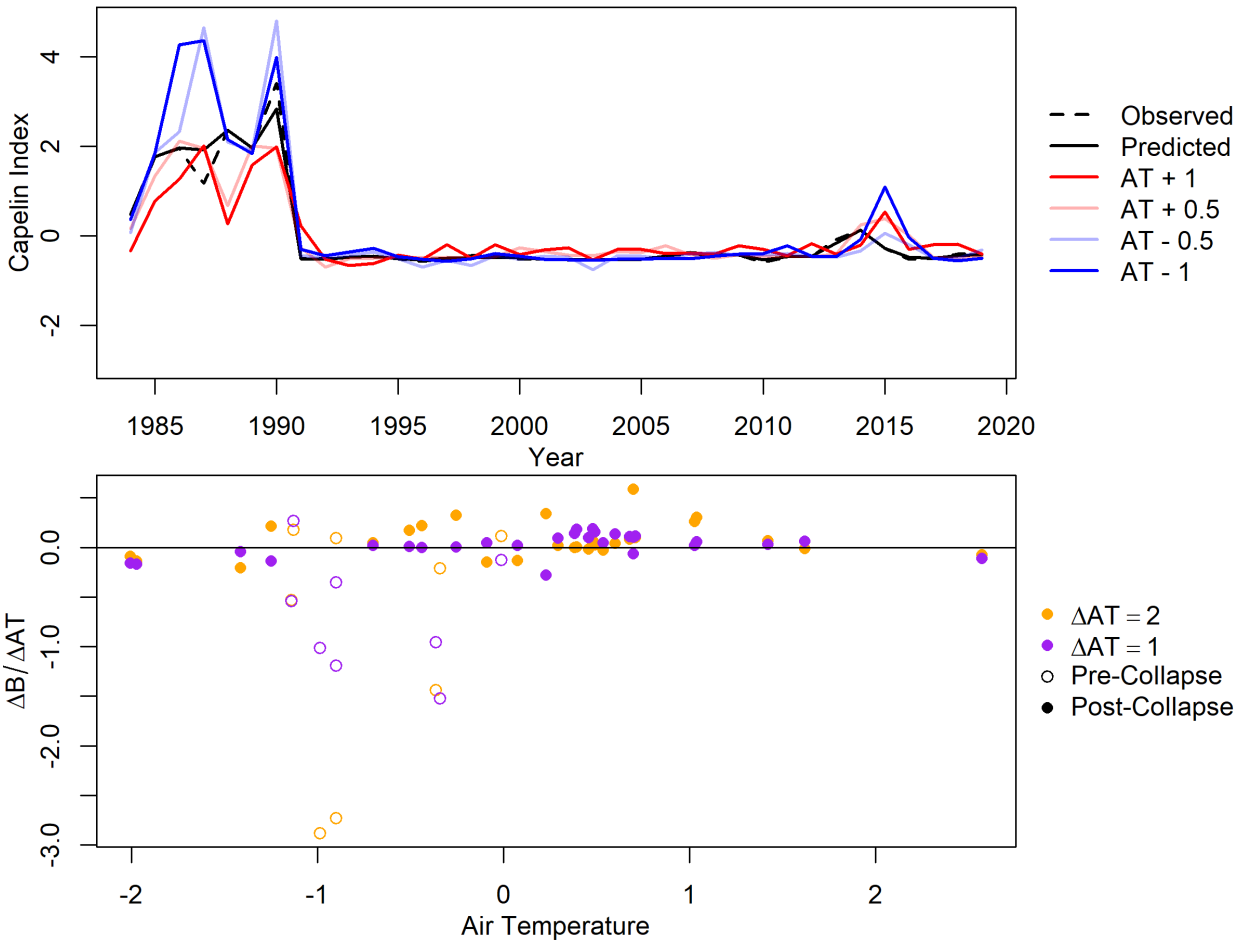

Figure 2.5 Predicted changes in the capelin acoustic index using S-Map scenario exploration with air temperature perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against normalized air temperature.

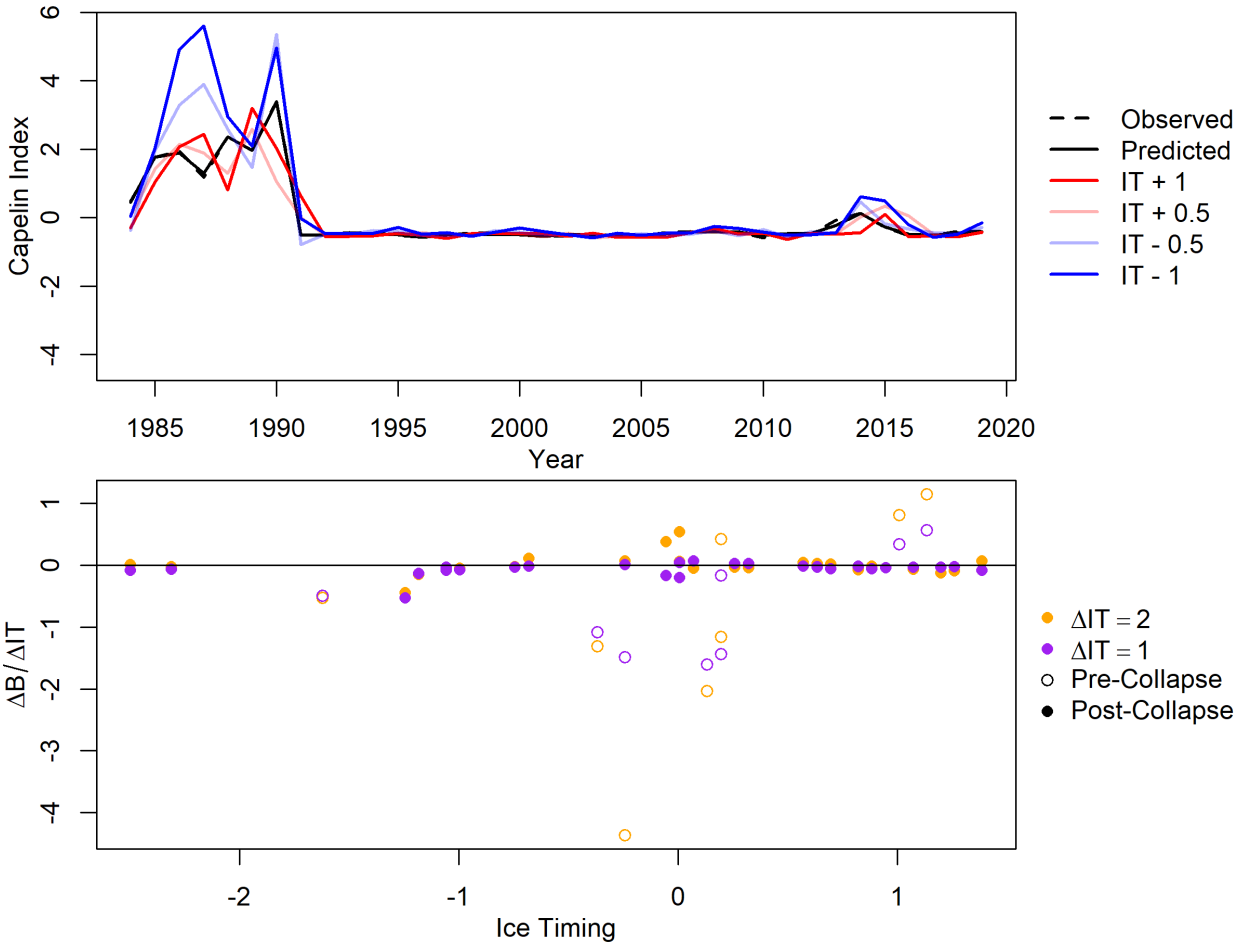

Figure 2.6 Predicted changes in the capelin acoustic index using S-Map scenario exploration with ice timing perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against normalized ice timing.

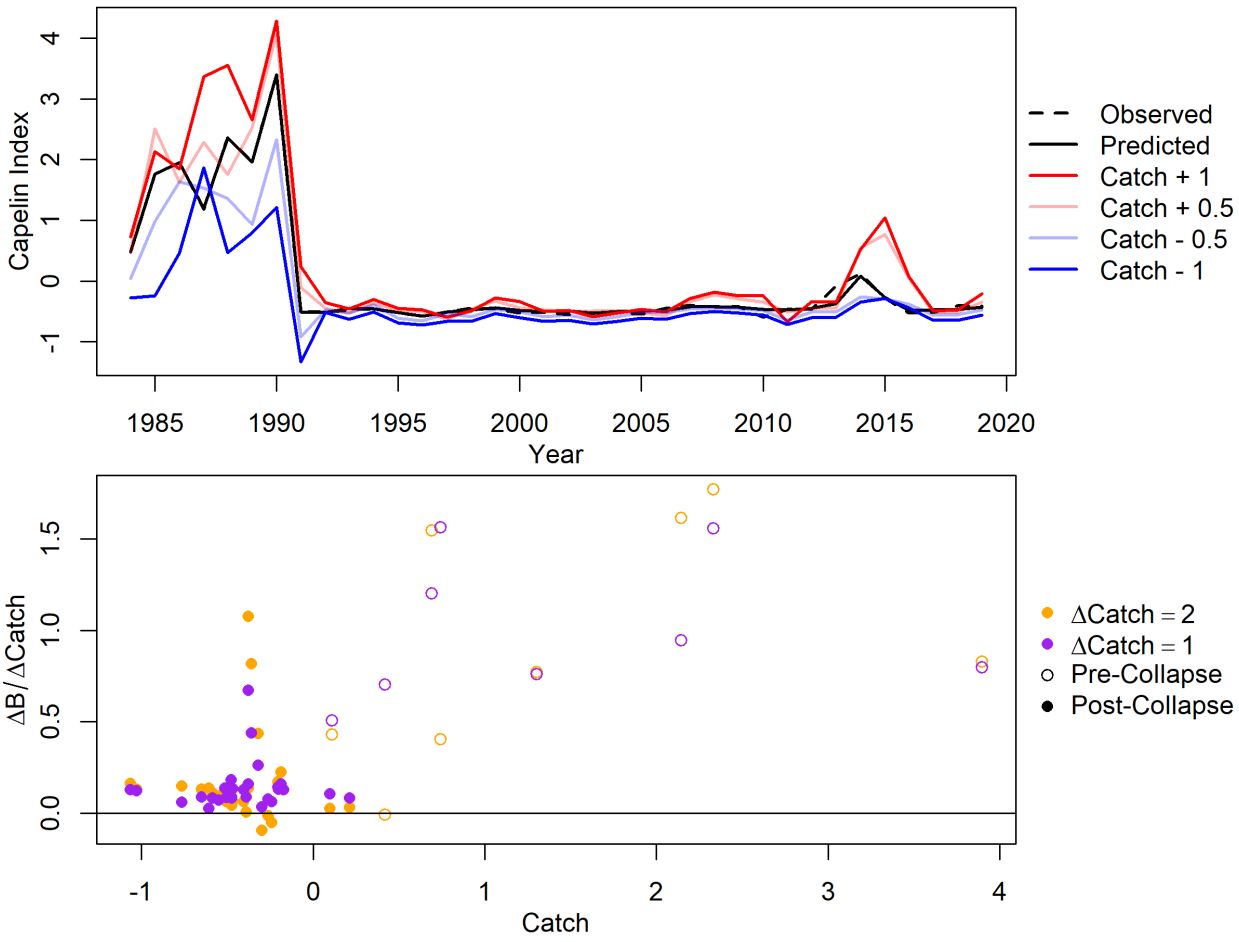

Figure 2.7 Predicted changes in the capelin acoustic index using S-Map scenario exploration with capelin catch perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against normalized capelin catch.

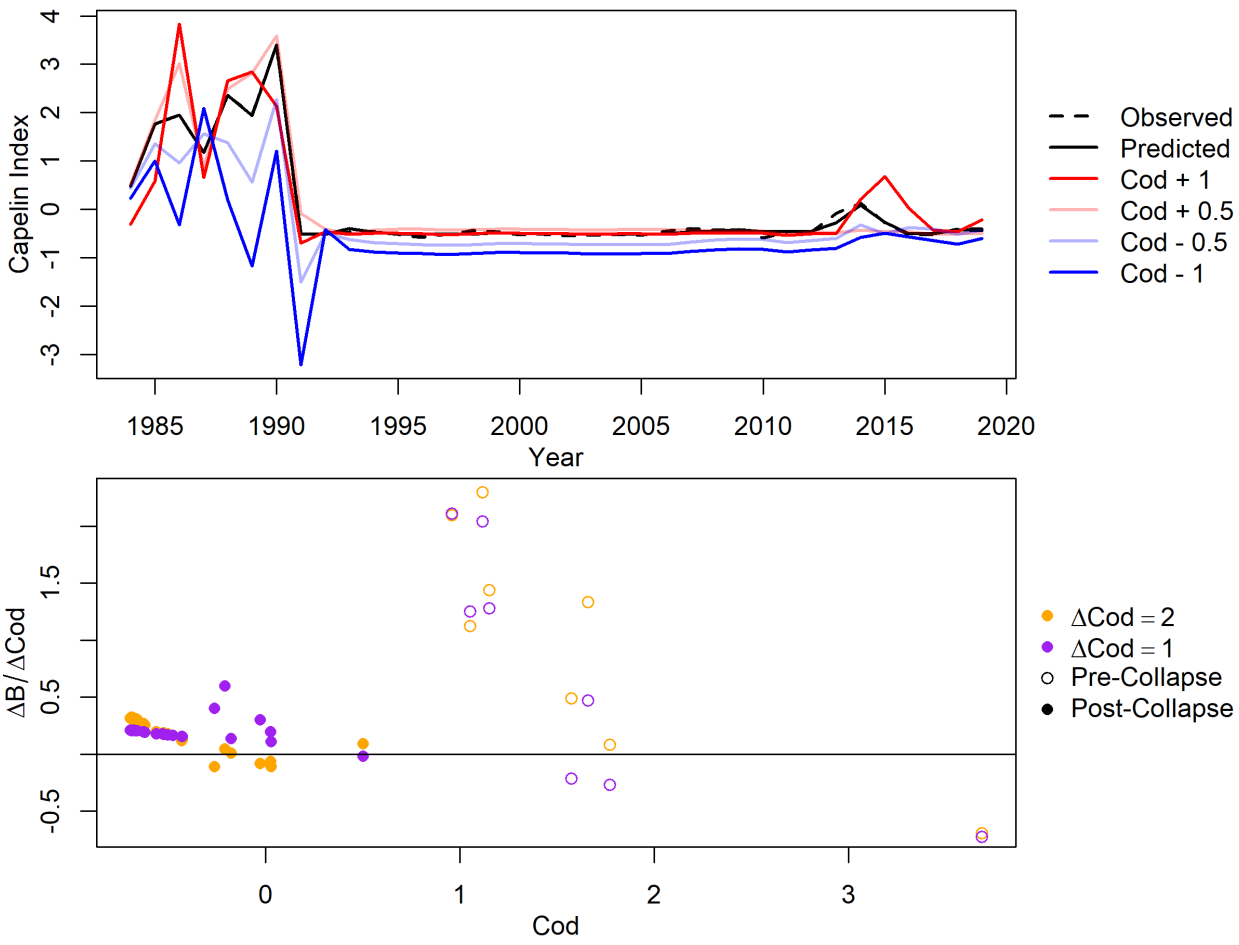

Figure 2.8 Predicted changes in the capelin acoustic index using S-Map scenario exploration with the Atlantic cod bottom trawl survey biomass index perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against the normalized cod biomass index.

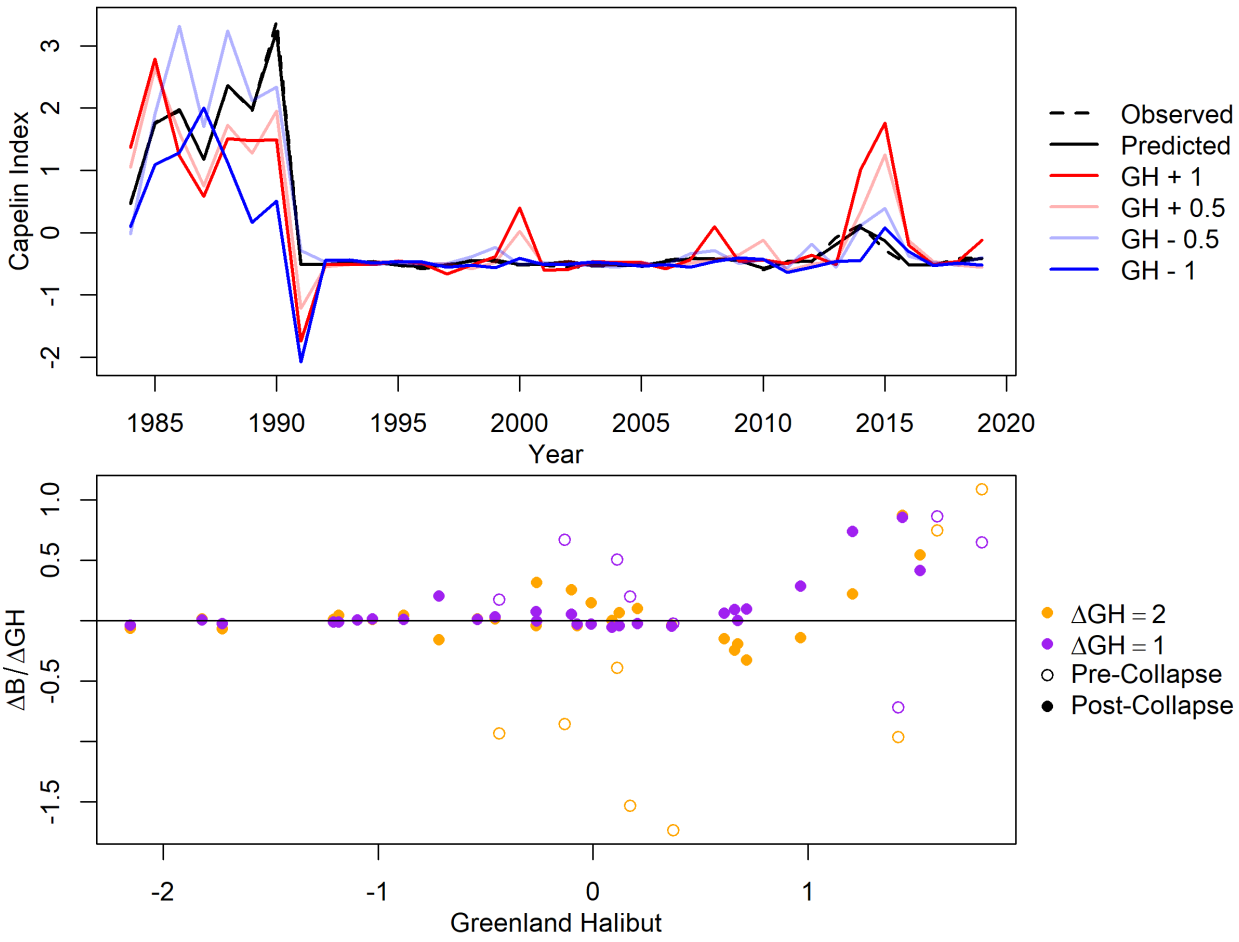

Figure 2.9 Predicted changes in the capelin acoustic index using S-Map scenario exploration with the Greenland halibut bottom trawl survey biomass index perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against the normalized Greenland halibut biomass index.

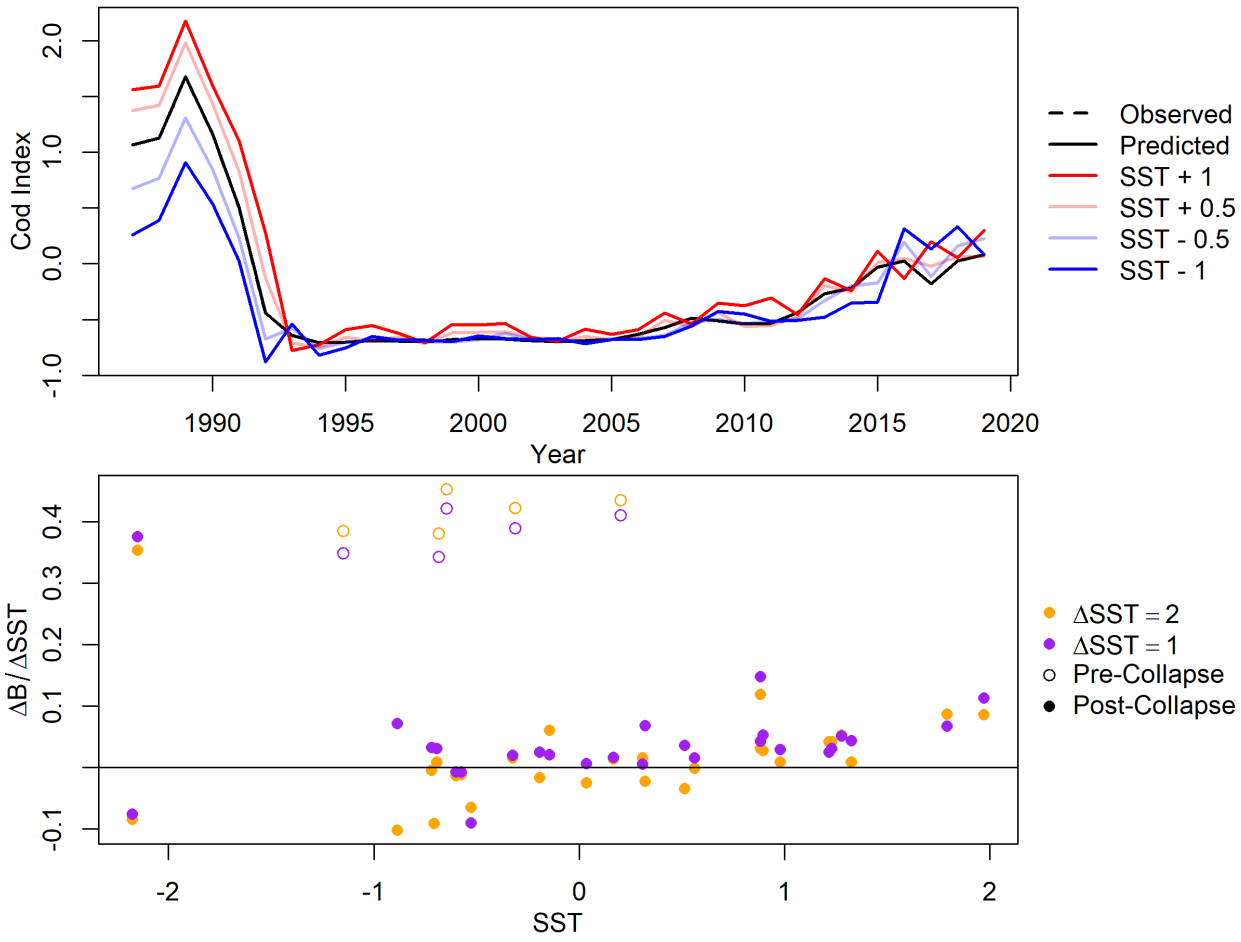

Figure 2.10 Predicted changes in the Atlantic cod bottom trawl survey biomass index using S-Map scenario exploration with SST perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against normalized SST.

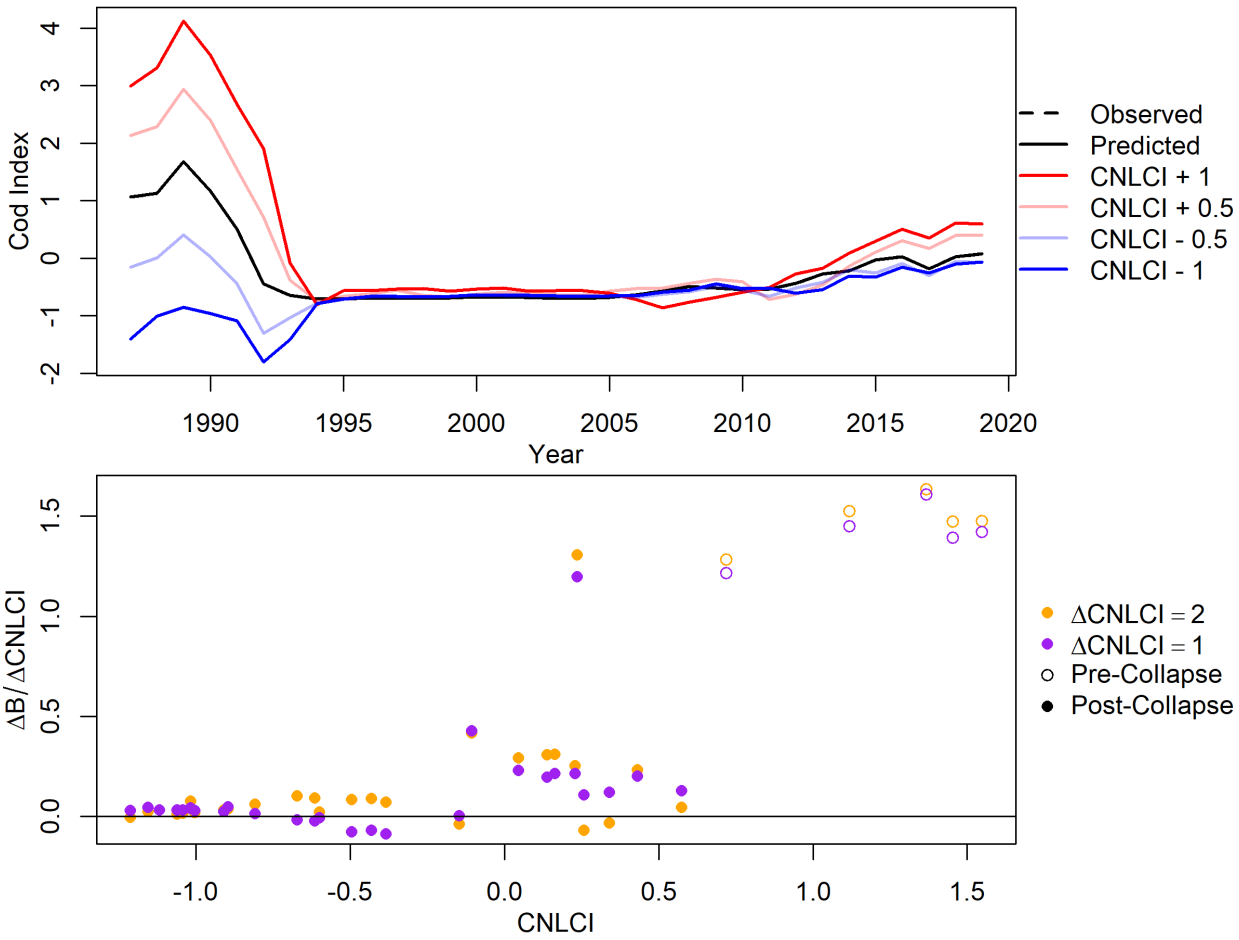

Figure 2.11 Predicted changes in the Atlantic cod bottom trawl survey biomass index using S-Map scenario exploration with the cumulative NLCI perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against the normalized cumulative NLCI.

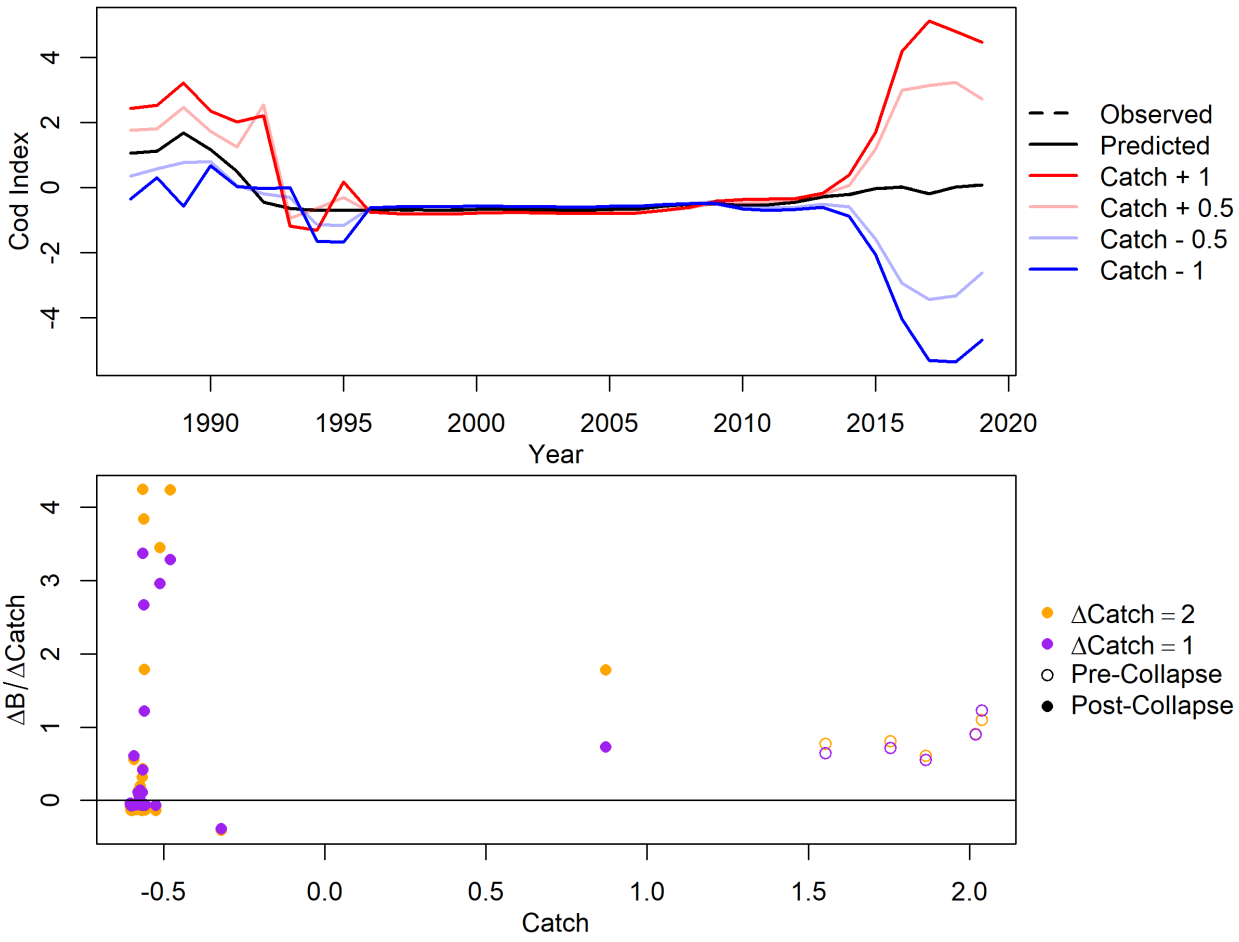

Figure 2.12 Predicted changes in the Atlantic cod bottom trawl survey biomass index using S-Map scenario exploration with cod catch perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against normalized cod catch.

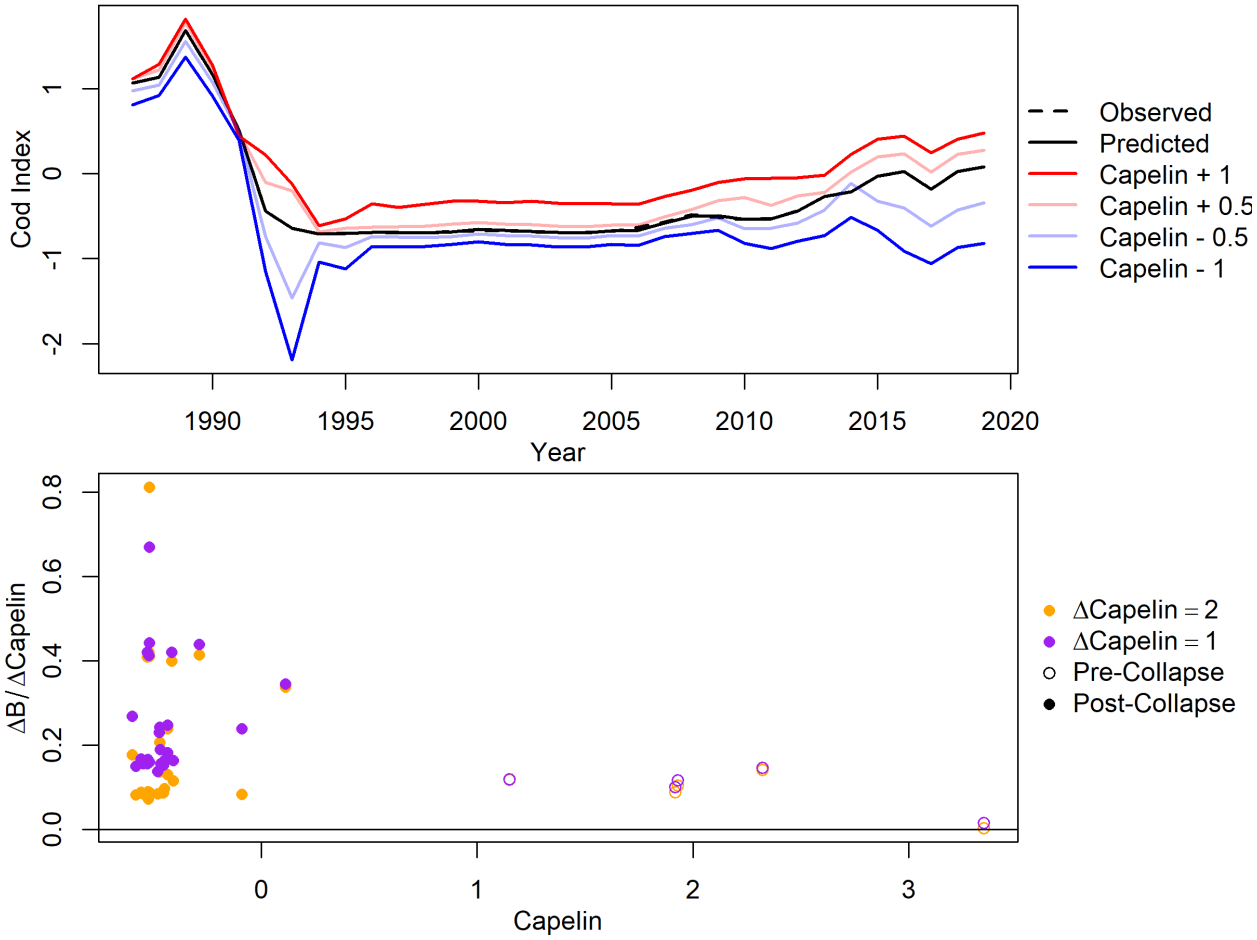

Figure 2.13 Predicted changes in the Atlantic cod bottom trawl survey biomass index using S-Map scenario exploration with the capelin acoustic index perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against the normalized capelin acoustic index.

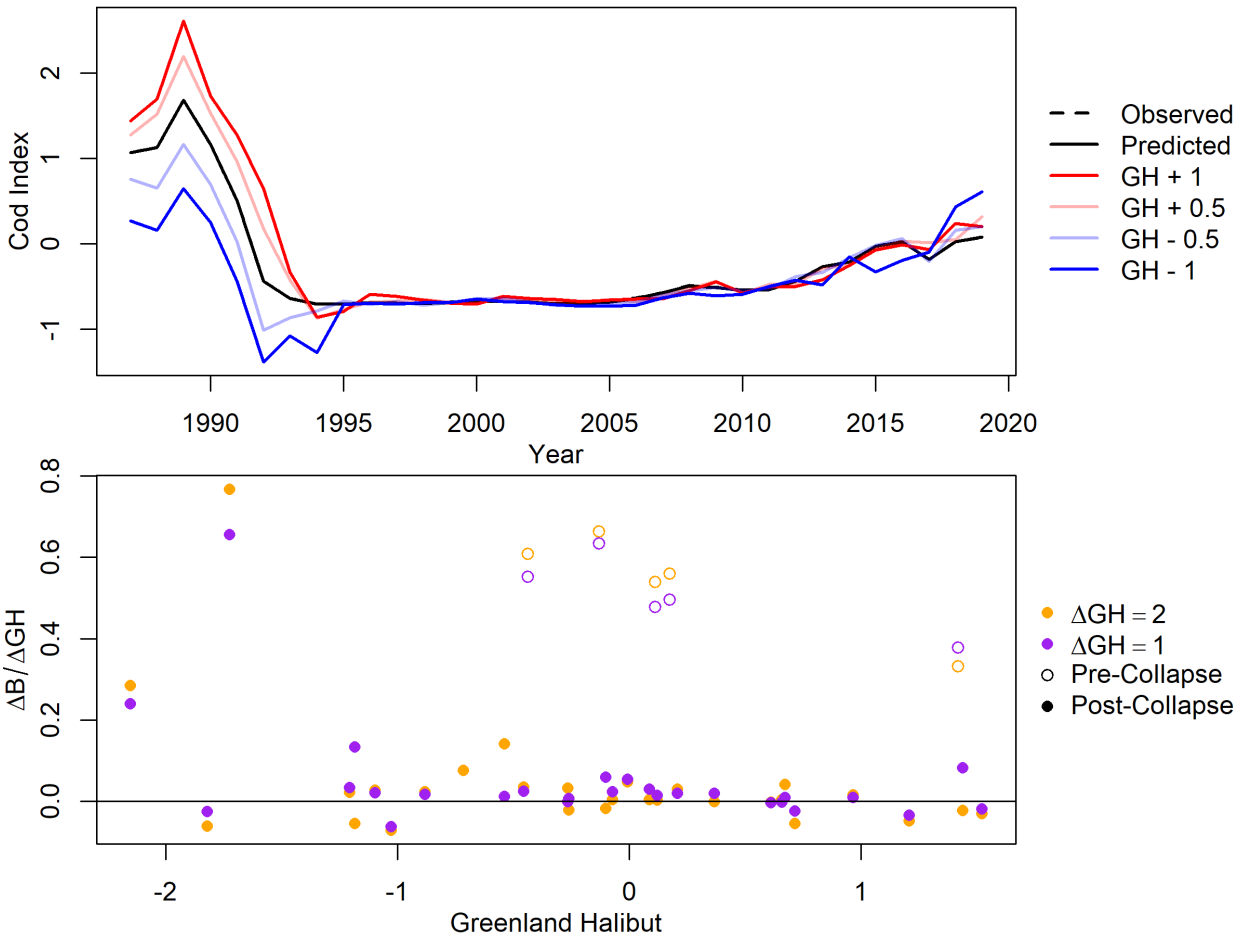

Figure 2.14 Predicted changes in the Atlantic cod bottom trawl survey biomass index using S-Map scenario exploration with the Greenland halibut bottom trawl survey biomass index perturbed positively and negatively by a half standard deviation and a full standard deviation from 1984-2019 (top), and scatterplot of the difference between positive perturbation predictions and negative perturbation predictions for each year in the time series plotted against the normalized Greenland halibut biomass index.

Summary and Conclusion

Overall, the results of this thesis show that EDM can be used to model and predict Newfoundland shelf capelin population dynamics as well as or better than linear models in most cases, though a longer time series and range of dynamics exhibited may be necessary to do so reliably in all cases. This thesis also shows EDM is capable of identifying and analyzing relationships between capelin dynamics and the forces which drive them. In Chapter 1, I show that capelin dynamics are nonlinear, and that in almost all cases, EDM MVE modelling draws with or outperforms linear models at predicting changes in capelin biomass using both fits and forecasts. Chapter 1 also identifies the cumulative NLCI, SST, ice timing, and Greenland halibut biomass as drivers of capelin biomass via CCM, while capelin catch and Atlantic cod biomass were in synchronicity with capelin biomass, indicating a strong causative relationship with unclear directionality. I expanded on this finding in chapter 2, showing that the cumulative NLCI was the strongest clear driver of both capelin and cod population dynamics, with warming climate improving capelin and cod biomass predictions. Increased SST resulted in a slight increase in capelin and cod biomass predictions, supporting this result. Both species exhibited strong positive relationships with each other and with their respective catch, with past literature indicating that capelin biomass drives cod biomass, and that catch is driven by biomass for both species.

We also found EDM exhibited similar weaknesses in both chapters. As already mentioned, EDM was not able to discern directionality in strong causative relationships without the context of past scientific literature. Additionally, the data requirements of EDM were difficult to meet in both chapters. Most EDM analyses require ~35-40 time points (Ye et al., 2015a), which is rare in time

series with one-year timesteps. This prevented the inclusion of some analyses (post-collapse CCM), and limited options for potential covariates. EDM can also struggle with noisy data, as evidenced in chapter 1 by the interaction sign and strength analyses and difficulty in predicting the mid 2010s recovery in capelin biomass. However, it should be noted that these problems are not necessarily limited to EDM. Assigning causality is a complex task, and models alone are not sufficient to properly identify and assign such relationships (e.g. Arif and MacNeil, 2022). Various improvements to EDM have been or are being developed to combat the limitations of EDM listed above and others, including multiview embedding (Ye and Sugihara, 2016), variable step-size EDM (Johnson and Munch, 2022), Empirical Dynamic Programming (Brias and Munch, 2021), pairwise asymmetric inference (McCracken and Weigel, 2014), multispatial CCM (Clark et al., 2015), cross map smoothness (Ma et al., 2014), continuity scaling (Ying et al., 2022), and others. Exploring the use of these methods in capelin stock assessment is a potential area for future research on this topic.

This thesis also identifies many other opportunities for further research into the use of EDM for stock assessment. For example, nonlinear forecasting models can be used to predict stock dynamics and compare results to more traditional forecasting and stock assessment models to identify which models are best for the target species and situation(Deyle et al., 2018; Munch et al., 2018; Sguotti et al., 2020; Ye et al., 2015a). Similarly, CCM and scenario exploration can be used to identify potential environmental, ecological, and anthropogenic covariates of interest and the ways in which they influence target species biomass, both of which can be incorporated into stock assessment models or management advice (Deyle et al., 2013; Giron-Nava et al., 2021; Wang et al., 2020; Wasserman et al., 2022; Zhang et al., 2022). EDM is a versatile set of tools and techniques which can be applied to solve a wide range of problems, including improving

forecast accuracy, identifying nonlinear effects, identifying driving effects of population dynamics, and predicting how those effects may change under different scenarios. Though care need be applied to work around its limitations and modelling approaches match objectives, EDM's versatility and ability to account for nonlinear effects make it a very powerful tool for fisheries science.

In the cases of capelin and cod on the Newfoundland shelf, the analyses in this thesis could easily be extended to include more or different covariates. This is particularly true of my ecological covariates, which I limited to only two species representing two trophic levels in this study. Capelin is a keystone prey species which supports a variety of fish, seabird, and mammal populations on the Newfoundland shelf, that is dependent on phytoplankton and zooplankton-derived food sources, and competes for those food sources with other forage fish species and larval or juvenile fish of other species. Testing the effects of some or all of these species on capelin, and vice versa, is a potential area for future research. This logic also extends to the use of other environmental and/or socioeconomic covariates, such as fishing effort, to combinations of covariates for predictive modelling and scenario exploration, which I limited to four covariates and one covariate at a time respectively, and to perturbation values for scenario exploration. As more years of data are collected, more potential covariates and species become available for use by EDM, and times series of already available covariates and species become more informative.

In the end, this thesis shows that EDM is a potentially useful tool for Newfoundland shelf capelin stock assessment due to its ability to detect and account for nonlinear dynamics, reliably outperform linear models at predicting capelin biomass, and identify and clearly define both linear and nonlinear relationships between capelin, climate, and other species in the ecosystem.

Though EDM has some weaknesses, they are clearly defined, not unique, and can be mitigated through the use of advanced EDM methods, application of ecological principles and past scientific literature, and by using EDM as part of an ensemble approach alongside other stock assessment models and methodologies.

References

Arif, S., MacNeil, M.A., 2022. Predictive models aren't for causal inference. Ecology Letters 25, 1741–1745. https://doi.org/10.1111/ele.14033

Bartsev, S., Saltykov, M., Belolipetsky, P., Pianykh, A., 2021. Imperfection of the convergent cross-mapping method. IOP Conf. Ser.: Mater. Sci. Eng. 1047, 012081. https://doi.org/10.1088/1757-899X/1047/1/012081

Bogstad, B., Gjøsæter, H., 2001. Predation by cod (*Gadus morhua*) on capelin (*Mallotus villosus*) in the Barents Sea: implications for capelin stock assessment. Fisheries Research 53, 197–209. https://doi.org/10.1016/S0165-7836(00)00288-5

Bourne, C., Murphy, H., Adamack, A.T., Lewis, K., 2021. Assessment of capelin (*Mallotus villosus*) in 2J3KL to 2018. Can. Sci. Advis. Sec. Res. Doc. 2021/055, iv + 39p.

Bowering, W.R., Lilly, G.R., 1992. Greenland halibut (*Reinhardtius hippoglossoides*) off Southern Labrador and Northeastern Newfoundland (Northwest Atlantic) feed primarily on capelin (*Mallotus villosus*). Netherlands Journal of Sea Research 29, 211–222. https://doi.org/10.1016/0077-7579(92)90021-6

Brias, A., Munch, S.B., 2021. Ecosystem based multi-species management using Empirical Dynamic Programming. Ecological Modelling 441, 109423. https://doi.org/10.1016/j.ecolmodel.2020.109423

Buren, A.D., Koen-Alonso, M., Pepin, P., Mowbray, F., Nakashima, B., Stenson, G., Ollerhead, N., Montevecchi, W.A., 2014. Bottom-up regulation of capelin, a keystone forage species. PLOS ONE 9, e87589. https://doi.org/10.1371/journal.pone.0087589

Buren, A.D., Murphy, H.M., Adamack, A.T., Davoren, G.K., Koen-Alonso, M., Montevecchi, W.A., Mowbray, F.K., Pepin, P., Regular, P.M., Robert, D., Rose, G.A., Stenson, G.B., Varkey, D., 2019. The collapse and continued low productivity of a keystone forage fish species. Marine Ecology Progress Series 616, 155–170. https://doi.org/10.3354/meps12924

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

Chavez, F.P., Ryan, J., Lluch-Cota, S.E., Ñiquen C., M., 2003. From anchovies to sardines and back: Multidecadal change in the Pacific Ocean. Science 299, 217–221. https://doi.org/10.1126/science.1075880

Clark, A.T., Ye, H., Isbell, F., Deyle, E.R., Cowles, J., Tilman, G.D., Sugihara, G., 2015. Spatial convergent cross mapping to detect causal relationships from short time series. Ecology 96, 1174–1181. https://doi.org/10.1890/14-1479.1

Clark, T.J., Luis, A.D., 2020. Nonlinear population dynamics are ubiquitous in animals. Nat Ecol Evol 4, 75–81. https://doi.org/10.1038/s41559-019-1052-6

Cyr, F., Galbraith, P.S., 2021. A climate index for the Newfoundland and Labrador shelf. Earth Syst. Sci. Data 13, 1807–1828. https://doi.org/10.5194/essd-13-1807-2021

Dakos, V., Glaser, S.M., Hsieh, C., Sugihara, G., 2017. Elevated nonlinearity as an indicator of shifts in the dynamics of populations under stress. Journal of The Royal Society Interface 14, 20160845. https://doi.org/10.1098/rsif.2016.0845

Dawe, E., Koen-Alonso, M., Chabot, D., Stansbury, D., Mullowney, D., 2012. Trophic interactions between key predatory fishes and crustaceans: comparison of two Northwest

Atlantic systems during a period of ecosystem change. Mar. Ecol. Prog. Ser. 469, 233–248. https://doi.org/10.3354/meps10136
Deyle, E., Schueller, A.M., Ye, H., Pao, G.M., Sugihara, G., 2018. Ecosystem-based forecasts of recruitment in two menhaden species. Fish and Fisheries 19, 769–781. https://doi.org/10.1111/faf.12287
Deyle, E.R., Bouffard, D., Frossard, V., Schwefel, R., Melack, J., Sugihara, G., 2022. A hybrid empirical and parametric approach for managing ecosystem complexity: Water quality in Lake Geneva under nonstationary futures. Proceedings of the National Academy of Sciences 119, e2102466119. https://doi.org/10.1073/pnas.2102466119
Deyle, E.R., Fogarty, M., Hsieh, C., Kaufman, L., MacCall, A.D., Munch, S.B., Perretti, C.T., Ye, H., Sugihara, G., 2013. Predicting climate effects on Pacific sardine. Proceedings of the National Academy of Sciences 110, 6430–6435. https://doi.org/10.1073/pnas.1215506110
Deyle, E.R., Maher, M.C., Hernandez, R.D., Basu, S., Sugihara, G., 2016a. Global environmental drivers of influenza. Proceedings of the National Academy of Sciences 113, 13081–13086. https://doi.org/10.1073/pnas.1607747113
Deyle, E.R., May, R.M., Munch, S.B., Sugihara, G., 2016b. Tracking and forecasting ecosystem interactions in real time. Proc. R. Soc. B. 283, 20152258. https://doi.org/10.1098/rspb.2015.2258
DFO, 2019. 2019 Status of Northwest Atlantic Harp Seals, *Pagophilus groenlandicus*. DFO Can Sci Advis Sec Sci Advis Rep 2020/020, 14.
DFO, F. and O.C., 2021. 2020 Stock Status Update for Northern Cod. DFO Can Sci Advis Sec Sci Advis Rep 2021/004, 18.
DFO, F. and O.C., 2012. Results and recommendations from the Ecosystem Research Initiative - Newfoundland and Labrador's Expanded Research on Ecosystem Relevant but Under-Surveyed Splicers. DFO Can Sci Advis Sec Sci Advis Rep 2012/058, 15.
Doubleday, W.G., 1981. Manual on Groundfish Surveys in the Northwest Atlantic. NAFO Sci. Coun. Studies 2, 7–55.
Dwyer, K.S., Buren, A., Koen-Alonso, M., 2010. Greenland halibut diet in the Northwest Atlantic from 1978 to 2003 as an indicator of ecosystem change. Journal of Sea Research, Proceedings of the Seventh International Symposium on Flatfish Ecology, Part II 64, 436–445. https://doi.org/10.1016/j.seares.2010.04.006
Ebisuzaki, W., 1997. A Method to Estimate the Statistical Significance of a Correlation When the Data Are Serially Correlated. Journal of Climate 10, 2147–2153. https://doi.org/10.1175/1520-0442(1997)010<2147:AMTETS>2.0.CO;2
Giron-Nava, A., Ezcurra, E., Brias, A., Velarde, E., Deyle, E., Cisneros-Montemayor, A.M., Munch, S.B., Sugihara, G., Aburto-Oropeza, O., 2021. Environmental variability and fishing effects on the Pacific sardine fisheries in the Gulf of California. Can. J. Fish. Aquat. Sci. 78, 623–630. https://doi.org/10.1139/cjfas-2020-0010
Gjøsæter, H., 2002. Assessment methodology for Barents Sea capelin, *Mallotus villosus* (Müller). ICES Journal of Marine Science 59, 1086–1095. https://doi.org/10.1006/jmsc.2002.1238
Gjøsæter, H., Bogstad, B., Tjelmeland, S., Subbey, S., 2015. A retrospective evaluation of the Barents Sea capelin management advice. Marine Biology Research 11, 135–143. https://doi.org/10.1080/17451000.2014.928414

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

Glaser, S.M., Ye, H., Sugihara, G., 2014b. A nonlinear, low data requirement model for producing spatially explicit fishery forecasts. Fisheries Oceanography 23, 45–53. https://doi.org/10.1111/fog.12042

Gomes, M.C., Haedrich, R.L., Villagarcia, M.G., 1995. Spatial and temporal changes in the groundfish assemblages on the north-east Newfoundland/Labrador Shelf, north-west Atlantic, 1978–1991. Fisheries Oceanography 4, 85–101. https://doi.org/10.1111/j.1365-2419.1995.tb00065.x

Hsieh, C., Glaser, S.M., Lucas, A.J., Sugihara, G., 2005. Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean 435, 5.

ICES, 2021. Arctic Fisheries Working Group (AFWG). ICES Scientific Reports 3, 817. https://doi.org/10.17895/ices.pub.8196

Johnson, B., Munch, S.B., 2022. An empirical dynamic modeling framework for missing or irregular samples. Ecological Modelling 468, 109948. https://doi.org/10.1016/j.ecolmodel.2022.109948

Kuriyama, P.T., Sugihara, G., Thompson, A.R., Semmens, B.X., 2020. Identification of shared spatial dynamics in temperature, salinity, and ichthyoplankton community diversity in the California Current system with Empirical Dynamic Modeling. Frontiers in Marine Science 7. https://doi.org/10.3389/fmars.2020.557940

Lewis, K.P., Buren, A.D., Regular, P.M., Mowbray, F.K., Murphy, H.M., 2019. Forecasting capelin *Mallotus villosus* biomass on the Newfoundland shelf. Marine Ecology Progress Series 616, 171–183. https://doi.org/10.3354/meps12930

Lilly, G.R., Parsons, D.G., Kulka, D.W., 2000. Was the increase in shrimp biomass on the Northeast Newfoundland Shelf a consequence of a release in predation pressure from cod? J. Northw. Atl. Fish. Sci. 27, 45–61. https://doi.org/10.2960/J.v27.a5

Ma, H., Aihara, K., Chen, L., 2014. Detecting causality from nonlinear dynamics with short-term time series. Sci Rep 4, 7464. https://doi.org/10.1038/srep07464

McCracken, J.M., Weigel, R.S., 2014. Convergent cross-mapping and pairwise asymmetric inference. Phys. Rev. E 90, 062903. https://doi.org/10.1103/PhysRevE.90.062903

Mønster, D., Fusaroli, R., Tylén, K., Roepstorff, A., Sherson, J.F., 2017. Causal inference from noisy time-series data — Testing the Convergent Cross-Mapping algorithm in the presence of noise and external influence. Future Generation Computer Systems 73, 52–62. https://doi.org/10.1016/j.future.2016.12.009

Mowbray, F.K., 2012. Some results from spring acoustic surveys for capelin (*Mallotus villosus*) in NAFO Division 3L between 1982 and 2010. DFO Can Sci Advis Sec Sci Advis Rep 2012/143, 36.

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

Munch, S.B., Rogers, T.L., Sugihara, G., 2023. Recent developments in empirical dynamic modelling. Methods in Ecology and Evolution 14, 732–745. https://doi.org/10.1111/2041-210X.13983

Murphy, H.M., Adamack, A.T., Cyr, F., 2021. Identifying possible drivers of the abrupt and persistent delay in capelin spawning timing following the 1991 stock collapse in

Newfoundland, Canada. ICES Journal of Marine Science 78, 2709–2723. https://doi.org/10.1093/icesjms/fsab144
NAFO, 2021. STATLANT 21A Database.
NAFO, 2010. Report of the NAFO Scientific Council Working Group on Ecosystem Approaches to Fisheries Management (WGEAFM),. NAFO Scientific Council Summary Document 10/10, 1–101.
Olsen, E.M., Heino, M., Lilly, G.R., Morgan, M.J., Brattey, J., Ernande, B., Dieckmann, U., 2004. Maturation trends indicative of rapid evolution preceded the collapse of northern cod. Nature 428, 932–935. https://doi.org/10.1038/nature02430
Rose, G., 2002. Capelin are good for cod: can the northern stock rebuild without them? ICES Journal of Marine Science 59, 1018–1026. https://doi.org/10.1006/jmsc.2002.1252
Rose, G.A., 2005. Capelin (*Mallotus villosus*) distribution and climate: a sea "canary" for marine ecosystem change. ICES Journal of Marine Science 62, 1524–1530. https://doi.org/10.1016/j.icesjms.2005.05.008
Schwartzlose, R.A., Alheit, J., Bakun, A., Baumgartner, T.R., Cloete, R., Crawford, R.J.M., Fletcher, W.J., Green-Ruiz, Y., Hagen, E., Kawasaki, T., Lluch-Belda, D., Lluch-Cota, S.E., MacCall, A.D., Matsuura, Y., Nevárez-Martínez, M.O., Parrish, R.H., Roy, C., Serra, R., Shust, K.V., Ward, M.N., Zuzunaga, J.Z., 1999. Worldwide large-scale fluctuations of sardine and anchovy populations. South African Journal of Marine Science 21, 289–347. https://doi.org/10.2989/025776199784125962
Sguotti, C., Otto, S.A., Cormon, X., Werner, K.M., Deyle, E., Sugihara, G., Möllmann, C., 2020. Non-linearity in stock–recruitment relationships of Atlantic cod: insights from a multi-model approach. ICES Journal of Marine Science 77, 1492–1502. https://doi.org/10.1093/icesjms/fsz113
Skagseth, Ø., Slotte, A., Stenevik, E.K., Nash, R.D.M., 2015. Characteristics of the Norwegian Coastal Current during years with high recruitment of Norwegian spring spawning herring (*Clupea harengus L.*). PLoS ONE 10, e0144117. https://doi.org/10.1371/journal.pone.0144117
Smith, S.J., Somerton, G.D., 1981. STRAP: a user-oriented computer analysis system for groundfish research trawl survey data. Canadian Technical Report of Fisheries and Aquatic Sciences No. 1030.
Sugihara, G., 1994. Nonlinear forecasting for the classification of natural time series. Philosophical Transactions: Physical Sciences and Engineering 348, 477–495.
Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., Munch, S., 2012. Detecting causality in complex ecosystems. Science 338, 496–500.
Sugihara, G., May, R.M., 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature 344, 734–741. https://doi.org/10.1038/344734a0
Takens, F., 1981. Detecting strange attractors in turbulence, in: Rand DA, Young LS (Eds) Dynamic Systems and Turbulance. Springer, New York, pp. 366–381.
Toresen, R., Østvedt, O. j., 2000. Variation in abundance of Norwegian spring-spawning herring (*Clupea harengus*, Clupeidae) throughout the 20th century and the influence of climatic fluctuations. Fish and Fisheries 1, 231–256. https://doi.org/10.1111/j.1467-2979.2000.00022.x
Ushio, M., Hsieh, C., Masuda, R., Deyle, E.R., Ye, H., Chang, C.-W., Sugihara, G., Kondoh, M., 2018. Fluctuating interaction network and time-varying stability of a natural fish community. Nature 554, 360–363. https://doi.org/10.1038/nature25504

Wang, J.-Y., Kuo, T.-C., Hsieh, C., 2020. Causal effects of population dynamics and environmental changes on spatial variability of marine fishes. Nat Commun 11, 2635. https://doi.org/10.1038/s41467-020-16456-6

Wasserman, B.A., Rogers, T.L., Munch, S.B., Palkovacs, E.P., 2022. Applying empirical dynamic modeling to distinguish abiotic and biotic drivers of population fluctuations in sympatric fishes. Limnology and Oceanography 67, S403–S415. https://doi.org/10.1002/lno.12042

Worm, B., Myers, R.A., 2003. Meta-analysis of cod–shrimp interactions reveals top-down control in oceanic food webs. Ecology 84, 162–173. https://doi.org/10.1890/0012-9658(2003)084[0162:MAOCSI]2.0.CO;2

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

Ye, H., Deyle, E.R., Gilarranz, L.J., Sugihara, G., 2015b. Distinguishing time-delayed causal interactions using convergent cross mapping. Sci Rep 5, 14750. https://doi.org/10.1038/srep14750

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

Ying, X., Leng, S.-Y., Ma, H.-F., Nie, Q., Lai, Y.-C., Lin, W., 2022. Continuity Scaling: A Rigorous Framework for Detecting and Quantifying Causality Accurately. Research 2022, 1–10. https://doi.org/10.34133/2022/9870149

Zhang, Z., Yang, W., Ding, J., Sun, T., Liu, H., Liu, C., 2022. Identifying changes in China's Bohai and Yellow Sea fisheries resources using a causality-based indicator framework, convergent cross-mapping, and structural equation modeling. Environmental and Sustainability Indicators 14, 100171. https://doi.org/10.1016/j.indic.2022.100171