Multispecies methods to facilitate the transition from heuristics to statistics in ecosystem-based fisheries management

Kelli Faye Johnson

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

André E. Punt, Chair

Isaac C. Kaplan

Timothy E. Essington

Program Authorized to Offer Degree:

School of Aquatic and Fishery Sciences

ProQuest Number: 10841823

ProQuest 10841823

University of Washington

Abstract

Multispecies methods to facilitate the transition from heuristics to statistics in ecosystem-based fisheries management

Kelli Faye Johnson

Chair of the Supervisory Committee:
Professor André E. Punt
School of Aquatic and Fishery Sciences

Since 1985, fisheries management has undergone vast changes. Requests for new policies that evaluate impacts beyond individual removals of fish have outpaced the scientific tools available to implement them. Tools that were or became available were often implemented without rigorous testing. Here, simulations were used to assess the ability of two multivariate time-series models to provide information relevant to ecosystem-based fisheries management. Simulations highlighted the need to account for time-series properties of data. Multivariate autoregressive state-space models successfully estimated interactions between two time series when observation error was small and the length of the time series was sufficient. Prewhitening procedures were also successful in estimating interactions. The framework proved useful for estimating synchrony between indicators of ecosystem status and the true state of the ecosystem. Spatiotemporal multivariate models successfully recovered trends in the data. Models were robust to model misspecification and estimates of covariates related to linear relationships between habitat and relative abundance were unbiased. In contrast, squared terms were biased, particularly for the catch-rate component of the model. Incorrectly including a habitat covariate when it did not

govern the data-generating process was less problematic than not including the covariate when it should have been included. Results from simulations guided the selection of quantitative tools used to provide inference on management uncertainty related to gear switching in the US West Coast Groundfish LE trawl fishery. Fishers in this fishery were allowed to switch from using trawl-gear to fixed-gear to land sablefish in 2011, and it was expected that bycatch of species with restrictive quotas would influence their decision to switch. However, bycatch species were not a major factor in their decision. Instead, unmodeled factors related to the major port groups included in the analysis were the best predictor of the proportion of landings that were caught using fixed-gear compared to trawl-gear. It was hypothesized that social and logistical challenges specific to each port group could limit their ability or desire to switch gear. Additionally, vessels could be limited to using trawl-gear because ports may not have adequate facilities to process catch from fixed-gear.

TABLE OF CONTENTS

i

LIST OF FIGURES

vi

LIST OF TABLES

ACKNOWLEDGEMENTS

I was extremely grateful for being accepted into the School of Aquatic and Fishery Sciences at the University of Washington in 2012, but I had no idea how much more gratitude I would feel because of the assistance of so many individuals that helped in the development of my skills since then. Most of all, I will be forever thankful for all that Dr. André Punt, my major advisor, taught me and how he facilitated the exploration of projects that I did not know were even possible prior to my Dissertation. He was **always** available, he was beyond patient when I peppered him with trivial questions, and he took the time to introduce me to his colleagues that I would have otherwise never had the courage to talk to. Throughout my studies he made me feel like an equal and that I deserved to be his student.

My ability to contribute to science beyond that of fisheries management was greatly expanded through the mentorship of Dr. Isaac Kaplan, my National Oceanic and Atmospheric Administration mentor and committee member. He **always** took time to explain many bigger picture items to me and wrote notes in the margins of draft manuscripts that were positive as well as constructive. He was the most realistic member of my committee with respect to what was attainable within the given time frame and how much should be expected of graduate students. His encouragement and positive outlook on how much work I had done gave me the courage to speak up when I thought that I had done enough and to do a little bit more when I was exhausted.

The additional members of my committee, Dr. Trevor Branch, Dr. Timothy Essington, and Dr. Kerim Aydin, helped me see new perspectives and provided critical insight into fisheries management that would have taken me decades to learn on my own. Their comments on draft manuscripts were extensive, and learning how to incorporate their sometimes-differing suggestions improved my cognitive abilities more than any class offered in my program.

Many others involved in the field of fisheries also positively contributed to my scientific endeavors. Fellow students at the University of Washington, in particular the past and current members of the Punt, Branch, Hilborn, and Essington labs, were terrific learning partners and encouraged me to be the best scientist possible. Members of the Northwest Fisheries Science Center Fisheries Resource Analysis and Monitoring Division were kind enough to collaborate with me during my early days and hire me during my last year. The hands-on experience of working

with "real" data was instrumental in developing skills necessary to provide meaningful science. Participants of the International Whaling Commission's Scientific Committee were encouraging yet ruthless and quickly facilitated the development of my grammar and scientific skills. The analyses presented in this Dissertation would not have been possible without the work of B. Horness, K. Somers, E. Steiner, and A. Warlick who provided data and C. Ainsworth, E.A. Fulton, C. Hansen T. Ihde, J. Link, H. Morzaria Luna, K.N. Marshall, H. Townsend, and M. Weijerman who provided output from Atlantis models. O.S. Hamel, K.N. Marshall, J.F. Samhouri, M.D. Scheuerell, A.O. Shelton, J.T. Thorson, and E.J. Ward provided helpful feedback that improved the text and the analyses. H. Morzaria Luna assisted with the map of the Atlantis ecosystems.

My friends and family were tremendously supportive throughout these last six years. They provided constant encouragement, even when they didn't know it. Without them I would not have had the courage to try to become a "doctor" and for sure would not have been able to see the glass as half full. With thier support and love, my glass is all the way full. My parents instilled a work ethic deep inside my soul that I will never again take for granted. My experiences from years on the farm taking care of things that were not going to just magically happen, day or night, holiday or not, consistently gives me the courage to try anything I set my mind to.

My husband, Tim, endured the most pain throughout this process. He is a **wonderful** human on every level. He taught me how to see the best in things, even when I really did not want to. His level of patience with all things is a quality I strive to achieve. I will never be able to thank him enough for all that he has given me.

DEDICATION

to

those who believed in me

x

INTRODUCTION

Fisheries and aquaculture are important sources of food, nutrition, income, and livelihood for hundreds of millions of people across the globe. Harvests from the world's fisheries have steadily increased since the end of the Second World War (FAO, 2018). However, landings from distant-water fleets have declined since the late 1980s. In 1989, 26% of the world's fish stocks were estimated to be fished at biologically unsustainable levels (FAO, 2016). This overexploitation threatens biodiversity (Boehlert, 1996; Parsons, 1991) and the ability of the fishing industry to contribute to human nutrition, food security, and economic growth (Béné *et al*., 2016). Concerns about overexploitation, overcapacity, and illegal unregulated and unreported fishing led to calls for fundamental changes in fisheries management. These concerns are still of interest today with 31.4% of the world's fish stocks estimated to be overfished as of 2013.

Declines in the landings of wild-caught fisheries, paralleled with increasing concerns about the sustainability of fisheries, led to an emphasis on fisheries management in national policies and legislation (e.g., UN, 1992). By this time, fisheries management was already far more holistic than it was at the turn of the 20th century when the potential for the depletion of fish populations was debated (Huxley, 1883). The 1990s legislation renewed the impetus for the precautionary approach to management and the need for the participation of all concerned citizens. The precautionary approach emphasizes the need to err on the side of caution. Specifically, when information is sparse or uncertain this uncertainty should help inform decision making and lead to more stringent management (UN, 1992). This call for the precautionary approach was followed by the collapse of the Canadian Atlantic northern cod fishery despite it being managed under a framework that set harvest levels at those predicted to be sustainable (Myers *et al*., 1997). The collapse and subsequent slow recovery of the cod fishery highlighted the need to translate international policies focused on the precautionary approach to the national and local level.

These international policies included broad objectives such as to protect rare or fragile ecosystems and to protect and maintain relationships and dependencies among species (WCED, 1987; UN, 1992). Fisheries managers were subsequently tasked with explaining the consequences of management actions with respect to these broad objectives. Consequences, in terms of the perceived risk of failing to meet an objective, were best described using trade-offs because all fishing activities have at least some environmental impact. The precautionary approach applied to single-species fisheries management focused on maintaining the production of target species. A

logical extension of this has been the development of ecosystem-based fishery management (EBFM), which expands beyond targeted single species to recognize risks to non-target species (e.g., Walters *et al*., 2005) and the greater social-ecological system in which the species reside (Ostrom, 2009).

Ecosystem-based fishery management was thus coined as a holistic approach to maintaining ecosystem integrity and the benefits derived from such ecosystems (Larkin, 1996). The US National Oceanic and Atmospheric Administration and the Food and Agriculture Organization define EBFM as

"an approach that takes major ecosystem components and services (both structural and functional) into account in managing fisheries. It values habitat, embraces a multispecies perspective, and is committed to understanding ecosystem processes. Its goal is to rebuild and sustain populations, species, biological communities, and marine ecosystems at high levels of productivity and biological diversity so as not to jeopardize a wide range of goods and services from marine ecosystems while providing food, revenue, and recreation for humans" (NRC, 1998).

The same problems that faced traditional fisheries management (i.e., overfishing, overcapacity, and illegal unregulated and unreported fishing) are also problems in EBFM. Nevertheless, EBFM proposes to mitigate these problems in a more holistic manner. Consideration of the broader ecosystem is seen as one potential solution to sustainable fisheries (Pikitch *et al*., 2004).

Operationally, EBFM starts with the priorities of the ecosystem rather than target species. For example, total allowable catches (TACs) in the Bering Sea and Aleutian Islands groundfish fishery are first capped at 2 million mt (well below the sum of the individual TACs) and individual TACs estimated from single-species stock assessment models are subtracted from this cap (NPFMC, 2017). Some TACs may be set at values lower than what was estimated as corresponding to fishing at the maximum sustainable yield (or its proxy) to ensure the sum of all catches is below the cap. It can be argued that using a multispecies stock assessment model to simultaneously estimate TACs would be a better way forward rather than summing TACs based on single-species stock assessments (Burgess *et al*., 2017). However, multispecies stock assessment models are still works in progress and have their own set of associated biases (Punt and Hilborn, 1994; Hollowed *et al*., 2000; Plagányi *et al*., 2012). Off the US West Coast, marine protected areas with allowances for some catches were implemented by the Pacific Fisheries Management Council (PFMC) to limit bycatch of overfished species and to conserve essential fish habitat (66 FR 2338). These protected areas were seen as a way to keep the groundfish fishery open. Albeit at the potential expense of

not being able to fully achieve TACs of stocks that were not overfished, highlighting Gunderson and Holling's (2002) take that single-species approaches are not wrong, just incomplete.

Some goals of EBFM are easily defined, present few controversial trade-offs, and are achievable without the use of complex models. For example, reducing the bycatch of non-target species can sometimes be realized through the adoption of a relatively simple gear modification (e.g., Cooke *et al*., 2012). Goals can also be complex, controversial, fraught with trade-offs, and may threaten the ability of marine ecosystems to provide goods and services if left unaddressed. For example, forage fish support marine predators and contribute billions of dollars to the global economy as catches (Pikitch *et al*., 2014). Addressing complex goals in the best-case scenario may mean the use of highly parameterized models but available data are not always sufficient to do so. Nevertheless, the lack of "full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation" (principle 15 of UN, 1992).

Indicators of ecosystem status offer a means to evaluate changes in the ecosystem when there is insufficient knowledge or data to directly calculate ecosystem properties of interest (i.e., attributes). Changes in indicators can be compared to changes in fisheries management as a way to assess the implications of the current regime. Indicators such as maximum size-at-age can track changes in age- and size-composition and provide a useful way to assess how ecosystems respond to fishing pressure (Probst *et al*., 2013; Rice, 2000). Catches and the mean trophic level (TL) of the catch provide information on potential changes in predator-prey relationships (Christensen, 1996). Thus, time-series data on species that reside within an ecosystem can be informative about the effects of fishing and the status of biodiversity and ecosystem function without needing to conduct a stock assessment for each species (Greenstreet *et al*., 2012; Shannon *et al*., 2009). Furthermore, indicators that result from single-species stock assessments may not be informative about unwanted consequences to the ecosystem.

The number of potential indicators for EBFM seems to be ever-growing. Rice and Rochet (2005) outline a framework for choosing indicators. Reliable ecological indicators must possess a number of desirable properties, some of which are not mutually exclusive and lead to trade-offs (Rice and Rochet, 2005). For example, ecological indicators should minimize false positives (i.e., incorrectly identifying that an ecosystem has crossed the reference point) and false negatives (i.e., failing to identify that an ecosystem has crossed the reference point). Additionally, indicators should be robust across time and space (Noss, 1990) and drivers of change (Link *et al*., 2010).

Survey data are informative about community-level indicators (Jennings and Dulvy, 2005), such as size classes, size spectra, average size, species richness, and species evenness. These indicators are often noted as being increasingly robust compared to ecological indicators that aggregate coarsely over many taxa (e.g., total biomass; Methratta and Link, 2006) or single-species traits (Cottingham and Carpenter, 1998; Fulton *et al*., 2005). For instance, zooplankton biomass is known to exhibit little changes in biomass with increased ocean acidification (Schindler *et al*., 1985). Unfortunately, some community-level indicators can only be applied when species-specific composition data are available. Globally, age- or size-composition data are less likely to be collected than biomass or presence/absence data, thus limiting their use when looking beyond the first order effects of fishing (Jouffre *et al*., 2010).

Efficient indicators should be robust to issues of scale, temporal and spatial, and ecosystem characteristics. Minimally, it should be known which factors cause an indicator to deviate from the attribute it is hypothesized to track. Unfortunately, the majority of research on indicators relates to reporting empirical trends (e.g., Coll *et al*., 2016) and methods for determining reference points (e.g., Trenkel and Rochet, 2003) rather than indicator robustness (for exceptions see Fulton *et al*., 2010; Shin *et al*., 2010a). Fulton *et al*. (2005) paved the way for evaluating indicator robustness using a simulation-based approach. Correlations between indicators and attributes were used to assess indicator robustness across a range of fishing patterns and nutrient loads in two simulated ecosystems. Evaluating trade-offs without having to measure empirical data offers an efficient means to eliminate proposed indicators that are unlikely to be robust because it is unlikely that an indicator will perform well when applied to real data if it does not perform well in a simulated and simplified ecosystem. Furthermore, simulations offer a way to test indicator robustness under several management regimes (Samhouri *et al*., 2009) or levels of observation error, currently a less studied topic with respect to indicator robustness.

Data on indicators are typically time series or spatial replicates. However, their treatment as such is limited. For example, Fulton *et al.* (2005) ignored the time-series properties of the data on indicators and attributes and used correlation to assess relationships. Multivariate autoregressive state-space (MARSS) models offer one way to detect spatially- and temporally-explicit patterns of synchrony (Hinrichsen and Holmes, 2009). Their use in ecology began with time-series data for single populations (Dennis *et al*., 2006; Millar and Meyer, 2000; Staples *et al*., 2004) and later extended into two-dimensional space for movement data (Jonsen *et al*., 2003) and multiple

subpopulations (Ward *et al.*, 2010) and species (Mattsson *et al.*, 2013). The models inherently account for autocorrelation and can estimate both process and observation error in a single framework. Their simpler counterparts, multivariate autoregressive (MAR) models, have been used to estimate patterns of temporal correlation in abundances from species interactions (Ives, 1995; Ives *et al.*, 1999, 2003). Their utility in measuring interaction strengths between two time series was compared to correlation coefficients using simulation in Chapter 1.

Models that define simulation analyses can be as simple as a two linear equations like those used in Chapter 1. Conversely, they can be extremely complex. The end-to-end ecosystem model known as Atlantis can include human dynamics in addition to the marine ecosystem (e.g., van Putten *et al.*, 2013). Chapter 2 built upon the results of Fulton *et al.* (2005) using methods explored in Chapter 1 to provide inference on the robustness of indicators. Robustness was measured across ecosystems and drivers of change.

Simulated output used as data in Chapter 2 included information on relative abundance from fisheries-independent and -dependent sources. Both sources of data provide information in the form of catch rates that by themselves are seldom proportional to abundance and need to be standardized (Maunder and Punt, 2004). Standardization can account for factors such as differences among vessels in fishing power (Helser *et al.*, 2004; Robins *et al.*, 1998), the depth or location at which fish are caught (Bigelow and Maunder, 2007; Cao *et al.*, 2017), and sampling intensity (Cochran, 1977). The resulting standardized index of abundance can be used directly by management as an indicator or as input when fitting a stock assessment model. Recently, the use of spatiotemporal models for index standardization has increased because of their ability to account for spatial heterogeneity and improve precision relative to design-based estimators (Thorson *et al.*, 2015c). In theory, the inclusion of habitat covariates (e.g., depth) should also improve precision. Chapter 3 used simulation to assess the utility of including habitat covariates in a spatiotemporal index-standardization model that is currently used to provide relative indices of abundance for many US West Coast groundfish species managed by the PFMC. Determining if habitat covariates are informative could become increasingly relevant as species ranges are predicted to shift with respect to depth and latitude as oceans warm in response to climate change (Dulvy *et al.*, 2008; Nye *et al.*, 2009; Perry *et al.*, 2005).

Simulations can also be used in a dynamic sense to test the effectiveness of management strategies in meeting objectives (Punt *et al.*, 2016). A closed-loop simulation that includes each

step going forward in time can be used to identify management strategies that are more likely to lead to the achievement of objectives. The simulations decompose management strategies into three parts, the data used to inform ecosystem status, the analysis to determine the current status, and to implement actions based on current status. A key source of uncertainty in these closed-loop simulations is that of the management actions themselves, also known as management uncertainty. Management uncertainty is a pervasive problem in fisheries management, entering at multiple stages of the management cycle, working to undermine effective policies.

There are many examples where the outcomes of management actions are different than those that were hypothesized. For example, entry and exit behavior should be predictable given economic theory. In Iceland, fishers often fail to exit a fishery when it is economically rational to do so because of fear of low catches in a new fishery, where low catches are considered humiliating (Pálsson and Durrenberger, 1982). Choices of fishers can be guided by information on environmental factors, risk tolerance, personal experience, economic expectations, management constraints, etc. (Steelman and Wallace, 2001; van Putten *et al.*, 2012). Furthermore, choices are typically made in relation to multiple objectives, involve factors with varying levels of uncertainty, and may depend on the actions of other fishers (Allen and McGlade, 1987). Therefore, fisher behavior will be driven by much more than just economic objectives (e.g., profit maximization), and policies that fail to account for complexities arising from socioeconomic and cultural contexts may fail to reach objectives (Branch *et al.*, 2005; Mahon *et al.*, 2008).

Decades of research exists on the human dimensions of fisheries. Much of this research assumes fishers' reactions to management are driven by economic effects alone, and the knowledge gained from this research is almost never explicitly included in scientific advice to managers (Fulton *et al.*, 2010). Removing all sources of management uncertainty will never be possible. Nevertheless, quantitative methods that determine the drivers of fisher behavior, beyond economics, are needed. Chapter 4 used empirical data collected from the US West Coast Limited Entry Trawl Fishery to determine drivers of gear choice when landing sablefish (*Anoplopoma fimbria*). Results will be useful for parameterizing future closed-loop simulations to assess the effectiveness of EBFM strategies implemented by the Pacific Fisheries Management Council.

This dissertation highlights the growing recognition that EBFM relies on quantitative tools to provide strategic and tactical advice, involves multiple disciplines, and does not need to wait for the collection of more data. Assessment of whether a fishery is sustainable has evolved over the

last two decades, with an increasing focus on accounting for the ecological impacts of fishing. Unfortunately, science to inform assessments lags mandates. The main objective of this research is to increase the number of quantitative methods available to fisheries scientists engaged in EBFM. The work proposed here will facilitate inclusion of impacts beyond the direct removal of fish when assessing the sustainability of fisheries. Quantitative tools, such as those proposed here, facilitate the evaluation of trade-offs by providing reproducible results, being applicable to historical data, and providing measures of uncertainty.

Chapter 1. JUST SAY NO TO CORRELATION FOR TIME-SERIES DATA AND INSTEAD USE METHODS THAT ACCOUNT FOR AUTOCORRELATION

Abstract

Detecting ecological interactions (e.g., density-dependence) is an increasingly important task of ecologists aiming to increase their general understanding of intra- and interspecific interactions and how they are affected by the abiotic environment. Three methods (cross correlation, prewhitened cross correlation, and multivariate autoregressive state-space models) were fit to simulated data to assess which method best identified "true" interactions between time series. Simulated data included cases where the time series were positively or negatively correlated with themselves and with each other. Multivariate autoregressive state-space models offered an unbiased integrated approach to estimate interactions and observation error when the data were informative. Prewhitened cross correlation was more robust to time-series length, observation error, and interaction strength than multivariate autoregressive state-space models when the data were less informative. Prewhitened cross correlation provided unbiased estimates of interactions that corresponded to correlations in process errors. Correlations in the process errors can be thought of as interactions due to factors not included in the model. Consequently, the choice of an appropriate model depends on the interaction of interest and the information contained in the data. Simple correlation should not be used when the time series are thought to be autocorrelated.

1.1 INTRODUCTION

Understanding interactions between organisms and their environment and among groupings of organisms is important to ecologists and critical to the management of natural resources (Allen and Hoekstra 2015). Classic examples include interactions between lynx (*Lynx canadensis*) and snowshoe hare (*Lepus americanus*) in boreal forests of North America (Figure 1.1; Elton and Nicholson, 1942), the number of young individuals entering a population and mature biomass (e.g., Hjort, 1914), and the number of young individuals entering a population and environmental conditions (Szuwalski *et al.*, 2015). Correlation coefficients, which measure the extent to which

two variables fluctuate together, offer a means to quantify such interactions (Galton 1888). However, most ecological data are autocorrelated (e.g., Beamish, 1995; Hollowed *et al.*, 2001) and thus lack the statistical independence needed to infer the significance of correlations (Yule, 1926).

The usefulness of techniques proposed to reduce the probability of statistical errors when calculating correlation coefficients in the presence of autocorrelation depend on the sign and magnitude of the autocorrelation (Pyper and Peterman, 1998). First differencing has long been known to reduce the presence of Type-1 errors in econometrics (Granger and Newbold, 1974). However, first differencing can increase the probability of Type-2 errors when the driving force of the interaction is strongly autocorrelated (Pyper and Peterman, 1998). First differencing can also lead to the estimation of negative interactions when there are none (Buonaccorsi *et al.*, 2001). Time series models such as a first-order autoregressive models, AR(1), can be used to "prewhiten" the time series (Box *et al.*, 2015). The usefulness of prewhitening may depend on how well the prewhitening framework matches the true underlying process. The degrees of freedom or the variance of the correlation coefficient can be corrected using an estimate of the autocorrelation (Bartlett, 1946; Bayley and Hammersley, 1946). Unfortunately, autocorrelation is notoriously difficult to estimate when the time series includes fewer than 20 measurements (Thompson and Page, 1989) or the autocorrelation is weak (Ryding *et al.*, 2007).

Biased standard errors can also result when the time series are intercorrelated in multiple ways. Multicollinearity can lead to unstable correlation coefficients that are difficult to interpret. In ecology, multicollinearity has a history of being ignored (Graham, 2003) even though ecological responses are often the result of more than one interaction (Borcard *et al.*, 1992). For example, lynx visually appear to be correlated with hare at a negative lag (panel a in Figure 1.1) and lynx and hare appear to be correlated with themselves (panel b in Figure 1.1). The lag -1 correlation (solid line in panel c of Figure 1.1) was statistically significant after accounting for autocorrelation (i.e., prewhitening). Unfortunately, cross-correlation coefficients are estimated independently and thus the effect of other lags are not taken into account. Mutually experienced factors such as weather may be affecting the time series simultaneously, and prey availability could be contributing to lagged effects.

Many methods are available for estimating interactions besides correlation, but the benefits of their additional complexity are not always clear in an ecological context. Monte Carlo

simulations were used to quantify and compare the ability of cross correlation, prewhitened cross correlation, and multivariate autoregressive state-space models (MARSS: Holmes *et al.*, 2012) to estimate relationships between two time series affected by potentially more than one type of interaction. Simulations, for which the truth is known, allowed for differences among the results from each estimation method to be interpreted in terms of how time-series length, autocorrelation, direct interactions, mutual or co-interactions, and observation error affected each investigated method. Previous work has investigated some of these given a single estimation framework (e.g., Pyper and Peterman (1998) estimated cross correlations across a range of co-interactions), but the results presented here represent the first effort to compare the performance of the three estimation methods across a wide range of interaction types. Results provide clear examples of the dangers of not prewhitening and guidance regarding when using prewhitening or MARSS models is advantageous. The principal focus of this manuscript is on assessing interactions in the presence of temporal autocorrelation, but the same statistical issues apply to spatial autocorrelation (Moran, 1950).

1.2   METHODS

There are many types of ecological interactions, but for ease of understanding we limit their classification to just three types. The first, which we call direct-interactions, occur when time series change in a scaled, predictable way because the value of one time-series directly depends on the other. Changes in the independent time series could in theory be used to predict the dependent time series in direct-interactions. The second, co-interactions, occur when time series rise and fall together without a direct causative link. Lastly, time series can exhibit relationships with themselves, self-interactions, such as would be expected for time series of population size. Autocorrelation is typical of large-bodied species that experience less variability than smaller-bodied species (Pimm, 1991). Hypothetical data were simulated for combinations of interaction types and cross correlation, prewhitened cross correlation, and MARSS models were fit to simulated data to determine the circumstances in which each estimation methods adequately characterized the true dynamics.

1.2.1   *Data generation*

1.2.2 *Data generation*

Simulated data included two time series (x_1 and x_2) that potentially interacted through time (t),

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_t = \begin{bmatrix} b_{1,1} & 0 \\ b_{2,1} & b_{2,2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_{t-1} + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}_t \qquad (1.1).$$

Direct-interactions ($b_{2,1}$) were limited to one-way interactions with a lag of negative one and a strength ranging from -0.9 to 0.9. Self-interaction strengths were equal ($b_{1,1}=b_{2,2}$) and ranged from -0.9 to 0.9. Co-interactions (q) were implemented using multivariate normal process error ($\mathbf{w}_t$) with a mean of zero and covariance matrix $\mathbf{Q}$. The variance of the process errors ($q_{1,1} = q_{2,2}$) were equal and set to 1.0 (Ives *et al.*, 2003). The off-diagonal of $\mathbf{Q}$ ranged from -0.9 to 0.9 and was back-calculated from the correlation in $\mathbf{Q}$ ($q = q_{2,1}/(\sqrt{q_{1,1}}\sqrt{q_{2,2}})$. The model can be thought of as a multivariate Gompertz model if the time series are log-transformed abundances, where per-capita growth is a log-linear function of abundance (Ives *et al.*, 2003). Process error variance was fixed at 1.0 to ensure that there was sufficient information in the errors for estimation purposes, matching that used in Ives *et al.* (2003). Normally distributed observation error with a variance of 1.0 was added to some simulations to increase the applicability of the results to ecologists because ecological processes are typically measured imperfectly (Figure A.1). Data were simulated for 100 time steps, which may be unrealistic for some disciplines but it is not unheard of. For example, annual flow for the Main River at Wuerzburg, Germany started in 1823. Additional data sets were simulated with stronger interactions ($b_{2,1}$ of 1.35 and 1.8) and two observations for each time series per time step instead of just one. Increasing the number of observations led to two realizations of x_{1_t} that differed from each other because of added i.i.d. observation error and two realizations of x_{2_t} that also differed from each other because of observation error to emulate collecting data at two stations or by two observers simultaneously. Shorter data sets were created by truncating the full 100 time steps to the first 25 time steps.

1.2.3 *Estimation methods*

Buonaccorsi *et al.* (2001) highlighted the difficulty in estimating interactions between time series and concluded that prewhitened correlation (Shumway and Stoffer, 2011) leads to better characterization of interactions than correlation in the presence of autocorrelation. Multivariate

autoregressive models have also been suggested as a method to estimate interactions (Ives *et al.*, 2003). The two methods have yet to formally compared. Here, we investigated both methods as well as standard cross correlation using Pearson's correlation coefficients and multivariate autoregressive state-space (MARSS) models (Holmes *et al.*, 2012). See Appendix A for more details.

Cross correlation is a widely used statistic to investigate relationships between lagged time series. The statistic provides a measure of linear intensity bound between ±1, measuring the proportion of variation explained. We hypothesized that lag-zero and -negative one coefficients would reflect scaled versions of the true q and $b_{2,1}$ used to simulate the data because the data were generated using stationary process errors and one-way direct-interactions at a lag of negative one. Lags of ±2 were investigated to explore the repercussions of estimating lags beyond those used to generate the data (Olden and Neff, 2001).

Prewhitened cross correlation was included to investigate the benefits of attempting to remove autocorrelation prior to calculating correlation coefficients. Time series were "prewhitened" by finding the best-fitting autoregressive integrated moving average (ARIMA) model for the leading time series and applying that same model to the lagging time series. Prewhitened cross correlation coefficients were then calculated on the residuals in the hope that the leading time-series residuals were a white noise process. ARIMA models were used instead of AR(1) models because the true generating process would not be known outside of a simulation context, and thus, the process better replicated what could be done using empirical data. Estimated autoregressive, moving average, and differencing parameters are summarized in Table A.2.

Lastly, two configurations of MARSS models (Table A.1) were investigated. The first configuration ignored observation error (hereafter referred to as MAR models) and the second configuration estimated i.i.d. observation error (e.g., Ward *et al.*, 2010). The process component of the MARSS model specifies the unobserved state and was similar to Equation 1.1. The estimation method freely estimated all four parameters relating the leading and lagging time series in the interaction matrix, **B**, and an additional parameter vector, **u**, was added to describe the mean. Thus, $\mathbf{x}_t = \mathbf{B}\mathbf{x}_{t-1} + \mathbf{u} + \mathbf{w}_t$. The observation component, $\mathbf{y}_t = \mathbf{Z}\mathbf{x}_t + \mathbf{v}_t$, allowed for the estimation of multivariate normal observation error, $\mathbf{v}_t \sim MVN(0, \mathbf{R})$. Multiple observations of the state process were accommodated using **Z**, which is a model matrix of zeros and ones specifying how each element in the vector of observations at t. $\mathbf{y}_t$ correspond to the unobserved states.

Ignoring observation error involved fixing all elements of **R** to zero. **R** was specified diagonal matrix with one parameter when observation error was estimated (e.g., Ives and Dakos, 2012). Models were fit using the MARSS package (Holmes *et al*., 2012) in the R statistical language (R Core Team, 2018).

1.2.4 *Estimation method performance*

Cross correlation, prewhitened cross correlation, and MARSS models were evaluated for their ability to estimate interactions using median estimates across 100 simulations and visual inspection of their variability. Large sample sizes can lead to significant results for weak relationships (White *et al*., 2014), so significance was not reported. Results of the MARSS models are only reported for those that converged. Converged models were those that successfully estimated standard errors of parameters (Table A.3).

The data-generating process was the most similar to MARSS models. Comparisons of bias (estimated-true) were made between parameters of the data-generating process and all investigated models to emphasize the ability, or lack thereof, of each model to estimate each type of interaction. The off-diagonal elements of **B** can range beyond ±1 and are not directly comparable to correlation coefficients, which are bound between ±1. The scaling issue can be further complicated when data are standardized to have a mean of zero and a standard deviation of one. Ecological data are typically standardized in this manner (Schielzeth, 2010) and thus results are also included for models fit to standardized time series.

1.3 RESULTS

As expected, prewhitened cross correlation and MARSS models were on average better able to estimate interactions than cross correlation when the time series were autocorrelated (bottom row in Figures 1.2-1.4). Prewhitening sometimes decreased the magnitude of the direct-interaction leading to increased bias relative to correlation in a few instances. Prewhitened results were almost always more precise than correlation. Interpreting cross correlation coefficients was not straightforward (Figures A.2-A.4). The coefficients reflected combinations of the true interactions used to simulate the data. For example, when the direct-interaction was large the lag negative one correlation decreased with decreasing self-interaction strength (Figure A.2).

Estimates from models fit to data sets that were simulated without observation error provided inference regarding model performance for a best-case scenario. Bias was only measurable with respect to MAR (i.e., **R** fixed at zero) and MARSS models because their less complicated forms matched the model used to generate the data. The bias was negligible for this best-case scenario (Figure 1.2). Prewhitened cross correlation detected the presence of non-zero direct-interactions at a lag of negative one (Figure 1.2). The signal of the direct interactions was not present at other investigated lags (results not shown). Lag negative one correlation coefficients were more precise but less accurate than MAR models in estimating the strength of the direct-interaction when the self-interactions were zero (Figures 1.2 and A.5). Lag-zero correlation coefficients from prewhitened time series were more precise than those from data that were not prewhitened. Prewhitening reduced the amount of information present in the data about co-interactions as more types of interactions were present in the data-generating process (Figures 1.3 and A.6). Estimates of the co-interactions from MARSS models were unbiased for this best-case scenario. It was not surprising that MAR models better characterized the self-interactions than the unconstrained prewhitening procedure because they were limited to an AR(1) process (Figures 1.4 and A.7; Table A.1).

The presence of observation error led to increased bias and variance for co-interactions compared to when prewhitened cross correlation and MAR models were fit to data without observation error (Figure 1.5 versus Figure 1.3). Variance also increased when the length of the time series decreased. Many MARSS models did not converge (Table A.2) and those that did converge struggled to discern interactions unless all three types were present at extreme values (light gray in right panel in Figure 1.5) or multiple observations were available per state (i.e., a non-identity **Z** matrix; Figures A.8-A.9).

Observation error also induced bias in the characterization of direct-interactions (Figure A.10). As mentioned above, prewhitened cross correlation could not be assessed for bias in the same way as the output from MARSS models. In general, estimates of the direct-interaction from MARSS models increased and decreased relative to the true value more consistently than estimates from prewhitened cross correlation. However, estimates from MARSS models shrank towards zero when the data were standardized. Prewhitened cross correlation estimates were typically similar regardless of standardization (Figure A.11).

1.4   DISCUSSION

Ecological time series rarely consist of independent identically distributed random variables (e.g., Pyper and Peterman 1998, Hollowed *et al.*, 2001). Yet, ecological time series are commonly treated as though they were. Results presented here corroborate previous research demonstrating that autocorrelation can lead to biased estimates (e.g., Yule 1926, Katz 1988) and that prewhitening reduces this bias (Buonaccorsi *et al.*, 2001). More novel was the finding that MARSS models only outperformed prewhitened cross correlation when observation error was low and the state variables strongly interact. This result is surprising because the prewhitening procedure was not constrained to match the data-generating process. In contrast, MARSS models were constrained to an AR(1) process. Unfortunately, correlations cannot be used for forecasting, a popular practice in ecology, whereas well-informed MAR and MARSS models provide an ideal platform for forecasting time-series.

Ecologists are typically interested in direct-interactions and less interested in co-interactions. However, the latter can have important consequences such as increasing extinction risk (Petchey *et al.*, 1997; Heino, 1998). Additionally, the combination of co-interactions with direct- or self-interactions can lead to large and nonintuitive changes over time (Ripa and Ives, 2003). For example, time series that exhibit positive self-interactions are equally likely to be positively or negatively correlated if they react similarly to their environments (Figure A.1). Therefore, *a priori* knowledge of the biological mechanism causing the direct-interaction is not enough. Ecologists should also be aware of the potential for interactions driven by environmental conditions (Ripa and Ives, 2003), which are often what is being measured.

Informative time series led to very little bias in MARSS results. Ecological time series are typically less informative than what would be considered ideal. Consequently, results from MARSS models were often imprecise because many parameter combinations could explain the patterns in data. When too many parameter combinations fit the data the MARSS models failed to converge. Even MAR models failed to estimate standard errors when fit to some of the simulated data sets. The amount of information in the data decreased when the diagonal of the **Q** matrix was small (results not shown), observation errors were present, and the length of the time series was less than ideal. Convergence is a known issue for many time-series methods because time series are often short and subject to observation error (Ward *et al.*, 2017). Convergence improved when

MARSS models were fit to multiple observations of the same process (See and Holmes, 2015). It is unclear how to best accommodate known dependence structures such as sampling location using correlation. In contrast, MARSS models can be informed about dependence structures and may even be able to accommodate correlation structures that are likely to arise from spatial processes (Moran, 1950).

Prewhitening can be performed using several methods such as first differencing; smoothing; first-order autoregressive models fit using ordinary least squares, generalized least squares, maximum likelihood, or restricted maximum likelihood; or ARIMA models. Here, ARIMA models that differed in their degree of differencing, autoregressive structure, and moving-average parameters were fit to the leading time series and the best model was chosen via Akaike information criteria corrected for small sample sizes. This procedure was chosen to mimic an empirical analysis where the true structure is unknown and does not necessarily follow the standard AR(1) autoregressive model (see Appendix A for more details). Patterns in bias resulted when the ARIMA models failed to discern self-interactions from combinations of interactions, similar to estimates of correlation. These patterns may have been less prominent had the prewhitening procedure been more specific and matched the data-generating procedure. Additionally, some prewhitening procedures can accommodate state-space formulations to estimate observation error. However, canned software to do so is not as widely available or documented as thoroughly as MARSS models. Nevertheless, estimating observation error within the prewhitening procedure may decrease the likelihood of over smoothing. Over smoothing can lead to positively biased estimates of interactions at non-zero lags because it induces correlation in the data (Katz, 1988). In general, ARIMA models fit to the longer time series better estimated the AR(1) parameter than when they were fit to the shorter time series, corroborating previous research suggesting at least 50 data points are necessary for cross correlation (Box and Jenkins, 1976).

Choosing an appropriate model to account for autocorrelation is challenging, especially when data are influenced by several processes. Results indicate that MARSS models offer an integrated method for estimating multiple types of interactions should the data be informative. Cross correlation proved to be adequate for detecting relative differences in direct- and co-interactions. Future work on increasing the stability of MARSS models could increase their applicability within ecology where time series are short and fraught with observation error. Non-Gaussian distributions

and Bayesian methods could be a promising line of research to better estimate synchrony within environmental time series.

1.5 FIGURES
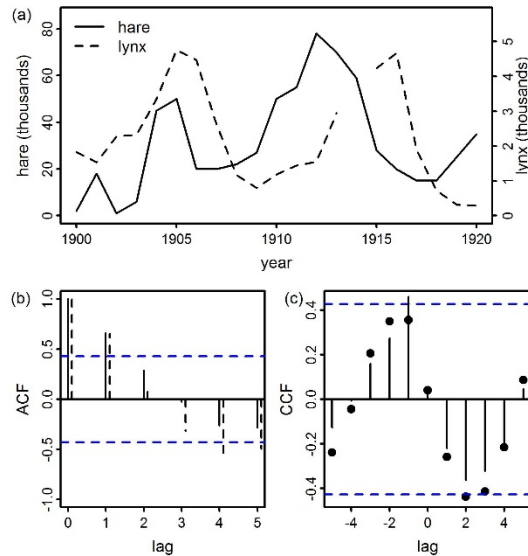
Figure 1.1. Lynx (*Lynx canadensis*) and snoeshoe hare (*Lepus americanus*) from 1900 to 1920. Panel (a) depicts counts (thousands) of hare (solid line) and lynx (dashed line) from 1900 to 1920. Both time series exhibit lag-1 autocorrelation (~0.7; panel b). Cross correlation was only significant at a lag of -1 after an AR(1) model was fit to the hare data to prewhiten the time series (solid vertical lines in panel c). Cross correlation of the raw time series (points in panel c) were largely not signficant at the 0.05 level. The dashed blue horizontal lines in (b) and (c) are the values beyond which autocorrelations are significantly different from zero.
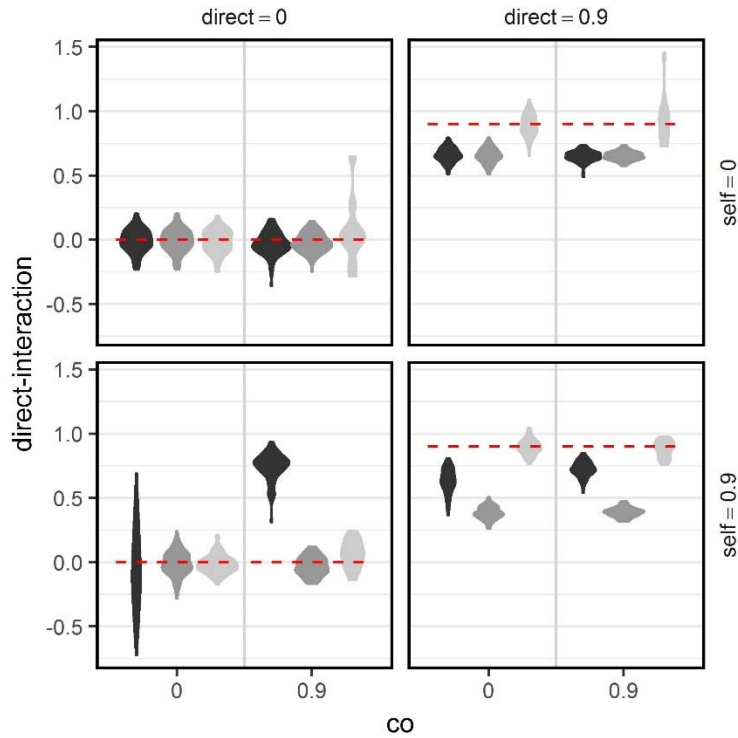
Figure 1.2. Violin plots of estimated direct-interactions from cross correlation (dark gray), prewhitened cross correlation (gray), and multivariate autoregressive state-space (MARSS) models (light gray) that did not estimate observation error (referred to as MAR models; light gray). Self-interactions are present in the simulated data in the bottom row. Direct-interactions are present in the right column. Co-interactions are present in the right side of each panel. Fitted time series included 100 points observed with error. Horizontal dashed, red lines indicate the true direct-interaction used to simulate the data. Results from MAR models are limited to those that successfully estimated standard errors for all included parameters.
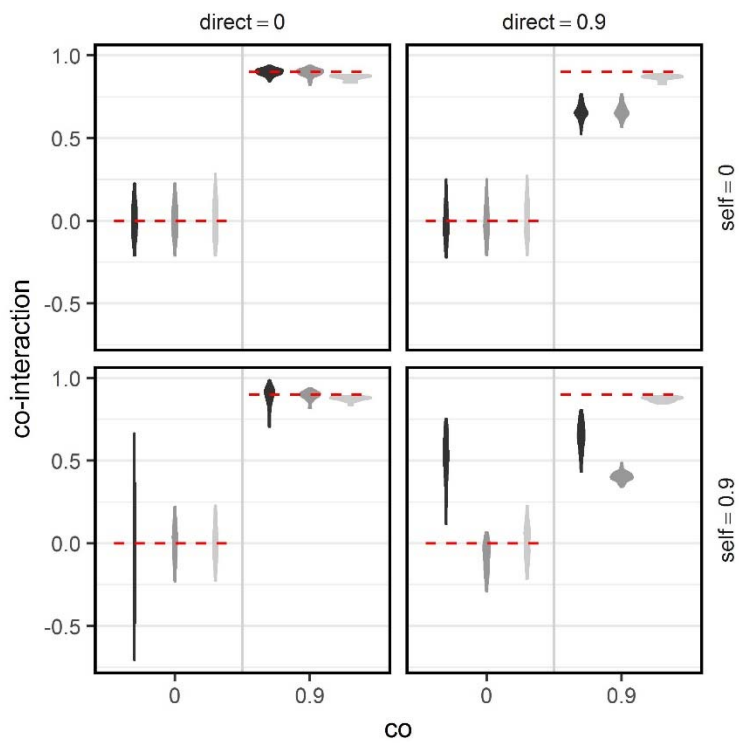
Figure 1.3. Same as Figure 1.2, except the y-axis displays estimates of co-interactions.
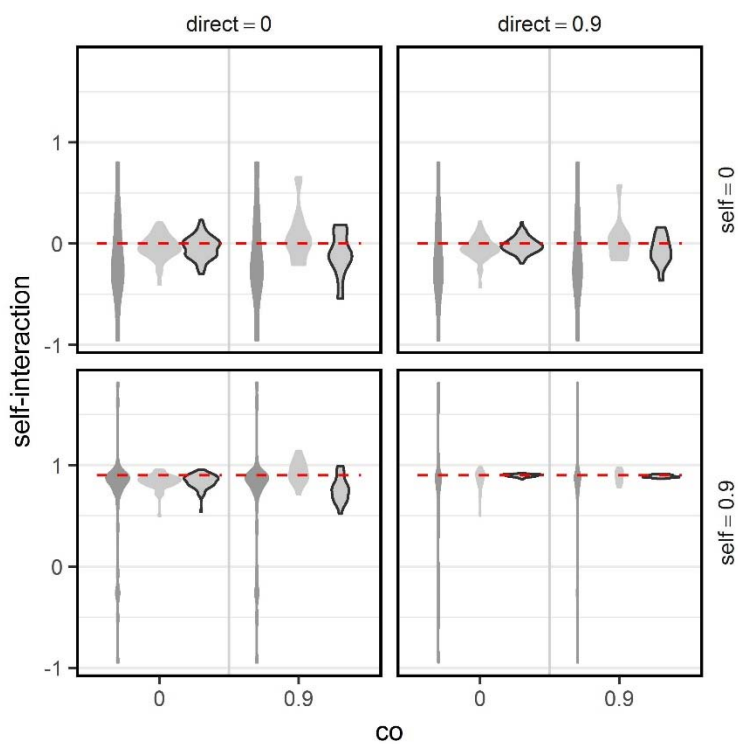
Figure 1.4. Same as Figure 1.2, except y-axis displays estimates of AR(1) from the prewhitening procedure (gray) and from MAR models for the leading (light gray) and lagging (light gray with black border) time series.
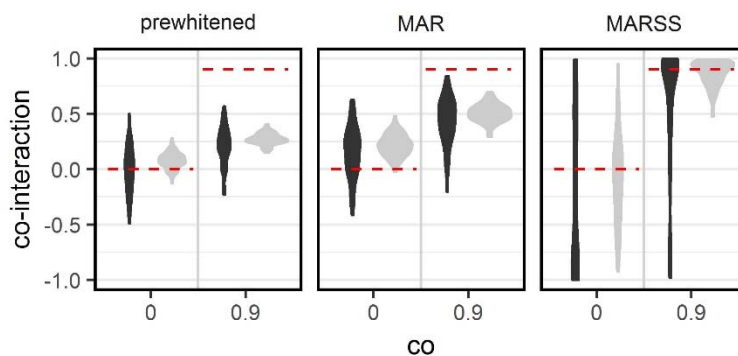
Figure 1.5. Violin plots of estimated co-interactions from prewhitened cross correlation, multivariate autoregressive models (MAR), and multivariate autoregressive state-space (MARSS) models fit to 25 and 100 years of data (black and gray, respectively) across two levels co-interactions (x-axis; horizontal dashed, red line). Time series were simulted with a direct-interaction of 0.9, autocorrelation of 0.9, and observation-error variance of 1.0. Only MARSS models estimated observation error. The prewhitening model was the model that best fit the leading time series as chosen by Akaike information criterion corrected for small sample size. Results from MAR models are limited to those that successfully estimated standard errors for all included parameters.

Chapter 2. TESTING IF INDICATORS OF MARINE ECOSYSTEMS PREDICT ECOSYSTEM ATTRIBUTES USING ECOSYSTEM MODELS FROM ACROSS THE GLOBE

Abstract

Ecosystem-based management in the ideal case is informed by knowledge of the state of the system relative to management targets and how the ecosystem may change with new regulations. Unfortunately, we cannot precisely measure many ecosystem properties ("attributes") and must rely on indicators or proxies of attributes to infer ecosystem status. Traditionally, correlation has been used to quantify how well indicators track attributes although this violates assumptions of correlation analyses because the data are typically autocorrelated. We used a simulation experiment to quantify indicator-attribute relationships using time-series models (i.e., multivariate autoregressive models) fitted to simulated output from eight ecosystem models. Biomass and mean trophic level from survey data best tracked ecosystem attributes. However, indicator performance varied across ecosystems. Indicators tracked attributes that were calculated in the same year more than attributes from the previous year. Accounting for the time-series dependencies inherent in the data was necessary to avoid bias. This work provides a framework for testing indicator-attribute relationships, which we recommend prior to using indicators in a management context.

2.1 INTRODUCTION

Ecosystem-based management (EBM) requires balancing multiple objectives while managing ecosystems subject to cumulative and potentially synergistic impacts from drivers of change (Crain *et al*., 2008; Darling and Côté, 2008; Coll *et al*., 2016). Potential drivers of change include those arising from the physical/chemical environment, anthropogenic activities, and ecological structure and function (e.g., predation). All of these drivers affect the ability to meet EBM goals, and Botsford *et al*. (1997) suggest that the historical lack of EBM is a major cause of unsustainable marine fisheries in many nations. Consequently, fishery managers, once charged with primarily setting catch limits, have now been tasked with addressing several objectives such as conserving

habitat, protecting endangered species, and maximizing catch while the oceans are experiencing increased use (Link, 2010).

Progress towards implementing EBM involves the following steps: development of ecosystem goals (Kerschner *et al.*, 2001), assessment of ecosystem status, and selection of decisions on how to best reach goals given the ecosystem status (Lackey, 1998; Levin *et al.*, 2018). Advancement towards reaching ecosystem goals can be measured using ecosystem attributes. For example, measurements of biodiversity (e.g., Noss, 1990; Heino, 2010) can provide information on how well ecosystem resilience is being maintained. Changes in attributes, if measurable (or within a simulation), provide relevant information on ecosystem structure and function and quantitative measures of ecological responses to drivers of change. Attributes are often challenging to measure directly and require the use of proxies (i.e., indicators). Indicators provide a means to reduce the complexity associated with observing systems into a small set of potentially useful, measurable characteristics (e.g., Fulton *et al.*, 2005; Methratta and Link, 2006; Shin *et al.*, 2010b; Boldt *et al.*, 2014). Indicators have been applied in the context of EBM to quantify ecosystem status and trends, assess management performance across disparate objectives, and evaluate the strength and synergies of drivers (e.g., Shin *et al.*, 2010a). For example, under the European Marine Strategy Framework Directive, decision makers use indicators to measure progress toward Good Environmental Status (Borja *et al.*, 2013). In the US, the status and five-year trend of indicators relative to the long-term mean are used as proxies for ecosystem status within US Integrated Ecosystem Assessments (Harvey and Garfield, 2017) and related report cards (Zador *et al.*, 2017).

The choice of which indicators to include in assessments of ecosystem status is a difficult, but essential, component of EBM. No single indicator sufficiently quantifies all relevant objectives (Kleisner *et al.*, 2015; Masi *et al.*, 2017) and not all indicators are applicable to all management systems (Blanchard *et al.*, 2010; Boldt *et al.*, 2014; Coll *et al.*, 2016). Indicators should be limited to those that are directly relatable to observations and theoretically sound. They should also be responsive, specific, and sensitive (Jennings, 2005; Rice and Rochet, 2005; Niemeijer and de Groot, 2008; Kershner *et al.*, 2011). To be *responsive* they should reflect ecosystem changes within a short time period (here we choose one to two years). Only those that respond in a known way are *specific*. *Sensitive* indicators are those with smooth, monotonic, and strong responses. Their cost effectiveness is often based historical availability, measurability, and quantifiability. Finally, their familiarity to the general public can be assessed using expert opinion and stakeholder input.

Historically, indicator-attribute relationships have been assessed for their responsiveness and sensitivity using correlation (e.g., Fulton *et al.*, 2004; Samhouri *et al.*, 2009). However, the significance of the correlation and potentially the correlations themselves will be biased if past values of the attribute affect future attribute values (i.e., autocorrelation; Orcutt and James, 1948; Shumway and Stoffer, 2011). The previous will also be true if the indicator is autocorrelated, which is often the case but its affect is often wrongly assumed to be negligible. Multivariate autoregressive (MAR) state-space models (Ives, 1995; Ives *et al.*, 2003) offer an unbiased approach to estimating interactions among time series (Chapter 1). The MAR approach provides inference on how much of the variability in the data can be explained by "direct" linear changes between the indicator and the attribute versus shared "co" variability due to factors external to themselves or their past "self" (Sugihara, 1995). Note that the partitioning of variability is different from inferring causation. Nevertheless, disentangling variability is an important step in assessing indicator-attribute relationships and could potentially reduce the number of contradictions observed when additional years of data are added (Sugihara *et al.*, 2012).

Here, we extend previous work on indicator-attribute relationships (e.g., Fulton *et al.*, 2005; Samhouri *et al.*, 2009; Shin *et al.*, 2018) to demonstrate the need to account for autocorrelation when assessing indicators for management purposes. We used output from eight simulations representing ecosystems from all over the globe (Olsen *et al.*, 2018) to identify if indicator performance varies by ecosystem. Additionally, future simulated ecosystem dynamics were forced, in part, using gradients of changes in fishing, ocean acidification, or marine protected areas to infer if indicator performance depends on drivers of change. The comparison was conducted across a range of drivers in hopes of providing inference that will be robust to a wide range of future conditions.

2.2 METHODS

We analyzed output from eight marine ecosystem models (Table 2.1; Figure 2.1; detailed in Olsen *et al.*, 2018) to quantify the sensitivity and specificity of 12 indicators (Table 2.2) relevant to EBM. As detailed below, analyses included the following steps: project each ecosystem model forward in time for fifty years under base-case conditions or one of three drivers of change, extract relevant model outputs to calculate indicators and attributes, calculate the sensitivity of indicator-attribute relationships, and investigate indicator specificity. Measures of sensitivity were

investigated using MAR models that quantify relationships among time series while accounting for autocorrelation (Chapter 1). Results from MAR models were compared to those based on Pearson cross correlation coefficients.

2.2.1   *Ecosystem models*

Detailed end-to-end ecosystem models (Rose *et al.*, 2010), such as Atlantis (Fulton *et al.*, 2004, 2011), allow for the model developer to control for factors that often plague data sampled from actual ecosystems, while systematically adding or removing complexity. Atlantis is a cost-effective tool for evaluating management strategies (Fulton *et al*., 2014; Dichmont *et al*., 2016) and indicators (Fulton *et al*., 2005; Smith *et al*., 2015; Masi *et al*., 2017), allowing for the evaluation of management actions without having to implement them in reality. Typically, simulations have been parameterized for individual geographic regions. Recent research has highlighted the value of comparing a wide range of modelled ecosystems subject to similar drivers of change (e.g., Megrey *et al*., 2009; Olsen *et al*., 2018). The included ecosystems spanned four ecosystem types, coral reef (n = 1), open ocean (n = 4), gulf (n = 2), and brackish waters (n = 1).

Atlantis was used to represent the dynamics of each ecosystem. Briefly, Atlantis is a 3-dimensional, end-to-end ecosystem modelling framework. The deterministic model allows for two-way coupling between environmental and anthropogenic drivers of change and their direct and indirect effects on ecosystem components. Bottom-up and top-down forcing facilitate realistic, complex future ecosystem dynamics that cannot necessarily be predicted *a priori*. A system of forward differential equations simulates ecosystem dynamics specific to spatial polygons, typically in 12-hour time steps. Polygons are designed to match major geographical features of the ecosystem such that homogeneity can be assumed within a polygon. Polygons are vertically resolved with up to five water-column layers, a sediment layer, and an epibenthic layer. Movement of biological groups can occur within polygons, between polygons, and groups can migrate in and out of the model domain. Modelled biological groups often include multiple species centered on a specific guild (e.g., small pelagic fish). Model developers can also choose to make groups based on a single species. Detritus groups (labile detritus, refractory detritus, and carrion) are modelled as aggregate biomass pools. Vertebrate groups are structured by age. Other biological groups, such as invertebrates, can be modelled as aggregate biomass pools or structured by age.

Fishery dynamics can include multiple fleets, each with their own set of target and bycatch groups and fishing areas. Spatial closures to fishing can be included as marine protected areas, and fishing dynamics can range from simple harvest rates to complex agent-based parameterizations. Here, ecosystems were parameterized in terms of constant fishing mortality rates to generate catches that approximated recent observed time series and biomass levels that approximated relative biomasses estimated from survey data. In the case of the Nordic and Barents Sea ecosystem, catches were scaled in proportion to their maximum sustainable yield. Additional models parameterized in terms of dynamic effort fishing mortality rates were included for the NE US and SE Australia ecosystems to allow fishing rates to vary across time to better approximate observed time series (see Link *et al.*, 2011; Fulton *et al.*, 2014).

Key Atlantis outputs include time series of catches and abundances of biological groups. These outputs are compared to empirical data during the calibration stage such that modelled productivity and resilience closely replicate historical trends where appropriate (see Table 2.1 for relevant primary publications). The initial year for each ecosystem depended on data availability, but all ecosystems were projected forward for fifty years from their respective starting year. The first twenty years were considered a burn-in period to account for transient behavior, leaving thirty years for the analysis.

2.2.2 *Drivers of change*

Detailed predictions of marine ecosystem responses to anthropogenic drivers of change are challenging (Evans *et al.*, 2011), but scenario approaches (Maury *et al.*, 2017) are one tool that has been utilized to explore a range of future conditions. Previous simulation work in individual ecosystems has shown that the largest tradeoffs among stakeholder groups and among policy goals often result from scenarios related to ocean acidification, selective fishing, and marine protected areas (Kaplan *et al.*, 2012; Weijerman *et al.*, 2016b). Therefore, we chose to explore these three drivers. For simplicity, we considered the drivers independently rather than creating bundled scenarios with multiple drivers and cumulative effects. The eight ecosystem models were developed for a range of purposes and the simulated effects of drivers will be specific to some of the choices made by model developers. Base-case scenarios typically represent a recent year for each ecosystem. Drivers were parameterized as consistently as possible across ecosystems by implementing parallel relative changes to the ecosystem-specific base-case scenarios. For

example, the Gulf of Mexico base-case scenario represents conditions in year 2012 and includes 24 marine protected areas (Ainsworth *et al.*, 2015) and the N Gulf of California base-case scenario represents conditions in year 2008 and does not have marine protected areas (Morzaria-Luna *et al.*, 2013), so ensuring that 10% of the continental shelf was protected by marine protected areas led a larger percent increase in marine protected areas for the N Gulf of California than the Gulf of Mexico. Scenarios include the following and were also used in Olsen *et al.* (2018):

1. Base case: The base-case scenario represents business-as-usual for each ecosystem based on published values (Table 2.1) and served as the reference scenario. Fishing mortality was ecosystem-specific and specified at the level used for calibration. Therefore, fishing was constant with respect to time for all models except the two models parameterized using dynamic effort (NE US and SE Australia).

2. Fishing: Fishing mortality was altered (eliminated, halved, or doubled) from base-case levels. Fishing scenarios were implemented for each of the following groups in turn if they were included in the ecosystem: invertebrates, small-pelagic fish, demersal fish and sharks, large-pelagic and highly-migratory species, and all fished groups. Fishing remained at base-case levels for all other groups in an ecosystem that were not identified as the group of concern. For example, fishing of demersal sharks and fish were eliminated while all other fishing rates were kept the same as the base-case scenario was one fishing scenario. The Nordic and Barents Sea ecosystem was excluded from fishing scenarios because fishing mortality was parameterized in terms of maximum sustainable yield rather than historical fishing. The two dynamic effort models were also excluded because fishing mortality was time varying.

3. Ocean acidification: The two ocean-acidification scenarios increased natural mortality of calcifying algae, corals, coccolithophores, echinoderms, and mollusks (i.e., groups thought to be affected by ocean acidification) to mimic hypothesized changes in survival (Kroeker *et al.*, 2013). The more extreme scenario added 1% to the base-case natural-mortality rate (day^{-1}) for the previously listed taxa, and the less extreme scenario added 0.5%.

4. Marine protected areas: Three no-take marine protected area scenarios were investigated. Protected areas were initiated starting from shore and extending out towards the edge of the continental shelf (250 m) until 10, 25, and 50% of the

continental-shelf area was closed to fishing. These scenarios represent a reduction in total fishing mortality rather than a displacement because fishing rates outside of the closed areas were maintained at the same rates used in the base case. The constant-effort version of the NE US model was excluded from these scenarios because the dynamic-effort version was hypothesized to provide a more realistic representation of fishers' potential responses to protected areas.

2.2.3   *Attributes*

Attributes were chosen for their theoretical importance to ecosystem structure and function (Bradshaw, 1984). Investigated attributes included (a) biomass of groups targeted by fisheries, (b) total biomass, (c) mean trophic level (TL) of the ecosystem (Christensen, 1996), (d) net primary productivity (NPP; Odum, 1985), and (e) total biomass divided by NPP. The first three attributes summarize food-web structure and the latter two relate to energetic attributes of the community. Biomass of groups with a TL$\leq$1 were excluded from (b) and (c) and TL<1 were excluded from (d) because of their ephemeral nature and sometimes exceptionally large biomass. We acknowledge that additional or different attributes could have been examined, but we leave that for future research.

Attributes are not easily measured in the real world. For these simulated ecosystems, we could calculate their values using data from all species, including species not sampled by surveys, not targeted by fisheries, and not assessed using stock assessments because simulations allow the truth to be known. Indicator performance was then measured by their ability to track these ecosystem attributes of interest. In this way, simulation can thus act as an initial filter to eliminate indicators that are unlikely to track attributes in the real world.

2.2.4   *Indicators*

Of the many potential indicators, we chose to evaluate the suite of indicators that were investigated by the Indicator for the Seas (*IndiSeas*) Working Group (Table 2.2). Survey biomass, stability of the survey biomass, biomass-weighted mean size of surveyed fish, biomass-weighted mean life span of surveyed fish, proportion of the survey biomass comprised of predatory fish, biomass-weighted mean TL of the survey, proportion of those groups with an assessment that

shoed a declining trend, and the proportion of those groups that were estimated to be at greater than 60% of their unfished biomass were based on survey data. Biomass-weighted mean intrinsic vulnerability index (IVI; Cheung *et al.*, 2005) of fish catches, biomass-weighted mean TL, and biomass-weighted mean tropic index of the catch were derived from catch data. Inverse fishing pressure used both survey and catch data. Some of these indicators would also need other data to provide life-history information or for an assessment of their status if the data were empirically collected as opposed to simulated output. *IndiSeas* chose these 12 indicators for their ability to quantify the impact of fishing on ecosystem status based on several criteria (*sensu* Rice and Rochet, 2005). Criteria for their assessment included their hypothesized comparability across ecosystems (Shin and Shannon, 2010; Shin *et al.*, 2010b; Coll *et al.*, 2016). Indicators and attributes are highlighted using quotation marks rather than acronyms for clarity.

Some specifications of the *IndiSeas* calculations had to be adapted for Atlantis simulated output (Table B.1; see Appendix B for more details). For instance, data used by *IndiSeas* were of individual fish, but Atlantis models cohorts of fish. Consequently, "fish size" was changed to be the survey-biomass weighted mean of maximum fish size (Worm *et al.*, 2009) rather than individual fish size. Utilized life-history information for multi-species groups was from the most dominant species or from a biomass-weighted mean depending on choices made by the model developer. Estimation of self-, direct-, and co-interactions

Indicator-attribute relationships were quantified using MAR models, a tool for analyzing time-series data (Ives, 1995; Ives *et al.*, 2003). The approach regresses an indicator against an attribute to quantify interactions while accounting for relationships of each time series with their past self and common relationships of both time series with their external environment. Prior to fitting the models, indicators and attributes were standardized to have a mean of zero and a standard deviation of one to facilitate the comparison of values collected on different scales and to increase MAR model convergence compared to modelling values on their absolute scale. The model, $\mathbf{d}_t = \mathbf{B}\mathbf{d}_{t-1} + \boldsymbol{\eta}_t$, was fit to the indicator and attribute time series using the Multivariate Autoregressive State Space (MARSS) package (Holmes *et al.*, 2012) in R (R Core Team, 2018). Here, $\mathbf{d}_t$ is a vector of length two representing the attribute and indicator at time t, $\boldsymbol{\eta}_t$ is a vector of length two representing the factors not explicitly included in the model dynamics at time t, and $\mathbf{B}$ is a two by two matrix.

The **B** matrix included parameters to estimate interactions between each time series with itself at the previous time step (termed self-interactions) on the diagonal, interactions between the attribute at the previous time step and indicator in the lower off diagonal ($b_{2,1}$; termed direct-interactions), and could estimate the direct-interaction between the indicator at the previous time step and the attribute in the upper off diagonal ($b_{1,2}$), but $b_{1,2}$ was fixed at zero because it was assumed that indicators do not affect attributes. Any time lag or combination thereof could be investigated, but a 1-year lag was investigated because managers normally work within a short time horizon.

The variance-covariance matrix (**Q**) of the stochastic dynamics, $\eta_t \sim \text{MVN}(0, \mathbf{Q})$, included parameters to estimate the variance of the attribute ($q_{1,1}$) and the variance of the indicator ($q_{2,2}$) on the diagonal and a single parameter to estimate the covariance on the off diagonal (i.e., $q_{2,1} = q_{1,2}$). The standardized covariance (termed co-interaction), $q_{1,2}/(\sqrt{q_{1,1}}\sqrt{q_{2,2}})$, provided a measure of the synchrony in the temporal variability of the time series after accounting for the direct- and self-interaction(s). This interaction could include changes due to extrinsic environmental conditions that lead to changes in both the indicator and the attribute.

Parameter estimates from the MAR models were based on maximum likelihood estimation and confidence intervals were calculated using parametric bootstrapping based on 20 samples (Stoffer and Wall, 1991). Bootstrapping was used rather than asymptotic variances because the assumed symmetry of the asymptotic distribution can lead to biased estimates when variance parameters are small.

2.2.5   *Calculation of sensitivity from direct- and co-interactions*

An indicator-attribute relationship was considered sensitive if the direct- or co-interaction was positive and the confidence interval did not cross zero. Negative relationships were also considered for two attributes, "total biomass" (except between "biomass" and "total biomass") and "NPP" because the direction of movement in the indicator as the ecosystem experienced fishing could depend on the type of fishing (Samhouri *et al.*, 2009). Nevertheless, the signs of the non-zero co- and direct-interactions had to match for the relationship to be considered sensitive. We did not investigate ranges of sensitivity because we had no *a priori* knowledge of how strong an interaction should be to be of use to management, only that it should be non-zero.

2.2.6 *Calculation of specificity from sensitivity*

Specificity has been defined several ways, such as an indicator being specific if it is only sensitive to a single attribute (discussed below as coverage; Kershner *et al*., 2011), if it accounts for a single species (discussed below as representativeness; Hattam *et al*., 2015), or if it is affected by a single external pressure (Rice and Rochet, 2005). We adopted the following definition of specificity close to that of Rice and Rochet (2005) to disentangle the impacts of fishing from other drivers: the proportion of relationships for a given indicator and attribute combination that were sensitive out of all investigated relationships for that indicator-attribute combination across drivers of change. Specificity was also calculated across ecosystems for a given driver of change. The magnitude of specificity is meant to provide inference on how difficult it would be to infer that changes in an indictor were the result of changes in a specific driver in the presence of multiple drivers. For example, if 15 of the 22 simulated scenarios (1 base case, 16 fishing, 3 marine protected areas, and 2 ocean acidification) demonstrated a sensitive relationship between "predators" (indicator) and "target biomass" (attribute) in the California Current, then the specificity of the relationship would be 0.68. If 20 of the 22 scenarios demonstrated sensitive relationships the specificity would be 0.91, insinuating it would be difficult to attribute changes in the indicator to a specific driver.

2.2.7 *Calculation of representativeness and coverage*

Indicator performance was further summarized in terms of "representativeness" and "coverage". Representativeness was measured by investigating relationships between indicators and directly analogous attributes (a form of "self-test"). For example, TL of the catch (indicator) was tested against TL of ecosystem (attribute), highlighting how well (or poorly) fisheries data represented the ecosystem. Second, indicators that tracked more than one attribute related to ecosystem structure and function (Table B.2) were labeled as having high coverage. Indicators with low coverage may be preferred if the aim is to identify "diagnostic" indicators, determining changes in specific attributes, while indicators with high coverage may be useful as general bellwethers of ecosystem changes in structure and function.

2.2.8 *Correlation*

The analysis was repeated using Pearson's cross correlation coefficients as a cautionary example of the implications of ignoring the time-series nature of the "data". Here, sensitive relationships were those that were greater than (or less than for total biomass and NPP) what would be expected assuming a significance level of 0.05 and a sample size of 30 (i.e., 0.36). Lag-1 cross correlations were compared to direct-interactions because they were parameterized to investigate the same lag. Lag-0 cross correlations were compared to co-interactions. It would not be expected that the absolute values would be similar because cross correlation measures total variance and MAR models partition the variance into the three interactions (Chapter 1).

2.3 RESULTS

2.3.1 *Low sensitivity of indicators to attributes overall*

The indicators included in this analysis were expected to be good proxies for ecosystem attributes. Only one-third (33%) of the 10,965 investigated relationships were sensitive (Table B.3). Indicator-attribute relationships tended to be sensitive for either direct- or co-interactions but not both (see for example, upper and lower panels of Figures 2.2-2.3, respectively). Co-interactions were often stronger in magnitude than direct-interactions (upper versus lower panel of Figure 2.3, respectively), suggesting that indicators and attributes are synchronously responding to external pressures rather than attributes leading to lagged changes in indicators. Sometimes (7%) these relationships were opposite in sign, a result most common for the "fish size" and "fish life span" indicators. Relationships between "total biomass" and "TL of the survey" had the largest number of simultaneously positive direct- and co-interactions.

Plots of indicators versus attributes helped show why some interactions were negative and why some models did not converge. The provided example between "biomass" and "biomass of the ecosystem" in the SE Australia (Figure 2.2) and the Chesapeake Bay (Figure 2.3) ecosystems illustrate the prominent patterns seen. First, the time-series data extracted from the Atlantis output exhibited strong self-interactions that can be seen by following sequentially numbered points (estimates of the self-interactions are printed on the respective axes). Model output was relatively stable because it was reported for the entire spatial domain on yearly time steps. Second, estimates

of co- and direct-interactions were sometimes opposite in sign compared to lag-0 and lag-1 correlation coefficients (see Figure 2.3 and below for more details).

Estimated self-interactions were generally stronger than estimated direct-interactions. This was particularly true when the self-interaction of the indicator was large and positive (cone-shape pattern in upper right panel of Figure 2.4). For example, the self-interaction of "biomass" was greater than one and the direct-interaction with "biomass of the ecosystem" was less than 0.1 in the Chesapeake Bay ecosystem under the base-case scenario (Figure 2.3). The pattern was the opposite for attributes but weaker. That is, some of the strongest direct-interactions were estimated for attributes with large self-interactions (top left panel of Figure 2.4). Self-interactions were the most extreme for "biomass stability". Neither of these patterns were apparent between co-interactions and self-interactions (bottom row in Figure 2.4).

2.3.2 *Representativeness*

It was hypothesized that indicators would track similarly calculated attributes in a positive manner, but 24% of the relationships included to investigate representativeness were negative (Tables 2.3 and B.4-B.6). "Sustainable stocks", "TL of the survey", and "biomass stability" were the least representative indicators (Table B.3). For indicators that were representative, the sensitive relationship was more likely to be a co-interaction (Tables B.3 and B.4) rather than a direct-interaction (Tables B.5 and B.6), which was not surprising because indicators should represent the current state of the system and not the state of the ecosystem from the prior year. Representativeness appeared to depend on how thoroughly the groups included in that particular attribute were included in the catch or survey data (see % targeted and % surveyed in Table 2.1). For instance, "predators" and indicators related to TL failed to represent the Guam ecosystem (Table 2.3) because only two of the four predatory fish groups were surveyed and only one was targeted although almost half of the ecosystem was surveyed. In contrast, predators were more representative in the Gulf of California because some of the key predatory fish are surveyed although only 18% of all species were surveyed. Representativeness also depended on how fishing was implemented. For example, "trophic index of the catch" was sensitive to "trophic index of the ecosystem" in the NE US only when fishing was implemented using constant rates. Catch-based indicators were the most representative when fishing on all groups increased or fishing on small pelagics decreased (Table B.4).

2.3.3 *Specificity*

Indicator performance was ecosystem-specific but tended to not depend on the drivers of change. Consequently, specificity was generally higher (i.e., more models estimated sensitive interactions that did not conflict in sign) but more variable when calculated with respect to a given ecosystem across all drivers of change than driver of change across all ecosystems (Figure 2.5 and Figure B.1 as opposed to Figures A2-A3). Conversely, if an indicator tracked an attribute well in a given ecosystem, it did so irrespective of which driver of change acted on the system. Relationships within an ecosystem were generally centered on the base-case scenario with scenarios related to ocean acidification or fishing leading to the largest changes in the strength of the relationship (purple crosses and green triangles, respectively, in Figures B.4-B.15). These more extreme values within an ecosystem were generally opposite in sign compared to the median response. Additionally, specificity appeared to be independent of the strength of the relationship, where stronger relationships did not lead to relationships that were more likely to be sensitive across ecosystems or drivers of change. For example, the co-interaction between "TL of the catch" and "biomass" was similar in strength for Chesapeake Bay and the Gulf of California but more specific in the latter (bottom row, second panel from the right in Figure 2.5).

2.3.4 *Coverage*

Two indicators, "biomass" and "TL of the survey", had the highest coverage of the attributes (Table B.3). These indicators had the highest number of MAR models with sensitive direct- and co-interactions for three of the five attributes. Models investigating the "intrinsic vulnerability index" to "target biomass" led to many sensitive relationships, but "intrinsic vulnerability index" was ranked 4th in terms of coverage because it had few sensitive relationships with the remaining four attributes. "Fish life span" also had fairly high coverage, but its coverage was largely related to its sensitivity to "total biomass" and "NPP".

2.3.5 *Poor-performing indicators*

"Non-declining species" (Figure B.4) and "sustainable stocks" (Figure B.5) are based on counts (whole numbers) of species above specified thresholds. These indicators covered the least number of attributes and were the least representative of those investigated (Table B.3). Of the 954

potential indicator-attribute relationships investigated with "sustainable stocks" as the indicator, only 658 successfully lead to the estimation of an asymptotic variance-covariance matrix. The indicator was often constant over multiple years, and the MARSS framework (Holmes *et al.*, 2012) had difficulty parsing the small amount of variability in the data among the parameters. "Non-declining species" had relatively high model convergence (93%), but only 3% of the models estimated sensitive direct- and co-interactions.

2.3.6 *Correlation*

The sign and the magnitude of cross correlation coefficients did not always match the MAR results. This was not surprising because correlation coefficients are bound between [-1,1]. It was concerning that for some relationships the sign changed when the time-series properties of the data were accounted for. Many co-interactions had negative lag-0 cross correlation coefficients when the co-interaction was estimated as positive, but positive lag-1 cross correlation coefficients when the direct-interaction was negative (Figure B.16). For example, cross correlations between "biomass" and "biomass of the ecosystem" were negative while the direct- and co-interactions were positive (Figure 2.3).

2.4 DISCUSSION

There is not, and probably will not ever be, a universal indicator for EBM. Instead, multiple appropriate indicators must be chosen such that managers can be informed about the status of their goals. Choosing appropriate indicators requires understanding indicator-attribute relationships, which we show can be facilitated using MAR models. We assumed that indicators should track attributes relevant to EBM (Bradshaw, 1984), "TL of the ecosystem", "target biomass", "ecosystem biomass", "NPP", and the ratio of the latter two. "Biomass" tracked many of these attributes for most drivers of change and across most ecosystems. Fewer of the *IndiSeas* indicators (Coll *et al.*, 2016) tracked the relevant ecosystem properties than hypothesized and indicator performance was ecosystem specific. Thus, we suggest that indicator selection be ecosystem-specific and informed by simulation testing as well as empirical validation.

Multivariate autoregressive models provided descriptions of indicator and attribute time series while accounting for autocorrelation in the data. The framework decomposes temporal changes in the indicator and explicitly apportions them to (1) the indicator at the previous time step (self-

interactions), (2) the current environment (co-interactions), and (3) the attribute in the previous time step (direct-interactions). Including (1) is important because autocorrelation can increase the likelihood of estimating spurious relationships (Chapter 1). Self-interactions are similar to autocorrelation but they are not bound between [-1,1]. Direct-interactions were smaller on average when indicator time series exhibited a positive self-interaction. Previous work on indicator performance also documented decreased cross correlation between indicators and drivers of change when the indicator was autocorrelated (Probst *et al.*, 2012) and increased differences between cross correlation and MAR models (Chapter 1). Thus, previously documented correlations between indicators and attributes that assumed the effect of autocorrelation was negligible and may need to be revisited (e.g., Fulton *et al.*, 2005).

Applications of MAR models are well documented and have been well-used to highlight the intricacies of community dynamics. However, their use in ecology beyond understanding plankton communities is limited (Hampton *et al.*, 2012). Their use is becoming increasingly tractable as ecological time series become longer. Furthermore, no additional mechanistic understanding beyond that needed to calculate correlation coefficients is needed to fit the models. They offer the added benefit of being able to simultaneously test multiple hypotheses such as was done here by setting the effect of the indicator on the attribute ($b_{1,2}$) to zero. Lags other than one can be investigated, or multiple lags can simultaneously be investigated. Spatial complexity can be accounted for by including rather than summarizing across spatial replicates (e.g., Ward *et al.*, 2010), Bayesian estimation can allow for the inclusion of priors (MacNally *et al.*, 2010), etc. Perhaps most importantly, results include confidence intervals based on the amount of information in the data. This differs from correlation that uses sample size and significance level, which should be adjusted for the presence of autocorrelation. Adjusting for autocorrelation is widely recognized but no method is best for all situations (Pyper and Peterman, 1998) and many methods depend on estimating the degree of autocorrelation present in the data, which is difficult with short time series (Thompson and Page, 1989). We hope that the results of this manuscript encourage future analyses of indicator-attribute relationships to account for autocorrelation using MAR models because although MAR and correlation may provide similar point estimates the value of point estimates without measures of uncertainty "is of little use" (p. 637 in Jollife, 2007).

Indicator performance suggests conclusions about their potential utility. "Biomass" and "TL of the survey" were the most sensitive indicators of those investigated. Both indicators were

particularly sensitive to "biomass of the ecosystem" with respect to co-interactions. Previous work also highlighted "TL of the survey" (Jiming, 1982; Gascuel *et al*., 2016) as a good indicator. Coll *et al.* (2016) noted that "life span" and "predators" typically agreed with "TL of the survey" but we suggest that "predators" may be a better indicator in general because it does not depend on diet or life-history data (see below for more details). Negative relationships were more common between "TL of the survey" and "biomass of the ecosystem" than "biomass" and "biomass of the ecosystem". Changes in trophic structure often follow intense periods of fishing (Bell *et al*., 2014, 2018) and the negative relationships could have resulted from competitive releases due to declining predators. "Biomass" was a particularly sensitive indicator as measured by the co-interaction, and generally, this interaction (along with the self-interaction) seemed to explain more of the time-series patterns than the lagged direct-interaction. There are many instances in which managers might only be interested in indicators that are "immediately" synchronized in their responses to the ecosystem, such as temperature as a proxy for tuna distribution (Hobday and Hartmann, 2006). Perhaps future measures of sensitivity should not be informed by lagged responses. Care should be taken when interpreting direct- versus co-interactions because direct-interactions are more likely to be correlated with estimates of self-interactions than co-interactions because they are parameterized using the same matrix and both involve lagged interactions. We choose to base sensitivity on both the co- and direct-interactions because if both were sensitive then it decreased the chances for incorrect inferences when the data are not collected synchronously (e.g., a survey early in the year versus a fishery later in the year) or processes of interest occur at slightly mismatched time scales.

Indicators that were well sampled and based on direct output from the Atlantis model (and in the real world would be sampled at sea with minimal modeling required) performed better than those that need analysts to perform additional calculations. Specifically, "inverse fishing", "biomass stability", "sustainable stocks", and "non-declining species" were the least sensitive indicators. The latter three were also the least representative. These indicators required calculations beyond dividing by biomass such as calculating the coefficient of variation. The calculations themselves were not difficult but led to truncated data in terms of the length of the time series or number of groups represented. Additionally, a lack of variability in some of the data used as input to the MAR models decreased model convergence relative to output from the most variable ecosystems. For instance, stopping all fishing led to very few overfished stocks and measurements

of "sustainable stocks" that were constant across the entire time series. Relatively stable environmental forcing in some models also led to low variability in the output and hindered convergence in the MAR models.

Results regarding TL indicators should be interpreted with caution. Outside of a simulation context "TL of the survey", "TL of the catch", and "trophic index" would have required additional calculations based on diet data. Instead, these indicators are essentially manipulated versions of biomass, catches, and a subset of catches, respectively, because TL information was either extracted from www.fishbase.org or supplied by model developers. Calculating TLs is difficult because TLs can vary with size, time, and space (Jennings *et al*., 2002; Vinagre *et al*., 2012). Fulton *et al*. (2005) found that TL indicators were sensitive to decreased quality of abundance information as well. Results presented here did not include observation error, but they do highlight a third source of bias, selectivity (Trenkel and Rochet, 2003). The selective nature of fishery and survey protocols to not capture all available species with an equal probability led to lower indicator representativeness than what was expected. Results highlight how "marine trophic index" might be easier to interpret than all-encompassing TL indicators because it uses an explicit cut-off TL level to focus on high TL organisms that tend to be more threatened and potentially better sampled (Pauly and Watson, 2005).

The poor performance of several indicators thought *a priori* to be good indicators highlights the need for more work on indicator performance. The *IndiSeas* list can act as a starting point but needs to be followed up with additional screening. In this regard, simulations could be used as a tool to check that indicators are sensitive to attributes that are hard to measure empirically, instead of relying on correlations. For example, two ecosystems (Chesapeake Bay and California Current) demonstrated a high dependence of predators on productivity from shelled invertebrates, which declined under ocean acidification scenarios; on the other hand, in the NE US this change in benthic productivity was shunted to predators via a deposit feeder group (e.g., amphipods and isopods) that would be hard to survey accurately.

Fulton *et al*. (2005) highlighted differences in indicator-attribute relationships under two axes of uncertainty not investigated here, data aggregation and observation error. The latter highlights the usefulness of indicators based on data that are easier to collect such as biomass of groups targeted by fisheries rather than TL, which depends on diet data. Fulton *et al*. (2005) suggest that TL may only be a good indicator when observation error is low. Our results are based on zero

observation error. We left the addition of observation error for future work because one goal of this work was to introduce MAR models as a framework for assessing indicator-attribute relationships not to provide management with a list of operational indicators. Results are still informative for managers and offer a first pass at indicator selection because indicators should at a minimum track attributes in an observation-error-free context. Future MAR models can be parameterized to account for observation error using a state-space formulation (Hinrichsen and Holmes, 2009; Holmes *et al.*, 2012). The discrepancies in indicator performance across levels of data aggregation that Fulton *et al.* (2005) found could be a result of using correlation because aggregating data to the ecosystem level may have introduced autocorrelation (Katz, 1988; Chapter 1). Data provided for this analysis were summaries of the entire ecosystem but Atlantis can output more fine-scale data. Future work could assess the tradeoffs of data aggregation versus estimating spatial autocorrelation within an ecosystem.

Marine ecosystems across the globe differ markedly, thereby decreasing the likelihood that a one-size-fits-all set of indicators that capture all attributes of all systems will be found (Coll *et al.*, 2010; Heymans *et al.*, 2014; Pikitch *et al.*, 2014). Ecosystems with similar maximum depth tend to behave similarly (Olsen *et al.*, 2018) and could be a tangible way to summarize indicator recommendations. Open-ocean ecosystems have low productivity and include several guilds in moderation. Shelf systems tend to be dominated by benthic productivity. Consequently, biodiversity indicators weighted by guild may be sufficient for open-ocean systems but may fail to reflect losses of key benthic species in shelf systems. Results corroborate previous work showing that indicators are ecosystem-specific rather than driver-specific (Shannon *et al.*, 2010, 2014; Coll *et al.*, 2016) and the need for multiple comparisons (Murawski *et al.*, 2010; Essington and Plagányi, 2014). This fortunately suggests that previous work testing indicator performance under fishing scenarios (e.g., Samhouri *et al.*, 2009) will be relevant for most future ecosystems influenced by other drivers. Not having to test ecosystems in terms of all potential drivers of change could mean faster implementation of EBM.

Overall, our results illustrate that simulation and MAR models, in combination, offer a useful method for characterizing indicator-attribute relationships prior to their use in management. Atlantis also offers the ability to test indicator performance under dynamic management based on indicator status (i.e., management strategy evaluation; Punt *et al.*, 2016). Thus, dynamic simulations could be used in the future to determine the consequences of using indicator-thresholds

such that management actions based on indicator status produce desired responses within a level of acceptable risk (Fay *et al*., 2015). Simulations could also facilitate determining threshold levels of measurement error that can be present before MARSS models fail to estimate a signal through the noise. These simulations would be helpful for determining which empirical data sets can be used to further vet indicator-attribute relationships.

2.5 TABLES

Table 2.1. Details of the eight Atlantis ecosystem models. Model-specific spatial extents (area) are divided into a specified number of polygons (*n* polygons) and each polygon is comprised of a number of vertical layers (*n* layers). The number of modelled groups included in the ecosystem with a trophic level greater than one (n groups) is used to calculate the percent of included groups that are "fish" (i.e., bony and cartilaginous fish groups), groups that are adequately "surveyed", "assessed" with fishery-independent or -dependent data providing a measurement of their status, and the number of groups "targeted" by fisheries. Note that the table is continued below, and the following abbreviations are used: Canada (CAN), European Union (EU), Mexico (MEX), and United States of America (USA).

ecosystem	abbreviation	type	region	country	reference
SE Australia	AustSE	ocean	SW Pacific	Australia	Fulton *et al*., 2005
California Current	CalCu	ocean	E Pacific	CAN, MEX, USA	Marshall *et al*., 2017
Chesapeake Bay	CAM	brackish	W Atlantic	USA	Ihde and Townsend, 2017
N Gulf of California	GOC	gulf	NE Pacific	MEX	Morzaria-Luna *et al*., 2013
Gulf of Mexico	GoMex	gulf	Gulf of MEX	Cuba, MEX, USA	Ainsworth *et al*., 2015
Guam	Guam	reef	W Pacific	USA	Weijerman *et al*., 2016a
NE US	Neus	ocean	NW Atlantic	USA	Link *et al*., 2010
Nordic & Barents Sea	NOBA	ocean	NE Atlantic	EU	Hansen *et al*., 2016

area (km^2)	*n* polygons	*n* layers	*n* fleets	initial year	*n* groups	% fish	% surveyed	% assessed	% targeted
3000000	71	5	33	2005	54	57	31	54	54
1500000	89	7	20	2013	73	49	59	44	40
8896	97	5	1	2009	39	54	41	31	41
57800	66	7	33	2008	51	53	18	10	73
564200	66	7	21	2012	78	62	46	42	47
110	55	2	7	1985	32	59	47	0	47
293000	22	4	18	1964	37	51	70	76	68
3700000	60	7	22	1981	45	44	44	36	36

Table 2.2. Ecological indicators selected by the Indicator of the Seas (*IndiSeas*) working group to track the direct and broader impacts of fishing on exploited marine ecosystems (Coll *et al.*, 2016).

Abbreviation	Label	Definition
BS	Biomass stability	1 / Coefficient of variation (10-year based on TB)
invF	Inverse fishing	TB/catch
IVI	Mean vulnerability	Biomass-weighted mean intrinsic vulnerability index of fish catches (Cheung *et al.*, 2005)
LG	Fish size	Biomass-weighted mean max length of surveyed fish
LS	Life span	Biomass-weighted mean max life span of surveyed fish
MTI	Trophic index	Biomass-weighted mean trophic index of catch (Pauly and Palomares, 2005)
NDES	Non-declining species	Proportion of non-declining exploited species that have an assessment of their status (Kleisner *et al.*, 2015)
PF	Predators	Proportion of survey biomass that is predatory fish
SS	Sustainable stocks	Proportion of non-fully exploited stocks (i.e., biomass > 60% of unfished biomass) of those that could be assessed
TB	Biomass	Biomass of surveyed groups
TLc	Trophic level of catch	Biomass-weighted mean trophic level of catch
TLs	Trophic level of survey	Biomass-weighted mean trophic level of survey

Table 2.3. Proportion of positive and negative co-interactions out of the total number of indicator-attribute relationships investigated to determine indicator representativeness for each ecosystem (column; i.e., one indicator-attribute combination across all drivers of change). Positive proportions are in the top panel and negative proportions are in the bottom panel. See Table 2.1 for region acronyms and Table 2.2 for indicator acronyms. Proportions greater than 0.5 are highlighted in gray. The sum for a given row/column combination across the panels may not always equal one because estimates were only considered positive or negative if their confidence interval did not cross zero.

sign	indicator	region									
		AustSE	AustSEDE	CalCu	CAM	GOC	GoMex	Guam	Neus	NeusDE	NOBA
Positive	BS	0.00	0.17	0.29	0.14	0.19	0.24	0.00	0.00	0.00	1.00
	invF	0.95	1.00	0.95	0.86	0.62	0.95	0.00	0.17	1.00	1.00
	IVI	0.62	0.83	0.05	0.00	0.81	0.38	0.94	0.28	0.00	0.17
	LG	0.90	0.33	0.24	1.00	0.76	0.29	0.89	1.00	1.00	1.00
	LS	1.00	1.00	1.00	1.00	1.00	0.67	1.00	1.00	1.00	1.00
	MTI	0.05	0.00	0.05	0.52	0.00	0.81	0.00	0.78	0.00	0.00
	NDES	0.95	1.00	0.95	0.95	0.52	0.95	0.94	0.89	1.00	0.00
	PF	1.00	1.00	1.00	0.95	0.33	0.43	0.00	1.00	1.00	1.00
	SS	0.00	0.00	0.95	0.19	0.76	0.67	0.56	0.00	0.00	0.00
	TB	0.90	0.67	0.81	0.71	0.90	0.86	0.00	0.89	0.00	1.00
	TLc	0.10	0.00	0.00	0.00	0.00	0.81	0.11	0.72	0.17	0.00
	TLs	0.81	0.00	1.00	0.00	0.05	0.33	0.17	1.00	1.00	1.00
Negative	BS	0.05	0.00	0.00	0.00	0.00	0.76	0.00	0.22	0.00	0.00
	invF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00
	IVI	0.05	0.00	0.19	0.57	0.00	0.29	0.00	0.00	0.00	0.00
	LG	0.05	0.00	0.00	0.00	0.00	0.48	0.00	0.00	0.00	0.00
	LS	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.00
	MTI	0.24	0.00	0.00	0.00	0.05	0.05	0.00	0.00	0.00	0.00
	NDES	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	PF	0.00	0.00	0.00	0.00	0.05	0.33	0.00	0.00	0.00	0.00
	SS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	TB	0.00	0.00	0.00	0.00	0.00	0.10	0.06	0.00	0.00	0.00
	TLc	0.10	0.00	0.00	0.76	0.00	0.10	0.00	0.00	0.00	0.00
	TLs	0.10	0.33	0.00	0.90	0.10	0.57	0.00	0.00	0.00	0.00

2.6   FIGURES
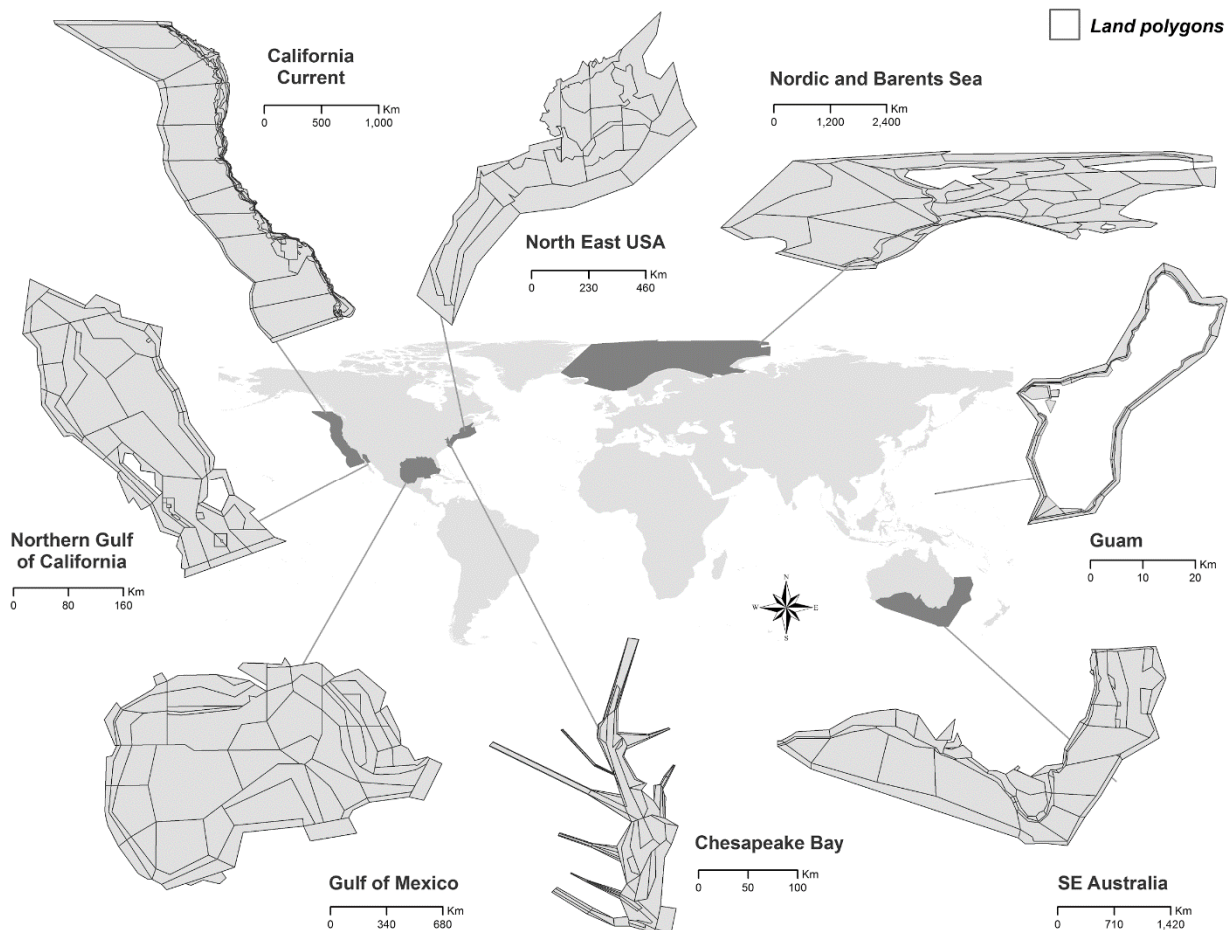
Figure 2.1. Spatial extents and polygon configurations for the eight Atlantis models (see Table 2.1 for more details).
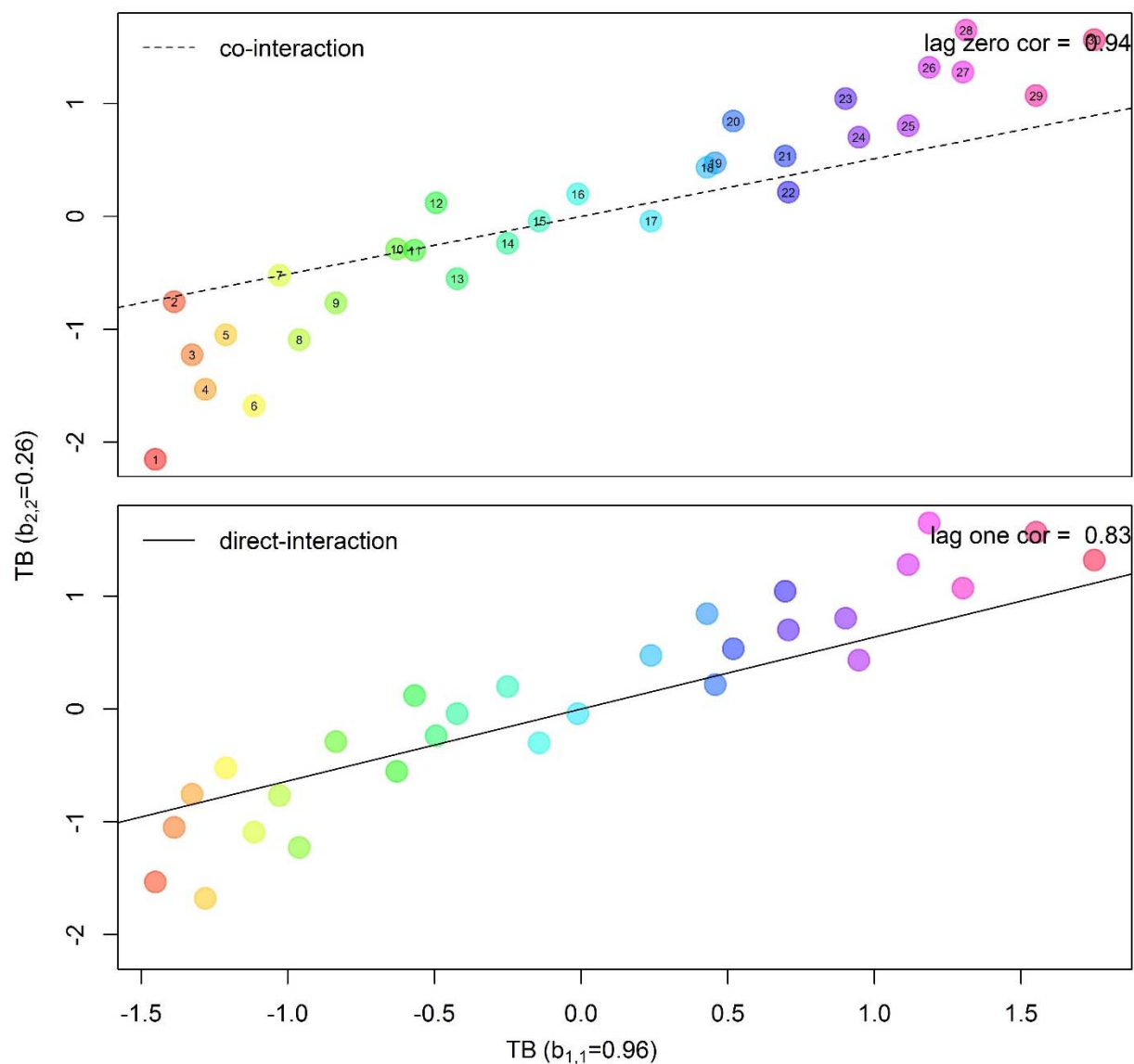
Figure 2.2. Standardized time series of the "total biomass" (TB) in the survey (y-axis; indicator) versus "TB in the ecosystem" (x-axis; attribute) plotted against each other at the same time step (upper panel) and with the indicator lagged by one year (lower panel) for the SE Australia ecosystem subject to base case (BC) conditions. The lines depict the slope of the co- (dashed) and direct-interactions (solid) and they are not intended as a regression on the points. The values in the upper right corners are the cross-correlation coefficients of the time series at the plotted lags. Autocorrelation of the indicator (b_{22}) and the attribute (b_{11}) are noted in parentheses in the axes labels. The points are colored according to their location in the time series with printed time series positions in the upper panel as well.
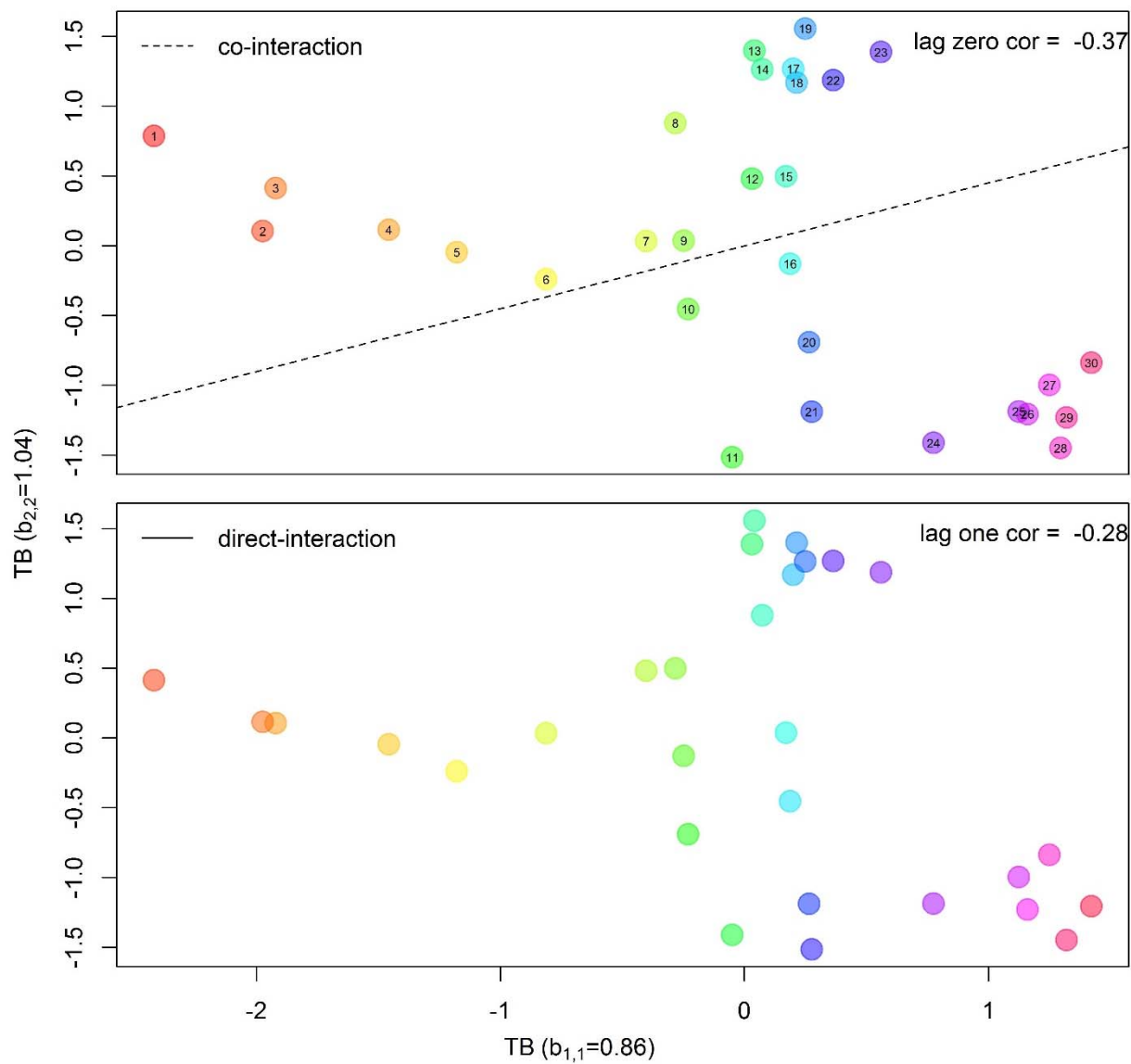
Figure 2.3. Same as Figure 2.2, except for the Chesapeake Bay ecosystem.
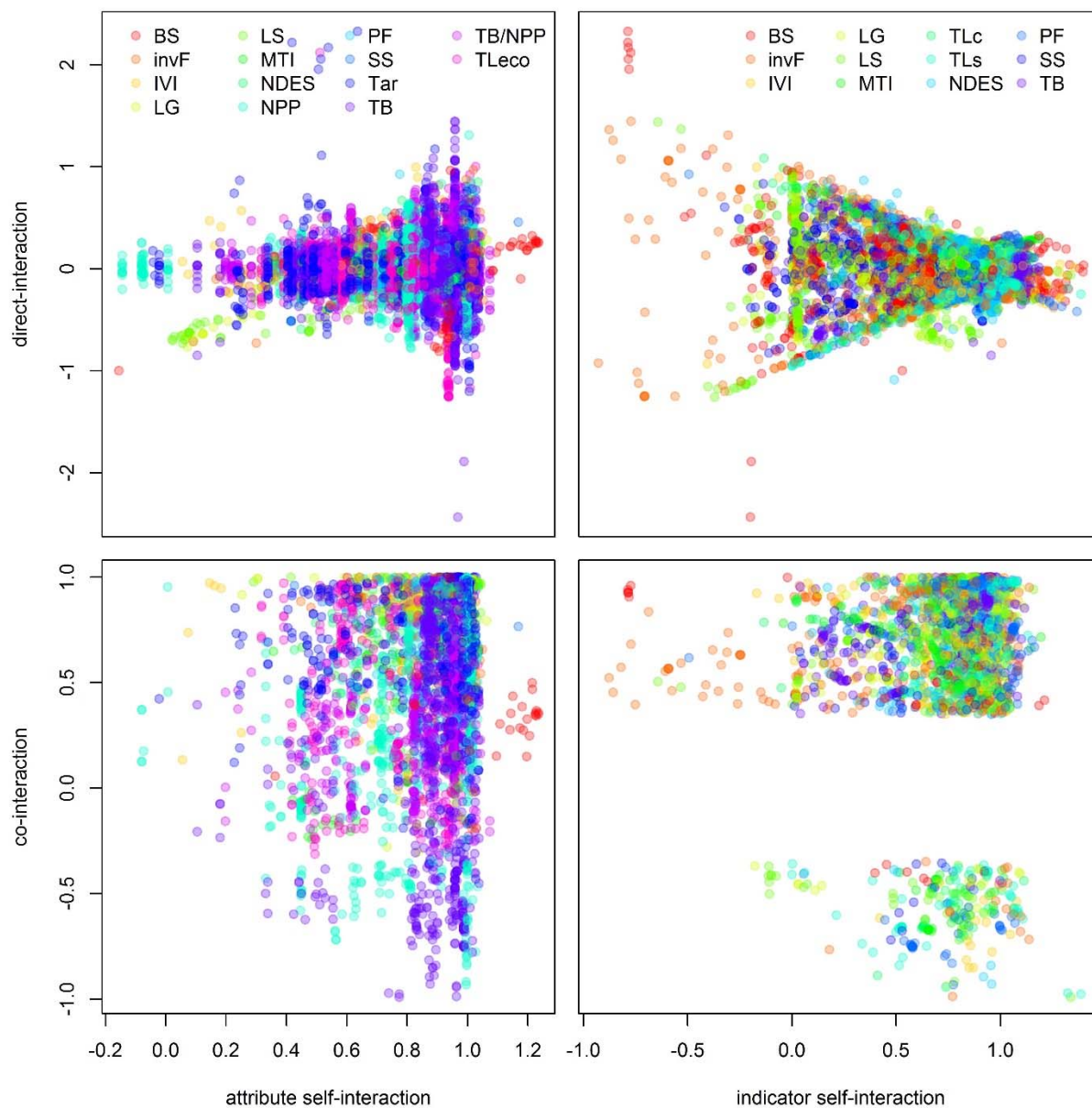
Figure 2.4. Direct- and co-interactions versus self-interactions (top and bottom row, respectively) for the attribute (left column) and the indicator (right column). Colors are used to display particular attributes and indicators. See Table 2.2 for acronyms. Parameter estimates that were not statistically different than zero were excluded and led to the horizontal gaps in the plots.
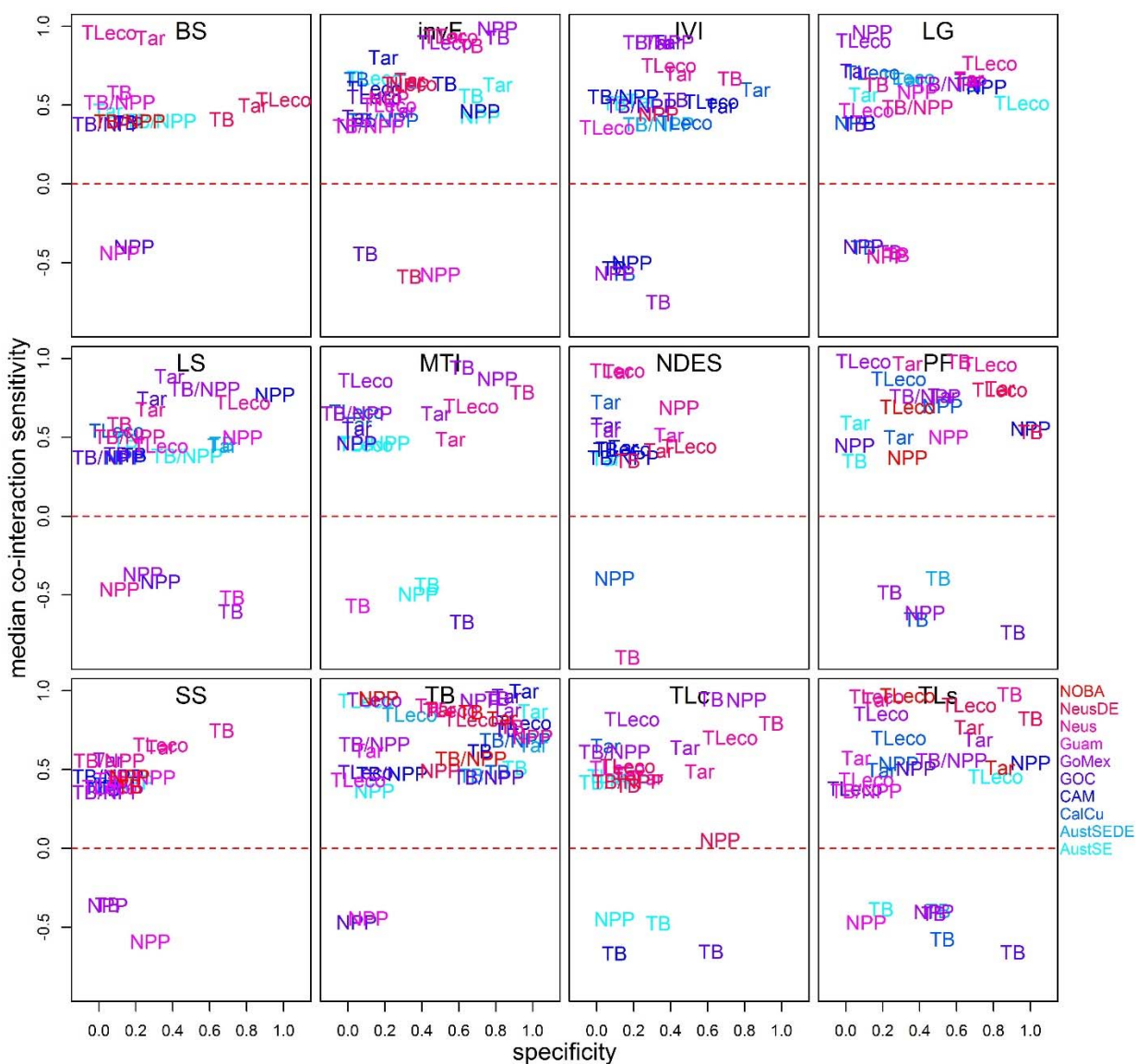
Figure 2.5. The median co-interaction across drivers of change of those that were sensitive versus the proportion of indicator-attribute relationships that were sensitive for a given indicator-attribute combination out of all investigated combinations for a given ecosystem (colors). The dashed horizontal line at zero was included to distinguish between positive and negative interactions. Refer to Table 2.1 for region abbreviations and Table 2.2 for acronyms. Attributes are abbreviated as follows: net primary productivity (NPP), total biomass (TB), mean TL of the ecosystem (TLeco), TB divided by NPP (TB/NPP), and biomass of groups targeted by fisheries (Tar).

Chapter 3. JUSTIFYING THE INCLUSION OF HABITAT COVARIATES IN SPATIOTEMPORAL INDEX STANDARDIZATION

Abstract

Estimates of current abundance are a primary need in managing marine fishes and invertebrates? Design-based estimators, which make inferences according to the randomness of the sampling protocol, have classically been used to provide a relative index of abundance. More recently, the use of spatiotemporal models has increased because of their ability to account for spatial heterogeneity and improve precision relative to design-based estimators. In theory, the inclusion of habitat covariates (e.g., depth) should also improve precision. We used a simulation experiment to evaluate the bias and precision of results from spatiotemporal index-standardization models when the true process was and was not governed by a habitat covariate. The simulation was conditioned on fits to data on darkblotched rockfish (*Sebastes crameri*) because of their known increased preference for deeper water with age and limited migration after recruiting to the fishery. In general, indices of abundance were estimated without bias, although parameters themselves were not necessarily unbiased. Incorrectly including a covariate when it did not govern the true process was less problematic than not including the covariate when it should have been included. In general, the Akaike Information Criteria correctly identified model misspecification. Results suggest that habitat covariates can improve the precision of abundance estimates. However, depth may not be the best variable to adequately capture ecological processes that regulate the distribution and abundance of groundfish off the US West Coast.

3.1 INTRODUCTION

The successful management of marine fishes and invertebrates depends, at least in part, on knowledge of abundance. Catch rates can be used to inform trends in abundance. However, catch rates by themselves are seldom proportional to abundance and need to be standardized (Maunder and Punt, 2004). Standardization can account for factors such as differences among vessels in fishing power (Helser *et al*., 2004; Robins *et al*., 1998), the depth or location at which fish are caught (Bigelow and Maunder, 2007; Cao *et al*., 2017), and sampling intensity (Cochran, 1977).

The resulting standardized index of abundance can be used directly by management or as input when fitting a stock assessment model.

Abiotic (e.g., depth and sediment type) and biotic (e.g., presence of coral or algae) habitat characteristics are fundamental drivers of local abundance. For example, subsurface poleward flow defines habitat for Pacific hake (*Merluccius productus*) in the California Current ecosystem (Agostini *et al*., 2006). Flatfish distributions in the eastern Bering Sea are related to temperature, which acts as a proxy for the presence of the cold pool (Kotwicki *et al*., 2005; Mueter and Litzow, 2008; Spencer, 2008). Local abundances can vary at seasonal or finer temporal scales because of changing habitat conditions. For example, water levels of wetlands offer essential breeding, rearing, and feeding grounds for many species and can vary within and between seasons (Johnson *et al*., 2005). Habitats can also vary spatially, and the most complex habitats vary spatially and temporally (Hinton and Maunder, 2004). Data collection typically occurs across multiple habitat types, and including habitat covariates in index-standardization models is most important when covariates explain a substantial portion of variability in catch rates (e.g., distance to rocky outcrop for darkblotched rockfish (*Sebastes crameri*); Shelton *et al*., 2014). Unfortunately, habitat information is not always available for all sampling locations, or even stratifications of the raw data, limiting its inclusion in traditional index-standardization models.

Recently, spatiotemporal models have been used to standardize catch-rate data because of their ability to estimate less biased and more precise indices than either design-based or model-based approaches (Shelton *et al*., 2014; Thorson *et al*., 2015a, Thorson and Haltuch, in press). The smoothed surface representing spatial variation in catch rates (i.e., densities) assumes that densities at nearby sites are more similar than densities farther apart (i.e., spatial autocorrelation). Spatial autocorrelation can arise from exogenous or endogenous processes. Exogenous processes typically stem from spatial autocorrelation in underlying covariates (Koenig, 1999) and endogenous processes stem from life-history characteristics (Levins, 1969). This autocorrelation was historically seen as a statistical nuisance, but research is now showing that autocorrelation can have strong predictive power regardless of its generating process (Bahn and McGill, 2007; Doorman, 2007). For example, a spatiotemporal model applied to bottom trawl survey-data for 28 groundfish species off the US West Coast generated lower levels of uncertainty compared to a model-based approach (Thorson *et al*., 2015c). Increases in precision are also being seen for spatially-explicit age/length composition data (Thorson and Haltuch, in press).

Spatiotemporal models can incorporate information on habitat using covariates. However, the statistical gains from explicitly including habitat in a model that already accounts for spatiotemporal effects has yet to be examined. Here, we used an existing geostatistical index-standardization model (Thorson and Barnett, 2017) to estimate the effect of depth on the density of darkblotched rockfish off the US West Coast. Increasing length with depth suggests that biological processes that coincide with depth at least partially regulate distribution in darkblotched rockfish and, potentially, their growth (Hamel, 2008). A simulation experiment was conditioned using these estimates. The simulation facilitated the exploration of statistical properties of the model when the true densities were and were not governed by a habitat covariate (nominally termed "depth," however, the results apply to any habitat covariate). Quantitative descriptions of relationships between marine species and habitat covariates such as those presented here provide partial explanations for species distributions. Perhaps more importantly, simulation results provide inference on when we can best estimate those relationships when the truth is uncertain.

3.2   METHODS

3.2.1   *Vector-autoregressive spatiotemporal (VAST) model*

Vector-autoregressive spatiotemporal (VAST; Thorson and Barnett, 2017; www.github.com/James-Thorson/VAST) models predict variation in density for multiple locations (s) across time (t). Year-specific indices of abundance can then be calculated by weighting estimated densities by the area of the spatial domain. Additionally, the models can accommodate multiple categories facilitating the simultaneous analysis of multiple species or age classes (see Thorson and Barnett, 2017). For this application, we grouped all ages into a single category and fit the model to data for just one species, removing the need to model categories.

The approach utilized by VAST accommodates observations of zero and is often referred to as a delta- or hurdle-model (Martin *et al.*, 2005). The probability distribution for each catch-rate data point (b_i) is decomposed into two components, the probability of encounter, $p_1(s_i, t_i)$, and the expected catch rate given the species is encountered, $p_2(s_i, t_i)$, at location s_i and year t_i of the ith sample.

The probability of encounter,

$$\text{logit}(p_1(i)) = \beta_1(t_i) + \sum_{p=1}^{n_p} \gamma_1(t_i, p) X(s_i, t_i, p) + \eta_1(v_i) + \omega_1(s_i) + \varepsilon_1(s_i, t_i), \quad (3.1)$$

was modelled using a linear relationship between $p_1(i)$, which is the predictor for observation i at location s_i in year t_i, and a year-specific intercept, $\beta_1(t_i)$, and covariates. Habitat covariates, $X(x_i, t_i, p)$, can be included as an array of n_p covariates that explain variation in density for time t and location s. The matrix $\gamma_1(t_i, p)$ is the estimated impact of covariates. Here, $n_p = 2$ because we investigated linear and quadratic terms for the habitat covariate, depth and depth^2. Random variability in catchability, $\eta_1(v_i)$, among vessel-year combinations (v) was accounted for using a normally distributed random effect with a mean of zero and an estimated variance ($\sigma_{v_1}^2$).

Spatial, $\omega_1(s_i)$, and spatiotemporal, $\varepsilon_1(s_i, t_i)$, variation at location s_i follow spatially correlated stochastic processes represented by zero-mean Gaussian random fields that were modelled using stochastic partial differential equations (Lindgren et al., 2011). These spatial fields facilitate modelling environmental and biological factors that are not directly included in the model but affect the density of the modelled species. The expected value, variance, and covariance at a set of fixed locations ($s = \langle x, y \rangle$) were assumed to follow multivariate normal (MN) distributions, $\omega_1(s) \sim MN(0, \sigma_{\omega_1}^2 \mathbf{R}_1)$ and $\varepsilon_1(s, t) \sim MN(0, \sigma_{\varepsilon_1}^2 \mathbf{R}_1)$. The easting and northing for each location are represented using x and y. Locations in s were determined by applying a k-means clustering algorithm to the empirical data. The resulting mesh, computed using the R-INLA software (Lindgren and Rue, 2015), was comprised of 250 knots distributed across the spatial domain such that their density was proportional to the density of the samples. The spatial ($\sigma_{\omega_1}^2$) and spatiotemporal ($\sigma_{\varepsilon_1}^2$) variance in p_1 scale the spatial correlation ($\mathbf{R}_1$) given the estimated decorrelation distance (κ_1). Correlation between two locations, s and $s + h$, was specified using a Matérn function, $\mathbf{R}_1(s, s + h) = \frac{1}{2^{\nu-1}\Gamma(n)} \times (\kappa_1|h\mathbf{H}|)^\nu \times K_\nu(\kappa_1|h\mathbf{H}|)$, with smoothness ($\nu$) fixed at 1.0 (Lindgren et al., 2011), a Bessel function (K_ν), and geometric anisotropy. Anisotropy is the tendency for correlations to decline faster in one direction than another and was estimated using a two-dimensional matrix ($\mathbf{H}$) with a determinant of 1.0.

Positive catch rates,

$$\log(p_2(i)) = \beta_2(t_i) + \sum_{p=1}^{n_p} \gamma_2(t_i, p)X(s_i, t_i, p) + \eta_2(v_i) + \omega_2(s_i) + \varepsilon_2(s_i, t_i), \quad (3.2)$$

were defined similarly to (3.1, except using different subscripts.

The probability of the data,

$$\Pr(b = B) = \begin{cases} 1 - p_1 & \text{if } B = 0 \\ p_1 \times Gamma(b = B; \sigma^{-2}, \exp(p_2)\sigma^2) & \text{if } B > 0 \end{cases}, \quad (3.3)$$

is a combination of the two components using logit and log links, respectively. The Gamma distribution, $Gamma(b; x, y)$, was used to represent the probability density function evaluated at b and with shape parameter x and positive scale parameter y (Cadigan and Myers, 2001). The Gamma distribution was parameterized in terms of the expected catch given that the species is encountered $\exp(p_2)$ and the coefficient of variation of measurement errors for positive catches (σ_m).

Parameter estimation was facilitated by maximizing the marginal likelihood of the fixed effects (intercepts, decorrelation distances, anisotropy parameters, and residual variation in positive catch rates) given the fitted data. The marginal likelihood was approximated using the Laplace approximation (Skaug and Fournier, 2006), which approximates the multidimensional integral of the joint likelihood (i.e., the product of the probability of random effects, given the fixed effects, and the probability of the data, given random and fixed effects), using Template Model Builder (Kristensen *et al.*, 2016). To improve computational efficiency, we used the stochastic partial differential equation approximation for all spatial and spatiotemporal random fields (Lindgren *et al.*, 2011) and a "predictive process" formulation wherein we estimate the value of spatial variables at 250 knots. It is assumed that variables are homogenous in the vicinity of each knot. A non-linear minimization routine available in the R statistical environment (R Core Team, 2018) was used to identify the maximum-likelihood estimate of the fixed effects given the gradient of the approximated marginal likelihood with respect to all fixed effects. We confirmed that all models had a maximum absolute gradient < 0.0001 and that the Hessian matrix was positive definite. Expected values of total relative abundance within a given year were corrected for bias to account for the nonlinear transformation of random effects (Thorson and Kristensen, 2016).

3.2.2 *Empirical data*

Darkblotched rockfish are found in the Pacific Ocean from the eastern Bering Sea to Santa Catalina Island, California (Figure 3.1; Eschmeyer *et al.*, 1983). Typically, adults rest on mud near rocky substrate in waters ranging from 25-550 m depth. Commercial catches began in appreciable quantities in the mid-1940s and peaked in 1966 (Wallace and Gertseva, 2018). Prior to the year 2000, the species was managed as part of a multispecies rockfish complex. The first single-species assessment for the darkblotched stock was conducted in 2000 and indicated it was overfished (Rogers *et al.*, 2000). Reduced catches, the implementation of Rockfish Conservation Areas, and

above average recruitment has since led to its recovered status as of 2017 (Wallace and Gertseva, 2018).

Data for darkblotched rockfish were obtained from the Northwest Fisheries Science Center, which has supported an annual bottom trawl survey to collect information on the distribution and abundance of groundfish since 2003. The survey covers depths from 55 to 1280 m, using standard gear and deployment methods aboard contracted fishing vessels (Keller *et al*., 2012). Data provided catch per unit effort for each sampled location. The number of locations sampled per year ranged from 494 to 758 (depending if one or two survey passes were conducted). Darkblotched rockfish are thought to be adequately sampled in the survey, such that the survey provides information used in the stock assessment model on the population trend from 2003 to present. On average, darkblotched rockfish were present in 17% of the locations sampled in a given year.

Two versions of the spatiotemporal index-standardization model were fit to the empirical data. The first version ignored habitat covariates and the second included depth as a fixed effect. Depth data (NMFS, 2013) were standardized to have a mean of zero and a standard deviation of one prior to finding the mean depth for each "knot". Quadratic functions of depth (i.e., depth^2) were also explored to investigate the importance of the covariate at intermediate depths. The knots were kept constant for both models and the fitting of simulated data (detailed below). The *VAST* package in R facilitated the estimation of parameters (see Table 3.1 for a list of estimated parameters). Marginal Akaike Information Criterion (AIC; Burnham and Anderson, 2002) was used to identify the model parameterization that most parsimoniously fit the data based on the marginal log likelihood and the number of estimated fixed effects.

3.2.3 *Simulation*

Simulations were used to evaluate the ability of spatiotemporal index-standardization models to (1) estimate the sign and magnitude of habitat covariates, (2) estimate spatial and spatiotemporal correlation parameters, (3) estimate a relative index of abundance, and (4) identify when the true process was and was not governed by a habitat covariate using standard model-selection techniques. One hundred simulated data sets were generated for each true state (i.e., with and without a habitat covariate) based on parameters that were estimated by fitting each estimation model to the empirical data (Table 3.1). Specifically, the simulated data sets were conditioned using estimated variance parameters, estimated depth effect (only when simulating data that

included a depth effect), and annual intercepts that had the same mean and standard deviation as the estimated set. One hundred data sets were included per scenario to ensure that the median parameter estimates across replicates were stable and that additional replicates did not alter estimation performance (Figure C.1). These two hundred data sets were fit to both estimation models leading to 400 estimated parameter sets. Each simulated data set had the same annual sample size and utilized the same sampling locations as the empirical data,

The overall design was factorial, where the true process generated by the operating model was and was not governed by depth and the estimation process did and did not include depth. The combination resulted in two misspecified and two correctly specified combinations of simulated data and estimation method. The operating model differed from the estimation method in that it utilized a delta-lognormal process ($N \sim$ Bernoulli(c_i); if N=0 then $c_i = 0$ and if N=1 then $c_i =$ lognormal$(\log(p_2) - \frac{\sigma_m^2}{2}, \sigma_m^2)$) to generate simulated data (c_i) rather than a delta-gamma process, thus ensuring that the estimation method could not fit the data perfectly.

Estimation performance was quantified using median absolute error ($MAE = |\hat{\theta} - \theta|$), where $\hat{\theta}$ and θ are the estimated and true values from the operating model used to generate the data, respectively. Bias in the trend of the index was assessed in two ways. In the first method, the difference was computed between the estimated log ratio of the first and last year and the true log ratio the first and last year. In the second method, log-linear models of the index, $\log(\hat{b}_{r,t}) = \alpha_r + \delta \log(b_{r,t}) + \varepsilon_{r,t}$, with a normal error structure ($\varepsilon_{r,t} \sim N(0, \sigma^2)$) were used to compute bias, where δ will equal one if the estimated index ($\hat{b}_{r,t}$) from replicate r in year t reflects changes in the true index ($b_{r,t}$). Hyperstable models will lead to $\delta < 1$ and the opposite will be true for hypersensitive models (Wilberg *et al.*, 2010). Estimates of one imply that the index is well calibrated. Coverage of the true index was assessed using 95% confidence intervals. The confidence interval is expected to include the true value 95% of the time. The best model, of the two investigated, was chosen using AIC (Burnham and Anderson, 2002). Differences in AIC of four correspond to an evidence ratio of 7.4 or an 88% chance that one model is better than the other, using a Bayesian interpretation of the likelihood and certain priors (Burnham *et al.*, 2011).

3.3 RESULTS

3.3.1 *Fits to empirical data on darkblotched rockfish*

The catch-rate data provided more support for a model that included depth over a model that lacked a habitat covariate for darkblotched rockfish off the US West Coast (ΔAIC=74.77). Both effects of depth in the catch-rate model were negative, suggesting a dome-shaped relationship as expected rather than an asymptotic one (Table 3.1). The effect of depth in the catch-rate model predicted a rate of change of -1.04 when depth is zero and a steepness/curvature of -0.64. This negative rate of change was stronger than that estimated for thirteen of fourteen other US West Coast groundfish species (results not shown). A stronger effect was estimated for longspine thornyhead (*Sebastolobus altivelis*).

The estimated trends in abundance indices were not greatly affected by the inclusion of depth. One effect of including depth was a reduction in the range of the confidence intervals (blue versus red in Figure 3.2). Another effect was decreased correlation of the residuals perpendicular to the coast (i.e., increases in κ_2 and changes to **H**; Table 3.1) for encounters and positive catch rates (Figure C.2). Including depth did not change the pattern in the residuals parallel to the coast for positive catch rates or change the ability of the estimation method to capture the overall shape of the dispersion in the positive catch-rate data (Figure C.3). Differences in the estimated indices were largely a result of differences in the estimated ranges and yearly intercepts because the changes in the random spatial and spatiotemporal effects were not large relative the estimated standard errors.

The empirical data supported the inclusion of quadratic depth effects. The addition of the two depth-squared terms decreased the AIC by 14.39 units relative to a model with only linear depth terms. Parameters not related to depth were relatively consistent between the linear and quadratic models and the resulting index was similar (Figure C.4). Both linear depth terms changed when depth-squared terms were included even though the independent variable was centered (Table 3.1). Centering was performed to reduce the correlation between the linear and quadratic terms. Correlations between linear and quadratic terms were common across other species as well (results not shown).

Increasing the number of knots increased the magnitude of most of the linear and quadratic depth effects (Table 3.1). Appropriately, it also decreased the range parameters that govern the

distance at which spatial correlation becomes negligible. It is unclear how to perform model selection between models that use a different number of spatial knots. However, the data supported the inclusion of linear and quadratic depth effects relative to a model without depth ($\Delta AIC=555.12$) when 500 knots were used. In this case, the trend of the index between the two models was almost identical, but the variance was smaller for the quadratic-depth model (results not shown). An increase in the number of knots did not always lead to a greater disparity between estimation methods with and without depth (results not shown).

3.3.2 *Simulation results*

Simulation results suggest that covariates affecting catch rates of marine fishes can be estimated using spatiotemporal models. Linear covariates related to habitat were estimated without bias for the positive catch rate model and tended to be positively biased for the encounter model component (points should be centred on the crosshairs in Figure 3.3). The squared effects of depth were biased towards zero (Figure 3.4). The bias was not removed when the simulated effect was modelled using 500 knots (Figure C.5). The estimated effect of depth was small, if not zero, when the true effect was zero for both linear and quadratic effects. Parameters became biased when there was a mismatch between the operating model that included depth and the estimation method that did not include a depth parameter.

The trend in the derived index of abundance as reflected by the difference between first and last values (on the log scale) was unbiased (Figure 3.5). Model misspecification did not lead to an increase in bias as might be expected. Instead, the model compensated for the misspecification by increasing the standard deviation of the spatial field of the encounter model (Figure 3.6). The bias in the range parameter of the encounter model increased when the model was misspecified and appears to be correlated with the bias in the standard deviation of the spatial field for encounters.

Indices of abundance from the spatiotemporal model were neither hyperstable nor hypersensitive (i.e., $\delta = 1$). Stability was similar for all investigated models regardless of how the operating model was defined or if the estimation method was misspecified. Examples of estimated trajectories are provided in the supplementary material showing the similarity in the outcomes of the estimation methods despite the model misspecification (Figure C.6).

3.4 DISCUSSION

Simulations were used to infer the costs and benefits of adding a habitat covariate to spatiotemporal index-standardization models that account for complex correlations in spatial structures. Previous work (e.g., Cao *et al.*, 2017; Thorson, 2017) predicted that their inclusion would result in better capture of the mean response by removing nonstationarity without explicitly including nonstationary Gaussian random fields (Fuglstad *et al.*, 2015). Results confirmed that habitat covariates can help explain spatial variation in catch rates and may decrease the variance of the derived index (see Figure 3.2). However, including a habitat covariate did not significantly decrease the variation in the spatial or spatiotemporal fields, even when AIC supported the additional parameters (results not shown).

The main goal in standardizing data is to provide a relative index of abundance. In doing so, analysts hope to remove the variation in the data that is not attributable to changes in the true population size. It was unclear prior to conducting this simulation experiment if the modelling framework could discern habitat effects from residual spatial variation. Simulations indicated that habitat covariates should be included and that their inclusion is particularly useful when the true variance of the spatial field is small. For the case study, depth explained longitudinal variation in darkblotched rockfish encounters and positive catch rates and led to slightly reduced standard errors of the derived index of abundance. We only tested a single covariate and leave it to future work to investigate multiple habitat covariates. Investigating multiple covariates could be complicated given that such covariates tend to be correlated and spatially autocorrelated themselves. Nevertheless, incorrectly including depth in the model when only one covariate was included in the operating model used to simulate data did not lead to bias in the derived index of abundance. A result that could change if the sampling scheme was not random with respect to the habitat covariate.

The presence of resources within a habitat is theorized to directly relate to abundance (i.e., ecological-niche theory; Hutchinson, 1957), but the exact mechanism regulating darkblotched rockfish distribution with depth remains unknown. Here, depth was included as a proxy for unmeasured habitat characteristics (e.g., vertical current, primary production, diel migration, community composition, etc.) that relate to ecological process affecting density, such as benthic-pelagic coupling (Giraldo *et al.*, 2017; Woodland and Secor, 2013). The relationships observed

between catch rates and depth could be the result of spatial distributions in prey. However, darkblotched rockfish can be opportunistic predators with the ability to engage in vertical, diel migrations to feed on the bottom during non-daylight hours (Hannah *et al.*, 2005). Alternatively, previous research has shown that older, mature individuals typically inhabit deeper waters (Hamel, 2008) and the negative relationship may reflect the age distribution of the population (Wallace and Gertseva, 2018). Future empirical work should investigate relationships between depth and age-specific indices of abundance (e.g., Kristensen *et al.*, 2014; Thorson and Haltuch, in press). We also noticed differences in the sign of estimated depth covariates with changes in the number of knots used in the analysis (results not shown).

It is possible that depth does not have major ecological significance and the model instead may be capturing other sources of variability in catch rates. The likelihood of the previous being true was a legitimate concern for the case study because of the large scale of the analysis. Estimated ranges (i.e., distance at which spatial correlation decreased to ~ 10%) were smaller when depth was included as a covariate than when it was not included. These ranges were larger than those estimated from a similar model applied to species with limited dispersal (e.g., Lecomte *et al.*, 2013) but smaller than the average dispersal distance of the species (Hamel, 2008). Larger true ranges resulted in less bias in misspecified models than when the true range was based on fits to the empirical data (results not shown), implying that including habitat covariates may be increasingly important for more sedentary species. Alternatively, it could be that the range parameter is not well estimated and would benefit from additional sampling to increase the number of data points that inform each knot or such that more knots can be modelled (Fuglstad *et al.*, 2018). More knots would decrease the feasibility of running the model, which already push the boundaries of available memory on standard desktop computers. Future work should investigate the benefit of decreasing the bias in the range parameter versus adding habitat covariates, particularly as the variance of the habitat data increases with respect to a given knot or how to summarize multiple measures of covariates per knot, where currently the mean is used.

The structure of spatiotemporal index-standardization models allows for flexibility in several ways. For example, other distributions can be used for positive catch rates, such as the lognormal distribution. Here, the gamma distribution was used because of greater efficiency relative to the lognormal distribution, which is known to be sensitive to deviations from lognormality (Syrjala, 2000) and misspecification (Firth, 1988). Zero-inflated negative binomial (e.g., Brodziak and

Walsh, 2013) and compound Poisson-gamma distributions (e.g., Lecomte *et al.*, 2013; Thorson, 2017) could also be explored. The former accommodates more zeros than expected under a negative binomial distribution and the latter mimics the process of capturing clusters of biomass. Choices regarding distributional assumptions typically involve hypotheses about the mean-variance relationship of the data, which we assume to be consistent, and inertia, which in fisheries research tends to favor delta-models (Steffanson, 1996). Nevertheless, the use of the delta-gamma model limited the analysis in terms of comparing the estimated parameters across the two components of the model because the logit-link used to model encounters requires the explanation of the expectation in terms of odds ratios compared to percent change as for positive catch rate model. This limitation becomes increasingly important when the signs of the effects are inconsistent across model components (e.g., Rubec *et al.* (2016) that showed depth was positively and negatively related to the occupancy and catch-rate models, respectively, for pink shrimp, *Farfantepenaeus duorarum*). The compound Poisson-gamma and Poisson-link models eliminate this problem by using a single model to both account for the probability of encounter and the number observed (Thorson, 2017). Thus, shared inference with respect to relationships between the habitat covariate and encounters and positive counts would result from the use of a single parameter. In contrast, under the delta-model shared inference across the model components is only possible using an additional covariance matrix, further decreasing parsimony (e.g., Thorson and Ward, 2013). The compound Poisson-gamma model is more efficient compared to the Poisson-link model, but limitations in its computational feasibility have yet to allow for its use in spatiotemporal models (Thorson, 2017).

In general, simpler models are often preferred over highly parameterized models for management purposes because they are easier to explain (Ludwig and Walters, 1985) and better at predicting. Nevertheless, simple linear or quadratic relationships may result in poor approximations of the impact of habitat on density (Harris, 2015). Bias is expected to decrease with increasing model complexity, however, often at the expense of increased imprecision (Burnham and Anderson, 2002). The tradeoffs in assuming simple versus complex relationships between habitat covariates and density in spatiotemporal index-standardization models has yet to be formally investigated. However, Shelton *et al.* (2014) suggest that the inclusion of an additional spatiotemporal effect can accommodate remaining nonlinear and unmeasured habitat associations. Therefore, if the primary covariate that governs species density does so in a nonlinear manner

(Sadykova *et al*., 2017), it may be best to investigate other weaker but linear relationships. Future studies should investigate methods used to generate the squared covariate data to decrease correlation between the two terms. We investigated depth because it is commonly available for most species from bathymetric charts or survey tows. Other commonly-used proxies for ecological processes include temperature as a proxy for the availability of thermoregulated prey (e.g., Maravelias and Reid, 1997), thermocline depth-to-bottom-depth ratio as a proxy for water column production (e.g., Rooper and Martin, 2011), and presence of chlorophyll *a* as a proxy for phytoplankton biomass and oceanic productivity (e.g., Heenan *et al*., 2016).

3.5   TABLES

Table 3.1. Parameter names and their corresponding symbols used in the spatiotemporal index-standardization model. Four versions of the model were fit to the empirical data for darkblotched rockfish. Estimates (est.) of fixed and random effects are given for each version. Models increase in complexity from left to right, no depth, linear depth terms, linear and squared depth terms, and linear and squared depth terms estimated for 500 knots. Standard errors are provided in parentheses. Standard errors are approximate for models with degenerate random effects. Variances of spatial and spatiotemporal fields are reported as standard deviations (SD).

Name	Symbol	Dimension	Est. (se)	Est. linear (se)	Est. quadratic (se)	Est. quadratic 500 knots (se)
Sample	i					
Location	s	250 or 500				
Year (2003-2015)	t	13				
Catch data for survey sample i	b_i					
Measured covariates for sample i	$\mathbf{X}_i$					
Fixed effects in p_1	$\boldsymbol{\beta}_1$	13x1				
Fixed effects in p_2	$\boldsymbol{\beta}_2$	13x1				
SD of random spatial effects affecting p_1	σ_{ω_1}		4.45 (0.66)	3.64 (0.52)	3.65 (0.56)	3.35 (0.43)
SD of random spatiotemporal effects affecting p_1	σ_{ε_1}		0.16 (0.21)	0.00 (0.34)	0.00 (0.57)	0.11 (0.29)
SD of random spatial effects affecting p_2	σ_{ω_2}		1.71 (0.19)	1.77 (0.33)	1.71 (0.32)	1.23 (0.17)
SD of random spatiotemporal effects affecting p_2	σ_{ε_2}		1.31 (0.09)	1.87 (0.30)	1.85 (0.30)	1.56 (0.17)
Random spatial effects affecting p_1	$\omega_1(s)$					
Random spatiotemporal effects affecting p_1	$\epsilon_1(s,t)$					
Random spatial effects affecting p_2	$\omega_2(s)$					
Random spatiotemporal effects affecting p_2	$\epsilon_2(s,t)$					
Depth effect for p_1	$\gamma_{1,1}$		NA	1.70 (0.36)	2.11 (0.41)	2.22 (0.31)
Depth effect for p_2	$\gamma_{2,1}$		NA	-1.45 (0.17)	-1.04 (0.31)	-0.75 (0.25)
Depth squared effect for p_1	$\gamma_{2,2}$		NA	NA	-1.62 (0.41)	-1.72 (0.35)
Depth squared effect for p_2	$\gamma_{2,2}$		NA	NA	-0.64 (0.42)	-1.24 (0.30)
Anisotropy	$\mathbf{H}$	2x2				
Spatial scale in p_1	κ_1		0.02 (0.18)	0.03 (0.21)	0.03 (0.22)	0.03 (0.18)
Spatial scale in p_2	κ_2		0.07 (0.24)	0.24 (0.28)	0.24 (0.29)	0.26 (0.19)
Dispersion in the probability density function of positive catch rates	σ_m		1.26 (0.03)	1.25 (0.02)	1.25 (0.02)	1.18 (0.03)
SD vessel year effect affecting p_1	σ_{v_1}		0.00 (0.88)	0.00 (0.08)	0.00 (0.09)	0.00 (0.36)
SD vessel year effect affecting p_2	σ_{v_2}		0.33 (0.07)	0.27 (0.07)	0.27 (0.07)	0.24 (0.07)

3.6   FIGURES
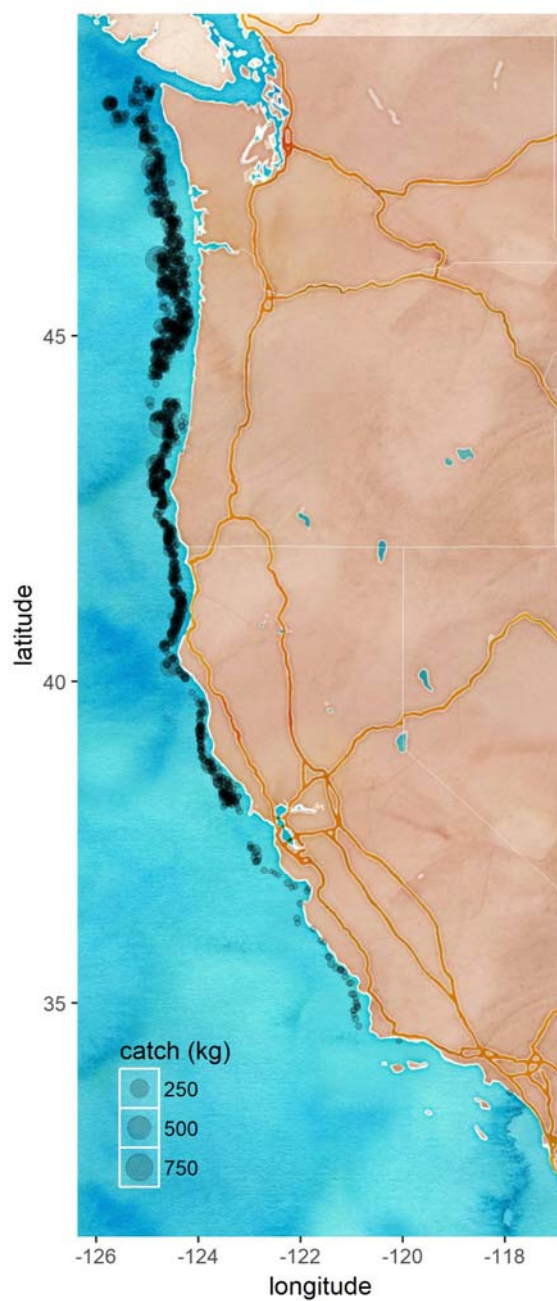
Figure 3.1. Map of darkblotched rockfish (*Sebastes crameri*) catches (kg) collected by the Northwest Fisheries Science Center during the groundfish bottom trawl survey. Coordinates are in decimal degrees.
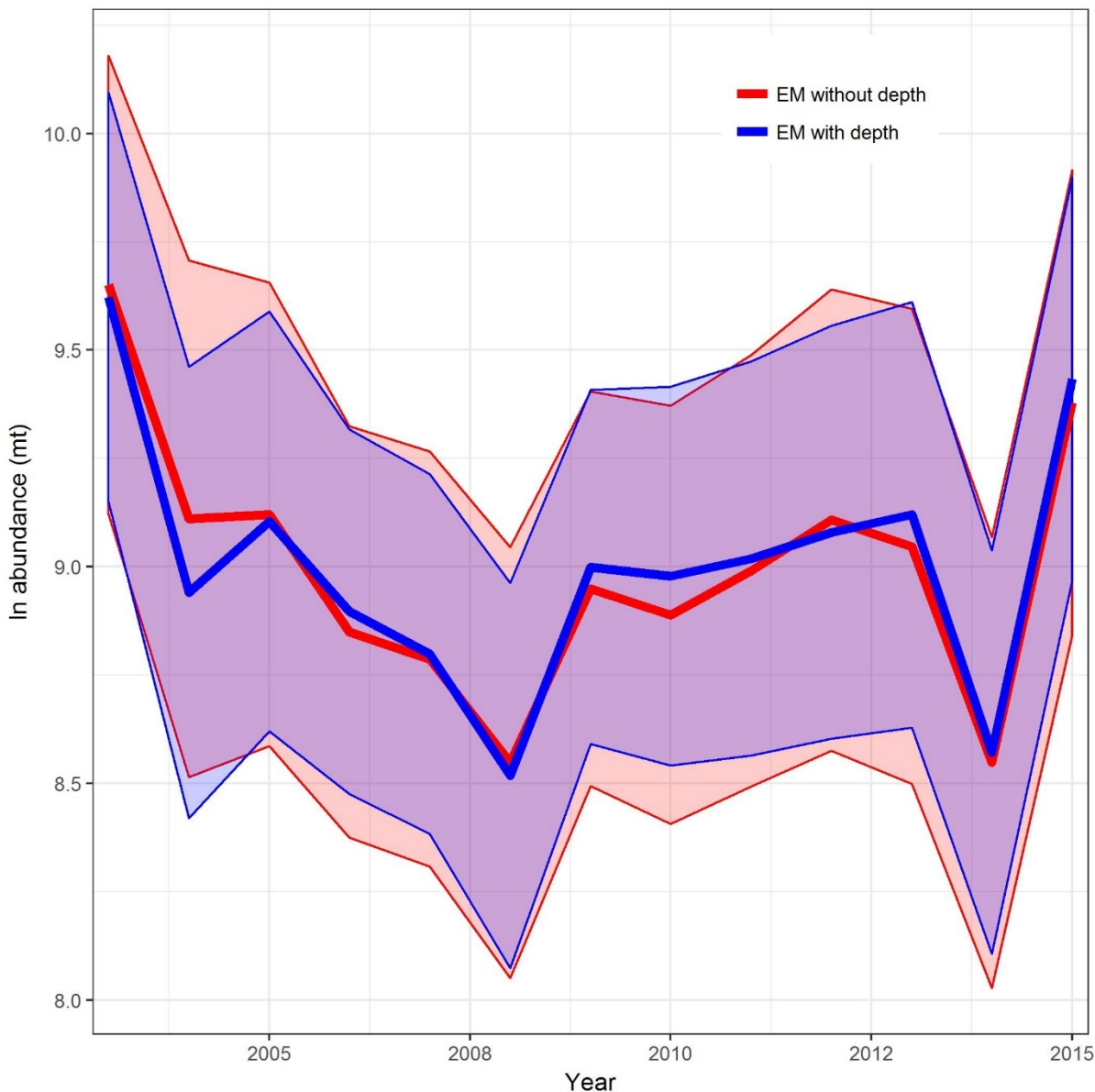
Figure 3.2. Comparison of derived indices of abundance (mt) in log space for darkblotched rockfish from 2003 to 2015. Data were fit to two estimation methods (EMs), one that included linear and quadratic depth terms (red) and one that did not include depth (blue). Polygons show the 95% confidence intervals for each model, which are purple when they overlap.
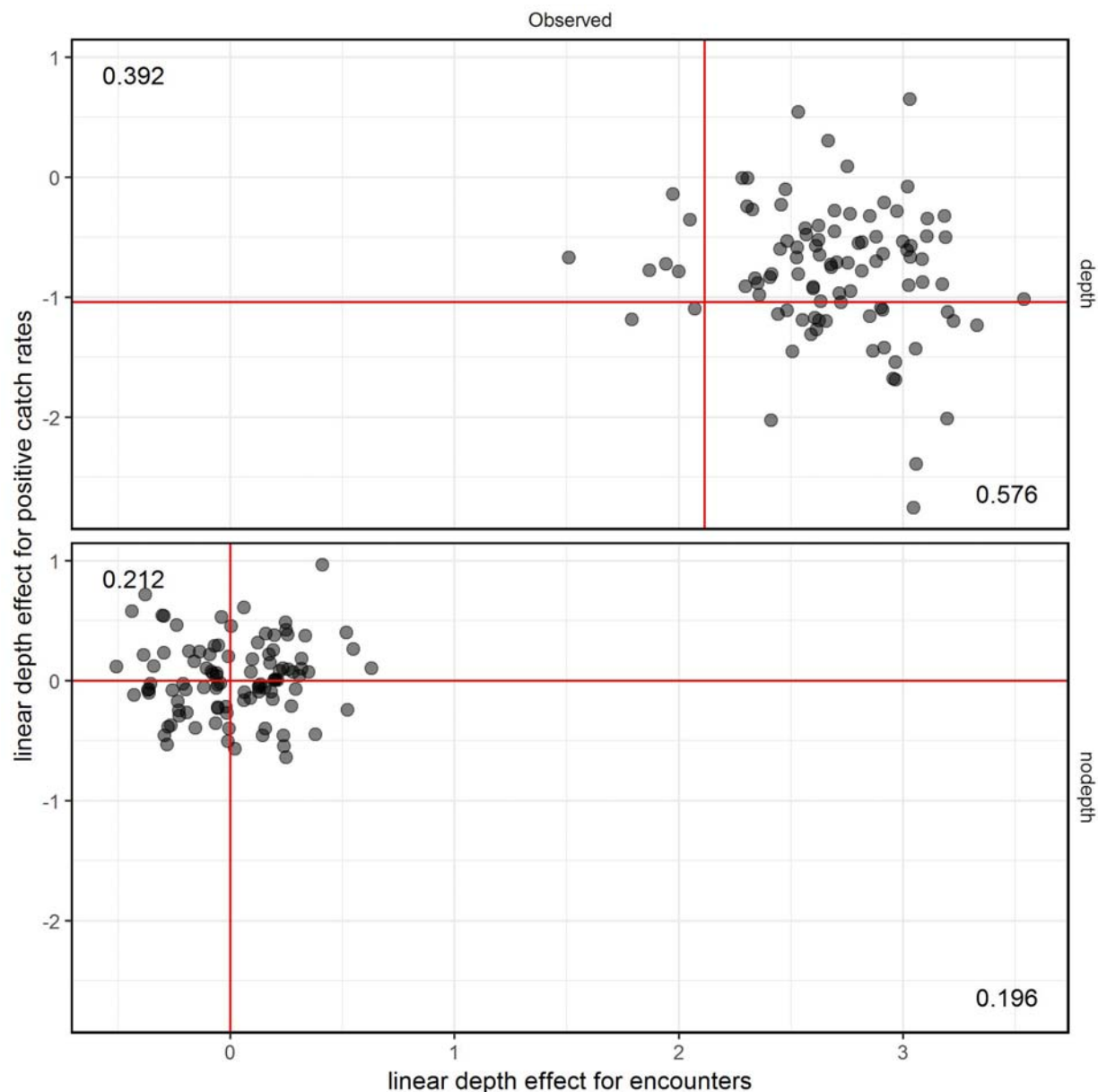
Figure 3.3. Linear effects of depth as estimated by each component of the delta-model (encounters on x-axis and positive catch rate on y-axis). Each point is a replicate of an estimation method fitted to simulated data that was (top row; "depth") and was not (bottom row; "nodepth") governed by depth. The median absolute error (MAE; median|estimated – true; printed near axis|) was smaller when the data were governed by depth. Unbiased results are located on the red cross-hairs.
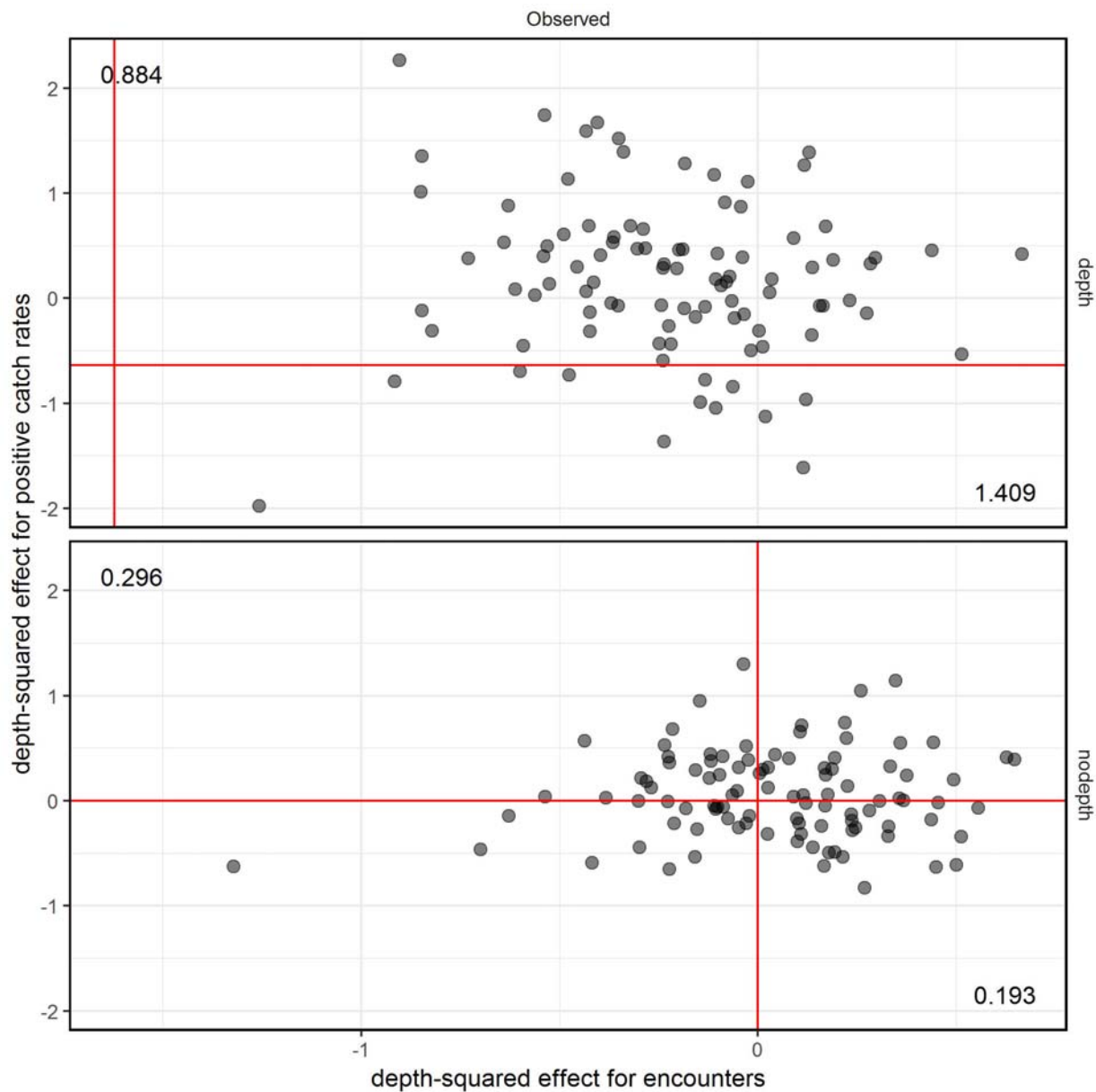
Figure 3.4. Simulation results illustrating bias in estimates of squared effect of depth for each model component (encounters on x-axis and positive catch rate on y-axis). Each point is a replicate of an estimation method fitted to simulated data that was (top row; "depth") and was not (bottom row; "nodepth") governed by depth. The median absolute error (MAE; median|estimated – true|) for each parameter is printed near its respective axis. Unbiased results are located on the red cross-hairs.
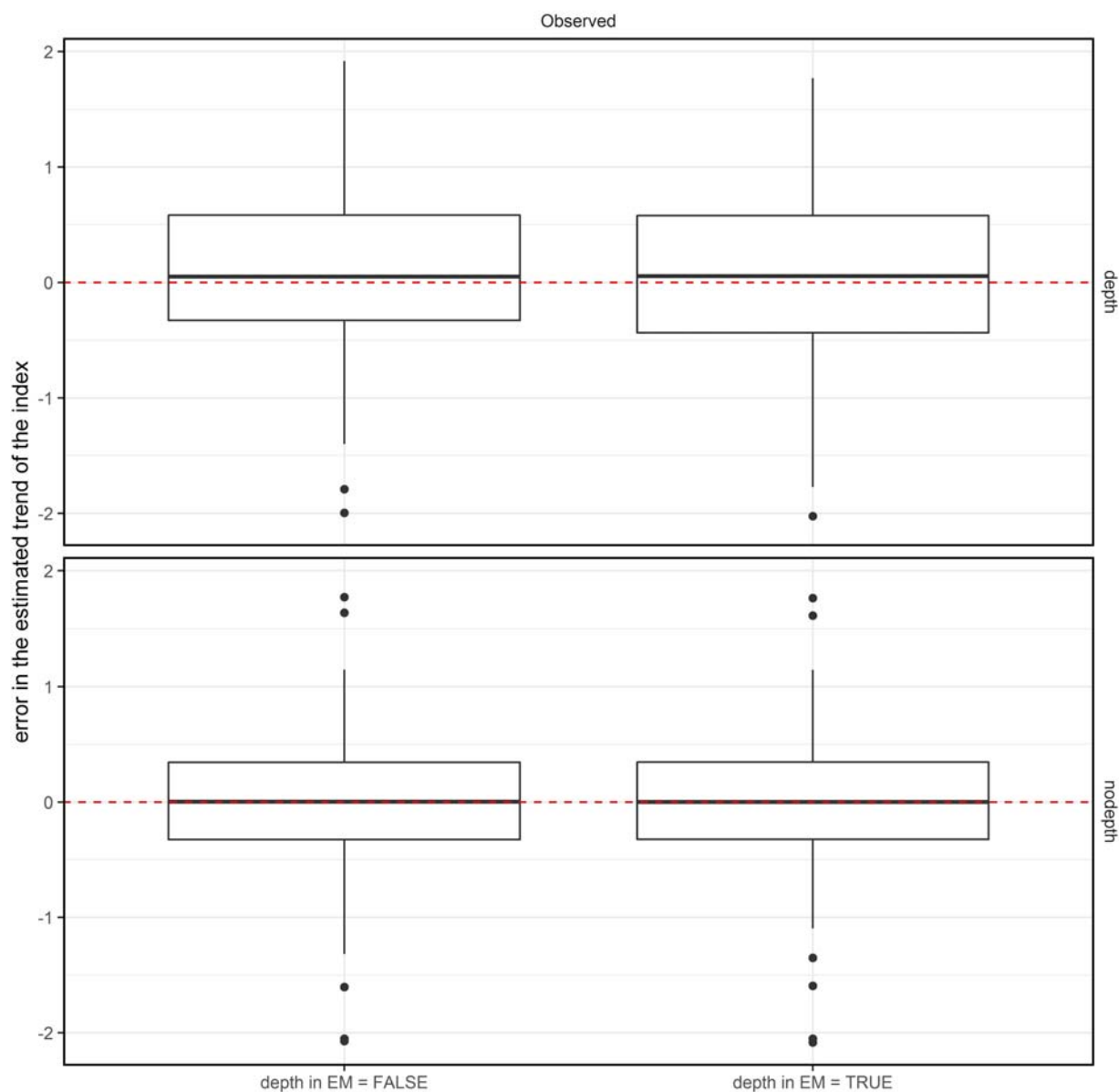
Figure 3.5. Boxplots of the error in the log ratio of the first and last years of the index of abundance. Panels define the presence of quadratic effect of depth in the operating model (rows) and if depth was included in the estimation method (EM; column). The red line indicates the location of unbiased estimates. Whiskers depict 1.5 times the first and third quartiles and outliers are represented using points.
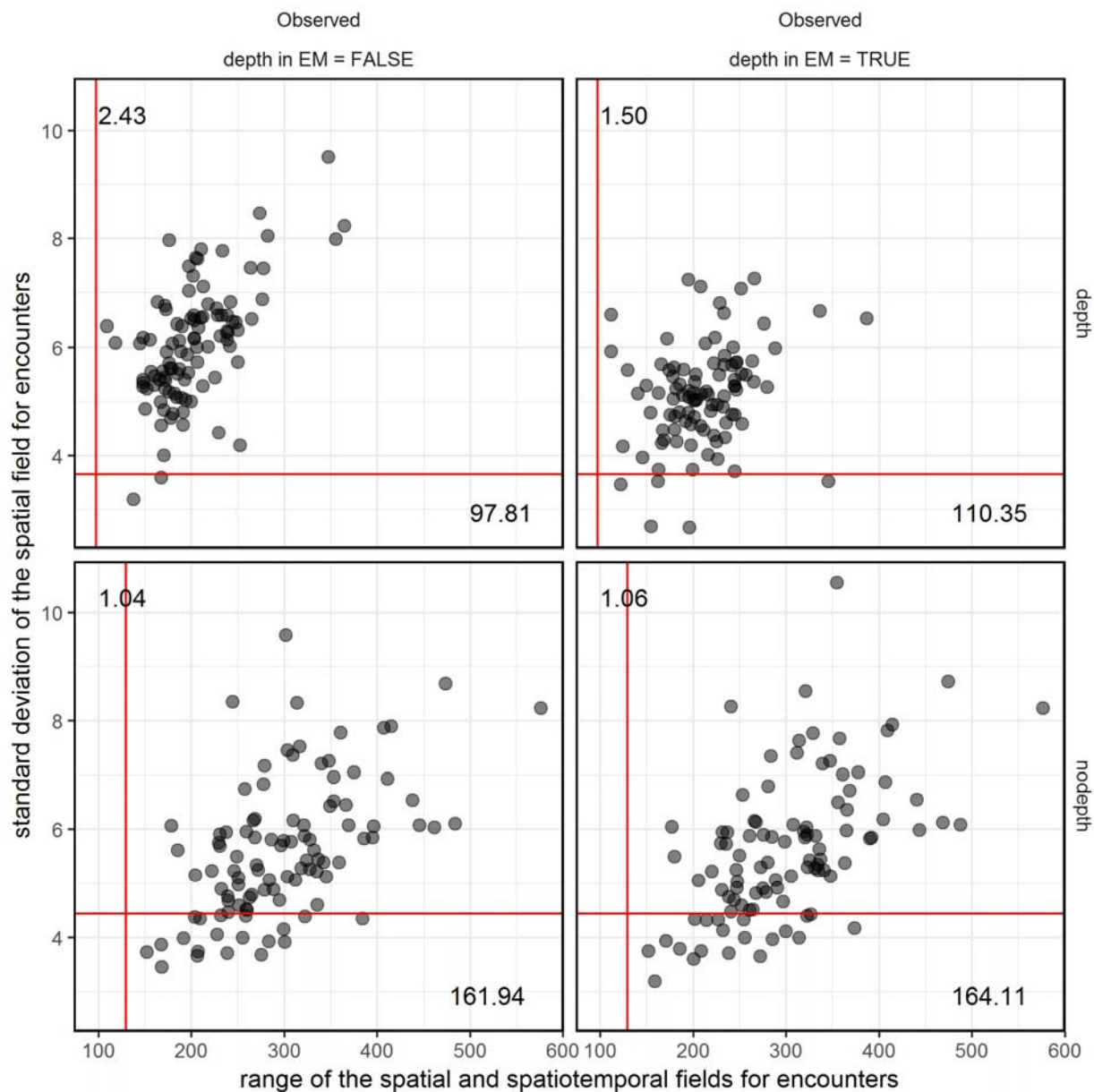
Figure 3.6. Estimates of the variance of the spatial field for encounters versus the range of the spatial and spatiotemporal fields for the encounter model. Linear and quadratic depth terms were included in the operating model used to simulate the data in the bottom row. Linear and quadratic depth terms were included in the estimation method in the right column. Thus, the estimation methods in the bottom-left and the top-right panels are misspecified. The median absolute error (MAE; median|estimated – true|) for each parameter is printed near its respective axis. Unbiased results would be located on the red cross-hairs.

Chapter 4. FACTORS DRIVING CHANGES IN FISHING GEAR IN THE US WEST COAST SABLEFISH FISHERY

Abstract

Reducing catch of non-target individuals and species (i.e., bycatch) remains a global priority of fisheries managers and an impediment to achieving sustainable fisheries. Switching to fishing gear with a lower rate of bycatch is often identified as a feasible method to reduce bycatch. However, fishers face many decisions, and it is often difficult to predict under what circumstances they will switch gear types when the management framework allows it. Allowances for changes from trawl to fixed gear in the US West Coast groundfish fishery were implemented in 2011 for fishers holding a Limited Entry trawl permit. Thus, it was expected that the proportion of sablefish (*Anoplopoma fimbria*) landings caught using fixed gear versus trawl gear would increase starting in 2011. Cluster analysis identified two years (1991 and 2005) since 1982 with major changes in gear types used to land sablefish. Changes in the proportion of landings caught using fixed gear versus trawl gear within the limited entry trawl fishery since 2011 were best explained by port of landing and not vessel characteristics or the relative abundance of bycatch species. Future work could explore the importance of local governance structures, such as membership in and/or rules of risk pools, to understand shifts in fishing gear pending the collection of such information. The results highlight difficulties of integrating data across multiple disciplines and the importance of doing so when analysing human choices.

4.1 INTRODUCTION

The management of natural resources occurs in complex socio-ecological systems (SESs) that are characterized by multiple connections and feedbacks (Ostrom, 2009). Large scale controlled, replicated experiments are nearly impossible within these complex systems (Jensen *et al.*, 2012), as is the complete elimination of uncertainty. Therefore, understanding uncertainty and

incorporating it into management frameworks is necessary for the development of a comprehensive and sustainable approach to the management of natural resources.

There are three broad categories of uncertainty that are relevant to manager in charge of natural resources, environmental, scientific, and implementation uncertainty (Mehta *et al.*, 1999). Environmental uncertainty refers to the variability of ecosystem processes. Scientific uncertainty arises from imperfect sampling and the need to simplify ecosystems into tractable models that will almost always lead to model misspecification. Lastly, implementation uncertainty refers to the inability to predict how users will respond to policies or regulations. Each type of uncertainty can act to undermine the effectiveness of a management framework. Implementation uncertainty has received far less attention than environmental and scientific uncertainty (Fulton *et al.*, 2011). Adequately characterizing implementation uncertainty in the management of marine fishes relies on honing our ability to predict fishers' responses to management frameworks. Informing predictions of fisher response can be accomplished by reviewing past fisher behavior under similar management frameworks or using logic to predict what fishers might do under a new management framework. Here, we review the behavior of fishers within a complex SES to inform future incentive-based management strategies.

Fisher behavior is the result of both long- and short-term decisions, which are highly contextual (Hart and Pitcher, 1998; Hunt, 2005). For example, long-term decisions can relate to capital investments such as purchasing a new vessel, whereas short-term decisions can pertain to a fishing trip such as whether or not to go fishing. Decisions are often shaped by what is allowed within the management framework and knowledge gained through experience or from others' experiences. Experiences can include factors such as environmental conditions, risk tolerance, costs of fishing, and past catch rates (Steelman and Wallace, 2001; van Putten *et al.*, 2012). Furthermore, decisions involve assessing trade-offs among multiple personal objectives, uncertainty in the information used to guide choices, and the actions of others (Allen and McGlade, 1987). Therefore, fisher behavior will be driven by much more than just rational economic objectives (e.g., profit maximization). Consequently, management frameworks that fail to account for the complex socioeconomic and cultural contexts of fisher decision-making processes may fail to achieve their objectives (Hilborn, 1985; Mahon *et al.*, 2008).

Insights on fisher behavior are often deduced from information meant to inform fisheries scientists about landings and effort. These data are typically available for commercial fisheries but

subject to varying levels of uncertainty. This indirect inference is needed because direct methods, such as interviewing fishers, often require an unreasonable amount of time and costs. Furthermore, the theory of reasoned actions proposes that an individual's attitudes cannot alone be used as a direct predictor of their behavior because intention may or may not result in a specified behavior (Ajzen, 1991). Few studies, at least with respect to large-scale commercial fisheries, rely on interviews to inform research on fisher behavior, but see Holland and Sutinen (1999), St. Martin and Hall-Arber (2008), and Andersen *et al.* (2012) for exceptions.

Identifying factors that affect fisher behavior, from any type of data, is challenging. Relationships are often non-linear and constantly evolving across space and time (Holling *et al.*, 1998). For example, purse-seine fishers in the Peruvian anchovy (*Engraulis ringens*) fishery use information from acoustic systems to adapt their travel and search durations based on the perceived spatial extent of the stock. Acoustic devices have improved over time leading to better predictions, but predictions also depend on environmental conditions which can vary on short time scales (Bertrand *et al.*, 2004). Formal and informal rules can also affect behavior, and the removal of any particular rule may affect the interactions of those that remain (Cox, 2011).

Finding a combination of rules that leads to the achievement of the desired objectives is challenging. Traditionally, commercial fisheries have been regulated through the combined use of input and output control rules. Input measures, such as limiting the number or size of fishing vessels, can increase the cost of fishing and frequently lead to unintended consequences (Holland *et al.*, 2010). Output control measures, such as total allowable catches (TACs), incentivize competition among fishers as they "race-to-fish." Conversely, incentive-based measures have been found to increase the likelihood of reaching multiple objectives desired by management and fishers (Hilborn *et al.*, 2005; Grafton *et al.*, 2006). For example, providing extra days at sea to fishers who switch to a gear with lower catch rates of non-target species (i.e., bycatch) can lead to decreased discards and longer seasons. Additionally, individual transferable quotas (ITQs) can reduce the need for input control measures while still achieving management goals (Branch, 2009). Low TACs for overfished species allocated to individual fishers may limit the overall catch without directly specifying how fishers stay within their limits (Casey, 1995; Hall *et al.*, 2000; Holland and Jannot, 2012). Unfortunately, it is difficult to predict how fishers will change their behavior once they have the enhanced security of owning their own share of the TAC. Furthermore, ITQs

may need to be modified once in place to ensure that management goals are being met and that the system is fair and socially acceptable for everyone involved (Branch *et al.*, 2006).

Much research exists on the effects of ITQ systems in general and for several case studies. The combined effects of ITQ systems and the relaxation of input control measures are less studied. Here, we use the US West Coast groundfish fishery as a case study to investigate drivers of fisher behavior regarding gear choice in an ITQ fishery that allows fishers to choose among gear types. Fishers target several species within this fishery, but we focus on those targeting sablefish (*Anoplopoma fimbria*) because it is the species within the fishery most frequently caught by non-trawl (i.e., fixed) gear and the most valuable species, by far. The objectives of this study, therefore, are to (1) identify and contextualize major changes in gear types used to land sablefish since the start of the fishery and (2) quantify drivers of changes in behavior for fishers engaged in gear switching since the implementation of the ITQ system.

4.2 METHODS

4.2.1 *Overview*

Two methods were used to understand the behavior of fishers targeting sablefish in the US West Coast groundfish fishery. First, cluster analyses were used to identify temporal shifts in gear choices. Second, generalized linear models (GLMs) were used to identify drivers of gear choice. Data used to fit the models included information on ecological, economic, and social aspects of the fishery, collected from both fishery-dependent and -independent sources. Analyses and steps required to merge the data that were provided by several groups to a useable format are detailed below, or in associated appendices.

4.2.2 *US West Coast groundfish fishery*

The US West Coast groundfish fishery is a federally managed fishery that operates between the US-Canada and the US-Mexico borders (Figure 4.1). Commercial harvests occur within four main sectors. The Limited Entry (LE) sector is composed of two sub-sectors, the LE trawl and LE fixed-gear sectors. The open-access sector consists of fishers who have a history of fishing but do not hold a LE permit. Tribal fishers off the coast of Washington and recreational fishers fishing along the entire coast comprise the last two sectors. All sectors are subject to annual catch limits

and area- and gear-specific restrictions. Major changes occurred in the fishery in 2011 when an ITQ system for the shore-based Pacific whiting (Pacific hake, *Merluccius productus*) and non-Pacific whiting sectors was implemented, a co-op provision for the at-sea Pacific whiting sector was established, discarding was no longer allowed, observer coverage increased from 20% to 100%, and an allowance for the harvesting of trawl ITQ using any gear that is legal within the fishery was implemented (PFMC, 2010). Prior to 2011, the fishery was managed using trip limits that did not include discards.

The fishery includes 90+ species and primarily targets demersal species such as sablefish, Dover sole (*Microstomus pacificus*), shortspine thornyhead (*Sebastolobus alascanus*), Petrale sole (*Eopsetta jordani*), and Pacific whiting (PFMC, 2016). Targeting occurs partially through spatial choices in fishing location to capitalize on species-specific habitat preferences (Branch *et al.*, 2005; Jannot and Holland, 2013). Fishers encounter several bycatch species including Pacific halibut (*Hippoglossus stenolepis*), an internationally-managed species, bocaccio rockfish (*S. paucispinis*), cowcod rockfish (*S. levis*), darkblotched rockfish (*S. crameri*), Pacific ocean perch (*S. alutus*), and yelloweye rockfish (*S. ruberrimus*). These species other than Pacific halibut include the majority of those declared overfished and under rebuilding plans since 2004 (PFMC, 2003). As of December 2017, only cowcod rockfish and yelloweye rockfish remain under rebuilding plans.

Sablefish is one of the most valuable species within the fishery with 2015 gross revenue of $28.7 million (NMFS, 2017). Ex-vessel prices peaked in 2011 at $6.99 / kg and have since declined to the mid-$5 / kg range. Catches from trawl gear generally fetch a lower ex-vessel price than those caught using fixed gear, and 27% of the 1,010 mt landed in 2011 was landed using fixed gear. Trawlers landing sablefish are typically targeting dover sole, longspine thornyhead (*S. altivelis*) and shortspine thornyhead along with sablefish. Conversely, almost all vessels using fixed gear are primarily targeting sablefish.

4.2.3 *Transition years among dominant gear types*

Historical landings were used to highlight temporal changes in the dominant gear type used to land sablefish. Long time series of sablefish landings were available because sablefish have consistently been identified to the species level since 1908 and gear-specific reconstructions extend the time series to 1900 (Johnson *et al.*, 2015). Landings were defined as fish brought to

shore, and landings plus discards equal catches. Landings by gear (line, pot/trap, and trawl gear) and sector were available since 2002 (PacFIN, 2016).

Transitions were investigated since 1982, the first year of federal management, using hierarchical clustering, Nonmetric Multidimensional Scaling, and Simpson Indexes of Diversity. Hierarchical clustering was performed on Bray-Curtis measures of distance between proportions of landings caught using each gear type out of the total yearly landings (Oksanen *et al.*, 2015). Year-specific proportions were used to emphasize differences in gear rather than absolute yearly differences that could be affected by annual catch limits. Differences of one indicate complete dissimilarity and differences of zero occur when proportions are equal between years. Clusters were defined using nearest neighbour-chain algorithms, which minimize variance using a "complete bottom up", or agglomerative, criterion. Average silhouette width provided a measure of cluster validity by measuring the degree of membership of an object to its cluster (Rousseeuw, 1987; Maechlet *et al.*, 2014). Degree of membership was calculated as the difference between the average distance between the given object and all objects of the cluster to which it belongs and the same measure computed for the next closest cluster. Two-dimensional Nonmetric Multidimensional Scaling (Oksanen *et al.*, 2015) provided a way to visualize dominate gear types across years. The Simpson Index of Diversity measurements highlighted years with relatively equal proportions by gear type (Oksanen *et al.*, 2015). A literature review of PFMC documents, non-governmental organization reports, and peer-reviewed literature was performed to provide context to identified transitions.

4.2.4 *Drivers of fisher behavior*

Investigated social and ecological drivers of fisher behavior within the LE trawl sector were chosen using Ostrom's (2009) framework for SESs. The framework facilitates the integration of data from multiple disciplines such that variables are sought from all influential aspects of the SES. Information on the resource unit, the resource system, and the governance system was sought to explain the behavior of resource users (McGinnis and Ostrom, 2014).

The dependent variable ("fixed landings") was the year- and port-group-specific proportion of sablefish landings within the LE trawl sector landed using fixed gear (PacFIN, 2016; www.psmfc.org/pacfin). Fixed-gear landings are assumed to represent the degree to which fishers within the sector chose to use fixed gear rather than trawl gear to land sablefish. Proportions were

calculated for each port group from 2011-2015 (Appendix C). Port-specific landings were assigned to port groups (Washington, Astoria and Tillamook, Newport, Coos Bay, Brookings and Crescent City, Eureka, Fort Bragg, San Francisco and Bodega Bay, and Monterey and Morro Bay) based on spatial proximity and civil boundaries (PFMC, 2004).

Generalized linear models were used to quantify the power of each independent variable to explain variability in fixed landings (Table 4.1). Multiple variables related to fishers (i.e., resources users) were available from NOAA's Economic Data Collection Program. Information on sablefish (i.e., resource unit) was available from the Northwest Fisheries Science Center (NWFSC) Shelf-Slope survey. The survey also provided information about other species inhabiting the California Current Large Marine Ecosystem (i.e., resource system). Information regarding the governance system was unavailable (see below).

The Economic Data Collection Program has been a mandatory component of the US West Coast groundfish fishery since 2009. The program collects social-ecological data, such as annual expenses and vessel characteristics, on fishers within the LE trawl sector. Information is collected directly from owners and operators who must certify, under penalty of perjury, that all information is true and complete to the best of their knowledge (USOFR 2014c). Year-specific information from each participating vessel was assigned to the port group for which that vessel had the highest ex-vessel revenue in that year. Yearly, port group-specific sums or means were provided for average daily fuel use (gal·day^{-1}); average fixed costs (USD), which included expenses for processing equipment, vessel equipment, insurance premiums, and moorage; average variable costs (USD), which included expenses for bait, captain salary, communications, crew salary, association dues, food, freight, fuel and lubrication, ice, licenses, observers, offloading, supplies, travel, and trucking; number of fish buyers; average crew size, excluding the captain; total number of vessels delivering to the port group; average horsepower of main engine (hp); average fuel capacity (gal); average towing speed while trawling (kn); and average vessel length (m).

The availability of sablefish to fishers delivering landings to each port group was characterized using year-specific estimates of local abundance. Estimates were derived using spatiotemporal delta-models (Thorson and Barnett, 2017) that separately estimate the probability that a geo-referenced sample will encounter the species and the expected catch rate given that the species is encountered (Maunder and Punt, 2004). Models were used rather than mean or stratified-mean abundance to account for sampling design and fish behavior (Thorson and Ward, 2013).

Spatiotemporal models were fit to relative biomass data on sablefish collected annually from 2003 to 2015 during the scientific survey for groundfish and invertebrates off the US West Coast (Bradburn *et al.*, 2011). Models provided a local index of abundance for each port group assuming that fishers on average travel no more than 200 nm north or south of the main port (see 3.2.1).

We assumed that fishers targeting sablefish attempt to avoid catching species managed under rebuilding plans because of the limited quota available and rules prohibiting discarding. Year-specific estimates of relative local abundance were derived for canary rockfish, darkblotched rockfish, and Pacific ocean perch in the same manner as was done for sablefish. Each of these species were managed under a rebuilding plan in 2011.

Information on local abundance was also sought for bocaccio rockfish, cowcod, widow rockfish, and yelloweye rockfish, additional species managed under rebuilding plans in 2011. Unfortunately, the survey had an insufficient number of positive tows per year to be informative on a coast-wide level for estimating local indices for these species (Taylor and Wetzel, 2011). Future work could investigate ways of estimating the local abundance using additional surveys such as the hook-and-line survey also operated by the Northwest Fisheries Science Center (Harms *et al.*, 2010). Fortunately, bycatch of cowcod has been low since 2002 after the implementation of Rockfish Conservation Areas (NMFS, 2004) and bycatch of bocaccio rockfish is small relative to some of the other species managed under a rebuilding plan. Thus, not including them as potential drivers of behavior in the model was not seen as an issue. However, bycatch of yelloweye and canary rockfish is possible when trawling or fishing with longlines (Jenkins and Garisson, 2013).

The dependent variable was regressed against potential independent variables using GLMs under the assumption that the errors were beta distributed. The beta distribution (see Appendix C for more details) was used because measures of the proportion of the landings of sablefish by fixed gear were non-integer values bound between zero and one (Smithson and Verkuilen, 2006; Schmid *et al.*, 2013). The model, $fixed\ landings = port\ group + SES\ variables$, was initialized to only depend on port group. Additional variables were added using forward step-wise model selection. Continuous fixed effects were transformed, $(x_i - mean(x))/sd(x)$, to have a mean of zero and a standard deviation of one. Standardization changed the range of all variables to approximately [-3,3] and facilitated the inclusion of variables measured in different units. Standardization was performed across port groups and years (i.e., transformations occurred after data were summarized to port groups).

Models were assessed for their convergence using standard residual plots. The best model was chosen using the Akaike information criterion (AIC; Burnham and Anderson, 2002).

4.3   RESULTS

4.3.1   *Dominant transitions*

Documented landings of sablefish caught along the US West Coast date back to 1908 (Figure 4.2 upper panel; Johnson *et al*., 2015). Historically, sablefish were mainly landed using line gear and the use of pots/traps started increasing in 1960. Pots/traps were mainly fished by foreign fleets. Landings from trawl gear began increasing in the late 1960s, but trawl landings did not proportionally exceed those caught by pots/traps until the late 1980s (lower panel in Figure 4.2).

Clustering algorithms identified three distinct groups of years since 1982 with similar proportions of landings by gear type, 1982-1990, 1991-2004, and 2005-2014 (Figure 4.3). Less dominant (i.e., dissimilarity metric < 0.04) transition points were also identified in 1987, 1997, and 2011 (Figure 4.3). Changing the start year from 1982 to any year between 1900 and 2002 did not result in 2011 being identified as a dominant transition year (results not shown). The first assemblage (1982-1990) was dominated by landings from pot/trap gear, whereas the third assemblage (2005-2014) was dominated by landings from line gear (Figure 4.4). Landings from trawl and line gear dominated the transitionary assemblage from 1991 to 2004. The mean proportion of landings from line gear calculated for each assemblage of years increased over time, whereas the opposite was true for landings from trawl gear. In 2005, the proportion of landings caught using pot/trap gear was higher than all previous years since 1994 (Figure 4.4).

In general, evenness among gear types increased over time (Figure 4.5). The Simpson Diversity Index identified four periods, since 1982, where a single gear type dominated the landings (Figure 4.5): 1982-1984 (pot/trap gear), 1989-1993 (trawl gear), 1997-1999 (line gear), and 2008 (line gear).

4.3.2   *Drivers of fisher behavior*

Data from the NWFSC Shelf-Slope trawl survey collected from 2003 to 2015 indicate sablefish are ubiquitous up and down the US West Coast. Larger than average tows of sablefish were observed north of central California but never south of San Francisco (Figure 4.6). The spatial

extents of darkblotched and canary rockfish were more similar to sablefish than that of Pacific ocean perch (Figure 4.6).

Port group and the number of vessels (Table 4.2) were the most influential variables, of those investigated (Table 4.1), in explaining fisher behavior. San Francisco and Bodega Bay had the highest and Coos Bay had the lowest proportion of landings from fixed gear. Three of the nine included port groups (Brookings and Crescent City, Coos Bay, and Eureka) had zero landings from fixed gear in at least some years. Proportions of zero were excluded from the analysis because of the use of the beta distribution, which led to 36 observations. The small sample size decreased the ability to estimate effect sizes and could have led to larger standard errors than what would have been estimated had the model been fit using more data. For example, the positive effect of the number of vessels was only weakly supported (difference in AIC of 1.78) and the lower bound of the confidence interval for the effect was close to zero.

Although the relative abundances of sablefish, canary rockfish, darkblotched rockfish and Pacific ocean perch varied spatially over time (Figure 4.7), this variation did not appear to drive fisher behavior. Additional species not included in the GLM could be more constraining.

4.4 DISCUSSION

Fishers are subject to a number of drivers that can and do affect their behavior. Understanding these drivers can help managers devise and improve management plans; particularly, plans that include incentives to change fisher behavior. This case study of the US West Coast groundfish fishery serves as an example of how retrospective analyses of fisher behavior can be used to inform future fisheries management. First, cluster analyses revealed that fishers may not be as apt to switch gears as hypothesized. Changes in gear types used to land sablefish between 2011 and 2015 were seen as small changes when measured against all changes since 1982. Second, results from GLMs suggest that future incentive-based management aimed at increasing the proportion of sablefish landed using fixed gears could be better informed if they accounted for current incentives that are more than likely to vary across the major port groups. We did not model incentives explicitly and thus cannot predict incentives that would lead to future change, only that behavior differs among the major port groups. These issues are explored in more detail below.

The LE sector of the US West Coast groundfish fishery experienced several management changes in 2011. The allowance for gear switching was implemented concurrently with the

implementation of the ITQ program for the LE trawl sector. This allowance enabled LE trawl permit holders to land quota using fixed gear. Additionally, fishers within the LE fixed-gear sector could acquire a trawl permit and fish for its associated quota using fixed gear. We thus hypothesized that a dominant transition in gear type would be measured for 2011. However, the transition identified for 2011 was less defined than the transitions identified for 1991 and 2005. The transition identified for 1991 was largely the result of increased output control measures (i.e., shortened fixed-gear seasons; Methot *et al.*, 1994). Conversely, the 2011 allowance for gear switching represents a relaxation of an input control measure. It is possible that the changes in gear type will be slow because fishers typically exhibit risk-averse behavior (e.g., Eggert and Martinsson, 2004). The definition of slow in this context is unclear. Rates of change may increase with time as fishers become more informed about the costs of the transition. For instance, the cost of new gear might not be outweighed by the higher price per pound when decreased landings of other species targeted with trawl gear are taken into account (Cinner, 2007). Decreased differences in prices between sablefish landed with fixed gear versus trawl starting in 2012 may have also discouraged fishers from using fixed gear.

Records of sablefish landings confirm that fixed-gear landings did increase. Records also highlighted a less publicized shift that occurred in 2008 when fixed gear was allowed within the LE trawl sector under an Exempted Fishing Permit. The Exempted Fishing Permit allows vessels fishing under six LE trawl permits owned by The Nature Conservancy to fish in waters outside of the seaward boundary of the Rockfish Conservation Area between 36ºN (Point Lopez) to 34º27'N (Point Conception) using fixed gears. Quotas for limiting species are aggregated across all permits; thereby, acting as a community-based management system (TNC, 2014). Consequently, the allowance for gear switching in 2011 had less of an abrupt impact on landings originating in central California (i.e., "Monterey and Morro Bay") than it did for the other port groups. Unfortunately, vessel data were aggregated to the port-group level without consideration of their participation in community-based programs. We were thus unable to tease apart how much of the quota starting in 2011 data from "Monterey and Morro Bay" included that from fishers operating under the Exempted Fishing Permit from those not under the special permit. Furthermore, some of the fishers that participate in the community-based program fish using vessels owned by The Nature Conservancy are not using vessels chosen by fishers as optimal for converting to fixed gear. Future

work could investigate fisher behavior using vessel-specific data. However, this vessel specific information is only available for some of the databases used in this analysis.

Vessel-specific data would address the issue of assigning vessels to a single port group. For this analysis, vessels were assigned to the port for which they had the highest ex-vessel revenue, even though vessels may deliver to multiple ports. In this way, port group acted as a proxy for spatial location and other variables not included in the model. For instance, we suspect that vessels assigned to Oregon ports may have experienced increased difficulty in switching gears, both socially and logistically, compared to other port groups because Astoria and Warrenton generate the highest trawl revenue of all of US West Coast port communities (Sepez *et al.*, 2006). These ports and others in the general vicinity may not be equipped to support vessels fishing with fixed-gear. Interviewing fishers could provide more insight on actual constraints. Furthermore, using vessel-specific data would eliminate the need to include the number of vessels per port group in the model. Assigning relative abundances of the resource and the environment to the user is difficult given the available data. Here, we determined the average distance fishers travel across years to assign relative abundance along the coast to each port group. Vessel monitoring systems or automatic identification systems could be an additional source of information on where fishers travel. Using data from monitoring systems would limit retrospective analyses because the requirement for this technology aboard fishing vessels is relatively new and not universal. In the future more precise and accurate estimates could be obtained if information from several surveys could be combined in a single analysis to increase spatial and temporal coverage.

Results demonstrate that fishers can adapt their behavior, within certain constraints, but it remains unclear if the increased use of fixed gear should be seen as successful because no target goals were set upon implementation. Jenkins and Garrison (2013) hypothesized that most fishers using trawl gear would likely not permanently convert to fixed gears and reductions in bycatch from using gear with a lower catch rates of non-target species compared to trawl gear would not be fully realized without appropriate incentives. Results presented here suggest that if managers wish to further increase the proportion of the sablefish TAC landed using fixed gear that future incentives could be more effective if they consider how the proposed incentives will lead to changes in revenue (Eliasen *et al.*, 2013). For example, allowing fishers to use fixed gear will only be fruitful if ports are equipped to buy catches from fixed gear or if other regulations allow them to fish in familiar areas with the new gear type (Manson *et al.*, 2012).

4.5 TABLES

Table 4.1. Hypothesized drivers of fisher behavior within the US West Coast LE trawl fishery. Variables are linked to one of Ostrom's (2009) four subsystems that structure social-ecological systems, resource system (RS), resource unit (RU), governance system (GS), and resource user (U). Data sources include the Northwest Fisheries Science Center (NWFSC) Shelf-Slope Trawl Survey conducted from 2003 to 2015, Economic Data Collection (EDC) data provided by the West Coast Fisheries Economic Program from 2009 to 2015, and Pacific Fisheries Information Network (PacFIN) archives provided by Pacific States Marine Fisheries Commission. Some variables were hypothesized to be drivers but spatially- or temporally-explicit information was lacking ("avail").

variable (unit)	system	source	interpretation	avail.
sablefish (mt)	RU	NWFSC	RU abundance can influence catch rates	yes
bocaccio rockfish (mt)	RS	NWFSC	potential for bycatch	no
canary rockfish (mt)	RS	NWFSC	potential for bycatch	yes
cowcod rockfish (mt)	RS	NWFSC	potential for bycatch	no
darkblotched rockfish (mt)	RS	NWFSC	potential for bycatch	yes
Pacific ocean perch (mt)	RS	NWFSC	potential for bycatch	yes
yelloweye rockfish (mt)	RS	NWFSC	potential for bycatch	no
community-based management	GS	lit	rules defined w/o gov. authority	no
vessel length (ft)	GS	EDC	permit transfers are limited by vessel size, larger vessels can fish farther offshore, and smaller vessels fish with fixed gear in near-shore rockfish conservation areas	yes
fuel (gal·day^{-1})	U	EDC	higher fuel prices may cause fishers to stay home or target other species	yes
crew size (individuals)	U	EDC	increased transaction costs with more crew	yes
num. of vessels	U	EDC	account for more permits	yes
horse power (hp)	U	EDC	ability to steam to more distant grounds	yes
fuel capacity (gal)	U	EDC	ability to steam to more distant grounds	yes

Table 4.2. Results from fitting generalized linear models to the proportion of sablefish landings within the limited entry trawl sector caught using fixed-gear out of all gear types for each port group-year combination ("fixed landings"). Parameters are reported for when other variables were held at their means, and port groups are listed geographically from north to south. The precision parameter (ϕ) was modelled without covariates. The column "year" provides the number of years that each port group had landings from fixed gear between 2011 and 2015. Port group x year combinations without fixed-gear landings were excluded from the analysis (e.g., Eureka in all years).

variable	number of years	parameters		
		estimate	se	p value
Washington	5	1.44	0.43	<0.01
Astoria and Tillamook	5	-0.79	0.30	<0.01
Newport	5	0.28	0.42	0.51
Coos Bay	4	-1.10	0.53	0.04
Brookings and Crescent City	2	0.53	0.56	0.34
Eureka	0			
Fort Bragg	5	-0.64	0.48	0.19
San Francisco and Bodega Bay	5	2.23	0.45	<0.01
Monterey and Morro Bay	5	1.46	0.52	<0.01
Number of vessels		0.41	0.19	0.03
ϕ		9.22	2.13	<0.01

4.6 FIGURES
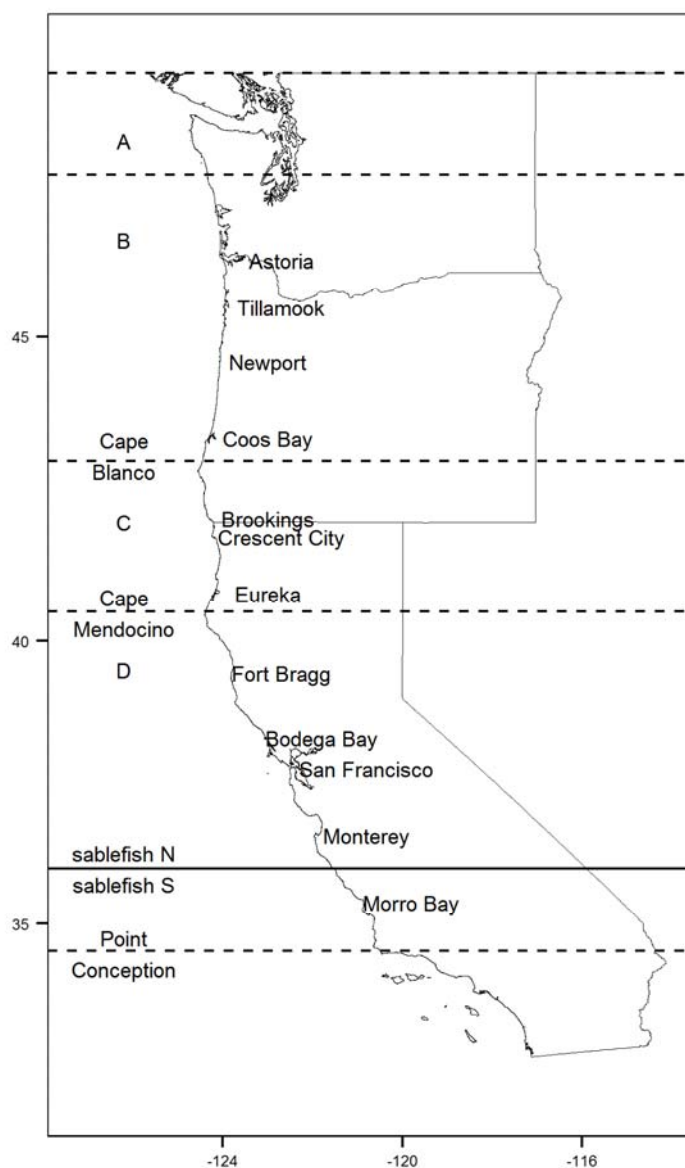
Figure 4.1. The geographical extent of the US West Coast groundfish fishery. Dashed, horizontal lines indicate locations of geophysical points of interest, Cape Blanco (42°50'N), Cape Mendocino (40°26'N), and Point Conception (34°27'N). Nine unique port groups were included in the analysis, Washington, Astoria and Tillamook, Newport, Coos Bay, Brookings and Crescent City, Eureka, Fort Bragg, San Francisco and Bodega Bay, and Monterey and Morro Bay. The solid, horizontal line at 36°N defines a management border that defines north from south quota allocations.
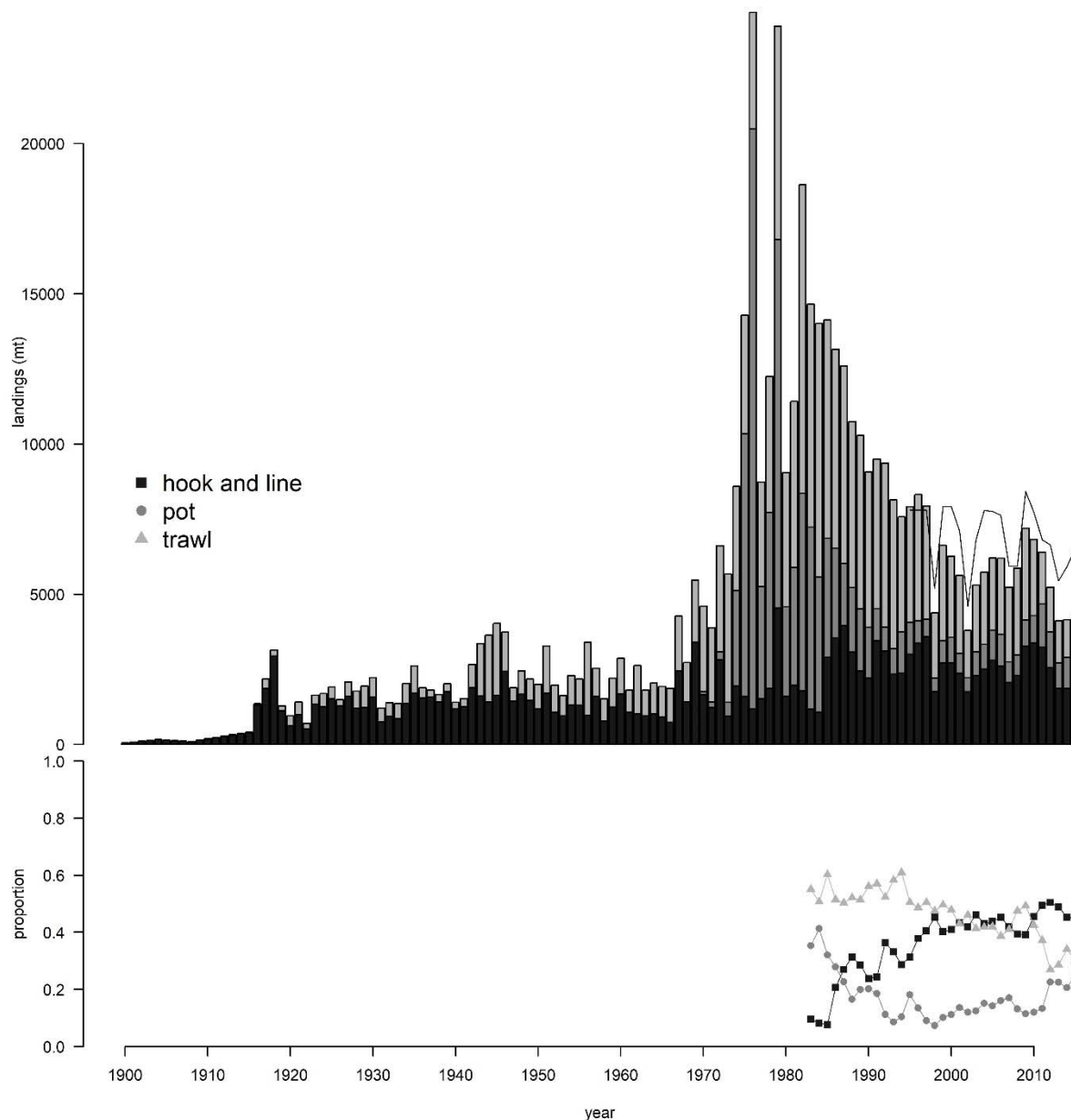
Figure 4.2. Reconstructed and documented US West coast sablefish landings (mt) for all fisheries and sectors from 1900 to 2015 by gear type (top panel). Gear types include line (dark gray), pot/trap (gray), and trawl (light gray) gear. Landings include those from foreign vessels, which are largely responsible for the peak landings in 1976 and 1979. The thin black line indicates the harvest guideline (1995-1997), optimum yield (1998-2010), or annual catch limit (2011-2015) summed across gear type and area from 1995 to 2015. Proportions by gear type are displayed in the bottom panel.
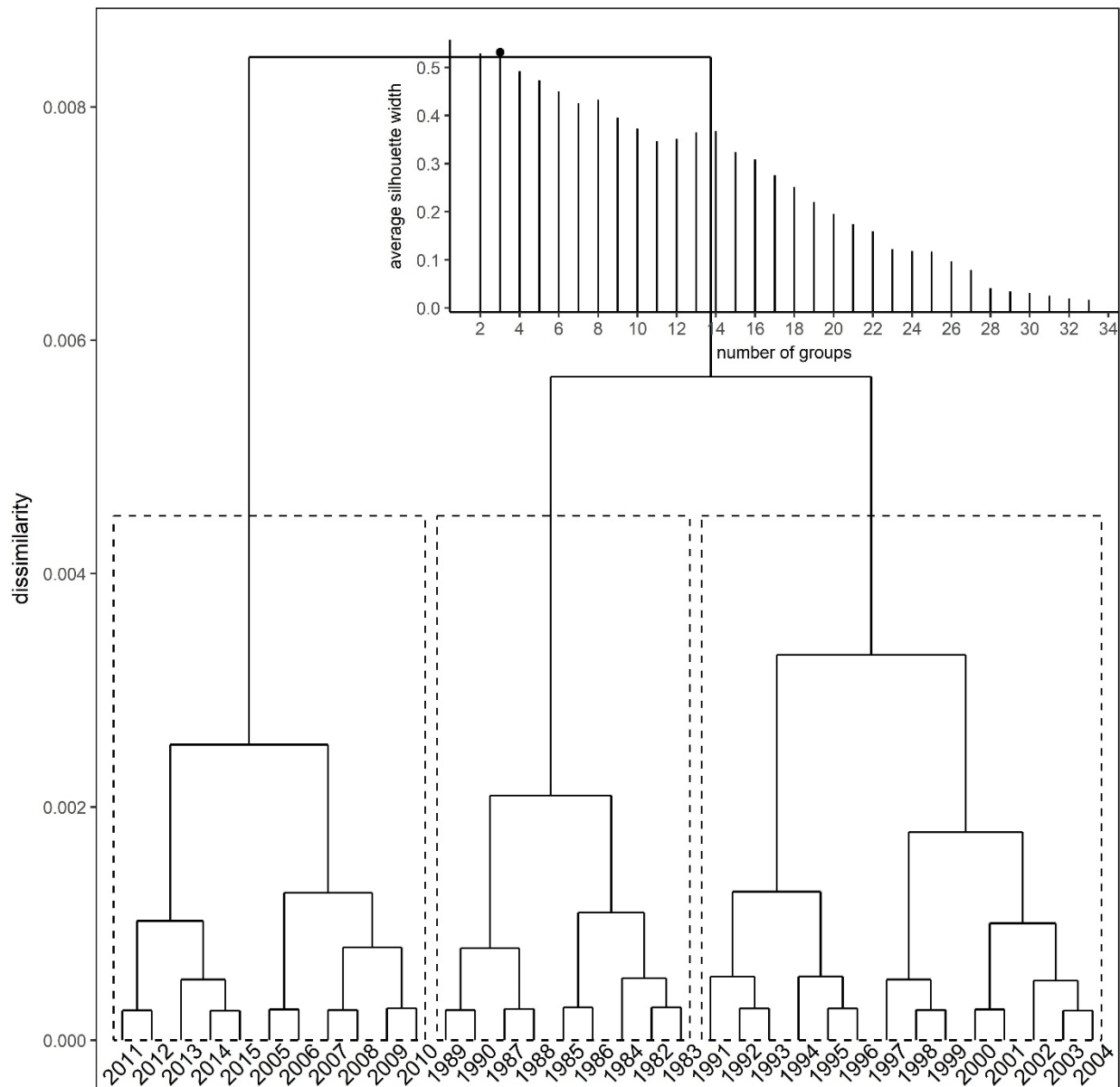
Figure 4.3. Three groups of similar years of gear-specific landings from the US West Coast sablefish fishery. Dashed, black boxes define the groups that were derived from a hierarchical clustering algorithm and average silhouette width (upper inset). Dissimilarities (y-axis) of zero indicates there is no difference between groups.
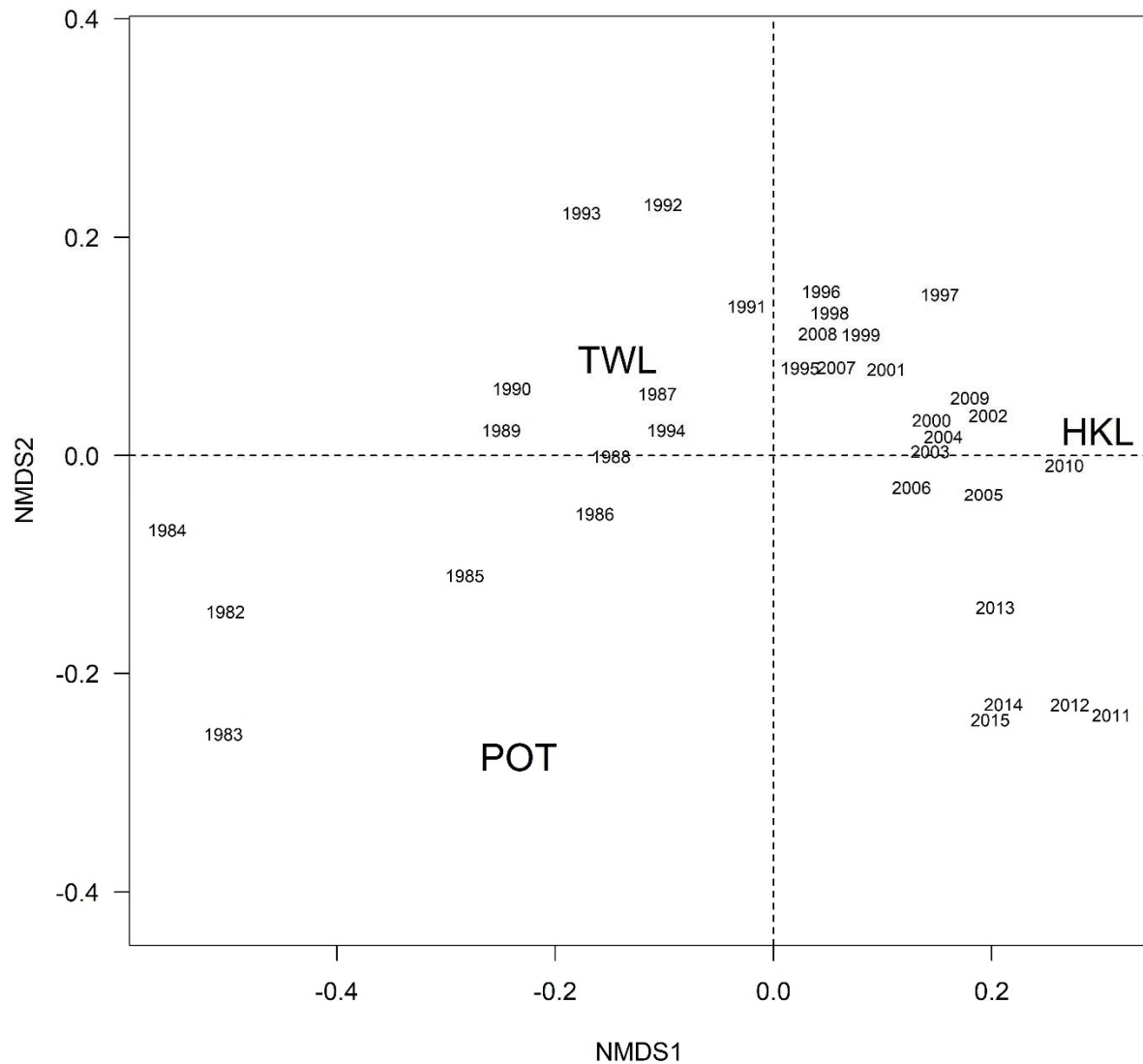
Figure 4.4. Two-dimensional non-metric multidimensional scaling (NMDS2) of gear-specific sablefish landings within the US West Coast groundfish fishery from 1982 to 2015. Gear types include pot/trap (POT), line (LNE), and trawl (TWL) gear.
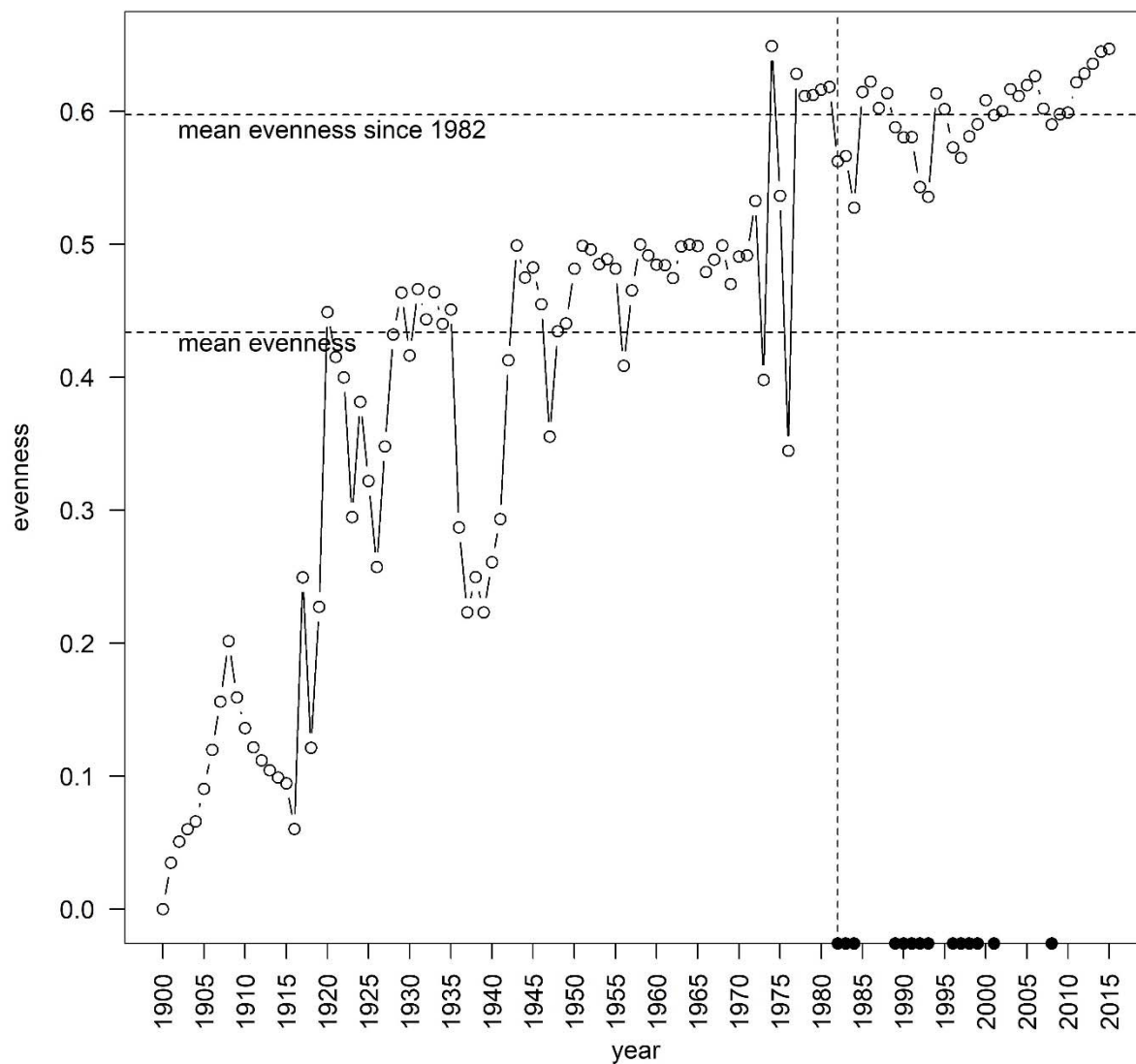
Figure 4.5. Simpson diversity index of landings by gear type for the US West Coast sablefish fishery. Higher values indicate increasing evenness among landings from each gear type. The vertical dashed line at 1982 indicates the first year of federal management regulations for the US West Coast groundfish fishery. Horizontal dashed lines indicate mean evenness for a given range of years: 1900 to present (lower) and 1982 to present (upper). Black marks on the x axis indicate years where the evenness is below the mean evenness since 1982.
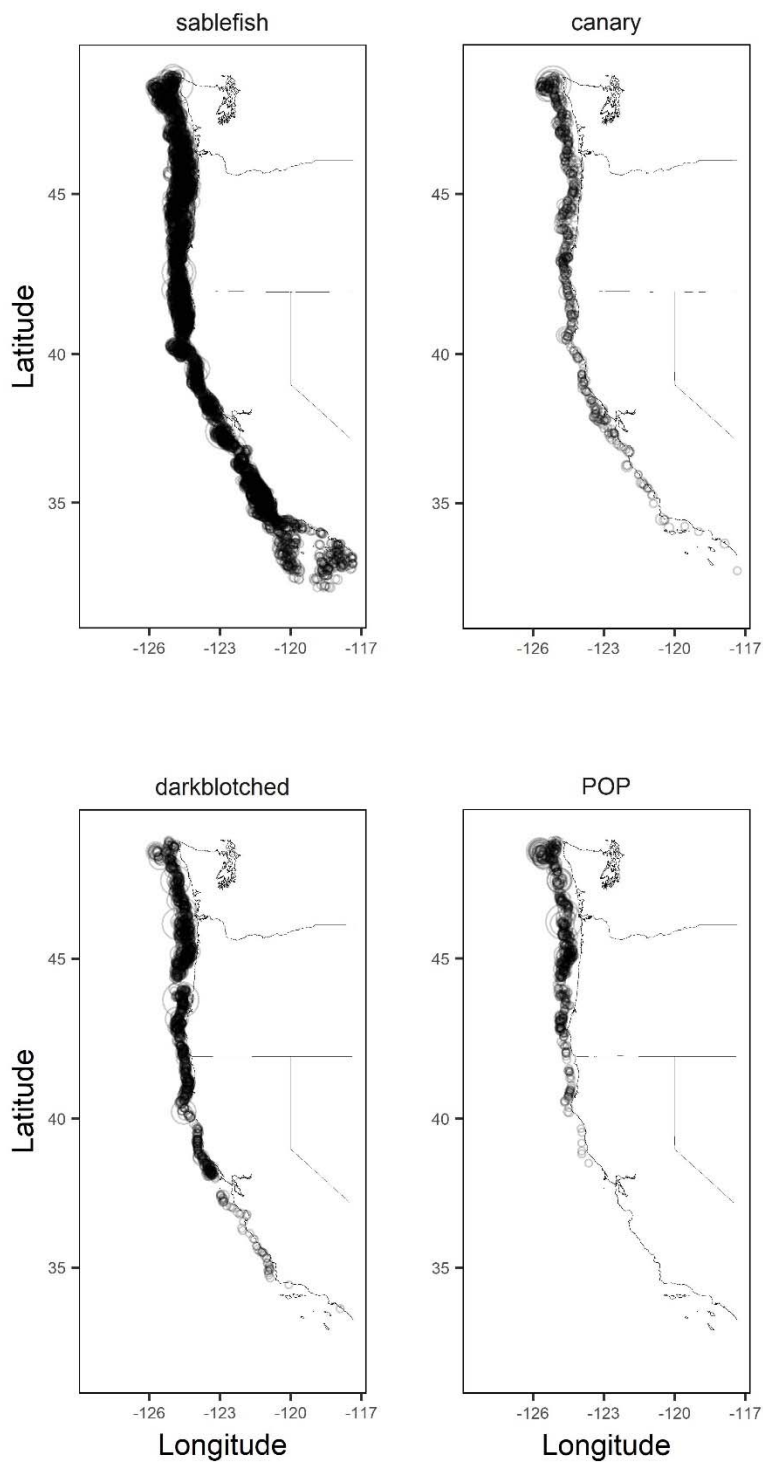
Figure 4.6. Positive tows (kg) of sablefish, canary rockfish, darkblotched rockfish, and Pacific ocean perch (POP) from the Northwest Fisheries Science Center Shelf-Slope survey from 2003 to

2015. Panels display data for a single species over all years, where the size of the transparent circle is relative to the species-specific weight in a given tow.
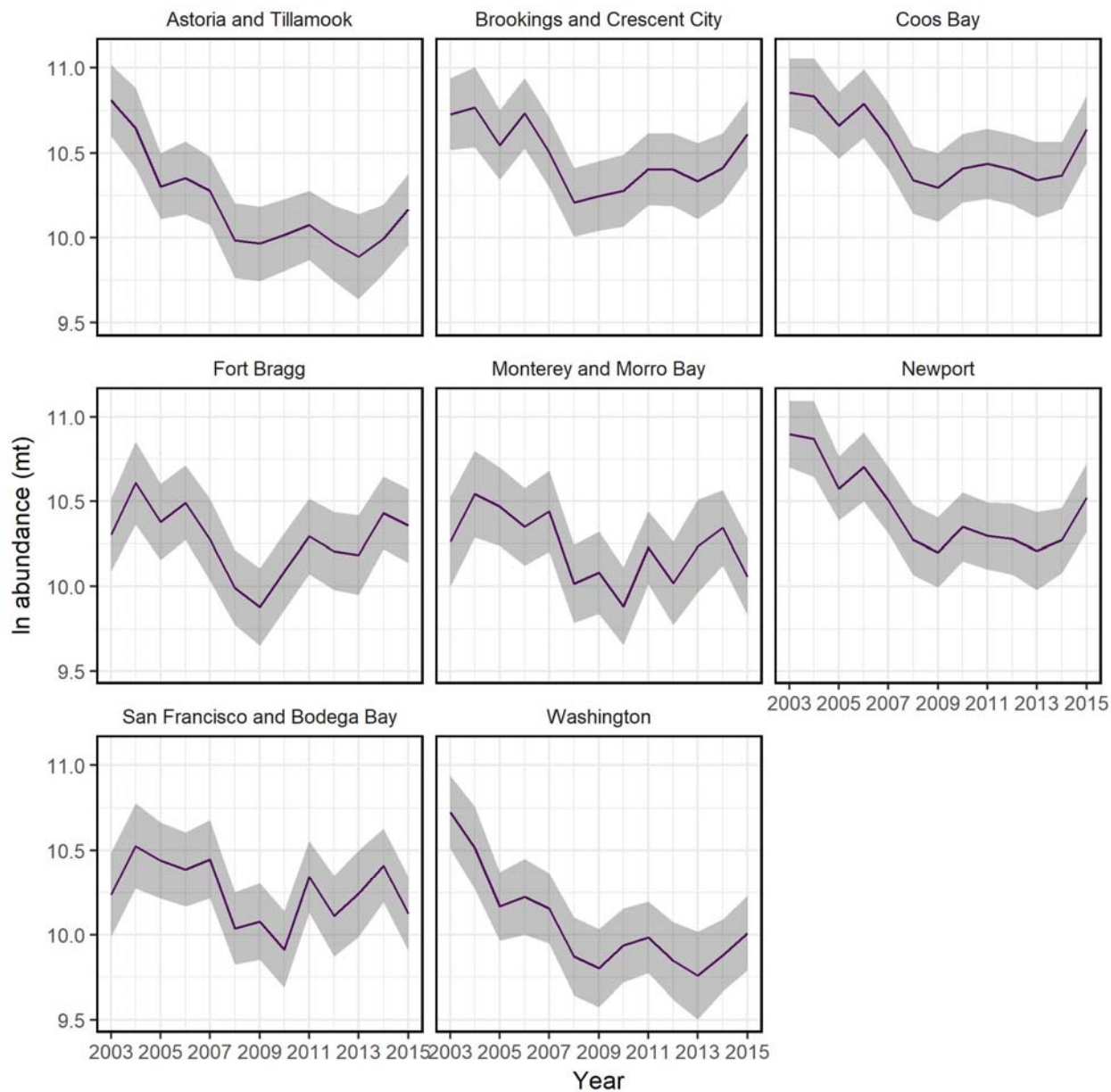
Figure 4.7. Relative indexes of abundances by port group of sablefish. Estimates were generated using spatiotemporal delta-models fit to data from the Northwest Fisheries Science Shelf-Slope trawl survey. Estimates of abundance are for 200 nm north and south of the major port for each port group. Lines are model estimates and dashed lines are 95% credible intervals.

Chapter 5. CONCLUSION

5.1   MAJOR CONCLUSIONS

Since 1985, fisheries management has undergone vast changes. Management complexity increased, additional stakeholders became part of the management process, and statistical methods to evaluate the status of marine species increased in their flexibility (Hilborn, 2012). Additionally, the push for ecosystem models to evaluate impacts beyond individual removals increased. Requests for new policies based on EBFM outpaced scientific tools needed to implement them and tools that were or became available were often implemented without rigorous testing (Smith *et al*., 2007). Here, two multivariate time-series models were assessed for their ability to provide information relevant to EBFM using simulation (Chapters 1 and 3) prior to using them to provide management advice (Chapters 2 and 4). Simulations highlighted the need to account for time-series properties of data and the ability of spatiotemporal models to recover trends in the data even when the spatially-explicit data-generating processes used to simulate the data were not included in the estimation framework.

Yule (1926) demonstrated the need to account for serial correlation in data when calculating correlation coefficients many years ago. Since then, a plethora of research has been published on best practices to avoid spurious correlations (e.g., Katz, 1988; Pyper and Peterman, 1998). For some reason, ecologists often choose to ignore these best practices. The use of MARSS models to estimate interactions is not new, nor are prewhitening procedures to create white-noise time series. The novelty of the work presented here comes from the comparison of the two methods, which had not previously been done. The MARSS framework outperformed prewhitening procedures when observation error was negligible and the time series was moderate in length. Consequently, MARSS models offered an ideal framework for investigating indicator-attribute relationships using output from ecosystem models as these did not include observation error. Results highlighted the need to engage in ecosystem-specific selection of indicators because of the inability of indicators to capture key losses across different types of ecosystems.

Biomass and TL from survey data were the best indicators of those investigated at tracking attributes thought to be of interest to managers responsible for EBFM. All objectives relevant to EBFM were not captured by the five attributes included in this analysis and the best indicators will

be situational given objectives and ecosystem type. Nevertheless, the goal was to provide a framework for assessing indicator-attribute relationships and demonstrate the implications of ignoring the time-series properties of attributes and indicators.

Spatial correlation is also frequently encountered and subsequently ignored in ecological data. There are many methods to account for space, where some detect large-scale patterns and others account for local autocorrelations (Legendre and Legendre, 1998; Haining, 2003). The benefits of accounting for space when data are georeferenced are well established (e.g., Thorson *et al.*, 2015b, 2015c; Thorson and Barnett, 2017). Benefits with respect to fisheries management include improved precision of estimates of relative abundance, as compared to traditional design-based estimators. Recently, the capacity has been added for spatial habitat covariates in spatiotemporal index standardization models or their inclusion has been advocated for. The practical implications of this were tested in Chapter 3.

Simulation was used to evaluate the benefits and repercussions of including a habitat covariate when the true process was and was not governed by habitat. Estimates of linear covariates were unbiased. In contrast, quadratic terms were biased, particularly for the catch-rate component of the model. Nevertheless, the trend of the index was well estimated and incorrectly including a habitat covariate when it did not govern the data-generating process was less problematic than not including the covariate when it should have been included.

The same framework was used to estimate indices of abundance for four species managed by the PFMC within the US West Coast Groundfish LE trawl fishery. Importantly, the use of the spatiotemporal-index standardization model allowed information on the biological status of bycatch species to be related to economic characteristics of vessels at the port-group level, rather than aggregated across the fishery as has previously been done.

Reducing bycatch of non-target individuals was emphasized as a key component of limiting the effects of fisheries to the broader ecosystem in early legislation regarding EBFM. Switching gears or shifting the spatial footprint of the fishery are often cited as the two most-feasible methods to limiting bycatch while allowing the fishery to remain open. Generalized linear models were used to estimate socio-economic factors that led fishers in the US West Coast Groundfish LE trawl fishery to switch from using trawl-gear to fixed-gear to land sablefish since management allowed such gear switching in 2011. Results indicated that the presence of bycatch species was not a major factor in their decisions to use fixed-gear. Instead, unmodeled factors related to the major port

groups included in the analysis were the best predictor of the proportion of landings that were caught using fixed-gear compared to trawl-gear. It was hypothesized that social and logistical challenges specific to each port group could limit fishers' ability or desire to switch gear. For example, vessels assigned to Oregon ports may feel pressured by local governance structures to fish using trawl-gear because Astoria and Warrenton generate the highest trawl revenue of all ports along the US West Coast (Sepez *et al.*, 2006). Additionally, vessels could be limited to trawl-gear because ports may not have adequate facilities to process catch from fixed-gear

This final chapter gathered lessons learned from each of the previous chapters to operationalize EBFM using statistical tools. Results will be of interest to those conducting future analysis on the implications of management strategies because they are informative about management uncertainty. Management uncertainty is the inability to predict how resource users will react to management strategies. Thus, under the precautionary approach to management, characterization of this uncertainty is vital to the sustainable management of marine fishes. Quantitative analyses such as those undertaken here to characterize management uncertainty related to gear switching in the US West Coast Groundfish LE trawl fishery are informative for shifting the burden of proof that all fisheries lead to ecosystem consequences that are to be avoided, thereby helping keep fisheries open.

5.2   FUTURE WORK

Multivariate auto-regressive state-space models implemented using the MARSS package in R are limited in their flexibility. The package allows for the use of the Gaussian distribution and no other options are currently parameterized because the framework uses a Kalman Filter. Bayesian implementation of the model-fitting procedure is also not available within the package. Lastly, model fitting can be prohibitively slow. Future work could investigate the use of the VAST package to fit similar models. VAST allows for several non-normal distributions and can be implemented with or without the associated spatial field. Most importantly, VAST is based on the Gompertz model (Dennis and Taper, 1994) the same as MARSS. Parameterizing the interactions between categories or species may not currently be as flexible as the MARSS framework but it could be explored.

Three additional lines of research should be investigated with respect to the VAST model. First, the compound Poisson-gamma distributions (e.g., Lecomte *et al.*, 2013; Thorson, 2017)

could be explored. The compound Poisson-gamma distribution mimics the process of capturing clusters of biomass, thereby changing the hypotheses about the mean-variance relationship of the data. The mean-variance relationship was assumed to be consistent in the simulations because fisheries research tends to favor delta-models (Steffanson, 1996). The delta-model limited the analysis in terms of comparing the estimated parameters across the two components of the model because the logit-link used to model encounters requires the explanation of the expectation in terms of odds ratios. In contrast, the catch-rate model can be explained in terms of percent change. This limitation becomes increasingly important when the signs of the effects are inconsistent across model components (e.g., depth was positively and negatively related to pink shrimp (*Farfantepenaeus duorarum*) occupancy and catch-rate models, respectively; Rubec *et al.*, 2016). The compound Poisson-gamma and Poisson-link models eliminate this problem by using a single model to both account for the probability of encounter and the number observed (Thorson, 2017). Thus, shared inference with respect to relationships between the habitat covariate and encounters and positive counts would result from the use of a single parameter.

Second, boundary conditions of the spatial knots used to approximate the spatial processes in the model should be explored. The number of knots were both increased and decreased when fitting the empirical data, but the extent of the study area was never changed. Additionally, the study area contains regions with islands that have positive depths. Knots can be configured such that portions of the study area are excluded from the analysis, but this was not done here.

Third, other proxies for habitat should be investigated. Depth was investigated because it is commonly available for most species through either navigational studies or as a measurement collected in conjunction with survey tows. Temperature as a proxy for the availability of thermoregulated prey (e.g., Maravelias and Reid, 1997), thermocline depth-to-bottom-depth ratio as a proxy for water column production (e.g., Rooper and Martin, 2011), and presence of chlorophyll *a* as a proxy for phytoplankton biomass and oceanic productivity (e.g., Heenan, *et al.*, 2016) could be explored.

A major limitation of Chapter 4 was the need to analyze fisher behavior at the port group level. Using vessel-specific data would address the issue of how to assign vessels to a single port group. For this analysis, vessel-port combinations were based on the port of highest ex-vessel revenue even though vessels may deliver to multiple ports. Vessel-specific information would also eliminate the need to include the number of vessels per port group in the model. However, it creates

a new problem of how to assign relative abundances of the resource and the environment to the user. Vessel monitoring systems could provide information on where fishers travel and thus the expected abundance of species they have access to. Using data from monitoring systems would limit retrospective analyses because the requirement for this technology aboard fishing vessels is relatively new and not universal.

BIBLIOGRAPHY

Agostini, V.N., Francis, R.C., Hollowed, A.B., Pierce, S.D., Wilson, C., and Hendrix, A.N. 2006. The relationship between Pacific hake (*Merluccius productus*) distribution and poleward subsurface flow in the California Current System. Canadian Journal of Fisheries and Aquatic Sciences, 63: 2648-2659. 10.1139/f06-139.

Ainsworth, C.H., Schirripa, M.J., and Morzaria-Luna, H. 2015. An Atlantis ecosystem model for the Gulf of Mexico supporting Integrated Ecosystem Assessment. U.S. Department of Commerce, NOAA Technical Memo NMFS-SEFSC-676, 161 p.

Ajzen, I. 1991. The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2): 179-211.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F. Csáki (Eds). 2nd International Symposium on Information Theory. Akadémiaj Kiadó, Budapest. pp 267-281.

Allen, P.M., and McGlade, J.M. 1987. Modelling complex human systems: a fisheries example. European Journal of Operational Research, 30: 147-167.

Andersen, B.S., Ulrich, C., Eigaard, O.R., and Christensen, A-S. 2012. Short-term choice behavior in a mixed-fishery: investigating metier selection in the Danish gillnet fishery. ICES Journal of Marine Science, 69: 131-143.

Axelsen, B.E., and Johnsen, E. 2015. An evaluation of the bottom trawl surveys in the Benguela Current Large Marine Ecosystem. Fisheries Oceanography, 24: 74-87.

Bahn, V., and McGill, B.J. 2007. Can niche-based distribution models outperform spatial interpolation? Global Ecology and Biogeography, 16(6): 733-742. 10.1111/j.1466-8238.2007.00331.x.

Bartlett, M.S. 1946. On the theoretical specification and sampling properties of autocorrelated time series. Supplement to the Journal of the Royal Statistical Society, 8(1): 27-41.

Bayley, G.V., and J.M. Hammersley. 1946. The "effective" number of independent observations in an autocorrelated time series. Supplement to the Journal of the Royal Statistical Society, 8(2): 184-197.

Beamish, R.J. 1995. Climatic change and northern fish populations. NRC Research Press, Victoria.

Bell, R.J., Fogarty, M.J., and Collie, J.S. 2014. Stability in marine fish communities. Marine Ecology Progress Series, 504: 221-239. 10.3354.meps10730.

Bell, R.J., Collie, J.S., Branch, T.A., Fogarty, M.J., Minto, C., and Ricard, D. 2018. Changes in the size structure of marine fish communities. ICES Journal of Marine Science, 75(1): 102-112. 10.1093/icesjms/fsx118.

Béné, C., Arthur, R., Norbury, H., Allison, E.H., Beveridge, M., Bush, S., Campling, L., *et al*. 2016. Contribution of fisheries and aquaculture to food security and poverty reduction: assessing the current evidence. World Development, 79: 177-196.

Bertrand, S., Díaz, E., and Ñiquen, M. 2004. Interactions between fish and fisher's spatial distribution and behavior: an empirical study of the anchovy (*Engraulis ringens*) fishery of Peru. ICES Journal of Marine Science, 61: 1127-1136.

Bigelow, K.A., and Maunder, M.N. 2007. Does habitat or depth influence catch rates of pelagic species? Canadian Journal of Fisheries and Aquatic Sciences, 64(11): 1581-1594. 10.1139/f07-115.

Blanchard, J.L., Coll, M., Trenkel, V.M., Vergnon, R., Yemane, D., Jouffre, D., Link, J.S., *et al*. 2010. Trend analysis of indicators: a comparison of recent changes in the status of marine ecosystems around the world. ICES Journal of Marine Science, 67: 732-744. 10.1093/icesjms/fsp282.

Boldt, J.L., Martone, R., Samhouri, J., Perry, R.I, Itoh, S., Chung, I.K., Takahashi, M. *et al*. 2014. Developing ecosystem indicators for responses to multiple stressors. Oceanography, 27: 116-133.

Borja, A., Elliott, M., Andersen, J.H., Cardoso, A.C., Carstensen, J., Ferreira, J.G., Heiskanen, A-S., *et al*. 2013. Good environmental status of marine ecosystems: what is it and how do we know when we have attained it? Marine Pollution Bulletin, 76(1-2): 16-27. 10.1016/j.marpolbul.2013.08.042.

Botsford, L.W., Castill, J.C., and Peterson, C.H. 1997. The management of marine ecosystems. Science, 27(5325): 509-515. 10.1126/science.277.5325.509.

Box, G.E.P., and G.M. Jenkins. 1976. Time Series Analysis: Forecasting and Control. Holden-Day San Francisco, California.

Bradshaw, A.D. 1984. Ecological principles and land reclamation practice. Landscape Planning, 11(1): 35-48. 10.1016/0304-3924(84)90016-9.

Brodziak, J., and Walsh, W.A. 2013. Model selection and multimodel inference for standardizing catch rates of bycatch species: a case study of oceanic whitetip shark in the Hawaii-based longline fishery. Canadian Journal of Fisheries and Aquatic Sciences, 70: 1723-1740. 10.1139/cjfas-2013-0111.

Bradburn, M.J., Keller, A., and Horness, B.H. 2011. The 2003 to 2008 US West Coast bottom trawl surveys of groundfish resources off Washington, Oregon, and California: estimates of distribution, abundance, length, and age composition. US Department of Commerce, NOAA Technical Memo NMFS-NWFSC-114, 323 p.

Branch, T.A., Hilborn, R., and Bogazzi, E. 2005. Escaping the tyranny of the grid: a more realistic way of defining fishing opportunity. Canadian Journal of Fisheries and Aquatic Sciences, 62: 631-642.

Branch, T.A., Hilborn, R., Haynie, A.C., Fay, G., Flynn, L., Griffiths, J., Marshall, K.N., *et al*. 2006. Fleet dynamics and fishermen behavior: lessons for fisheries managers. Canadian Journal of Fisheries and Aquatic Sciences, 63: 1647-1668.

Branch, T.A. 2009. How do individual transferable quotas affect marine ecosystems? Fish and Fisheries, 10: 39-57.

Buonaccorsi, J.P., J.S. Elkinton, S.R. Evans, and A.M. Liebhold. 2001. Measuring and testing for spatial synchrony. Ecology, 82(6): 1668-1679.

Burgess, M.G., Giacomini, H.C., Szuwalski, C.S., Costello, C., and Gaines, S.D. 2017. Describing ecosystem contexts with single-species models: a theoretical synthesis for fisheries. Fish and Fisheries, 18: 264-284. 10.1111/faf/12179.

Burnham, K.P., and Anderson, D.R. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, second ed. Springer, New York.

Burnham, K.P., Anderson, D.R., and Huyvaert, K.P. 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behavioral Ecology and Sociobiology, 65: 23-35. 10.1007/s00265-010-1029-6.

Cadigan, N.G., and Myers, R.A. 2001. A comparison of gamma and lognormal maximum likelihood estimators in a sequential population analysis. Canadian Journal of Fisheries and Aquatic Sciences, 58: 560-567. 10.1139/cjfas-58-3-560.

Cao, J., Thorson, J.T., Richards, A., and Chen, Y. 2017. Spatiotemporal index standardization improves the stock assessment of northern shrimp in the Gulf of Maine. Canadian Journal of Fisheries and Aquatic Sciences, 74: 1781-1793. 10.1139/cjfas-2916-0137.

Casey, K.E., Dewees, C.M., Turris, B.R., and Wilen, J.E. 1995. The effects of individual vessel quotas in the British Columbia Halibut Fishery. Marine Resource Economics, 110: 211-230.

Cheung, W.W.L., Pitcher, T.J., and Pauly, D. 2005. A fuzzy logic expert system to estimate intrinsic extinction vulnerabilities of marine fishes to fishing. Biological Conservation, 124(1): 97-111. 10.1016/j.biocon.2005.01.017.

Christensen, V. 1996. Managing fisheries involving predator and prey species. Reviews in Fish Biology and Fisheries, 6(4): 417-442.

Cinner, J.E. 2007. Designing marine reserves to reflect local socioeconomic conditions: lessons from long-enduring customary management systems. Coral Reef, 26: 1035-1045.

Cochran, W.G. 1977. Sampling Techniques, third ed. John Wiley and Sons, New York.

Coll, M., Shannon, L.J., Yemane, D., Link, J.S., Ojaveer, H., Neira, S., Jouffre, D. *et al*. 2010. Ranking the ecological relative status of exploited marine ecosystems. ICES Journal of Marine Science, 67: 769-786.

Coll, M., Shannon, M.J., Kleisner, K.M., Juan-Jordá, M.J., Bundy, A., Akoglu, A.G., Banaru, D. *et al*. 2016. Ecological indicators to capture the effects of fishing on biodiversity and conservation status of marine ecosystems. Ecological Indicators, 60: 947-962. 10.1016/j.ecolind.2015.08.048.

Cooke, S.J., Nguyen, V.M., Murchie, K., Danylchuk, A.J., and Suski, C.D. 2012. Scientific and stakeholder perspectives on the use of circle hooks in recreational fisheries. Bulletin of Marine Science, 88: 395-410.

Cottingham, K.L., and Carpenter, S.R. 1998. Population, community, and ecosystem variates as ecological indicators: phytoplankton responses to whole-lake enrichment. Ecological Applications, 8: 508-530.

Cox, M. 2011. Advancing the diagnostic analysis of environmental problems. International Journal of the Commons, 5: 346-363.

Crain, C.M., Kroeker, K., and Halpern, B.S. 2008. Interactive and cumulative effects of multiple human stressors in marine systems. Ecology Letters, 11: 1304-1315. 10.111/j.1461-0248.2008.01253.x.

Darling, E.S., and Côté, I.M. 2008. Quantifying the evidence for ecological synergies. Ecology Letters, 11: 1278-1286. 10.1111/j.1461-0248.2008.01243.x.

Dennis, B., and Taper, M.L. 1994. Density dependence in time series observations of natural populations: estimation and testing. Ecological Monographs, 64: 205-224.

Dichmont, C.M., Fulton, E.A., Gorton, R., Sporcic, M., Little, L.R., Punt, A.E., Dowling, N., *et al*. 2016. From data rich to data-limited harvest strategies—does more data mean better management?. ICES Journal of Marine Science, 74(3): 670-686. 10.1093.icesjms/fsw199.

Doorman, C.F. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Global Ecology and Biogeography, 16(2): 129-138. 10.1111/j.1466-8238.2006.00279.x.

Dulvy, N.K., Rogers, S.I., Jennings, S., Stelzenmüller, V., Dye, S.R., and Skjoldal, H.R. 2008. Climate change and deepening of the North Sea fish assemblage: a biotic indicator of warming seas. Journal of Applied Ecology, 45: 1029-1039.

Eggert, H., and Martinsson, P. 2004. Are commercial fishers risk-lovers?. Land Economics, 80: 550-560. 10.2307/3655810.

Eliasen, S.Q., Papadopoulou, K-N., Vassilopoulou, V., and Catchpole, T.L. 2013. Socio-economic and institutional incentives influencing fishers' behavior in relation to fishing practices and discard. ICES Journal of Marine Science, 71: 1298-1307.

Elton, C., and M. Nicholson. 1942. The ten-year cycle in numbers of the lynx in Canada. Journal of Animal Ecology, 11(2): 215-244.

Eschmeyer, W.N., Herald, E.S., and Hammann, H. 1983. A Field Guide to Pacific Coast Fishes of North America. Houghton Mifflin Company, Boston.

Essington, T.E, and Plagányi, E.E. 2014. Pitfalls and guidelines for "recycling" models for ecosystem-based fisheries management: evaluating model suitability for forage fish fisheries. ICES Journal of Marine Science, 71: 118-127. 10.1093/icesjms/fst047.

Evans, M.R., Norris, K.J., and Benton, T.G. 2011. Predictive ecology: systems approaches. Philosophical Transactions of the Royal Society B, 367(1586): 163-169. 10.1098/rstb.2011.0191.

Fay, G., Link, J.S., Large, S.I., and Gamble, R.J. 2015. Management performance of ecological indicators in the Georges Bank finfish fishery. ICES Journal of Marine Science, 72(5): 1285–1296. 10.1093/icesjms/fsu214.

Firth, D. 1988. Multiplicative errors: log-normal or gamma? Journal of the Royal Statistical Society Series B Statistical Methodology, 50(2): 266-268.

Food and Agriculture Organization of the United World Nations (FAO). 2018. FishStat Plus – universal software for fishery statistical time series. FAO Fisheries and Aquaculture Department, Rome.

Food and Agriculture Organization of the United World Nations (FAO). 2016. The state of world fisheries and aquaculture 2016. Contributing to food security and nutrition for all. FAO Fisheries and Aquaculture Department, Rome.

Fuglstad, G-A., Simpson, D., Lindgren, F., and Rue, H. 2015. Does non-stationary spatial data always require non-stationary random fields? Spatial Statistics, 14(C): 505-531. 10.1016.j.spasta.2015.10.001.

Fuglstad, G-A., Simpson, D., Lindgren, F., and Rue, H. 2018. Constructing priors that penalize the complexity of Gaussian random fields. Journal of the American Statistical Association. 10.1080/01621459.2017.1415907.

Fulton, E.A., Parslow, J.S., Smith, A.D.M., and Johnson, C.R. 2004. Biogeochemical marine ecosystem models II: the effect of physiological detail on model performance. Ecological Modelling, 173(4): 371-406. 10.1016/j.ecolmodel.2003.09.024.

Fulton, E.A., Smith, A.D.M., and Punt, A.E. 2005. Which ecological indicators can robustly detect effects of fishing? ICES Journal of Marine Science, 62: 540-551.

Fulton, E.A., Smith, A.D.M., and van Putten, I.E. 2010. Human behavior: the key source of uncertainty in fisheries management. Fish and Fisheries, 12: 2-17.

Fulton, E.A., Link, J.S, Kaplan, I.C, Savina-Rolland, M., Johnson, P., Ainsworth, C., Horne, P. *et al.* 2011. Lessons in modelling and management of marine ecosystems: the Atlantis experience. Fish and Fisheries, 12: 171-188. 10.1111/j.1467-2979.2011.00412.x.

Fulton, E.A., Smith, A.D.M., Smith, D.C., and Johnson, P. 2014. An integrated approach is needed for ecosystem based fisheries management: insights from ecosystem-level management strategy evaluation. PLoS ONE, 9(1): e84242. 10.1371/journal.pone.0084242.

Galton, F. 1888. Co-relations and their measurement, chiefly from anthropometric data. Proceedings of the Royal Society of London, 45: 135-145.

Gascuel, D., Coll, M., Fox, C., Guénette, S., Guitton, J., Kenny, A., Knittweis, L. *et al*. 2016. Fishing impact and environmental status in European seas: a diagnosis from stock assessments and ecosystem indicators. Fish and Fisheries, 17(1): 31-55. 10.1111/faf.12090.

Giraldo, C., Ernande, B., Cresson, P., Kopp, D., Cachera, M., Travers-Trolet, M., and Lefebvre, S. 2017. Depth gradient in the resource use of a fish community from a semi-enclosed sea. Limnology and Oceanography, 62: 2213-2226. 10.1002/lno.10561.

Grafton, R.Q., Arnason, R., Bjørndal, T., Campbell, D., Campbell, H.F., Clark, C.W., Connor, R., *et al*. 2006. Incentive-based approaches to sustainable fisheries. Canadian Journal of Aquatic and Fisheries Sciences, 63: 699-710. 10.1139/F05-247.

Graham, M.H. 2003. Confronting multicollinearity in ecological multiple regression. Ecology, 84(11): 2809-2815.

Granger, C.W.J., and P. Newbold. 1974. Spurious regressions in econometrics. Journal of Econometrics, 2: 111-120.

Greenstreet, S.P.R., Rossber, A.G., Fox, C.J., Le Quesne, W.J.F., Blasdale, T., Boulcott, P., Mitchell, I., Millar, C., and Moffat, C.F. 2012. Demersal fish biodiversity: species-level indicators and trends-based targets for the marine Strategy Framework Directive. ICES Journal of Marine Science, 69: 1789-1801.

Gunderson, L.H., and Holling, C.S. 2002. Panarchy: Understanding Transformations in Human and Natural Systems. Island Press, Washington, DC.

Haining, R. 2003. Spatial Data Analysis. Theory and Practice. Cambridge University Press, Cambridge.

Hall, M.A., Alverson, D.L., and Metuzals, K.I. 2000. By-catch: problems and solutions. Marine Pollution Bulletin, 41: 204-219.

Hamel, O.S. 2008. Status and future prospects for the darkblotched rockfish resource in waters off Washington, Oregon, and California as assessed in 2007. Pacific Fisheries Management Council, Portland.

Hampton, S.E., Holmes, E.E., Scheef, L.P., Scheuerell, M.D., Katz, S.L., Pendleton, D.E., and Ward, E.J. 2012. Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models. Ecology, 94(12): 2663-2669. 10.1890/13-0996.1.

Hannah, R.W., Parker, S.J., and Buell, T.V. 2005. Evaluation of a selective flatfish trawl and diel variation in rockfish catchability as bycatch reduction tools in the deepwater complex fishery off the U.S. West Coast. North American Journal of Fisheries Management, 25(2): 581-593. 10.1577/M04-126.1.

Hansen, C., Skern-Mauritzen, M., van der Meeren, G.I., Jähkel, A., and Drinkwater, J.A. 2016. Set-up of the Nordic and Barens Sea (NoBa) Atlantis model. Havforskningsinstituttet Institute of Marine Research, Bergen, Norway. 112 p.

Harms, J.H., Wallace, J.R., and Stewart, I.J. 2010. Analysis of fishery-independent hook and line-based data for use in the stock assessment of bocaccio rockfish (*Sebastes paucispinis*). Fisheries Research, 106(3): 298-309. 10.1016/j.fishres.2010.08.010.

Harris, D.J. 2015. Generating realistic assemblages with a joint species distribution model. Methods in Ecology and Evolution, 6(4): 465-473. 10.1111/2041-210X.12332.

Harvey, C., and Garfield, T. 2017. California Current Integrated Ecosystem Assessment (CCIEA) California Current ecosystem status report, 2017. Pacific Fisheries Management Council, Portland, OR. 22 p. downloaded from pcouncil.org/wp-content/uploads/2017/02/F1a_NMFS_Rpt1_2017IEA_Main_Rpt_Final_Mar2017BB.pdf.

Hart, P.J.B., and Pitcher, T.J. 1998. Conflict, consent and cooperation: the utility of an evolutionary perspective on individual human behavior in fisheries management. In Reinventing Fisheries Management, pp. 215-225. Ed by T. J. Pitcher, D. Pauly, and P. Hart. Kluwer Academic Publishers, Dordrecht, Netherlands, 415 p.

Hattam, C., Atkins, J.P., Beaumont, N., Börger, T., Böhnke-Henrichs, A., Burdon, D., de Groot, R. *et al*. 2012. Marine ecosystem services: linking indicators to their classification. Ecological Indicators, 49: 61-75. 10.1016/j.ecolind.2014.09.026.

Heenan, A., Hoey, A.S., Williams, G.J., and Wiliams, I.D. 2016. Natural bounds on herbivorous coral reef fishes. Proceedings of the Royal Society B, 283(1843): 20161716. 10.1098/rspb.2016.1716.

Heino, M. 1998. Noise colour, synchrony and extinctions in spatially structured populations. Oikos, 83: 368-375.

Heino, J. 2010. Are indicator groups and cross-taxon congruence useful for predicting biodiversity in aquatic ecosystems? Ecological Indicators, 10(2): 112-117. 10.1016/j.ecolind.2009.04.013.

Helser, T.E., Punt, A.E., and Methot, R.D. 2004. A generalized linear mixed model analysis of a multi-vessel fishery resource survey. Fisheries Research, 70(2-3):251-264. 10.1016/j.fishres.2004.08.007.

Heymans, J.J., Coll, M., Libralato, S., Morissette, L., and Christensen, V. 2014. Global patterns in ecological indicators of marine food webs: a modelling approach. PLoS ONE, 9(4): e95845. 10.1371/journal.pone.0095845.

Hilborn, R. 1985. Fleet dynamics and individual variation: why some people catch more fish than others. Canadian Journal of Fisheries and Aquatic Sciences, 42: 2-13.

Hilborn, R. 2012. The evolution of quantitative marine fisheries management 1985-2010. Natural Resource Modeling, 25(1): 122-144.

Hilborn, R., Orensanz, J.M., and Parma, A.M., 2005. Institutions, incentives and the future of fisheries. Philosophical Transaction of the Royal Society B, 360: 47-57.

Hinrichsen, R.A., and Holmes, E.E. 2009. Using multivariate state-space models to study spatial structure and dynamics. In: R.S. Cantrell, C. Cosner, and S. Ruan (Eds.), Spatial Ecology. Chapman Hall, Boca Raton.

Hinton, M.G., and Maunder, M.N. 2004. Status of striped marlin in the eastern Pacific Ocean in 2002 and outlook for 2003-2004. Inter-American Tropical Tuna Commission Stock Assessment Report, 4: 287-310.

Hjort, J. 1914. Fluctuations in the great fisheries of northern Europe, viewed in the light of biological research. Rapports et Procès-Verbaux Conseil Permanent International Pour L'Exploration de la Mer, 20: 1-228.

Hobday, A.J., and Hartmann, K. 2006. Near real-time spatial management based on habitat preferences for a longline bycatch species. Fisheries Management and Ecology, 13(6): 365-380. 10.1111/j.1365-2400.2006.00515.x.

Holland, D.S., Sanchirico, J.N., Johnston, R.J., and Joglekar, D. 2010. Economic Analysis for Ecosystem-Based Management: Applications to Marine and Coastal Environments. Earthscan, London.

Holland, D.S., and Sutinen, J.G. 1999. An empirical model of fleet dynamics in New England trawl fisheries. Canadian Journal of Fisheries and Aquatic Sciences, 56: 253-264.

Holland, D.S., and Jannot, J.E. 2012. Bycatch risk pools for the US West Coast Groundfish Fishery. Ecological Economics, 78:132-147.

Holling, C.S., Berkes, F., and Folke, C. 2000. Science, sustainability and resource management. In Linking Social and Ecological Systems: Management and Practices and Social mechanisms for Building Resilience, pp. 342-362. Ed. by F. Berkes, and C. Folke. Cambridge University Press, Cambridge, UK. 459 p.

Hollowed, A.B., Bax, N., Beamish, R., Collie, J., Fogarty, M., Livingston, P., Pope, J., and Rice, J.R. 2000. Are multispecies models an improvement on single-species models for measuring fishing impacts on marine ecosystems?. ICES Journal of Marine Science, 57: 707-719. 10.1006/jmsc.2000.0734.

Hollowed, A.B., S.R. Hare, and W.S. Wooster. 2001. Pacific Basin climate variability and patterns of Northeast Pacific marine fish production. Progress in Oceanography, 49(1-4): 257-282.

Holmes, E.E., E.J. Ward, and K. Wills. 2012. MARSS: Multivariate autoregressive state-space models for analyzing time-series data. The R Journal, 4(1): 11-19.

Hunt, L.M. 2005. Recreational fishing site choice models: insights and future opportunities. Human Dimensions of Wildlife, 10: 153-172.

Hutchinson, G.E. 1957. Concluding remarks. Cold Spring Harbor Symposia Quantitative Biology, 22: 415-427. 10.1101/SQB.1957.022.01.039.

Huxley, T.H. 1883. Address: inaugural meeting of the Fishery Congress. International Fisheries Exhibition, London.

Ihde, T.F., and Townsend, H.M. 2017. Accounting for multiple stressors influencing living marine resources in a complex estuarine ecosystem using an Atlantis model. Ecological Modelling, 365: 379-389.

Ives, A.R. 1995. Predicting the response of populations to environmental change. Ecology, 76: 926-941. 10.2307/1939357.

Ives, A.R., Carpenter, S.R., and Dennis, B. 1999. Community interaction webs and zooplankton response to planktivory manipulations. Ecology, 80: 1405-1421.

Ives, A.R., Dennis, B., Cottingham, K.L., and Carpenter, S.R. 2003. Estimating community stability and ecological interactions from time-series data. Ecological Monographs, 73: 301-330. 10.1890/0012-9615(2003)073[0301:ECSAEI]2.0.CO;2.

Ives, A.R., and V. Dakos. 2012. Detecting dynamical changes in nonlinear time series using locally linear state-space models. Ecosphere, 3: 1-15.

Jannot, J.E., and Holland, D.S. 2013. Identifying ecological and fishing drivers of bycatch in a US groundfish fishery. Ecological Applications, 23:1645-1658.

Jenkins, L.D., and Garrison, K. 2013. Fishing gear substitution to reduce bycatch and habitat impacts: an example of social-ecological research to inform policy. Marine Policy, 38: 293-303.

Jennings, S. 2005. Indicators to support and ecosystem approach to fisheries. Fish and Fisheries, 6(3): 212-232. 10.1111/j.1467-2979.2005.00189.x.

Jennings, S., and Dulvy, N.K. 2005. Reference points and reference directions for size-based indicators of community structure. ICES Journal of Marine Science, 62: 397-404.

Jennings, S., Greenstreet, S., Hill, L., Piet, G., Pinnegar, J., and Warr, K.J. 2002. Long-term trends in the trophic structure of the North Sea fish community: evidence from stable-isotope analysis, size-spectra and community metrics. Marine Biology, 141(6): 1085-1097. 10.1007%2Fs00227-002-0905-7.

Jiming, Y. 1982. A tentative analysis of the trophic levels of North Sea fish. Marine Ecology Progress Series, 7(3): 247-252.

Jonsen, I.D., Myers, R.A., and Flemming, J.M. 2003. Meta-analysis of animal movement using state-space models. Ecology, 84(11): 3055-3063. 10.1890/02-0670.

Johnson, K.F., Rudd, M.R., Pons, M., Axelrud, C.A., Lee, Q., Hurtado-Ferro, F., Haltuch, M.A., *et al*. 2015. Status of the US sablefish resource in 2015. Pacific Fisheries Management Council. Portland, Oregon. 243 p.

Johnson, W.C., Millett, B.V., Gilmanov, T., Voldseth, R.A., Guntenspergen, G.R., and Naugle, D.E. 2005. Vulnerability of northern prairie wetlands to climate change. BioScience, 55(10): 863-872. 10.1641/0006-3568(2005)055[0863:VONPWT]2.0.CO;2.

Jolliffe, I.T. 2007. Uncertainty and inference for verification measures. Weather and Forecasting, 22(3): 637-650. 10.1175/WAF989.1.

Jouffre., D., Borges, M.F., Bundy, A., Coll, M., Diallo, I., Fulton, E.A., Guitton, J., Labrosse, P., Mohamed Abdellahi, K.O., Masumbuko, B., and Thiao, D. 2010. Estimating EAF indicators from scientific trawl surveys: theoretical and practical concerns. ICES Journal of Marine Science, 67: 796-806.

Kaplan, I.C., Horne, P.J., and Levin, P.S. 2012. Screening California Current fishery management scenarios using the Atlantis end-to-end ecosystem model. Progress in Oceanography, 102: 5-18. 10.1016/j.pocean.2012.03.009.

Katz, R.W. 1988. Use of cross correlations in the search for teleconnections. *Journal of* Climatology, 8: 241-253.

Keller, A.A., Wallace, J.R., Horness, B.H., Hamel, O.S., and Stewart, I.J. 2012. Variations in eastern North Pacific demersal fish biomass based on the U.S. west coast groundfish bottom trawl survey (2003-2010). Fisheries Bulletin, 110: 205-222.

Kershner, J., Samhouri, J.F., James, C.A., and Levin, P.S. 2011. Selecting indicator portfolios for marine species and food webs: a Puget Sound case study. PLOS ONE, 6(10): e25248. 10.1371/journal.pone.0025248.

Kleisner, K.M., Coll, M., Lynam, C.P., Bundy, A., Shannon, L., Shin, Y-J., Boldt, J.L. *et al.* 2015. Evaluating changes in marine communities that provide ecosystem services through comparative assessments of community indicators. Ecosystem Services, 16: 413-429. 10.1016/j.ecoser.2015.02.002.

Koenig, W.D. 1999. Spatial autocorrelation of ecological phenomena. Trends in Ecology and Evolution, 14(1): 22-26. 10.1016/S0169-5347(98)01533-X.

Kotwicki, S., Buckley, T.W., Honkalehto, T., and Walters, G. 2005. Variation in the distribution of walleye pollock (*Theragra chalcogramma*) with temperature and implications for seasonal migration. Fisheries Bulletin, 103(4): 574-587.

Kristensen, K., Thygesen, U.H., Andersen, K.H., and Beyer, J.E. 2014. Estimating spatio-temporal dynamics of size-structured populations. Canadian Journal of Fisheries and Aquatic Sciences, 71(2): 326-336. 10.1139/cjfas-2013-0151.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell, B.M. 2016. TMB: Automatic differentiation and Laplace approximation. Journal of Statistical Software, 70(5): 1-21. 10.18637/jss.v070.i05.

Kroeker, K.J., Kordas, R.L., Crim, R., Hendriks, I.E., Ramojo, L., Duarte, C.M., and Gattuso, J-P. 2013. Impacts of ocean acidification on marine organisms: quantifying sensitivities and interaction with warming. Global Change Biology, 19: 1884-1896. 10.1111/gcb.12179.

Lackey, R.T. 1998. Seven pillars of ecosystem management. Landscape and Urban Planning, 40: 21-30. 10.1016/S0169-2046(97)00095-9.

Larkin, 1996. Concepts and issues in marine ecosystem management. Reviews in Fish Biology and Fisheries, 6(2): 139-164.

Lecomte, J.B., Benoît, H.P., Etienne, M.P., Bel, L., Parent, E., 2013. Modeling the habitat associations and spatial distribution of benthic macroinvertebrates: a hierarchical Bayesian model for zero-inflated biomass data. Ecol. Modell. 265, 74-84. 10.1016/j.ecolmodel.2013.06.017.

Legendre, P., and Legendre, L. 1998. Numerical Ecology, second edition. Elsevier Science, Amsterdam.

Levin, P.S., Essington, T.E., Marshall, K.N., Koehn, L.E., Anderson, L.G., Bundy, A., Carothers, C., *et al*. 2018. Building effective fishery ecosystem plans. Marine Policy, 92: 48-57. 10.1016/j.marpol.2018.01.019.

Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. Bulletin of the Entomological Society of America, 15(3): 237-240. 10.1093/besa/15.3.237.

Lindgren, F., Rue, H., and Lindström, J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal

Statistical Society Series B Statistical Methodology, 73(4): 423-498. 10.1111/j.1467-9868.2011.00777.x.

Link, J.S. 2010. Ecosystem-Based Fisheries Management: Confronting Tradeoffs. Cambridge University Press, New York.

Link, J.S., Fulton, E.A., and Ramble, R.J. 2010. The Northeast US application of ATLANTIS: a full system model exploring marine ecosystem dynamics in a living marine resource management context. Progress in Oceanography, 87: 214-234. 10.1016/j.pocean.2010.09.020.

Link, J.S., Yemane, D., Shannon, L.J., Coll, M., Shin, Y-J., Hill, L., and Borges, M.F. 2010. Relating marine ecosystem indicators to fishing and environmental drivers: an elucidation of contrasting responses. ICES Journal of Marine Science, 67: 787-795.

Link, J.S., Gamble, R.J., and Futon, A.E. 2011. NEUS − Atlantis: construction, calibration, and application of an ecosystem model with ecological interactions, physiographic conditions, and fleet behavior. U.S. Department of Commerce, NOAA Technical Memo NFMS-NE-218, 254 p.

Lindgren, F., and Rue, H. 2015. Bayesian spatial modelling with R-INLA. Journal of Statistical Software, 63(19): 1-25. 10.18637/jss.v063.i19.

Ludwig, D., Walters, C.J. 1985. Are age-structured models appropriate for catch-effort data? Canadian Journal of Fisheries and Aquatic Sciences, 42: 1066-1072. 10.1139/f85-132.

MacNally, R., Thomson, J.R., Kimmerer, W.J., Feyrer, F., Newman, K.B., Sih, A., Bennett, W.A., *et al*. 2010. Analysis of pelagic species decline in the upper San Francisco Estuary using multivariate autoregressive modeling (MAR). Ecological Applications, 20(5): 1417-1430. 10.1890/09-1724.1.

Mahon, R., McConney, P., and Roy, R.N. 2008. Governing fisheries as complex adaptive systems. Marine Policy, 32: 104-112.

Manson, J. Kosaka, R., Mamula, A., and Speir, C. 2012. Effort changes around a marine reserve: the case of the California Rockfish Conservation Area. Marine Policy, 36: 1054-1063.

Maravelias, C.D., and Reid, D.G. 1997. Identifying the effects of oceanographic features and zooplankton on prespawning herring abundance using generalized additive models. Marine Ecology Progress Series, 147(1/3): 1-9.

Marshall, K.N., Kaplan, I.C., Hodgson, E.E., Hermann, A., Busch, D.S., McElhany, P., Essington, T.E., *et al*. 2017. Risks of ocean acidification in the California Current food web and fisheries: ecosystem model projections. Global Change Biology, 23: 1525-1539. 10.1111/gcb.13594.

Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J., *et al*. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecology Letters, 8(11): 1235-1246. 10.1111/j.1461-0248.2005.00826.x.

Maunder, M.N., and Punt, A.E. 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries Research, 70: 141-159. 10.1016/j.fishres.2004.08.002.

Maury, O., Campling, L., Arrizabalaga, H., Aumont, O., Bopp, L., Merino, G., Squires, D. *et al*. 2017. From shared socio-economic pathways (SSPs) to oceanic system pathways (OSPs): building policy-relevant scenarios for global oceanic ecosystems and fisheries. Global Environmental Change, 45: 203-216. 10.1016/j.gloenvcha.2017.06.007.

Masi, M.D., Ainsworth, C.H., and Jones, D.L. 2017. Using a Gulf of Mexico Atlantis model to evaluate ecological indicators for sensitivity to fishing mortality and robustness to observation error. Ecological Indicators, 74: 516-525. 10.1016/j.ecolind.2016.11.008.

McGinnis, M.D., and Ostrom, E. 2014. Social-ecological system framework: initial changes and continuing challenges. Ecology and Society, 19(2): 30. 10.5751/ES-06387-190230.

Megrey, B.A., Link, J.S., Hunt, G.L., Jr., and Moksness, E. 2009. Comparative marine ecosystem analysis: applications, opportunities, and lessons learned. Progress in Oceanography, 81(1-4): 2-9. 10.1016/j.pocean.2009.04.002.

Mehta, L., Leach, M., Newell, P., Scoones, I., Sivaramakrishnan, K., and Way, S-A. 1999. Exploring understandings of institutions and uncertainty: new directions in natural resource management. Institute of Development Studies, University of Sussex, Brighton, UK.

Methot, R.D., Lauth, R., Shaw, F., and Wilkins, M. 1994. Assessment of the West Coast sablefish stock in 1994. Pacific Fisheries Management Council, Portland.

Methratta, E.T., and Link, J.S. 2006. Evaluation of quantitative indicators for marine fish communities. Ecological Indicators, 6: 575-588.

Moran, P.A.P. 1950. Notes on continuous stochastic phenomena. Biometrika, 37(1-2): 17-23.

Morzaria-Luna, H., Ainsworth, C.H., Kaplan, I.C., Levin, P.S., and Fulton, E.A. 2013. Indirect effects of conservation policies on the coupled human natural ecosystem of the Gulf of California. PLOS ONE, 8(5): e64085. 10.1371/journal.pone.0064085.

Mueter, F.J., and Litzow, M.A. 2008. Sea ice retreat alters the biogeography of the Bering Sea continental shelf. Ecological Applications, 18(2): 309-320. 10.1890/07-0564.1.

Murawski, S.A., Steele, J.H., Taylor, P., Fogarty, M.J., Sissenwine, M.P., Ford, M., and Suchman, C. 2010. Why compare marine ecosystems? ICES Journal of Marine Science, 67: 1-9. 10.1093/icesjms/fsp221.

Myers, R.A., Hutchings, J.A., and Barrowman, N.J. 1997. Why do fish stocks collapse? The example of cod in Atlantic Canada. Ecological Applications, 7(1): 91-106.

National Marine Fishery Service (NMFS). 2004. West Coast Groundfish Observer Program data report and summary analyses for 2001-2003 coverage of sablefish-endorsed fixed-gear permits. NOAA National Marine Fisheries Service, Seattle.

National Marine Fisheries Service (NMFS), 2013. Groundfish essential fish habitat synthesis: a report to the Pacific Fishery Management Council. NOAA NMFS Northwest Fisheries Science Center, Seattle.

National Marine Fishery Service (NMFS). 2017. Fisheries economics of the United States, 2015. U.S. Department of Commerce, NOAA Technical Memo NFMS-F/SPO-170, 247 p.

National Research Council (NRC). 1998. A report of the Committee on Ecosystem Management for Sustainable Fisheries. National Academy Press, Washington D.C.

Niemeijer, D., and de Groot, R.S. 2008. A conceptual framework for selecting environmental indicator sets. Ecological Indicators, 8(1): 14-25. 10.1016/j.ecolind.2006.11.012.

North Pacific Fishery Management Council (NPFMC). 2017. Fishery Management Plan for groundfish of the Bering Sea and Aleutian Islands management area. North Pacific Fishery Management Council, Anchorage.

Noss, R.F. 1990. Indicators for monitoring biodiversity: a hierarchical approach. Conservation Biology, 4: 355-364.

Nye, J.A., and Link, J.S. 2009. Changing spatial distribution of fish stocks in relation to climate and population size on the Northeast United States continental shelf. Marine Ecology Progress Series, 393: 111-129.

Odum, E.P. 1985. Trends expected in stressed ecosystems. BioScience, 35(7): 419-422.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G. L., Solymos, P., Stevens, M.H.H., and Wagner, H. 2015. vegan: community ecology package. R package version 2.2-2/r2945. http://R-Forge.R-project.org/projects/vegan.

Olden, J.D., and B.D. Neff. 2001. Cross-correlation bias in lag analysis of aquatic time series. Marine Biology, 138: 1063-1070.

Olsen, E., Kaplan, I.C., Ainsworth, C., Fay, G., Gaichas, S., Gamble, R., Girardin, R. *et al*. 2018. Ocean futures as explored using a worldwide suite of ecosystem models. Frontiers in Marine Science, 10.3389/fmars.2018.00064.

Orcutt, G.H. and James, S.F. 1948. Testing the significance of correlation between time series. Biometrika, 35(3-4): 397-413. 10.2307/2332358.

Ostrom, E. 2009. A general framework for analyzing sustainability of social-ecological systems. Science, 325(5939): 419-422. 10.1126/science.1172133.

Pacific Fishery Management Council (PFMC). 2003. Amendment 16-2 to the Pacific Coast Groundfish Fishery Management Plan. Pacific Fishery Management Council, Portland.

Pacific Fishery Management Council (PFMC). 2004. Proposed acceptable biological catch and optimum yield specifications and management measures for the 2005-2006 Pacific Coast groundfish fishery. Final environmental impact statement and regulatory analyses. Pacific Fishery Management Council, Portland.

Pacific Fishery Management Council (PFMC). 2016. Pacific Coast Groundfish fishery management plan: for the California, Oregon, and Washington groundfish fishery. Pacific Fishery Management Council, Portland.

Pacific Fishery Management Council (PFMC). 2010. Groundfish fishery management plan amendment 20 (trawl rationalization). Pacific Fishery Management Council, Portland.

PacFIN. 2016. PacFIN (Pacific Fisheries Information Network). Pacific States Marine Fisheries Commission, Portland. www.psmfc.org.

Pálsson, G., and Durrenberger, E.P. 1982. To dream of fish: the causes of Icelandic skippers' fishing success. Journal of Anthropology Research, 38: 227-242.

Pauly, D., and Palomares, M.L. 2005. Fishing down marine food web: It is far more pervasive than we thought. Bulletin of Marine Science, 76: 197-211.

Pauly, D., and Watson, R. 2005. Background and interpretation of the 'Marine Trophic Index' as a measure of biodiversity. Philosophical Transactions of the Royal Society B, 360(1454): 415-423. 10.1098/rstb.2004.1597.

Perry, A.L., Low, P.J., Ellis, J.R., and Reynolds, J.D. 2005. Climate change and distribution shifts in marine fishes. Science, 308: 1912-1915.

Petchey, O.L., A. Gonzalez, and H.B. Wilson. 1997. Effects on population persistence: the interaction between environmental noise colour, intraspecific competition and space. Proceedings of the Royal Society London B Biological Sciences, 264: 1841-1847.

Pikitch, E.K., Santora, C., Babcock, E.A., Bakun, A., Bonfil, R., Conover, D.O., Dayton, P., *et al*. 2004. Ecosystem-based fishery management. Science, 305: 346-347.

Pikitch, E.K., Rountos, K.J., Essington, T.E., Santora, C., Pauly, D., Watson, R., Sumaila, U.R., *et al*. 2014. The global contribution of forage fish to marine fisheries and ecosystems. Fish and Fisheries, 15: 43-64.

Pimm, S.L. 1991. The Balance of Nature? Ecological Issues in the Conservation of Species and Communities. The University of Chicago Press, Chicago.

Plagányi, É.E., Punt, A.E., Hillary, R., Morello, E.B., Thebaud, O., Hutton, T., Pillans, R.D., *et al*. 2012. Multispecies fisheries management and conservation: tactical applications using models of intermediate complexity. Fish and Fisheries, 15: 1-22.

Probst, W.N., Stelzenmüller, V., and Fock, H.O. 2012. Using cross-correlations to assess the relationship between time-lagged pressure and state indicators: an exemplary analysis of North Sea fish population indicators. ICES Journal of Marine Science, 69: 670-681. 10.1093/icesjms/fss015.

Probst, W.N., Stelzenmüller, V., and Kraus, G. 2013. A simulation-approach to assess the size structure of commercially exploited fish populations within the European Marine Strategy Framework Directive. Ecological Indicators, 24: 621-632.

Punt, A.E., and Hilborn, R. 1994. A comparison of fishery models with and without cannibalism with implications for the management of the Cape hake resource off southern Africa. ICES Journal of Marine Science, 51: 19-29.

Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A., and Haddon, M. 2016. Management strategy evaluation: best practices. Fish and Fisheries, 17(2): 303-334. 10.1111/faf.12104.

Pyper, B.J., and R.M. Peterman. 1998. Comparison of methods to account for autocorrelation in correlation analyses of fish data. Canadian Journal of Fisheries and Aquatic Sciences, 55: 2127-2140.

R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

Rice, J.C. 2000. Evaluating fishery impacts using metrics of community structure. ICES Journal of Marine Science, 57: 201-352.

Rice, J.C., and Rochet, M-J. 2005. A framework for selecting a suite of indicators for fisheries management. ICES Journal of Marine Science, 62: 516-527.

Ripa, J., and A.R. Ives. 2003. Food web dynamics in correlated and autocorrelated environments. Theoretical Population Biology, 64: 369-384.

Robins, C.M., Wang, Y.G., and Die, D. 1998. The impact of global positioning systems and plotters on fishing power in the northern prawn fishery, Australia. Canadian Journal of Fisheries and Aquatic Sciences, 55(7): 1645-1651. 10.1139/f98-037.

Rochet, M-J., and Rice, J. 2005. Do explicit criteria help in selecting indicators for ecosystem-based fisheries management? ICES Journal of Marine Science, 62: 528-539. 10.1016/j.icesjms.2005.01.007.

Rogers, J.B., Methot, R.D., Builder, T.L, Piner, K., and Wilkins, M. 2000. Status of the darkblotched rockfish (*Sebastes crameri*) resource in 2000. Pacific Fisheries Management Council, Portland.

Rooper, C.N., and Martin, M.H. 2011. Comparison of habitat-based indices of abundance with fishery-independent biomass estimates from bottom trawl surveys. Fisheries Bulletin, 110: 21-35.

Rose, K.A., Allen, J.I., Artioli, Y., Barange, M., Blackford, J., Carlotti, F., Cropp, R. *et al.* 2010. End-to-end models for the analysis of marine ecosystems: challenges, issues, and next steps. Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science, 2: 115-130. 10.1577/C09-059.1.

Rousseeuw, P.J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computation and Applied Mathematics, 20: 53-65.

Rubec, P.J., Kiltie, R., Leone, E., Flamm, R.O., McEachron, L, and Santi, C. 2016. Using delta-generalized additive models to predict spatial distributions and population abundance of juvenile pink shrimp in Tampa Bay, Florida. Marine and Coastal Fisheries, 8(1): 232-243. 10.1080/19425120.2015.1084408.

Ryding, K.E., and J.J. Millspaugh, and J.R. Skalski. 2007. Using time series to estimate rates of population change from abundance data. Journal of Wildlife Management, 71(1): 202-207.

Sadykova, D., Scott, B.E., De Dominicis, M., Wakelin, S.L., Sadykov, A., and Wolf, J. 2017. Bayesian joint models with INLA exploring mobile predator-prey and competitor species habitat overlap. Ecology and Evolution, 7(14): 5212-5226. 10.1002/ece3.3081.

Samhouri, J.F., Levin, P.S., and Harvey, C.J. 2009. Quantitative evaluation of marine ecosystem indicator performance using food web models. Ecosystems, 12: 1283-1298. 10.1007/s10021-009-9286-9.

Schielzeth, H. 2010. Simple means to improve the interpretability of regression coefficients. Methods in Ecology and Evolution, 1(2): 103-113.

Schindler, D.W., Mills, K.H., Malley, D.F., Findlay, D.L., Shearer, J.A., Davies, I.J., Turner, M.A., *et al*. 1985. Long-term ecosystem stress: the effects of years of experimental acidification on a small lake. Science, 228: 1395-1401.

Schmid, M., Wickler, F., Maloney, K.O., Mitchell, R., Fenske, N., and Mayr, A. 2013. Boosted Beta regression. PLoS ONE, 8: e61623.

See, K.E., and E.E. Holmes. 2015. Reducing bias and improving precision in species extinction forecasts. Ecological Applications, 25(4): 1157-1165.

Sepez, J., Norman, K., Poole, A., and Tilt, B. 2006. Fish scales: scale and method in social science research for North Pacific and West Coast fishing communities. Human Organization, 65: 280-293.

Shannon, L.J., Coll, M., and Neira, S. 2009. Exploring the dynamics of ecological indicators using food web models fitted to time series of abundance and catch data. Ecological Indicators, 9: 1078-1095.

Shannon, L.J., Coll, M., Yemane, D., Jouffre, D., Neira, S., Bertrand, A., Diaz, E. *et al.* 2010. Comparing data-based indicators across upwelling and comparable systems for communicating ecosystem states and trends. ICES Journal of Marine Science, 67: 807-832. 10.1039/icesjms/fsp270.

Shannon, L.J., Coll, M., Bundy, A., Gascuel, D., Heymans, J.J., Kleisner, K., Lynam, C.P. *et al*. 2014. Trophic level-based indicators to track fishing impacts across marine ecosystems. Marine Ecology Progress Series, 512: 115-140. 10.3346/meps10821.

Shelton, A.O., Thorson, J.T., Ward, E.J., and Feist, B.E. 2014. Spatial semiparametric models improve estimates of species abundance and distribution. Canadian Journal of Fisheries and Aquatic Sciences, 71: 1-12. 10.1139/cjfas-2013-0508.

Shin, Y-J., Bundy, A., Shannon, L.J., Simier, M., Coll, M., Fulton, E.A., Link, J.S., *et al*. 2010a. Can simple be useful and reliable? Using ecological indicators to represent and compare the state of marine ecosystems. ICES Journal of Marine Science, 67: 717-731.

Shin, Y-J., Shannon, L.J., Bundy, A., Coll, M., Aydin, K., Bez, N., Blanchard, J.L. *et al.* 2010b. Using indicators for evaluating, comparing, and communicating the ecological status of exploited marine ecosystems. 2. Setting the scene. ICES Journal of Marine Science, 67: 692-716. 10.1093/icesjms/fsp294.

Shin, Y-J., and Shannon, L.J. 2010. Using indicators for evaluating, comparing, and communicating the ecological status of exploited marine ecosystems. 1. The *IndiSeas* project. ICES Journal of Marine Science, 67: 686-691. 10.1093/icesjms/fsp273.

Shin, Y-J., Houle, J.E., Akoglu, E., Blanchard, J.L., Bundy, A., Coll, M., Demarcq, H., *et al*. 2018. The specificity of marine ecological indicators to fishing in the face of environmental change: a multi-model evaluation. Ecological Indicators, 89:317-326. 10.1016/j.ecolind.2018.01.010.

Shumway, R.H., and D.S. Stoffer. 2011. Time Series Analysis and Its Application: With R Examples. Springer, New York.

Skaug, H.J., and Fournier, D.A. 2006. Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. Computational Statistics & Data Analysis, 51(2): 699-709. 10.1016/j.csda.2006.03.005.

Smith, A.D.M., Fulton, E.A., Hobday, A.J., Smith, D.C., and Shoulder, P. 2007. Scientific tools to support the practical implementation of ecosystem-based fisheries management. ICES Journal of Marine Science, 64: 633-639.

Smith, M.D., Fulton, E.A., Day, R.W., Shannon, L.J., and Shin, Y-J. 2015. Ecosystem modelling in the southern Benguela: comparisons of Atlantis, Ecopath with Ecosim, and OSMOSE under fishing scenarios. African Journal of Marine Science, 37: 65-78. 10.2989/1814232X.2015.1013501.

Smithson, M., and Verkuilen, J. 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychological Methods, 11: 54.

Spencer, P.D. 2008. Density-independent and density-dependent factors affecting temporal changes in spatial distributions of eastern Bering Sea flatfish. Fisheries Oceanography, 17(5): 396-410. 10.1111/j.1365-2419.2008.00486.x.

St. Martin, K., and Hall-Arber, M. 2008. The missing layer: geo-technologies, communities, and implications for marine spatial planning. Marine Policy, 32: 779-786.

Steelman, T.A., and Wallace, R. 2001. Property rights and property wrongs: why context matters in fisheries management. Policy Science, 34: 357-379.

Steffanson, G. 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES Journal of Marine Science, 53(3): 577-588. 10.1006.jmsc.1996.0079.

Stoffer, D.S., and Wall, K.D. 1991. Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter. Journal of the American Statistical Association, 86(416): 1024-1033. 10.1080/01621459.1991.10475148.

Sugihara, G. 1995. From out of the blue. Nature, 378: 559-560.

Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., and Munch, S. 2012. Detecting causality in complex ecosystems. Science, 338(6106): 496-500. 10.1126/science.1227079.

Syrjala, S.E. 2000. Critique of the use of the delta distribution for the analysis of trawl survey data. ICES Journal of Marine Science, 57: 831-842. 10.1006/jmsc.2000.0571.

Szuwalski, C.S., K.A. Vert-Pre, A.E. Punt, T.A. Branch, and R. Hilborn. 2015. Examining common assumptions about recruitment: a meta-analysis of recruitment dynamics for worldwide marine fisheries. Fish and Fisheries, 16(4): 633-648.

Taylor, I.G., and Wetzel, C. 2011. Status of the US yelloweye rockfish resource in 2011 (Update of 2009 assessment model). Pacific Fishery Management Council, Portland.

The Nature Conservancy (TNC). 2014. California Risk Pool: A Co-management Model to advance fisheries Resource Stewardship. The Nature Conservancy.

Thompson, K.R., and F.H. Page. 1989. Detecting synchrony of recruitment using short, autocorrelated time series. Canadian Journal of Fisheries and Aquatic Sciences, 46: 1831-1838.

Thorson, J.T., 2017. Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative. Canadian Journal of Fisheries and Aquatic Sciences. 10.1139/cjfas-2017-0266.

Thorson, J.T., and Ward, E.J. 2013. Accounting for space-time interactions in index standardization models. Fisheries Research, 147: 426-433. 10.1016/j.fishres.2013.03.012.

Thorson, J.T., Scheuerell, M.D., Shelton, A.O., See, K.E., Skaug, H.J., and Kristensen, K. 2015a. Spatial factor analysis: a new tool for estimating joint species distributions and correlations in species range. Methods in Ecology and Evolution, 6(6): 627-637. 10.1111/2041-210X.12359.

Thorson, J.T., Shelton, A.O., Ward, E.J., and Skaug, H.J. 2015b. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES Journal of Marine Science, 72(5): 1297-1310. 10.1093/icesjms/fsu243.

Thorson, J.T., Skaug, H.J., Kristensen, K., Shelton, A.O., Ward, E.J., Harms, J.H., and Benante, J.A. 2015c. The importance of spatial models for estimating the strength of density dependence. Ecology, 96(5): 1202-1212. 10.1890/14-0739.1.

Thorson, J.T., and Kristensen, K. 2016. Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. Fisheries Research, 175: 66-74. 10.1016/j.fishres.2015.11.016.

Thorson, J.T., and Barnett, L.A.K. 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES Journal of Marine Science, 74(5): 1311-1321. 10.1093/icesjms/fsw193.

Thorson, J.T., and Haltuch, M. in press. Spatio-temporal analysis of compositional data: increased precision and improved workflow using model-based inputs to stock assessment. Canadian Journal of Fisheries and Aquatic Sciences. 10.1139/cjfas-2018-0015.

Trenkel, V.M., and Rochet, M-J. 2003. Performance of indicators derived from abundance estimates for detecting the impact of fishing on a fish community. Canadian Journal of Fisheries and Aquatic Sciences, 60: 67-85. 10.1139/f02-163.

United Nations (UN). 1992. United Nations Conference on environment and development. United Nations General Assembly Official Records, 46th Session, Agenda Item 21, UN Doc A/Conf.151/26.

U.S. Office of the Federal Register (USOFR). 2014. Trawl fishery economic data collection program. Code of Federal Regulations, Title 50, Part 660.114. U.S. Government Printing Office, Washington, D.C.

van Putten, I.E., Kulmala, S., Thébaud, O., Dowling, N., Hamon, K.G., Hutton, T., and Pascoe, S. 2012. Theories and behavioral drivers underlying fleet dynamics models. Fish and Fisheries, 13: 216-235.

van Putten, I.E., Gorton, R.J., Fulton, E.A., and Thebaud, O. 2013. The role of behavioral flexibility in a whole of ecosystem model. ICES Journal of Marine Science, 70: 150-163.

Vinagre, C., Salgado, J.P., Mendonca, V., Cabral, H., and Costa, M.J. 2012. Isotopes reveal fluctuation in trophic levels of estuarine organisms, in space and time. Journal of Sea Research, 72: 49-54. 10.1016/j.seares.2012.05.010.

Wallace, J.W., and Gertseva, V. 2018. Status of the darkblotched rockfish resource off the continental U.S. Pacific Coast in 2017 (Update of 2015 assessment model). Pacific Fishery Management Council, Portland.

Walters, C.J., Christensen, V., Martell, S.J., and Kitchell, J.F. 2005. Possible ecosystem impacts of applying MSY policies from single-species assessment. ICES Journal of Marine Science, 62: 558-568. 10.1016/j.icesjms.2004.12.005.

Ward, E.J., H. Chirakkal, M. Gonzálex-Suárez, D. Aurioles-Gamboa, E.E. Holmes, and L. Gerber. 2010. Inferring spatial structure from time-series data: using multivariate state-space models to detect metapopulation structure of California sea lions in the Gulf of California, Mexico. Journal of Applied Ecology, 47: 47-56. 10.1111/j.1365-2664.2009.01745.x.

Ward, E.J., E.E. Holmes, J.T. Thorson, and B. Collen. 2017. Complexity is costly: a meta-analysis of parametric and non-parametric methods for short-term population forecasting. Oikos, 123: 652-661.

Weijerman, M., Fulton, E.A., and Brainard, R.E. 2016a. Management strategy evaluation applied to coral reef ecosystems in support of ecosystem-based management. PLOS ONE, 11: e0152577.

Weijerman, M., Link, J.S., Fulton, E.A., Olsen, E., Townsend, H., Gaichas, S., Hansen, C. *et al*. 2016b. Atlantis Ecosystem Model Summit: report from a workshop. Ecological Modelling, 335: 35-38. 10.1016/j.ecolmodel.2016.05.007.

White, J.W., A. Rassweiler, J.F. Samhouri, A.C. Stier, and C. White. 2014. Ecologists should not use statistical significance test to interpret simulation model results. Oikos, 123: 385-388.

World Commission on Environment and Development (WCED). 1987. Our Common Future. Oxford University Press, Oxford.

Wilberg, M.J., Thorson, J.T., Linton, B.C., and Berkson, J. 2009. Incorporating time-varying catchability into population dynamic stock assessment models. Reviews in Fisheries Science, 18(1): 7-24. 10.1080/10641260903294647.

Woodland, R.J., and Secor, D.H. 2013. Benthic-pelagic coupling in a temperate inner continental shelf fish assemblage. Limnology Oceanography, 58(3): 966-976. 10.4319/lo.2013.58.3.0966.

Worm, B., Hilborn, R. Baum, J.K., Branch, T.A., Collie, J.S., Costello, C., Fogarty, M.J., *et al*. 2009. Rebuilding global fisheries. Science, 31(325): 578-585. 10.1126/science.1173146.

Yule, G.U. 1926. Why do we sometimes get nonsense-correlations between time-series?—A study in sampling and the nature of time series. Journal of the Royal Statistical Society, 89(1): 1-63.

Zador, S.G., Holsman, K.K., Aydin, K.Y., and Gaichas, S.K. 2017. Ecosystem considerations in Alaska: the value of qualitative assessments. ICES Journal of Marine Science 74(1): 421-430. 10.1093/icesjms/fsw144.

Appendix A. SUPPLEMENTARY INFORMATION FOR CHAPTER 1

A.1 ESTIMATION METHODS

Estimates of cross correlation were included in the manuscript despite the fact that most simulated time series were autocorrelated thus violating the assumption of independence needed to assess the significance of the measurement. Consequently, estimates demonstrated the repercussions of not having i.i.d. data and the potential benefits of more appropriate methods. It was hypothesized that cross correlation would estimate co-interactions rather than direct-interactions.

Pearson's correlation coefficients provided a measure ranging between ±1 of the intensity of the linear relationship between the time series for each investigated lag (Zar, 1999). Potential errors introduced by intra-multiplicity (i.e., investigation of numerous k values, such that the probability of at least one null hypothesis being rejected is increased beyond the desired Type-I error rate) were minimized by restricting the investigation of k to ±2 (Olden and Neff, 2001). Pearson's correlation coefficients were reported instead of Spearman's non-parametric rank order correlation because Pearson is the default method in the R statistical environment (R Core Team 2018) and it tends to be the default method in ecological manuscripts. Nevertheless, differences between the two methods proved to be trivial. The maximum absolute difference in Pearson versus Spearman coefficients across all explored factors was 0.28 and the median was 0.0006.

Prewhitening included fitting autoregressive integrated moving average (ARIMA) models to the leading time series to estimate autocorrelation, differencing, and moving average parameters that best fit the data. ARIMA models were fit using the forecast::auto.arima function (Hyndman and Khandakar 2008) in the R statistical environment. The model that provided the best fit of those investigated was chosen using Akaike information criterion corrected for small sample sizes (Burnham and Anderson 2002). Parameter estimates from the best fit model were then applied to the second time series and cross correlation coefficients, cross correlation$^k = \frac{\sum(z_{1,t}-\overline{z_1})(z_{2,t-k}-\overline{z_2})}{\sqrt{\sum(z_{1,t}-\overline{z_1})^2}\sqrt{\sum(z_{2,t}-\overline{z_2})^2}}$, were computed on the residuals, $z_{1,t}$ and $z_{2,t}$, of the respective time series at lags, k, where $\overline{z_1}$ and $\overline{z_2}$ are the means of the time series of residuals.

To fit a MARSS model one must specify the process and observation models. Specifically, the call to MARSS::MARSS has the following default arguments model = list(Z = "identity", B = "identity", U = "unconstrained", A = "scaling", Q = "diagonal and unequal", Z = "diagonal and unequal", V0 = "zero", x0 = "unconstrained"). "Unconstrained" estimates all parameters, "diagonal and equal" estimates one parameter for the diagonal, "diagonal and unequal" estimates each diagonal element, "identity" is a matrix with ones on the diagonal and zeros elsewhere, "zero" fixes all parameters at zero, and "scaling" estimates an additional parameter for all observed time series beyond the primary time series per unobserved state. Additionally, users can specify individual parameters of the parameter matrices such as would be the case if one only wanted to estimate one-way direct-interactions (i.e., matrix(list("1:1", "2:1", 0, "2:2"), 2, 2)). The flexibility inherent in the MARSS package is facilitated through the use of an Expectation-Maximization (EM) algorithm, where the ability to fix at least some parameters can increase identifiability, especially with relatively short time series (Zuur *et al.*, 2003).

Parameters of the investigated MARSS models were specified in the following manner (Table A.1). Observation error was either assumed to be zero or estimated as "diagonal and equal". The interaction matrix, **B**, was estimated as an unconstrained matrix, although exploration of fixing the upper off-diagonal element at zero was explored via sensitivity testing. Thus, b_{21} is the effect of the leading time series on the lagging time series and b_{12} is the effect of the lagging time series on the leading time series. Process errors, **Q**, were estimated using an unconstrained matrix. The initial state vector included two parameters drawn from a multivariate normal distribution, $\mathbf{x}_0 \sim MVN(0, 1)$, and results were similar when the parameters were fixed at zero, but they were estimated here for completeness. Finally, multiple observations of the unobserved states, when provided to the estimation method, were assumed to be on the same scale, and thus, **Z** became a 4x2 matrix and **A** was fixed at zero.

A.2   REFERENCES

Burnham, K.P., and Anderson, D.R. 2002. Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach. Springer-Verlag, New York, New York. 488 pp.
Holmes, E.E., Ward, E.J. and Wills, K. 2012. MARSS: Multivariate autoregressive state-space models for analyzing time-series data. The R Journal, 4(1): 11-19.

Hyndman, R.J., and Khandakar, Y. 2008. Automatic time series forecasting: the forecast package for R. Journal of Statistical Software, 26(3): 1-22.

Olden, J.D., and Neff, B.D. 2001. Cross-correlation bias in lag analysis of aquatic time series. Marine Biology, 138: 1063-1070. 10.1007/s002270000517.

R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

Zar, J.H. 1999. Biostatistical Analysis. Prentice-Hall, Inc. Upper Saddle River, New Jersey.

Zuur, A.F., Fryer, R.J., Jollifee, I.T., Dekker, R., and Beukema, J.J. 2003. Estimating common trends in multivariate time series using dynamic factor analysis. Environmetrics, 14(7): 665-685. 10.1002/env.611.

A.3 TABLES

Table A.1. Estimation methods included cross correlation (ρ^k) at lags (k) ±2, prewhitened cross correlation at lags ±2 (ρ_p^k), and multivariate autoregressive state-space (MARSS) models. Cross correlation involved prewhitening by estimating first- through fifth-order autoregressive (AR), moving average (MA), and differencing (D) parameters that best fit the leading time series and applying those to the lagging time series. Four MARSS models were included that differed in how observation errors and the model matrix ($\mathbf{Z}$) were parameterized. Blank entries were specified the same as the model in the previous row and NA indicates that the parameters did not apply. Quoted words such as "unconstrained" are how MARSS models are implemented in the MARSS package (Holmes *et al.*, 2012).

model	co-interaction	self-correlation	moving average	differencing	direct-interaction	observation error	$\mathbf{Z}$
Cross correlation	$\rho^{-2}; \rho^{-1}; \rho^0;$ $\rho^1; \rho^2$	NA	NA	NA	NA	NA	NA
Prewhitened cross correlation	$\rho_p^{-2}; \rho_p^{-1}; \rho_p^0;$ $\rho_p^1; \rho_p^2$	$AR(1); AR(2)$ $AR(3); AR(4)$ $AR(5)$	$MA(1); MA(2)$ $MA(3); MA(4)$ $MA(5)$	$D(1); D(2)$ $D(3); D(4)$ $D(5)$	NA	NA	NA
MARSS models that use observed time series as observations of unobserved states, and the $\mathbf{Z}$ matrix maps observations to these states.	$\begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}$ $(q_{21} = q_{12})$	$b_{11}; b_{22}$	NA	NA	$b_{21}; b_{12}$ ("unconstrained")	0 ("zero")	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
						0	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$
						$\begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}$ ("equal")	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
						$\begin{bmatrix} r & 0 & 0 & 0 \\ 0 & r & 0 & 0 \\ 0 & 0 & r & 0 \\ 0 & 0 & 0 & r \end{bmatrix}$ ("equal")	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$

Table A.2. Prewhitening included fitting an autoregressive integrated moving average model (ARIMA) to the leading time series. Autoregressive (p), moving average (q), and differencing (d) orders were limited to five for each parameter and parameters were chosen using Akaike information criteria corrected for small sample sizes. Proportions of models that utilized a given order are reported for levels self-interactions, observation-error variance (z), length of the time series (n), and number of observations per state (ns). Input values to the simulation for co-interactions and direct-interactions did not lead to changes in the proportions.

self	z	n	ns	p.1	p.2	p.3	p.4	q.1	q.2	q.3	q.4	d.1	d.2	d.3
-0.9	0	25	1	0.04	0.75	0.21	0.00	0.93	0.07	0.00	0.00	1.00	0.00	0.00
-0.9	0	100	1	0.00	0.87	0.13	0.00	0.89	0.06	0.04	0.01	1.00	0.00	0.00
0	0	25	1	0.93	0.04	0.03	0.00	0.91	0.09	0.00	0.00	0.96	0.04	0.00
0	0	100	1	0.91	0.05	0.04	0.00	0.92	0.05	0.03	0.00	0.97	0.03	0.00
0.9	0	25	1	0.65	0.29	0.05	0.01	0.85	0.14	0.01	0.00	0.46	0.53	0.01
0.9	0	100	1	0.55	0.32	0.12	0.01	0.88	0.09	0.01	0.02	0.39	0.61	0.00
-0.9	1	25	1	0.22	0.66	0.11	0.01	0.86	0.14	0.00	0.00	0.94	0.06	0.00
-0.9	1	25	2	0.11	0.65	0.21	0.02	0.71	0.22	0.06	0.01	1.00	0.00	0.00
-0.9	1	100	1	0.00	0.72	0.25	0.03	0.42	0.46	0.08	0.04	1.00	0.00	0.00
-0.9	1	100	2	0.00	0.57	0.29	0.11	0.27	0.43	0.23	0.04	0.99	0.01	0.00
0	1	25	1	0.93	0.03	0.04	0.00	0.88	0.09	0.03	0.00	0.95	0.05	0.00
0	1	25	2	0.93	0.05	0.01	0.01	0.87	0.10	0.03	0.00	1.00	0.00	0.00
0	1	100	1	0.89	0.08	0.02	0.01	0.86	0.11	0.02	0.01	0.98	0.02	0.00
0	1	100	2	0.67	0.14	0.16	0.01	0.71	0.12	0.14	0.03	0.99	0.01	0.00
0.9	1	25	1	0.63	0.36	0.01	0.00	0.84	0.16	0.00	0.00	0.52	0.47	0.01
0.9	1	25	2	0.22	0.61	0.16	0.01	0.80	0.12	0.08	0.00	0.94	0.06	0.00
0.9	1	100	1	0.35	0.48	0.16	0.01	0.30	0.56	0.13	0.01	0.46	0.54	0.00
0.9	1	100	2	0.13	0.54	0.25	0.06	0.23	0.45	0.26	0.06	0.79	0.21	0.00

Table A.3. Convergence rates of multivariate autoregressive state-space (MARSS) models across investigated levels of direct-interactions and co-interactions, self-interactions, observation error variance (z), length of time series (n), and number of observations per state (ns). The proportion of models that converged out of the 100 fit per scenario are reported for models that were fit to the "raw" and standardized ("std"; mean of zero and standard deviation of one) simulated data. Convergence was defined as the ability to successfully produce and invert the hessian matrix such that estimate standard errors were estimated for each parameter. Multivariate autoregressive (MAR) models did not estimate observation error and MARSS models assumed observation was uncorrelated with equal variance.

type	direct	co	self	z	n	ns	raw	std
MAR	-0.90	-0.90	-0.90	0	25	1	0.33	0.79
MAR	-0.90	-0.90	-0.90	0	50	1	0.29	0.6
MAR	-0.90	-0.90	-0.90	0	100	1	0.2	0.33
MAR	-0.90	-0.90	0.00	0	25	1	0.32	0.4
MAR	-0.90	-0.90	0.00	0	50	1	0.26	0.42
MAR	-0.90	-0.90	0.00	0	100	1	0.18	0.38
MAR	-0.90	-0.90	0.90	0	25	1	0.44	0.86
MAR	-0.90	-0.90	0.90	0	50	1	0.33	0.64
MAR	-0.90	-0.90	0.90	0	100	1	0.22	0.29
MAR	-0.90	0.00	-0.90	0	25	1	0.97	0.99
MAR	-0.90	0.00	-0.90	0	50	1	0.92	0.94
MAR	-0.90	0.00	-0.90	0	100	1	0.84	0.95
MAR	-0.90	0.00	0.00	0	25	1	0.79	0.81
MAR	-0.90	0.00	0.00	0	50	1	0.87	0.88
MAR	-0.90	0.00	0.00	0	100	1	0.74	0.82
MAR	-0.90	0.00	0.90	0	25	1	0.96	0.99
MAR	-0.90	0.00	0.90	0	50	1	0.94	1
MAR	-0.90	0.00	0.90	0	100	1	0.86	0.98
MAR	-0.90	0.90	-0.90	0	25	1	0.31	0.76
MAR	-0.90	0.90	-0.90	0	50	1	0.24	0.51
MAR	-0.90	0.90	-0.90	0	100	1	0.14	0.26
MAR	-0.90	0.90	0.00	0	25	1	0.34	0.43
MAR	-0.90	0.90	0.00	0	50	1	0.25	0.39
MAR	-0.90	0.90	0.00	0	100	1	0.12	0.35
MAR	-0.90	0.90	0.90	0	25	1	0.38	0.84
MAR	-0.90	0.90	0.90	0	50	1	0.25	0.68
MAR	-0.90	0.90	0.90	0	100	1	0.18	0.35
MAR	0.00	-0.90	-0.90	0	25	1	0.32	0.33

MAR	0.00	-0.90	-0.90	0	50	1	0.31	0.28
MAR	0.00	-0.90	-0.90	0	100	1	0.24	0.23
MAR	0.00	-0.90	0.00	0	25	1	0.28	0.32
MAR	0.00	-0.90	0.00	0	50	1	0.21	0.25
MAR	0.00	-0.90	0.00	0	100	1	0.12	0.12
MAR	0.00	-0.90	0.90	0	25	1	0.4	0.45
MAR	0.00	-0.90	0.90	0	50	1	0.27	0.31
MAR	0.00	-0.90	0.90	0	100	1	0.18	0.24
MAR	0.00	0.00	-0.90	0	25	1	0.95	0.99
MAR	0.00	0.00	-0.90	0	50	1	0.91	0.9
MAR	0.00	0.00	-0.90	0	100	1	0.77	0.9
MAR	0.00	0.00	0.00	0	25	1	0.77	0.86
MAR	0.00	0.00	0.00	0	50	1	0.77	0.89
MAR	0.00	0.00	0.00	0	100	1	0.66	0.78
MAR	0.00	0.00	0.90	0	25	1	0.97	0.92
MAR	0.00	0.00	0.90	0	50	1	0.91	0.97
MAR	0.00	0.00	0.90	0	100	1	0.86	0.92
MAR	0.00	0.90	-0.90	0	25	1	0.32	0.36
MAR	0.00	0.90	-0.90	0	50	1	0.25	0.3
MAR	0.00	0.90	-0.90	0	100	1	0.16	0.23
MAR	0.00	0.90	0.00	0	25	1	0.31	0.32
MAR	0.00	0.90	0.00	0	50	1	0.2	0.22
MAR	0.00	0.90	0.00	0	100	1	0.1	0.16
MAR	0.00	0.90	0.90	0	25	1	0.43	0.33
MAR	0.00	0.90	0.90	0	50	1	0.29	0.25
MAR	0.00	0.90	0.90	0	100	1	0.21	0.2
MAR	0.90	-0.90	-0.90	0	25	1	0.3	0.8
MAR	0.90	-0.90	-0.90	0	50	1	0.28	0.57
MAR	0.90	-0.90	-0.90	0	100	1	0.2	0.28
MAR	0.90	-0.90	0.00	0	25	1	0.3	0.47
MAR	0.90	-0.90	0.00	0	50	1	0.21	0.42
MAR	0.90	-0.90	0.00	0	100	1	0.19	0.44
MAR	0.90	-0.90	0.90	0	25	1	0.4	0.85
MAR	0.90	-0.90	0.90	0	50	1	0.33	0.68
MAR	0.90	-0.90	0.90	0	100	1	0.24	0.37
MAR	0.90	0.00	-0.90	0	25	1	0.94	0.96
MAR	0.90	0.00	-0.90	0	50	1	0.93	0.97
MAR	0.90	0.00	-0.90	0	100	1	0.85	0.94
MAR	0.90	0.00	0.00	0	25	1	0.86	0.88
MAR	0.90	0.00	0.00	0	50	1	0.81	0.83
MAR	0.90	0.00	0.00	0	100	1	0.77	0.8

MAR	0.90	0.00	0.90	0	25	1	0.96	0.97
MAR	0.90	0.00	0.90	0	50	1	0.92	0.98
MAR	0.90	0.00	0.90	0	100	1	0.88	0.97
MAR	0.90	0.90	-0.90	0	25	1	0.3	0.81
MAR	0.90	0.90	-0.90	0	50	1	0.25	0.59
MAR	0.90	0.90	-0.90	0	100	1	0.14	0.32
MAR	0.90	0.90	0.00	0	25	1	0.31	0.41
MAR	0.90	0.90	0.00	0	50	1	0.25	0.39
MAR	0.90	0.90	0.00	0	100	1	0.14	0.38
MAR	0.90	0.90	0.90	0	25	1	0.41	0.81
MAR	0.90	0.90	0.90	0	50	1	0.28	0.64
MAR	0.90	0.90	0.90	0	100	1	0.17	0.25
MARSS	0.00	-0.90	-0.90	1	25	1	0.01	0.02
MARSS	0.00	-0.90	-0.90	1	25	2	0.03	0.14
MARSS	0.00	-0.90	-0.90	1	100	1	0	0.1
MARSS	0.00	-0.90	-0.90	1	100	2	0.05	0.13
MARSS	0.00	-0.90	0.00	1	25	1	0.05	0.09
MARSS	0.00	-0.90	0.00	1	25	2	0.11	0.11
MARSS	0.00	-0.90	0.00	1	100	1	0.01	0.04
MARSS	0.00	-0.90	0.00	1	100	2	0.03	0.12
MARSS	0.00	-0.90	0.90	1	25	1	0.06	0.11
MARSS	0.00	-0.90	0.90	1	25	2	0.13	0.21
MARSS	0.00	-0.90	0.90	1	100	1	0.24	0.42
MARSS	0.00	-0.90	0.90	1	100	2	0.46	0.5
MARSS	0.00	0.00	-0.90	1	25	1	0	0.07
MARSS	0.00	0.00	-0.90	1	25	2	0.06	0.26
MARSS	0.00	0.00	-0.90	1	100	1	0	0.31
MARSS	0.00	0.00	-0.90	1	100	2	0.16	0.66
MARSS	0.00	0.00	0.00	1	25	1	0.03	0.08
MARSS	0.00	0.00	0.00	1	25	2	0.2	0.36
MARSS	0.00	0.00	0.00	1	100	1	0.02	0.04
MARSS	0.00	0.00	0.00	1	100	2	0.3	0.4
MARSS	0.00	0.00	0.90	1	25	1	0.14	0.38
MARSS	0.00	0.00	0.90	1	25	2	0.38	0.68
MARSS	0.00	0.00	0.90	1	100	1	0.84	0.9
MARSS	0.00	0.00	0.90	1	100	2	0.79	0.82
MARSS	0.00	0.90	-0.90	1	25	1	0	0.04
MARSS	0.00	0.90	-0.90	1	25	2	0.04	0.06
MARSS	0.00	0.90	-0.90	1	100	1	0	0.07
MARSS	0.00	0.90	-0.90	1	100	2	0.07	0.12
MARSS	0.00	0.90	0.00	1	25	1	0.03	0.03

MARSS	0.00	0.90	0.00	1	25	2	0.06	0.1
MARSS	0.00	0.90	0.00	1	100	1	0.01	0.03
MARSS	0.00	0.90	0.00	1	100	2	0.09	0.08
MARSS	0.00	0.90	0.90	1	25	1	0.08	0.16
MARSS	0.00	0.90	0.90	1	25	2	0.13	0.21
MARSS	0.00	0.90	0.90	1	100	1	0.26	0.38
MARSS	0.00	0.90	0.90	1	100	2	0.38	0.49
MARSS	0.90	-0.90	-0.90	1	25	1	0.04	0.15
MARSS	0.90	-0.90	-0.90	1	25	2	0.02	0.31
MARSS	0.90	-0.90	-0.90	1	100	1	0.05	0.59
MARSS	0.90	-0.90	-0.90	1	100	2	0	0.45
MARSS	0.90	-0.90	0.00	1	25	1	0.01	0.11
MARSS	0.90	-0.90	0.00	1	25	2	0.15	0.2
MARSS	0.90	-0.90	0.00	1	100	1	0.03	0.07
MARSS	0.90	-0.90	0.00	1	100	2	0.08	0.11
MARSS	0.90	-0.90	0.90	1	25	1	0.28	0.7
MARSS	0.90	-0.90	0.90	1	25	2	0.33	0.8
MARSS	0.90	-0.90	0.90	1	100	1	0.32	0.98
MARSS	0.90	-0.90	0.90	1	100	2	0.15	0.86
MARSS	0.90	0.00	-0.90	1	25	1	0.01	0.21
MARSS	0.90	0.00	-0.90	1	25	2	0.04	0.43
MARSS	0.90	0.00	-0.90	1	100	1	0.02	0.61
MARSS	0.90	0.00	-0.90	1	100	2	0.09	0.75
MARSS	0.90	0.00	0.00	1	25	1	0.02	0.08
MARSS	0.90	0.00	0.00	1	25	2	0.3	0.49
MARSS	0.90	0.00	0.00	1	100	1	0.05	0.15
MARSS	0.90	0.00	0.00	1	100	2	0.34	0.53
MARSS	0.90	0.00	0.90	1	25	1	0.36	0.6
MARSS	0.90	0.00	0.90	1	25	2	0.56	0.76
MARSS	0.90	0.00	0.90	1	100	1	0.78	0.95
MARSS	0.90	0.00	0.90	1	100	2	0.78	0.87
MARSS	0.90	0.90	-0.90	1	25	1	0.01	0.14
MARSS	0.90	0.90	-0.90	1	25	2	0.04	0.4
MARSS	0.90	0.90	-0.90	1	100	1	0	0.57
MARSS	0.90	0.90	-0.90	1	100	2	0.03	0.62
MARSS	0.90	0.90	0.00	1	25	1	0	0.12
MARSS	0.90	0.90	0.00	1	25	2	0.13	0.24
MARSS	0.90	0.90	0.00	1	100	1	0.04	0.09
MARSS	0.90	0.90	0.00	1	100	2	0.14	0.16
MARSS	0.90	0.90	0.90	1	25	1	0.18	0.5
MARSS	0.90	0.90	0.90	1	25	2	0.32	0.71

| MARSS | 0.90 | 0.90 | 0.90 | 1 | 100 | 1 | 0.27 | 0.99 |
| MARSS | 0.90 | 0.90 | 0.90 | 1 | 100 | 2 | 0.23 | 0.51 |

A.4   FIGURES
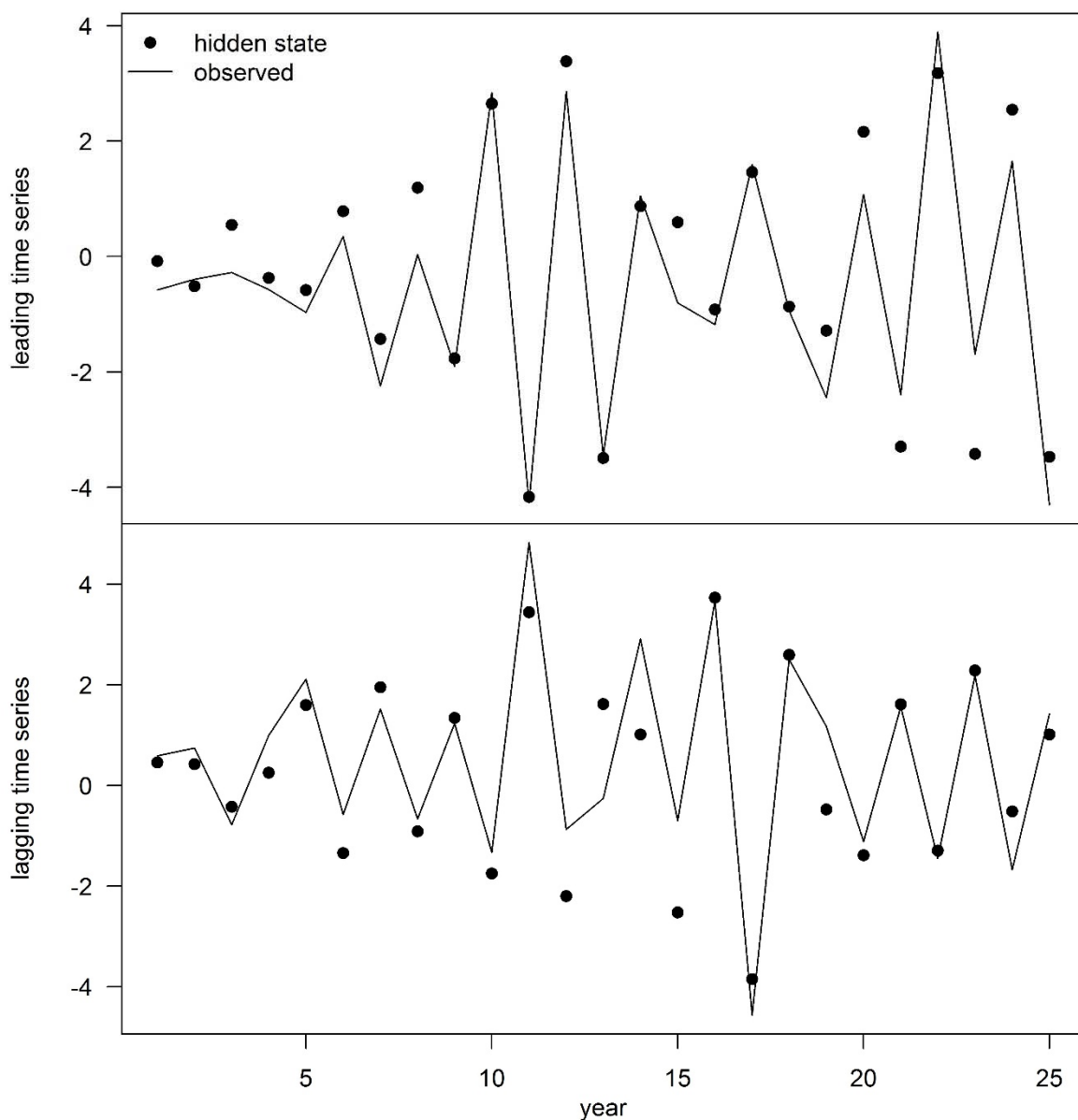
Figure A.1. Examples of a leading and lagging time series (top and bottom panels, respectively) generated from the simulation process when there was no direct-interaction and the co-interaction was -0.9. Each time series exhibited an self-interaction of -0.9. The points are the unobserved states and the lines are the observed time series with observation error, where the variance of the observation error is one.
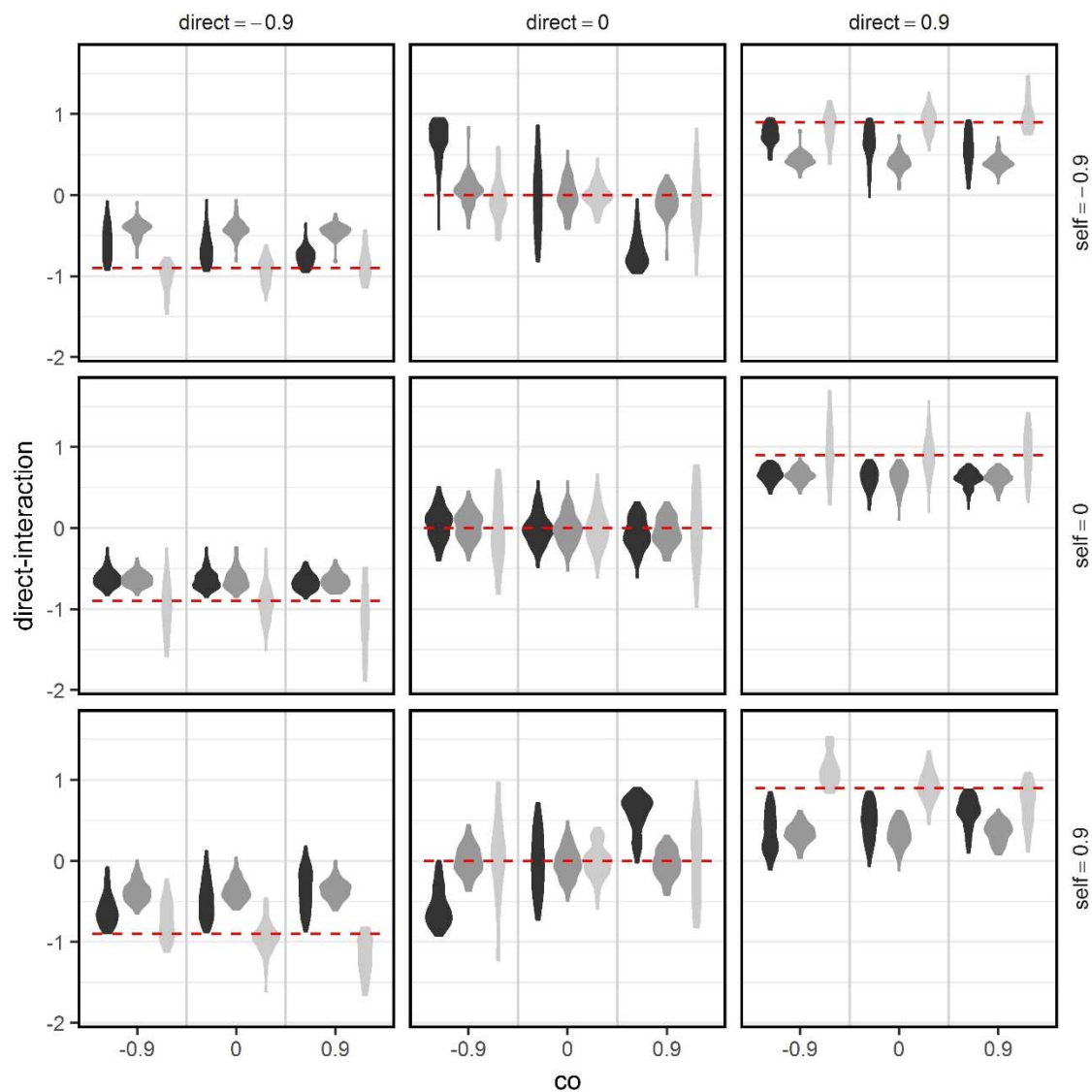
Figure A.2. Violin plots of estimated direct-interactions from cross correlation (dark gray), prewhitened cross correlation (gray), and multivariate autoregressive state-space (MARSS) models that did not estimate observation error (referred to as MAR models; light gray) across three levels of self-interactions (rows), direct-interactions (columns), and co-interactions (x-axis). Fitted time series included 25 data points observed without error. Horizontal dashed, red lines indicate the true value used to simulate the data. The prewhitening model was the model that best fit the leading time series as chosen by Akaike information criteria corrected for small sample size. Results from MAR models are limited to those that successfully estimated standard errors for all included parameters.
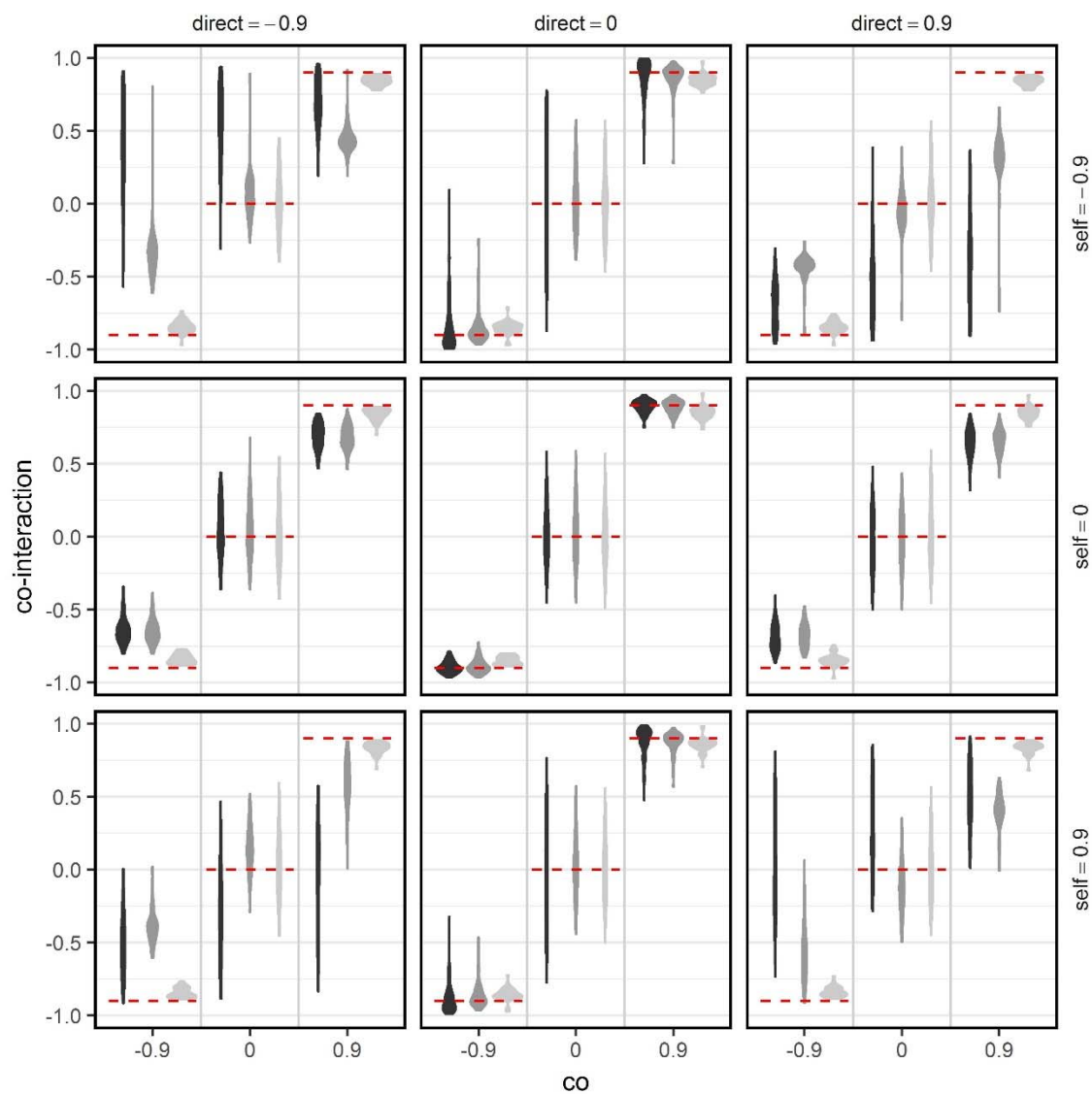
Figure A.3. Same as Figure A.2, except y-axis displays estimates of co-interactions.
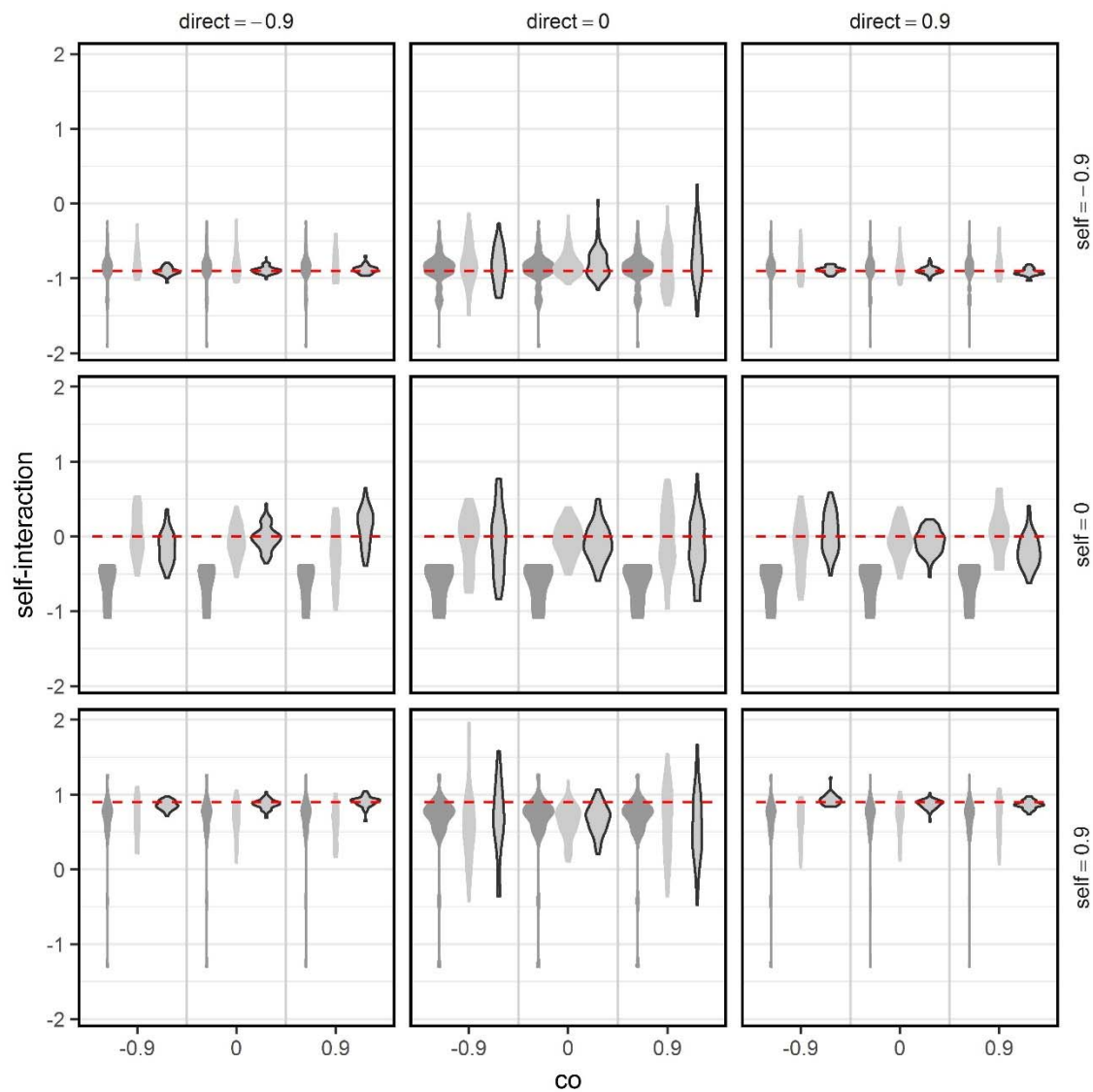
Figure A.4. Same as Figure A.2, except y-axis displays estimates of AR(1) from the prewhitening procedure (gray) and from MAR models for the leading (light gray) and lagging (light gray with black border) time series.
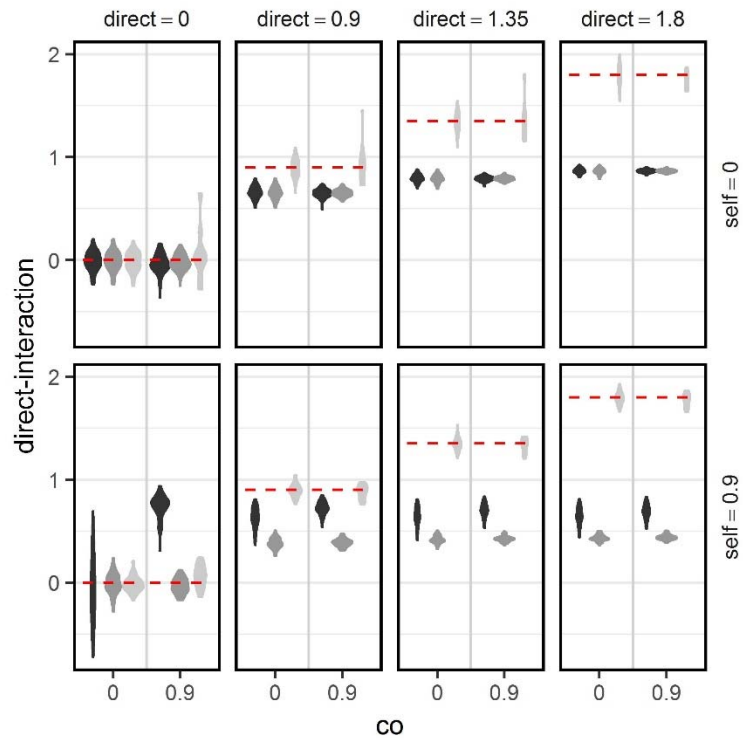
Figure A.5. Same as Figure A.2, except for models fit to 100 data points rather than 25 and across a wider range of input values for the direct-interaction (columns).
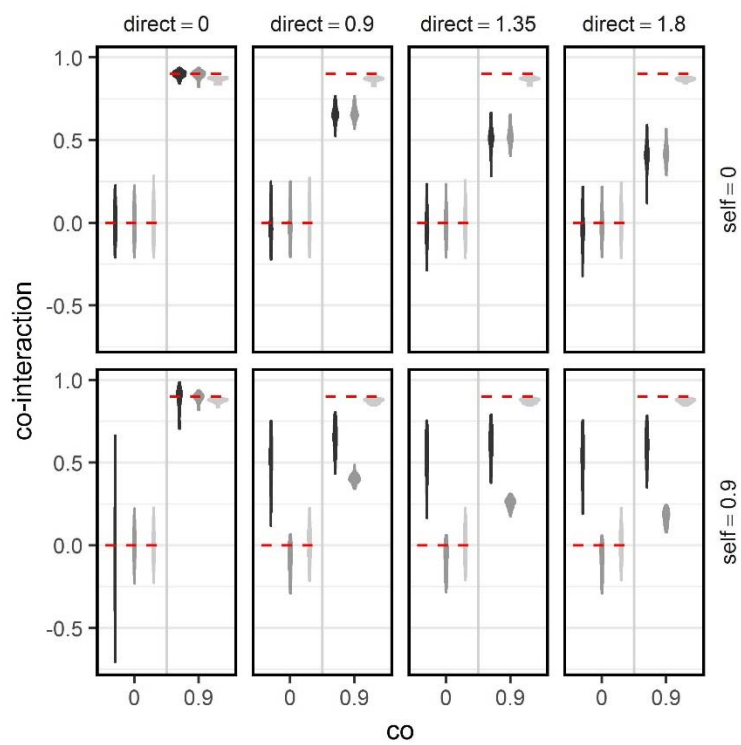
Figure A.6. Same as Figure A.5, except the y-axis for estimates of co-interactions.
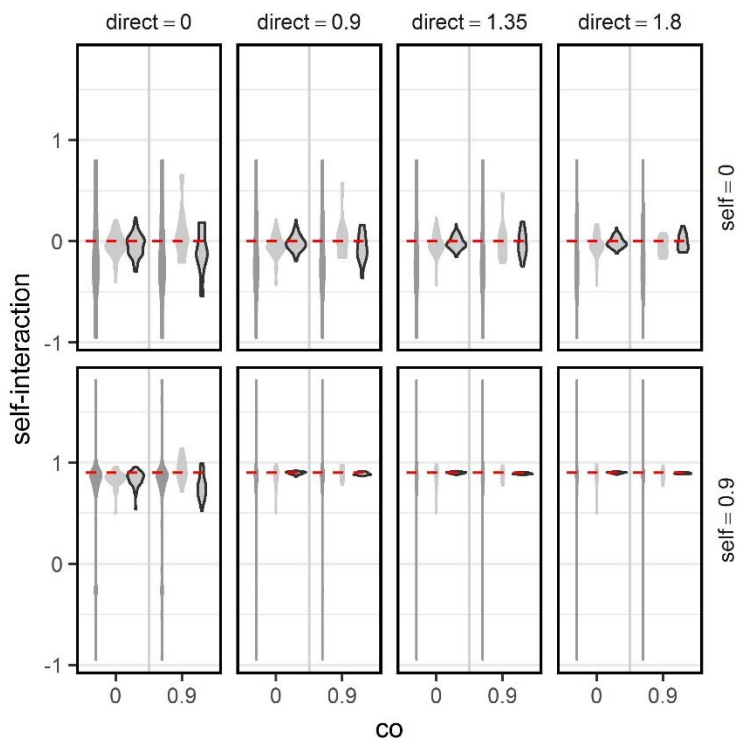
Figure A.7. Same as Figure A.5, except the y-axis displays estimates of AR(1) from the prewhitening procedure (gray) and from MAR models for the leading (light gray) and lagging (light gray with black border) time series.
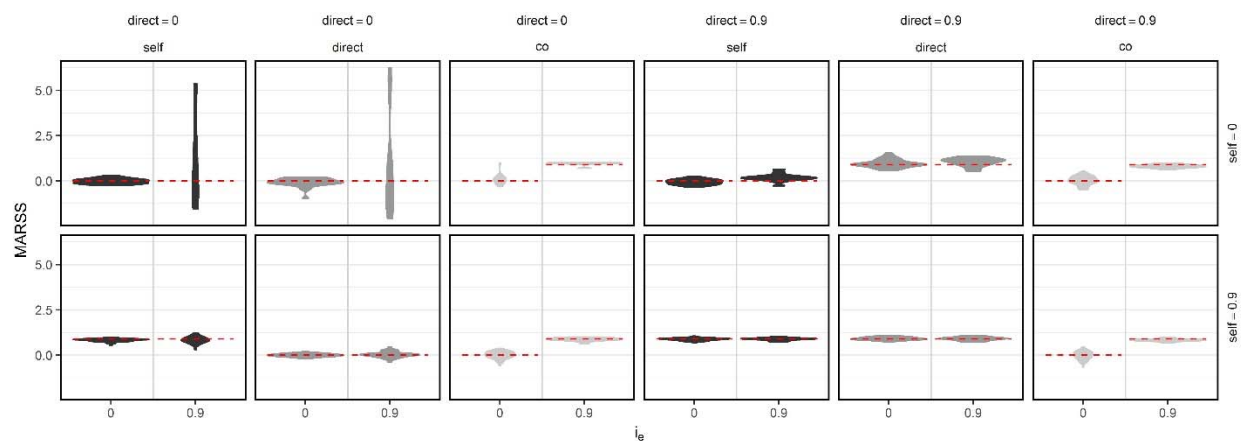
Figure A.8. Violin plots of estimated self-interactions of the leading time series (dark gray), direct-interactions (gray), and co-interactions (light gray) from multivariate autoregressive state-space (MARSS) models fit two 100-year time series per state. Data were simulated using two levels of self-interactions (rows), direct-interactions (columns), and co-interactions (x-axis). Horizontal dashed, red lines indicate the true values for each parameter of interest (panels). Time series were simulted with an observation-error variance of 1.0. Results are limited to those that successfully estimated standard errors for all included.
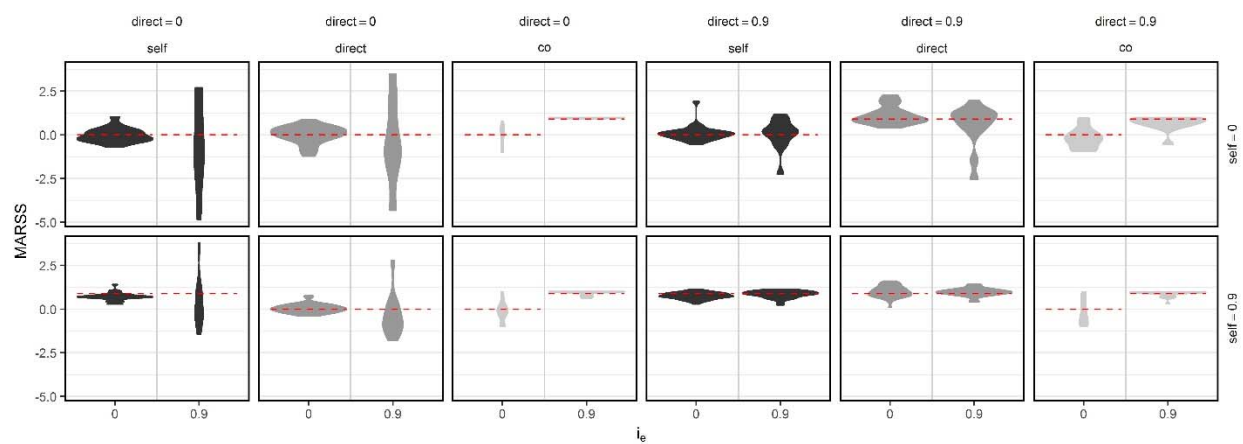
Figure A.9. Same as Figure A.8, except for time series with 25 data points.
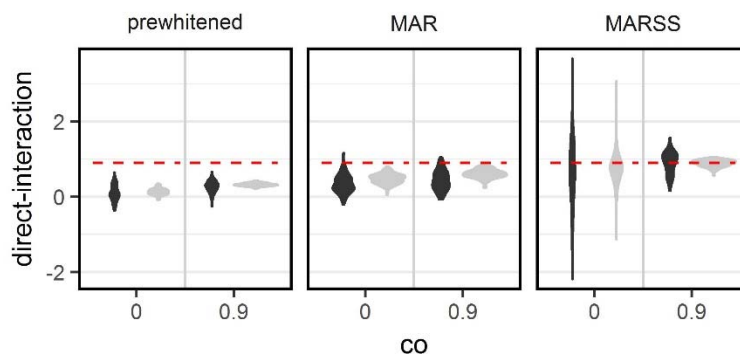
Figure A.10. Violin plots of estimated direct-interactions from prewhitened cross correlation, multivariate autoregressive (MAR) models, and multivariate autoregressive state-space (MARSS) models fit to 25 and 100 years of data (black and gray, respectively) across two levels co-interactions (x-axis). Time series were simulted with a direct-interaction of 0.9 (horizontal dashed, red line), self-interactions of 0.9, and observation-error variance of 1.0. Only MARSS models estimated observation error. The prewhitening model was the model that best fit the leading time series as chosen by Akaike information criterion corrected for small sample size. Results from MAR models are limited to those that successfully estimated standard errors for all included parameters.
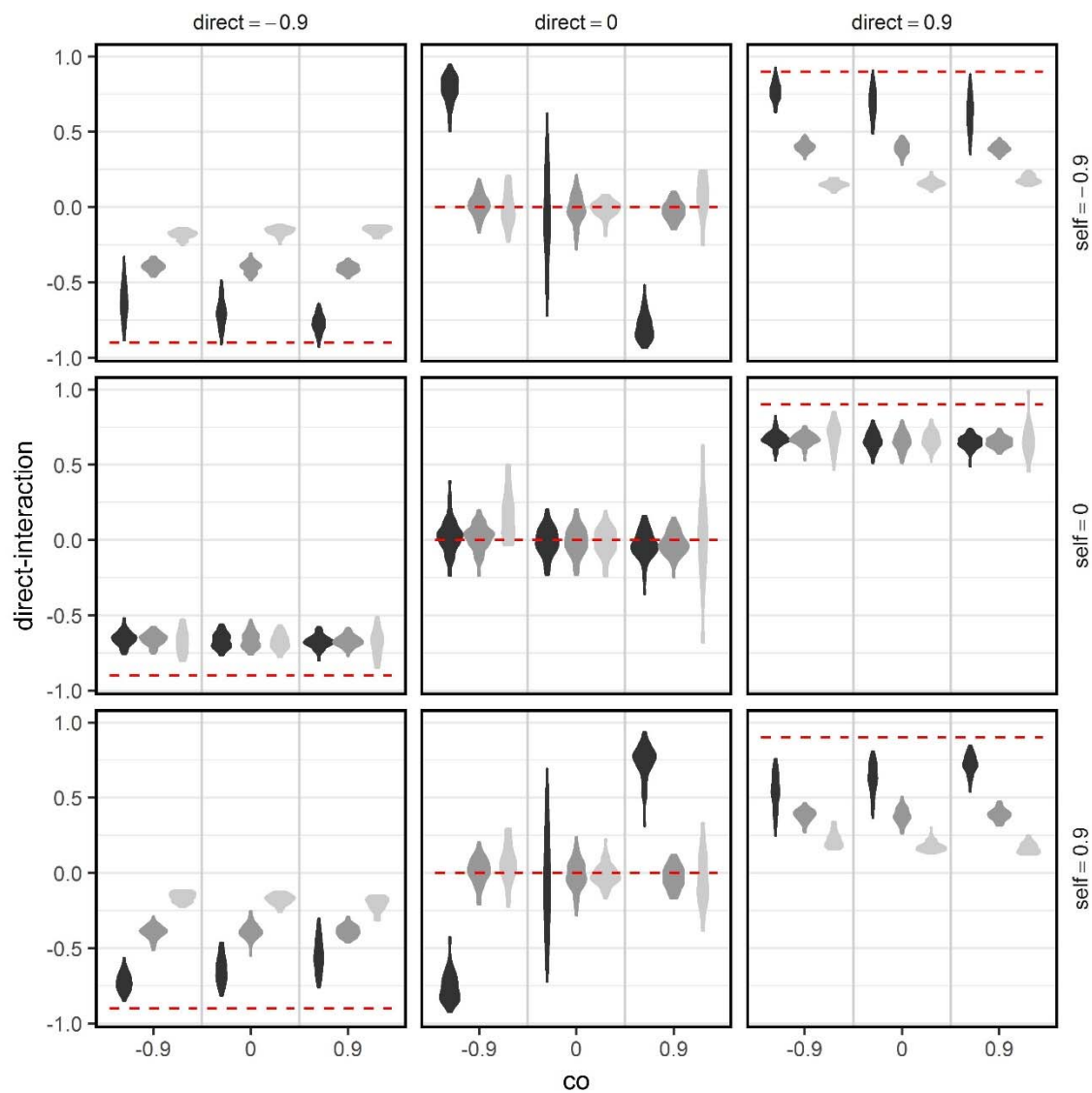
Figure A.11. Same as Figure A.2, except for standardized time series of 100 data points. Standardization led to the data having a mean of zero and a standard deviation of one.

Appendix B. SUPPLEMENTARY INFORMATION FOR CHAPTER 2

B.1 METHODS

The twelve included indicators are part of a suite of indicators selected by the Indicators for the Seas (*IndiSeas*) Working Group (Table A1), an international collaborative developed to evaluate the status of exploited marine ecosystems (Coll *et al.*, 2016). Indicators were selected from a list proposed to facilitate ecosystem-based management (EBM) by *IndiSeas* because of their expected ability to assess the sustainability of harvests and incorporate broader ecosystem considerations. As listed in the main text, the twelve indicators are biomass of the surveyed community (TB), inverse of the coefficient of variation of the surveyed-community biomass over the last 10 years (BS), inverse fishing pressure (invF), biomass-weighted mean ("mean") intrinsic vulnerability index of fish in the catch (IVI), mean fish length in the surveyed community (LG), mean maximum fish life span in the surveyed community (LS), mean marine trophic index of the catch (MTI), proportion of non-declining exploited species that have an assessment of their status (NDES), proportion of predatory fish in the surveyed community (PF), proportion of non-fully-exploited stocks (i.e., current biomass greater than 60% of unfished biomass) of those that are or could be assessed (SS), mean trophic level (TL) of the catch (TLc), and mean TL of the surveyed community (TLs). Some indicators were represented as inverses of more well-known indicators, such that all indicators decreased with increasing fishing pressure

Indicators were calculated from catch and survey data. In the context of this simulation study, survey data corresponded to model output (i.e., biomass time series) of groups defined by model developers as those that have an appreciable amount of their biomass surveyed in actuality during routine surveys, such that the resulting empirical data could be used to generate a reliable index of abundance for an assessment of their status. Survey-based indicators were based only on output from these groups. *IndiSeas* calculated their survey-based indicators using data collected from fishery-independent sources on adult demersal and pelagic bony and cartilaginous fish and commercially important invertebrates (Coll *et al.*, 2016). Simulated catches consisted of all groups marked as "caught" by model developers, where this could include both targeted and bycatch groups, but always targeted groups. This definition largely matches *IndiSeas'* definition of catch

data that included landed species and discards if discards were from a target species for which certain size classes were not profitable to retrain. Simulated model output necessary to calculate two additional indicators proposed by *IndiSeas*, mean TL of the modelled community estimated from Ecosim models fitted to time-series data and proportion of discards in the fishery, were not available within the context of this simulation.

Life-history information for modelled groups that contained multiple species were based on the most dominant species within the group or a biomass-weighted mean. Information sources varied amongst the ecosystems, but largely relied on information contained in Fishbase (fishbase.org; sealifebase.org). Maximum, theoretical length (L_{inf}; cm) was derived from von Bertalanffy (von Bertalanffy, 1938) relationships or using Fishbase, if it was not available from local sources. Maximum age (A_{max}; year) and TL were also determined from Fishbase, if they were not available from local sources. Both L_{inf} and A_{max} were fixed parameters in their respective indicator calculations for each modelled group (Table B.1). The intrinsic vulnerability index (IVI; Cheung *et al.*, 2005), which can range from 0 to 100, is based on life-history traits and ecological characteristics from Fishbase.

Predatory fish were defined as those with a trophic level greater than 3.5 rather than using direct information on diets. *IndiSeas* defined predatory fishes as those that were largely piscivorous or fed on invertebrates larger than 2 cm in length.

Mean length of fish in the surveyed community had to be adapted to work with simulated output. Atlantis outputs included weights of modelled groups at specific times and not of individual fish. Thus, weights of modelled groups were used to calculate a biomass-weighted maximum length, where maximum length of a group was a fixed parameter defined by model developers.

Management targets were standardized based on Food and Agriculture Organization of the United Nations (FAO) definition of exploitation levels. Non-fully exploited is defined by FAO as a stock with a current biomass estimate greater than 0.6 of unfished biomass (FAO, 2011). Only stocks that were or could be assessed to provide an estimate of unfished biomass were included in the sustainable stock indicator. Therefore, both stocks that are assessed using fishery-dependent or -independent methods and stocks that are surveyed but do not currently have an assessment were included in the metric. *IndiSeas* based their indicator on the same metric of sustainability as defined by FAO but used estimates of current depletion available in the latest FAO report (2011;

http://www.fao.org/docrep/015/i2389e/i2389e.pdf, part D), augmented with expert opinion and information gleaned from white papers, if estimates were not available.

The proportion of non-declining species (NDES), originally developed to circumvent the statistical fallacies of repeatedly testing regression and correlation coefficients in community analyses (Lynam *et al.*, 2010), was calculated using Kendall's tau over ten-year time periods to provide a time-series measure. Thus, the metric provides a moving window of trends in biomass for each ecosystem. Kendall's tau was calculated from time series of biomass outputs and their corresponding years for groups that were exploited. The final metric, NDES, was then computed by adding up the number of groups with a tau greater than or equal to zero divided by the number of estimated tau parameters within an ecosystem for each ten-year time period.

Analyses were performed on results from eight simulated ecosystems. Two versions of SE Australia (AustSE) and NE US (Neus) were included where the first version was parameterized in terms of constant fishing effort similar to recent years and the second version was parameterized in terms of dynamic effort (i.e., fishing rates varied with time) to match the time-series data on which the ecosystem model was based. All other ecosystems, California Current (CalCu), Chesapeake Bay (CAM), Northern Gulf of California (GOC), Gulf of Mexico (GoMex), Guam, and Nordic and Barents Sea (NOBA), had one version and were parameterized in terms of constant fishing rates.

B.2   REFERENCES

Cheung, W.W.L, Pitcher, T.J., and Pauly, D. 2005. A fuzzy logic expert system to estimate intrinsic extinction vulnerabilities of marine fishes to fishing. Biological Conservation, 124(1): 97-111.

Coll, M., Shannon, M.J., Kleisner, K.M., Juan-Jordá, M.J., Bundy, A., Akoglu, A.G., Banaru, D. *et al.* 2016. Ecological indicators to capture the effects of fishing on biodiversity and conservation status of marine ecosystems. Ecological Indicators, 60: 947-962. 10.1016/j.ecolind.2015.08.048.

FAO. 2011. Review of the state of world marine fishery resources. FAO Fisheries and Aquaculture Technical Paper, 569. Rome. pp. 334.

Lynam, C.P., Cusack, C., and Stokes, D. 2010. A methodology for community-level hypothesis testing applied to detect trends in phytoplankton and fish communities in Irish waters. Estuarine, Coastal and Shelf Science, 87(3): 451-462. 10.1016/j.ecss.2010.01.019.

von Bertalanffy, L. 1938. A quantitative theory of organic growth (inquiries on growth laws. II). Human Biology, 10: 181-213.

B.3   TABLES

Table B.1. Indicators from those proposed by Indicator for the Seas (*IndiSeas*) Working Group used to track the direct and broader impacts of fishing on exploited marine ecosystems, along with their definitions (Coll *et al.*, 2016). Some indicator calculations were modified to work with the Atlantis output as noted in "attribute definition". A blank entry indicates no difference between the *IndiSeas* definition and the definition used here. Corresponding attributes are calculated from the total relevant (i.e., TL > 1) biomass rather than the respective survey or catch biomasses. Abbreviations are as follows: coefficient of variation (CV), biomass (B), equilibrium biomass (B_0), fished group (f), intrinsic vulnerability (IV), maximum age (A_{max}), maximum length (L_{inf}), modelled group (g), predatory fish groups with a trophic level > 3.5, standard deviation (sd), and trophic level (TL).

Label	Calculation	*IndiSeas* definition	Definition used here	Attribute definition
Biomass, B	$\sum_g B_g$	Total survey biomass		Total ecosystem biomass groups with TL > 1
Biomass stability	$\dfrac{B[last\ 10\ years]}{sd(B[last\ 10\ years])}$	1 / CV (10-year survey biomass)		1/CV (10-year ecosystem biomass)
Inverse fishing	$\dfrac{B}{catch}$	Survey biomass / catch		Ecosystem biomass / cat...
Mean vulnerability	$\dfrac{\sum_g(IV_g * B_g)}{\sum_g B_g}$	Mean intrinsic vulnerability index of retained fish	Biomass-weighted mean intrinsic vulnerability index of caught fish	Biomass-weighted mean intrinsic vulnerability ind... all fish
Fish size	$\dfrac{\sum_g\left(L_{inf_g} * B_g\right)}{\sum_g B_g}$	Mean fish length in the surveyed community	Biomass-weighted mean maximum length of surveyed fish	Biomass-weighted mean maximum length of all f...
Life span	$\dfrac{\sum_g\left(A_{max_g} * B_g\right)}{\sum_g B_g}$	Mean max life span of surveyed fish		Biomass-weighted mean life span of all fish
Trophic index	$\dfrac{\sum_g(Trophic\ index_g * B_g)}{\sum_g B_g}$	Catch-based marine trophic index	Catch-based marine trophic index	Ecosystem-based marine trophic index
Non-declining species	$\dfrac{\sum_f ifelse(cor(1:10, B_f) \geq 0, 1, 0)}{\sum_f 1}$	Proportion of non-declining exploited species	Proportion of non-declining exploited groups	Proportion of non-declin... species
Predators	$\dfrac{\sum_{pf}(B_{pf})}{\sum_g B_g}$	Proportion of predatory fish in the survey	Proportion of fish with TL > 3.5 in the survey	Proportion of fish with a... 3.5 in the ecosystem
Sustainable stocks	$\dfrac{\sum_g ifelse(B_g > 0.6B_0, 1, 0)}{\sum_g ifelse(g\ is\ or\ could\ be\ assessed, 1, 0)}$	Proportion of non-fully exploited stocks	Proportion of exploited stocks that are or could be assessed with B>0.6B_0	Proportion of exploited s... with B>0.6B_0
TL of catch	$\dfrac{\sum_g(TL_g * Catch_g)}{\sum_g Catch_g}$	Mean TL of catch	Biomass-weighted mean TL of caught groups	Biomass-weighted mean... the ecosystem
TL of survey	$\dfrac{\sum_g(TL_g * B_g)}{\sum_g B_g}$	Mean TL of surveyed community		Biomass-weighted mean... the ecosystem

Table B.2. Definitions of the attributes selected for their importance to ecosystem structure and function. Attributes are calculated from the total relevant biomass (i.e., trophic level (TL) > 1), except for net primary productivity (NPP), which includes all autotrophic groups with TL $\geq$ 1. Note that total biomass (B) and TL of the ecosystem directly correspond to indicators and are attributes included in Table B.1.

Label	Calculation	Definition
Net primary productivity (NPP)	$\sum_{autorophs}(B_{autotrophs})$	Biomass produced by autotrophs
Total biomass	$\sum_{g\ with\ TL>1}(B_{g\ with\ TL>1})$	Total biomass of all modelled groups (g) with a TL > 1
TL of the ecosystem	$\dfrac{\sum_{g\ with\ TL>1}(TL_g * B_g)}{\sum_{g\ with\ TL>1}(B_g)}$	Biomass-weighted mean TL of all modelled groups with a TL > 1
Total biomass / NPP	$\dfrac{\sum_{g\ with\ TL>1}(B_{g\ with\ TL>1})}{\sum_{autorophs}(B_{autotrophs})}$	Biomass of the ecosystem (i.e., TL > 1) divided by NPP
Target biomass	$\sum_{f}(B_f)$	Biomass of groups targeted by fisheries (f)

Table B.3. The number of sensitive indicator-attribute relationships , where the confidence interval that did not cross zero for the direct- and the co-interactions. The twelve indicators are as follows: inverse of the coefficient of variation of the surveyed-community biomass over the last 10 years (BS), inverse fishing pressure (invF), biomass-weighted mean ("mean") intrinsic vulnerability index of fish in the catch (IVI), mean fish length in the surveyed community (LG), mean maximum fish life span in the surveyed community (LS), mean marine trophic index of the catch (MTI), proportion of non-declining exploited species that have an assessment of their status (NDES), proportion of predatory fish in the surveyed community (PF), proportion of non-fully-exploited stocks (i.e., current biomass greater than 60% of unfished biomass) of those that are or could be assessed (SS), biomass of the surveyed community (TB), mean trophic level (TL) of the catch (TLc), and mean TL of the surveyed community (TLs). Additional attributes not included in self-tests include primary productivity (NPP), TB/NPP, and target biomass (Tar). The number of possible scenarios differed for each scenario type; where, the base case, fishing, marine protected area, and ocean acidification drivers of change had 10, 112 (16x7), 27 (3x9), and 20 (2x10) potential estimates per relationship, respectively. All ten ecosystems were included in the BC and OA scenarios, but the two ecosystems with dynamic fishing and the Nordic and Barents Sea were excluded from the fishing scenarios, and the NE US ecosystem was excluded from the marine protected area driver of change. Missing values indicate that the indicator-attribute combination was not investigated.

indicator	interaction		attribute													
	direct	co	TB	NPP	TB/NPP	TLeco	Tar	BS	invF	IVI	LG	LS	MTI	NDES	PF	SS
BS	TRUE	FALSE	44	33	26	4	22	18								
BS	TRUE	TRUE	4		8	4	4	13								
BS	FALSE	TRUE	5	6	3	3	7	9								
invF	TRUE	FALSE	4	13	23	5	1		7							
invF	TRUE	TRUE	27	28	5	31	27		59							
invF	FALSE	TRUE	32	35	1	5	15		35							
IVI	TRUE	FALSE	5	16	12	26	7			15						
IVI	TRUE	TRUE	11		7	15	23			25						
IVI	FALSE	TRUE	25	8	10	16	37			35						
LG	TRUE	FALSE	29	9	23	21	12				9					
LG	TRUE	TRUE	5	10	6	21	33				51					
LG	FALSE	TRUE	13	23	13	22	12				25					
LS	TRUE	FALSE	48	29	33	12	3					3				
LS	TRUE	TRUE	7	16	15	15	13					71				
LS	FALSE	TRUE	36	35	11	7	10					40				

MTI	TRUE	FALSE	24	37	24	47	1	48			
MTI	TRUE	TRUE	14	4	3	12	19	22			
MTI	FALSE	TRUE	39	22	1	5	4	9			
NDES	TRUE	FALSE	24	8	41	2	21		1		
NDES	TRUE	TRUE	1	3	3	6	17		108		
NDES	FALSE	TRUE	4	7		2	1		20		
PF	TRUE	FALSE	3	19	13	25	2			16	
PF	TRUE	TRUE	15	18		28	21			81	
PF	FALSE	TRUE	38	36	9	1	8			27	
SS	TRUE	FALSE	22	10	32	16	12				
SS	TRUE	TRUE	13	5	2	5	7				41
SS	FALSE	TRUE	2	7	5	2	2				11
TB	TRUE	FALSE	5	3	22						
TB	TRUE	TRUE	61	8	24	28	82				
TB	FALSE	TRUE	49	35	23	16	19				
TLc	TRUE	FALSE	25	28	23	48	3				
TLc	TRUE	TRUE	12	3	2	13	23				
TLc	FALSE	TRUE	41	20	3	9	5				
TLsc	TRUE	FALSE	15	15	25	10	6				
TLsc	TRUE	TRUE	20	11		20	18				
TLsc	FALSE	TRUE	49	36	14	19	14				

Table B.4. Proportion of positive and negative co-interactions (top and bottom panels, respectively) out of the total number of indicator-attribute relationships investigated to determine indicator representativeness for each driver of change (column; i.e., one indicator:attribute combination across all ecosystems). Drivers of change include the base case (BC); fishing (F) at three levels of the BC on all, demersal (dem), invertebrate (inv), large pelagic (lpel), and small pelagic (spel) groups; three levels of marine protected areas (MPAs); and two levels of ocean acidification (OA). See caption of Table B.3 for definitions of the indicator acronyms. Proportions greater than 0.5 are highlighted in gray. The sum for a given row/column combination across the panels may not always equal one because estimates were only considered positive or negative if their confidence interval did not cross zero.

sign	indicator	BC	F_all			F_dem			F_inv			F_lpel			F_spel			MPA			OA	
			0%	50%	200%	0%	50%	200%	0%	50%	200%	0%	50%	200%	0%	50%	200%	10%	25%	50%	0.05	0.01
positive	BS	0.20	0.14	0.29	0.00	0.14	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.43	0.14	0.00	0.22	0.22	0.33	0.40	0.30
	invF	0.80	0.00	0.57	0.71	0.71	0.71	0.71	0.71	0.43	0.71	0.71	0.71	0.71	0.71	0.71	0.57	0.78	0.78	0.89	0.80	0.80
	IVI	0.30	0.00	0.43	0.71	0.43	0.43	0.57	0.57	0.43	0.57	0.43	0.29	0.14	0.57	0.29	0.43	0.33	0.22	0.33	0.70	0.50
	LG	0.70	0.71	0.57	0.86	0.71	0.71	1.00	0.57	0.57	0.71	0.57	0.57	0.43	1.00	0.71	0.71	0.78	0.78	0.78	0.70	0.70
	LS	1.00	0.86	1.00	1.00	1.00	1.00	1.00	0.86	0.86	1.00	0.86	0.86	0.86	0.86	0.86	1.00	1.00	1.00	1.00	0.90	0.90
	MTI	0.30	0.00	0.43	0.29	0.43	0.43	0.43	0.14	0.43	0.14	0.43	0.29	0.43	0.29	0.29	0.29	0.22	0.11	0.11	0.20	0.20
	NDES	0.80	0.00	0.86	1.00	0.86	1.00	0.86	1.00	0.86	0.86	0.71	0.86	0.71	0.86	1.00	0.86	0.89	0.89	0.89	0.90	0.90
	PF	0.70	0.71	0.86	0.57	0.57	0.57	0.57	0.71	0.86	0.57	0.86	0.57	0.57	0.57	0.57	0.71	0.78	0.67	0.78	0.90	0.90
	SS	0.40	0.00	0.43	0.57	0.43	0.43	0.43	0.43	0.57	0.57	0.29	0.57	0.29	0.57	0.57	0.57	0.44	0.11	0.22	0.30	0.30
	TB	0.80	0.86	0.86	0.57	0.86	0.86	0.86	0.86	0.71	0.57	0.86	0.86	0.86	0.71	0.71	0.86	0.56	0.44	0.67	0.40	0.40
	TLc	0.20	0.00	0.29	0.14	0.29	0.29	0.29	0.00	0.29	0.00	0.29	0.29	0.29	0.29	0.29	0.29	0.11	0.11	0.22	0.30	0.30
	TLs	0.50	0.57	0.43	0.57	0.57	0.43	0.43	0.57	0.43	0.71	0.43	0.43	0.29	0.57	0.43	0.43	0.44	0.44	0.44	0.50	0.60
negative	BS	0.10	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.00	0.00	0.00	0.30	0.20
	invF	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00
	IVI	0.10	0.00	0.14	0.00	0.00	0.29	0.00	0.14	0.29	0.14	0.00	0.14	0.14	0.43	0.14	0.29	0.22	0.22	0.22	0.10	0.00
	LG	0.10	0.14	0.14	0.14	0.00	0.14	0.00	0.14	0.14	0.00	0.00	0.14	0.14	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.10
	LS	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.10

MTI	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.11	0.00	0.11	0.10	0.10
NDES	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PF	0.00	0.14	0.00	0.14	0.14	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.14	0.14	0.14	0.00	0.11	0.00	0.00	0.00	0.00
SS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.11	0.00	0.00	0.10
TLc	0.10	0.00	0.14	0.29	0.14	0.29	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.00	0.22	0.11	0.11	0.00	0.00
TLs	0.20	0.29	0.14	0.14	0.14	0.29	0.29	0.14	0.29	0.14	0.29	0.29	0.29	0.14	0.14	0.29	0.33	0.33	0.22	0.20	0.20

Table B.5. Proportion of positive and negative direct-interactions (top and bottom panels, respectively) out of the total number of indicator-attribute relationships investigated to determine indicator representativeness for each ecosystem (column; i.e., one indicator:attribute combination across all drivers of change). See caption of Table B.3 for definitions of the indicator acronyms. Region acronyms are as follows: SE Australia (AustSE), California Current (CalCu), Gulf of California (GOC), Gulf of Mexico (GOM), NE US (Neus), and Nordic and Barent's Sea (NOBA). Additional models were included for AustSE and Neus that used dynamic fishing effort (DE) instead of constant fishing effort. Proportions greater than 0.5 are highlighted in gray. The sum for a given row/column combination across the panels may not always equal one because estimates were only considered positive or negative if their confidence interval did not cross zero.

sign	indicator	region									
		AustSE	AustSEDE	CalCu	CAM	GOC	GoMex	Guam	Neus	NeusDE	NOBA
positive	BS	0.00	0.00	0.90	0.00	0.24	0.62	0.06	0.00	0.17	0.17
	invF	0.19	0.00	0.05	0.86	0.00	0.57	0.00	0.33	0.17	0.00
	IVI	0.86	0.00	0.00	0.05	0.52	0.14	0.67	0.11	0.67	0.00
	LG	0.38	0.33	0.00	0.52	0.29	0.57	0.06	0.00	0.00	0.00
	LS	0.00	0.67	0.43	0.67	0.29	0.52	0.00	0.00	0.00	0.00
	MTI	0.38	0.83	0.86	0.14	0.05	0.00	0.89	0.33	0.17	0.00
	NDES	0.29	0.33	0.05	0.33	0.05	0.00	0.17	0.00	0.17	0.00
	PF	0.24	0.67	0.14	0.10	0.76	0.67	0.00	0.00	0.00	0.33
	SS	0.00	0.00	0.00	0.10	0.00	0.29	0.17	0.00	0.00	0.00
	TB	0.90	0.67	0.19	0.10	0.86	0.19	0.22	0.00	0.00	0.00
	TLc	0.38	1.00	0.90	0.10	0.43	0.05	0.17	0.33	0.67	0.00
	TLs	0.90	0.67	0.00	0.00	0.14	0.05	0.06	0.00	0.00	0.17
negative	BS	0.57	0.83	0.00	0.14	0.10	0.14	0.00	0.39	0.17	0.00
	invF	0.10	0.17	0.10	0.00	0.38	0.05	0.22	0.11	0.00	0.50
	IVI	0.00	0.50	0.62	0.76	0.05	0.19	0.00	0.56	0.17	0.50
	LG	0.00	0.00	1.00	0.05	0.57	0.14	0.28	0.89	0.17	0.00
	LS	0.14	0.00	0.05	0.00	0.62	0.00	0.11	0.94	0.00	0.83
	MTI	0.00	0.00	0.00	0.00	0.14	0.76	0.00	0.00	0.00	0.00
	NDES	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.39	0.00	0.00
	PF	0.00	0.00	0.00	0.00	0.00	0.10	0.06	0.33	0.00	0.00
	SS	0.00	0.00	0.05	0.00	0.48	0.00	0.11	0.00	0.00	0.00
	TB	0.10	0.17	0.00	0.00	0.05	0.14	0.00	0.11	0.50	0.33
	TLc	0.00	0.00	0.00	0.43	0.00	0.76	0.72	0.06	0.33	0.00
	TLs	0.10	0.33	0.71	0.57	0.00	0.81	0.28	0.83	0.33	0.67

Table B.6. Same as Table B.4, except for direct-interactions.

sign	indicator	BC	F_all 0%	F_all 50%	F_all 200%	F_dem 0%	F_dem 50%	F_dem 200%	F_inv 0%	F_inv 50%	F_inv 200%	F_lpel 0%	F_lpel 50%	F_lpel 200%	F_spel 0%	F_spel 50%	F_spel 200%	MPA 10%	MPA 25%	MPA 50%	OA 0.05	OA 0.01
positive	BS	0.20	0.14	0.14	0.43	0.29	0.29	0.43	0.43	0.29	0.29	0.29	0.29	0.29	0.14	0.14	0.29	0.22	0.22	0.11	0.20	0.20
	invF	0.20	0.00	0.29	0.29	0.14	0.14	0.57	0.14	0.29	0.43	0.43	0.43	0.43	0.43	0.43	0.14	0.11	0.22	0.22	0.20	0.10
	IVI	0.30	0.00	0.57	0.43	0.00	0.29	0.29	0.43	0.43	0.43	0.57	0.14	0.14	0.14	0.43	0.29	0.44	0.33	0.33	0.40	0.20
	LG	0.30	0.29	0.29	0.14	0.14	0.43	0.29	0.43	0.29	0.14	0.14	0.14	0.29	0.29	0.14	0.29	0.22	0.11	0.22	0.40	0.20
	LS	0.20	0.29	0.29	0.14	0.29	0.14	0.14	0.57	0.43	0.29	0.14	0.29	0.43	0.29	0.29	0.29	0.33	0.44	0.22	0.10	0.20
	MTI	0.50	0.00	0.29	0.43	0.43	0.43	0.43	0.29	0.57	0.43	0.43	0.29	0.29	0.29	0.29	0.43	0.56	0.33	0.33	0.30	0.20
	NDES	0.10	0.00	0.14	0.00	0.00	0.00	0.00	0.29	0.00	0.29	0.29	0.00	0.00	0.14	0.00	0.29	0.00	0.22	0.22	0.30	0.30
	PF	0.20	0.29	0.29	0.14	0.29	0.43	0.43	0.14	0.29	0.29	0.43	0.29	0.29	0.14	0.29	0.29	0.22	0.33	0.33	0.50	0.10
	SS	0.00	0.00	0.00	0.14	0.14	0.00	0.14	0.00	0.14	0.29	0.00	0.14	0.00	0.29	0.00	0.14	0.11	0.00	0.00	0.00	0.00
	TB	0.30	0.29	0.57	0.43	0.29	0.57	0.29	0.29	0.29	0.43	0.29	0.29	0.29	0.43	0.29	0.43	0.22	0.22	0.33	0.40	0.30
	TLc	0.50	0.00	0.29	0.14	0.71	0.29	0.57	0.14	0.14	0.29	0.43	0.29	0.43	0.14	0.43	0.57	0.44	0.44	0.56	0.30	0.30
	TLs	0.20	0.29	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.29	0.14	0.14	0.14	0.29	0.14	0.14	0.22	0.22	0.33	0.20	0.00
negative	BS	0.30	0.29	0.43	0.43	0.14	0.29	0.14	0.14	0.14	0.29	0.14	0.00	0.14	0.57	0.14	0.00	0.22	0.22	0.22	0.00	0.10
	invF	0.20	0.00	0.00	0.14	0.29	0.14	0.29	0.00	0.00	0.00	0.14	0.14	0.00	0.00	0.00	0.14	0.11	0.11	0.22	0.40	0.40
	IVI	0.30	0.00	0.29	0.29	0.29	0.43	0.29	0.43	0.29	0.43	0.14	0.29	0.29	0.29	0.29	0.43	0.44	0.22	0.56	0.20	0.40
	LG	0.30	0.14	0.29	0.71	0.43	0.43	0.57	0.29	0.29	0.43	0.43	0.43	0.43	0.43	0.57	0.43	0.22	0.22	0.22	0.30	0.30
	LS	0.30	0.14	0.29	0.43	0.14	0.29	0.43	0.14	0.14	0.29	0.14	0.29	0.43	0.29	0.29	0.29	0.22	0.11	0.22	0.20	0.30
	MTI	0.10	0.00	0.14	0.14	0.14	0.14	0.14	0.29	0.14	0.14	0.14	0.14	0.00	0.14	0.14	0.14	0.00	0.11	0.00	0.20	0.10
	NDES	0.10	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.14	0.00	0.00	0.14	0.00	0.00	0.14	0.14	0.11	0.00	0.00	0.00	0.10
	PF	0.00	0.14	0.00	0.14	0.00	0.00	0.00	0.14	0.14	0.00	0.14	0.00	0.14	0.14	0.00	0.14	0.00	0.00	0.00	0.10	0.00
	SS	0.10	0.00	0.14	0.14	0.14	0.29	0.00	0.00	0.14	0.14	0.14	0.14	0.14	0.00	0.00	0.14	0.00	0.00	0.00	0.10	0.00
	TB	0.00	0.14	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.11	0.11	0.50	0.40
	TLc	0.30	0.00	0.43	0.43	0.14	0.43	0.43	0.43	0.29	0.29	0.29	0.14	0.14	0.29	0.43	0.29	0.11	0.22	0.11	0.10	0.20
	TLs	0.60	0.43	0.43	0.43	0.57	0.57	0.43	0.43	0.43	0.29	0.57	0.43	0.57	0.29	0.57	0.43	0.33	0.44	0.33	0.40	0.60

B.4 Figures

Figure B.1. Median, sensitive direct-interactions across drivers of change versus the proportion of indicator-attribute relationships that were sensitive for a given combination out of all investigated combinations within an ecosystem (colors; refer to text for abbreviations). Dashed line at zero distinguishes between positive and negative interactions. Attributes are abbreviated as follows: net primary productivity (NPP), total biomass (TB), mean TL of the ecosystem (TLeco), TB divided by NPP (TB/NPP), and biomass of groups targeted by fisheries (Tar).
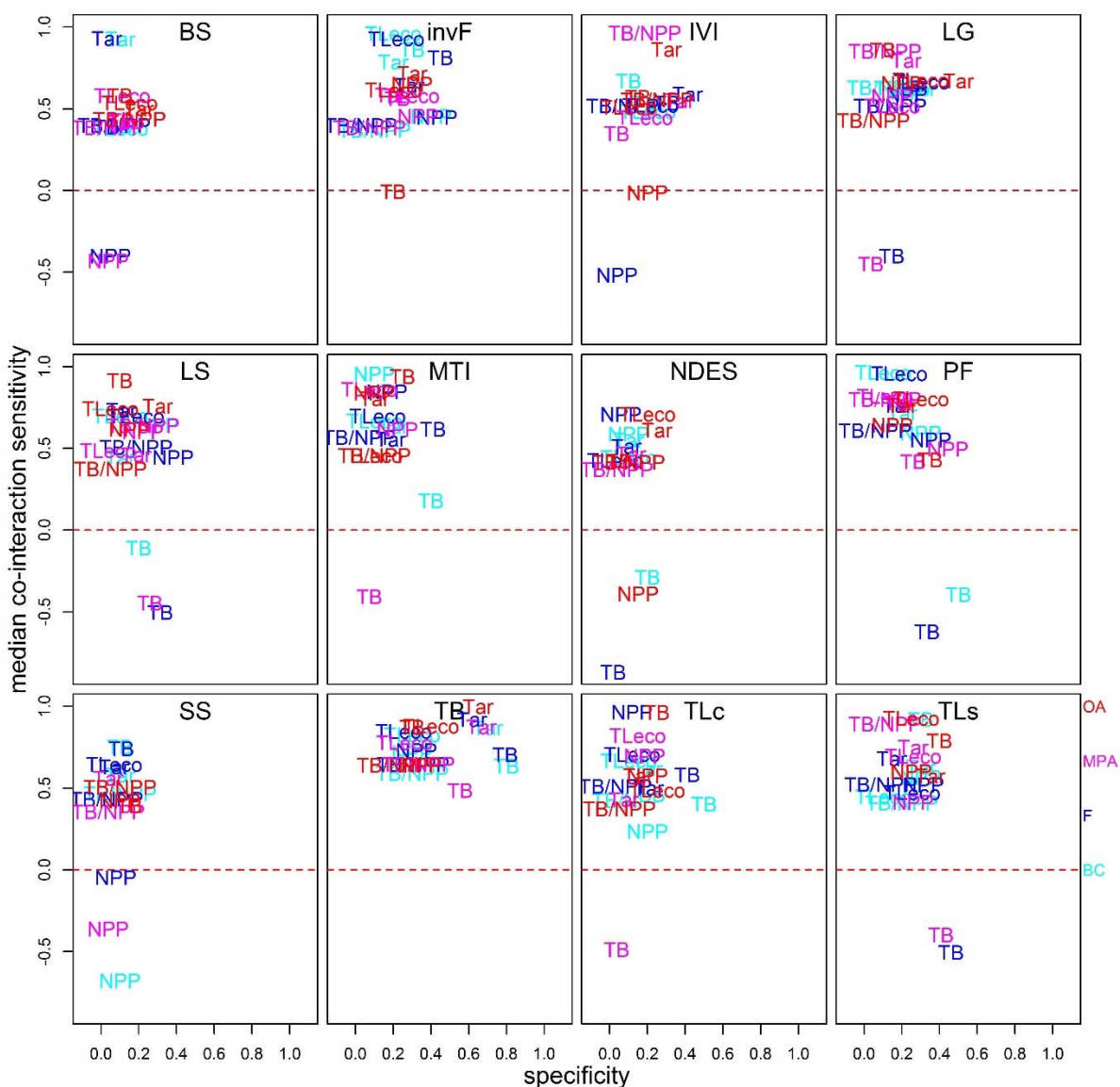
Figure B.2. Same as Figure B.1, except for the co-interaction across scenario groups (i.e., the base case (BC), ocean acidification (OA), marine protected area (MPA), and fishing (F), which are depicted using colors).
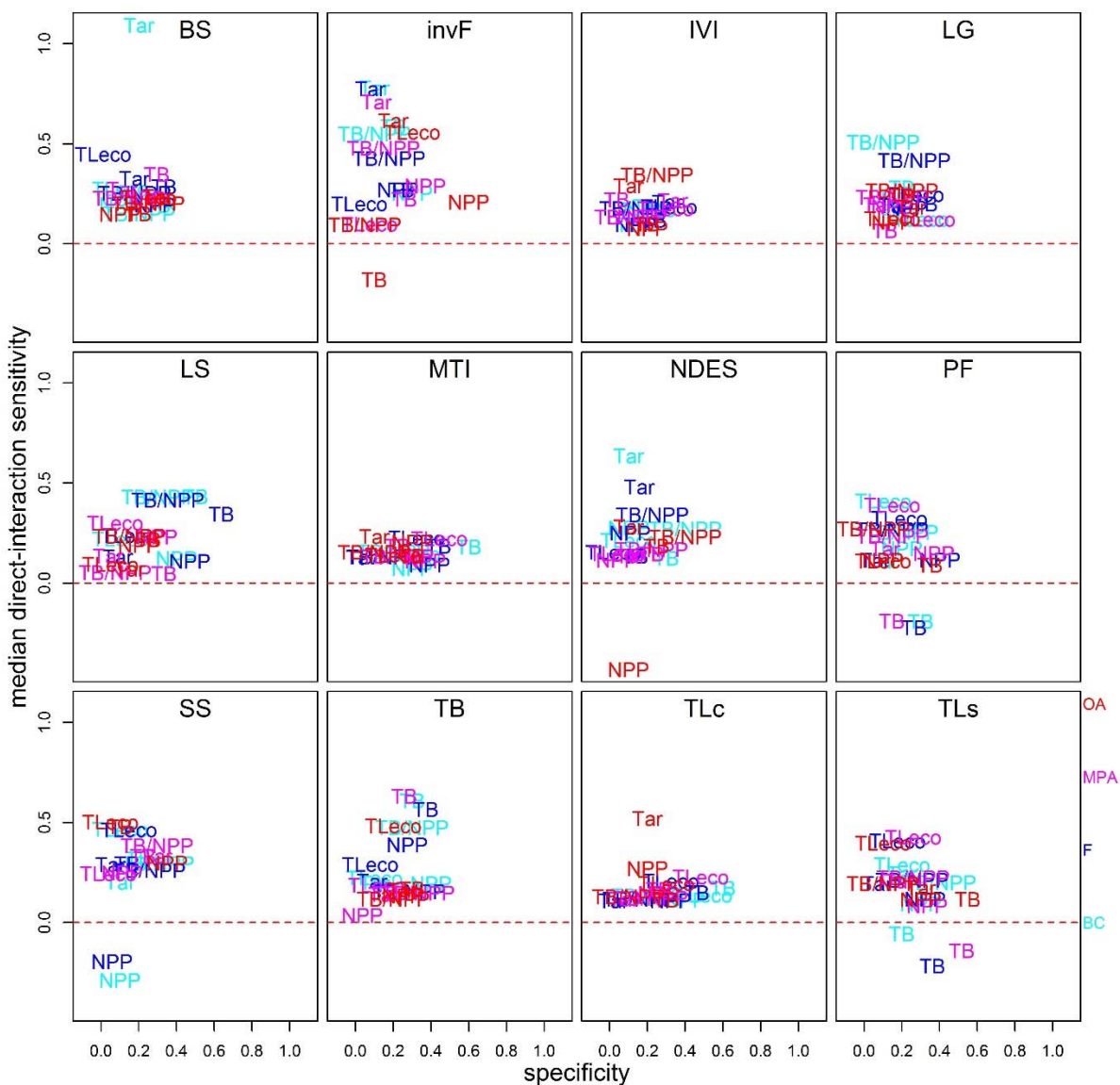
Figure B.3. Same as Figure B.1, except for the direct-interaction across scenario groups (i.e., the base case (BC), ocean acidification (OA), marine protected area (MPA), and fishing (F), which are depicted using colors).
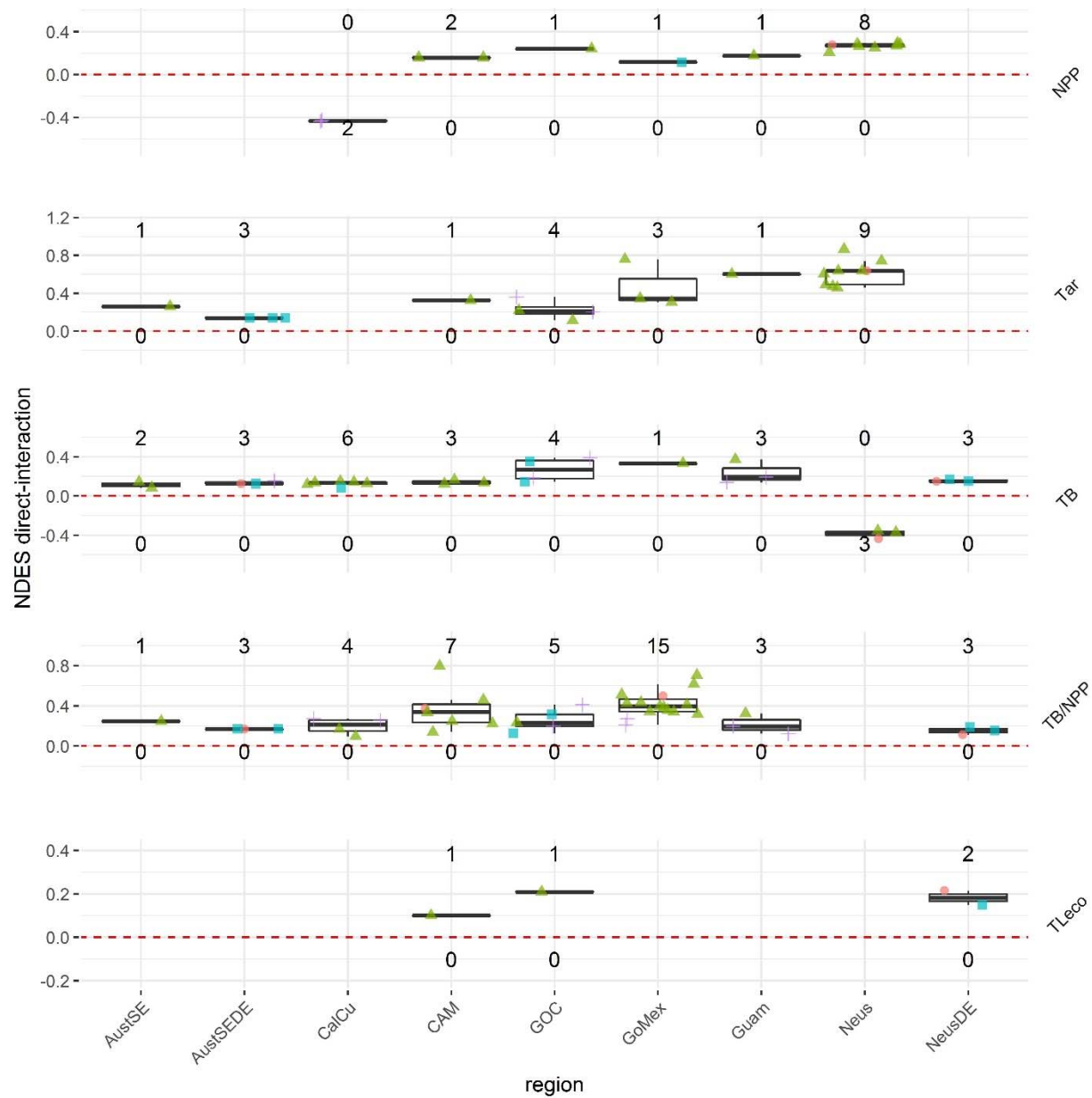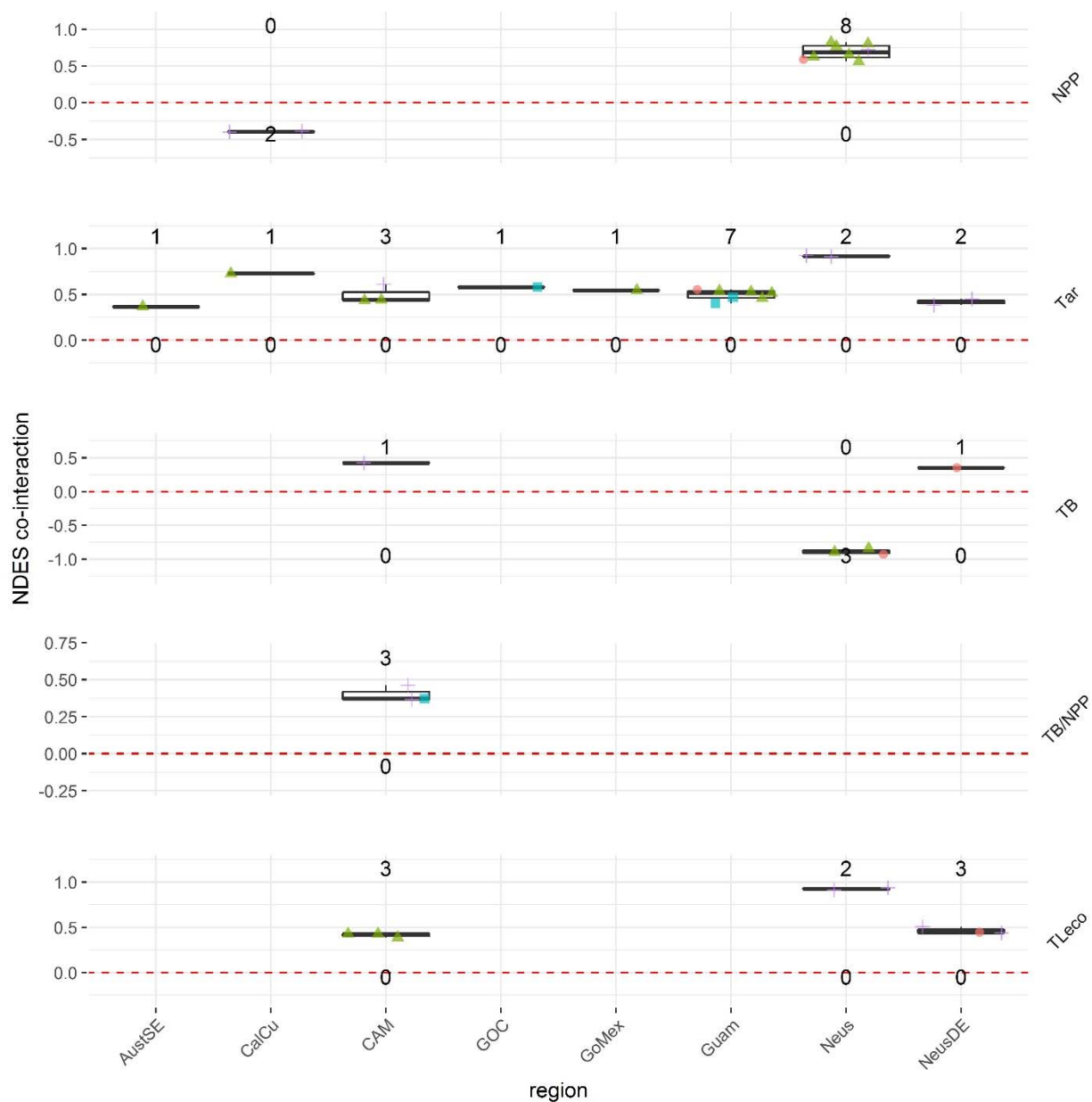
Figure B.4. Interactions for proportion of non-declining species of those with an assessment of their status within a region (NDES) for a given attribute (rows). Upper and lower panel are direct- and co-interactions, respectively. Whiskers depict 1.5 times the first and third quartiles and colored shapes are overlaid such that drivers of change within an ecosystem can be visualized. Red circles are the base case, green triangles are fishing scenarios, blue squares are marine protected area scenarios, and purple crosses are ocean acidification scenarios. The top row of numbers in each panel are the number of relationships that were sensitive and greater than zero and the bottom row is the number of relationships that were sensitive and less than zero. Refer to text in Appendix for

region abbreviations. Attributes are abbreviated as follows: net primary productivity (NPP), total biomass (TB), mean trophic level of the ecosystem (TLeco), TB divided by NPP (TB/NPP), and biomass of groups targeted by fisheries (Tar).
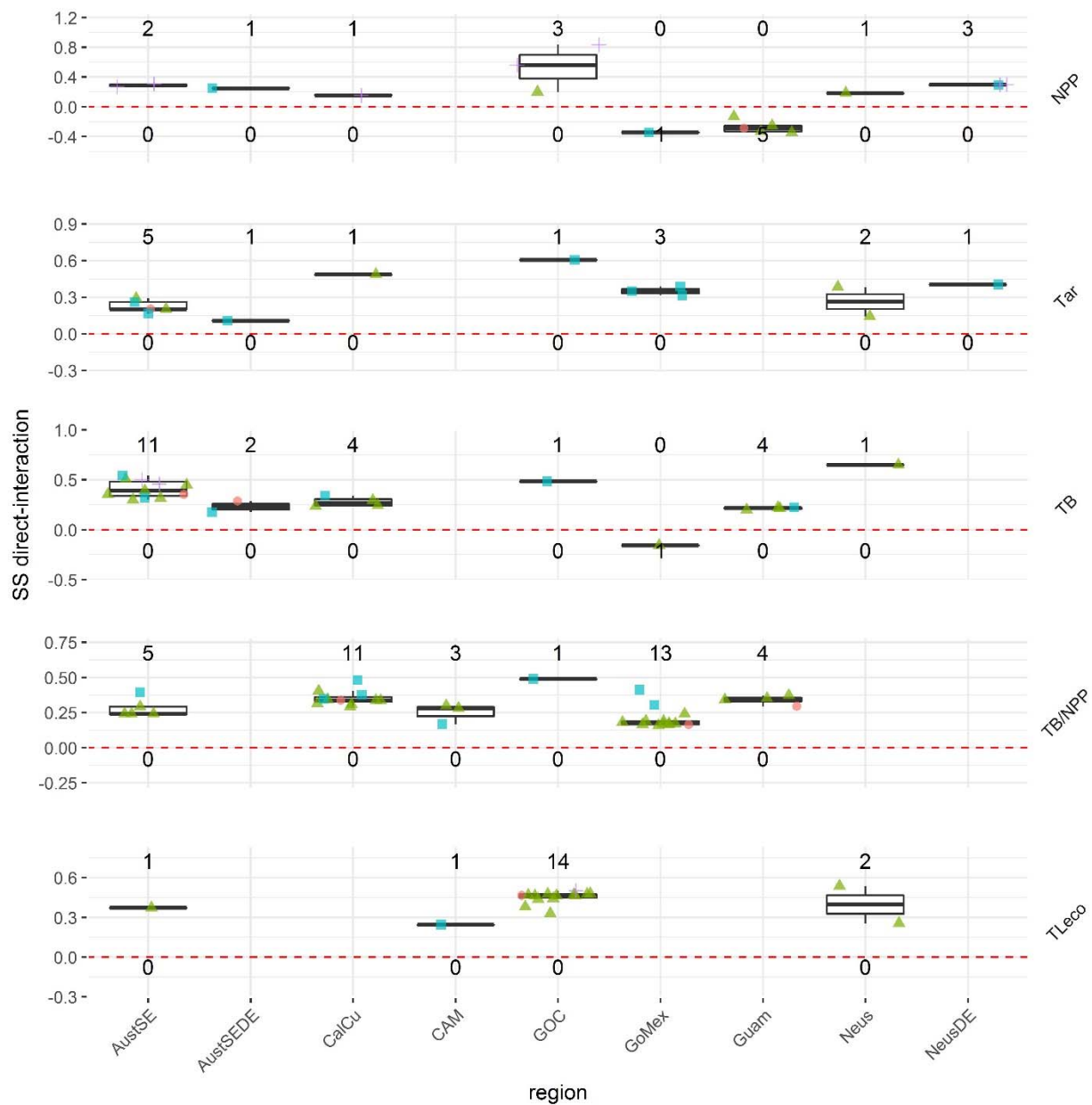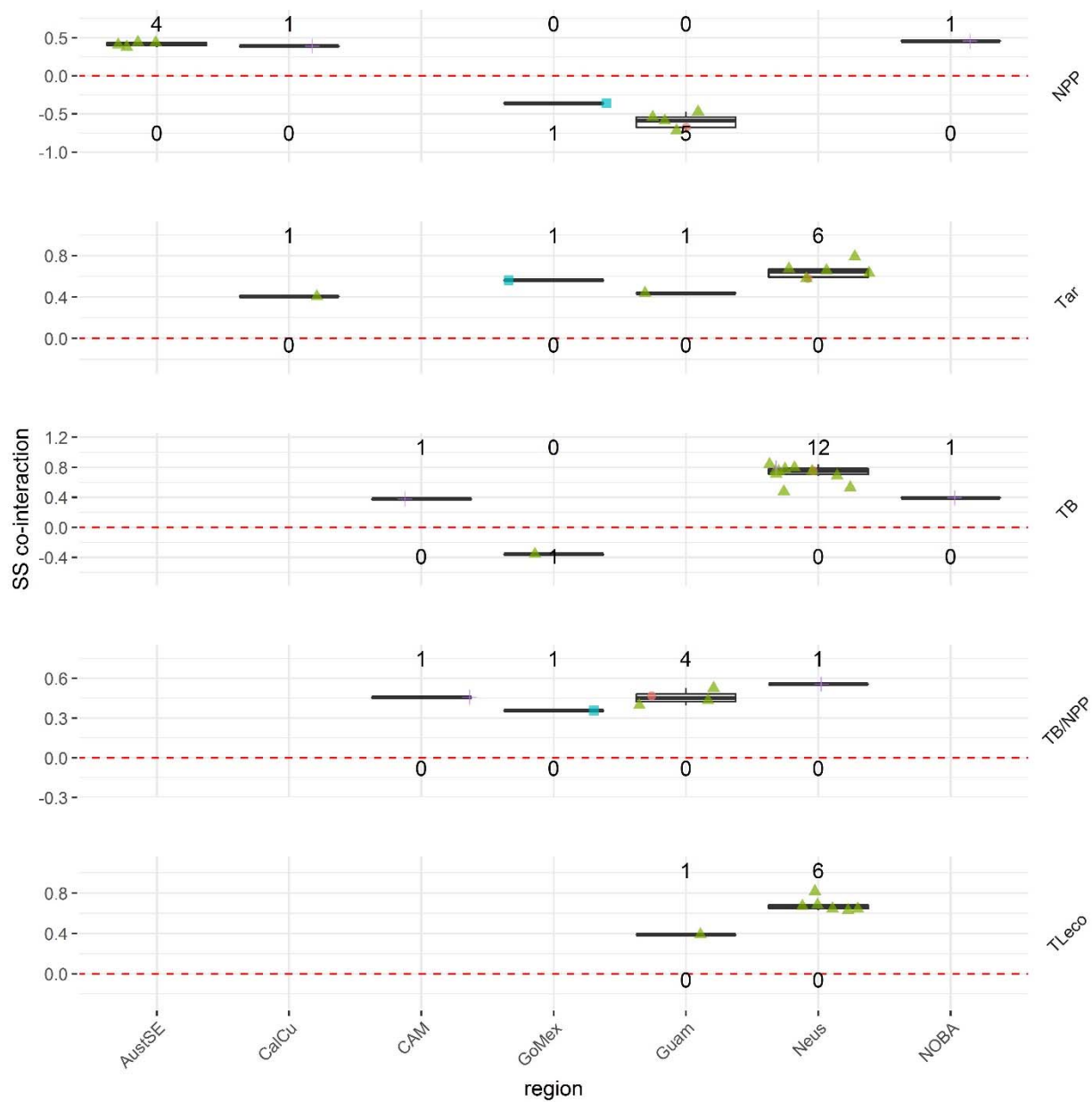
Figure B.5. Same as Figure B.4 except for sustainable stocks.
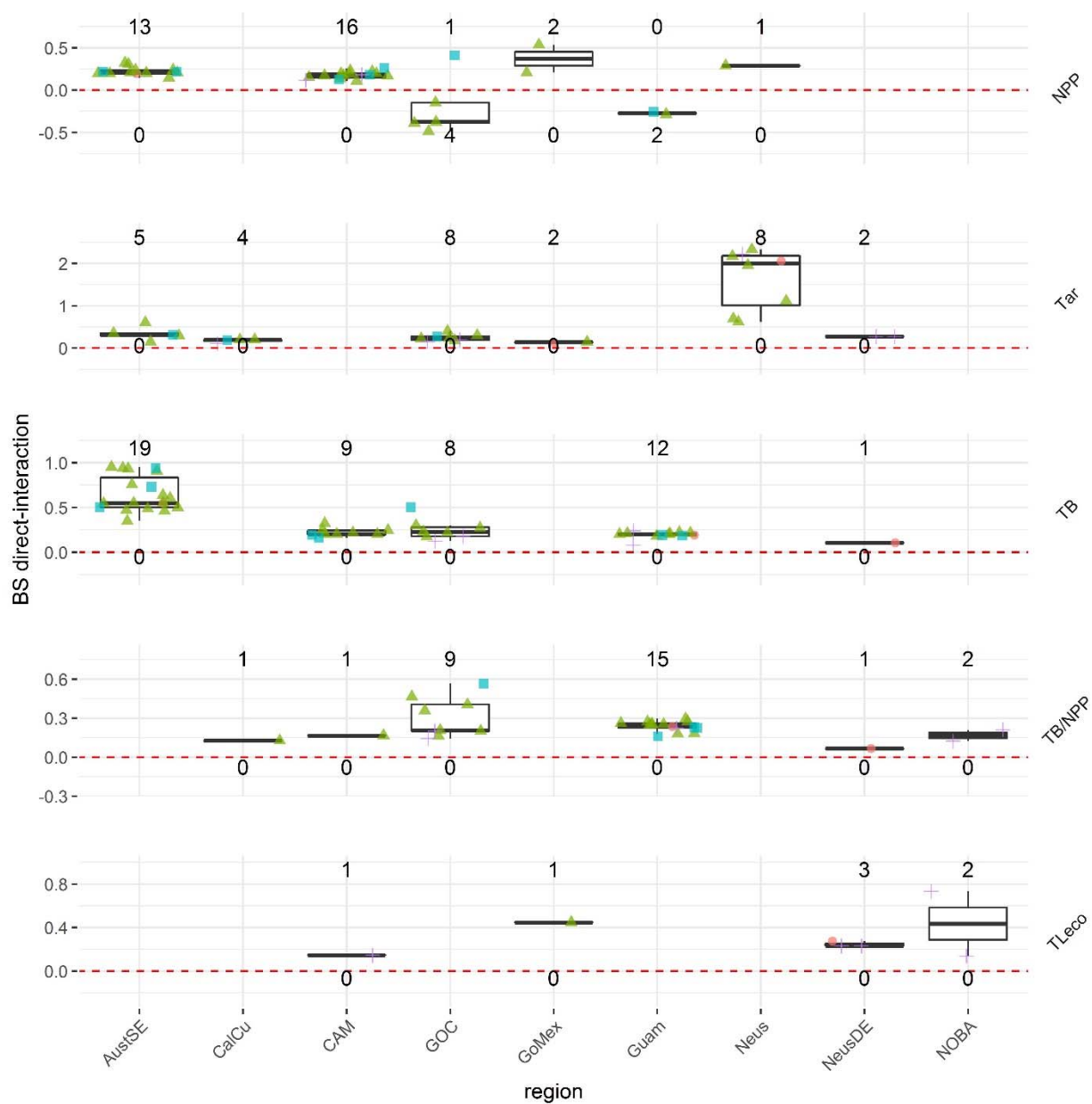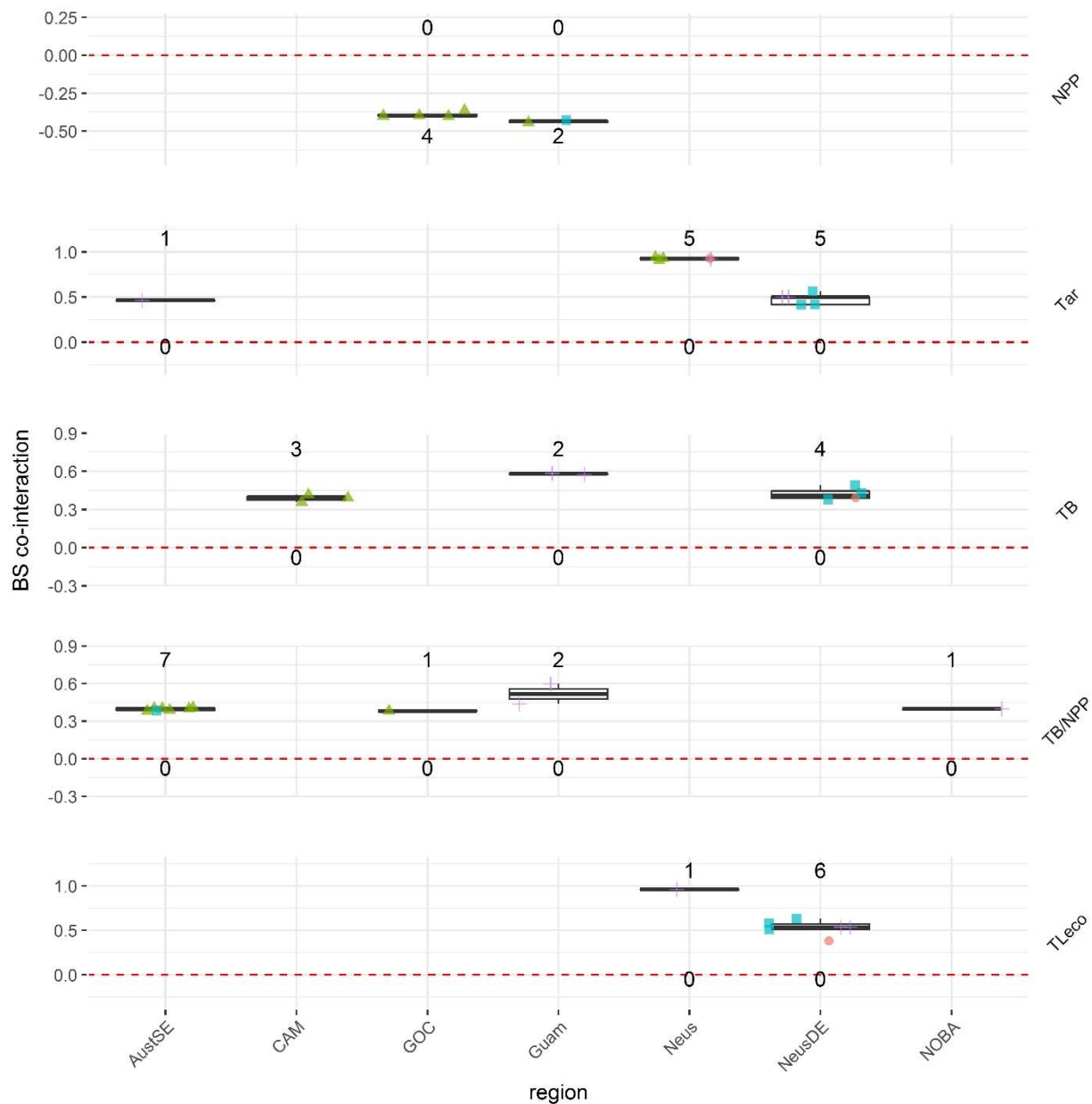
Figure B.6. Same as Figure B.4 except for biomass stability.
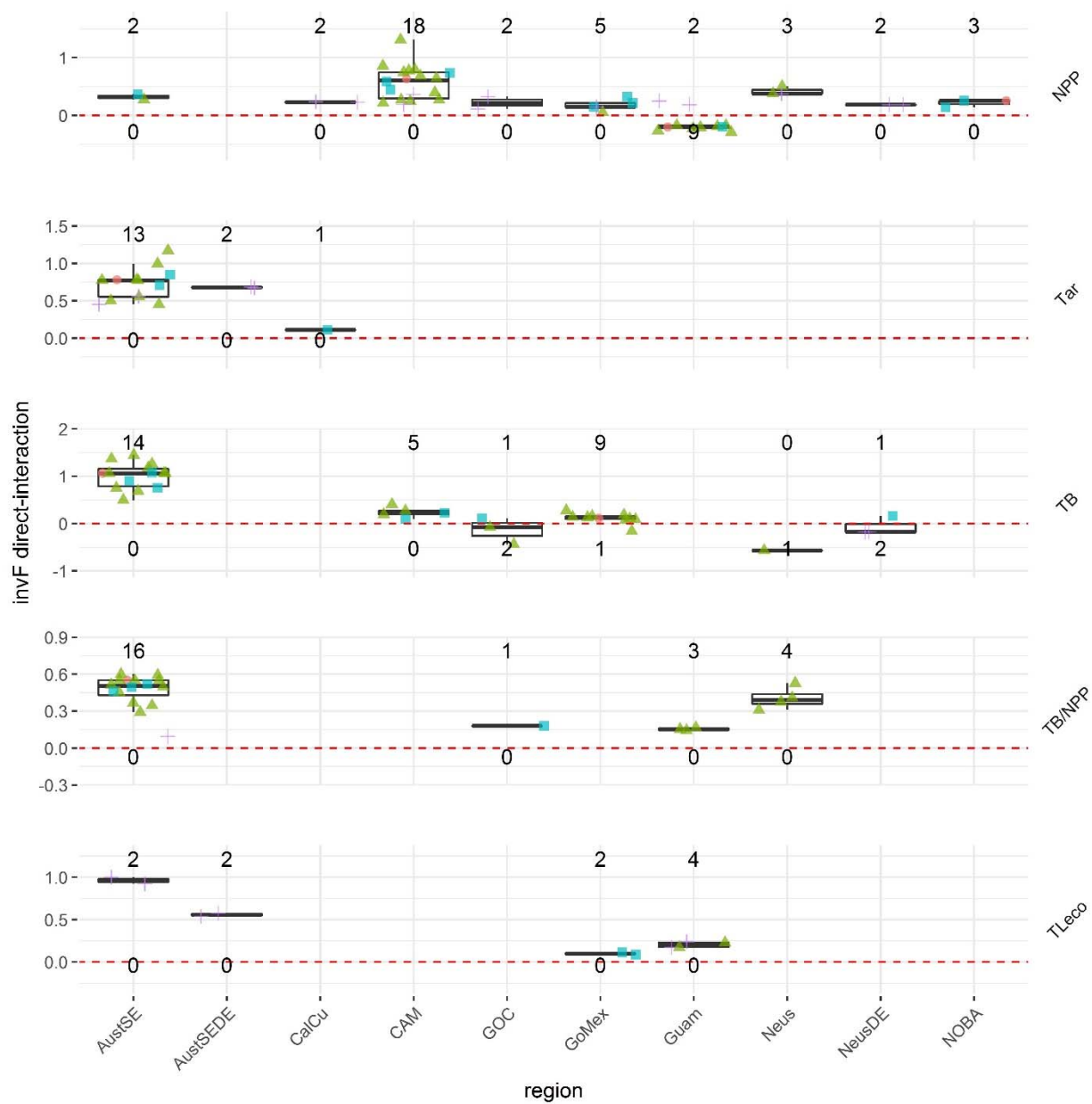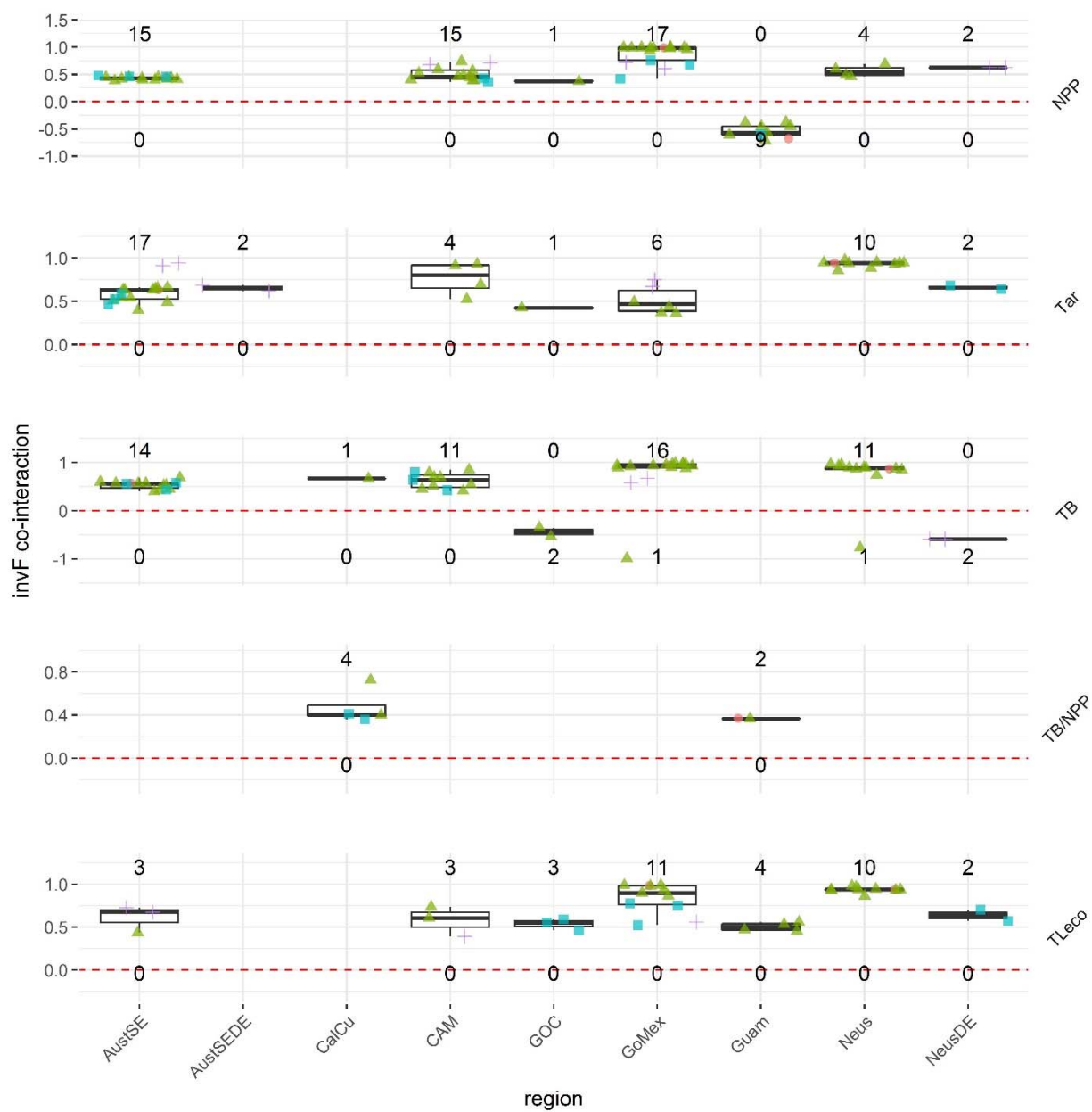
Figure B.7. Same as Figure B.4 except for inverse fishing pressure.
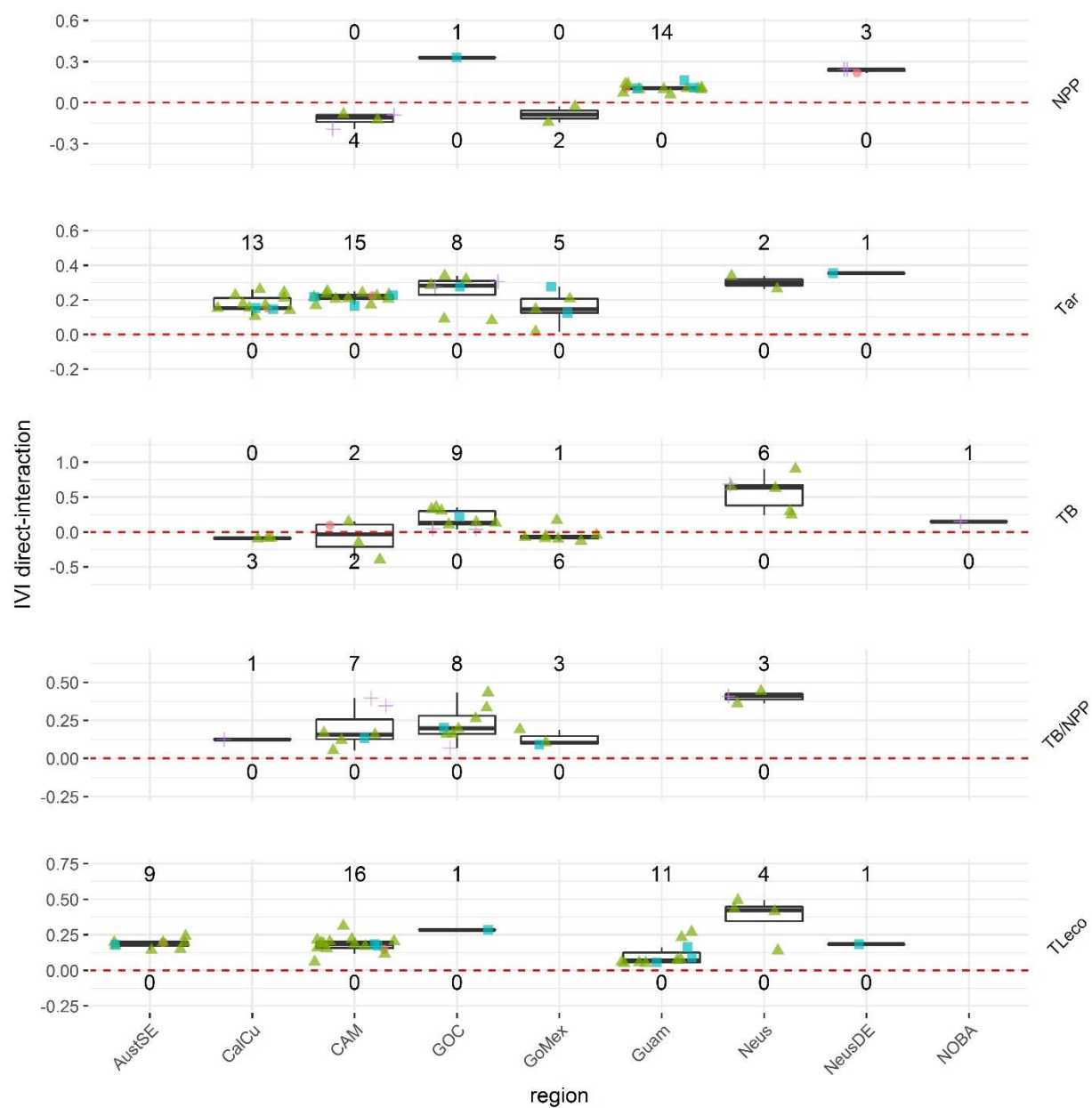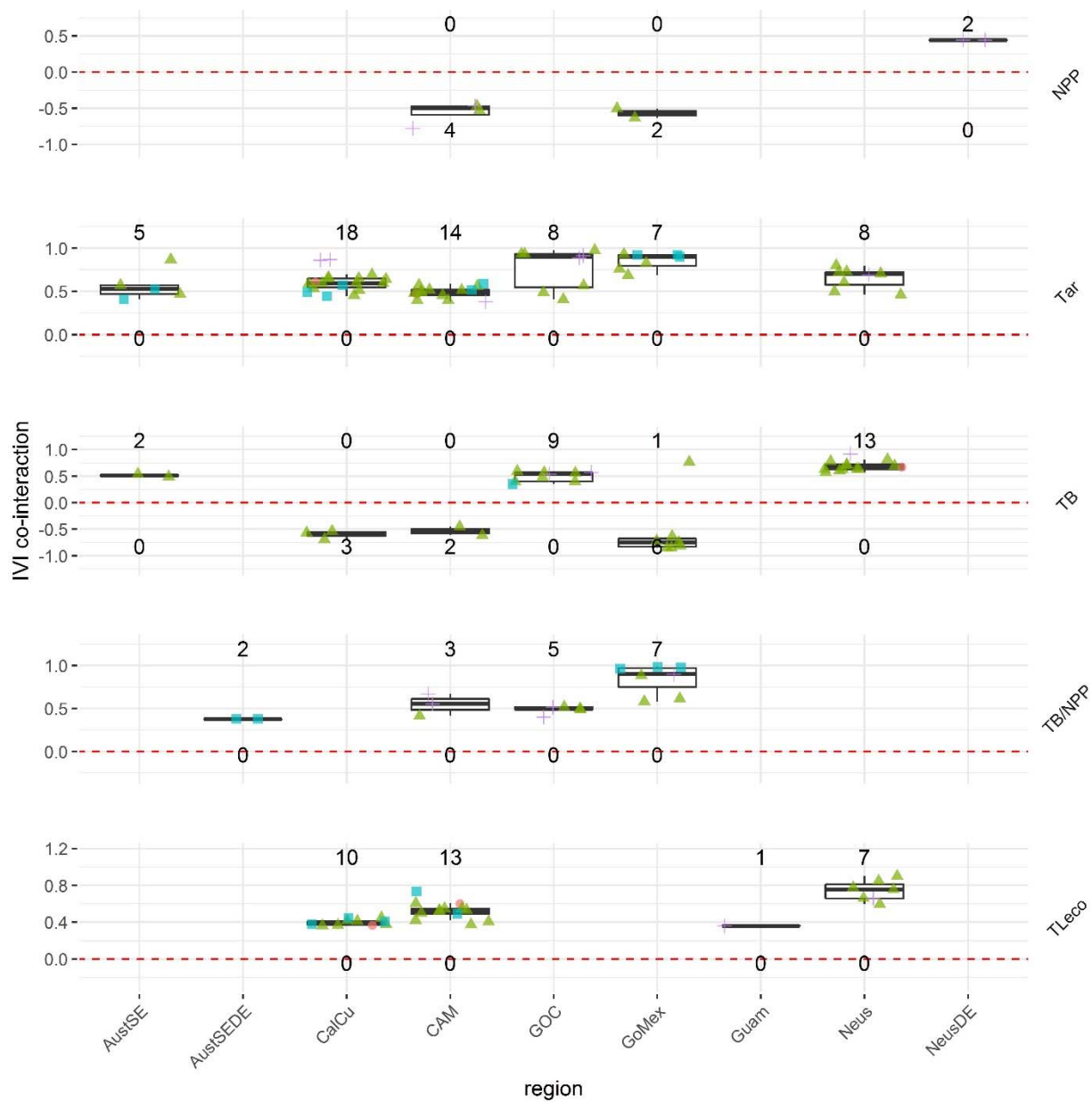
Figure B.8. Same as Figure B.4 except for intrinsic vulnerability index.
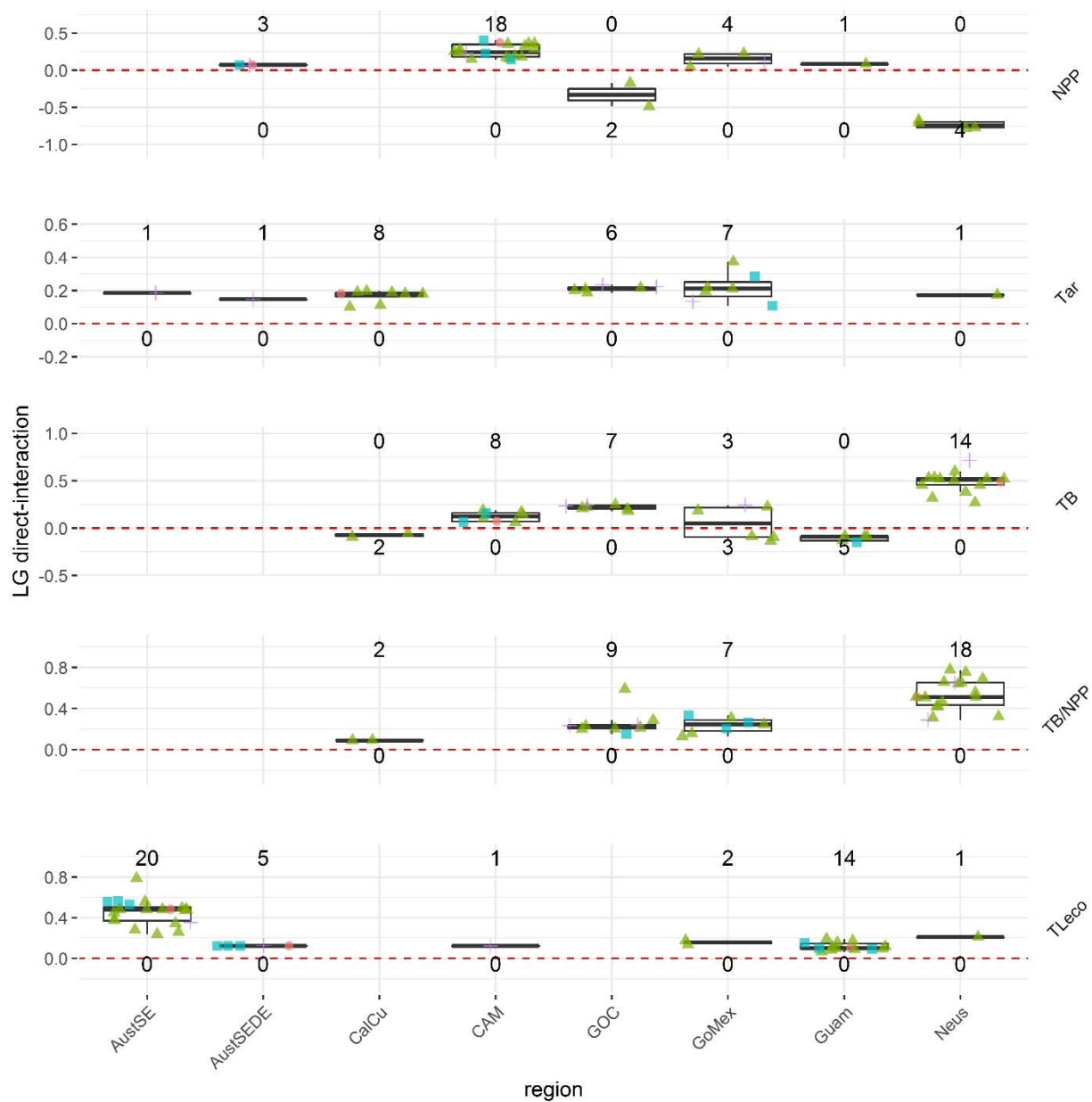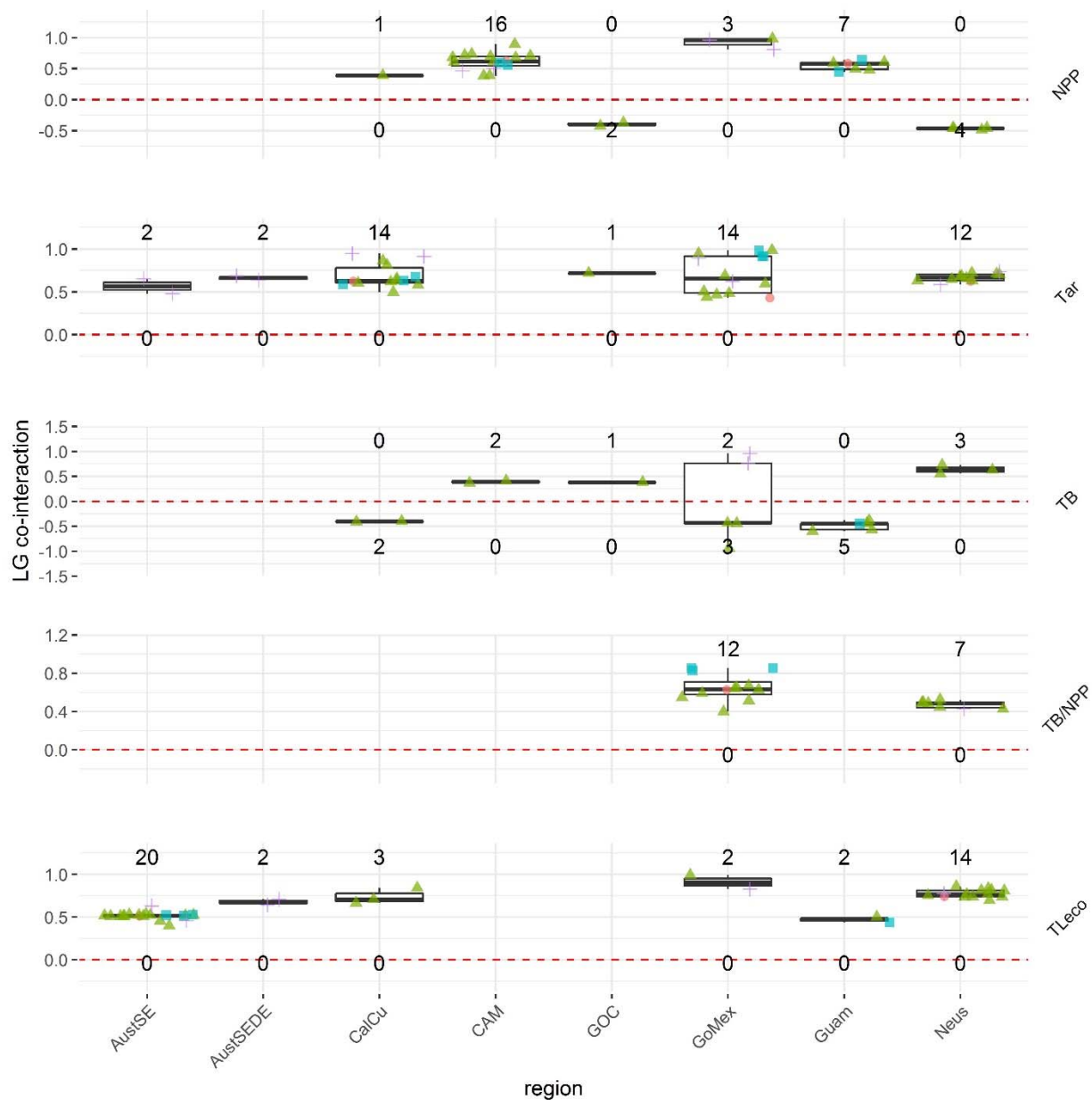
Figure B.9. Same as Figure B.4 except for biomass mean maximum length of fish in the surveyed community.
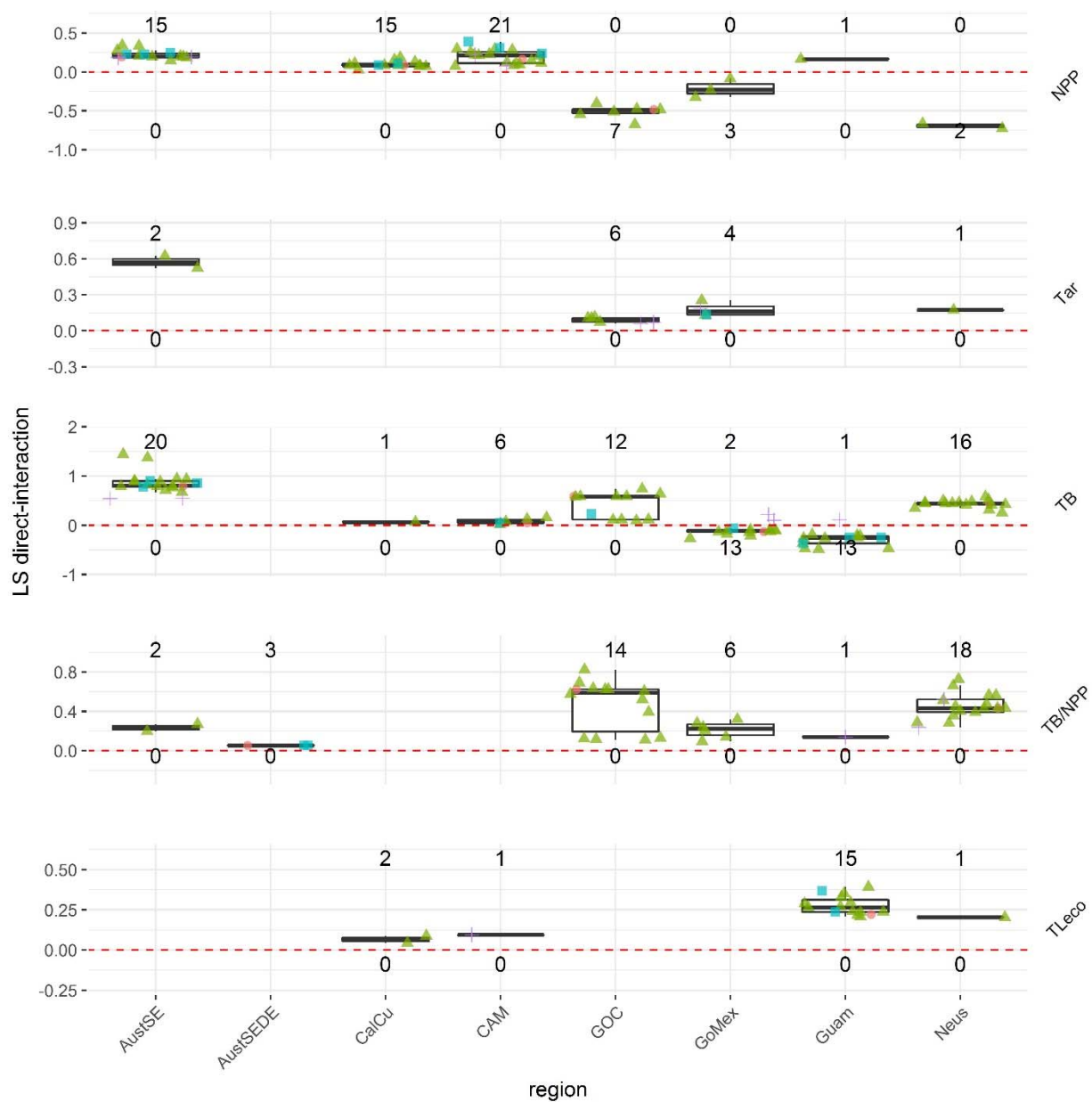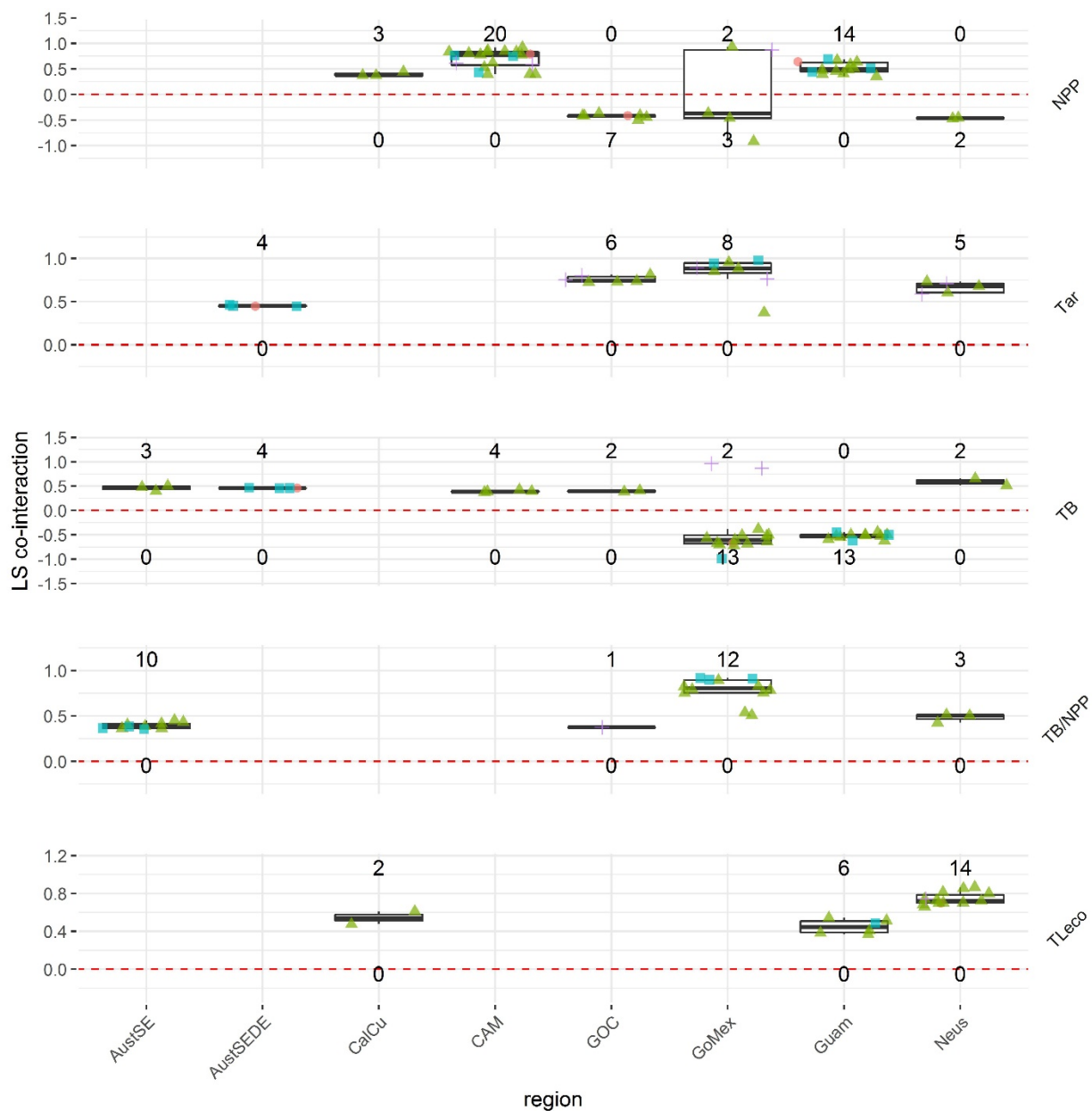
Figure B.10. Same as Figure B.4 except for mean life span of fish in the surveyed community.
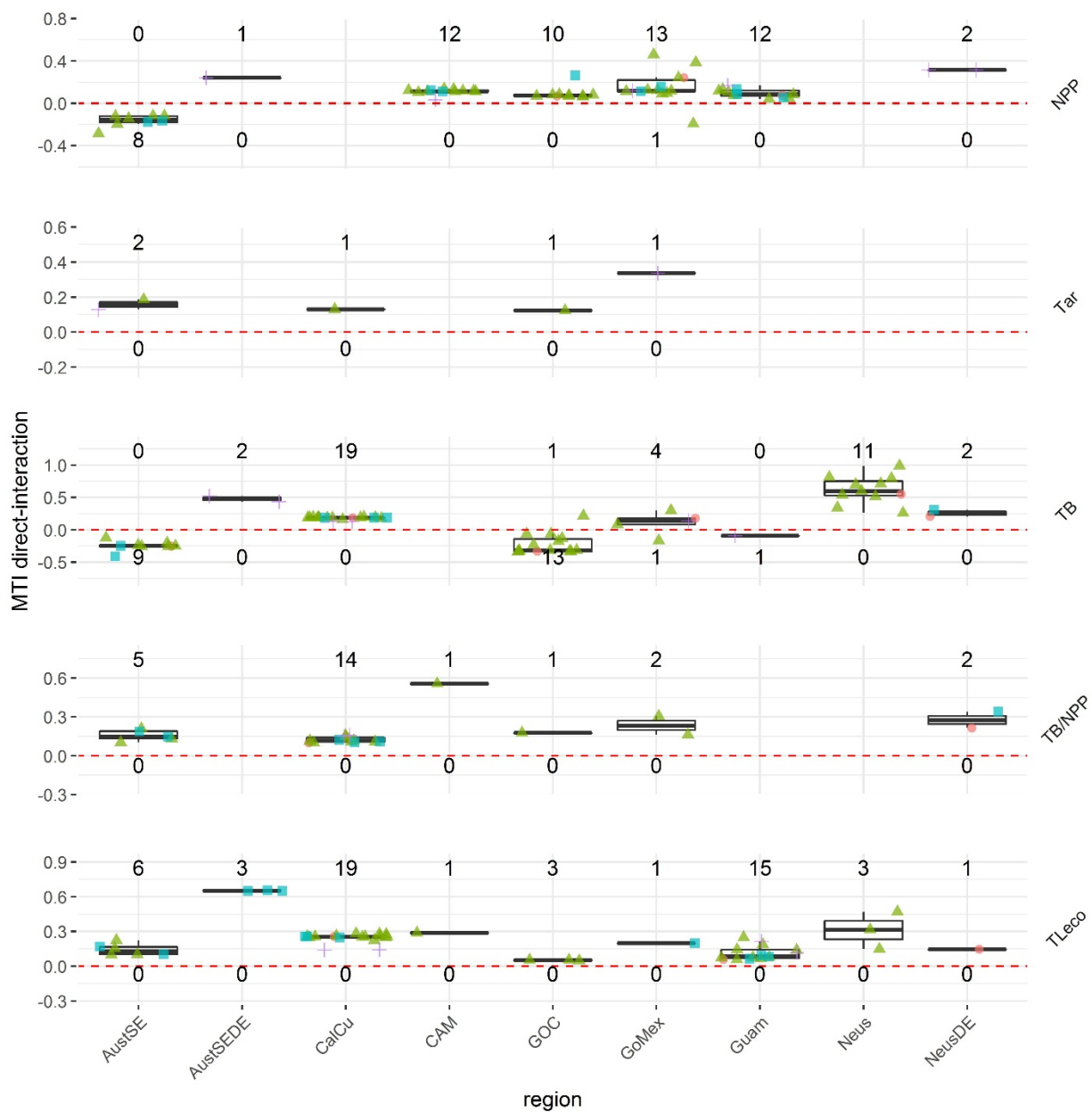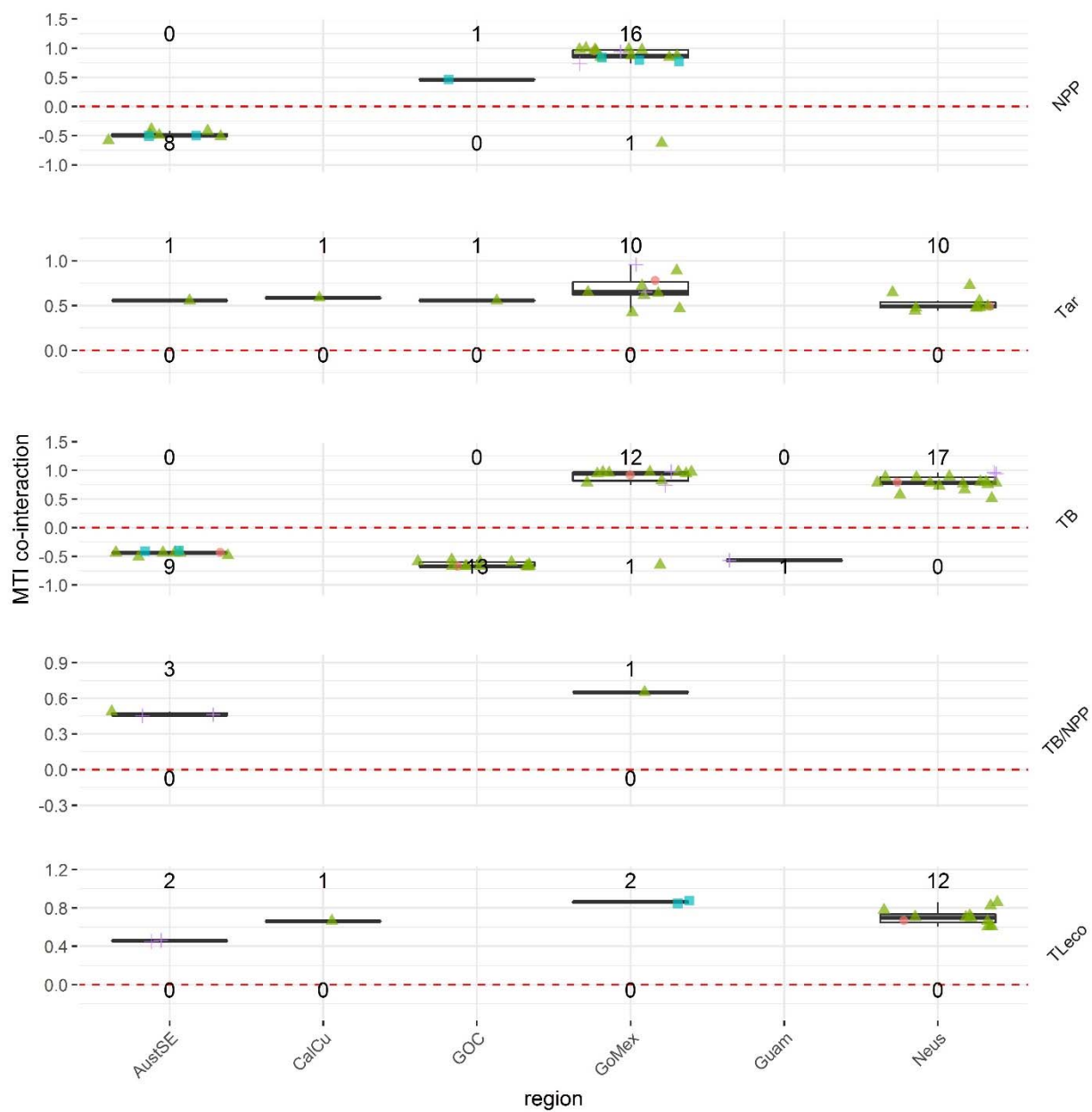
Figure B.11. Same as Figure B.4 except for biomass weighted mean trophic index of the catch.
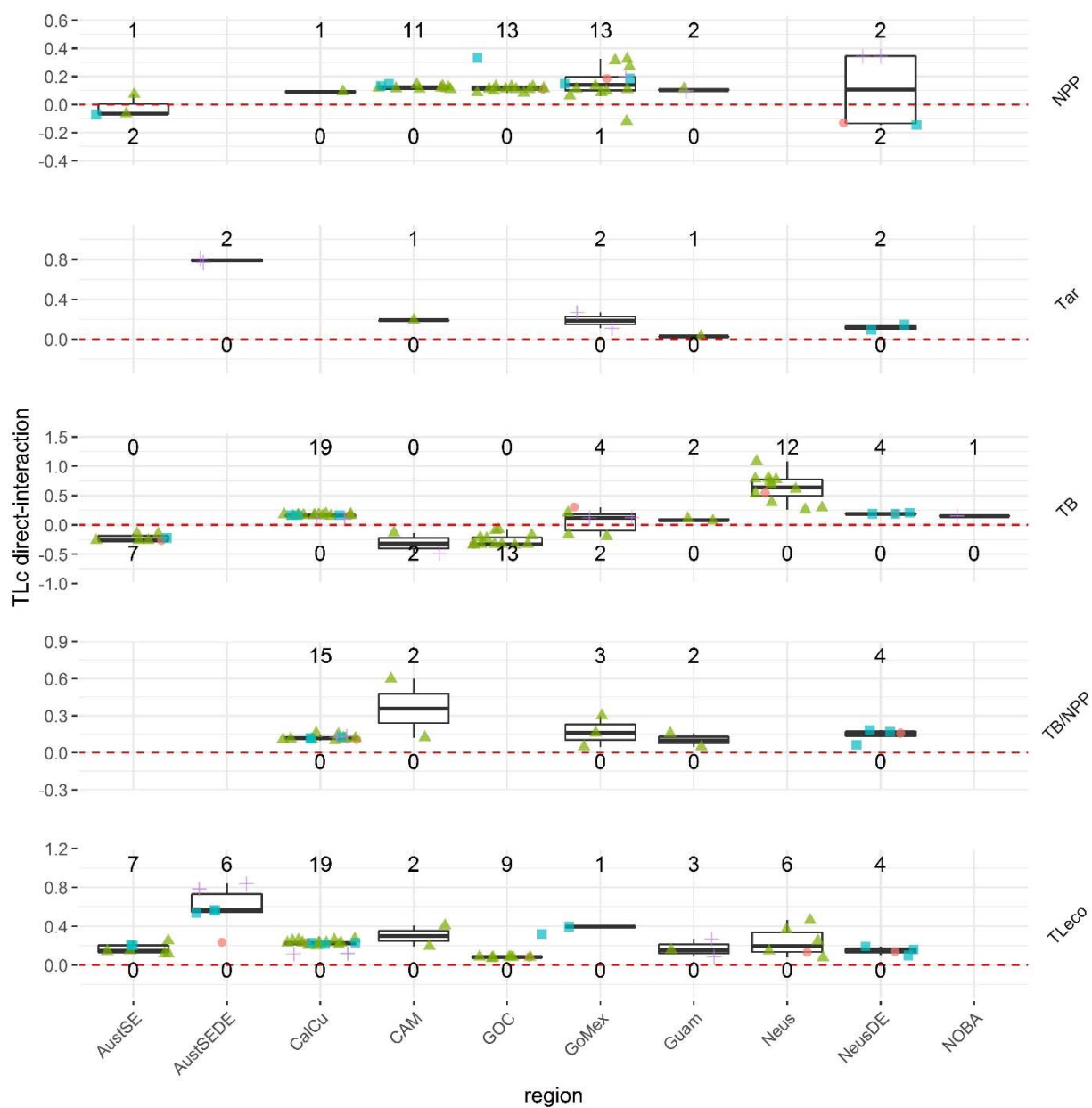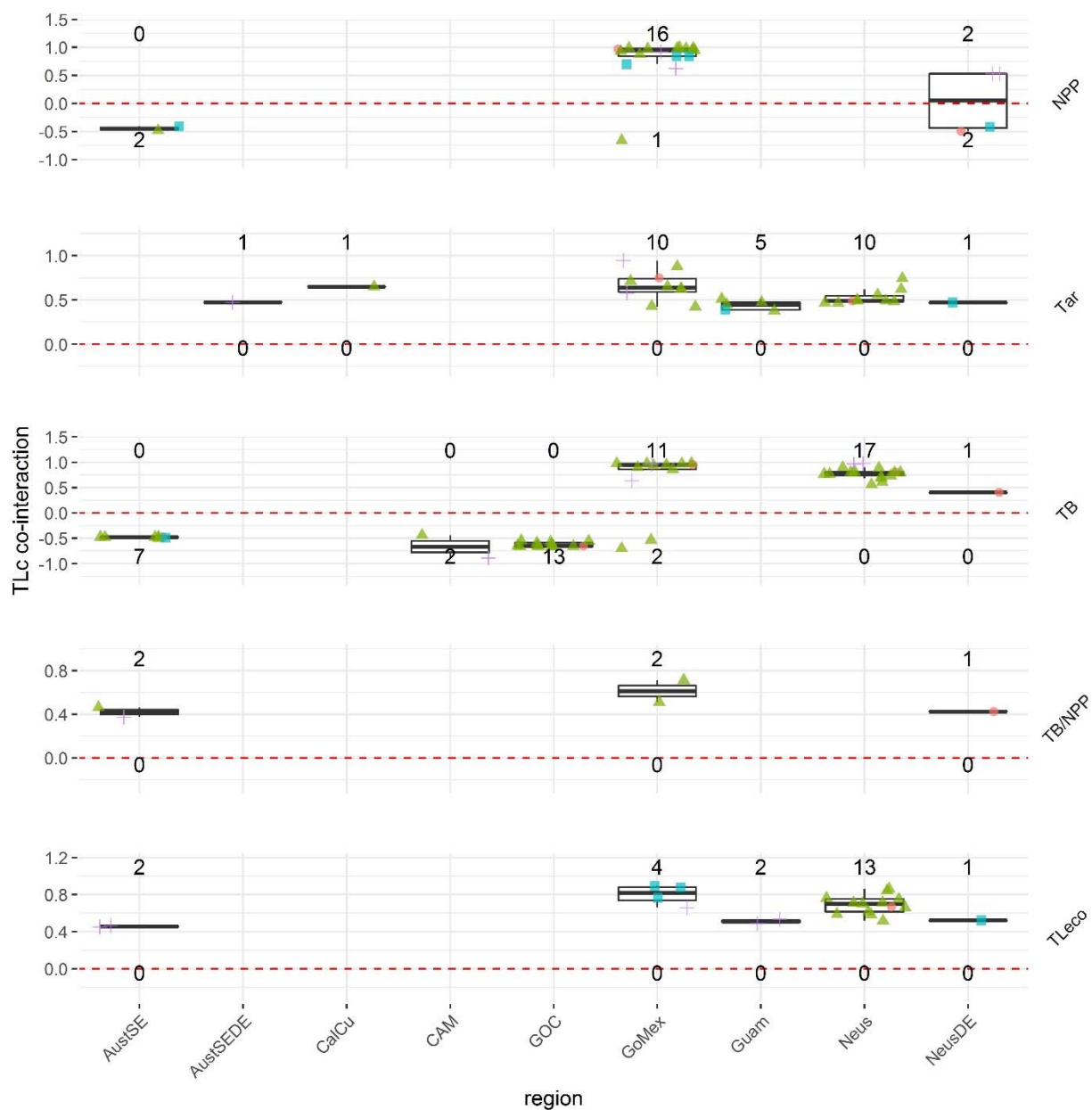
Figure B.12. Same as Figure B.4 except for trophic level (TL) of the catch.
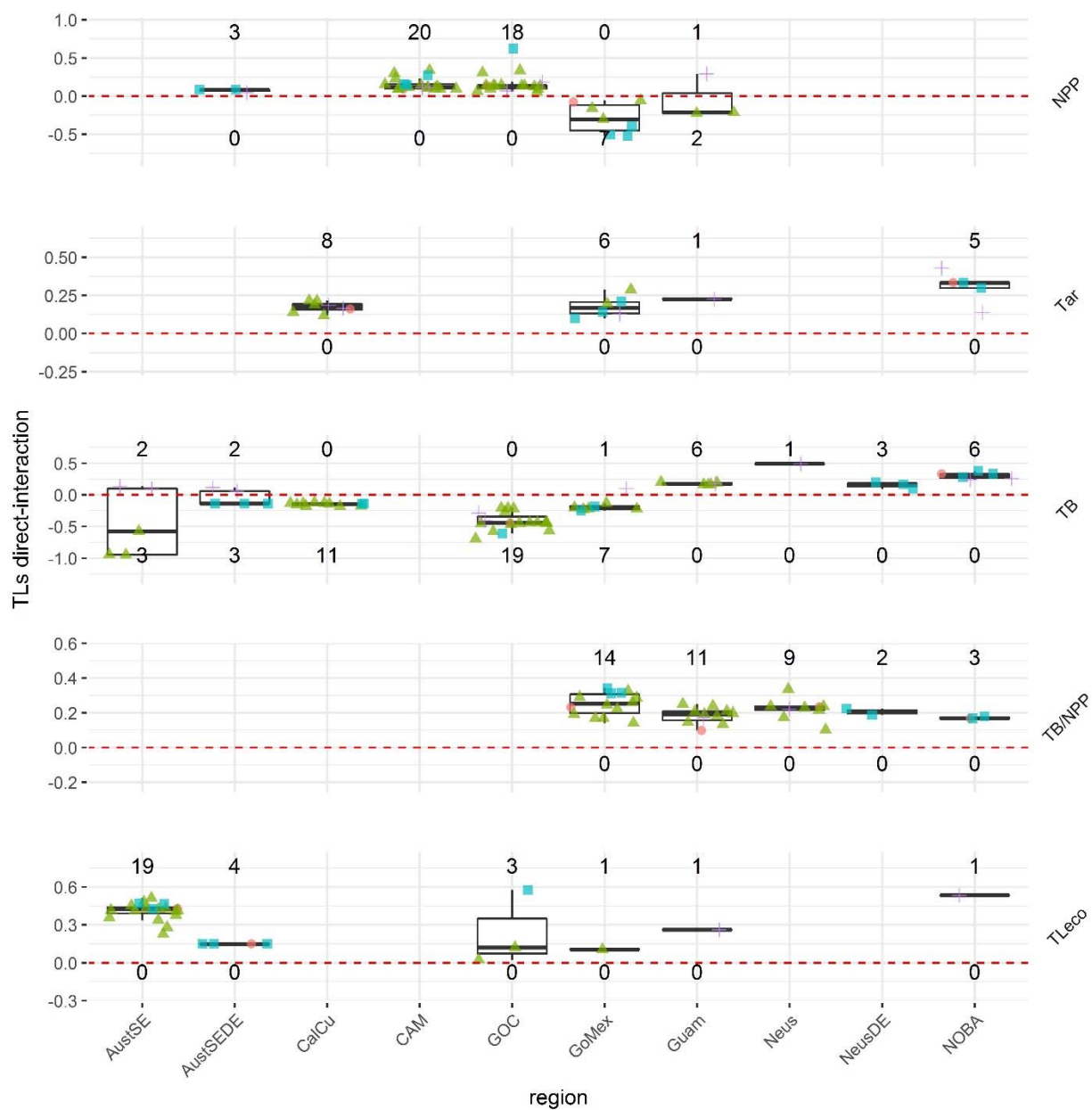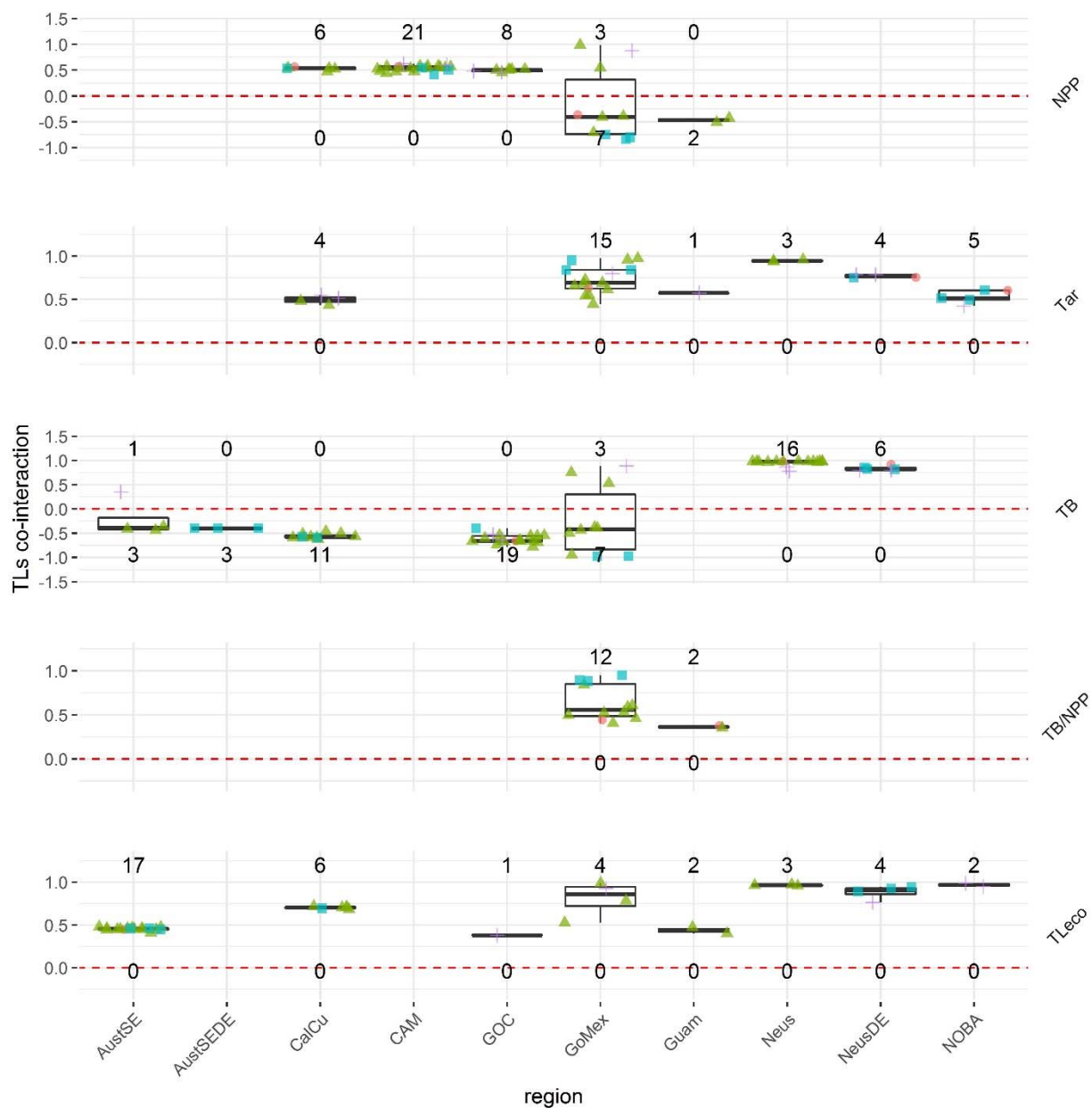
Figure B.13. Same as Figure B.4 except for trophic level (TL) of the survey.
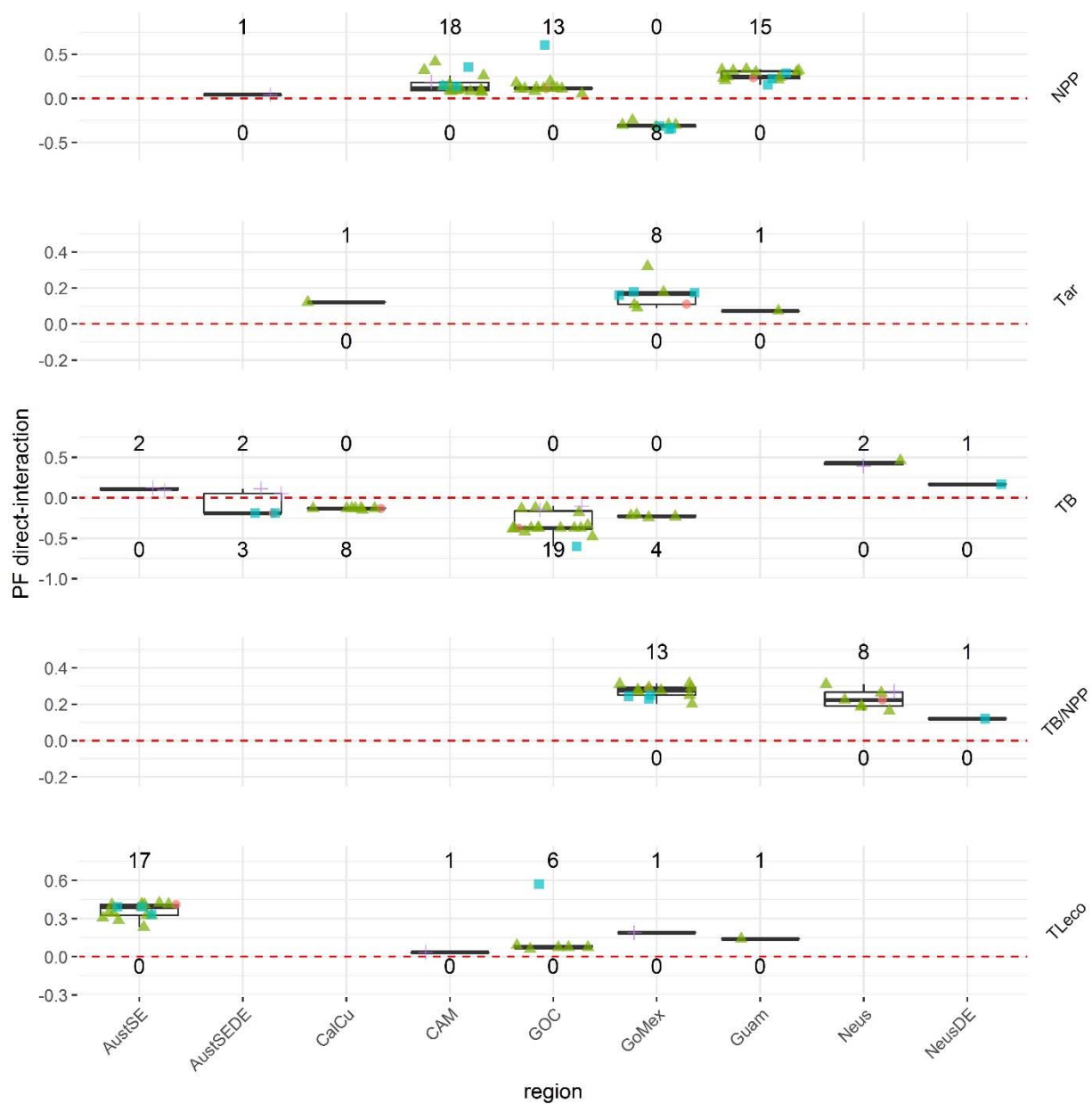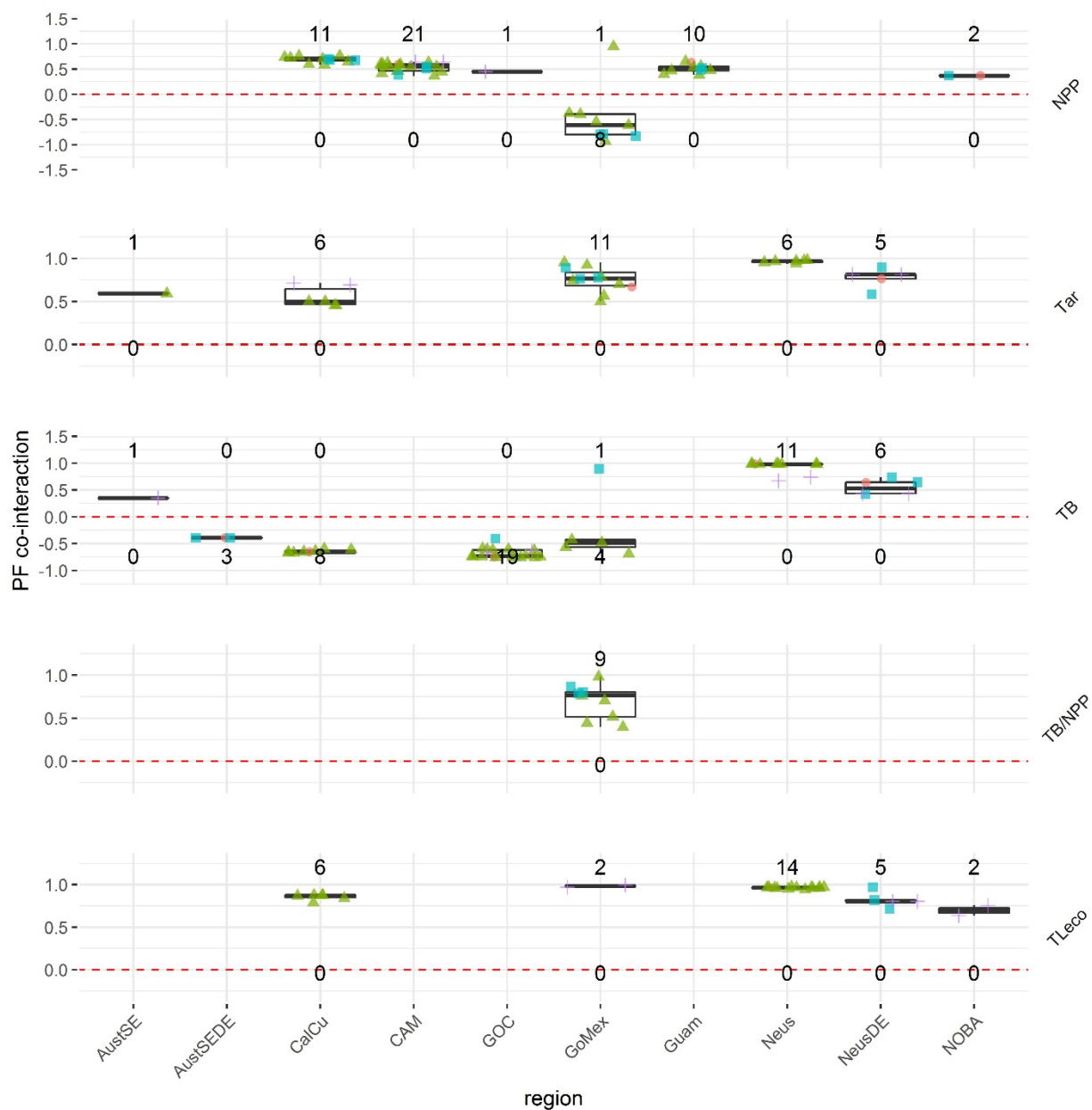
Figure B.14. Same as Figure B.4 except for the proportion of predatory fish in the surveyed community.
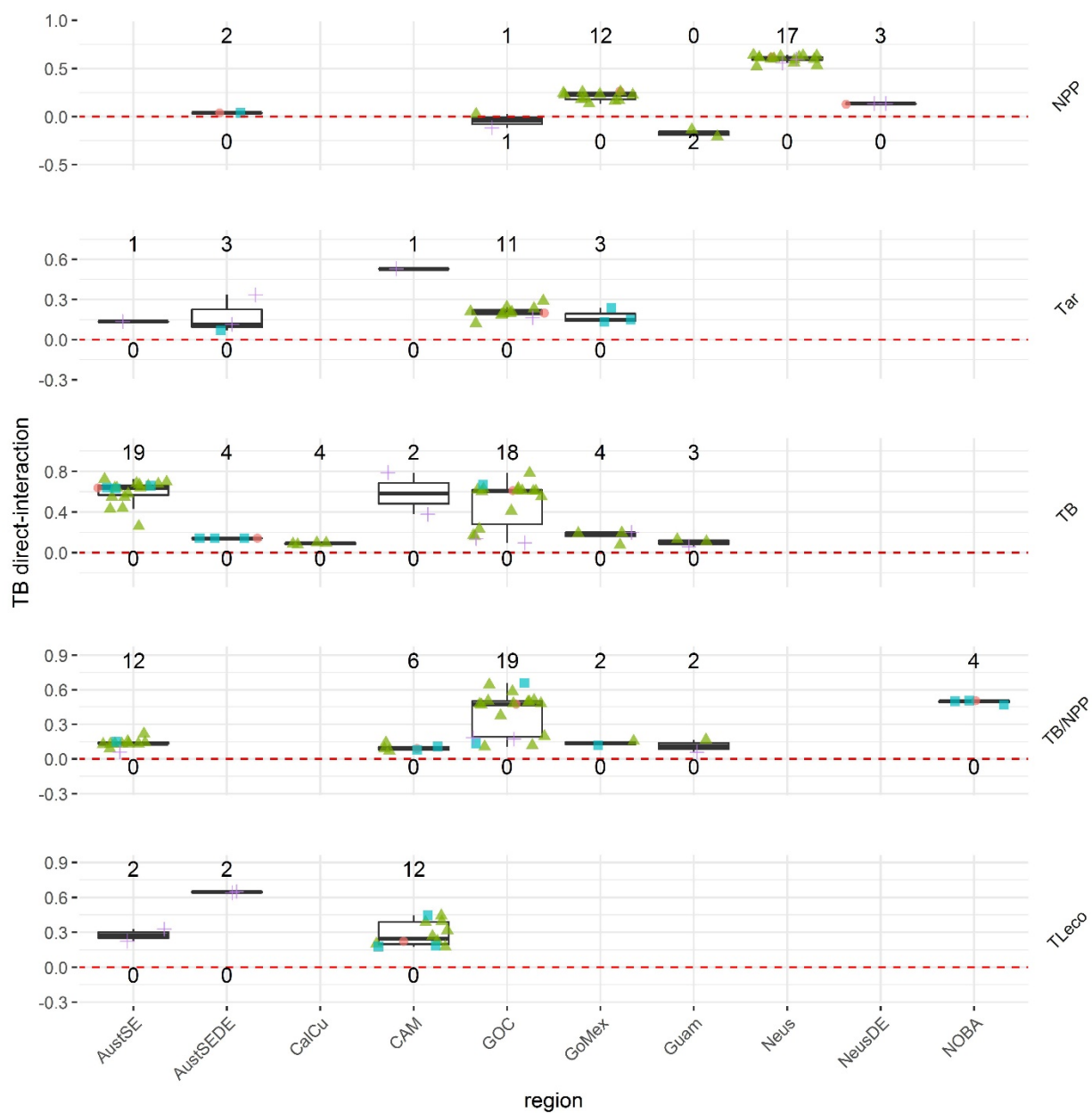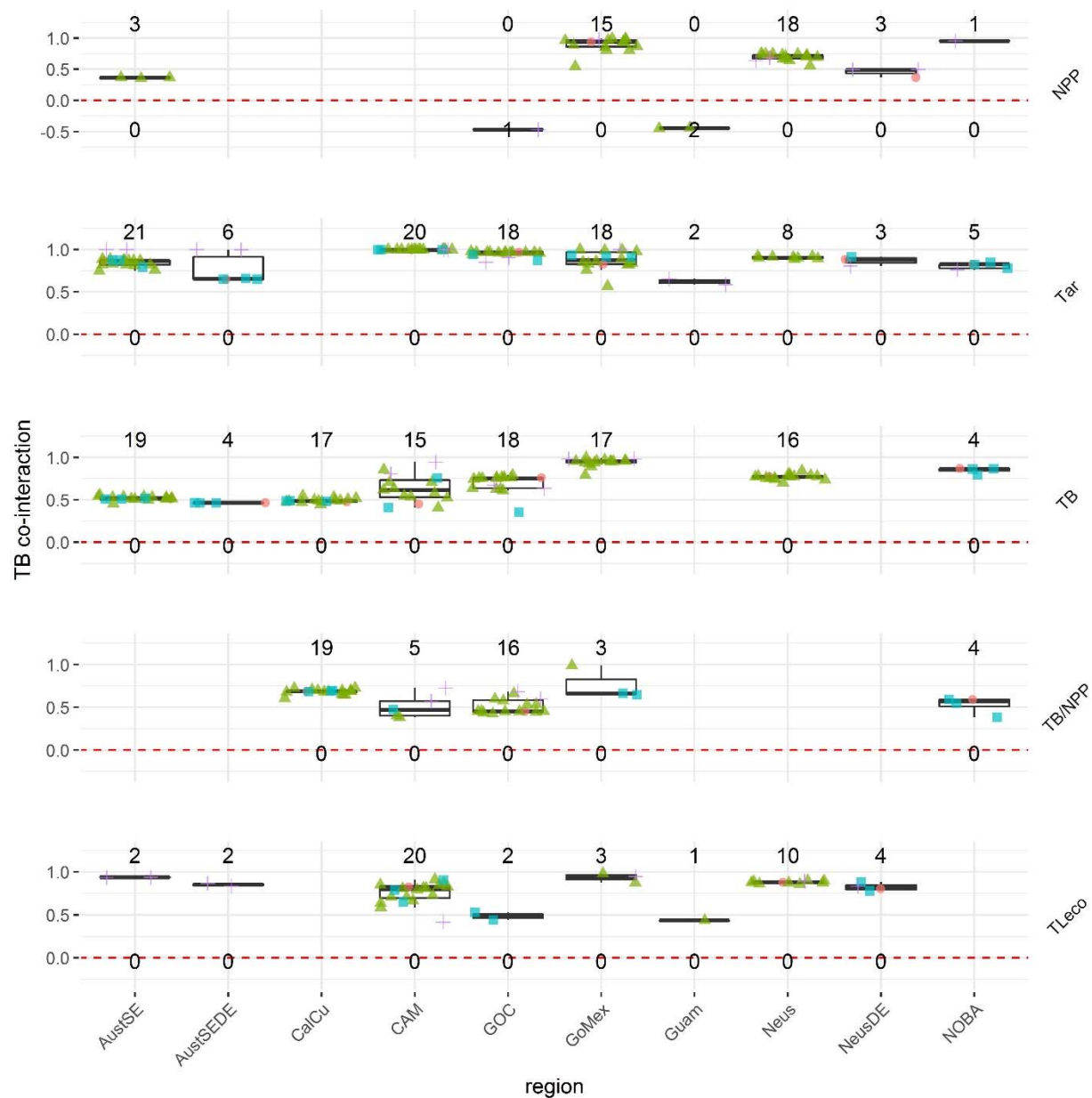
Figure B.15. Same as Figure B.4 except for biomass of the surveyed community.
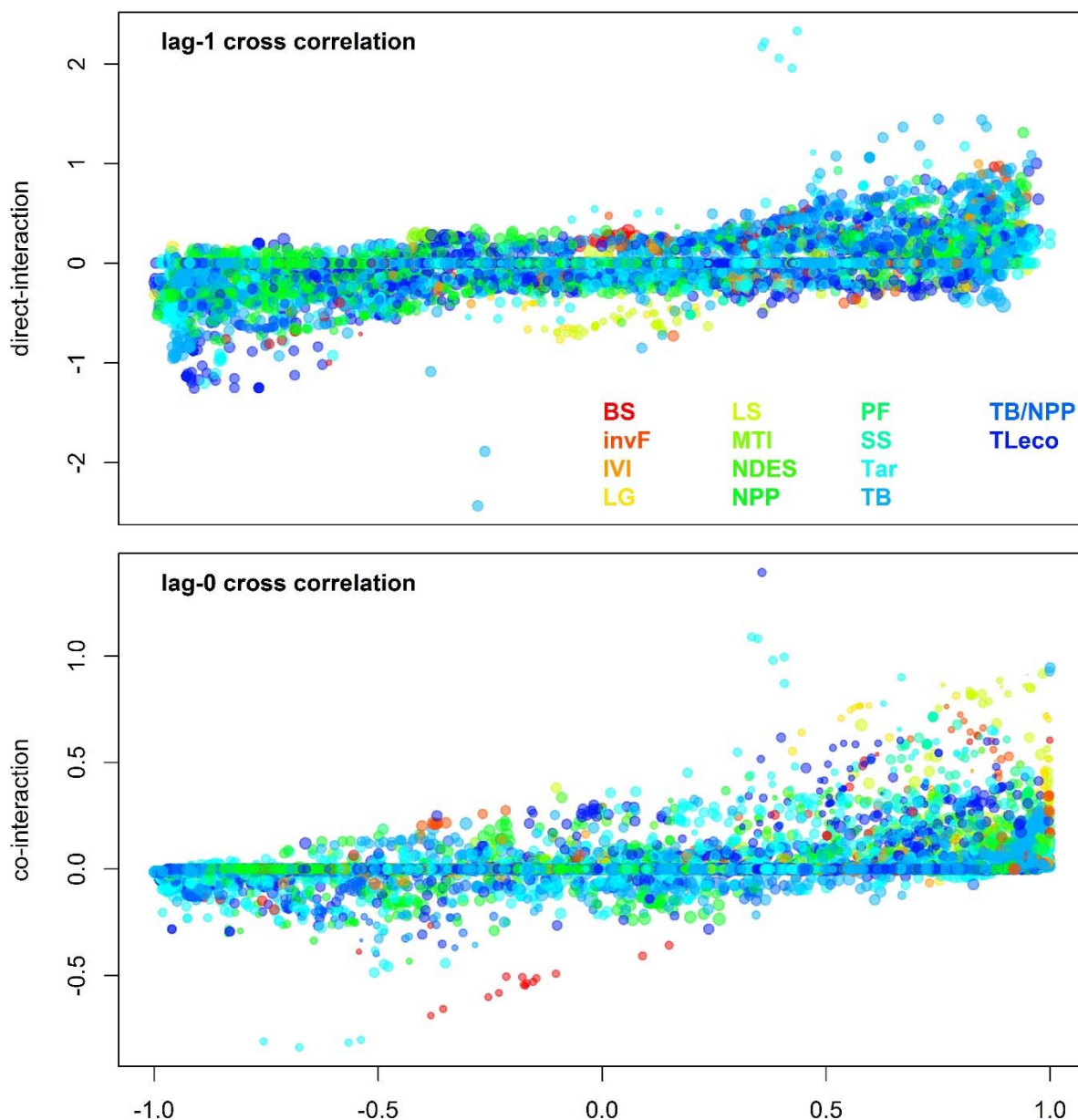
Figure B.16. Direct- and co-interactions versus lag-1 and lag-0 cross correlations (top and bottom panels, respectively). Points are colored according to the indicator and the size of the point refers to the "self-interaction" of the indicator or attribute, whichever was larger in terms of absolute value. Cross correlations are only included if their absolute value was greater than 0.35 because smaller values may not be statistically different than zero and all interactions from multivariate autoregressive (MAR) models that were not statistically different than zero are included as zero.

Appendix C. SUPPLEMENTARY FIGURES FOR CHAPTER 3

C.1 FIGURES
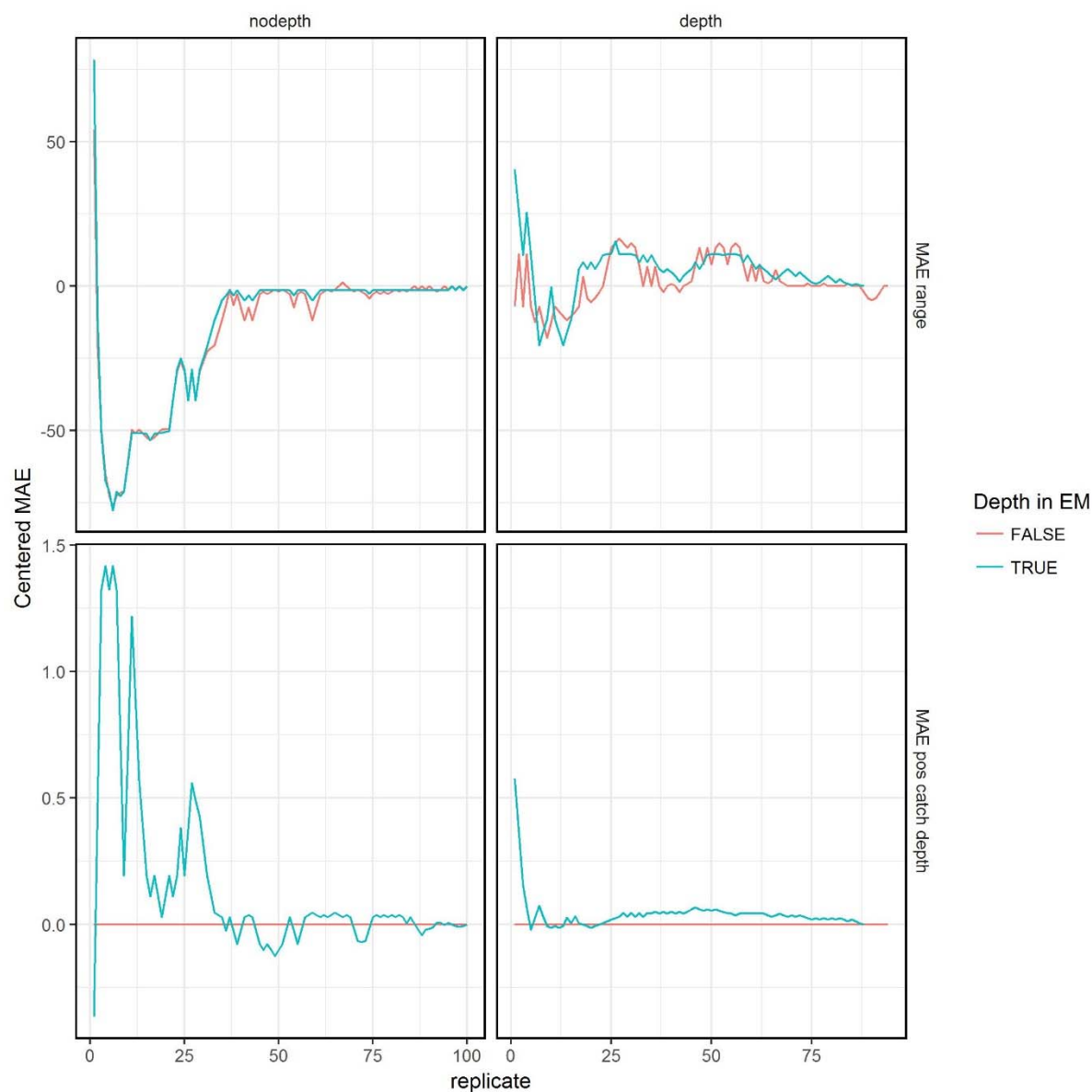
Figure C.1. Centered median absolute error (MAE) for two estimated parameters: range of the spatial and spatiotemporal fields for the encounter model (top row) and the linear depth term for positive catch rates (bottom row). Results are from simulations that did (right column) and did not (left column) include quadratic depth in the operating model. Colors are used to define if depth was included in the estimation method (orange does not and blue does).
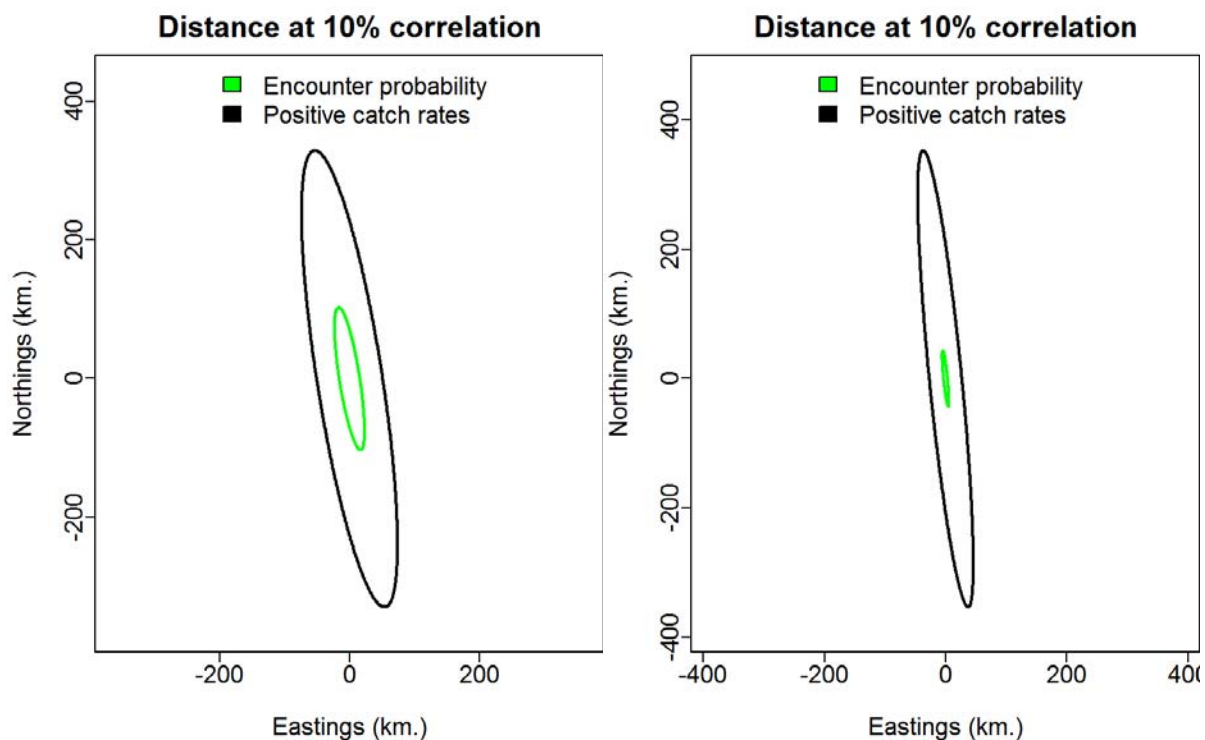
Figure C.2. Geometric anisotropy specifying the distances leading to 10% correlation in two-dimensional space for encounters (green) and catch-rates (black) as determined using Eigen-values of the anisotropy matrix (**H**). Estimates are from models fit to empirical data for darkblotched rockfish, where the left panel did not include depth and the right panel included quadratic depth. Units are in km for east-west (x-axis) and north-south (y-axis) distances. Vertically stretched ellipses indicate correlations stretch over larger distances going from north to south compared to east to west.
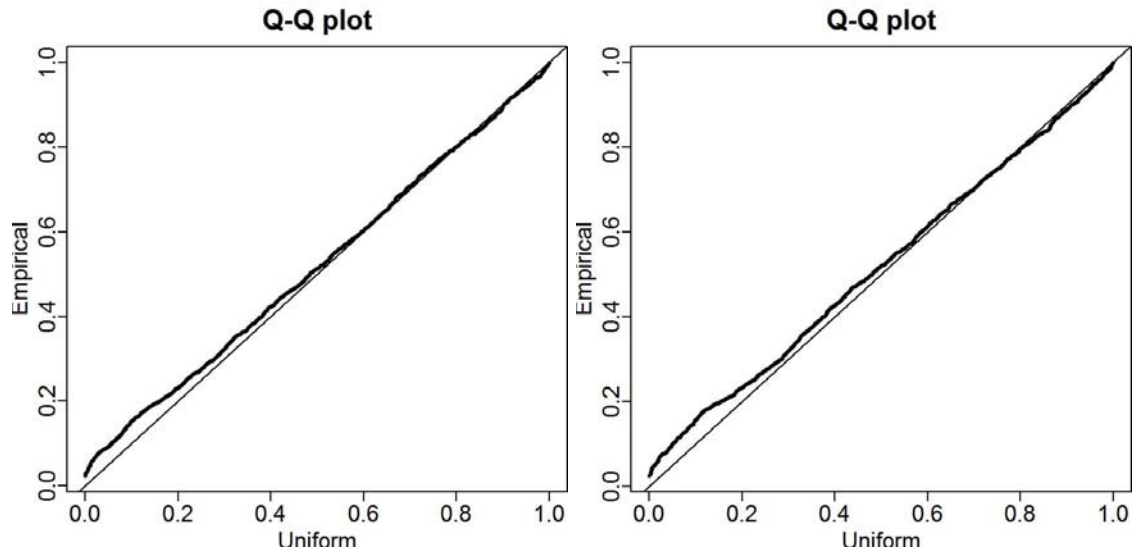
Figure C.3. Empirical versus uniform quantiles for fits of the empirical data to a model that did not include depth (left) and a model that included quadratic depth (right). Points should fall close to the 1:1 line.
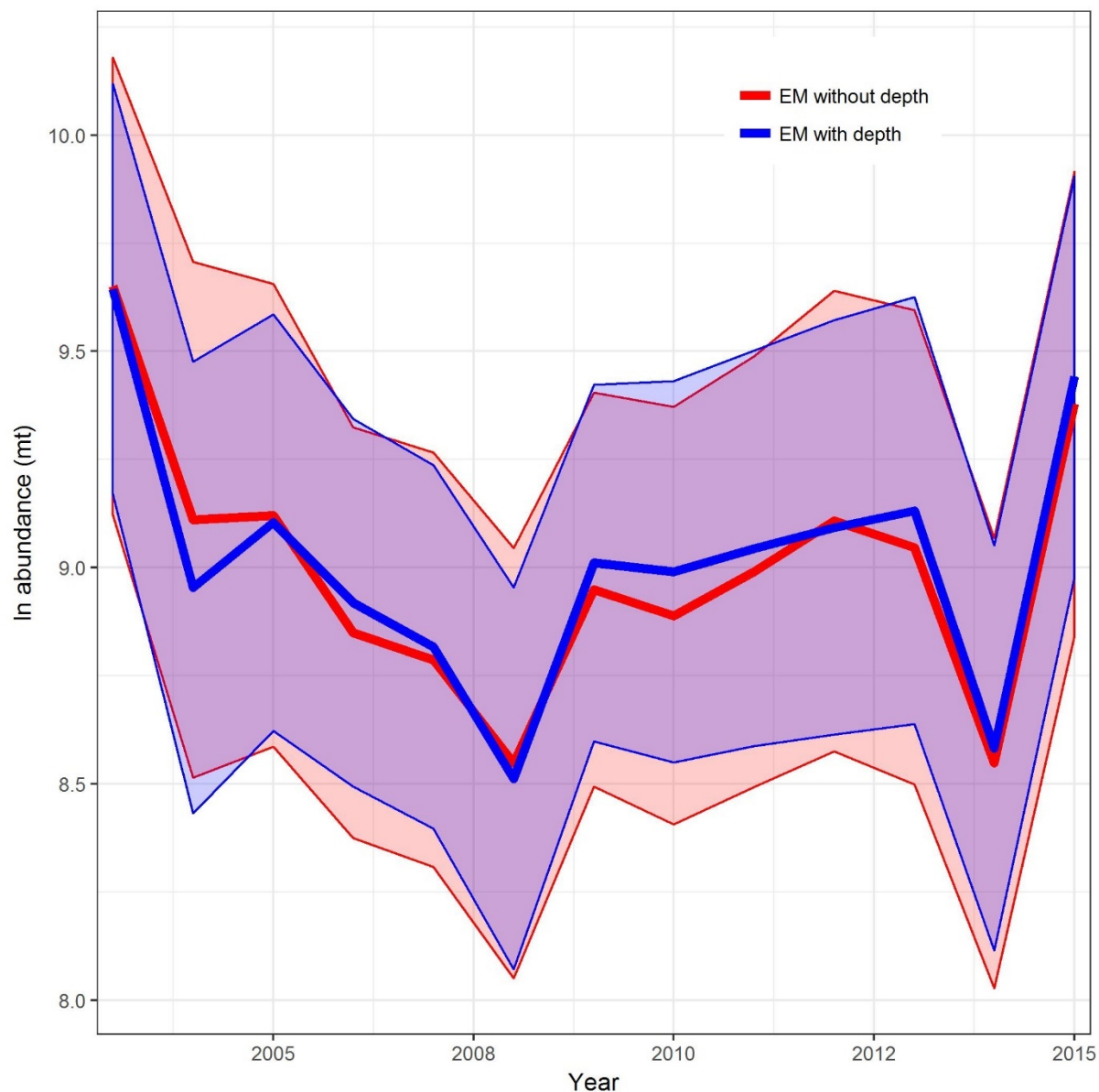
Figure C.4. Comparison of estimated relative indices of abundance (mt) in log space for darkblotched rockfish from 2003 to 2015. Data were fit to two estimation methods (EMs), one that included linear depth (blue) and one that did not include depth (red). Polygons show the 95% confidence intervals for each model, which are purple when they overlap.
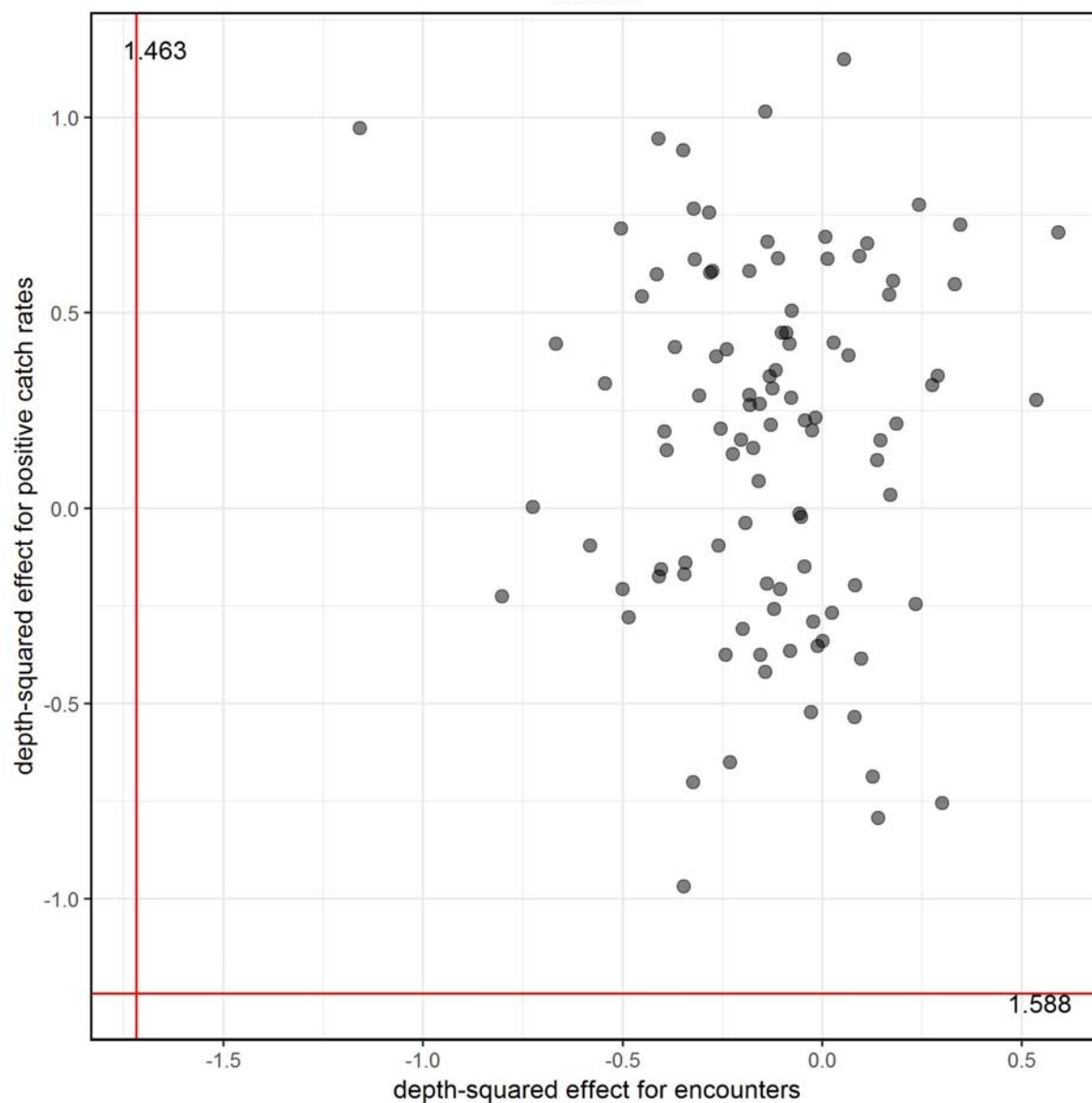
Figure C.5. Effect of depth squared as estimated by each component of the delta-model (encounters on x-axis and positive catch rate on y-axis) when the true process was governed by quadratic depth using 500 knots. Each point is a replicate of an estimation method fitted to simulated data that was governed by depth. The median absolute error (MAE; median|estimated − true|) for each parameter is printed near its respective axis. Unbiased results would be located on the red cross-hairs.
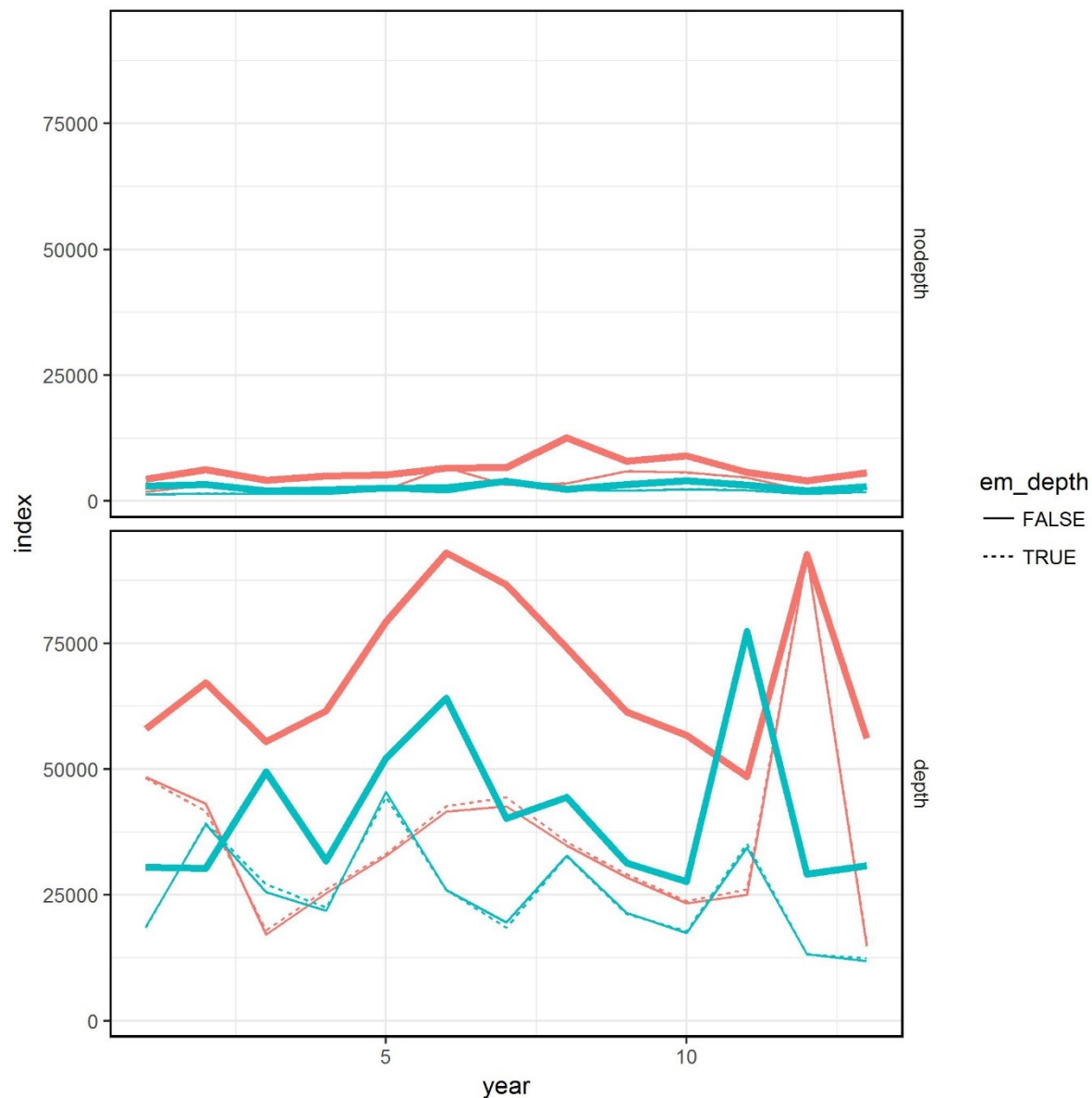
Figure C.6. Example simulated time series (thick lines). Depth was included in the operating model in the bottom row but not in the top row. Fits to the data (thin lines) are provided for two estimation methods. The thin, dashed line represents estimates from the estimation method that included linear depth and the solid line is from a misspecified estimation model that did not include depth. Colors are used to indicate unique replicates, where just two replicates are shown per panel.

Appendix D. SUPPLEMENTARY INFORMATION FOR CHAPTER 4

D.1   ESTIMATION METHODS

Recent literature on whether to transform the response variables that are proportions or use the beta distribution suggest the latter (Warton and Hui, 2011; Schmid *et al*., 2013; Herpigny and Gosselin, 2015). The binomial distribution is also an option when working with proportion data but requires information on the sample size. The beta distribution is a continuous distribution with finite support on [0, 1] and is governed by two shape parameters, μ and ϕ, $f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}$, where μ is the mean of **y**, $\phi > 0$ is the precision parameter and $\Gamma(\cdot)$ is the gamma function. The variance of **y** is given by $\mu(1-\mu)/(1+\phi)$, which is a scaled version of binomial variance, $\mu(1-\mu)$, allowing for more variation than would be expected by a binomial model (i.e., "overdispersion"). The beta distribution was chosen for these data as it appears to limit the number of assumptions that are violated (Figure D.1).

D.2   REFERENCES

Herpigny, B., and Gosselin, F. 2015. Analyzing plant cover class data quantitatively: customized zero-inflated beta distributions show promising results. Ecological Informatics, 26: 18-26.

Schmid, M., Wickler, F., Maloney, K. O., Mitchell, R., Fenske, N., and Mayr, A. 2013. Boosted Beta regression. PLoS ONE, 8: e61623.

Warton, D.I., and Hui, F.K.C. 2011. The arcsine is asinine: the analysis of proportions in ecology. Ecology, 92: 3-10.
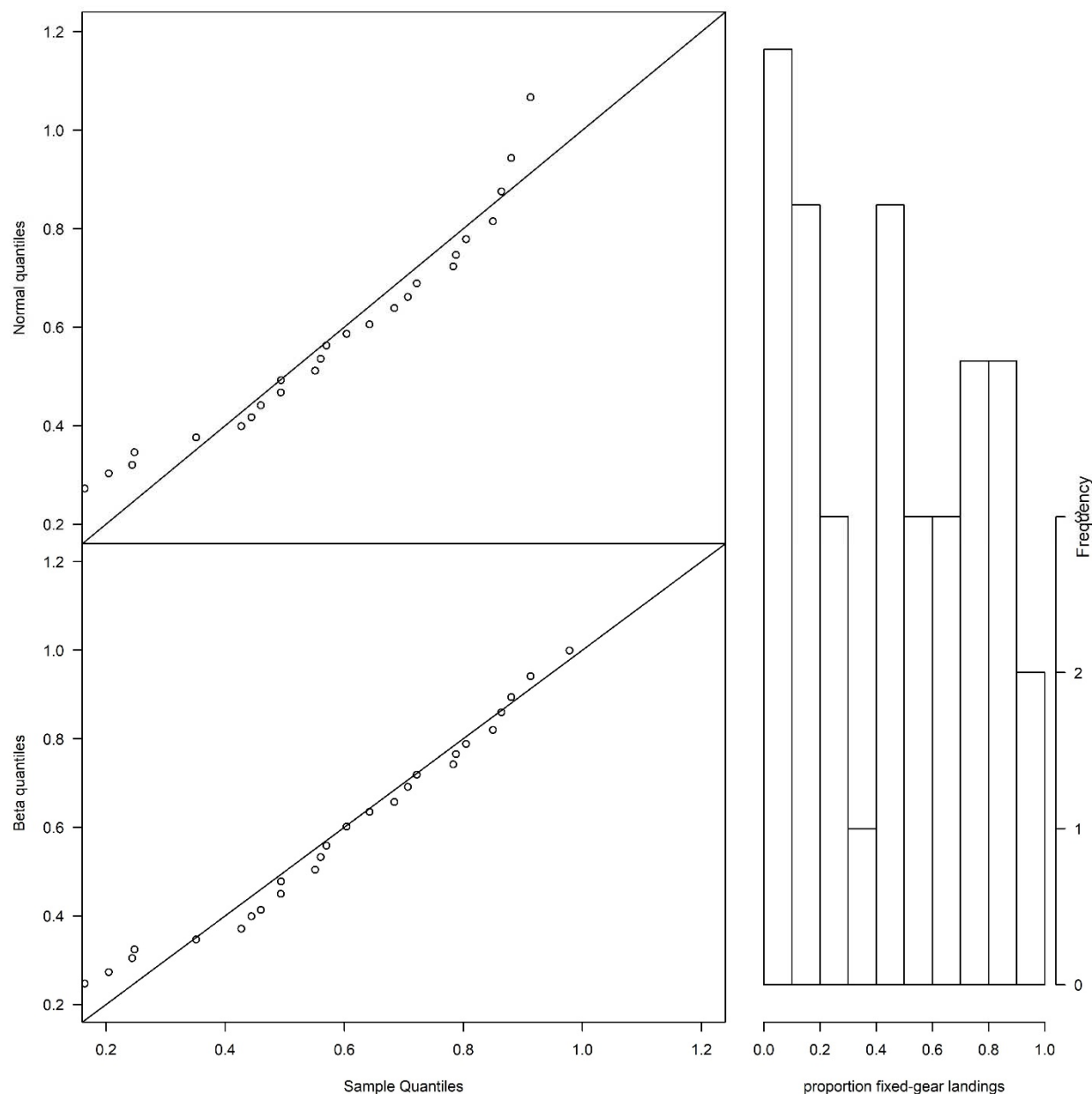
D.3 FIGURES

Figure D.1. Comparisons between sample and theoretical quantiles for a standard normal distribution (upper) and a beta distribution with mean of 0.083 and variance of 0.0035 (lower). The beta distribution provides a better fit to the data. The histogram displays the frequencies of the proportion of the yearly annual catch limit of sablefish assigned to the LE trawl sector landed in a port group using trawl gear within the limited entry (LE) groundfish trawl fishery.

VITA

Kelli Faye Johnson obtained her B.S. from the University of Puget Sound in Tacoma, Washington. She obtained a M.R.M. from the School of Resource and Environmental Management at Simon Fraser University in Burnaby, British Columbia, Canada. While at the University of Washington, she completed her Dissertation under the guidance of André Punt. Since May of 2017, she has worked at the Northwest Fisheries Science Center within the Fishery Resource Analysis and Monitoring Division as Research Fish Biologist.