University of California
Santa Barbara

Uncertainty analysis in fisheries science–an interdisciplinary approach

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Environmental Science & Management

by

Laura C. Urbisci

Committee in charge:

Professor Steve Gaines, Chair
Professor Wendy Meiring
Doctor Kevin Piner

December 2018

The Dissertation of Laura C. Urbisci is approved.

Professor Wendy Meiring

Doctor Kevin Piner

Professor Steve Gaines, Committee Chair

December 2018

Uncertainty analysis in fisheries science–an interdisciplinary approach

by

Laura C. Urbisci

I dedicate my dissertation to the countless cups of coffee, wine, and whiskey I consumed in the past 5.25 years. I could be nowhere with out you.

Acknowledgements

There is a seemingly countless number of people I want to acknowledge who have supported me throughout my time in graduate school.

- Family: Thank you mummy and fasha for giving birth to me. You did a great job.
- PhD Committee: Thank you for all of your time, advice, and for putting up with all of my questions–both the smart and the dumb ones.
- Steve Gaines: You were my Thanksgiving toast. I'm so thankful for you.
- Hunter Lenihan: Even though our paths diverged, thank you for all of the lessons you taught me.
- My roommates: My roommates, my psuedo family, and my "German friends". I love and thank you so much for everything you've done for me. House Dad (Spencer), House Child (Katrina), and House Goldfish (Mark), don't die without your House Mom. She's really going to miss you.
- Cubicle mate: Thank you Brian for all of the puns, for being an awesome study buddy through all of the PSTAT classes we took together, for being a patient sounding board to run ideas by, and for listening to me vent all my frustrations. Someday we'll win the lottery.
- Past and current lab mates: Thank you for all of your support the hugs and the listening. I'll always have chocolate for you.
- Bren buddies: Special shout out to Ian and Timbo. Thanks guys!
- Bren staff: Thank you Sage for all of your help my first year. Thank you Corlei, I know you no longer are at Bren, but you helped me so much when you were at Bren. You were the ultimate Bren Mom. Thank you Dee for being so helpful and a sweetheart. Thank you Kim for being the wonderful, unique you. Never change. Thank you Casey for being a friend in need. Thank you Kristine for being

so supportive. Thank you Satie for guiding me along the way. Thank you Onella and Yoda. I loved getting to know you both. Thank you Doris for answering all of my emails. Sorry to bombard you all the time. Thank you Dave for all of your career advice. Thank you Geoff, Brad, and Steve for all of your computer help. Thank you Beth for joining me at all of our pro fem events. Fun times. Thank you Aleah for helping me set up for my defense. Thank you Lisa for always taking the time to help me, no matter how busy you were. I really appreciate everything you've done for me.

- Bren faculty: Thank you Andrew Plantinga. I had a great experience working with you as a mentor on the MESM GP. Thank you Sarah Anderson for being super supportive.
- PSTAT faculty: Thank you Prof. Jammalamadaka for being the professor of my dreams. I will always remember your catch phrase, "Let's see what's cooking!" Thank you Prof. Wang for being such a great teacher. I really enjoyed your lectures. Thank you Prof. Hsu. I still remember the time you thought I was doing my PhD in the PSTAT Department. I was so flattered.
- PSTAT staff: Thank yooou Jamie. You're the best.
- NOAA/IATTC staff: Thank you Hoo Hoo and Carolina for always taking the time to help me.
- ERI staff: Thank you Erik Fields, Stéphane Maritorena, and Dave Siegel for all of your help with SeaWiFS.
- Ladies crew: Ladies, I love and appreciate you so much. I can't even put in words how much you mean to me. Timnit, Jessica, Phoebe, Julia, Lewam, and Alice. So much love going your way.
- Potluck and karaoke buddies: Mengya, Yuwei, Rungsheng, Ying, Yang, Jiajia, Zhitong, and Yuxiong–I've had such a great time hanging out with all of you. I've

enjoyed all of our time together as friends.

– Non-Bren grad school friends: Tara/Yuanbo, Zach, Ya, Megan, Anna, Sergio, and Ben–thank you for being such great friends and for all of the laughs. Cool, cool.
– My mentors: Matt Burgess and Cody Szuwalski–I'm wishing the best for you in all of your endeavors.
– The Chalet Crew: Jen, Carl, Myley, Matt, Molly, Jacquie, Joe, and Mike–I've never been a group person, but you changed my mind
– Graduate Scholars family: my lovely mentoring family: Timnit, Terence, Natasha, and Xochitl and the GSP faculty and staff: Carlos, Michele, and Miros. I'm always here for you even if we are far apart.
– Other UCSB folks: Lana Hale-Smith and Megan Unden - I love you ladies. Thank you for being so helpful and validating.
– Thank you Deborah and Hap for being amazing role models and mentors.
– Funding sources: Dr. Daniel Vapnek Fellowship and Award, NMFS-Sea Grant Fellowship in Population and Ecosystem Dynamics (NOAA Grant $\#NA14OAR4170211$, California Sea Grant College Program Project $\#E/PD-13$), the Bren School, and the PSTAT Department.
– To all the dogs I ever dog sat–never forget me. And the owners too. Thanks for letting me get my doggo snuggles.

Curriculum Vitæ

Laura C. Urbisci

EDUCATION

2018	Ph.D. in Environmental Science & Management, University of California, Santa Barbara.
2016	M.A. in Applied Statistics, University of California, Santa Barbara.
2012	BS in Environmental Science and Management *with honors* Emphasis in Ecology, Biodiversity and Conservation, Minor in Spanish University of California, Davis
2010	Education Abroad Program Universidad de Carlos III – Madrid, Spain

RELEVANT COURSEWORK

Probability & Statistics (3 part course)	Regression Analysis
Statistical Theory (2 part course)	Statistical Consulting
Advanced Statistical Methodology (3 part course)	Data Mining
Design and Analysis of Experiments	Time Series Analysis
Linear and Nonlinear Mixed Effects Modeling	Data Science
Bayesian Data Analysis	Machine Learning (audit)

RELEVANT STATISTICAL EXPERIENCE

Quantitative Consultant, Santa Barbara, CA

Bren School of Environmental Science & Management 9/16 – 3/17

- Assisted all graduate students, faculty, postdocs, and visiting researchers with their quantitative needs
- Advised the most appropriate statistical methods to use given the data set and research questions
- Explained how to code statistical models and interpret model results

Probability and Statistics Department, UCSB

Group Projects 4/16 – 6/18

- Worked on a multiple projects that utilized different data sets including: biological and political
- Select skills applied include: principal component analysis, categorical KNN, classification tree analysis (with pruning, bagging, and random forests), and Naïve Bayes
- Lead group by setting goals and determined course of action for a group of 4
- Presented ideas effectively and wrote a report which received positive feedback from instructor

Probability and Statistics Department, UCSB

Individual Projects 12/14 – 6/16

– Worked on a variety of projects to analyze results from various data sets including: medical, economic, and biological
– Overview of skills and select a few used: fitted a time series model (Seasonal ARIMA) to data and forecasted into the future to predict values, compared the efficacy of three weight-loss programs using linear mixed effects models, looked at the economic relationship between the 48 contiguous states using multivariate analysis methods and linear models
– Presented ideas effectively in project interview and wrote a report which received positive feedback

STATISTICAL LEADERSHIP EXPERIENCE

Teaching Assistant (TA), Santa Barbara, CA
University of California, Santa Barbara 3/17 – 12/18
– Gave multiple guest lectures to ∼150 students on linear regression, binomial proportion test, and chi-squared tests
– Managed computer labs and showed students how to program in R, the R GUI R Commander, SAS, and Excel
– Taught material ranging from basic statistical concepts to advanced statistical theory to students who came from a wide range of backgrounds

Statistics Tutor, Santa Barbara, CA
Probability and Statistics Department 9/15 – 12/18
– Aided undergraduate students with their quantitative coursework and taught R
– Helped students understand difficult concepts by explaining it to them in novel ways

Master's Group Project PhD Mentor, Santa Barbara, CA
Bren School of Environmental Science & Management 3/16 – 3/17
– Helped master student's set attainable goals, define research questions, and develop a feasible project timeline
– Reviewed and provided feedback on drafts of reports and presentations
– Recommended appropriate statistical analysis given data

Intern, La Jolla, CA
NOAA Southwest Fisheries Science Center 8/12 – 8/13
– Analyzed scientific data and presented findings at a professional meeting
– Trained lab assistants

PUBLICATIONS

Urbisci, L. C., Stohs, S. M., and Piner, K. P. 2017. From sunrise to sunset in the California drift gillnet fishery: An examination of the effects of time and area closures on the catch and catch rates of four key pelagic species: thresher shark (Alopias vulpinus), swordfish (Xiphias gladius), blue shark (Prionace glauca), and shortfin mako (Isurus oxyrinchus). Marine Fisheries Review. 78(3-4):1-12.

Ayres, A., Degolia, A., Fienup M., Kim J., Sainz, J., **Urbisci, L. C.**, Viana, D., Wesolowski, G., Plantinga, A. J., Tague, C. 2016. Social science/natural science perspectives on wildfire and climate change. Geography Compass. 10.2: 67-86.

Urbisci, L. C., Sippel, T., Teo, L. H., Piner, K. R., and Kohin, S. 2013 Size composition and spatial distribution of shortfin mako sharks by size and sex in U.S. West Coast fisheries. Submitted to ISC Shark Working Group Workshop July 6-11, 2013.

Urbisci, L. C., Runcie, R., Sippel, T., Piner, K., Dewar, H., and Kohin, S. 2012 Examining size-sex segregation among blue sharks (Prionace glauca) from the Eastern Pacific Ocean using drift gillnet fishery and satellite tagging data. Submitted to ISC Shark Working Group Workshop January 7-14, 2013.

Urbisci, L. C. 2011. Testing the unknown: the distribution, size and abundance of intertidal Haliotis rufescens (red abalone) and Haliotis cracherodii (black abalone) within Marine Protected Areas. (Unpublished student report. On file at the Cadet Hand Library, U.C. Davis Bodega Marine Laboratory).

PRESENTATIONS

Urbisci, L.C. 2018. Untangling uncertainty in food webs. Presented to Schmidt Family Foundation on March 9, 2018, Santa Barbara, CA.

Urbisci, L.C. 2017. Fishing through the food web leads to systematic overestimation of maximum sustainable yield. Presented at the NMFS-SG Annual Fellows Meeting on May 8-10, 2017, Beaufort, NC.

Urbisci, L.C., 2016. Developing an alternative estimate for virgin biomass using food web dynamics. Presented at the NMFS-SG Annual Fellows Meeting on June 28-30, 2016, Santa Cruz, CA.

Urbisci, L.C., 2016. Developing an alternative estimate for virgin biomass using food web dynamics. Presented at the Bren School PhD Symposium on February 19, 2016, Santa Barbara, CA.

Urbisci, L.C., 2015. Developing a new ecosystem-based management approach: using ecosystem model to calculate a better estimate of population scale for single-species models. Presented at the NMFS-SG Annual Fellows Meeting on June 9-11, 2015, Miami, FL.

Urbisci, L.C., Stohs, S. M., and Piner, K. P. 2014. From sunrise to sunset in the California drift gillnet fishery: An examination of the effects of time and area closures on the catch and catch rates of four key pelagic species: thresher shark (Alopias vulpinus), swordfish (Xiphias gladius), blue shark (Prionace glauca), and shortfin mako (Isurus oxyrinchus). Presented at the Highly Migratory Species Management Team Meeting on January 22, 2014, La Jolla, CA.

Urbisci, L.C. Runcie, R., Sippel, T., Piner, K., Dewar, H., and Kohin, S. 2012 Examining size-sex segregation among blue sharks (Prionace glauca) from the Eastern Pacific Ocean using drift gillnet fishery and satellite tagging data. Presented at the ISC Shark Working Group Workshop January 10, 2013.

Urbisci, L.C. 2011. Testing the unknown: the distribution, size and abundance of intertidal Haliotis rufescens (red abalone) and Haliotis cracherodii (black abalone) within

Marine Protected Areas. Presented at the Sequence One and Two Student Symposium 2011, Bodega Bay, CA.

Abstract

Uncertainty analysis in fisheries science–an interdisciplinary approach

by

Laura C. Urbisci

My dissertation is an interdisciplinary approach that combines fisheries science, ecological theory, and applied statistics. My first chapter is a meta-analysis on transfer efficiency that describes and quantifies the variation in transfer efficiency. My second chapter assesses uncertainty in food web models by creating multiple Monte Carlo simulations to test various ecological assumptions about net primary production and transfer efficiency. My final chapter is a comparative analysis of two Bayesian models: a classic Bayesian surplus production model and a Bayesian surplus production model that incorporates ecological information. This chapter examines if the inclusion of ecological information informs and alters fisheries assessment models, with a focus on data-limited fisheries. Ultimately my work bridges the gap between applied statistics and ecological theory and encourages the use of uncertainty analysis to make more robust predictions in food web models.

Contents

Chapter 1

Introduction

Fisheries modeling take the complexity of a single heterogeneous stock and simplify these diverse dynamics into a cohesive model. These stock assessment models look at data and attempt to predict how these attributes will respond to fishing over time. Depending on the available data and level of complexity desired, we can model open ocean ecosystems using several different approaches. Most models focus on an individual stock and fall under the category of a single-species model. The simplest single-species models look only at abundance and are referred to as biomass dynamic or production models. However if sufficient data is unavailable to model the individual species dynamics, a common workaround is to cluster the targeted species in a fishery into a complex and model their dynamics in one production model. Doing this though, comes with a set of assumptions that when invalidated can have drastic consequences on the sustainability of some or all of the targeted species.

A multitude of ecosystem-based models have been developed within the past three decades to address the need to incorporate ecosystem-based science into fisheries management. These models help inform decision-makers about the effects of fishing mortality and

the indirect trophic implications of fishing in changing ecological environments. There are various types of ecosystem-based models. All of these classes of models aim to simulate the environment by including species interactions and environmental fluctuations.

Instead of attempting to explain all the ecological processes in one model, a new approach that is outlined in this dissertation is to move away from focusing on small-scale details and look at the ecosystem in a broader context. We can combine ecosystem-knowledge to improve upon single-species models. For instance by applying ecological theory such as food web dynamics, we can develop a more feasible approach to estimate the unfished biomass and carrying capacity. By taking the amount of net primary production that enters into the system, we can use the principle behind energy transfer in food webs to approximate the amount of biomass at each trophic level. By taking a bottom-up approach, we can ensure that our estimates of unfished biomass are feasible, because we account for how much energy goes into the system. We can additionally include sensitivity analysis in our model to account for the natural variation in the environment.

Chapter 2

Tangled is the web we weave

2.1 Introduction

One of the crucial, and at times, most puzzling concepts in food web dynamics research is the transfer efficiency–the movement of production between trophic levels. This paper addresses two aspects of transfer efficiency: first, we seek to provide clarity and untangle the web of confusion surrounding the conceptualization of transfer efficiency. Second, we address the often-cited claim that transfer efficiency is a constant 10%. We analyze extant research to show that transfer efficiency varies substantially across systems, trophic levels, and taxa.

2.1.1 Origin and Conceptualization of Transfer Efficiency

The definition of transfer efficiency has been somewhat muddled since its inception. We refer to transfer efficiency as the fraction of production passing from one trophic level to the next (Slobodkin 1959). At times it has also been referred to as trophic (transfer) efficiency (Chapman and Reiss 1998). This term is often confused with other non-equivalent efficiencies such as ecological efficiency, assimilation efficiency, and con-

sumption efficiency (Iverson 1990, Hairston 1993). However, each of these efficiencies addresses distinct ecological questions and thus require different data for their calculations. Slobodkin (1959) theorized a food chain efficiency metric and defined it as the ratio of the number of organisms removed from the targeted population to the food consumed by the targeted population (Slobodkin 1960, 1962). Removal includes both natural mortality and human harvesting. He subsequently renamed the concept "ecological efficiency" in his 1962 and 1972 papers (Slobodkin 1962, 1972). The energy budget requires a balance between inputs and outputs. When energy is ingested, some of that energy is lost to respiration and excretion. Then, the remaining energy that is assimilated is divided amongst basal maintenance, growth, and reproduction. Assimilation efficiency is defined as the percentage of energy ingested at trophic level n that is assimilated at trophic level n (Hairston 1993). The consumption efficiency measures the number of organisms from the prey population that is consumed by its predators and is defined as the percentage of net production at trophic level n that is consumed by trophic level $n+1$ (Hairston 1993). In an attempt to make the differences clearer, we provide a simple cartoon of a food web (Figure 2.1) that visualizes the definitions of four of the commonly used efficiencies.

Availability of data differs between ecosystems. It is difficult in aquatic systems, especially marine systems, to gather enough data on every species in order to calculate the assimilation and consumption efficiencies. Terrestrial studies, on the other hand, can collect detailed population data much easier. Thus, terrestrial studies do not need to rely as much on inferential techniques, like the transfer efficiency, and have the ability to calculate species-specific metrics, such as the assimilation and consumption efficiency. To clarify, the assimilation and consumption efficiencies can also be calculated at the trophic-level in addition to the species-level.

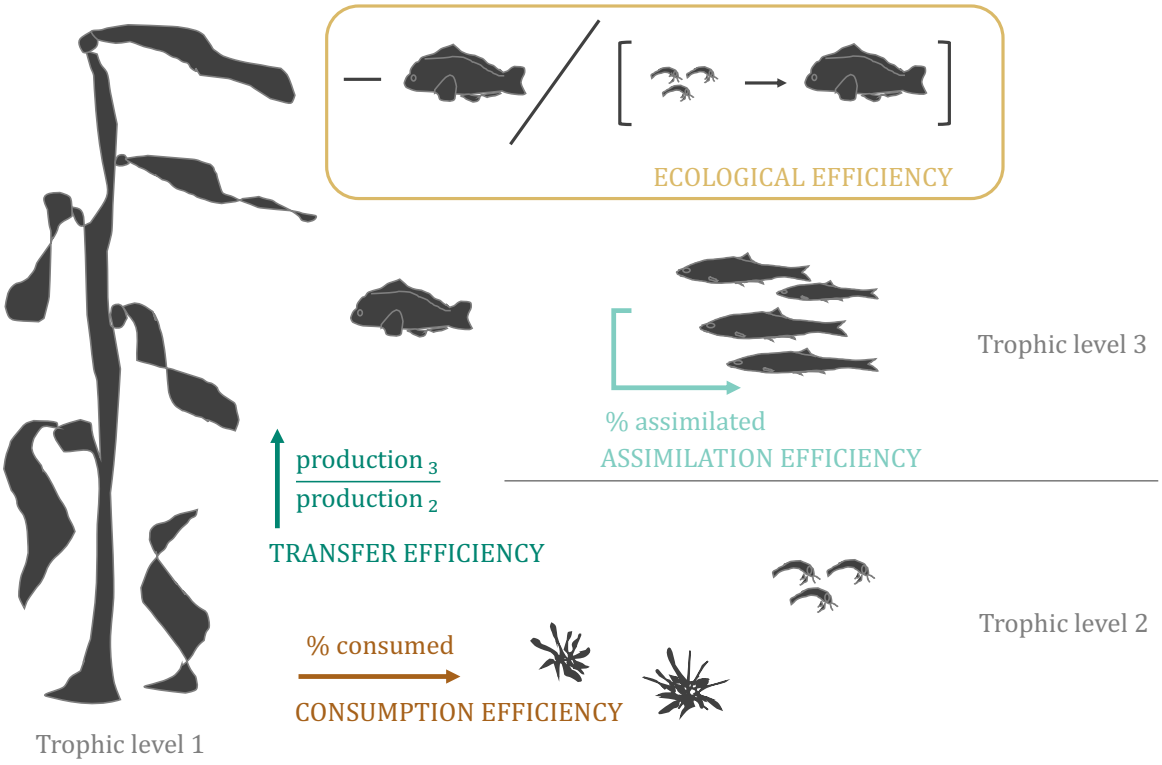

Figure 2.1: Cartoon of a food web that visualizes different efficiencies. The consumption efficiency is in brown, the transfer efficiency is in teal, the assimilation efficiency is in light blue, and the ecological efficiency (food chain efficiency) is in tan. In the consumption and ecological efficiency, the head of the arrow indicates the direction of consumption, where the species at the arrow head represent the species consuming the species at the arrow's origin. The diagram of the ecological efficiency includes a negative sign, division sign, and parentheses. Plot created using Microsoft office 2013.

The confusion around the definition is not the only complication with transfer efficiency– the values themselves have been disputed over the years and remain a point of contention. It is surprising that some scholars treat transfer efficiency as a fixed constant (i.e., 10%) for all trophic levels in light of the fact that other scholars have found that physiological, and potentially behavioral, characteristics influence transfer efficiency (May 1983, Pauly and Christensen 1995, Ware 2000, Cury et al. 2005, Libralato et al. 2008, Chassot et al. 2010, Trebilco et al. 2013, Watson et al. 2014).

2.1.2 Physiological and Behavioral Characteristics of Transfer Efficiency in Freshwater and Marine Ecosystems

Multiple factors have been shown to affect transfer efficiency in both freshwater and marine ecosystems. In freshwater systems, the sources of variability in transfer efficiencies include the body of water, season, trophic level, and species composition (Lindeman 1942, Gaedke and Straile 1994, Rybarczyk and Elkaım 2003, Karlsson et al. 2007). In marine systems, transfer efficiency varies by ecological system, geographic location, trophic level, metabolic strategy, and species composition (May 1983, Persson et al. 2007, Libralato et al. 2008, Barnes et al. 2010).

Ecological System Within Freshwater and Marine Ecosystems

Transfer efficiency has been found to be specific to the geographical region. Multiple marine studies found distinct transfer efficiencies between upwelling, temperate and tropical ecosystems (i.e., 5% upwelling, 10% temperate, and 14% tropical) (Libralato et al. 2008, Coll et al. 2008, Chassot et al. 2010). Even within a single ecosystem, Baird et al. (2004) found that each community within an intertidal ecosystem had unique transfer efficiency values. Distinct transfer efficiency values have also been found to occur not only between lakes and within trophic levels in freshwater ecosystems (Lindeman 1942), but also in bays and estuaries as well (Rybarczyk and Elkaım 2003).

Additionally, research has found that the amount of sunlight a region receives affects transfer efficiency. San Martin et al. (2006a) suggest that transfer efficiency from phytoplankton to zooplankton in marine ecosystems decreases as latitude increases due to the decrease in sunlight. Gaedke and Straile (1994) found seasonal variation in transfer efficiency between the first and second trophic level in lakes, with transfer efficiency rising

in the summer and fall and decreasing in the winter and spring. The seasonal variation in transfer efficiency can be attributed to the decrease in phytoplankton abundance in winter due to limited sunlight. As daylight increases in early spring, there is a gradual increase in phytoplankton blooms–culminating in the maximum phytoplankton production in summer (i.e., peak hours of sunlight). As the days become shorter in fall and the hours of sunlight decreases, there is a decrease in the amount of phytoplankton. There is a time lag corresponding to the change in sunlight in the spring and fall seasons. Therefore, the amount of sunlight indirectly influences transfer efficiency between the first and second trophic level by directly impacting the phytoplankton abundance.

Trophic Level

Size spectrum studies report that transfer efficiency decreases with body size, and by association, trophic level (Barnes et al. 2010). Therefore, the size ratio of prey to predators (e.g., phytoplankton to zooplankton) impacts transfer efficiency and trophic structure (Havens 1998, García-Comas et al. 2016).

Metabolic Strategy

Furthermore, May (1983) found ectotherms are more efficient than endotherms in transferring energy from trophic level n to trophic level $n + 1$, with energy transfer efficiencies around 20-50% for invertebrate ectotherms, around 10% for vertebrate ectotherms and less than 2% for endotherms. This discrepancy in transfer efficiency is due to the metabolic efficiency: ectotherms rely on environmental heat sources and therefore have a lower metabolic cost in comparison to endotherms. Much of the metabolic energy in endotherms goes to the production of heat. Therefore, transfer of energy in the higher trophic levels where endotherms are prominent is less than the lower trophic levels were ectotherms make up more of the composition in marine ecosystems (McGarvey et al.

2018).

Species Composition

Consuming nutritionally imbalanced food has been shown to lead to large respiratory losses, which negatively affect transfer efficiency (Persson et al. 2007). Karlsson et al. (2007) and von Elert et al. (2003) found that the species composition of prey, in particular different species of zooplankton crustaceans and the absence of long-chain polyunsaturated fatty acid in cyanobacteria, influence transfer efficiency. In addition, the presence of jellyfish blooms has been found to reduce the transfer of energy to higher trophic levels (Condon et al. 2011).

Trussell et al. (2006) and Schmitz et al. (2008) found that the risk of predation modifies prey conversion efficiencies and biomass production, which could therefore influence trophic structure and energy transfer. While these results refer specifically to assimilation and consumption efficiencies, it is plausible that this behavior influences transfer efficiencies as well. While the specific factors previously discussed influence transfer efficiency individually, these components interact in the natural environment. Because of this interaction, researchers must consider the impacts of the synergistic effects of these factors on the variability of the transfer efficiency and in turn how to account for them in the modeling process.

2.1.3 The 10% Transfer Efficiency

Although the studies above highlight that a number of factors can greatly affect trophic efficiencies, we still see broad use of the assumption of a constant value of 10%. To explore how (un)reasonable this assumption might be in different contexts, we synthesize the pattern of variation that has been observed in empirical studies that measured transfer

efficiencies. Our goal is to provide guidance for what is reasonable to assume and what is necessary to measure.

It is unclear where the 10% transfer efficiency assumption came from. Looking back at the historical records, we find a "tangled web" of misattributions and a general lack of empirical evidence. Semper (1881) might have come up with the theory that there is a 10% transfer between trophic levels, but he lacked empirical evidence to back this claim (McIntosh 1986). Lindeman (1942) developed more general theory by looking at energy flow diagrams and mentioned a progressive efficiency which is currently known as transfer efficiency. However, no explicit mention of a 10% value shows up in this work even though he is often credited for it (i.e., Lindeman's law of trophic transfer efficiency– Chapman and Reiss 1998). Slobodkin (1959, 1972) stated that "the values mentioned by Lindeman, as well as other values presented by other field workers, for ecological efficiency tended to cluster around 10%." Yet, Lindeman never explicitly discusses the ecological efficiency. He talked about the progressive efficiency, which as mentioned previously is a different concept. Regardless, Slobodkin and his students used laboratory experiments to formalize the hypothesis that there was an approximately 10% transfer between trophic levels (Slobodkin 1959). He referred to this as the food chain efficiency, which was later renamed to the ecological efficiency. However in a later study, Slobodkin (1972) found empirical and theoretical objections to the 10% ecological efficiency and rejected the theory. According to McIntosh (1986), "Nevertheless, May (1967b) in pursuit of the 'perfect crystals' of ecology, included Slobodkin's 1961 hypothesis in a series of community properties he described as 'constant and predictable'." In a more recent edition of May's Theoretical Ecology textbook, however, the authors reach a very different conclusion:

> One such [generalization] in the early 1960s suggested that the food-chain efficiency for transfer of energy from one trophic level to the next was generally

> around 10%. Subsequent studies showed that such food-chain efficiencies can vary over two or more orders of magnitude, from less than 0.1% to significantly more than 10%. Some evidence suggests such efficiencies may, other things being equal, be higher for carnivores and detritus feeders than for herbivores, possibly because biochemical conversion efficiencies are higher for animals eating plants. (May and McLean 2007)

In Figure 2.2, we present a flowchart showing the muddled origin of this concept. Given the unclear origin and application of the 10% transfer efficiency assumption, this assumption warrants further analysis, which is the focus of this current study. In the following section, we synthesize studies that provide empirical estimates of transfer efficiencies.

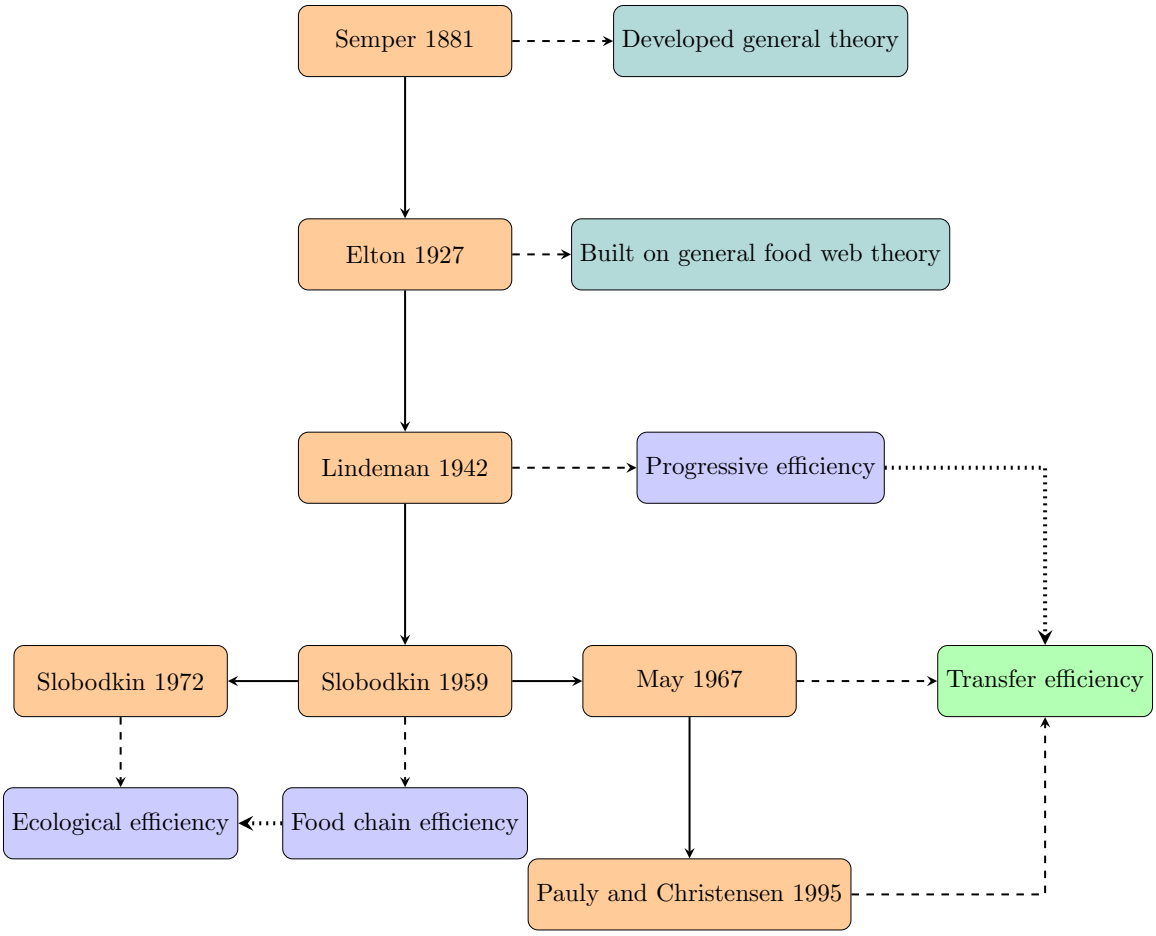

Figure 2.2: Flow chart of the origin of the 10% transfer efficiency theory (green icon). A subset of key journal articles are highlighted in the orange icons. The dashed lines point to the efficiency mention in a particular article. The dotted lines represent a change in the name of an efficiency. Teal icons denote that the article discussed general theory, while periwinkle blue icons represent the discussion of a type of efficiency. The arrowhead attached to the solid line denotes the downstream flow direction of citations. Plot created using LaTeX v.2.9.6211 (Lamport 1994) package tikz v.3.0.1a (Tantau 2015).

2.2 Methods

To explore the empirical distributions of transfer efficiencies, we collected articles that mentioned both food web and transfer efficiency. We then selected from these studies

those that included relevant data, whether model-based or empirical. While we initially hoped to include terrestrial as well as freshwater and marine studies, we found nearly all of the terrestrial studies were on the consumption and assimilation efficiency, not the transfer efficiency. Therefore, we broke our analysis into two sections: freshwater and marine and ignored terrestrial. We primarily applied exploratory data analysis techniques such as summary statistics to distinguish patterns between the systems.

If the sample size was sufficient large, we also used decision tree analysis (i.e., regression trees with pruning, bagging, and random forests) and Monte Carlo simulations (See Table 2.1 and 2.2). Decision tree analysis was employed to determine which factors had the largest impact on transfer efficiency. Using an approach similar to (Libralato et al. 2008), we clustered the marine ecosystems into the following regions: temperate shelves and seas, tropical shelves and seas, lagoons, upwelling ecosystems, and open oceans. Although we also clustered the freshwater ecosystems into lakes, springs, and ponds, the sample size of the freshwater transfer efficiency data was too small to run regression tree analysis. We used regression trees with pruning, bagging, and random forests on the marine transfer efficiency data set and used relative importance plots to determine which factors accounted for the largest sources of variation and were most useful in predicting transfer efficiency. In the discussion, we used Monte Carlo simulations to aid in the conversation.

2.2.1 Freshwater

Many of the preliminary studies on transfer efficiency occurred in freshwater systems. All of the early studies were empirical (i.e., data used to calculate the transfer efficiency were collected either through laboratory experiments or from the field), but over time studies shifted to being increasingly model-based (i.e., data used to calculate the transfer

efficiency were generated as the product of computer models). We found a total of 11 systems with transfer efficiency data (Table 2.1). Only the empirical studies reported transfer efficiency values for multiple trophic levels. The model-based studies reported the system-wide average. Most of the transfer efficiency data is empirically based.

The distributions of the freshwater transfer efficiencies are given in Figure 2.3. The empirical observations are skewed-right, while the model-based observations appear bi-modal (albeit with a small sample size–$n = 4$). Combining the empirical and model-based estimates, the collective freshwater transfer efficiencies ($n = 19$) range from 0.1% to 22.3% with a median of 8.4%. When we calculate the average transfer (progressive) efficiency values provided in Lindeman (1942), we found that the average actually is 9%. If we consider just transfer efficiencies between phytoplankton (trophic level 1) and zooplankton (trophic level 2), we found the median transfer efficiency to be 12.2%. Unfortunately, the sample size in each group (i.e., trophic level and geographical region) is insufficiently large to draw any strong conclusions with relative certainty about which factors are the biggest sources of variation.

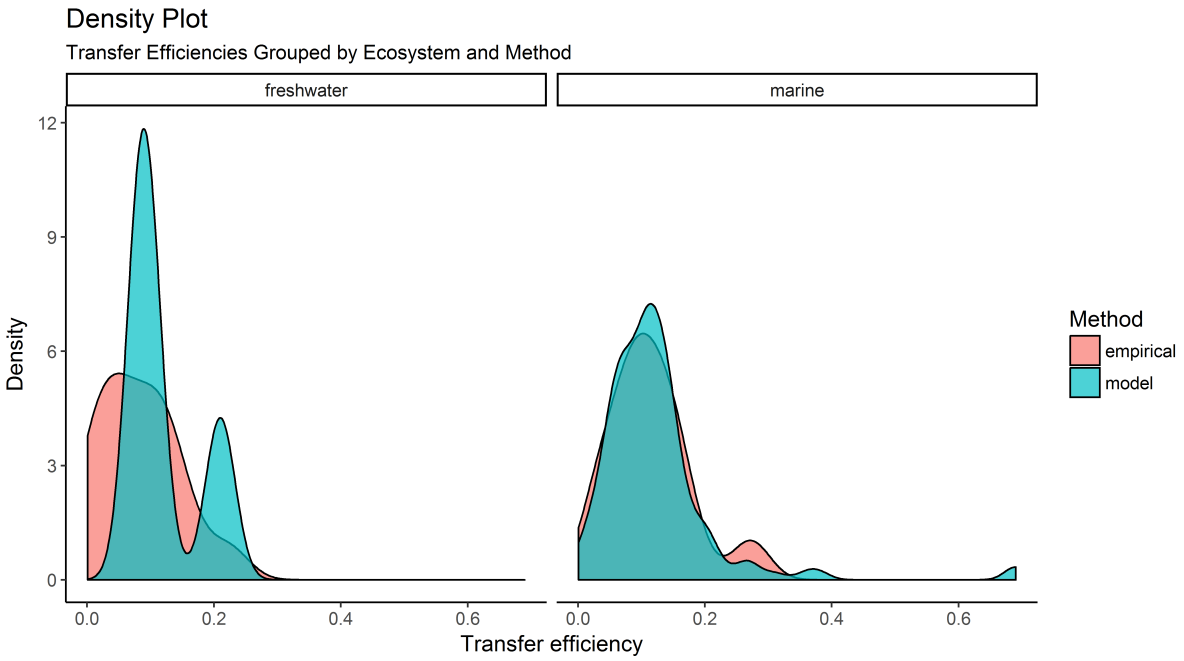

Figure 2.3: Density plot of transfer efficiencies where transfer efficiencies are grouped by ecosystem (i.e., freshwater vs. marine) and method (i.e., empirical and model-based). There are a total of 15 transfer efficiency observations gathered from freshwater empirical studies, 4 observations from freshwater model-based studies, 13 from marine empirical studies, and 134 from marine model-based studies. Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

Articles	Region	Clustered Region	Method	Trophic Level	Transfer Efficiency
Chea et al. 2016	Tonle Sap Great Lake	lakes	model	average	8.3
Gaedke 1993	Lake Constance	lakes	model	average	21
Lindeman 1942	Cedar Bog Lake	lakes	empirical	producers	0.1
Lindeman 1942	Cedar Bog Lake	lakes	empirical	primary consumers	13.3
Lindeman 1942	Cedar Bog Lake	lakes	empirical	secondary consumers	22.3

Lindeman 1942	Lake Mendota	lakes	empirical	producers	0.4
Lindeman 1942	Lake Mendota	lakes	empirical	primary consumers	8.7
Lindeman 1942	Lake Mendota	lakes	empirical	secondary consumers	5.5
Lindeman 1942	Lake Mendota	lakes	empirical	tertiary consumers	13
Odum 1959	Silver Springs	springs	empirical	producers	1.2
Odum 1959	Silver Springs	springs	empirical	primary consumers	16
Odum 1959	Silver Springs	springs	empirical	secondary consumers	11
Odum 1959	Silver Springs	springs	empirical	tertiary consumers	5
Rand and Stewart 1998	Lake Michigan	lakes	empirical	tertiary consumers	3.2
Rand and Stewart 1998	Lake Ontario	lakes	empirical	primary consumers	11.1
Rand and Stewart 1998	Lake Ontario	lakes	empirical	secondary consumers	8.3
Rand and Stewart 1998	Lake Ontario	lakes	empirical	tertiary consumers	4.6
Villaneuva et al. 2008	Lake Kivu	lakes	model	average	8.4
Villaneuva et al. 2006	Lake Nokoue	lakes	model	average	10.3

Table 2.1: Freshwater transfer efficiency data

2.2.2 Marine

Marine studies on transfer efficiency did not commence until decades after the start of freshwater studies. The popularity of marine transfer efficiency research has increased rapidly in the past 20 years and has overall now exceeded the number of freshwater studies. A total of 115 sites have transfer efficiency data (Table 2.2). In contrast to the freshwater studies, most marine transfer efficiency data ($n = 134$) come from model-based studies rather than empirical experiments ($n = 13$). Most marine studies report the average value for an entire system ($n = 94$). In studies that focused on individual transfer efficiencies between specific trophic levels, most focused on the transfer efficiency between phytoplankton (trophic level 1) and zooplankton (trophic level 2).

The empirical and model-based transfer efficiency data both form skewed-right distributions with large amounts of dispersion around the 10% value (Figure 2.3). For model-based observations, there are a few outlying points from a study on bays and estuaries that skew the distribution (Figure 2.3). The combined transfer efficiency data ranges from 0.2% to 69% with a median of 10.6%, while the range constricts with a minimum of 3.12% to a maximum of 27.2% for just the marine empirical studies (Figure 2.3).

To explore the potential drivers of variation in transfer efficiencies, we calculated importance plots from the decision tree analysis. When interpreting importance plots, the larger the score, the more influential the variable. A number close to zero indicates the variables is not important and could be dropped. When determining the importance

of a variable, the mean decrease in accuracy (i.e., mean square error, MSE) or the mean decrease in node impurity are used to measure how well the trees split the data. Thus, the relative importance plots from the decision tree analysis indicate that trophic level had the greatest influence on transfer efficiency, followed by the clustered region (i.e., temperate shelves and seas, tropical shelves and seas, lagoons, upwelling ecosystems, and open oceans) (Figure 2.4). The method employed (i.e., empirical or model-based) did not appear to be a useful predictor.

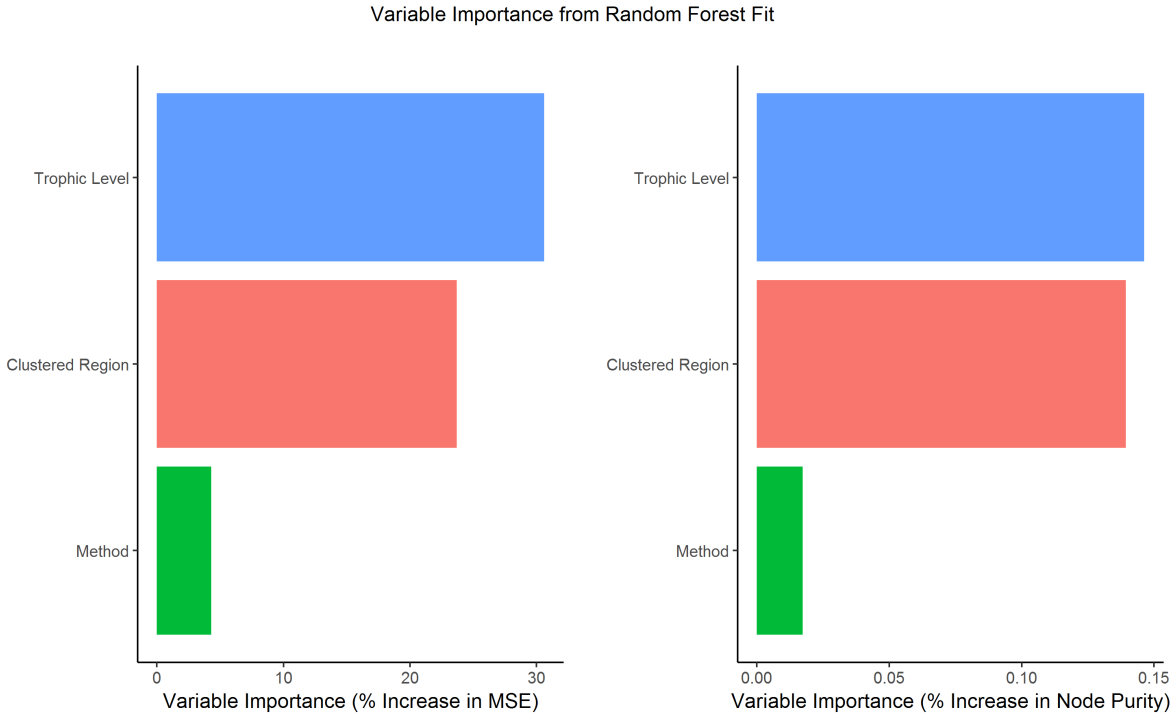

Figure 2.4: Relative importance plots for random forest fit on marine transfer efficiency observations. In panel a), the predictors are ordered by percent increase in Mean Square Error (MSE). In panel b), the predictors are ordered by increase in node purity. Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

We combined results across both marine and freshwater systems to examine differences among trophic levels (Figure 5). We combined data for all the systems to increase

the sample size. Whether the different densities are due to the sensitivity to small samples sizes or systematic differences is unclear. Nonetheless, our results support the hypothesis that transfer efficiency decreases as trophic level increases (see Garcia et al. 2012). This in turn supports the results from May (1983) that in the marine environment ectotherms (invertebrates then vertebrates), which dominate the lower trophic levels, are more efficient than endotherms, which are more prominent at higher trophic levels (Figure 2.5).

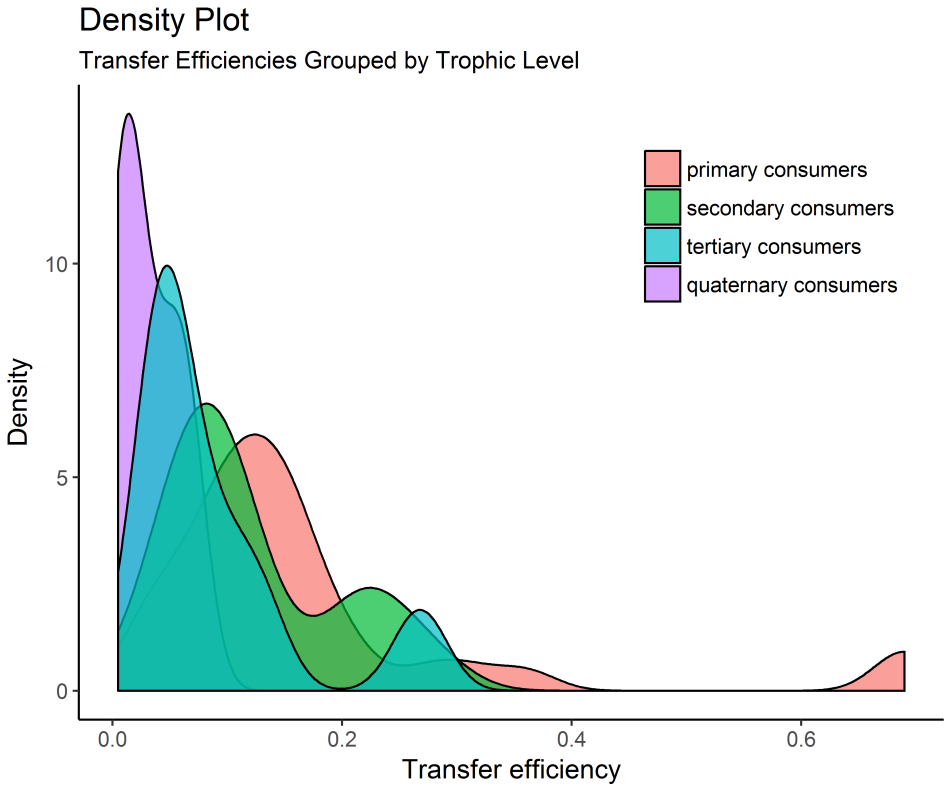

Figure 2.5: Density plot of transfer efficiencies grouped by trophic level. The data in this figure includes both ecosystem (i.e., freshwater and marine) and empirical and model-based transfer efficiency data. Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

Articles	Region	Clustered Region	Method	Trophic Level	Transfer Efficiency
Akoglu et al. 2014	Black Sea	temperate seas	model	primary consumers	3
Akoglu et al. 2014	Black Sea	temperate seas	model	secondary consumers	3.8
Akoglu et al. 2014	Black Sea	temperate seas	model	tertiary consumers	7.4
Akoglu et al. 2014	Black Sea	temperate seas	model	quaternary consumers	0.5
Anjusha et al. 2013	Gulf of Mannar	tropical seas	empirical	primary consumers	13
Anjusha et al. 2013	Gulf of Mannar	tropical seas	empirical	primary consumers	12
Anjusha et al. 2013	Gulf of Mannar	tropical seas	empirical	primary consumers	14.6
Anjusha et al. 2013	Gulf of Mannar	tropical seas	empirical	primary consumers	9.1
Anjusha et al. 2013	Gulf of Mannar	tropical seas	empirical	primary consumers	6.8
Anjusha et al. 2013	Palk Bay	tropical seas	empirical	primary consumers	27.2
Anjusha et al. 2013	Palk Bay	tropical seas	empirical	primary consumers	17.6
Anjusha et al. 2013	Palk Bay	tropical seas	empirical	primary consumers	9.39

Articles	Region	Clustered Region	Method	Trophic Level	Transfer Efficiency
Anjusha et al. 2013	Palk Bay	tropical seas	empirical	primary consumers	7.19
Anjusha et al. 2013	Palk Bay	tropical seas	empirical	primary consumers	3.12
Anjusha et al. 2013	Palk Bay	tropical seas	empirical	primary consumers	3.23
Baird et al. 2004	Sylt-Romo Bight	temperate seas	model	average	2.61
Baird et al. 2007	Sylt-Romo Bight: arenicola flats	temperate seas	model	average	3.47
Baird et al. 2007	Sylt-Romo Bight: dense zostera noltii beds	temperate seas	model	average	5.58
Baird et al. 2007	Sylt-Romo Bight: mud flats	temperate seas	model	average	6.13
Baird et al. 2007	Sylt-Romo Bight: muddy sand flats	temperate seas	model	average	7.31
Baird et al. 2007	Sylt-Romo Bight: mussel beds	temperate seas	model	average	14.92
Baird et al. 2007	Sylt-Romo Bight: pelagic domain	temperate seas	model	average	1
Baird et al. 2007	Sylt-Romo Bight: sandy beaches	temperate seas	model	average	6.5
Baird et al. 2007	Sylt-Romo Bight: sandy shoals	temperate seas	model	average	3.3

Articles	Region	Clustered Region	Method	Trophic Level	Transfer Efficiency
Baird et al. 2007	Sylt-Romo Bight: sparse zostera noltii beds	temperate seas	model	average	5.06
Barnes et al. 2010	summary of 21 locations	-	model	average small sizes	13.8
Barnes et al. 2010	summary of 21 locations	-	model	average large sizes	5.8
Baumann 1995	general claim	-	empirical	average	15
Bradford-Grieve et al. 2003	Southern Plateau, New Zealand	temperate seas	model	secondary consumers	23
Chassot et al. 2010	temperate	temperate seas	model	average	10
Chassot et al. 2010	tropical	tropical seas	model	average	14
Chassot et al. 2010	upwelling	upwelling ecosystems	model	average	5
Cornejo-Donoso and Antezana 2008	Antarctic Peninsula	temperate seas	model	primary consumers	21
Cornejo-Donoso and Antezana 2008	Antarctic Peninsula	temperate seas	model	secondary consumers	20

Articles	Region	Clustered Region	Method	Trophic Level	Transfer Efficiency
Cornejo-Donoso and Antezana 2008	Antarctic Peninsula	temperate seas	model	tertiary consumers	10
Cornejo-Donoso and Antezana 2008	Antarctic Peninsula	temperate seas	model	quaternary consumers	5
D'Alelio et al. 2016	Gulf of Naples	temperate seas	model	primary consumers	20
Duan et al. 2009	Pearl River Estuary	tropical seas	model	average	10.2
Gamito and Erzini 2004	Ria Formosa lagoon, south Portugal	lagoons	model	primary consumers	4.8
Gamito and Erzini 2004	Ria Formosa lagoon, south Portugal	lagoons	model	secondary consumers	6
Gamito and Erzini 2004	Ria Formosa lagoon, south Portugal	lagoons	model	tertiary consumers	3.5
Gamito and Erzini 2004	Ria Formosa lagoon, south Portugal	lagoons	model	quaternary consumers	1.2
Gamito and Erzini 2004	Ria Formosa lagoon, south Portugal	lagoons	model	quinary consumers	0.2

Articles	Region	Clustered Region	Method	Trophic Level	Transfer Efficiency
Libralto et al. 2008	Atlantic coast of Morocco	temperate seas	model	average	10.9
Libralto et al. 2008	Azores archipelago	temperate seas	model	average	10.5
Libralto et al. 2008	Bali Strait	tropical seas	model	average	11.7
Libralto et al. 2008	Baltic sea	temperate seas	model	average	25.9
Libralto et al. 2008	Bay of Bengal	tropical seas	model	average	9
Libralto et al. 2008	Bay of Biscay	temperate seas	model	average	16.5
Libralto et al. 2008	Bay of Revellata, Corsica	temperate seas	model	average	18.8
Libralto et al. 2008	Bolinao reef flat	tropical seas	model	average	10.4
Libralto et al. 2008	Brunei Darussalam	tropical seas	model	average	12.9
Libralto et al. 2008	California upwelling	upwelling ecosystems	model	average	4
Libralto et al. 2008	Campeche Bank of Yucatan shelf	tropical seas	model	average	17.6
Libralto et al. 2008	Cantabric Sea	temperate seas	model	average	38.1

Articles	Region	Clustered Region	Method	Trophic Level	Transfer Efficiency
Libralto et al. 2008	Celestun lagoon, Mexico	lagoons	model	average	6.4
Libralto et al. 2008	Central North Pacific Ocean	temperate seas	model	average	4.4
Libralto et al. 2008	Chesapeake Bay	temperate seas	model	average	12.5
Libralto et al. 2008	Chiku lagoon, Taiwan	lagoons	model	average	13.1
Libralto et al. 2008	Coast of Western Gulf of Mexico	tropical seas	model	average	16.2
Libralto et al. 2008	Coastal areas and reefs	-	model	average	13
Libralto et al. 2008	Continental shelf of southern Brazil	tropical seas	model	average	11.8
Libralto et al. 2008	Eastern Bering Sea	temperate seas	model	average	13.2
Libralto et al. 2008	Eastern Scotian shelf	temperate seas	model	average	11
Libralto et al. 2008	Eastern Tropical Pacific Ocean	temperate seas	model	average	20.4
Libralto et al. 2008	Etang de Thau, France	lagoons	model	average	10.8
Libralto et al. 2008	Faroe Islands	temperate seas	model	average	14.4

Articles	Region	Clustered Region	Method	Trophic Level	Transfer Efficiency
Libralto et al. 2008	Faroe Islands	temperate seas	model	average	15.4
Libralto et al. 2008	Floreana rocky reef Galapagos	tropical seas	model	average	13
Libralto et al. 2008	Georgia Strait	temperate seas	model	average	9.5
Libralto et al. 2008	Great Barrier Reef	tropical seas	model	average	11.5
Libralto et al. 2008	Gulf of Lingayen	tropical seas	model	average	13.5
Libralto et al. 2008	Gulf of Maine - Georges Bank	temperate seas	model	average	15.6
Libralto et al. 2008	Gulf of Mexico continental shelf	tropical seas	model	average	9.7
Libralto et al. 2008	Gulf of Thailand	tropical seas	model	average	10.4
Libralto et al. 2008	Hong Kong	tropical seas	model	average	9.1
Libralto et al. 2008	Icelandic fisheries	temperate seas	model	average	14.2
Libralto et al. 2008	Kuala Trengganu	tropical seas	model	average	17.8
Libralto et al. 2008	Laguna de Bay, Philippines	lagoons	model	average	12.4

Articles	Region	Clustered Region	Method	Trophic Level	Transfer Efficiency
Libralto et al. 2008	Lancaster Sound Region	temperate seas	model	average	8.2
Libralto et al. 2008	Maputo Bay	temperate seas	model	average	7.6
Libralto et al. 2008	Newfoundland	temperate seas	model	average	14.3
Libralto et al. 2008	North Benguela upwelling	upwelling ecosystems	model	average	7.9
Libralto et al. 2008	North Coast of Central Java	tropical seas	model	average	11.8
Libralto et al. 2008	North Sea	temperate seas	model	average	11.6
Libralto et al. 2008	Northern British Columbia	temperate seas	model	average	14.2
Libralto et al. 2008	Northern Gulf of Saint Lawrence	temperate seas	model	average	12
Libralto et al. 2008	Northern-central Adriatic Sea	temperate seas	model	average	10
Libralto et al. 2008	Norwegian and Barents Sea	temperate seas	model	average	10.5
Libralto et al. 2008	NW Africa upwelling	upwelling ecosystems	model	average	6.1

Articles	Region	Clustered Region	Method	Trophic Level	Transfer Efficiency
Libralto et al. 2008	Open oceans	open oceans	model	average	12
Libralto et al. 2008	Orbetello lagoon	lagoons	model	average	9.6
Libralto et al. 2008	Peru upwelling	upwelling ecosys-tems	model	average	6.6
Libralto et al. 2008	Prince William Sound, Alaska	temperate seas	model	average	14.1
Libralto et al. 2008	San Miguel Bay	tropical seas	model	average	20.6
Libralto et al. 2008	San Pedro Bay	tropical seas	model	average	9.4
Libralto et al. 2008	Schlei Fjord	temperate seas	model	average	7.4
Libralto et al. 2008	Shallow areas of Gulf of Thailand	tropical seas	model	average	6.8
Libralto et al. 2008	South Catalan Sea	temperate seas	model	average	12.6
Libralto et al. 2008	South China Deep Sea	tropical seas	model	average	10.6
Libralto et al. 2008	Southern Brazil	tropical seas	model	average	6.3
Libralto et al. 2008	Southwest coast of India	tropical seas	model	average	14

Articles	Region	Clustered Region	Method	Trophic Level	Transfer Efficiency
Libralto et al. 2008	Tampa Bay	tropical seas	model	average	8.6
Libralto et al. 2008	Temperate shelves	temperate seas	model	average	14
Libralto et al. 2008	Tropical shelves	tropical seas	model	average	10
Libralto et al. 2008	Upwellings	upwelling ecosys-tems	model	average	5
Libralto et al. 2008	Venezuela north-eastern shelf	tropical seas	model	average	7.3
Libralto et al. 2008	Venice lagoon	lagoons	model	average	14.5
Libralto et al. 2008	Vietnam-China shelf	tropical seas	model	average	7.5
Libralto et al. 2008	West Coast of Vancouver Island	temperate seas	model	average	13.7
Libralto et al. 2008	West Greenland coast	temperate seas	model	average	12.1
Libralto et al. 2008	West Greenland trawling area	temperate seas	model	average	7.1
Lin et al. 2006	Tapong Bay, southwestern Taiwan	tropical seas	model	average	5.5

Articles	Region	Clustered Region	Method	Trophic Level	Transfer Efficiency
Liu et al. 2009	Nanwan Bay, southern Taiwan	tropical seas	model	primary consumers	13.9
Liu et al. 2009	Nanwan Bay, southern Taiwan	tropical seas	model	secondary consumers	6.6
Liu et al. 2009	Nanwan Bay, southern Taiwan	tropical seas	model	tertiary consumers	5.2
Liu et al. 2009	Nanwan Bay, southern Taiwan	tropical seas	model	quaternary consumers	2
Manickchand-Heileman et al. 2003	Gulf of Paria, Venezuela and Trinidad	tropical seas	model	average	12.2
Neira and Arancibia 2004	upwelling Central Chile	upwelling ecosystems	model	primary consumers	8.1
Neira and Arancibia 2004	upwelling Central Chile	upwelling ecosystems	model	secondary consumers	27.4
Neira and Arancibia 2004	upwelling Central Chile	upwelling ecosystems	model	tertiary consumers	26.8
Neira and Arancibia 2004	upwelling Central Chile	upwelling ecosystems	model	quaternary consumers	6.7

Articles	Region	Clustered Region	Method	Trophic Level	Transfer Efficiency
Neira and Arancibia 2004	upwelling Central Chile	upwelling ecosystems	model	quinary consumers	7.4
Pauly and Christensen 1995	general claim	NA	empirical	average	10.13
Rybarczyk and Elkaim 2003	Bay of Somme	temperate seas	model	primary consumers	15.6
Rybarczyk and Elkaim 2003	Chesapeake Bay	temperate seas	model	primary consumers	31
Rybarczyk and Elkaim 2003	Delaware Bay	temperate seas	model	primary consumers	69
Rybarczyk and Elkaim 2003	Narragansett Bay	temperate seas	model	primary consumers	69
Rybarczyk and Elkaim 2003	Seine Estuary	temperate seas	model	primary consumers	36.05
Sheldon et al. 1977	ocean pelagic	open oceans	model	primary consumers	15
Tsagarakis et al. 2010	N. Aegean	temperate seas	model	average	17.4
Tsagarakis et al. 2010	N.C. Adriatic	temperate seas	model	average	10
Tsagarakis et al. 2010	S. Catalan	temperate seas	model	average	12.6
Villaneuva et al. 2006	Ebrie lagoon	lagoons	model	average	15.5

Articles	Region	Clustered Region	Method	Trophic Level	Transfer Efficiency
Ware 2000	Georges Bank	temperate seas	model	primary consumers	15.9
Ware 2000	Gulf of Maine	temperate seas	model	primary consumers	11.9
Ware 2000	Gulf of Maine	temperate seas	model	secondary consumers	10.5
Ware 2000	Mid-Atlantic Shelf	tropical seas	model	primary consumers	10.2
Ware 2000	Mid-Atlantic Shelf	tropical seas	model	secondary consumers	10.1
Ware 2000	Nova Scotia Shelf	temperate seas	model	primary consumers	12.1
Ware 2000	Nova Scotia Shelf	temperate seas	model	secondary consumers	12.3
Ware 2000	Oyashio current model	temperate seas	model	primary consumers	17
Ware 2000	Oyashio current model	temperate seas	model	secondary consumers	7.9
Ware 2000	SW Britisth Columbia model	temperate seas	model	primary consumers	11.1
Ware 2000	SW Britisth Columbia model	temperate seas	model	secondary consumers	8

Table 2.2: Marine transfer efficiency data

2.3 Discussion

Overall, these studies show similarities between the two systems. The results give visual evidence that although the average transfer efficiency does not differ greatly from 10%, there is substantial variation in transfer efficiency in both systems (Figure 2.3). Additionally our study was able to identify that trophic level and the general geographic location of the ecosystem impacts the variability of the transfer efficiency. Some of this large variation seems to have predictable patterns, but the potential sources of this variation can only be explored for the larger sample sizes from marine systems. Thus in the absence of data, the distributions generated in the current study are good starting points to model the variation in transfer efficiency.

We hypothesize the difference in transfer efficiency between trophic levels could be partially attributed to the composition of the taxa. Additionally, the differences could be due to the mobility of the organisms at each trophic levels and the amount of energy expedited to capture their prey. Both of these points highlight the need for additional research that can distinguish between the different mechanisms that influence transfer efficiency (e.g., endotherms vs. ectotherms).

We encountered two main difficulties in this study due to the different naming conventions in the early states (Figure 2.1 and 2.2) and the different efficiencies of interest amongst different fields. We bring this point up to highlight the challenges in conducting interdisciplinary work. We found freshwater studies report either assimilation and consumption efficiencies or transfer efficiency, while the majority of marine papers focused on transfer efficiency. As previously mentioned, we found very few terrestrial studies that reported results on transfer efficiency. It is intriguing that the focus on transfer efficiency is an aquatic phenomenon. While we are unable to say with certainty why this is, we

can speculate that in aquatic systems, and especially marine systems, it is extremely difficult to gather data on most species in an ecosystem. As a result, it is challenging to calculate the assimilation and consumption efficiencies in these systems. The great ease in counting terrestrial populations means the studies do not need to rely on more inferential, trophic level wide, techniques.

Despite the aforementioned limitations, we are able to demonstrate the 10% value is problematic given the substantial variation that exists in the transfer efficiency. Even though the average and median values that emerge from the synthesis are not dramatically different from 10%–which may suggest that an assumption of 10% would be reasonable–by continuing to use a 10% transfer efficiency researchers are eschewing the large variation and predictable patterns within this variation, which will impact a food web model?s ability to provide realistic results. From here on, we explore implications of applying a fixed 10% transfer efficiency in ecology, fisheries, and aquaculture.

2.3.1 Applications in Ecology

When it comes to ecology, the transfer efficiency is used to understand food web dynamics and how various species interact and influence one another. Most commonly, it is used in size spectrum studies that explore predator-prey relationships (Barnes et al. 2010). In general though, the more detailed and species-specific assimilation efficiency is used more frequently in such analyses. Since many issues within ecology and conservation biology focus on individual species patterns, it is valuable to be able to calculate species-specific efficiencies. The transfer efficiency is perhaps too broad of a metric for many questions. While using a 10% transfer efficiency value can have drastic underrepresentation and lead to severely mismanaged systems, given the preference for the assimilation efficiency the implications of a 10% transfer efficiency most likely has not

been directly measured in many cases in ecology. Previous research has shown distinct assimilation efficiencies between trophic levels and for terrestrial (i.e., temperate forests, deciduous forests, and grasslands) and freshwater (i.e., lakes) systems (Hairston 1993). Even though the assimilation and consumption efficiencies are different measures, there is comparable variation in them as there is in the transfer efficiency data, and therefore the implications of applying fixed values for these efficiencies could be the same as those for the transfer efficiency.

2.3.2 Applications in Fisheries

In fisheries, transfer efficiency is mostly used to determine the impact of fishing on a population. It can be used to estimate various metrics, such as biomass and the primary production required (PPR) given fisheries catch. Primary production required estimates the amount of net primary production needed to replace the biomass removed by fisheries landings. The idea is that primary production is a major limiting factor of fisheries catch. If biomass or the primary production required is incorrectly estimated, we risk the potential of under- or overfishing.

What is the cost of ignoring variability in transfer efficiencies for fisheries applications? To explore this question, we reexamined the analyses by Chassot et al. (2010) and Watson et al. (2014) of the primary production required to produce the biomass of fish that were caught. As an exploration, we recreated the analysis for one marine region, the California Current. We gathered annual catch data from Sea Around Us (`http://www.seaaroundus.org/`) from 1950 to 2014, trophic level data on fishes from Fishbase data base (`http://www.fishbase.org/`), and net primary production data on SeaWiFS from the Ocean Productivity web site (`http://www.science.oregonstate.edu/ocean.productivity/`). Watson et al. (2014) also included trophic level data on

invertebrates from SeaLifeBase; however, SeaLifeBase does not include trophic level data for the California Current. We first replicated the previously published results using a fixed 10% transfer efficiency using the following equation[1].

$$PPR_t = \sum_{i=1}^{n} \frac{C_{i,t}}{CR} * \frac{1}{TE}^{TL_i - 1} \tag{2.1}$$

PPR_t is the primary production required in year t to produce the observed catch, $C_{i,t}$ is the biomass of catch for species i in year t, CR is the conversion rate of carbon to wet weight, TE is the transfer efficiency, TL_i is trophic level for species i, and n is the number of species within a region.

To explore the impact of ignoring variability in transfer efficiencies, we used the data we gathered on marine transfer efficiencies to test for sensitivities to transfer efficiency variability. We fit an approximate distribution to the marine transfer efficiency data using goodness-of-fit criterion so that we could randomly sample a value from the observed distribution instead of using a fixed 10% value in the above equation (See Appendix A for details). This approach allowed us to incorporate variability in the transfer efficiency. While the time span of the Sea Around Us catch data ran for over 60 years, we explored the distribution at 20 year intervals to get a sample of the changes in transfer efficiency. We constructed Monte Carlo simulations and ran 10,000 simulations for the years 1950, 1970, 1990, and 2010. Although it is difficult to compare our results directly to Watson et al. (2014), since their results are broken up by continental fishing fleets, we found that when transfer efficiency is allowed to vary, the projected PPR varies dramatically. Although the majority of the time the PPR is a sustainable fraction of the total production of the California Current, on average 47.18% of the time (i.e., 46.93% in 1950, 47.21% in 1970, 47.35% in 1990, and 47.22% in 2010) the simulations suggest annual landings

[1]Chassot et al. (2010) specified ecosystem specific transfer efficiencies.

could exceed total primary production (Figure 2.6). Therefore, the application of a 10% transfer efficiency in fisheries management has a high chance of leading to unsustainable fishing practices.

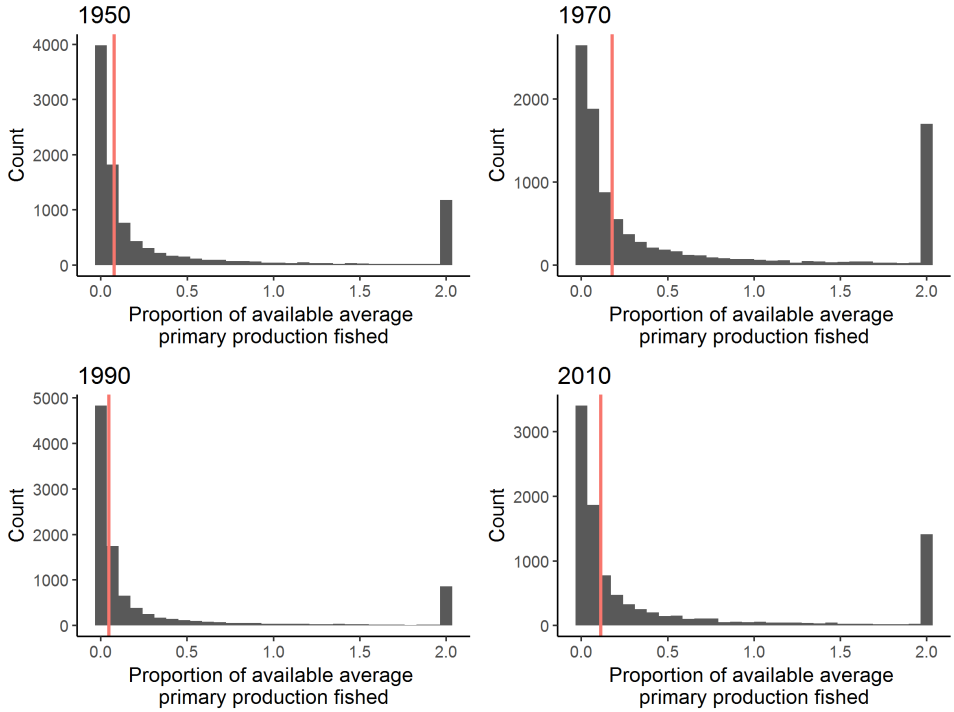

Figure 2.6: Histogram of simulated primary production required given catch divided by the average primary production in 1950, 1970, 1990, and 2010 in the California Current using a random transfer efficiency. The vertical red line indicates the primary production required divided by primary production using a fixed 10% transfer efficiency in the calculations. We adjusted the far right bin width due to rare events. The original range extends out further. Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

2.3.3 Applications in Aquaculture

Aquaculture already cultivates hundreds of different marine and freshwater species. The relevance of transfer efficiencies to production decisions depends on whether the species requires additional feed or not. Non-fed aquaculture includes either primary producers or primary consumers that are not provided additional feeds by the aquaculturist.

Fed aquaculture, on the other hand, involves consumer species that are typically fed compound feeds designed to meet their specific nutritional requirements. Consumption of autotrophs, usually phytoplankton by filter feeders, is studied extensively in non-fed aquaculture to track carrying capacity of ecosystems with added aquaculture (Banas et al. 2007). These carrying capacities are estimated using assimilation efficiencies of a specific target species (Rosland et al. 2009, Irisarri et al. 2013, Srisunont and Babel 2016) or as broader estimates of transfer efficiencies (Simenstad and Fresh 1995, Sommer 1998, Byron et al. 2011, Han et al. 2017).

More and more fed species are given specialized compound feeds rather than whole organisms in order to increase efficiency and sustainability of feed resources. To assess the sustainability of feeds, the field has begun to track transfer efficiencies by dissecting the feed into its compositional parts and estimating the efficiencies for each of the components using the method devised by Pauly and Christensen (1995). While a 10% transfer efficiency was initially employed to estimate the primary production requirements or biotic resource use in environmentally-based aquaculture assessments (e.g., Papatryphon et al. 2004, Pelletier and Tyedmers 2007, Pelletier et al. 2009), more recent studies have incorporated species-specific efficiencies in their analyses to better understand the environmental tradeoffs between different feed compositions (Cashion et al. 2016). Additionally, Cashion et al. (2016) found that the use of a 10% transfer efficiency has led to an underestimation of the impacts of salmon aquaculture on natural marine biomass resources. Actual impacts are likely three times greater. As with the case for wild fisheries, ignoring the variability in transfer efficiencies can have negative impacts on the conservation of limited resources and management of human activities.

2.4 Concluding Remarks

Our results raise the question why the use of an assumed transfer efficiency value of 10% is still so prevalent despite widespread evidence of substantial variability and examples of where the consequences of ignoring such variability have been documented? One issue is clearly convenience. The overall mean of observed values is not dramatically different from the assumed 10% value. Nonetheless, given the scope of observed variation and uncertainty, using a 10% value for the sake of arithmetic convenience carries large risks. Our study is not the first to raise this point, but synthesizing the full scope of evidence around the levels of variability and its potential consequences for decisions will hopefully highlight the costs of ignoring variability in this key ecosystem parameter.

Another reason for the continued assumption of a constant value is the lack of relevant data for most systems. But our synthesis of the distributions of observed values for transfer efficiency in freshwater and marine ecosystems provides an opportunity to draw from this synthetic distribution rather than assuming a constant value. In cases where locally relevant data are infeasible to collect, this synthetic distribution may provide a better platform for decisions.

Finally, one other issue is that the studies that have addressed the variability in transfer efficiency specifically have typically been in the context of narrower questions (e.g., aquaculture feeds). These narrower studies may not catch the attention of people using transfer efficiencies in other ways. By pulling together information from diverse studies from different fields, we hope this synthesis will generate a broader discussion.

In the absence of specific data for a broader array of systems, instead of using a fixed 10% constant, we suggest using simulations or Bayesian models drawing on these

synthetic distributions, which are great tools for incorporating variation and uncertainty. This approach would take us a long ways towards creating more valid food web models, and as such, improve our understanding on how food web dynamics impact community structure.

Chapter 3

Untangling uncertainty in food web models

3.1 Introduction

Many managed fisheries constitute data-limited cases (Farmer et al. 2016) where there is insufficient information to evaluate population health and the sustainability of fishing practices with a full stock assessment. One common approach to overcome data limitations has been to aggregate multiple species together into stock complexes and create a single aggregate population dynamic model for the group (Cope et al. 2011). Typically, similar species, or the species that are targeted together by the same fleets, are aggregated into a common model. However, this approach is based on strong assumptions which can compromise the validity or interpretability of the model results.

Perhaps the largest limitation to using a species complex model is that clustering species leads to a loss of information at the individual-species level. This is especially true when aggregated complexes are assessed using simple biomass dynamics models. In

those cases the aggregated species are assumed to share life history traits (e.g., carrying capacity, expected lifespan, growth rates, reproductive rates, etc.), geographical ranges, and vulnerabilities. Ignoring the potential differences in vulnerabilities and sensitivities to fishing pressure is a serious concern when modeling species together as a complex and can lead to inaccurate and unsustainable predictions of biomass for some or all the stocks. Catch levels set to maximize one species could drive a "weaker stock" to unsustainably low levels (Hastings et al. 2017).

Traditional single-species assessment modeling estimates the impacts of fishing using fisheries-related data such as catch, composition of the catch, and indices of abundance. These single-species models often lack the capability to incorporate broader-scale ecosystem information. Food web models, which are a type of ecosystem-based model, are capable of providing estimates of quantities such as carrying capacity, tropic level abundance, and transfer rates that are conceptually similar to concepts of single-species assessment modeling of carrying capacity, abundance, and intrinsic rates of increase. Incorporating ecosystem concepts into single-species assessments may offer a way forward for data-limited situations as well as provide a pathway from single-species to ecosystem-based management.

A problem in incorporating ecological concepts into statistical models of applied fisheries assessments is the quantification of uncertainty. Fisheries models are increasingly using state-space modeling approaches and with that comes an array of possibilities. Measures of the confidence in model results are important in setting management decisions. Most existing food web models are deterministic (do not take into account variation). A way of incorporating uncertainty in models is through simulation. We developed an approach using trophic pyramid food webs combined with Monte Carlo simulations to

simulate values of trophic level biomass under different ecological assumptions. By considering uncertainty under different scenarios, we are able to quantify parameter and model output uncertainties in a way that has not previously been done in existing ecosystem models.

Applying this technique on a data-limited fishery gives insight into the rigor of information we can expect to obtain when ecological concepts are incorporated into fisheries models and from there, how this information can be applied in fisheries assessments. The past few assessments of the main Hawaiian Islands Deep7 Bottomfish Complex groundfish fishery used an aggregate surplus production model for the complex. However there is a concern that life history traits may differ amongst the targeted species, which could ultimately affect the vulnerability to fishing pressure (Brodziak et al. 2011, Langseth et al. 2018). We used the main Hawaiian Islands (MHI) as the setting for our simulations to not only gain insight on the fishery and region, but to also minimize assumptions in our model construction, particularly those related to the movement of migratory species and larval dispersal. We needed to select a marine ecosystem that was as close to a naturally closed system as possible. Hawaii does not lie directly in the path of any current system, and the local wind and current patterns create a system of cyclonic eddies. Due to these properties, prior studies concluded that there is an oceanic barrier to larval dispersal (Lobel and Robinson 1986, Vermeij 1987, Hourigan and Reese 1987). Therefore, we considered the MHI a naturally enclosed marine ecosystem due to the containment of larval drift.

3.2 Methods

3.2.1 Base Equation

Both Lalli and Parsons (1997) and Libralato et al. (2008) outlined that production at each trophic level (h) can be estimated from net primary production (NPP, denoted ν) multiplied by the trophic transfer efficiencies (τ_h). In line with extant research on this topic (see Cury et al. 2005, Libralato et al. 2008, Chassot et al. 2010, Trebilco et al. 2013, Watson et al. 2014) we define the transfer efficiency (τ_h) as the fraction of production passing from trophic level $h-1$ to h, where h is the higher trophic level in the food web. In many cases net primary production is represented in metric tons C/year and is multiplied by 9 to convert from organic carbon (metric tons C) to wet weight (metric tons) (Strathmann 1967, Pauly and Christensen 1995, Chassot et al. 2010). We include the lifespan term λ_h to convert from the weight of individuals produced in a year to the overall standing stock in Eq. (3.1) as we are interested in solving for trophic level biomass (γ_h) instead of production. See Table 3.1 for defined variables and parameters definitions within Eq. (3.1).

$$\gamma_h = \nu * 9 * \left(\prod_{j=2}^{h} \tau_j \right) \lambda_h \quad \text{for} \quad h \in \{2, 3, 4\} \tag{3.1}$$

Table 3.1: Definition of notation in Eq. (3.1) and scientific units used

Terms	Description	Units
h	Trophic level in MHI, where $h = 2, 3, 4$	
γ_h	Trophic level biomass at trophic level h	metric tons
ν	Net primary production (NPP)	metric tons C/year
9	Carbon to wet weight conversion ratio	
τ_h	Transfer efficiency between trophic level $h-1$ and h	
λ_h	Lifespan of a species found at trophic level h	years

3.2.2 Data

We built Eq. (3.1) for our case study of the MHI using the following data sources: We used the 8-day time series Sea-viewing Wide Field-of-view Sensor (SeaWiFS) chlorophyll a data from 1997 to 2010 that was transformed using the *Eppley*-Vertically Generalized Production Model (VGPM) to estimate NPP from chlorophyll a from the Ocean Productivity group at Oregon State University (`http://www.science.oregonstate.edu/ocean.productivity/`) (Fig. 3.1) (Behrenfeld and Falkowski 1997). Estimated trophic levels (rational numbers) were calculated from the FishBase database (`http://www.fishbase.org/`) of Froese and Pauly (2017) for each species and truncated to assign species into integer-valued trophic levels. Truncated trophic levels range in the MHI from $h = 1, \ldots, 4$ with $h = 1$ representing phytoplankton, $h = 2$ primary consumers, $h = 3$ secondary consumers, and $h = 4$ tertiary consumers. In this ecosystem, the maximum trophic level is 4. However, this value can vary in other ecosystems depending on the species composition. The Froese and Pauly (2017) FishBase data set was also used to estimate the lifespan (λ_h) across all species at trophic level h. In this database Froese and Pauly (2017) defines the term lifespan as "the maximum expected age, on average, for a species, cohort, stock, or a population in the absence of fishing. Smaller than maximum age although may be used in this sense."

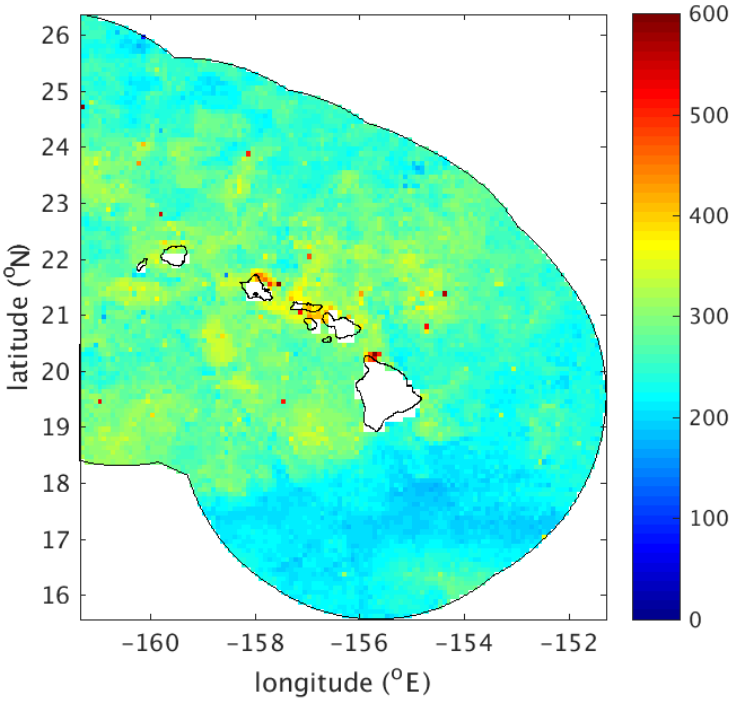

Figure 3.1: A single 8-day time frame of the SeaWiFS *Eppley*-VGPM NPP data in January 1998 for the main Hawaiian Islands EEZ. The color scale shows the amount of estimated NPP in total gigatons of Carbon per year per pixel. The pixel size of the SeaWiFS data set is 9 by 9 km. Image created by Erik Fields (Earth Research Institute, ERI).

3.2.3 Assumptions and General Model

Assumptions in Model

In Eq. (3.2), we present the decomposition of the joint distribution for the general ecological model. Here we assume that the random variables NPP (ν), transfer efficiencies (τ_2, τ_3, and τ_4), and lifespans (λ_2, λ_3, and λ_4) are independent. This assumption is ecologically reasonable, because the amount of energy going into the system does not depend upon the transfer of energy or an organism's lifespan. Additionally, we assume that the distribution of transfer efficiency at trophic level h, (τ_h), is independent of trans-

fer efficiency at the previous trophic level (τ_{h-1}). Throughout this paper, we therefore assume the joint distribution for $\nu, \tau_2, \tau_3, \tau_4, \lambda_2, \lambda_3$, and λ_4 can be broken down as:

$$f(\nu, \tau_2, \tau_3, \tau_4, \lambda_2, \lambda_3, \lambda_4) = f(\nu)f(\tau_2)f(\lambda_2)f(\tau_3)f(\lambda_3)f(\tau_4)f(\lambda_4) \tag{3.2}$$

Since the trophic level biomass, γ_h, for trophic levels $h \in \{2, 3, 4\}$ in Eq. (3.1) is a deterministic function of the random variables from Eq. (3.2), given the distributional assumptions detailed both above and in Table 3.2, it is possible in some cases to find the distribution of γ_h for $h \in \{2, 3, 4\}$. Our assumptions in Eq. (3.2) implies that we are assuming if trophic level h_x does not equal to trophic level h_y, the lifespan (λ_{h_x}) is independent of trophic level biomass (γ_{h_y}) for $h_x, h_y \in \{2, 3, 4\}$.

Random Distributions on Random Variables

Variation in ecosystem-based models is a contested area of research in fisheries science. Many studies use a default 10% transfer efficiency for τ_h for each $h \in \{2, 3, 4\}$ (i.e., $\tau_h = 0.10 \ \forall \ h \in \{2, 3, 4\}$), thereby assuming this is constant across all trophic levels (Lindeman 1942, Slobodkin 1962, May 1976, Pauly and Christensen 1995). However, other researchers have estimated that the true transfer efficiency may be higher than 10% (Schaefer 1965, Ryther et al. 1969, Sheldon et al. 1977, Baumann 1995), may vary depending on the size and type of ecosystem (Libralato et al. 2008, Heymans et al. 2011), and decrease with increasing trophic levels (May 1983, Barnes et al. 2010). As the current study is applied to a data-limited case, we are unsure of the values of NPP (ν), transfer efficiency (τ_h), and lifespans (λ_h) for $h \in \{2, 3, 4\}$ in the model. Therefore, we choose to study sensitivities of biomass (γ_h) to each of the three components.

Distributional Assumptions

Several distributional and conditional assumptions we placed on the distributions of the random variables NPP (ν), transfer efficiencies (τ_h), and lifespans (λ_h) for $h \in \{2, 3, 4\}$ are described in Table 3.2 and visualized in Fig. 3.2 as special cases of our general model Eq. (3.2). In Case 1 of $f(\nu)$, a value of $c_1 = 87339861.7$ metric tons C/ year was estimated by fitting a time series seasonal autoregressive integrated moving average model (SARIMA model) to the SeaWiFS data using R v.3.4.3 (R Core Team 2017) forecast package v8.2 and function auto.arima() (Hyndman 2017, Hyndman and Khandakar 2008). Case 2 of $f(\nu)$ was approximated based on goodness-of-fit tests fitted to the SeaWiFS data (Delignette-Muller and Dutang 2015) (Fig. 3.2, column 1).

We based the distributional assumptions for Case 1 of the transfer efficiencies τ_h for $h \in \{2, 3, 4\}$ from Pauly and Christensen (1995). We used Pauly and Christensen (1995) average transfer efficiency value of 0.1013 for c_2 in our study. We utilized data gathered from a meta analysis on marine transfer efficiencies to fit approximate distributions based on goodness-of-fit tests (Delignette-Muller and Dutang 2015) for Case 2 and 3 of the transfer efficiencies τ_h (Fig. 3.2, column 2). Case 2 has the transfer efficiency (τ_h) come from the same distribution for all $h \in \{2, 3, 4\}$, while Case 3 places distinct distributions on each transfer efficiency.

In Case 1 for $f(\lambda_2)$, $f(\lambda_3)$, and $f(\lambda_4)$, the values of $c_3 = 5.75$ years, $c_4 = 6$ years, and $c_5 = 11.95$ years were calculated by taking the median of the maximum expected lifespan data at each trophic level h. When lifespans (λ_h for $h \in \{2, 3, 4\}$) are treated as random variables (non-constant), we chose approximate distributions based on goodness-of-fit tests (Delignette-Muller and Dutang 2015) (Fig. 3.2, column 3). Details of the SARIMA model and the steps taken to estimate the parameters in the Lognormal and

Beta distributions in Table 3.2 are in Appendix B.

Table 3.2: Distributional assumption cases are special cases for the components of the joint distribution (i.e., Eq. (3.2)). We treat c_1, c_2, c_3, c_4, and c_5 as constants. Lack of subscripts means the parameter has the same value across all distributions.

	Case 1	Case 2	Case 3
$f(\nu)$	1 if $\nu = c_1$ 0 else	$Lognormal(\mu_\nu, \sigma^2_\nu)$	
$f(\tau_2)$	1 if $\tau_2 = c_2$ 0 else	$Beta(\alpha, \beta)$	$Beta(\alpha_{\tau_2}, \beta_{\tau_2})$
$f(\tau_3)$	1 if $\tau_3 = c_2$ 0 else	$Beta(\alpha, \beta)$	$Beta(\alpha_{\tau_3}, \beta_{\tau_3})$
$f(\tau_4)$	1 if $\tau_4 = c_2$ 0 else	$Beta(\alpha, \beta)$	$Beta(\alpha_{\tau_4}, \beta_{\tau_4})$
$f(\lambda_2)$	1 if $\lambda_2 = c_3$ 0 else	$Lognormal(\mu_{\lambda_2}, \sigma^2_{\lambda_2})$	
$f(\lambda_3)$	1 if $\lambda_3 = c_4$ 0 else	$Lognormal(\mu_{\lambda_3}, \sigma^2_{\lambda_3})$	
$f(\lambda_4)$	1 if $\lambda_4 = c_5$ 0 else	$Lognormal(\mu_{\lambda_4}, \sigma^2_{\lambda_4})$	

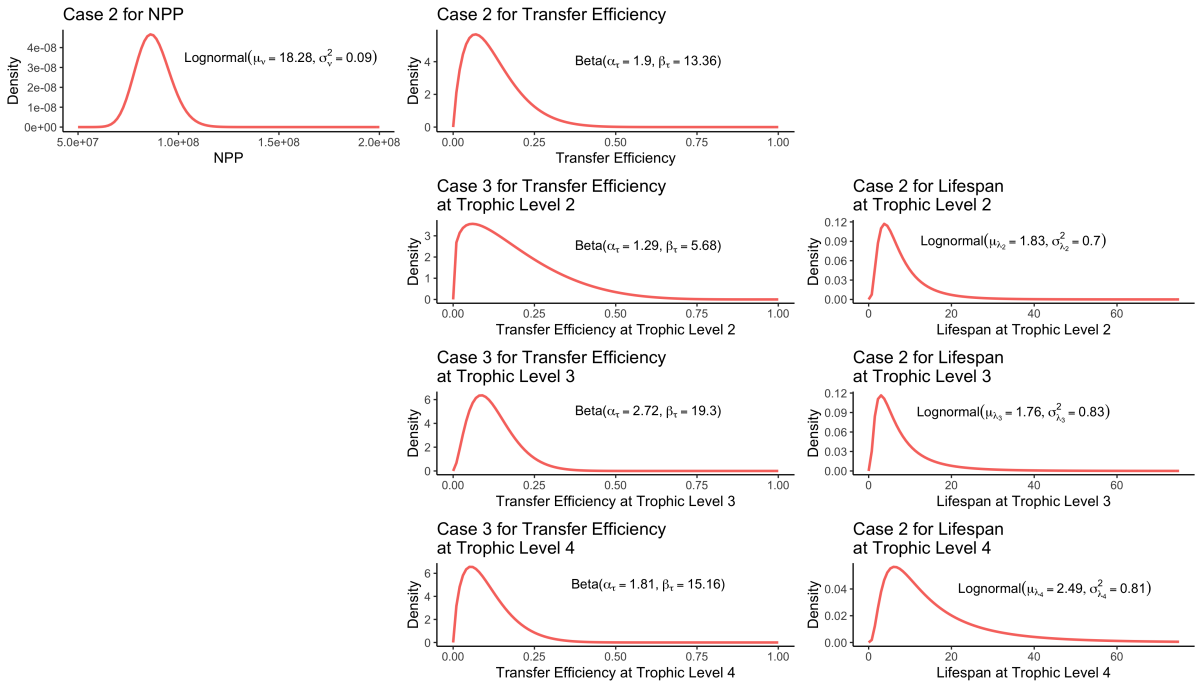

Figure 3.2: Probability distributions for the distributional assumption cases described in Table 3.2 where NPP (ν), transfer efficiencies (τ_h for $h \in \{2,3,4\}$), and lifespans (λ_h for $h \in \{2,3,4\}$) are not fixed values. Column 1 visualizes the distributional assumption placed on NPP. Column 2 breaks down the two distributional assumptions cases placed on transfer efficiencies (τ_h for $h \in \{2,3,4\}$), and column 3 shows the distributional assumptions placed on lifespans (λ_h for $h \in \{2,3,4\}$). Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

3.2.4 Simulation Scenarios

We combine distributional assumptions on the components of Eq. (3.2) corresponding to NPP $f(\nu)$, transfer efficiencies $f(\tau_h)$, and lifespans $f(\lambda_h)$ for $h \in \{2,3,4\}$ from Table 3.2 to create different scenarios that represent various ecological assumptions (See Table 3.3 for the explanation of scenario code names and Table 3.4 for description of scenarios). In Table 3.4, we interpret each cell as the i^{th} distributional case outlined in Table 3.2 for the x^{th} variable. For example, we would interpret $f^1(\nu)$ as the distributional assumption under the column Case 1 for $f(\nu)$ in Table 3.2. The transfer efficiencies (τ_2, τ_3, τ_4) and

maximum expected lifespans ($\lambda_2, \lambda_3, \lambda_4$) are independent of each other in all cases. We construct different scenarios to not only incorporate uncertainty in the NPP (denoted by superscript "2" in Table 3.2, $f^2(\nu)$) and lifespans ($f^2(\lambda_h)$ for $h \in \{2,3,4\}$), but also to encompass contrasting values for the transfer efficiency. The three areas focused on in this manuscript are: constant transfer efficiencies ($f^1(\tau_h)$ for $h \in \{2,3,4\}$), non-constant but independent transfer efficiencies ($f^2(\tau_h)$ for $h \in \{2,3,4\}$), and the constriction on transfer efficiencies distributions at successive trophic levels ($f^3(\tau_h)$ for $h \in \{2,3,4\}$). Scenario NTL (i.e., fixed NPP (ν), fixed transfer efficiencies (τ_h), and fixed lifespans (λ_h)) is a simplistic model intended to be used as a comparison to other relatively similar simple equations (Pauly and Christensen 1995, Cury et al. 2005, Chassot et al. 2010, Watson et al. 2014).

Table 3.3: Key of the scenario code names used to identify the various ecological and distributional assumptions.

Types of Transfer Efficiency Variation	NPP (ν)		Transfer efficiency (τ_h)		Lifespan (λ_h)	
	Fixed	Random	Fixed	Random	Fixed	Random
Fixed 10%	N	n	T		L	l
Allowed to vary	N	n		t	L	l
Decreases as move up food web	**N**	**n**		**t**	**L**	**l**

Table 3.4: Scenarios are combinations of the distributional assumption cases described in Table 3.2, where $f^i(x)$ indicates the i^{th} distributional assumption for the variable x. Table 3.3 defines the code names for the scenarios.

Scenarios	$f(\nu)$	$f(\tau_2)$	$f(\tau_3)$	$f(\tau_4)$	$f(\lambda_h)$
NTL	$f^1(\nu)$	$f^1(\tau_2)$	$f^1(\tau_3)$	$f^1(\tau_4)$	$f^1(\lambda_h)$
NTl	$f^1(\nu)$	$f^1(\tau_2)$	$f^1(\tau_3)$	$f^1(\tau_4)$	$f^2(\lambda_h)$
nTl	$f^2(\nu)$	$f^1(\tau_2)$	$f^1(\tau_3)$	$f^1(\tau_4)$	$f^2(\lambda_h)$
NtL	$f^1(\nu)$	$f^2(\tau_2)$	$f^2(\tau_3)$	$f^2(\tau_4)$	$f^1(\lambda_h)$
Ntl	$f^1(\nu)$	$f^2(\tau_2)$	$f^2(\tau_3)$	$f^2(\tau_4)$	$f^2(\lambda_h)$
ntl	$f^2(\nu)$	$f^2(\tau_2)$	$f^2(\tau_3)$	$f^2(\tau_4)$	$f^2(\lambda_h)$
NtL	$f^1(\nu)$	$f^3(\tau_2)$	$f^3(\tau_3)$	$f^3(\tau_4)$	$f^1(\lambda_h)$
Ntl	$f^1(\nu)$	$f^3(\tau_2)$	$f^3(\tau_3)$	$f^3(\tau_4)$	$f^2(\lambda_h)$
ntl	$f^2(\nu)$	$f^3(\tau_2)$	$f^3(\tau_3)$	$f^3(\tau_4)$	$f^2(\lambda_h)$

3.2.5 Simulation Details

We were able to find equations for the distribution of γ_h for $h \in \{2, 3, 4\}$ based on the transformation of parameters in Eq. (3.2) for Scenarios NTl and nTl. For details see Appendix B. Given the current distributional assumptions placed on NPP ($f(\nu)$), transfer efficiencies ($f(\tau_h)$), and lifespans ($f(\lambda_h)$ for $h \in \{2, 3, 4\}$), we used simulations for the other six scenarios. Scenario NTL is a deterministic equation and therefore does not need simulation. We ran one million simulations for the other scenarios.

In order to support simulation as a reasonable option, we examine the densities of Scenario NTl and nTl where we were able to derive the distributions of γ_h and compare them to a version where we instead used simulations. We find that the analytical densities of the derived distributions and simulations are visually very similar to each other (Fig. 3.3).

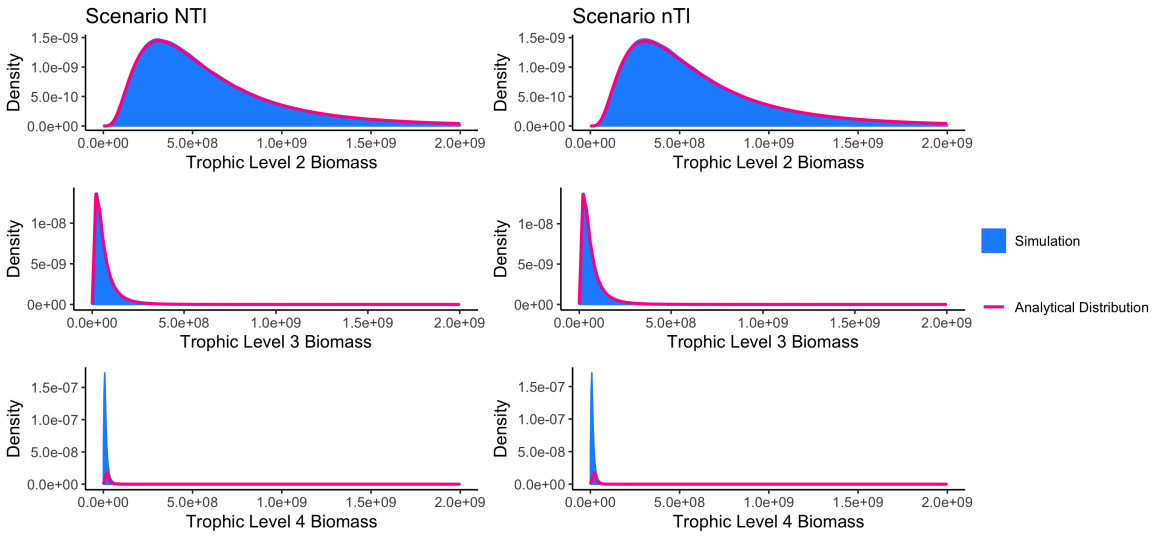

Figure 3.3: Comparison of the densities plots where we used simulations (blue) to the analytically derived probability distributions of γ_h (pink line) in Scenarios NTl and nTl at each trophic level. We limited the upper bound of the x-axis to improve visualization. Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

3.3 Results

At the primary consumer level ($h = 2$), all scenarios exhibit skewed distributions of trophic level biomass (γ_2) (Fig. 3.4). When the transfer efficiency is treated as a random variable in Scenarios *NtL*, *Ntl*, *ntl*, **NtL**, **Ntl**, and **ntl**, both the dispersion and interquartile ranges increase. Out of all of the scenarios, Scenarios NTl, nTl, *NtL*, and **NtL** have the smallest coefficients of variation (CV) while Scenarios **Ntl** and **ntl** have the largest CV (Table 3.5). The medians of the scenarios that represent the constricting distribution of transfer efficiencies as we move up the food web (i.e., **NtL**, **Ntl**, and **ntl**) are higher than Scenario NTL's trophic level biomass estimate (Fig. 3.5).

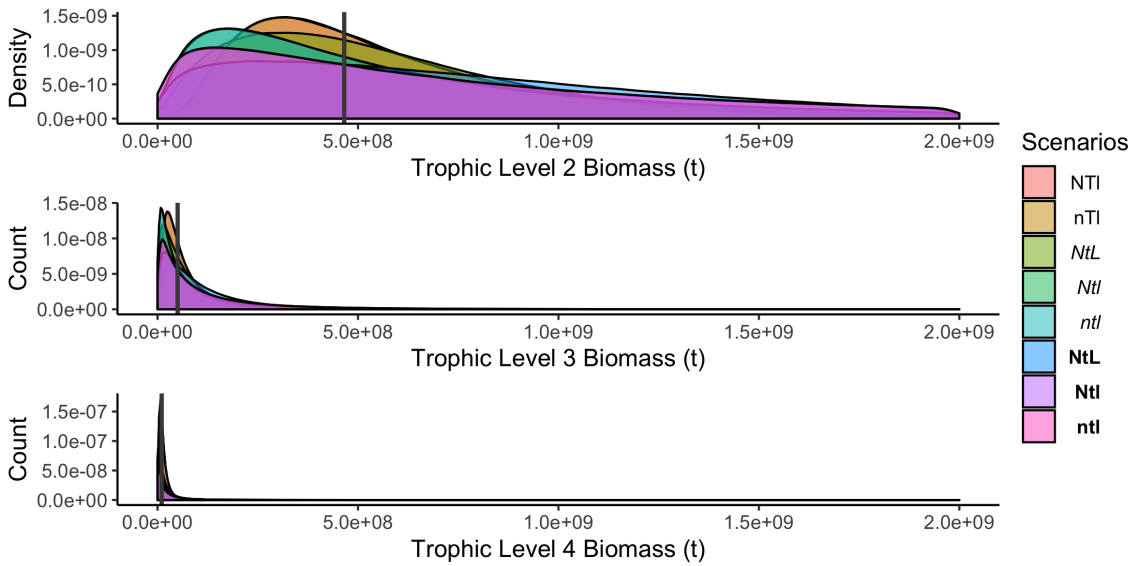

Figure 3.4: Density plots of the trophic level biomasses (γ_2, γ_3, γ_4, metric tons) for all scenarios except Scenario NTL. The plots have a vertical gray line representing the estimated trophic level biomass from Scenario NTL. Scenarios NTl and nTl show independent draws from analytical distributions, while the remaining scenarios show draws from simulations. We limited the upper bound of the x-axis to improve visualization. Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

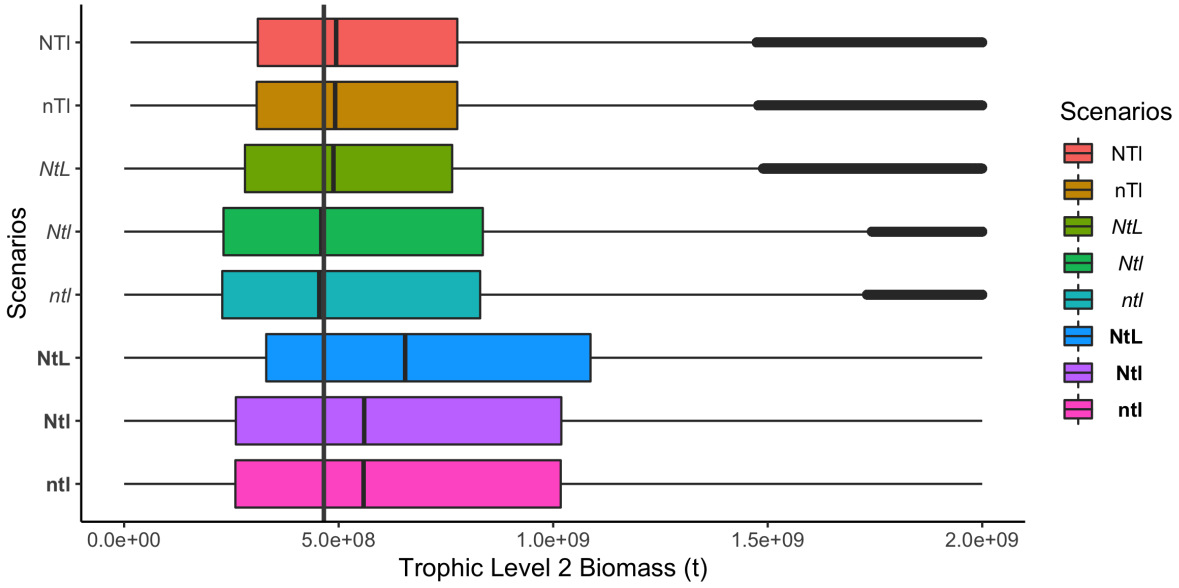

Figure 3.5: Box plots of the trophic level 2 biomass (γ_2, metric tons) for all scenarios except Scenario NTL. The plot has a vertical gray line representing the estimated trophic level biomass from Scenario NTL. Scenarios NTl and nTl show independent draws from analytical distributions, while the remaining scenarios show draws from simulations. The median of each distribution is denoted as a solid vertical black line inside each box. Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

Table 3.5: Coefficient of variation for each trophic level biomass and scenario combination. Scenario 1 is not included because it is a deterministic equation.

	Trophic level 2	Trophic level 3	Trophic level 4
Scenario NTl	0.79	1.00	0.97
Scenario nTl	0.80	1.01	0.98
Scenario *NtL*	0.66	1.02	1.38
Scenario *Ntl*	1.15	1.77	2.16
Scenario *ntl*	1.16	1.77	2.17
Scenario **NtL**	0.74	1.01	1.41
Scenario **Ntl**	1.23	1.73	2.18
Scenario **ntl**	1.24	1.76	2.20

At the secondary consumer level ($h = 3$), similar visual and mathematical patterns emerge at trophic level 3 that were found at trophic level 2. The distributions of trophic

level biomass (γ_3) for all scenarios are right-skewed with heavy tails (Fig. 3.4). The median of Scenarios **NtL** is a bit larger than Scenario NTL's trophic level biomass estimate, however the medians for the other scenarios are slightly lower (Fig. 3.6). The scenarios where transfer efficiency is a fixed 10% between successive trophic levels (i.e., Scenarios NTl and nTl) and the scenarios where the maximum expected lifespan is kept as a fixed constant (i.e., Scenarios *NtL* and **NtL**) have the smallest CVs. This is in contrast to the scenarios that allow the transfer efficiency to vary which have the largest CVs (Table 3.5). In general when the transfer efficiency is treated as a random variable (i.e., Scenarios *NtL*, *Ntl*, *ntl*, **NtL**, **Ntl**, and **ntl**), the variance and interquartile range increases. This is mostly attributed to the fact that these equations have more random values. Nevertheless, at trophic level 3 the dispersion has increased overall for all scenarios in comparison to the dispersion at trophic level 2.

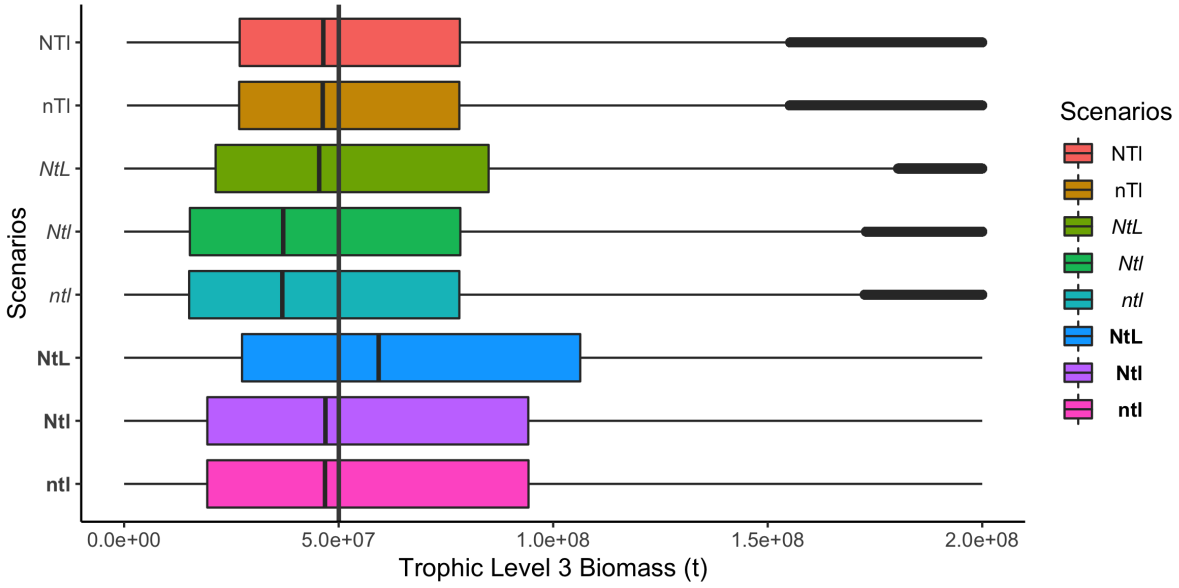

Figure 3.6: Box plots of the trophic level 3 biomass (γ_3, metric tons) for except Scenario NTL. The plot has a vertical gray line representing the estimated trophic level biomass from Scenario NTL. Scenarios NTl and nTl show independent draws from analytical distributions, while the remaining scenarios show draws from simulations. The median of each distribution is denoted as a solid vertical black line inside each box. Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

At the tertiary consumer level ($h = 4$), most of the same distributional and dispersion patterns emerge at trophic level 4 that were found at the previous two trophic levels. The distributions of trophic level biomass (γ_4) for all scenarios are right-skewed with heavy tails (Fig. 3.4). However at trophic level 4, the medians for all scenarios where transfer efficiency is treated as a random variable are lower than the case where the three components are fixed (i.e., Scenario NTL) (Fig. 3.7). Scenarios NTl and nTl, which both treat the transfer efficiency as a fixed 10% constant, have the smallest interquartile ranges and have medians that are relatively similar to Scenario NTL, but they have smaller CVs at trophic level 4 than at trophic level 3 (Table 3.5). For the other scenarios though, the simulations show an increase in the CVs at trophic level 4 in comparison to the CVs at

trophic level 3.

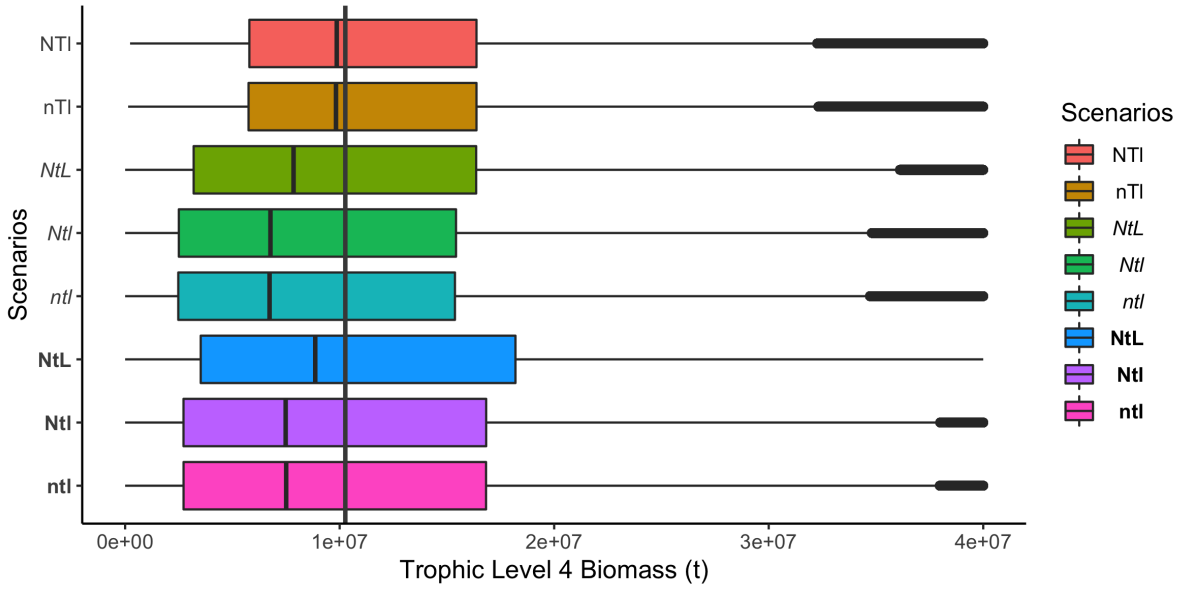

Figure 3.7: Box plots of the trophic level 4 biomass (γ_4, metric tons) for except Scenario NTL. The plot has a vertical gray line representing the estimated trophic level biomass from Scenario NTL. Scenarios NTl and nTl show independent draws from analytical distributions, while the remaining scenarios show draws from simulations. The median of each distribution is denoted as a solid vertical black line inside each box. Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

3.4 Discussion

Even though we placed different distributional assumptions on each of the scenarios, some common visual and mathematical patterns emerge in Fig. 3.5, 3.6, and 3.7 and in Table 3.5. The level of variability differed depending on the distributional assumptions. The simulations demonstrate that we can cluster the scenarios into 3 groups: fixed transfer efficiency (i.e., NTl and nTl), random transfer efficiency (i.e., *NtL*, *Ntl*, and *ntl*), and decreasing transfer efficiency (i.e., **NtL**, **Ntl**, and **ntl**). These communal patterns indicate that the distributional assumptions placed on the transfer efficiency affects the

distribution of trophic level biomasses more than the assumptions placed on NPP and maximum expected lifespan. Therefore by ignoring the variability in transfer efficiency, researchers risk drawing incorrect conclusions about biomass.

A common pattern emerged in the medians as trophic level increased. At trophic level two, most the the scenarios were similar to the case were the three components were fixed numbers (i.e., Scenario NTL), except with the cases were transfer efficiency decreased as trophic level increased (Fig. 3.5. That particular distributional assumption on transfer efficiency gave results that at the lower trophic level more biomass could be available than the fixed model Scenario NTL was predicting. However as trophic levels increased, all scenarios gradually had medians that were less than the fixed model Scenario NTL (Fig. 3.6 and 3.7). Implying, the fixed model NTL could be giving estimates of biomass that are greater than the ecosystem can sustain at higher trophic levels.

The results of the simulations provides direct insight on the MHI, which can in turn be used to inform management on the data-limited MHI Deep7 Bottomfish Complex groundfish fishery. This trophic pyramid food web model, which incorporates variability as an essential feature, now allows fisheries scientists to estimate trophic level biomasses with uncertainty assessment. Fisheries scientists now have a better idea of the biomasses that the MHI ecosystem can support, and they can utilize this information to help ground truth carrying capacity estimates for the groundfish fishery. Since the boxes in Fig. 3.5, 3.6, and 3.7 representing the various combinations of ecological assumptions mostly overlap, scientists can essentially choose any of them to estimate trophic level biomass in a data-limited case. However if this were a data-rich situation and fisheries scientists knew the ecological assumptions, they should choose the scenario that is most appropriate for their region.

The results of the simulations demonstrate the need for incorporating variation in the modeling process, especially in data-limited cases. When we allowed the parameters (i.e., NPP, transfer efficiency, and maximum expected lifespan) in the trophic pyramid food web model to vary, we obtained a large range of potential trophic level biomasses from the simulations. We recommend that scientists use the scenarios that incorporate variation, because otherwise the model could potentially not encapsulate all potential biomass values. There is a higher probability that the true biomass of some species might fall above or below the estimated value if a single point estimate was used instead. Especially in data-limited cases, fisheries scientists should always model variability since little ecological and biological information is known about the system. Doing this can help avoid disastrous situations (e.g., population and fisheries collapse). From there, depending how conservative the fisheries management bodies are, fisheries scientists can advise catch limits where the catch cannot exceed the maximum estimated trophic level biomass or even half of that. If the goal is conservation, such as in places with multiple endangered species, managers can error on side of caution and pick lower bounds.

Another reason scientists should want to incorporate variability is that the environment is not static. Even though it might appear on average to be acceptable to use a fixed NPP or transfer efficiency, there is year-to-year variability in everything from the input to the amount of energy that flows. It is well documented and understood that marine primary production changes seasonally and annually. Most notably within the past few decades, research has also examined how climate change has and will impact primary production (Gregg et al. 2003) and species composition (Walther et al. 2002). Research also shows that transfer efficiency varies by ecological system, geographic location, trophic level, and species composition (Baird et al. 2004, Barnes et al. 2010, Libralato et al. 2008, San Martin et al. 2006b). Condon et al. (2011) even found that the

presence of jellyfish blooms, which have been increasing worldwide, have been found to restrict the transfer of energy to higher trophic levels. Therefore, we want models that are adjustable, especially in context of data-limited fisheries. Error bounds give a more realistic picture of what is happening in the system.

In addition, when summarizing data sets it is common practice to record only the mean and standard deviation of the given data set. However, this proves problematic for other researchers needing to draw on data from multiple sources as these two summary statistics alone are insufficient to produce robust models. When choosing what to archive and make public, we argue that more information needs to be included in the databases (e.g., approximate distributions and parameter estimates). Knowing the specific distributions increases the accuracy of simulations, leading to more useful models.

Furthermore, we acknowledge that our NPP uncertainty estimate is underestimated due in part to the incomplete data set gathered by the SeaWiFS satellite. The satellite has missing information due to cloud cover, and the data set gathered has a shorter time series than the original time it was in operation due to parts malfunctioning. As a result, our model does not account for all of the NPP that is produced in the ocean. Additionally, the time series model (i.e., SARIMA) we used was simplistic. In future work on this topic, a more complicated time series approach will be utilized, as well as more specific estimates of NPP by season and by year.

The results of our simulations gives insight into the rigor of information we obtain when ecological concepts are incorporated into fisheries models and additionally how this information can be applied in fisheries assessments. Estimating trophic level biomass in data-limited scenarios is feasible, and our simulations yield a plausible range of biomass values. The multiple scenarios provide different options for tackling the same problem,

and fisheries scientists can choose between them based on their knowledge of the system. The general ecosystem information provided from the results of the trophic pyramid food web model can help ground truth data-limited single-species estimates by giving an estimate of the biomass that the ecosystem can support. We therefore suggest using deterministic food web models results with caution and encourage future development on food web models to improve their usefulness in data-limited scenarios.

Chapter 4

Ecosystem knowledge in Bayesian surplus production models–what can it tell us?

4.1 Introduction

Ecosystem models make use of different types of data than traditional stock assessment models and have the ability to describe a wide range of environmental states. While many ecosystem models do not provide the kind of results applicable to fisheries management, they may be able to provide valuable information on data-limited systems such as an estimate of the biomass available at an ecosystem-level. For example in Chapter 3: "Untangling uncertainty in food web models", the trophic pyramid food web model I developed cannot tell the commercial fishing operations exactly how many fish to catch, but the model can inform fisheries management on a realistic carrying capacity for the ecosystem, which may be useful in determining catch limits at a large scale. Other existing ecosystem models are complex and require similar data types to those used in

traditional stock assessment models. Often, the data required to run these ecosystem models far exceed those required to run stock assessment models. We theorize that in some situations combining a simpler ecosystem model that uses readily available data with data-limited stock assessment models may improve stock assessment reliability.

Data-limited assessment models struggle due to limited data. When species are aggregated and modeled at the aggregate level (i.e., as a complex) in data-limited assessments, they are assumed to share similar life history traits. When this assumption is not met, inaccurate predictions of biomass can result, leading to the setting of unsustainable levels of catch for some or all of the species in the complex. We hypothesize that one way to provide more information in data-limited assessments is to link ecosystem models with the data-limited stock assessment models with ecosystem models.

Bayesian methods may incorporate additional expert opinion and information in the form of prior distributions placed on some or all of the parameters (McAllister and Kirkwood 1998). Prior information is valuable since many fish stocks often contain little information about the key parameters found in population dynamics models. In conventional stock assessment approaches, uncertainties in the parameters are often ignored and point estimates or assumed values for these parameters are plugged into equations. However, values for such parameters may be similar among ecologically and taxonomically similar populations or may be assumed to follow a distribution across these populations. Parameter uncertainty could therefore be incorporated into Bayesian stock assessment in the form of prior probability distributions, alleviating data limitations (McAllister and Kirkwood 1998). Using ecosystem information as prior information to inform the posterior distribution in stock assessment models may provide an alternative way of estimating some population dynamic parameters.

In this study, we investigated the feasibility of incorporating ecosystem knowledge into existing single-species models, specifically for data-limited fisheries. We created two Bayesian models: a Bayesian surplus production model (McAllister and Ianelli 1997) and a Bayesian surplus production model that includes ecological information (See Chapter 3). We then used a trophic pyramid food web model to estimate the distribution of trophic level biomass for the ecosystem. Finally, within a Bayesian framework, we linked that information with the carrying capacities from multiple single-species models. By combining these approaches, we investigated if the indirect use of ecological information can help inform the estimation of the carrying capacity in stock assessment models.

4.2 Methods

4.2.1 Case Study

We selected a data-limited fishery in the main Hawaiian Islands (MHI) for our case study. The Hawaii bottomfish fishery complex uses traditional deep handline capture methods for commercial and recreational harvest of the thirteen species of snappers, jacks, and groupers inhabiting the Hawaiian Archipelago. The species within this complex occupy different ecological niches, including both shallow- and deep-water habitats. The subset show in Table 4.1 of the Hawaii bottomfish complex were initially believed to have similar life history traits and distributions. These species have been clustered together into their own complex called the Hawaii Deep7 Bottomfish and have been the focus of fisheries management for the past few assessments (see Langseth et al. 2018). However, the information known about these species is very limited. Scientists assessing this fishery are concerned that life history traits (e.g., intrinsic growth rate, carrying capacity, etc.) vary drastically amongst the Deep7 Bottomfish complex, which could

detrimentally impact the legitimacy of the projected biomass for one or all of the bottomfish species. Nonetheless, the last few published stock assessments used a single production model to determine the impacts of fishing for all seven species in the Hawaii Deep7 Bottomfish complex (Brodziak et al. 2009, Langseth et al. 2018).

Table 4.1: Common, Hawaiian, and Scientific names of the Hawaii Deep7 Bottomfish complex

Common name	Hawaiian Name	Scientific Name
Sea bass	Hapuupuu	*Hyporthodus quernus*
Snapper	Kalekale	*Pristipomoides sieboldii*
Pink snapper	Opakapaka	*Pristipomoides filamentosus*
Squirrelfish snapper	Ehu	*Etelis carbunculus*
Longtail snapper	Onaga	*Etelis coruscans*
Silver jaw jobfish	Lehi	*Aphareus rutilans*
Snapper	Gindai	*Pristipomoides zonatus*

4.2.2 Data from the Main Hawaiian Islands Deep7 Bottomfish Complex 2018 Assessment

Catch and CPUE Data

We utilized fishery catch and standardized catch-per-unit-effort (CPUE) indices data from the stock assessment on the Main Hawaiian Islands Deep7 Bottomfish Complex in 2018 with catch projections through 2022 (Langseth et al. 2018) in our analysis. The assessment listed species-specific total reported and unreported catch from 1949-2016 and aggregated CPUE data from 1948-2015. To keep a consistent time line, we only used catch and CPUE data that fell between 1949-2015 (Fig. 4.1).

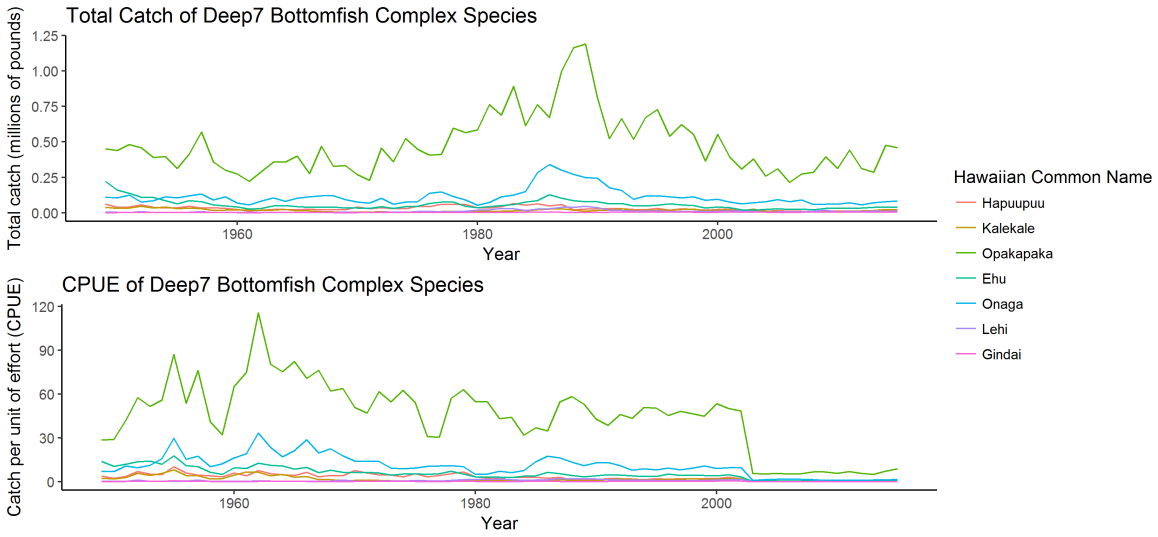

Figure 4.1: Time series of total catch and CPUE for years 1949-2015 for each of the seven Bottomfish Species in the Deep7 Bottomfish Complex using data from (Langseth et al. 2018). In the bottom panel that depicts the time series of CPUE, effort was defined as the number of days spent fishing in the years prior to 2002, while in 2002 and later effort was defined as the number of hours spent fishing. Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

In order to obtain species-specific estimates of CPUE per year ($CPUE_{i,t}$) from the aggregated CPUE data ($CPUE^*_t$) from (Langseth et al. 2018), we assumed that catch in year t for species i is proportional to abundance (Eq. 4.1). In Eq. 4.1, $C_{i,t}$ is the catch of species i in year t for species i where $i = 1, \ldots, n$ (upper panel of Fig. 4.1). In this initial work, we assume independence across years. The proportion shown in Eq. 4.1 scales the aggregated CPUE data to obtain species-specific estimates of CPUE per year (i.e., $CPUE_{i,t}$) (lower panel of Fig. 4.1).

$$CPUE_{i,t} = \frac{C_{i,t}}{\sum_{j=1}^{n} C_{j,t}} * CPUE^*_t \tag{4.1}$$

4.2.3 Logistic Surplus Production Models

We fitted logistic (Schaefer) surplus production models with relative abundance indices as our biomass dynamics models (Eq. 4.2) (Schaefer 1954). In Eq. 4.2, the *ith* species' biomass in year t ($B_{i,t}$ for $i = 1, \ldots, n$) is dependent on the previous year's biomass ($B_{i,t-1}$), the species' intrinsic growth rate (r_i), the species' carrying capacity (K_i), and the previous year's fishery catch ($C_{i,t-1}$). To clarify, $B_{i,t}$ and K_i is a scalar with dimension i. We assume that $r_i = r_j \ \forall \ i, j$. Thus for the sake of brevity, we will refer to the intrinsic growth rate simply as r for all species. However, the value of K_i is unique for species i. We also assume independence across years. In year zero (i.e., 1949), we set the estimated biomass equal to the virgin unfished biomass K_i. The two parameters that require estimation in Eq. (4.2) are r, where $0 < r < 1$ and K_i, where $K_i > 0$. See Table 4.2 for defined variables and parameters definitions within Eq. (4.2).

$$B_{i,t} = B_{i,t-1} + r * B_{i,t-1}\left(1 - \frac{B_{i,t-1}}{K_i}\right) - C_{i,t-1} \tag{4.2}$$

Table 4.2: Definition of notation in Eq. (4.2) and scientific units used

Terms	Description	Units
$B_{i,t}$	Predicted biomass in year t where $t \in \{1950, \ldots, 2015\}$ for species i where $i = 1, \ldots, n$	millions of lbs
$B_{i,t-1}$	Predicted biomass in year $t-1$ for species i	millions of lbs
r	Intrinsic growth rate for species (same $\forall$ species)	
K_i	Carrying capacity for species i	millions of lbs
$C_{i,t-1}$	Total fishery catch of species i in year $t-1$	millions of lbs

Catch per Unit Effort (CPUE)

We use the Langseth et al. (2018) definitions of catch per unit effort (CPUE, $CPUE_{i,t}$) and effort (E_t). CPUE ($CPUE_{i,t}$) is defined as the total weight caught in pounds for species i in year t divided by the unit of effort in year t (Eq. 4.3). Effort (E_t) is defined

as the number of days fishing in each year prior to 2002 ($E_{t<2002}$) and the number of hours fishing in each year since 2002 ($E_{t>=2002}$). Catch is recorded daily. Both catch and effort are aggregated to obtain yearly estimates. Effort is assumed to be the same for all species i. We assume independence across years. Table 4.3 describes notation and units for Eq. 4.3.

$$CPUE_{i,t} = C_{i,t}/E_t \quad \text{where } E_t = \begin{cases} E_{t<2002} & t < 2002 \\ E_{t>=2002} & t >= 2002 \end{cases} \tag{4.3}$$

Table 4.3: Definition of notation in Eq. (4.3) and scientific units used

Terms	Description	Units
$CPUE_{i,t}$	CPUE in year t for species i	millions of lbs/time spent fishing
$C_{i,t}$	Total catch of species i in year t	millions of lbs
$E_{t<2002}$	Effort in each year t if $t < 2002$ where $\in \{1949, \ldots, 2001\}$	number of days spent fishing
$E_{t>=2002}$	Effort in each year t if $t >= 2002$ where $\in \{2002, \ldots, 2015\}$	number of hours spent fishing

Catchability

Catchability is defined as the number of fish caught per fish available per unit of effort and per time unit. The catchability coefficient is defined as q_i for species i. We use Hilborn and Walters (1992) assumption that the catchability coefficient q_i is time invariant. We estimate q_i (i.e., $\widehat{q_i}$) in terms of $CPUE_{i,t}$ and $B_{i,t}$ in Eq. 4.4, where data are available on the total catch per species i in year t ($C_{i,t}$) and the observed relative abundance indices ($CPUE_{i,t}$) for species i in year t for years $1949 - 2015$. Since we have two measures of effort, we have two estimates of catchability: one for prior to 2002 ($\widehat{q_{i,t<2002}}$) and one for after 2002 ($\widehat{q_{i,t>=2002}}$). We assume independence across years and let the total number of years in each time frame of the data set be represented as $m_{t<2002}$

and $m_{t>=2002}$ respectively.

$$\widehat{q_{i,t}} = CPUE_{i,t}/B_{i,t} = CPUE_{i,t}/\left(B_{i,t-1} + r_i * B_{i,t-1}\left(1 - \frac{B_{i,t-1}}{K_i}\right) - C_{i,t-1}\right) \quad (4.4)$$

$$\widehat{q_{i,t<2002}} = e^{\frac{\sum_{t=1949}^{2001} ln(\widehat{q_{i,t}})}{m_{t<2002}}} \quad (4.5)$$

$$\widehat{q_{i,t>=2002}} = e^{\frac{\sum_{t=2002}^{2015} ln(\widehat{q_{i,t}})}{m_{t>=2002}}} \quad (4.6)$$

Regression methods

Using the relationship between Eq. 4.2 and 4.5, we can substitute the predicted biomass $B_{i,t}$ in Eq. 4.2 with $B_{i,t<2002} = \frac{CPUE_{i,t}}{\widehat{q_{i,t<2002}}}$ and $B_{i,t>=2002} = \frac{CPUE_{i,t}}{\widehat{q_{i,t>=2002}}}$. We outline the theory below discussed in Hilborn and Walters (1992). For the sake of brevity, we will not distinguish between the two estimates of the catchability coefficient in the following equations in this section and instead will just write $\widehat{q_i}$ to denote the estimate of q_i.

From Eq. 4.2

$$B_{i,t} = B_{i,t-1} + r * B_{i,t-1}\left(1 - \frac{B_{i,t-1}}{K_i}\right) - C_{i,t-1}$$

Substituting $B_{i,t} = \frac{CPUE_{i,t}}{\widehat{q_i}}$ gives

$$\frac{CPUE_{i,t}}{\widehat{q_i}} = \frac{CPUE_{i,t-1}}{\widehat{q_i}} + r * \frac{CPUE_{i,t-1}}{\widehat{q_i}}\left(1 - \frac{CPUE_{i,t-1}}{\widehat{q_i}} * \frac{1}{K_i}\right) - CPUE_{i,t-1} * E_{t-1}$$

This can be simplified as follows:

$$CPUE_{i,t} = CPUE_{i,t-1} + r * CPUE_{i,t-1}\left(1 - \frac{CPUE_{i,t-1}}{\widehat{q_i} * K_i}\right) - CPUE_{i,t-1} * \widehat{q_i} * E_{t-1}$$
$$CPUE_{i,t} = CPUE_{i,t-1} + r * CPUE_{i,t-1} - \frac{r * CPUE_{i,t-1}^2}{\widehat{q_i} * K_i} - CPUE_{i,t-1} * \widehat{q_i} * E_{t-1}$$
$$CPUE_{i,t} = CPUE_{i,t-1}\left(1 + r - \frac{r * CPUE_{i,t-1}}{\widehat{q_i} * K_i} - \widehat{q_i} * E_{t-1}\right)$$
$$\frac{CPUE_{i,t}}{CPUE_{i,t-1}} = 1 + r - \frac{r * CPUE_{i,t-1}}{\widehat{q_i} * K_i} - \widehat{q_i} * E_{t-1}$$
$$\frac{CPUE_{i,t}}{CPUE_{i,t-1}} - 1 = r - \frac{r}{\widehat{q_i} * K_i} * CPUE_{i,t-1} - \widehat{q_i} * E_{t-1}$$

Thus in Eq. 4.7, $\frac{CPUE_{i,t}}{CPUE_{i,t-1}} - 1$ is the dependent variable, the independent variables are $CPUE_{i,t-1}$ and E_{t-1}, and the parameters are $r, -\frac{r}{\widehat{q_i} * K_i}$, and $-\widehat{q_i}$ where r is the intercept and $-\frac{r}{\widehat{q_i} * K_i}$ and $-\widehat{q_i}$ are the slopes. Hilborn and Walters (1992) noted however that this method does not always provide reliable parameter estimates and is biased. Thus, including prior information on the parameters can improve their estimation.

$$\frac{CPUE_{i,t}}{CPUE_{i,t-1}} - 1 = r - \frac{r}{\widehat{q_i} * K_i} * CPUE_{i,t-1} - \widehat{q_i} * E_{t-1} \tag{4.7}$$

Optimization for r and K_i

We re-arrange the terms in Eq. 4.7 and treat the quantity defined in Eq. 4.8 as a random variable $w_{i,t}$. We define $w_{i,t}$ as

$$w_{i,t} = \frac{CPUE_{i,t}}{CPUE_{i,t-1}} - 1 - r - \frac{r}{\widehat{q_i} * K_i} * CPUE_{i,t-1} + \widehat{q_i} * E_{t-1} \tag{4.8}$$

We assume that the log of the random variable $w_{i,t}$ follows a $Normal(0, \sigma_i^2)$. Therefore, we use the maximum likelihood estimate (MLE) of σ_i^2 based on a Normal distribution. Before solving for σ_i^2, we define the sum of squares for each species i as SSQ_i as

the difference between the predicted and observed CPUE.

$$SSQ_i = \sum_{t=1949}^{2015} [ln(\widehat{q_i} * B_{i,t}) - ln(CPUE_{i,t})]^2$$

We let the total number of years in the data set be represented as m and take the square root of the sum of squares for each species divided by the total number of years in the data set.

$$\widehat{\sigma_i} = \sqrt{SSQ_i/m}$$

We assume $ln(w_{i,t}) \sim N(0, \sigma_i^2)$. We plug in the value of zero for μ_i and the MLE estimate of $\widehat{\sigma_i^2}$ into the Normal probability distribution function (PDF) to evaluate the negative log likelihood at MLE $\widehat{\sigma_i^2}$.

$$L(\mu_i = 0, \sigma_i^2 = \widehat{\sigma_i^2} | w_{i,1949}, \ldots, w_{i,2015}) = \prod_{t=1949}^{2015} \frac{1}{\sqrt{2\pi \widehat{\sigma_i}^2}} e^{-\frac{ln(w_{i,t}^2)}{2\widehat{\sigma_i}^2}} \tag{4.9}$$

$$-logL(\mu_i = 0, \sigma_i^2 = \widehat{\sigma_i^2}) = -\sum_{t=1949}^{2015} \left[-(m/2) * ln(2\pi) - (m/2) * ln(\widehat{\sigma_i}^2) - \frac{ln(w_{i,t}^2)}{2\widehat{\sigma_i}^2} \right] \tag{4.10}$$

The optimal values of r and K_i for species i minimizes the log likelihood. We used the limited-memory modification of the Broyden, Fletcher, Goldfarb and Shannon quasi-Newton method for optimization (Shanno 1970, Byrd et al. 1995). This method allows box constraints meaning the parameters being optimized (i.e., r and K_i) can be given lower and/or upper bounds.

These surplus production models work only if fishing catch ($C_{i,t}$) is causing the change in the biomass index value ($B_{i,t}$). When this is the case, then there is enough information

in the model to estimate r and K_i. If the change in catch per unit effort ($CPUE_{i,t}$) is proportional to abundance, then the model has no information to estimate r and K_i, because the changes in abundance are being dominated by the recruitment variability. $w_{i,t}$ will simply follow a random distribution.

4.2.4 Bayesian model

We set up two Bayesian frameworks in our analysis: a Bayesian surplus production model (McAllister and Ianelli 1997) and a Bayesian surplus production model that incorporates ecological information (See Chapter 3: "Untangling uncertainty in food web models"). The Bayesian surplus production models were applied on the individual populations and not the aggregated fishery complex. The posterior distribution for both cases are as follows:

$$\begin{aligned}\pi(\theta_i|w_{i,1949},\ldots,w_{i,2015}) &\propto \pi(\theta_i)L(\theta_i|w_{i,1949},\ldots,w_{i,2015}) \\ &\propto \pi(r)\pi(K_i)L(r,K_i|w_{i,1949},\ldots,w_{i,2015})\end{aligned}$$

The posterior distribution $\pi(\theta_i|w_{i,t})$ is proportional to the prior information $\pi(\theta_i)$ and the sampling information $L(\theta_i|w_{i,t})$. The parameter vector θ_i for species i ($i = 1,\ldots,n$) can be partitioned as $\theta_i = (r, K_i, q_{i,t<2002}, q_{i,t>=2002})$ where r and K_i are the parameters of interest and $q_{i,t<2002}$ and $q_{i,t>=2002}$ are the nuisance parameters. We place prior information on the intrinsic growth rate (r) and carrying capacity (K_i). The likelihood for the natural log of the random variable $ln(w_{i,t})$ is outlined in Eq. 4.9.

In both of the Bayesian frameworks, we ran 10,000 Monte Carlo simulations per species with the sampling/importance resampling algorithm (SIR) algorithm. Monte

Carlo simulations with the SIR algorithm are a simple and versatile method for drawing a sample approximately from a target distribution (McAllister and Ianelli 1997, Givens and Hoeting 2012). We based our approach off of the methods described in McAllister and Ianelli (1997). Samples from the posterior distribution of the logistic (Schaefer) production model parameters were simulated in order to make inferences.

4.2.5 Prior Information

We place prior information from Langseth et al. (2018) on the intrinsic growth rate (r) and carrying capacity (K_i) for species i where $i = 1, \ldots, n$. In Langseth et al. (2018), the intrinsic growth rate (r) and aggregate carrying capacity for all species ($K^* = \sum_{i=1}^{n} K_i$) parameters are assessed to follow Lognormal distributions with mean values (0.11 and 27.55) and standard deviation (0.028 and 9.69) for r and K^* respectively. We used the mean and variance equations of a Lognormal distribution to calculate the values of the parameters μ and σ^2 for the Lognormal distribution (See Appendix C for more details on the Lognormal distribution). In the prior, we assume r and K_i for species i where $i = 1, \ldots, n$ are independent from each other. While the prior information from Langseth et al. (2018) for the intrinsic growth rate (r) can be directly applied to individual surplus production models, the prior distribution for the carrying capacity (K^*) needed to be adjusted, as described in the next paragraph, since it represents the carrying capacity estimate for the sum of the seven groundfish species. Figure 4.2 visualizes the distribution of random draws from the prior distribution on the intrinsic growth rate r.

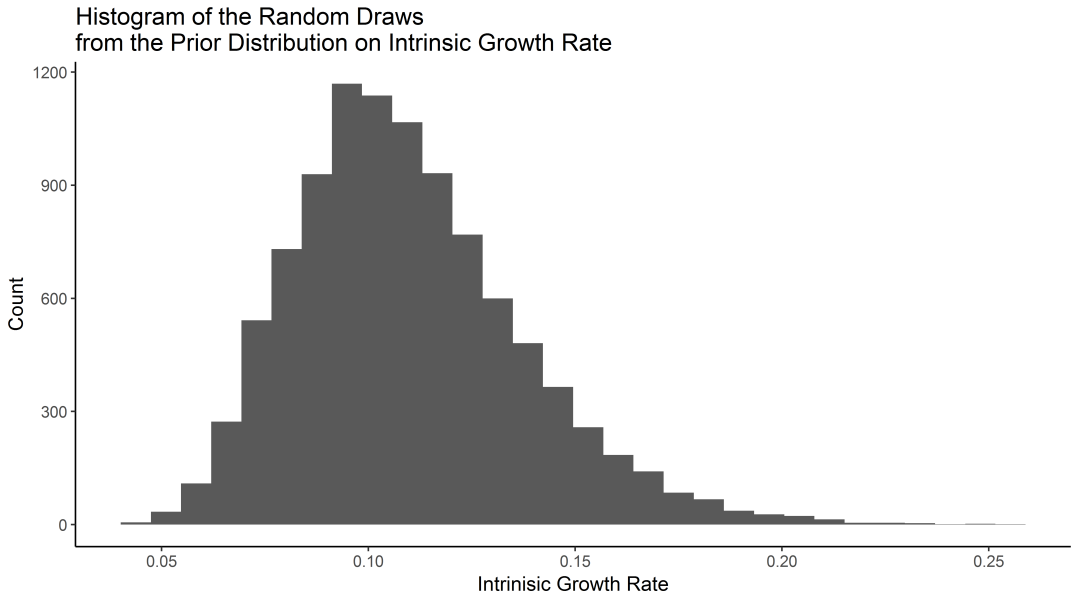

Figure 4.2: Histogram of random values drawn from the Lognormal prior distribution on the intrinsic growth rate r. Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

In order to put a prior distribution on the carrying capacities for each of the species, we created a hierarchical framework using hyperparameters and hyperpriors. We let the individual carrying capacities $K_1, ..., K_n$ for species $i = 1, \ldots, n$ come from a Multinomial distribution. We placed prior distributions on the parameters N and the vector $\boldsymbol{\delta}$, where N is the total tonnage (integer) and the ith element of the vector $\boldsymbol{\delta}$ is the proportion for species i. We used the proportion of each rounded values of the Lognormal distribution of the aggregate carrying capacity (K^*) be the hyperprior for the number of trials parameter (N) in the Multinomial distribution. We used our assumption that catch $C_{i,t}$ was proportional to abundance and calculated the average proportion of abundance for all species from the aggregate carrying capacity (i.e., 27.55 million lbs) to parameterize a Dirichlet distribution proportion parameter $\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is a vector and the ith element of the vector is the proportion for species i. We define the total number of years as m.

$$\boldsymbol{\alpha} = \frac{\sum_{t=1949}^{2015} \frac{C_{j,t}}{\sum_{j=1}^{n} C_{i,t}}}{m}$$

We then set $\boldsymbol{\alpha}$ as the vector of hyperparameter values driving the Dirichlet distribution for the proportion parameter $\boldsymbol{\delta}$ in the Multinomial distribution. Figure 4.3 visualizes the prior distribution of the carrying capacity K_i for species $i = 1, \ldots, n$.

$$r \sim Lognormal(\mu_r, \sigma_r^2)$$

$$(K_1, \ldots, K_n) \sim Multinomial(N = round(K^*), \boldsymbol{\delta})$$

$$K^* \sim Lognormal(\mu_{K^*}, \sigma_{K^*}^2)$$

$$\boldsymbol{\delta} \sim Dirichlet(\boldsymbol{\alpha})$$

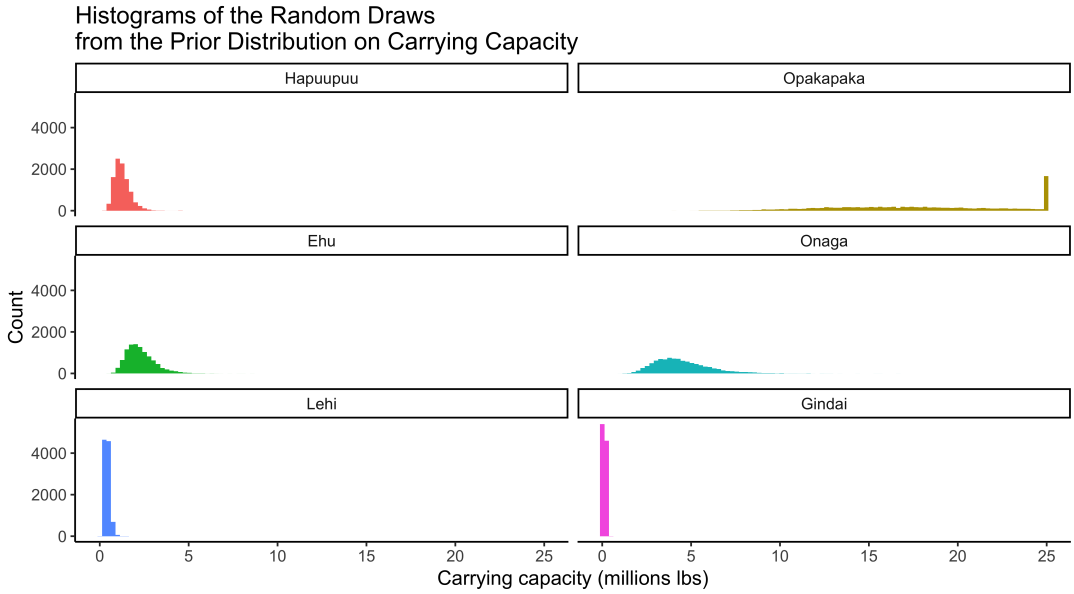

Figure 4.3: Histogram of random values that come from the prior distribution on the carrying capacity K_i for species $i = 1, \ldots, 6$. We adjusted the x-axis for visualization purposes and placed all the outlying points of Opakapaka in the far right bin. The original range extends out further. Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

Bayesian Surplus Production Model

Before fitting individual logistic (Schaefer) surplus production models, we clustered the bottomfish species by trophic level. We used the Fishbase database (`http://www.fishbase.org/`) of Froese and Pauly (2017) to obtain information on trophic level. Out of the seven species, only Kalekale (*Pristipomoides sieboldii*) occupies the third trophic level, with the remaining six occupying trophic level four. Therefore, we focused only on the six species that occupied trophic level four since our main interest in this study was to test if the carrying capacity parameter could be "constrained" by ecological data.

Thus, the current set of Bayesian models consisted of individual Bayesian logistic surplus production models for the six species found at trophic level four. We set the values for the parameters to the fixed values described below for the prior distributions.

$$
\begin{aligned}
r &\sim Lognormal(\mu_r = -2.24, \sigma_r^2 = 0.06) \\
K_1, \ldots, K_6 &\sim Multinomial(N = round(K^*), \boldsymbol{\delta}) \\
K^* &\sim Lognormal(\mu_K = 3.26, \sigma_K^2 = 0.17) \\
\boldsymbol{\delta} &\sim Dirichlet(\boldsymbol{\alpha} = (0.0446, 0.6697, 0.0811, 0.1590, 0.01460, 0.0046))
\end{aligned}
$$

Bayesian Surplus Production Model with Environmental Link

The second set of Bayesian surplus models uses the same Bayesian logistic surplus production models described earlier, but add on a layer of complexity. These models also estimates the relative abundance of the six species found at trophic level four, but in addition these Bayesian models include ecosystem-based information intended to limit the upper bounds of the carrying capacity. We applied an ecosystem-based constraint

that the sum of the production model's carrying capacity (i.e., $K_n = \sum_{i=1}^{n} K_i$) cannot exceed the estimated trophic level biomass. In other words, the individual population's carrying capacity is limited by the ecological carrying capacity.

We used a previously developed trophic pyramid food web model as presented in Chapter 3: "Untangling uncertainty in food web models" to estimate trophic level biomass (γ_h) from net primary production (ν), transfer efficiencies (τ_2, τ_3, and τ_4), and the lifespan of organisms in trophic level h (λ_h) (Eq. 4.11). See Table 4.4 for variable and parameter definitions within Eq. (4.11). While the original food web model analysis included multiple scenarios, we decided to only consider the scenario had fixed net primary production, random transfer efficiency, and random expected lifespan as this case accounted for more of the natural variation in the ecosystem than our other scenarios. Since our study investigates only the six species that occupy trophic level four, we only focus on trophic level four biomass (γ_4) here (i.e., .only on $h = 4$ in Eq. 4.11).

$$\gamma_h = \nu * 9 * \left(\prod_{j=2}^{h} \tau_j \right) \lambda_h \quad \text{for} \quad h \in \{2, 3, 4\} \tag{4.11}$$

Table 4.4: Definition of notation in Eq. (4.11) and scientific units used

Terms	Description	Units
h	Trophic level in MHI, where $h = 2, 3, 4$	
γ_h	Trophic level biomass at trophic level h	millions of lbs
ν	Net primary production (NPP)	millions of lbs C/year
9	Carbon to wet weight conversion ratio	
τ_h	Transfer efficiency between trophic level $h-1$ and h	
λ_h	Lifespan of a species found at trophic level h	years

We assume that the random variables NPP (ν), transfer efficiencies (τ_2, τ_3, and τ_4), and lifespan (λ_4) are independent. Here we set ν to equal the mean of the SeaWiFS time series (87339861.7 metric tons C/year) within the EEZ of the MHI found by fitting a sea-

sonal autoregressive integrated moving average model (SARIMA model) to the SeaWiFS data using R v.3.4.3 (R Core Team 2017) forecast package v8.2 and function auto.arima() (Hyndman 2017, Hyndman and Khandakar 2008). In this initial work, we ignore uncertainties in the ν value. The constant 9 is used to convert from organic carbon (metric tons C) to wet weight (metric tons) (Strathmann 1967, Pauly and Christensen 1995, Chassot et al. 2010). We based the distributional assumptions of the transfer efficiencies τ_h for $h \in \{2, 3, 4\}$ on data gathered from a literature review (Chapter 2: "Tangled is the web we weave") on transfer efficiencies. We assume the transfer efficiency (τ_h) at each trophic levels $h \in \{2, 3, 4\}$ comes from the same distribution. We chose an approximate distribution (i.e., $Beta(\alpha_\tau = 1.9, \beta_\tau = 13.36)$) for the transfer efficiency data based on goodness-of-fit tests (Delignette-Muller and Dutang 2015). The Froese and Pauly (2017) FishBase data set was used to estimate the lifespan (λ_4) across all species at trophic level 4. In this database Froese and Pauly (2017) defines the term lifespan as "the maximum expected age, on average, for a species, cohort, stock, or a population in the absence of fishing. Smaller than maximum age although may be used in this sense." We chose an approximate distribution (i.e., $Lognormal(\mu_{\lambda_4} = 2.49, \sigma^2_{\lambda_4} = 0.81)$) for the lifespan data based on goodness-of-fit tests (Delignette-Muller and Dutang 2015). Details of the SARIMA model fitting and the Beta and Lognormal approximate distributions are in Appendix C.

To create the environmental link in the simulations, we included a constraint within the SIR algorithm that the sum of the carrying capacities for the six species must not exceed the trophic level four biomass ($\sum_{i=1}^{6} K_i \leq \gamma_4$). We performed this by randomly drawing a value of γ_4 and comparing it to the sum of the joint draw for the carrying capacities $[K_1, \ldots, K_6]$ from the individual Bayesian surplus production models.

4.3 Results

We can break down our methods and thus our results into 3 parts that build in complexity: a logistic surplus production model without prior information (section Logistic Surplus Production Models), a Bayesian surplus production model with the prior for the carrying capacity broken out for each of the six species at trophic level four (section Bayesian Surplus Production Model), and then Bayesian surplus production model with ecological information (section Bayesian Surplus Production Model with Environmental Link). The results of the simulations indicate that the use of priors and then specifically, trophic level information, can improve our understanding of carrying capacity (K_i) for species that occupy the higher trophic levels (Fig. 4.4). For some species i.e., Lehi and Gindai), the surplus production model without prior information gave contradicting estimates of carrying capacity in comparison to the two Bayesian models (Fig. 4.4. This indicates that the priors provided some beneficial information about the carrying capacity. In addition, the density plot shows a slight shift left in distribution for the Bayesian model that incorporated ecological information (Fig. 4.4, which means the inclusion of ecological information did shift the posterior distribution of carrying capacity K_i.

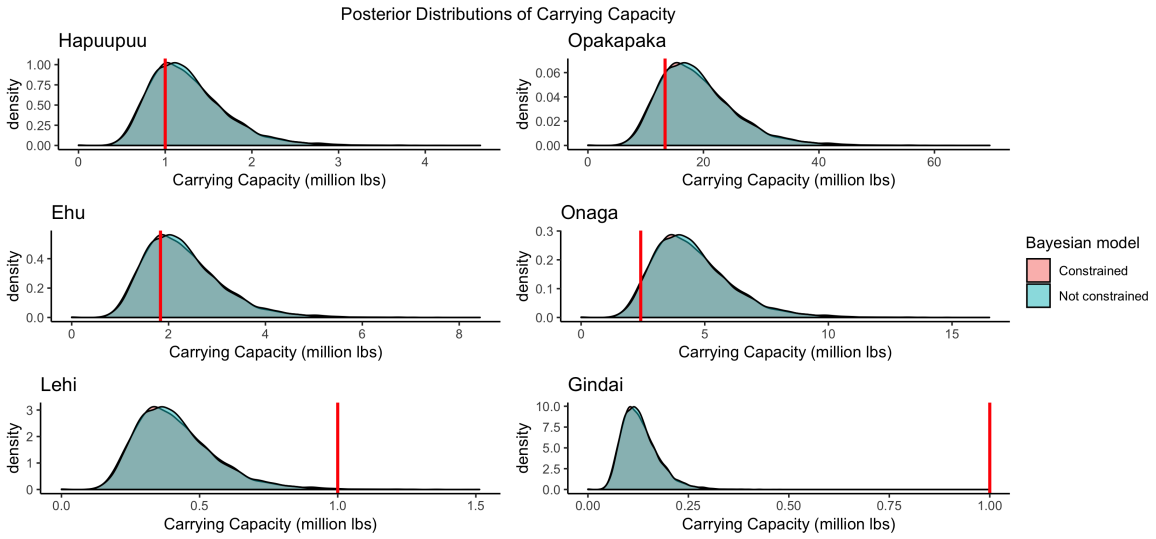

Figure 4.4: Gaussian kernal density plot of the posterior distribution for the carrying capacities (K_i) for the six species at trophic level 4. The vertical red line represents the carrying capacity predicted from the surplus production model with no prior information on it (i.e., the value of N in the Multinomial distribution). Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

The simulations also demonstrate that the use of priors and once again specifically, trophic level information, may have higher uncertainty estimates of biomass in the terminal year (i.e., 2015) (B_{2015}) for the six species that occupy the higher trophic levels since they account for more sources of variation (Fig. 4.5 and Table 4.5). Our surplus production model without prior information would give very low and almost unrealistic estimates of biomass for some species (i.e., Opakapaka, Lehi, and Onaga) (Fig. 4.5). Then for Gindai, the surplus production model without prior information gave an estimate of biomass that was much higher than the two Bayesian models (Fig. 4.5). Both Bayesian models had bimodial distributions for most species (i.e., Hapuupuu, Ehu, Onaga, and Gindai) (Fig. 4.5), but for all species the Bayesian surplus production model with the ecological information shifted the posterior distribution (Fig. 4.5 and Table 4.5). Thus the Bayesian surplus production model with the constrained SIR posterior distribu-

tion resulted in a shifted posterior distribution of the biomass predicted in the terminal year (i.e., 2015) in comparison to the Bayesian surplus production model without the constraint (Fig. 4.5 and Table 4.5).

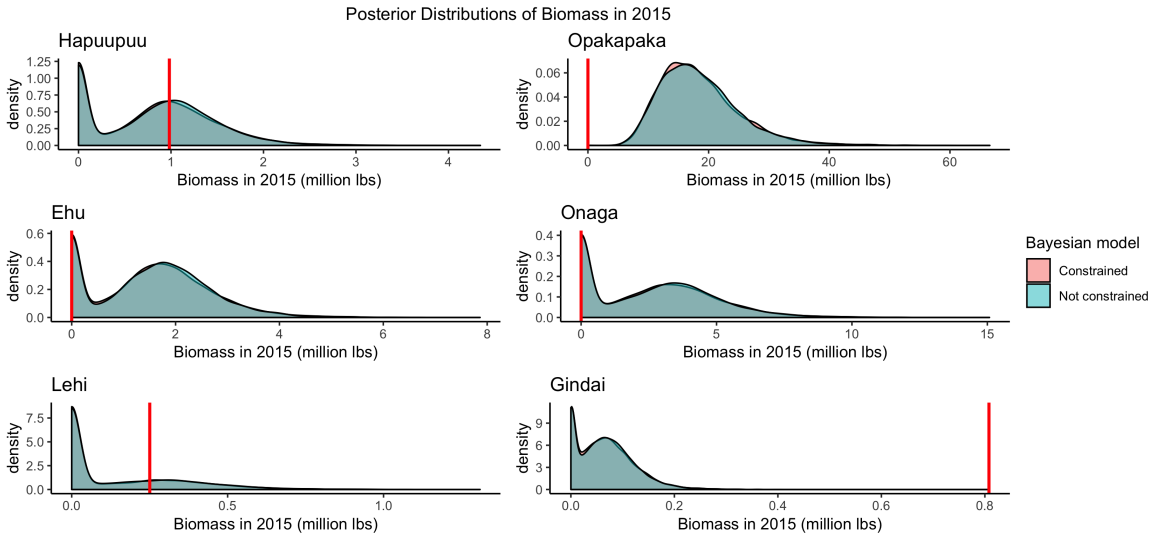

Figure 4.5: Gaussian kernal density plot of the predicted biomass values (B_{2015}) in the terminal year (i.e., year 2015) from the posterior distribution for the six species at trophic level 4. These values were calculated from both Bayesian model that included and did not include the ecological constraint in the posterior imposed within the SIR algorithm. The vertical red line represents the biomass predicted in year 2015 from the surplus production model with no prior information on it. Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

Table 4.5: Summary statistics of biomass in 2015 (B_{2015}) in millions of pounds

Species	Ecological Constraint	1st Quartile	Median	3rd Quartile	Maximum
Hapuupuu	Yes	0.000	0.831	1.226	3.846
	No	0.000	0.859	1.240	4.345
Opakapaka	Yes	13.730	17.390	21.940	59.50
	No	13.780	17.599	22.102	66.598
Ehu	Yes	0.388	1.523	2.197	6.747
	No	0.450	1.577	2.232	7.862
Onaga	Yes	0.000	2.443	4.043	13.545
	No	0.000	2.535	4.071	15.077
Lehi	Yes	0.000	0.000	0.258	1.201
	No	0.000	0.000	0.263	1.309
Gindai	Yes	0.017	0.060	0.098	0.364
	No	0.018	0.061	0.098	0.407

We also examined the depletion estimate or the ratio of the biomass in 2015 divided by the carrying capacity. A value below 1 indicates biomass is less than carrying capacity, whereas a value above 1 means more biomass is predicted than the population can realistically support. A value close to 1 means the predicted biomass is equivalent to the carrying capacity. For some species (i.e., Opakapaka, Ehu, and Onaga), the model with no prior information predicted a ratio of about zero. However, we draw different conclusions when we focus on the results from the Bayesian models (Fig. 4.6). The Bayesian models take into account more components of variation on the depletion estimate in comparison to the point estimate from the model with no prior information (Fig. 4.6). Overall, the posterior distributions on the depletion estimate were similar from both Bayesian models.

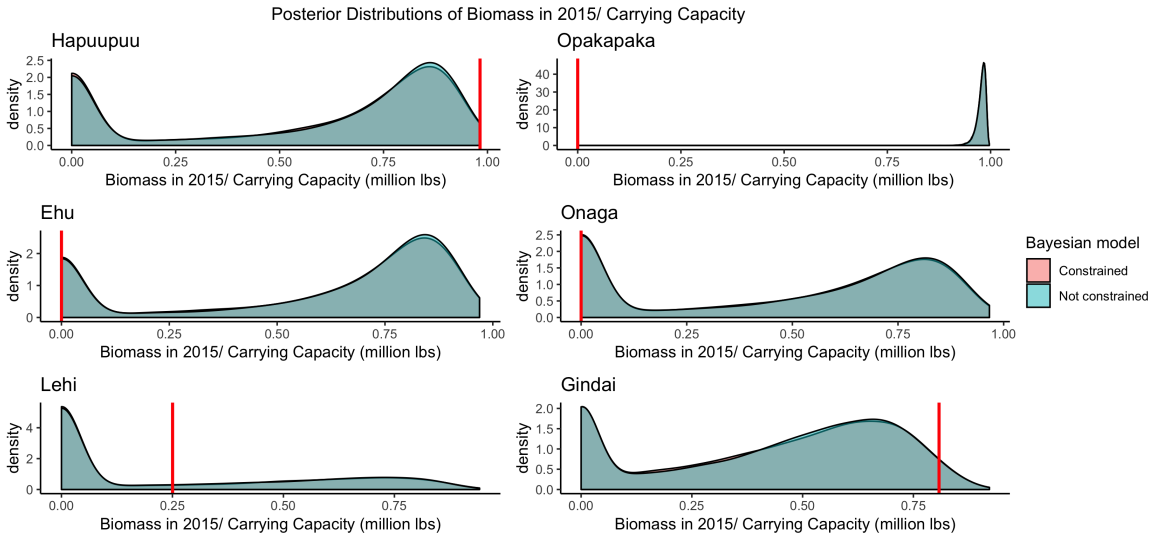

Figure 4.6: Gaussian kernal density plot of the posterior distribution of the depletion estimate or the predicted biomass values in the terminal year (i.e., year 2015) divided by the carrying capacities (i.e., $B_{i,2015}/K_i$) for the six species at trophic level 4. The vertical red line represents the biomass in year 2015 divided by the carrying capacity predicted from the surplus production model with no prior information on it. Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

We investigated the two estimates of the catchability coefficient to see how the Bayesian models would inform the estimation of the parameters. Remembering since we have two measures of effort, we have two estimates of catchability: one for prior to October 2002 ($\widehat{q_{i,t<2002}}$) and one for after October 2002 ($\widehat{q_{i,t>=2002}}$). We examined the catchability coefficients, because in the setup for the model, we define carrying capacity based on catch and have catch in the surplus production models too. In other words, we use catch twice to get at the question on carrying capacity. When we use the same set of information twice, we risk obtaining results that are reflecting patterns that are not we think they are. The results from Fig. 4.7 show that there are distinct distributions of the catchability coefficiencts for each species. This demonstrates that the Bayesian models are providing some information on these parameters, and we are not just seeing the catch ratios we defined in Eq. 4.1.

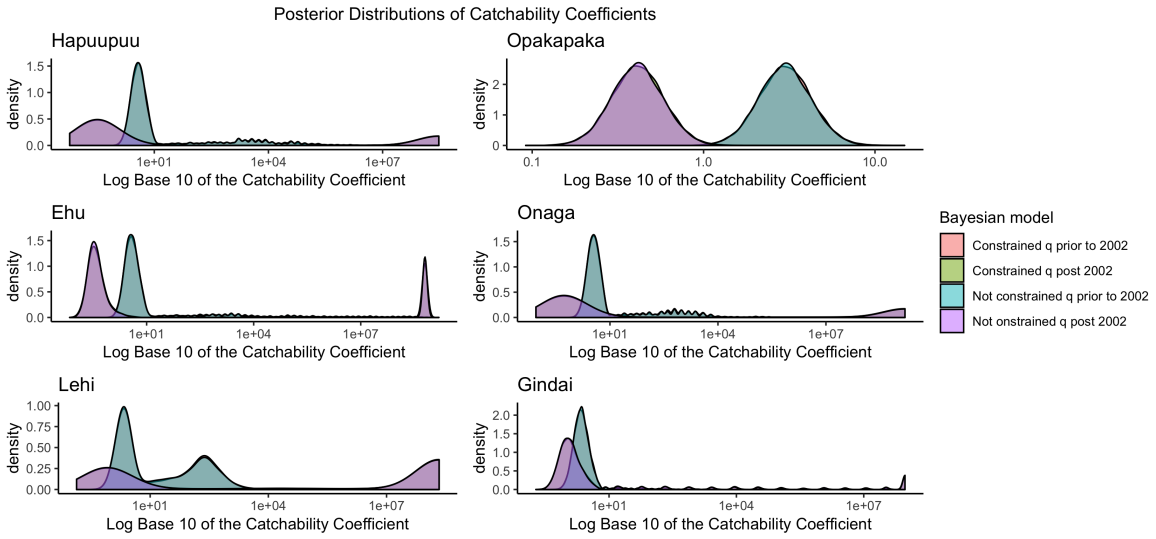

Figure 4.7: Gaussian kernal density plot of the two sets of catchability coefficients (i.e., one for prior to October 2002 ($\widehat{q_{i,t<2002}}$) and one for after October 2002 ($\widehat{q_{i,t>=2002}}$)) from each Bayesian model for each of the six species at trophic level 4. The log base 10 was applied on the x-axis to improve visualization. Plot created using R v.3.4.3 (R Core Team 2017) ggplot2 package v.2.2.1 (Wickham 2009).

While the published stock assessment for the Hawaii Deep7 Bottomfish complex used a single production model for the aggregated seven species, we created individual production models for six out of the seven species while relying on a set of underlying assumptions. By making a simple assumption that abundance was proportional to the total catch, we were able to obtain individual estimates of CPUE and thus "create individual production models". We set up a method for empirically choosing prior distributions based on other sources of information on the carrying capacity for the individual species in the complex through the use of hyperpriors. Thus for each of the six species at trophic level four, we were able to obtain species-specific prior distributions for the carrying capacity K_i (i.e., $i \in \{1, \ldots, 6\}$). The results of the simulations show five out of the six species at trophic level four have relatively more similar distributions for biomass (Fig. 4.5). However, the abundance of Opakapaka is drastically different from the others.

4.4 Discussion

One of the Deep7 Bottomfish species, Opakapaka, makes up the majority of the catch composition (Fig. 4.1). In addition, based upon the assumptions in our model our results suggests that Opakapaka has a drastically different distribution of abundance in relation to the other five species found at trophic level four. While the catch rates may be set at a level that is sustainable for Opakapaka, there is a chance that the "weaker" stocks (e.g., Hapuupuu, Ehu, Onaga, Lehi, and Gindai) in this complex could be overfished and driven to unsustainably low levels (Hastings et al. 2017).

We found that the simulation algorithms are extremely sensitive. The initial seed set can change the outcome of the simulations, which has the potential of giving inconsistent results on the effectiveness of the inclusion of ecological information in the model. We theorize this could be due to a few reasons. We hypothesize that we need to improve our understanding of ecological data. From Chapter 3: "Untangling uncertainty in food web models", we obtained a relatively large distribution of trophic level biomass. Our results could be more consistent, if we had a narrower range for the trophic level biomass. Another explanation is that this method would be better suited for data-rich fisheries. Even without the ecological knowledge, the posterior distribution for the six species at trophic level 4 gave large distributions of biomass. Future work will test this modeling approach on a data-rich fishery.

The way we split up the aggregated data has potential application in fisheries. We pulled apart and obtained species-level information when the data was originally clustered. We were able to calculate species-specific estimates of carrying capacity in contrast with the aggregated estimate currently available for the fishery through the use of hyperpriors and hyperparameters. We argue the use of hyperpriors and hyperparameters for

breaking apart the aggregate carrying capacity into its individual components can help data-limited fisheries move away from assessing the fishery as an aggregated complex to instead assessing the individual species.

The inclusion of ecological information in the Bayesian production models did shift the posterior distributions. The presence of ecological information truncated off some of the largest estimated carrying capacity and terminal (i.e., year = 2015) biomass values (Fig. 4.5). This finding supports what many ecologists and fisheries scientists believe. In theory, ecological knowledge would inform fisheries models and help ground truth the recommended sustainable removals by removing the unrealistic estimates. Our results found that incorporating ecosystem knowledge into fisheries models can influence the posterior distribution.

4.4.1 Limitations of Model

We also imposed an assumption about the starting biomass in the production model. Even though the locals have fished the Hawaiian bottomfish for hundreds of years, we assume unfished/virgin biomass in the start of the commercial fishery in year 1949. We are assuming that before modern technology, the effect of small-scale and artisanal fishing did not impact the populations enough to move the ecosystem away from its natural unfished state.

Our current carrying capacity estimates include only a subset of species. In the constrained Bayesian framework, we argued that the carrying capacity for the six bottomfish species could not exceed the trophic level four biomass. However, we acknowledge that there are more than six species at trophic level four. If more species from trophic level four were included, we may have additional information to inform reasonable carrying

capacity posterior distributions.

Appendix A

Appendix for Chapter 2: Tangled is the web we weave

A.1 Decision trees

We used regression trees with pruning, bagging, and random forests on the marine data set and trained the classifiers on 75% of the marine transfer data. We tried different tree-based methods which involve stratifying or segmenting the predictor space (i.e., transfer efficiency) into a number of simple regions. Though the simple tree-building process may produce good predictions on our training data, it is likely to over fit the data, leading to poor test set performance. Thus, a better strategy is to grow a very large tree, and then prune it in order to obtain a subtree. We also tried other tree methods, such as bagging and random forests. These methods grow multiple trees which are then combined to yield a single consensus prediction. Combining a large number of trees can often result in dramatic improvements in prediction accuracy and reduce variance, at the expense of some loss in interpretation. For bagging, we construct regression trees using 1,000 bootstrapped training datasets. Random forest builds on the idea of bagging, but

de-correlates the trees, thus leading to more reduction in variance. We build a 1,000 decision trees on bootstrapped training sample, but in the tree building process, each time a split in a tree is considered, a random sample of 2 predictors is chosen as split candidates from the full set of 3 predictors.

We selected the optimal number of number of nodes using cross validation. The best subtree with the minimized error had three nodes. The final subtree is visualized in Figure A.1 and illustrates that combinations of specific factors can lead to distinct transfer efficiencies. It shows that trophic level is the most important factor in predicting transfer efficiency and that certain trophic levels have higher and lower transfer efficiencies. It also indicates that different regions in the ocean impact transfer efficiency at the primary consumer level.

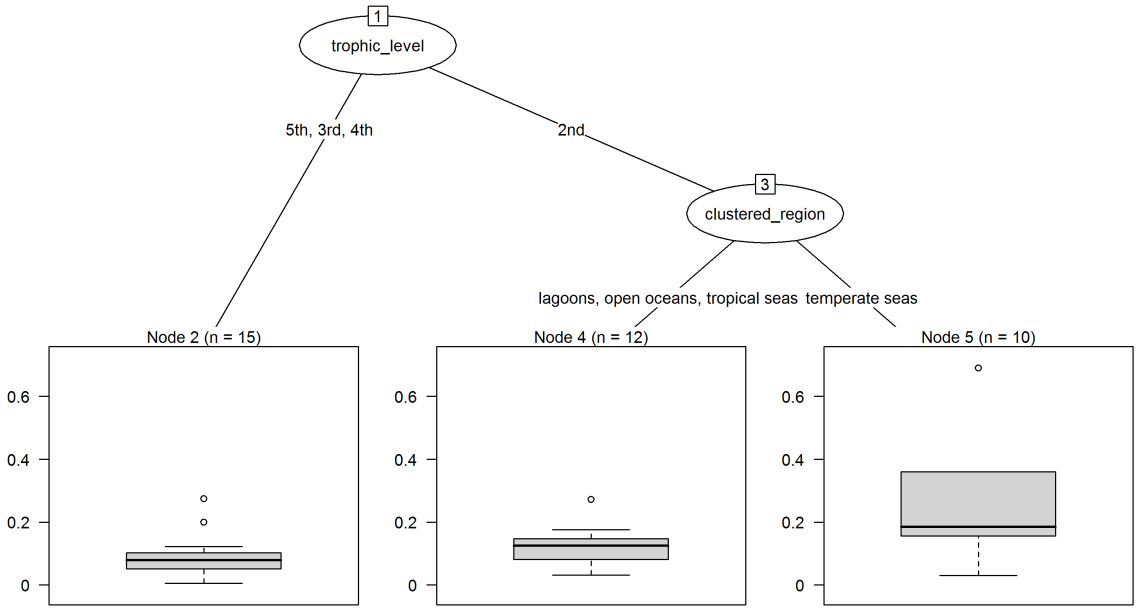

Figure A.1: Visualization of regression tree with pruning ending in three terminal nodes. The tree stratifies transfer efficiency into three segments: when the transfer efficiency applies to either secondary, tertiary, and quaternary consumers, when the transfer applies to primary consumers and the region is either a lagoon, open ocean, or tropical shelves and seas, and when the transfer efficiency applies to primary consumers and the region is temperate shelves and seas. Plot created using R v.3.4.3 (R Core Team 2017) partykit package v.1.2-2 (Hothorn and Zeileis 2015).

A.2 Fitting approximate distributions

In the Monte Carlo simulations, we randomly drew from an approximate distribution that was fitted to the marine transfer efficiency data. In order to determine an approximate distribution for the marine transfer efficiency data, we ran goodness-of-fit tests. We started off with a skewness-kurtosis plot (Figure A.2) to initially decide which distributions to consider (i.e., Beta and Gamma) (Delignette-Muller and Dutang 2015). Then we fitted individual distributions to the data using maximum likelihood estimation

and compared density plots of the fitted distributions to the histogram of the empirical distribution, a cumulative distribution (CDF) plot of both the empirical distribution and the fitted distributions, Q-Q plots, and P-P plots (Figure A.3). We chose the Beta distribution amongst the approximate distributions because, like percentages, it is defined within the range [0, 1]. Therefore, we concluded that a $Beta(\alpha = 1.90, \beta = 13.36)$ distribution was the most appropriate approximate distribution.

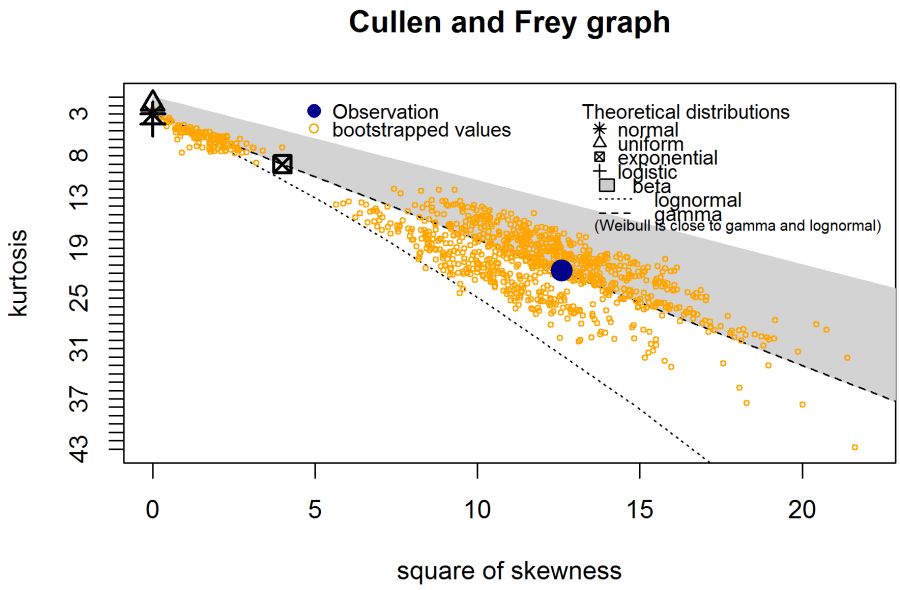

Figure A.2: Visualizes several potential continuous distributions against the marine transfer data and bootstrapped data. The figure shows it potentially follows a Beta and Gamma distribution. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

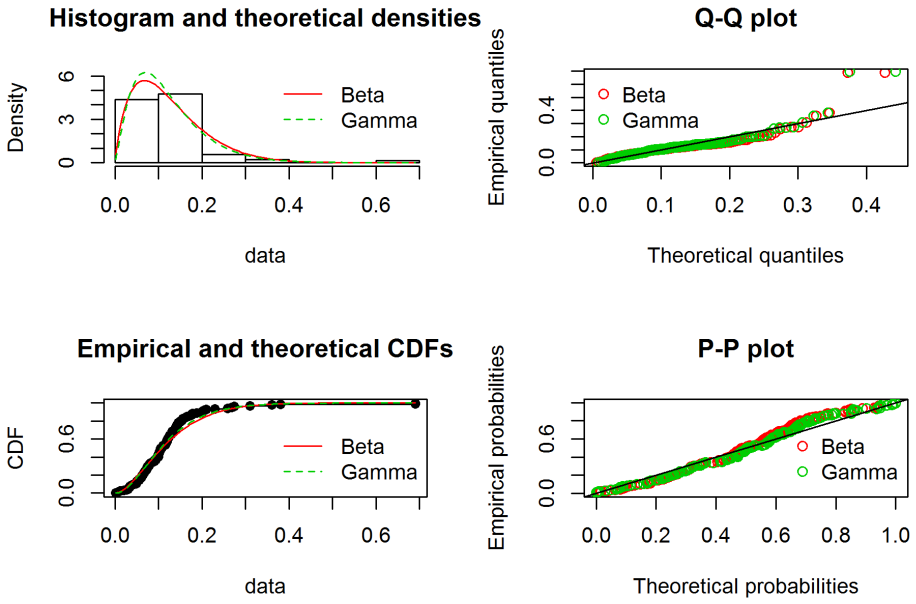

Figure A.3: Density plots of the fitted distributions to the histogram of the empirical distribution using the marine transfer efficiency data, a CDF plot of both the empirical distribution and the fitted distributions (for the Beta and Gamma choices), Q-Q plots, and P-P plots. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

A.3 SeaWiFS

We used 8-day time series Sea-viewing Wide Field-of-view Sensor (SeaWiFS) chlorophyll *a* data from 1997 to 2010 that was transformed using the Vertically Generalized Production Model (VGPM) to estimate net primary production (NPP) from chlorophyll *a* (Behrenfeld and Falkowski 1997). The SeaWiFS data were originally obtained from the Oregon State Ocean Productivity website (`http://www.science.oregonstate.edu/ocean.productivity`). They use a gap filling algorithm to populate missing pixels due to cloud coverage; however if no good data is available, pixels remain empty. We segmented the SeaWiFS data using the boundaries of the California Current and assumed a closed system at equilibrium. The data then were converted from 8 day averages in mg

C / m^2/ day per 9 km x 9 km pixel into total metric tons of Carbon per year total across the entire region. Although from 1997 to 2010 SeaWiFS collected data every 8 days per 9 km x 9 km pixel, the time series had gaps due to machine parts malfunctioning. Therefore, the updated data set used for our analysis contained observations only from 1998-2007. From there we fit an ARIMA model to calculate the mean (505210291 tons of C/year) of the time series.

A.4 R packages and versions

To promote reproducibility, we include a list of all R packages and versions used in this analysis.

- astsa package v.1.8 (Stoffer 2017)
- dplyr package v.0.7.5 (Wickham et al. 2015)
- fitdistrplus package v.1.0 − 9 (Delignette-Muller and Dutang 2015)
- forecast package v8.2 (Hyndman 2017, Hyndman and Khandakar 2008)
- ggplot2 package v.2.2.1 (Wickham 2009)
- gridExtra package v.2.3 (Auguie 2017)
- partykit package v.1.2-2 (Hothorn and Zeileis 2015)
- R v.3.4.3 (R Core Team 2017)
- randomForest v.4.6-12 (Liaw and Wiener 2016)
- rpart package v.4.1-11 (Therneau et al. 2015)
- tree package v.1.0-37 (Ripley 2016)

Appendix B

Appendix for Chapter 3: Untangling uncertainty in food web models

B.0.1 SeaWiFS

Time Series Analysis

To create Case 1 ($f^1(\nu)$) in the trophic pyramid food web model where ν is a fixed constant, we used 8-day time series Sea-viewing Wide Field-of-view Sensor (SeaWiFS) chlorophyll *a* data from 1997 to 2010 that was transformed using the *Eppley*-Vertically Generalized Production Model (VGPM) to estimate net primary production (NPP) from chlorophyll *a*. The *Eppley*-VGPM estimates were used rather than the VGPM data since temperatures surrounding the main Hawaiian Islands (MHI) are above 20°C (Morel and André 1991, Antoine et al. 1996, Stock et al. 2017). The SeaWiFS data were originally obtained from the Oregon State Ocean Productivity website (`http://www.science.oregonstate.edu/ocean.productivity/`). They use a gap filling algorithm to populate missing pixels due to cloud coverage; however if no good data is available, pixels remain empty. We segmented the SeaWiFS data using the exclusive

economic zone (EEZ) boundaries of the MHI and assumed a closed system at equilibrium (Fig. B.1). The data then were converted from 8 day averages in mg C / m^2 / day per 9 km x 9 km pixel into total metric tons of Carbon per year total across the entire MHI. Although from 1997 to 2010 SeaWiFS collected data every 8 days per 9 km x 9 km pixel, the time series had gaps due to machine parts malfunctioning. Therefore, the updated data set used for our analysis contained observations only from 1998-2007.

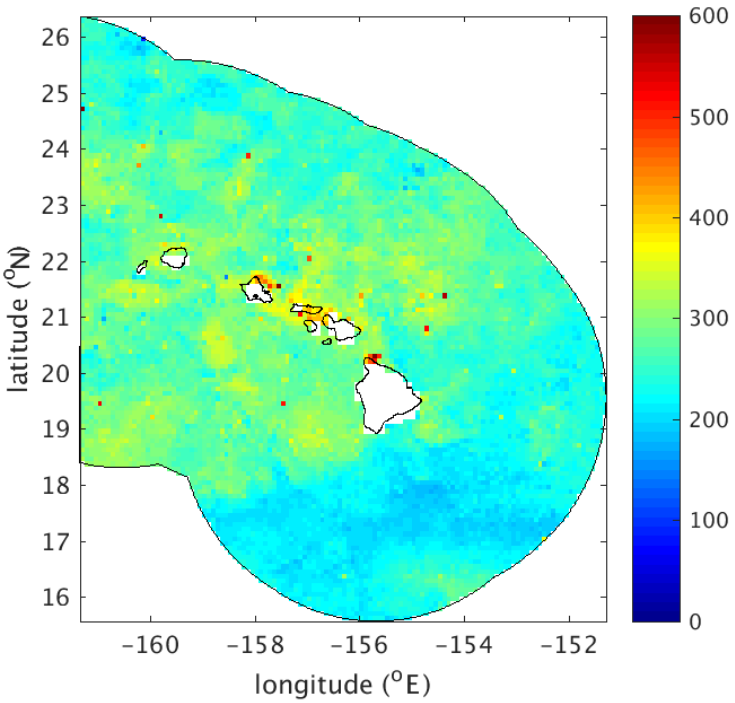

Figure B.1: A single 8-day time frame of the SeaWiFS *Eppley*-VGPM NPP data in January 1998 for the main Hawaiian Islands EEZ. The color scale shows the amount of estimated NPP in total gigatons of Carbon per year per pixel. The pixel size of the SeaWiFS data set is 9 by 9 km.

Since the SeaWiFS data were collected over time, we started off by verifying that there was a violation of independence and then chose a usable time series model. The NPP data demonstrated a strong annual frequency, a smaller six month frequency, and were non-stationary in the trend and seasonality (See Fig. B.2 and B.4). This was not

surprising since most biological data sets are seasonal and are influenced by the time of year.

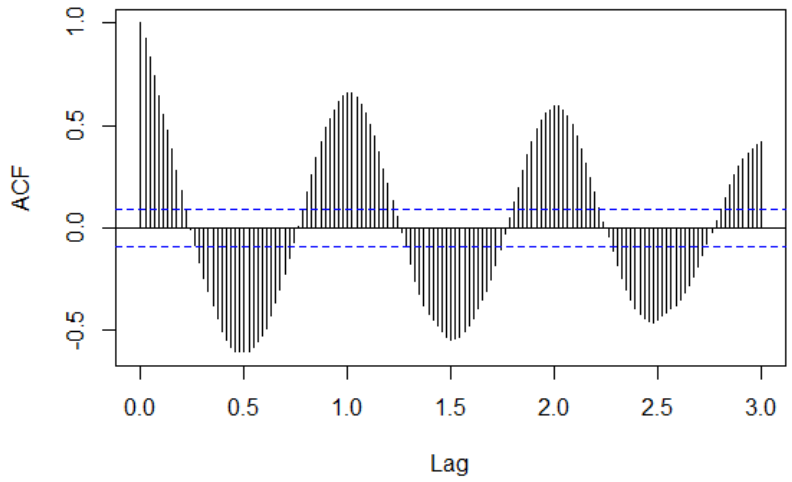

Figure B.2: ACF plot of SeaWiFS data. Plots created using R v.3.4.3 (R Core Team 2017) forecast package v8.2 (Hyndman 2017, Hyndman and Khandakar 2008).

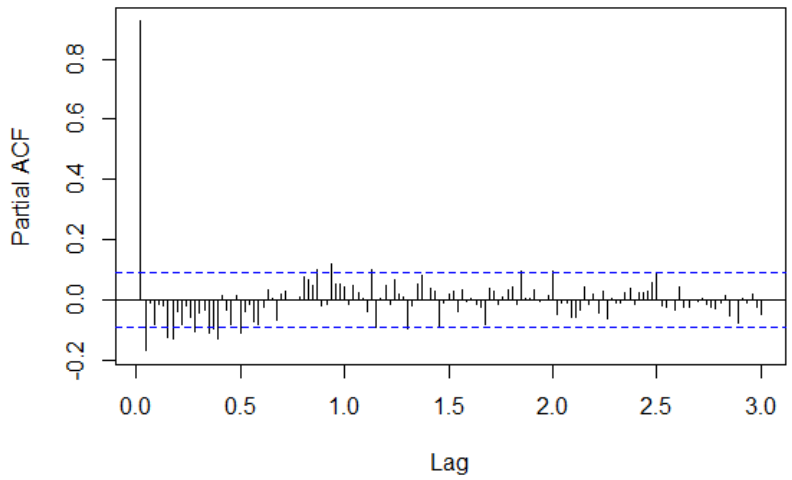

Figure B.3: PACF plots of SeaWiFS data. Plots created using R v.3.4.3 (R Core Team 2017) forecast package v8.2 (Hyndman 2017, Hyndman and Khandakar 2008).

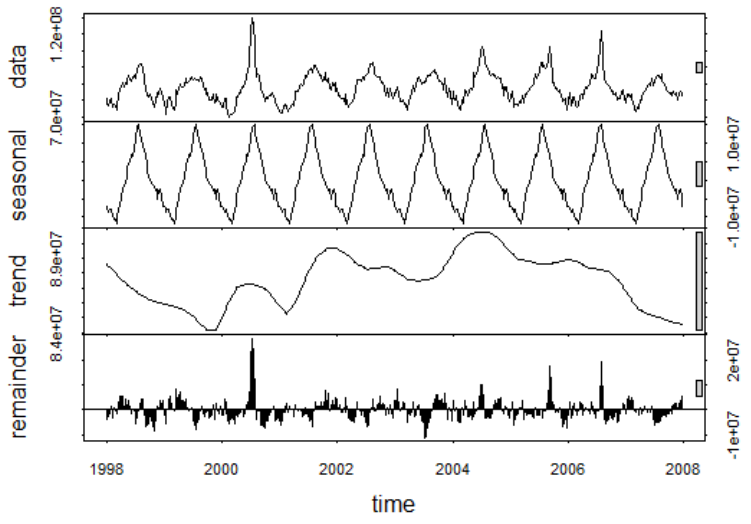

Figure B.4: Exploratory time series plots of SeaWiFS data–seasonal and trend components. Plots created using R v.3.4.3 (R Core Team 2017) forecast package v8.2 (Hyndman 2017, Hyndman and Khandakar 2008).

We fitted a seasonal autoregressive integrated moving average model, or

SARIMA$(0,0,2)(0,0,2)_{46}$ to the pre-processed SeaWiFS NPP data (i.e., moving average order $q = 2$, seasonal moving average $Q = 2$, and seasonal component $s = 46$ (365 days/8-day time series = 46)).

$$X_t = \delta + (1 + \theta_1 B + \theta_2 B^2)(1 + \Theta_1 B^{46} + \Theta_2 B^{92})W_t \tag{B.1}$$

In Eq. B.1, X_t is the total NPP estimated across the MHI in 8-day time periods from 1998-2007 where $t \in \{1, 460\}$, δ represents the mean of the time series, θ_1 and θ_2 are the moving average parameters, Θ_2 and Θ_2 are the seasonal moving average parameters, and $W_t \overset{iid}{\sim} N(0, \sigma_w^2)$. The backshift operator, B, is defined as $BW_t = W_{t-1}$, $B^2W_t = W_{t-2}$, $B^{46}W_t = W_{t-46}$, and $B^{92}W_t = W_{t-92}$. Our model, Eq. B.1, was chosen based on information criteria. Parameters were estimated using maximum likelihood estimation (Hyndman 2017, Hyndman and Khandakar 2008). The parameter estimates can be found in Table B.1. The estimate of parameter δ was used for the value of ν when it was treated as a fixed constant in the hierarchical food web model (c_1 in $f^1(\nu)$).

Table B.1: Seasonal ARIMA parameter estimates for Eq. B.1 and their respective standard error calculated using R v.3.4.3 (R Core Team 2017) forecast package v8.2 and function auto.arima() (Hyndman 2017, Hyndman and Khandakar 2008)

Parameters	Estimates	Standard error
δ	87339861.7	696375.7
θ_1	0.9806	0.0413
θ_2	0.4817	0.0359
Θ_1	0.3379	0.0470
Θ_2	0.3392	0.0449
σ_w^2	1.494306e+13	

Simulation Setting

While in Case 1 ($f^1(\nu)$) ν is treated as a fixed constant, in Case 2 ($f^2(\nu)$) ν is a random variable. In order to determine an approximate distribution for ν, we ran goodness-of-fit

tests on the SeaWiFS NPP data. We started off with a skewness-kurtosis plot (Fig. B.5) to initially decide which distributions to consider (i.e., Lognormal, Weibull, and Gamma) (Delignette-Muller and Dutang 2015). Then we fitted individual distributions to the data using maximum likelihood estimation and compared density plots of the fitted distributions to the histogram of the empirical distribution, a cumulative distribution (CDF) plot of both the empirical distribution and the fitted distributions, Q-Q plots, and P-P plots (Fig. B.6). From there, we concluded that the Lognormal distribution was the most appropriate approximate distribution amongst the distributions we considered. The Lognormal distribution has two parameters μ and σ^2 and is defined within the range for $\nu > 0$. The probability distribution function (PDF) and equations for the mean and variance of a Lognormal distribution can be found in Table B.2.

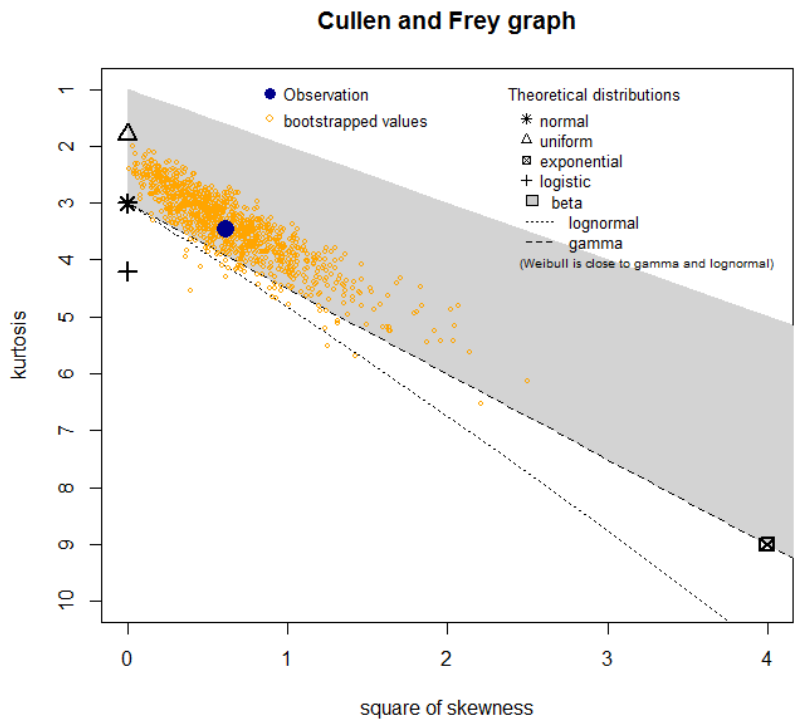

Figure B.5: Visualizes several potential continuous distributions against the SeaWiFS data and bootstrapped data. The figure shows it potentially follows a Gamma, Weibull, and Lognormal distribution. It also appears Beta could be a reasonable choice, but the SeaWiFS data is not constricted between 0 and 1. Therefore Beta is ruled out as a potential distribution. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

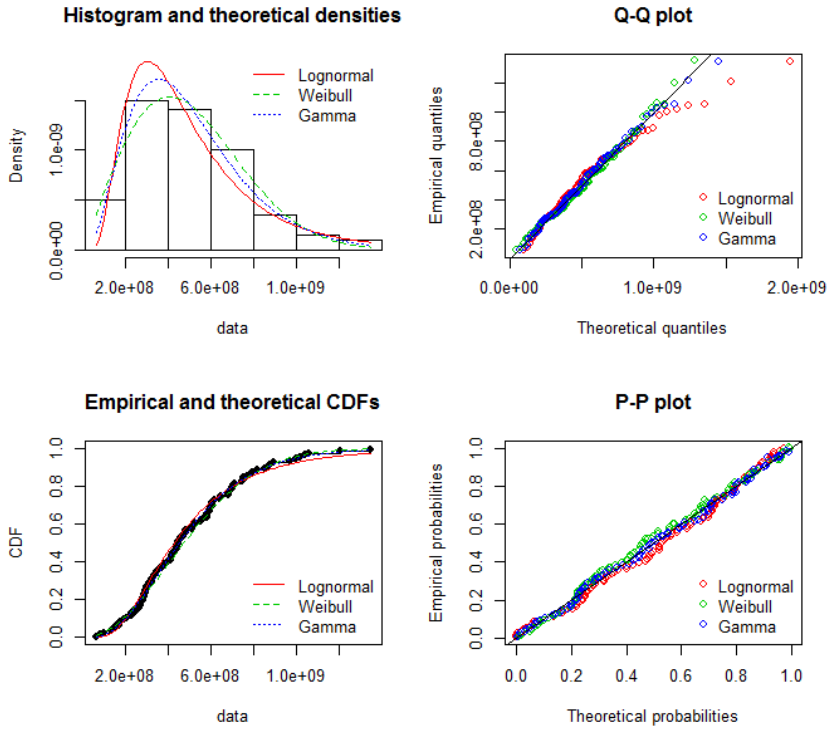

Figure B.6: Density plots of the fitted distributions to the histogram of the empirical distribution, a CDF plot of both the empirical distribution and the fitted distributions (Gamma, Weibull, and Lognormal), Q-Q plots, and P-P plots. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 − 9 (Delignette-Muller and Dutang 2015).

B.0.2 Transfer Efficiency

We utilized data gathered from a literature review on transfer efficiencies to fit approximate distributions for Case 2 and 3 of the transfer efficiencies (τ_h). The data includes articles that mentioned both food web and transfer efficiency and then selected from these studies those that included data whether model-based or empirical. Case 2 used all of the marine data (i.e., both model-based and empirical) to fit a distribution. The resulting distributional assumption was placed on transfer efficiencies (τ_h) for all trophic levels $h \in \{2, 3, 4\}$. Case 3 subsetted the marine transfer efficiency data by trophic level h in order to place distinct distributional assumptions on transfer efficiencies at each

trophic level. In order to determine approximate distributions for each of the cases, we ran goodness-of-fit tests. We started off with skewness-kurtosis plots (Fig. B.7, B.9, B.11, and B.13) to initially decide which distributions to consider (i.e., Beta and Gamma) (Delignette-Muller and Dutang 2015). Then we fitted individual distributions to the data using maximum likelihood estimation and compared density plots of the fitted distributions to the histogram of the empirical distribution, a cumulative distribution (CDF) plot of both the empirical distribution and the fitted distributions, Q-Q plots, and P-P plots (Fig. B.8, B.10, B.12, and B.14). We chose the Beta distribution, because the transfer efficiency is a percentage and the Beta distribution is defined within the range [0, 1]. Therefore, we concluded that for both Case 2 and 3 the Beta distribution was the most appropriate amongst the approximate distributions (versus Gamma distribution). Each data set used in Case 2 and 3 have distinct values for the shape parameters α and β. The probability distribution function (PDF) and equations for the mean and variance of a Beta distribution can be found in Table B.2.

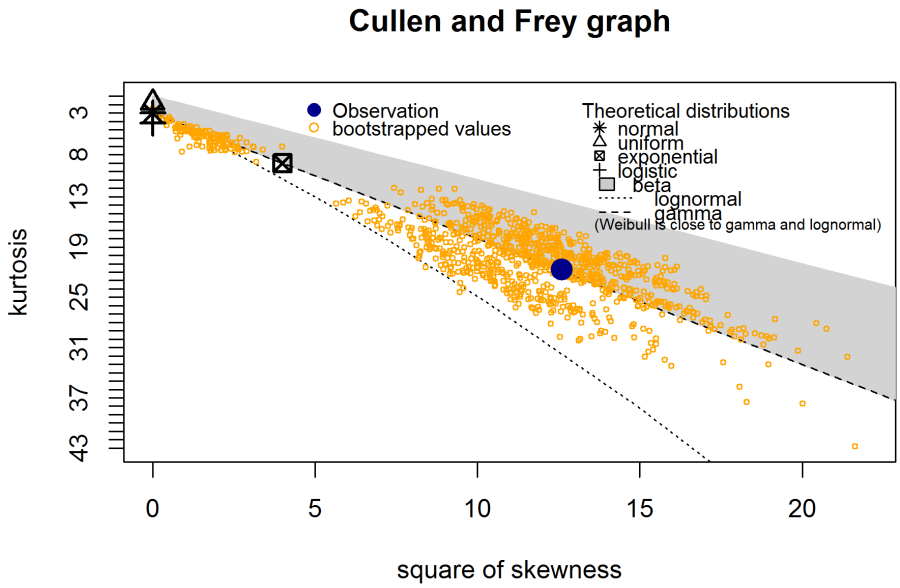

Figure B.7: Visualizes several potential continuous distributions against the marine transfer data (Case 2) and bootstrapped data. The figure shows it potentially follows a Beta and Gamma distribution. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

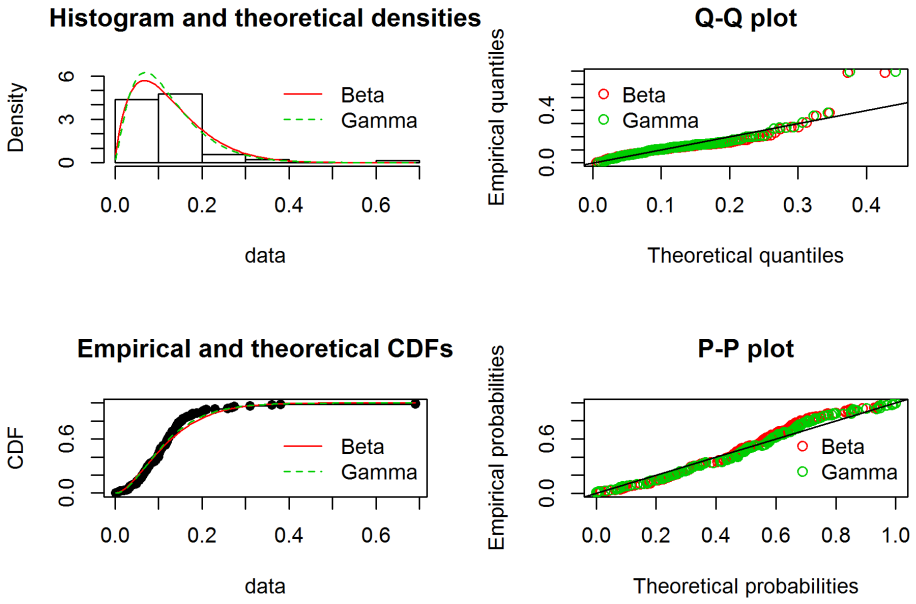

Figure B.8: Density plots of the fitted distributions to the histogram of the empirical distribution using the marine transfer efficiency data (Case 2), a CDF plot of both the empirical distribution and the fitted distributions (Beta and Gamma), Q-Q plots, and P-P plots. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

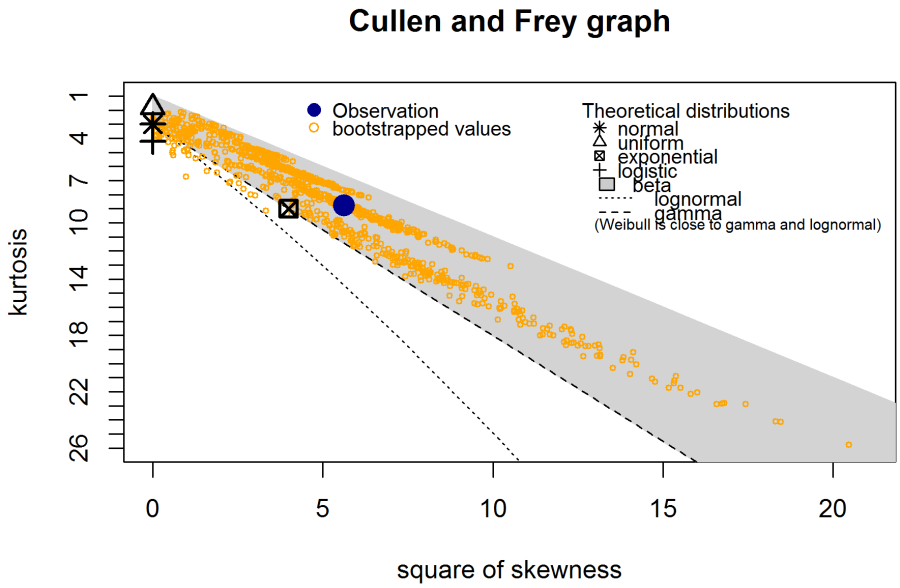

Figure B.9: Visualizes several potential continuous distributions against the marine transfer data for trophic level 2 (Case 3) and bootstrapped data. The figure shows it potentially follows a Beta distribution. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

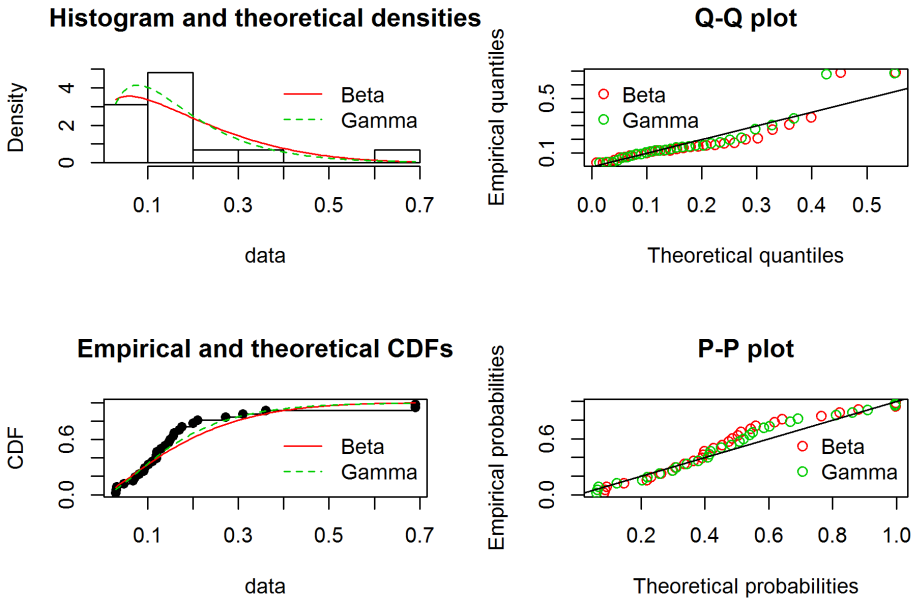

Figure B.10: Density plots of the fitted distributions to the histogram of the empirical distribution using the marine transfer efficiency data for trophic level 2 (Case 3), a CDF plot of both the empirical distribution and the fitted distributions (Beta and Gamma), Q-Q plots, and P-P plots. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

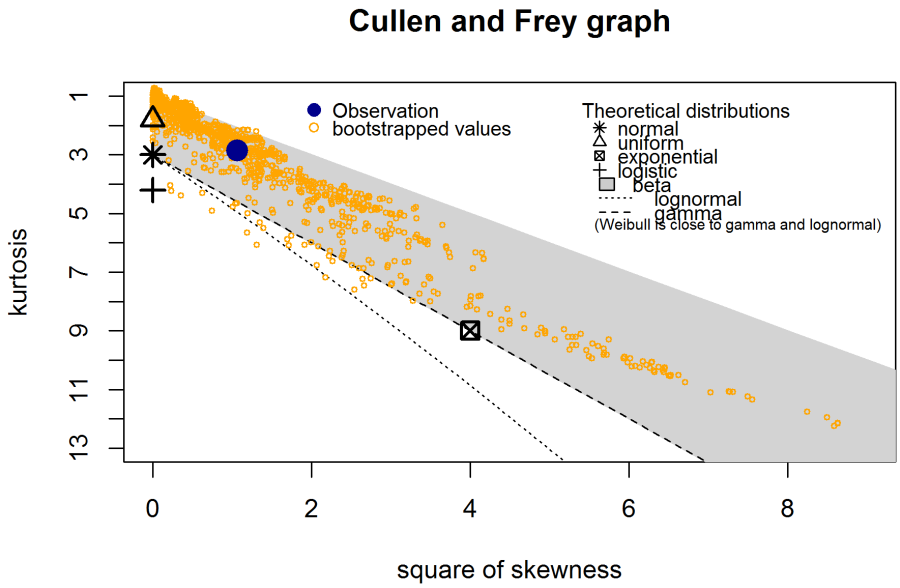

Figure B.11: Visualizes several potential continuous distributions against the marine transfer data for trophic level 3 (Case 3) and bootstrapped data. The figure shows it potentially follows a Beta distribution. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

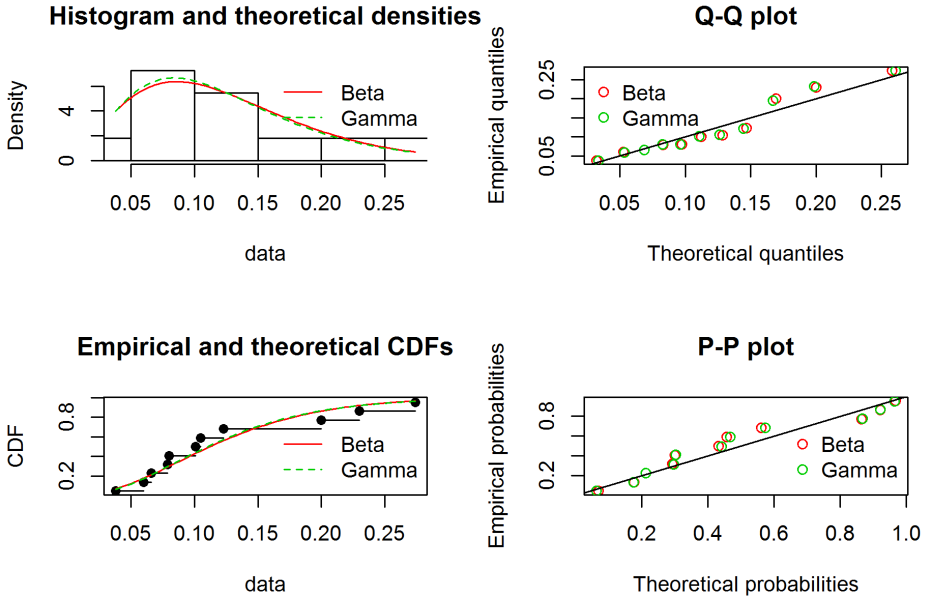

Figure B.12: Density plots of the fitted distributions to the histogram of the empirical distribution using the marine transfer efficiency data for trophic level 3 (Case 3), a CDF plot of both the empirical distribution and the fitted distributions (Beta and Gamma), Q-Q plots, and P-P plots. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

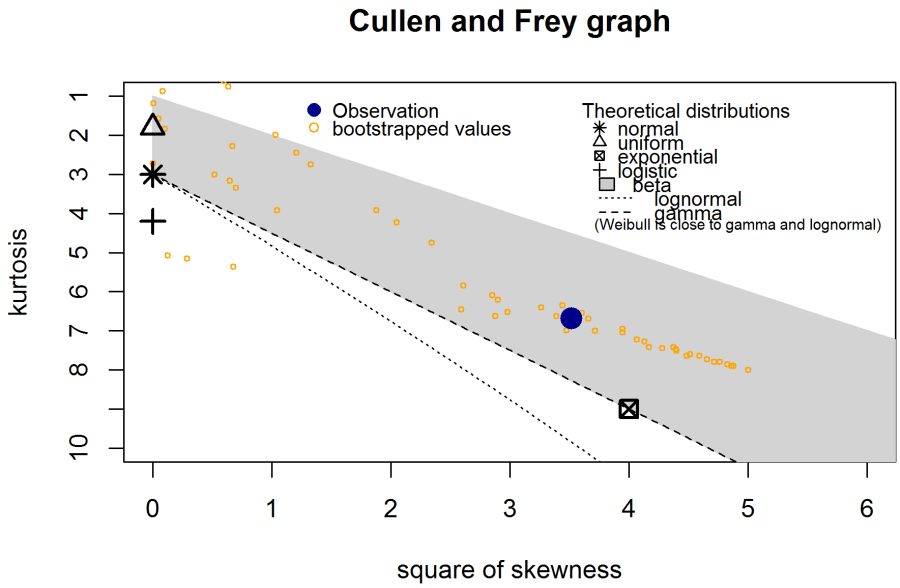

Figure B.13: Visualizes several potential continuous distributions against the marine transfer data for trophic level 4 (Case 3) and bootstrapped data. The figure shows it potentially follows a Beta distribution. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

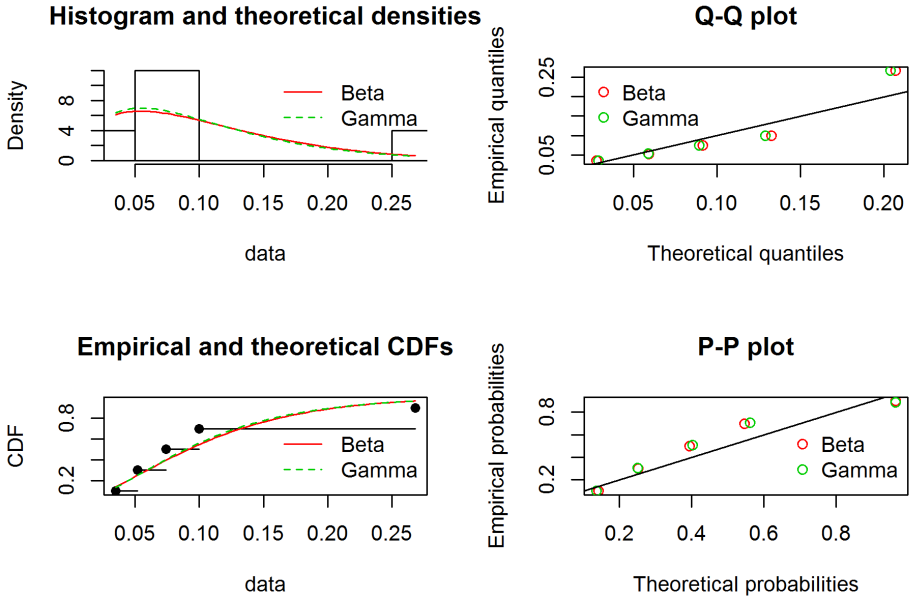

Figure B.14: Density plots of the fitted distributions to the histogram of the empirical distribution using the marine transfer efficiency data for trophic level 4 (Case 3), a CDF plot of both the empirical distribution and the fitted distributions (Beta and Gamma), Q-Q plots, and P-P plots. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

B.0.3 FishBase

Trophic Level

Estimated trophic levels (rational numbers) were calculated from the FishBase database (`http://www.fishbase.org/`) for each species and truncated to cluster species into integer-valued trophic levels. Truncated trophic levels range in the MHI from $h = 1, \ldots, 4$ with $h = 1$ representing phytoplankton, $h = 2$ primary consumers, $h = 3$ secondary consumers, and $h = 4$ tertiary consumers. If the rational number was greater than or equal to 4, then the organism was put into trophic level 4, if it was less than 4 and greater than or equal to 3 it was put into trophic level 3, and if it was less than 3 it was put into trophic level 2. The number of species found at each trophic level were not equal.

Maximum Expected Lifespan

The maximum expected lifespan λ_h for species at trophic level $h = 2, 3, 4$ was treated as a random value in Case 2. We ran goodness-of-fit tests for each level of h to find an approximate distribution. We started off with a skewness-kurtosis plot (Fig. B.15, B.17, and B.19) to initially decide which distributions to consider (i.e., Lognormal, Weibull, and Gamma) (Delignette-Muller and Dutang 2015). Then we fitted individual distributions to the data using maximum likelihood estimation and compared density plots of the fitted distributions to the histogram of the empirical distribution, a cumulative distribution (CDF) plot of both the empirical distribution and the fitted distributions, Q-Q plots, and P-P plots (Fig. B.16, B.18, and B.20). We chose Lognormal distribution as our approximate distribution for all trophic levels, where each trophic level has distinct values for μ and σ^2.

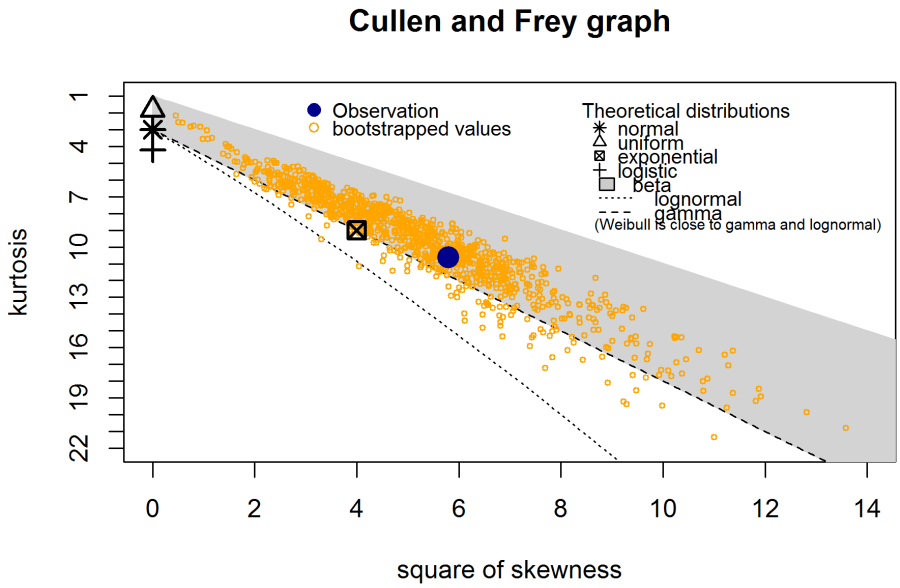

Figure B.15: Visualizes several potential continuous distributions against the maximum expected lifespan data for trophic level 2 and bootstrapped data. The figure shows it potentially follows a Gamma, Weibull, and Lognormal distribution. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette–Muller and Dutang 2015).

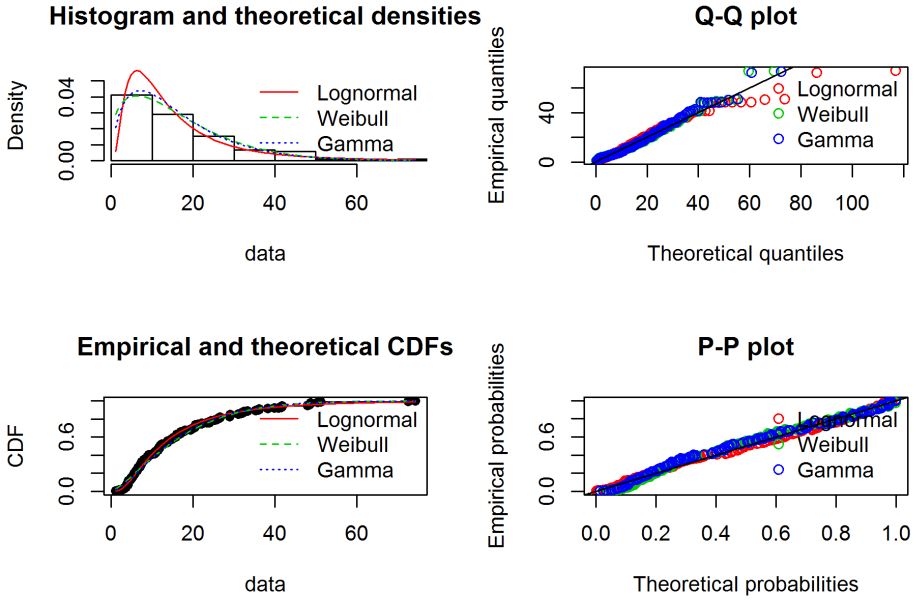

Figure B.16: Density plots of the fitted distributions to the histogram of the empirical distribution using the maximum expected lifespan data for trophic level 2, a CDF plot of both the empirical distribution and the fitted distributions (Gamma, Weibull, and Lognormal), Q-Q plots, and P-P plots. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

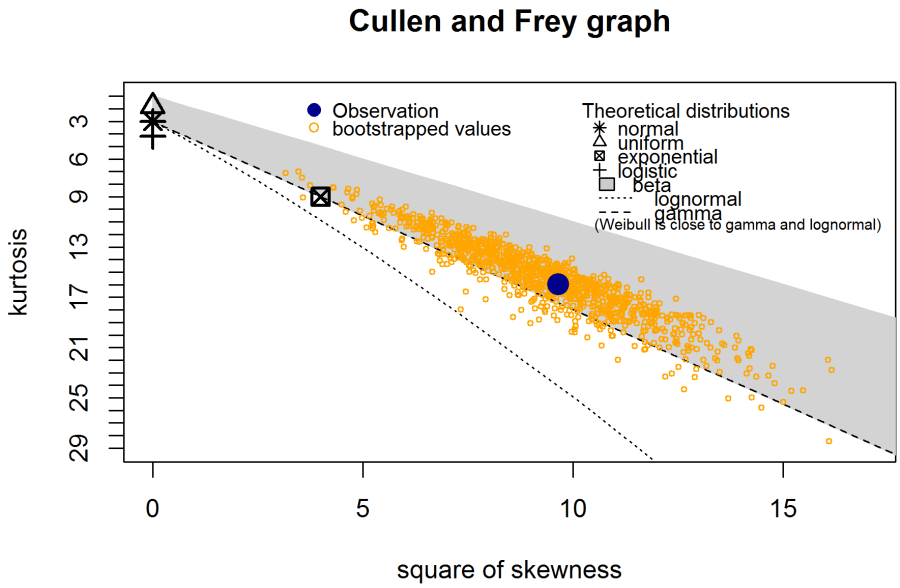

Figure B.17: Visualizes several potential continuous distributions against the maximum expected lifespan data for trophic level 3 and bootstrapped data. The figure shows it potentially follows a Gamma, Weibull, and Lognormal distribution. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette–Muller and Dutang 2015).

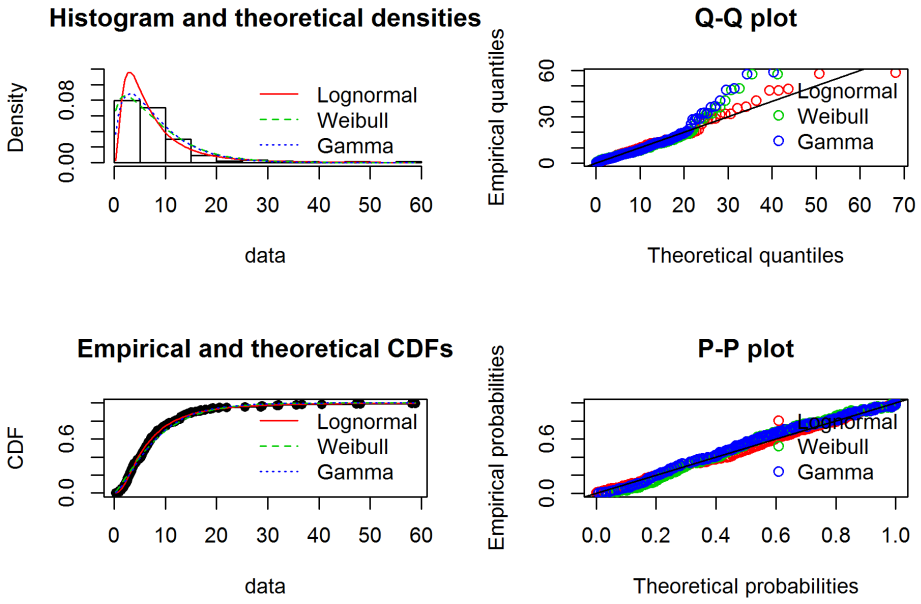

Figure B.18: Density plots of the fitted distributions to the histogram of the empirical distribution using the maximum expected lifespan data for trophic level 3, a CDF plot of both the empirical distribution and the fitted distributions (Gamma, Weibull, and Lognormal), Q-Q plots, and P-P plots. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

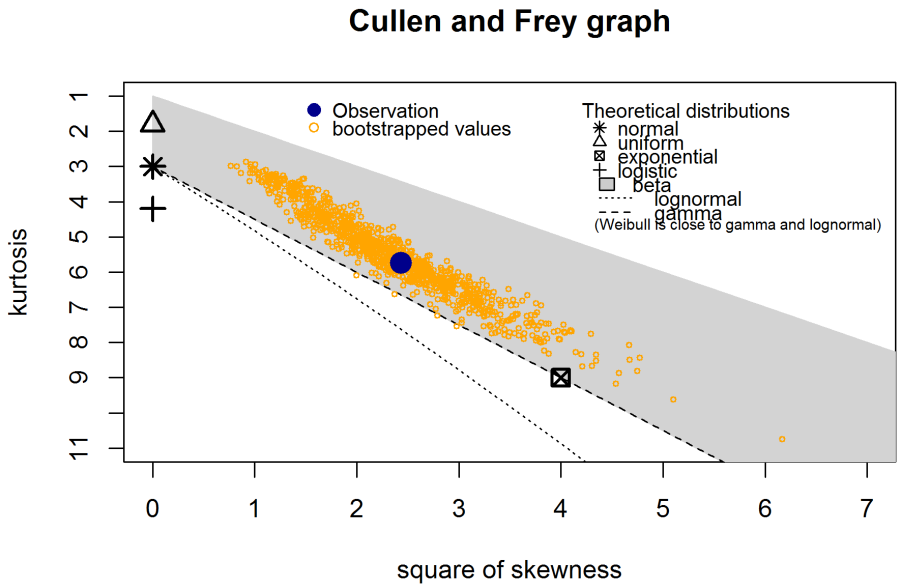

Figure B.19: Visualizes several potential continuous distributions against the maximum expected lifespan data for trophic level 4 and bootstrapped data. The figure shows it potentially follows a Gamma, Weibull, and Lognormal distribution. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette–Muller and Dutang 2015).

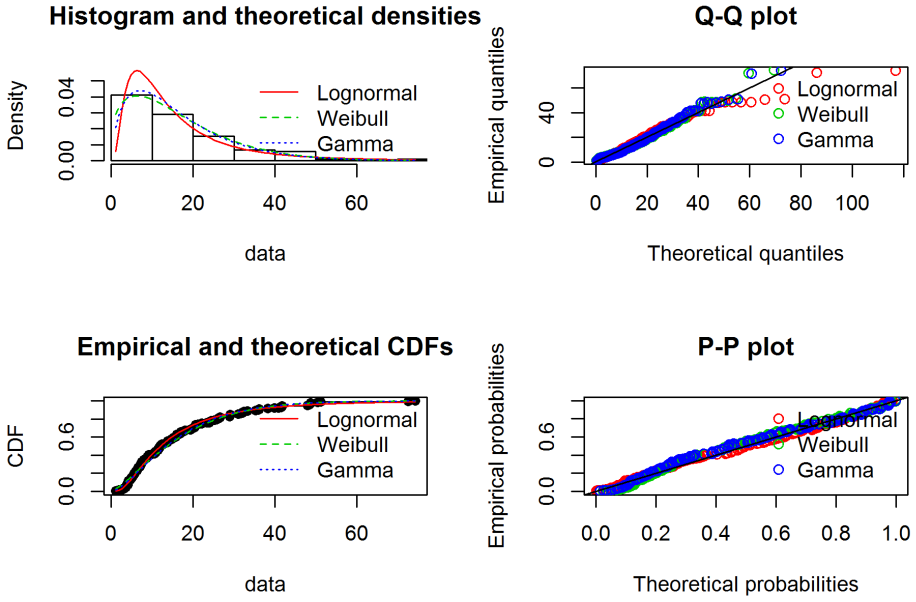

Figure B.20: Density plots of the fitted distributions to the histogram of the empirical distribution using the maximum expected lifespan data for trophic level 4, a CDF plot of both the empirical distribution and the fitted distributions (Gamma, Weibull, and Lognormal), Q-Q plots, and P-P plots. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

B.1 Marginal Distributions

When attempting to solve for the equations for the distributions of γ_h, we started off with the simplest stochastic scenario and progressed toward more complex cases.

B.1.1 Deriving Equations for Scenario NTl

In Scenario NTl, both ν and τ_h are treated as fixed constants. We let $\lambda_h \sim Lognormal(\mu_{\lambda_h}, \sigma^2_{\lambda_h})$ for $h \in \{2, 3, 4\}$ where μ_{λ_h} and $\sigma^2_{\lambda_h}$ are known constants.

Beginning with the top layer in the hierarchical model (i.e., trophic level 2). We write Eq. (1) from the main paper under Scenario NTl's assumptions where $c_{\nu 9 \tau_2} = \nu * 9 * \tau_2$.

$$\gamma_2 = \nu * 9 * \tau_2 * \lambda_2 = c_{\nu 9 \tau_2} * \lambda_2$$

From here, we are interested in determining the distribution of $c * \lambda_h$ assuming $c > 0$. Since we let $\lambda_h \sim Lognormal(\mu_{\lambda_h}, \sigma^2_{\lambda_h})$, we can take the log to obtain $log(\lambda_h) \sim N(\mu_{\lambda_h}, \sigma^2_{\lambda_h})$. Therefore, $\gamma_h = c * \lambda_h$ can be transformed into $log(\gamma_h) = log(c * \lambda_h) = log(c) + log(\lambda_h)$. Using Jacobian transformations, we find that $log(\gamma_h) \sim N(\mu_{\lambda_h} + log(c), \sigma^2_{\lambda_h})$. Finally using the same logic as earlier, we conclude that $\gamma_h \sim Lognormal(\mu_{\lambda_h} + log(c), \sigma^2_{\lambda_h})$.

Therefore in Scenario NTl, trophic level biomass (γ_h) at each trophic level h follows the following distributions:

$$\begin{aligned}
&\gamma_2 \sim Lognormal(\mu_{\lambda_2} + log(c'), \sigma^2_{\lambda_2}) \quad \text{where} \quad c' = \nu * 9 * \tau_2 \\
&\gamma_3 \sim Lognormal(\mu_{\lambda_3} + log(c''), \sigma^2_{\lambda_3}) \quad \text{where} \quad c'' = \nu * 9 * \tau_2 * \tau_3 \\
&\gamma_4 \sim Lognormal(\mu_{\lambda_4} + log(c'''), \sigma^2_{\lambda_4}) \quad \text{where} \quad c''' = \nu * 9 * \tau_2 * \tau_3 * \tau_4
\end{aligned} \tag{B.2}$$

B.1.2 Deriving Equations for Scenario nTl

In Scenario nTl, τ_h is treated as a fixed constant, and we treat $\nu \sim Lognormal(\mu_\nu, \sigma^2_\nu)$ and $\lambda_h \sim Lognormal(\mu_{\lambda_h}, \sigma^2_{\lambda_h})$ for $h \in \{2, 3, 4\}$ where we assume the parameters for both ν and λ_h are known constants.

Beginning with the top layer in the hierarchical model (i.e., trophic level 2), we write Eq. (1) from the main paper under Scenario nTl's assumptions where $c_{9\tau_2} = 9 * \tau_2$.

$$\gamma_2 = \nu * 9 * \tau_2 * \lambda_2 = \nu * c_{9\tau_2} * \lambda_2$$

We want to determine the distribution when two random variables that follow distinct Lognormal distributions are multiplied together. If we let $\lambda_h' = ln(\lambda_h) \sim N(\mu_{\lambda_h}, \sigma^2_{\lambda_h})$ and $\nu' = ln(\nu) \sim N(\mu_\nu, \sigma^2_\nu)$, then using properties of Normal distributions we obtain that $Y' = \lambda_h' + \nu' \sim N(\mu_{\lambda_h} + \mu_\nu, \sigma^2_{\lambda_h} + \sigma^2_\nu)$. Remembering that $Y' = ln(\lambda_h) + ln(\nu)$, we can then raise it to the exponent fo find that $e^{Y'} = e^{ln(\lambda_h * \nu)} = \lambda_h * \nu$. We then find that $e^{Y'} \sim Lognormal(\mu_{\lambda_h} + \mu_\nu, \sigma^2_{\lambda_h} + \sigma^2_\nu)$. Using the theory outline in Scenario NTl pertaining constants ($c > 0$) and Lognormal distributions, we conclude that $c * \lambda_h * \nu \sim Lognormal(\mu_{\lambda_h} + \mu_\nu + log(c), \sigma^2_{\lambda_h} + \sigma^2_\nu)$.

Therefore in Scenario nTl, trophic level biomass (γ_h) at each trophic level h follows the following distributions:

$$\begin{aligned}
\gamma_2 &\sim Lognormal\left(\mu_\nu + \mu_{\lambda_2} + log(c'), \sigma^2_\nu + \sigma^2_{\lambda_2}\right) \quad \text{where} \quad c' = 9 * \tau_2 \\
\gamma_3 &\sim Lognormal\left(\mu_\nu + \mu_{\lambda_3} + log(c''), \sigma^2_\nu + \sigma^2_{\lambda_3}\right) \quad \text{where} \quad c'' = 9 * \tau_2 * \tau_3 \\
\gamma_4 &\sim Lognormal\left(\mu_\nu + \mu_{\lambda_4} + log(c'''), \sigma^2_\nu + \sigma^2_{\lambda_4}\right) \quad \text{where} \quad c''' = 9 * \tau_2 * \tau_3 * \tau_4
\end{aligned} \tag{B.3}$$

B.1.3 Attempting to Derive Equations for Scenario *NtL*

In Scenario *NtL*, ν and λ_h for $h \in \{2, 3, 4\}$ are treated as fixed constants. We let $\tau_h \sim Beta(\alpha_h, \beta_h)$ for $h \in \{2, 3, 4\}$ and assume α_h and β_h are known constants for τ_h.

Beginning with the top layer in the hierarchical model (i.e., trophic level 2), we write Eq. (1) from the main paper under Scenario *NtL*'s assumptions where $c_{9\nu\lambda_2} = 9 * \nu\lambda_2$.

$$\gamma_2 = \nu * 9 * \tau_2 * \lambda_2 = c_{\nu 9 \lambda_2} * \tau_2$$

While we are interested in determining the distribution of $c * \tau_2$ assuming $c > 0$, no

theorems exist that prove the beta distribution has the scaling property. In addition, at higher trophic levels we will be multiplying multiple independent Beta distributions. Therefore given the lack of extant research on the topic, we decided simulation was a reasonable option.

B.1.4 Attempting to Derive Equations for Scenario *Ntl*

In Scenario *Ntl*, ν is treated as a fixed constant. We treat $\tau_h \sim Beta(\alpha_h, \beta_h)$ and $\lambda_h \sim Lognormal(\mu_{\lambda_h}, \sigma^2_{\lambda_h})$ for $h \in \{2, 3, 4\}$. We assume μ_{λ_h} and $\sigma^2_{\lambda_h}$ are known constants for λ_h and α_h and β_h are known constants for τ_h.

Beginning with the top layer in the hierarchical model (i.e., trophic level 2), we write Eq. (1) from the main paper under Scenario *Ntl*'s assumptions where $c_{9\nu} = 9 * \nu$.

$$\gamma_2 = \nu * 9 * \tau_2 * \lambda_2 = c_{\nu 9} * \tau_2 * \lambda_2$$

While our goal was to determine the exact distribution of the product of two random variables that follow a Beta and a Lognormal distribution, extant research (e.g., Casella and Berger 2002, Rohatgi and Saleh 2015) on the topic provides no clear path forward. The closest match were two papers on the distribution of the product of a Beta, Gamma, and mean zero Normal (Springer and Thompson 1970, Gaunt et al. 2018).

Gaunt et al. 2018 extends Stein's method to products of independent Beta, Gamma, generalized Gamma and mean zero Normal random variables. Gaunt defined a characteristic function outlined in Corollory 3.3 for the Product Beta, Product Gamma, and Product Normal where $W \sim N(0, \sigma^2)$. This Corollory derived from Gaunt et al. 2018 state that a characteristic function can be found for the Product Beta, Product Gamma, and Product central Normal. However, we believe this proof is going to break down

when moving from a symmetric to a non-symmetric distribution. Providing additional complexity, in Scenario *Ntl* and the in the remaining scenarios, Lognormal distributions are used (as opposed to Normal distributions).

Corollary 3.3. *The characteristic function of W is given by*

$$\phi(t) = MG^{2m+2n+N-1,1}_{2m+1,2m+2n+N-1}\left(\frac{\lambda^{2n}}{2^{2n+N-2}\sigma^2 t^2}\;\middle|\;\begin{matrix}1, \frac{a_1+b_1+1}{2}, \ldots, \frac{a_m+b_m+1}{2}, \cdots \\ \frac{a_1+1}{2}, \ldots, \frac{a_m+1}{2}, \frac{a_1}{2}, \ldots, \frac{a_m}{2}, \cdots \\ \cdots \frac{a_1+b_1}{2}, \ldots, \frac{a_m+b_m}{2} \\ \frac{r_1+1}{2}, \ldots, \frac{r_n+1}{2}, \frac{r_1}{2}, \ldots, \frac{r_n}{2}, \frac{1}{2}, \ldots, \frac{1}{2}\end{matrix}\right),$$

where

$$M = \frac{1}{\pi^{(n+N-1)/2}} \prod_{i=1}^{m} \frac{\Gamma(a_i+b_i)}{2^{b_i}\Gamma(a_i)} \prod_{j=1}^{n} \frac{2^{r_j-1}}{\Gamma(r_j)}.$$

Proof. Since the distribution of W is symmetric about the origin, it follows that the characteristic function $\phi(t)$ is given by

$$\phi(t) = \mathbb{E}[\mathrm{e}^{itW}] = \mathbb{E}[\cos(tW)] = 2\int_0^\infty \cos(tx)p(x)\,\mathrm{d}x.$$

B.2 Table of Distributions

Table B.2: Table of Distributions

Distribution	PDF	Mean & Variances
Lognormal	$f(\nu;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma\nu} exp\left(-\frac{(log(\nu)-\mu)^2}{2\sigma^2}\right)$ for $\nu > 0$	$E[\nu] = e^{\mu+\frac{\sigma^2}{2}}$ $Var[\nu] = e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$
Beta	$f(\tau_h;\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\tau_h^{\alpha-1}(1-\tau_h)^{\beta-1}$ $0 \leq \tau_h \leq 1$	$E[\tau_h] = \frac{\alpha}{\alpha+\beta}$ $Var[\tau_h] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

B.3 R packages and versions

To promote reproducibility, we include a list of all R packages and versions used in this analysis.

- astsa package v.1.8 (Stoffer 2017)
- data.table package v.1.10.4 − 3 (Dowle and Srinivasan 2017)
- ExtDist package v.0.6 − 3 (Wu et al. 2015)
- fitdistrplus package v.1.0 − 9 (Delignette-Muller and Dutang 2015)
- forecast package v8.2 (Hyndman 2017, Hyndman and Khandakar 2008)
- ggplot2 package v.2.2.1 (Wickham 2009)
- gridExtra package v.2.3 (Auguie 2017)
- gtable package v.0.2.0 (Wickham 2016)
- logspline package v.2.1.9 (Kooperberg 2016)
- R v.3.4.3 (R Core Team 2017)
- reshape package v.0.8.7 (Wickham 2007)

Appendix C

Appendix for Chapter 4: Ecosystem knowledge in Bayesian surplus production models–what can it tell us?

C.1 Data

C.1.1 SeaWiFS

Time Series Analysis

In the trophic pyramid food web model, ν is the estimated mean of the SeaWiFS time series (87339861.7 metric tons C/year) within the EEZ of the MHI found by fitting a seasonal autoregressive integrated moving average model (SARIMA model) to the SeaWiFS data using R v.3.4.3 (R Core Team 2017) forecast package v8.2 and function auto.arima() (Hyndman 2017, Hyndman and Khandakar 2008). To estimate ν, we used

8-day time series Sea-viewing Wide Field-of-view Sensor (SeaWiFS) chlorophyll *a* data from 1997 to 2010 that was transformed using the *Eppley*-Vertically Generalized Production Model (VGPM) to estimate net primary production (NPP) from chlorophyll *a*. The *Eppley*-VGPM estimates were used rather than the VGPM data since temperatures surrounding the main Hawaiian Islands (MHI) are above 20°C (Morel and André 1991, Antoine et al. 1996, Stock et al. 2017). The SeaWiFS data were originally obtained from the Oregon State Ocean Productivity website (`http://www.science.oregonstate.edu/ocean.productivity/`). They use a gap filling algorithm to populate missing pixels due to cloud coverage; however if no good data is available, pixels remain empty. We segmented the SeaWiFS data using the exclusive economic zone (EEZ) boundaries of the MHI and assumed a closed system at equilibrium (Fig. C.1). The data then were converted from 8 day averages in mg C / m^2 / day per 9 km x 9 km pixel into total metric tons of Carbon per year total across the entire MHI. Although from 1997 to 2010 SeaWiFS collected data every 8 days per 9 km x 9 km pixel, the time series had gaps due to machine parts malfunctioning. Therefore, the updated data set used for our analysis contained observations only from 1998-2007.

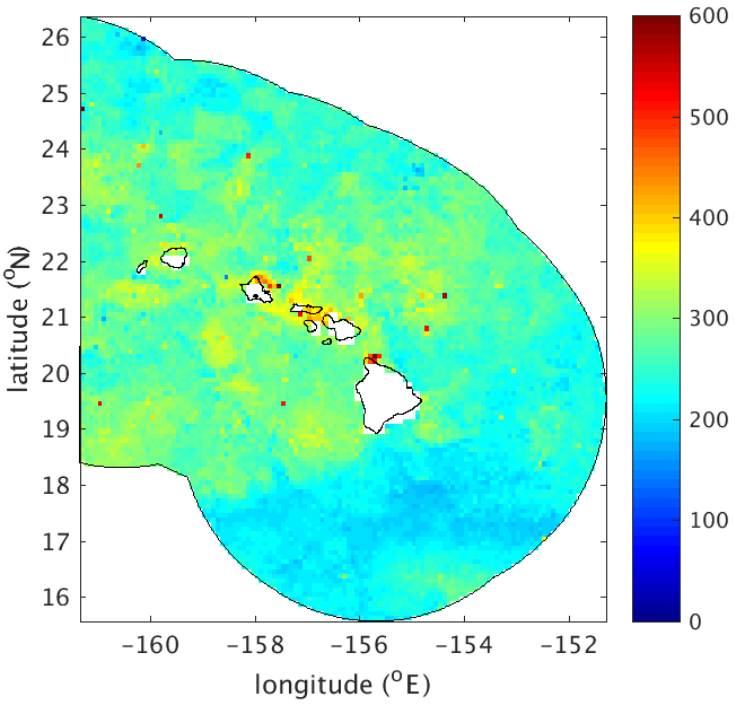

Figure C.1: A single 8-day time frame of the SeaWiFS *Eppley*-VGPM NPP data in January 1998 for the main Hawaiian Islands EEZ. The color scale shows the amount of estimated NPP in total gigatons of Carbon per year per pixel. The pixel size of the SeaWiFS data set is 9 by 9 km.

Since the SeaWiFS data were collected over time and exhibited strong temporal dependence, we started off choosing a time series model on which to base our estimate. The NPP data demonstrated a strong annual frequency, a smaller six month frequency, and were non-stationary in the trend and seasonality (See Fig. C.2 and C.4). This was not surprising since most biological data sets are seasonal and are influenced by the time of year.

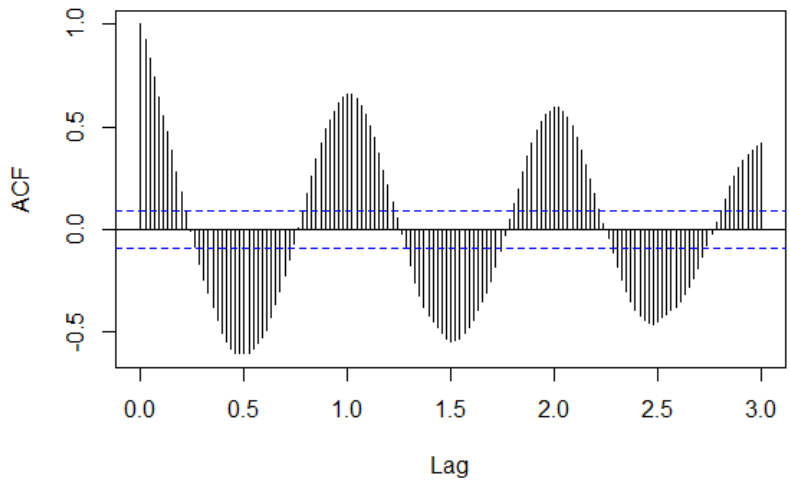

Figure C.2: ACF plot of SeaWiFS data. Plots created using R v.3.4.3 (R Core Team 2017) forecast package v8.2 (Hyndman 2017, Hyndman and Khandakar 2008).

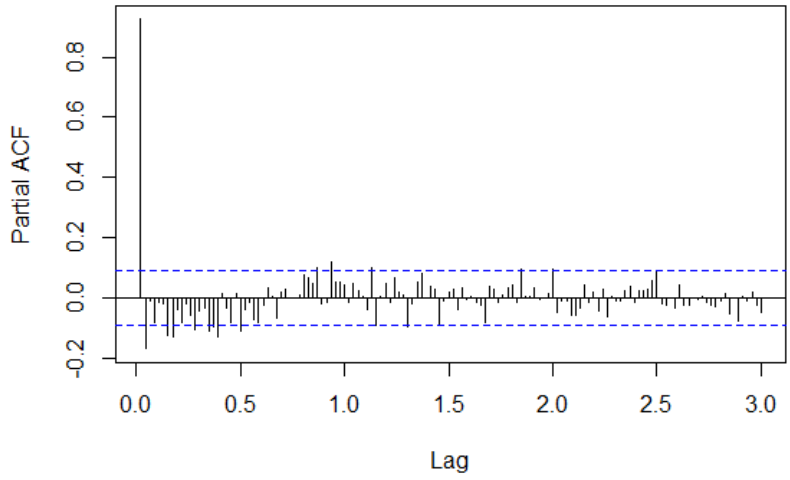

Figure C.3: PACF plots of SeaWiFS data. Plots created using R v.3.4.3 (R Core Team 2017) forecast package v8.2 (Hyndman 2017, Hyndman and Khandakar 2008).

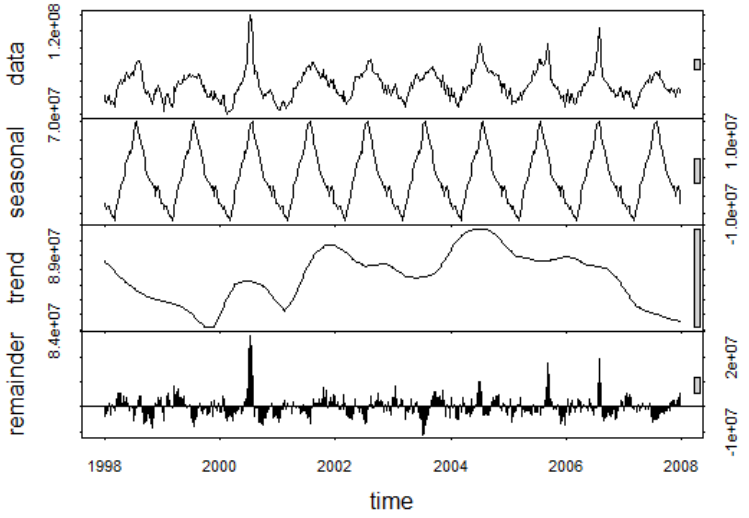

Figure C.4: Exploratory time series plots of SeaWiFS data–seasonal and trend components. Plots created using R v.3.4.3 (R Core Team 2017) forecast package v8.2 (Hyndman 2017, Hyndman and Khandakar 2008).

We fitted a seasonal autoregressive integrated moving average model, or SARIMA$(0,0,2)(0,0,2)_{46}$ to the pre-processed SeaWiFS NPP data (i.e., moving average order $q = 2$, seasonal moving average $Q = 2$, and seasonal component $s = 46$ (365 days/8-day time series = 46)).

$$X_t = \delta + (1 + \theta_1 B + \theta_2 B^2)(1 + \Theta_1 B^{46} + \Theta_2 B^{92})W_t \tag{C.1}$$

In Eq. C.1, X_t is the total NPP (millions of lbs C/year) estimated across the MHI in 8-day time periods from 1998-2007 where $t \in \{1, 460\}$, δ represents the mean of the time series, θ_1 and θ_2 are the moving average parameters, Θ_2 and Θ_2 are the seasonal moving average parameters, and $W_t \overset{iid}{\sim} N(0, \sigma_w^2)$. The backshift operator, B, is defined as $BW_t = W_{t-1}$, $B^2W_t = W_{t-2}$, $B^{46}W_t = W_{t-46}$, and $B^{92}W_t = W_{t-92}$. Our model, Eq. C.1, was chosen based on information criteria. Parameters were estimated using maximum likelihood estimation (Hyndman 2017, Hyndman and Khandakar 2008). The parameter

estimates can be found in Table C.1. The estimate of parameter δ was used for the value of ν when it was treated as a fixed constant in the hierarchical food web model.

Table C.1: Seasonal ARIMA parameter estimates for Eq. C.1 and their respective standard error calculated using R v.3.4.3 (R Core Team 2017) forecast package v8.2 and function auto.arima() (Hyndman 2017, Hyndman and Khandakar 2008)

Parameters	Estimates	Standard error
δ	87339861.7	696375.7
θ_1	0.9806	0.0413
θ_2	0.4817	0.0359
Θ_1	0.3379	0.0470
Θ_2	0.3392	0.0449
σ_w^2	1.494306e+13	

C.1.2 Transfer Efficiency

We utilized data gathered from a literature review (i.e., Chapter 2: "Tangled is the web we weave") on transfer efficiencies to fit an approximate distribution (τ_h). The data includes articles that mentioned both food web and transfer efficiency, and we then selected from these studies only those that included data whether model-based or empirical. We used all of the marine data (i.e., both model-based and empirical) to choose an approximate distribution. The resulting distributional assumption was placed on transfer efficiencies (τ_h) for all trophic levels $h \in \{2, 3, 4\}$. In order to determine an approximate distribution, we ran goodness-of-fit tests. We started off with skewness-kurtosis plots (Fig. C.5) to initially decide which distributions to consider (i.e., Beta and Gamma) (Delignette-Muller and Dutang 2015). Then we fitted individual distributions to the data using maximum likelihood estimation and compared density plots of the fitted distributions to the histogram of the empirical distribution, as well as a cumulative distribution (CDF) plot of both the empirical distribution and the fitted distributions, Q-Q plots, and P-P plots (Fig. C.6). We choose the Beta amongst the approximate distributions, because the transfer efficiency is a percentage which immediately scales to

a proportion and the Beta distribution is defined within the range [0, 1]. Therefore, we concluded that the Beta distribution was the most appropriate amongst our approximate distributions (versus the Gamma distribution). The probability distribution function (PDF) and equations for the mean and variance of a Beta distribution can be found in Table C.2.

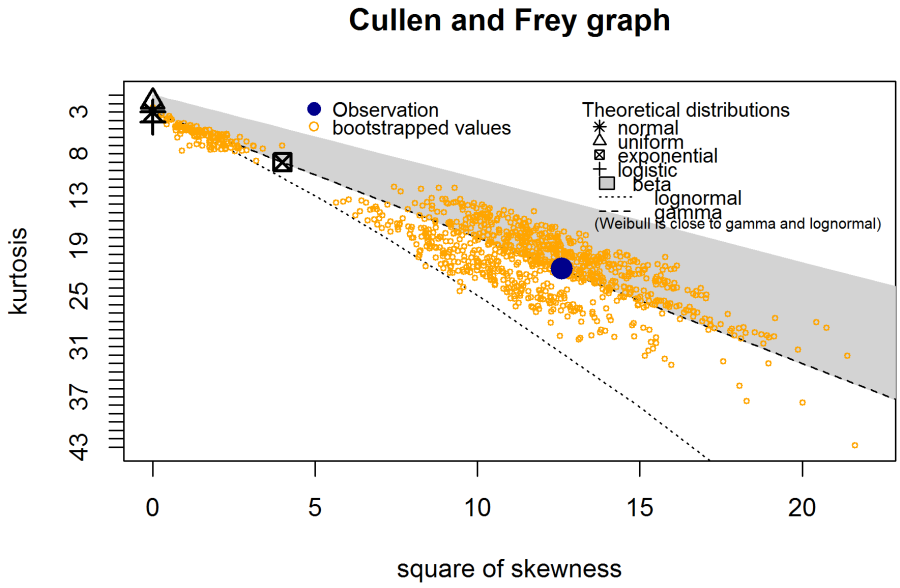

Figure C.5: Visualizes several potential continuous distributions against the marine transfer data (Case 2) and bootstrapped data. The figure shows it potentially follows a Beta and Gamma distribution. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

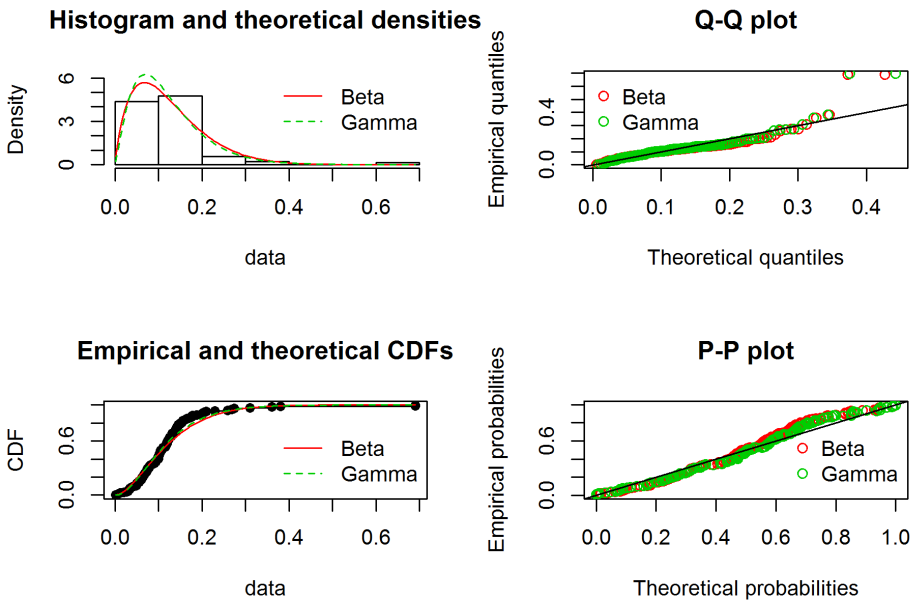

Figure C.6: Density plots of the fitted distributions to the histogram of the empirical distribution using the marine transfer efficiency data (Case 2), a CDF plot of both the empirical distribution and the fitted distributions (Beta and Gamma), Q-Q plots, and P-P plots. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

C.1.3 FishBase

Trophic Level

Estimated trophic levels (rational numbers) were calculated from the FishBase database (`http://www.fishbase.org/`) for each species and truncated to cluster species into integer-valued trophic levels. Truncated trophic levels range in the MHI from $h = 1, \ldots, 4$ with $h = 1$ representing phytoplankton, $h = 2$ primary consumers, $h = 3$ secondary consumers, and $h = 4$ tertiary consumers. If the rational trophic level was greater than or equal to 4, then the organism was put into trophic level 4.

Maximum Expected Lifespan

The maximum expected lifespan λ_4 for species at trophic level $h = 4$ was treated as a random value. We ran goodness-of-fit tests to find an approximate distribution. We started off with a skewness-kurtosis plot (Fig. C.7) to initially decide which distributions to consider (i.e., Lognormal, Weibull, and Gamma) (Delignette-Muller and Dutang 2015). Then we fitted individual distributions to the data using maximum likelihood estimation and compared density plots of the fitted distributions to the histogram of the empirical distribution, a cumulative distribution (CDF) plot of both the empirical distribution and the fitted distributions, Q-Q plots, and P-P plots (Fig. C.8). We found for all trophic levels the most appropriate approximate distribution amongst those we considered was a Lognormal distribution, where each trophic level has distinct values for μ and σ^2.

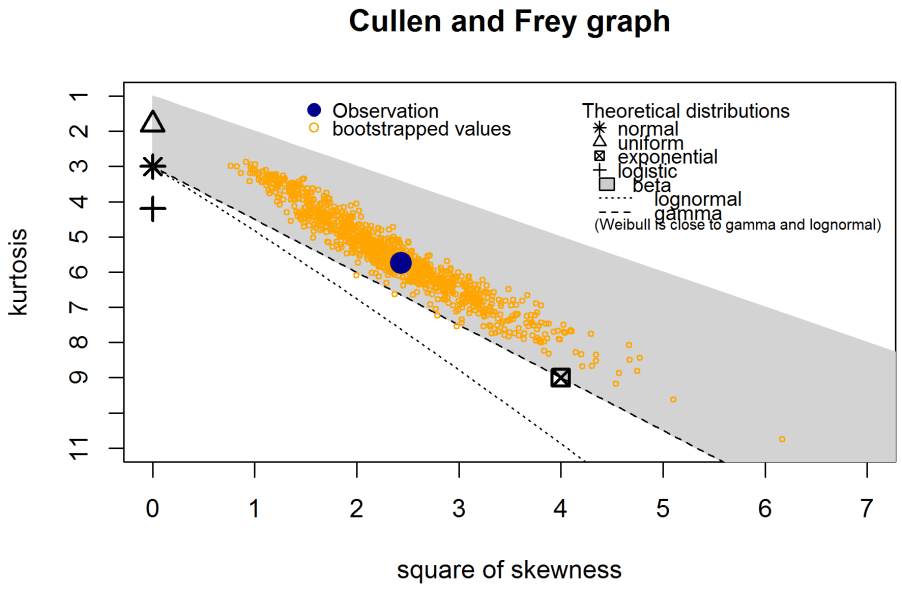

Figure C.7: Visualizes several potential continuous distributions against the maximum expected lifespan data for trophic level 4 and bootstrapped data. The figure shows it potentially follows a Gamma, Weibull, and Lognormal distribution. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

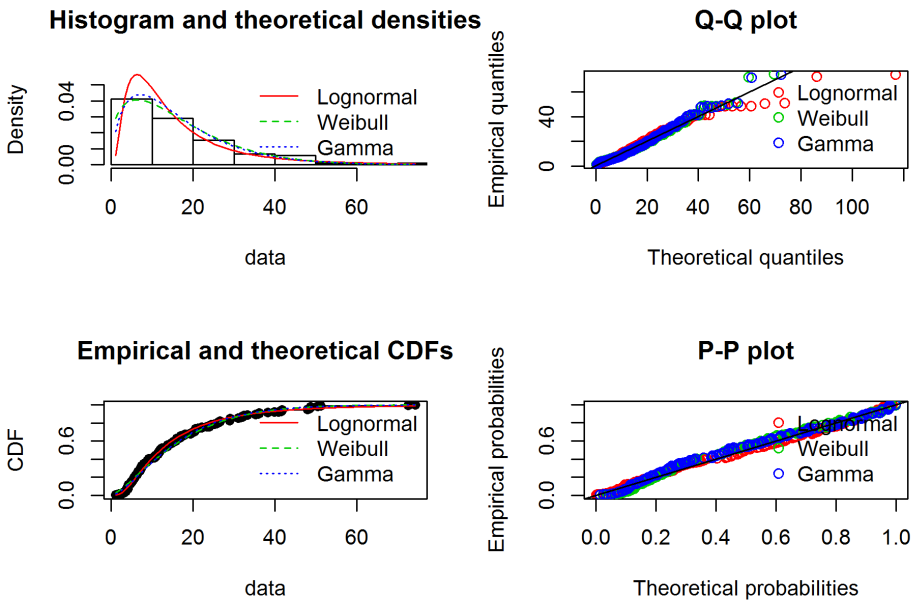

Figure C.8: Density plots of the fitted distributions to the histogram of the empirical distribution using the maximum expected lifespan data for trophic level 4, a CDF plot of both the empirical distribution and the fitted distributions (Gamma, Weibull, and Lognormal), Q-Q plots, and P-P plots. Plot created using R v.3.4.3 (R Core Team 2017) fitdistrplus package v.1.0 – 9 (Delignette-Muller and Dutang 2015).

C.2 Table of Distributions

Table C.2: Table of Distributions

Distribution	PDF	Mean & Variances
Lognormal	$f(\nu;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma\nu} exp\left(-\frac{(log(\nu)-\mu)^2}{2\sigma^2}\right)$ for $\nu > 0$	$E[\nu] = e^{\mu+\frac{\sigma^2}{2}}$ $Var[\nu] = e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$
Beta	$f(\tau_h;\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\tau_h^{\alpha-1}(1-\tau_h)^{\beta-1}$ $0 \leq \tau_h \leq 1$	$E[\tau_h] = \frac{\alpha}{\alpha+\beta}$ $Var[\tau_h] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Normal	$f(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$ for $y > 0$	$E[y] = \mu$ $Var[y] = \sigma^2$

C.3 R packages and versions

To promote reproducibility, we include a list of all R packages and versions used in this analysis.

- astsa package v.1.8 (Stoffer 2017)
- data.table package v.1.10.4 − 3 (Dowle and Srinivasan 2017)
- ExtDist package v.0.6 − 3 (Wu et al. 2015)
- fitdistrplus package v.1.0 − 9 (Delignette-Muller and Dutang 2015)
- forecast package v8.2 (Hyndman 2017, Hyndman and Khandakar 2008)
- ggplot2 package v.2.2.1 (Wickham 2009)
- gridExtra package v.2.3 (Auguie 2017)
- gtable package v.0.2.0 (Wickham 2016)
- MCMCpack package v.1.4 − 2 (Martin et al. 2011)
- logspline package v.2.1.9 (Kooperberg 2016)
- R v.3.4.3 (R Core Team 2017)
- reshape package v.0.8.7 (Wickham 2007)
- xtable package v.1.8 − 2 (Dahl 2009)

Bibliography

Antoine, D., André, J.-M., and Morel, A. (1996). Oceanic primary production 2. *Estimation at global scale from satellite (coastal zone color scanner) chlorophyll. Global Biochemical Cycles*, 10(1):57–69.

Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. R package version 2.3.

Baird, D., Asmus, H., and Asmus, R. (2004). Energy flow of a boreal intertidal ecosystem, the Sylt-Rømø Bight. *Marine Ecology Progress Series*, 279:45–61.

Banas, N., Hickey, B., Newton, J., and Ruesink, J. (2007). Tidal exchange, bivalve grazing, and patterns of primary production in Willapa Bay, Washington, USA. *Marine Ecology Progress Series*, 341:123–139.

Barnes, C., Maxwell, D., Reuman, D. C., and Jennings, S. (2010). Global patterns in predator–prey size relationships reveal size dependency of trophic transfer efficiency. *Ecology*, 91(1):222–232.

Baumann, M. (1995). A comment on transfer efficiencies. *Fisheries Oceanography*, 4(3):264–266.

Behrenfeld, M. J. and Falkowski, P. G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and oceanography*, 42(1):1–20.

Brodziak, J., Courtney, D., Wagatsuma, L., O'Malley, J., Lee, H.-H., Walsh, W., Andrews, A., Humphreys, R., and DiNardo, G. (2011). Stock assessment of the main Hawaiian Islands Deep 7 bottomfish complex through 2010. Technical Report NOAA Tech. Memo. NOAA-TM-NMFS-PIFSC-29, U.S. Dep. Commer.

Brodziak, J., Moffitt, R., and DiNardo, G. (2009). Hawaiian bottomfish assessment update for 2008. Technical Report Pacific Islands Fish. Sci. Cent. Admin Rep. H-09-02, Pacific Islands Fish. Sci. Cent., Natl. Mar. Fish. Ser.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.

Byron, C., Link, J., Costa-Pierce, B., and Bengtson, D. (2011). Calculating ecological carrying capacity of shellfish aquaculture using mass-balance modeling: Narragansett Bay, Rhode Island. *Ecological Modelling*, 222(10):1743–1755.

Casella, G. and Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.

Cashion, T., Hornborg, S., Ziegler, F., Hognes, E. S., and Tyedmers, P. (2016). Review and advancement of the marine biotic resource use metric in seafood LCAs: a case study of Norwegian salmon feed. *The International Journal of Life Cycle Assessment*, 21(8):1106–1120.

Chapman, J. L. and Reiss, M. J. (1998). *Ecology: principles and applications*. Cambridge University Press.

Chassot, E., Bonhommeau, S., Dulvy, N. K., Mélin, F., Watson, R., Gascuel, D., and Le Pape, O. (2010). Global marine primary production constrains fisheries catches. *Ecology letters*, 13(4):495–505.

Coll, M., Libralato, S., Tudela, S., Palomera, I., and Pranovi, F. (2008). Ecosystem overfishing in the ocean. *PLOS ONE*, 3(12):e3881.

Condon, R. H., Steinberg, D. K., Giorgio, P. A. d., Bouvier, T. C., Bronk, D. A., Graham, W. M., and Ducklow, H. W. (2011). Jellyfish blooms result in a major microbial respiratory sink of carbon in marine systems. *PNAS*, 108(25):10225–10230.

Cope, J. M., DeVore, J., Dick, E., Ames, K., Budrick, J., Erickson, D. L., Grebel, J., Hanshew, G., Jones, R., Mattes, L., et al. (2011). An approach to defining stock complexes for US West Coast groundfishes using vulnerabilities and ecological distributions. *North American Journal of Fisheries Management*, 31(4):589–604.

Cury, P., Shannon, L., Roux, J., Daskalov, G., Jarre, A., Moloney, C., and Pauly, D. (2005). Trophodynamic indicators for an ecosystem approach to fisheries. *ICES Journal of Marine Science: Journal du Conseil*, 62(3):430–442.

Dahl, D. B. (2009). xtable: Export tables to latex or html. *R package version*, pages 1–5.

Delignette-Muller, M. L. and Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, 64(4):1–34.

Dowle, M. and Srinivasan, A. (2017). *data.table: Extension of 'data.frame'*. R package version 1.10.4-3.

Farmer, N. A., Malinowski, R. P., McGovern, M. F., and Rubec, P. J. (2016). Stock complexes for fisheries management in the Gulf of Mexico. *Marine and coastal fisheries*, 8(1):177–201.

Froese, R. and Pauly, D. (2017). FishBase. World Wide Web electronic publication. `www.fishbase.org`. Online; accessed 06/2017.

Gaedke, U. and Straile, D. (1994). Seasonal changes of trophic transfer efficiencies in a plankton food web derived from biomass size distributions and network analysis. *Ecological Modelling*, 75:435–445.

Garcia, S., Kolding, J., Rice, J., Rochet, M.-J., Zhou, S., Arimoto, T., Beyer, J., Borges, L., Bundy, A., Dunn, D., et al. (2012). Reconsidering the consequences of selective fisheries. *Science*, 335(6072):1045–1047.

García-Comas, C., Sastri, A. R., Ye, L., Chang, C.-Y., Lin, F.-S., Su, M.-S., Gong, G.-C., and Hsieh, C.-h. (2016). Prey size diversity hinders biomass trophic transfer and predator size diversity promotes it in planktonic communities. *Proc. R. Soc. B*, 283(1824):20152129.

Gaunt, R. E. et al. (2018). Products of normal, beta and gamma random variables: Stein operators and distributional theory. *Brazilian Journal of Probability and Statistics*, 32(2):437–466.

Givens, G. H. and Hoeting, J. A. (2012). *Computational statistics*, volume 710. John Wiley & Sons.

Gregg, W. W., Conkright, M. E., Ginoux, P., O'Reilly, J. E., and Casey, N. W. (2003). Ocean primary production and climate: Global decadal changes. *Geophysical Research Letters*, 30(15).

Hairston, N. G. (1993). Cause-effect relationships in energy flow, trophic structure, and interspecific interactions. *The American Naturalist*, 142(3):379–411.

Han, D., Chen, Y., Zhang, C., Ren, Y., Xue, Y., and Wan, R. (2017). Evaluating impacts of intensive shellfish aquaculture on a semi-closed marine ecosystem. *Ecological Modelling*, 359:193–200.

Hastings, A., Gaines, S. D., and Costello, C. (2017). Marine reserves solve an important bycatch problem in fisheries. *Proceedings of the National Academy of Sciences*, 114(34):8927–8934.

Havens, K. E. (1998). Size structure and energetics in a plankton food web. *Oikos*, 81(2):346–358.

Heymans, J., Coll, M., Libralato, S., and Christensen, V. (2011). 9.06 - Ecopath theory, modeling, and application to coastal ecosystems. In Wolanski, E. and McLusky, D., editors, *Treatise on Estuarine and Coastal Science*, pages 93 – 113. Academic Press, Waltham.

Hilborn, R. and Walters, C. (1992). *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty*. Springer.

Hothorn, T. and Zeileis, A. (2015). partykit: A modular toolkit for recursive partytioning in R. *The Journal of Machine Learning Research*, 16(1):3905–3909.

Hourigan, T. F. and Reese, E. S. (1987). Mid-ocean isolation and the evolution of Hawaiian reef fishes. *Trends in ecology & evolution*, 2(7):187–191.

Hyndman, R. J. (2017). *forecast: Forecasting functions for time series and linear models*. R package version 8.2.

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

Irisarri, J., Fernández-Reiriz, M. J., Robinson, S. M., Cranford, P. J., and Labarta, U. (2013). Absorption efficiency of mussels Mytilus edulis and Mytilus galloprovincialis cultured under integrated multi-trophic aquaculture conditions in the Bay of Fundy (Canada) and Ría Ares-Betanzos (Spain). *Aquaculture*, 388:182–192.

Iverson, R. L. (1990). Control of marine fish production. *Limnology and Oceanography*, 35(7):1593–1604.

Karlsson, J., Lymer, D., Vrede, K., and Jansson, M. (2007). Differences in efficiency of carbon transfer from dissolved organic carbon to two zooplankton groups: an enclosure experiment in an oligotrophic lake. *Aquatic Sciences; Basel*, 69(1):108–114.

Kooperberg, C. (2016). *logspline: Logspline Density Estimation Routines*. R package version 2.1.9.

Lalli, C. and Parsons, T. R. (1997). *Biological Oceanography: An Introduction*. Butterworth-Heinemann.

Lamport, L. (1994). *LATEX: a document preparation system: user's guide and reference manual*. Addison-wesley.

Langseth, B., Syslo, J., Yau, A., Kapur, M., and Brodziak, J. (2018). Stock assessment for the Main Hawaiian Islands Deep 7 Bottomfish Complex in 2018, with catch projections through 2022. Technical Report NOAA Tech. Memo. NOAA-TM-NMFS-PIFSC-69, U.S. Dep. Commer.

Liaw, A. and Wiener, M. (2016). Classification and regression by randomforest. r news. 2002; 2 (3): 18–22.

Libralato, S., Coll, M., Tudela, S., Palomera, I., and Pranovi, F. (2008). Novel index for quantification of ecosystem effects of fishing as removal of secondary production. *Marine Ecology Progress Series*, 355:107–129.

Lindeman, R. L. (1942). The trophic-dynamic aspect of ecology. *Ecology*, 23(4):399–417.

Lobel, P. and Robinson, A. (1986). Transport and entrapment of fish larvae by ocean mesoscale eddies and currents in Hawaiian waters. *Deep Sea Research Part A. Oceanographic Research Papers*, 33(4):483–500.

Martin, A. D., Quinn, K. M., and Park, J. H. (2011). Mcmcpack: Markov chain monte carlo in r.

May, R. (1976). *Theoretical ecology: principles and applications*. Philadelphia: Saunders.

May, R. and McLean, A. R. (2007). *Theoretical Ecology: Principles and Applications*. OUP Oxford.

May, R. M. (1983). Ecology: The structure of food webs. *Nature*, 301(5901):566–568.

McAllister, M. and Kirkwood, G. (1998). Bayesian stock assessment: a review and example application using the logistic model. *ICES Journal of Marine Science*, 55(6):1031–1060.

McAllister, M. K. and Ianelli, J. N. (1997). Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Sciences*, 54(2):284–300.

McGarvey, R., Dowling, N., and Cohen, J. E. (2018). Two processes regulating trophic energy flow in pelagic and terrestrial ecosystems: Trophic efficiency and body size–dependent biomass production: (a reply to Giacomini). *The American Naturalist*, 191(3):364–367.

McIntosh, R. P. (1986). *The Background of Ecology: Concept and Theory*. Cambridge University Press.

Morel, A. and André, J.-M. (1991). Pigment distribution and primary production in the western mediterranean as derived and modeled from coastal zone color scanner observations. *Journal of Geophysical Research: Oceans*, 96(C7):12685–12698.

Papatryphon, E., Petit, J., Kaushik, S. J., and van der Werf, H. M. (2004). Environmental impact assessment of salmonid feeds using life cycle assessment (LCA). *AMBIO: A Journal of the Human Environment*, 33(6):316–323.

Pauly, D. and Christensen, V. (1995). Primary production required to sustain global fisheries. *Nature*, 374(6519):255–257.

Pelletier, N. and Tyedmers, P. (2007). Feeding farmed salmon: Is organic better? *Aquaculture*, 272(1-4):399–416.

Pelletier, N., Tyedmers, P., Sonesson, U., Scholz, A., Ziegler, F., Flysjo, A., Kruse, S., Cancino, B., and Silverman, H. (2009). Not all salmon are created equal: life cycle assessment (LCA) of global salmon farming systems.

Persson, J., Brett, M. T., Vrede, T., and Ravet, J. L. (2007). Food quantity and quality regulation of trophic transfer between primary producers and a keystone grazer (Daphnia) in pelagic freshwater food webs. *Oikos*, 116(7):1152–1163.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ripley, B. (2016). Tree: Classification and regression trees. r package version 1.0-37. *Avaliable at https://CRAN. R-project. org/package= tree. Accessed August*, 3:2017.

Rohatgi, V. K. and Saleh, A. M. E. (2015). *An introduction to probability and statistics*. John Wiley & Sons.

Rosland, R., Strand, Ø., Alunno-Bruscia, M., Bacher, C., and Strohmeier, T. (2009). Applying dynamic energy budget (DEB) theory to simulate growth and bio-energetics of blue mussels under low seston conditions. *Journal of Sea Research*, 62(2-3):49–61.

Rybarczyk, H. and Elkaım, B. (2003). An analysis of the trophic network of a macrotidal estuary: the Seine Estuary (Eastern Channel, Normandy, France. *Estuarine, Coastal and Shelf Science*, 58(4):775–791.

Ryther, J. H. et al. (1969). Photosynthesis and fish production in the sea. the production of organic matter and its conversion to higher forms of life vary throughout the world ocean. *Science (Washington)*, 166:72–76.

San Martin, E., Harris, R. P., and Irigoien, X. (2006a). Latitudinal variation in plankton size spectra in the Atlantic Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 53(14):1560–1572.

San Martin, E., Harris, R. P., and Irigoien, X. (2006b). Latitudinal variation in plankton size spectra in the atlantic ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 53(14-16):1560–1572.

Schaefer, M. B. (1954). Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. *Inter-American Tropical Tuna Commission Bulletin*, 1(2):23–56.

Schaefer, M. B. (1965). The potential harvest of the sea. *Transactions of the American Fisheries Society*, 94(2):123–128.

Schmitz, O. J., Grabowski, J. H., Peckarsky, B. L., Preisser, E. L., Trussell, G. C., and Vonesh, J. R. (2008). From individuals to ecosystem function: toward an integration of evolutionary and ecosystem ecology. *Ecology*, 89(9):2436–2445.

Semper, C. (1881). *Animal Life as Affected by the Natural Conditions of Existence*. D. Appleton.

Shanno, D. F. (1970). Conditioning of quasi-newton methods for function minimization. *Mathematics of computation*, 24(111):647–656.

Sheldon, R., Sutcliffe Jr, W., and Paranjape, M. (1977). Structure of pelagic food chain and relationship between plankton and fish production. *Journal of the Fisheries Board of Canada*, 34(12):2344–2353.

Simenstad, C. A. and Fresh, K. L. (1995). Influence of intertidal aquaculture on benthic communities in Pacific Northwest estuaries: scales of disturbance. *Estuaries*, 18(1):43–70.

Slobodkin, L. B. (1959). Energetics in Daphnia pulex populations. *Ecology*, 40(2):232–243.

Slobodkin, L. B. (1960). Ecological energy relationships at the population level. *The American Naturalist*, 94(876):213–236.

Slobodkin, L. B. (1962). Energy in animal ecology. *Advances in ecological research*, 1:69–101.

Slobodkin, L. B. (1972). On the inconstancy of ecological efficiency and the form of ecological theories. *Transactions of the Connecticut Academy of Arts and Sciences*, 44:293–305.

Sommer, U. (1998). From algal competition to animal production: enhanced ecological efficiency of Brachionus plicatilis with a mixed diet. *Limnology and Oceanography*, 43(6):1393–1396.

Springer, M. and Thompson, W. (1970). The distribution of products of beta, gamma and gaussian random variables. *SIAM Journal on Applied Mathematics*, 18(4):721–737.

Srisunont, C. and Babel, S. (2016). Estimating the carrying capacity of green mussel cultivation by using net nutrient removal model. *Marine pollution bulletin*, 112(1-2):235–243.

Stock, C. A., John, J. G., Rykaczewski, R. R., Asch, R. G., Cheung, W. W., Dunne, J. P., Friedland, K. D., Lam, V. W., Sarmiento, J. L., and Watson, R. A. (2017). Reconciling fisheries catch and ocean productivity. *Proceedings of the National Academy of Sciences*, 114(8):E1441–E1449.

Stoffer, D. (2017). *astsa: Applied Statistical Time Series Analysis*. R package version 1.8.

Strathmann, R. R. (1967). Estimating organic carbon content of phytoplankton from cell volume or plasma volume. *Limnol Oceanogr*, 12(3):411–418.

Tantau, T. (2015). *The TikZ and PGF Packages*.

Therneau, T., Atkinson, B., and Ripley, B. (2015). rpart: Recursive partitioning and regression trees. r package version 4.1–10.

Trebilco, R., Baum, J. K., Salomon, A. K., and Dulvy, N. K. (2013). Ecosystem ecology: size-based constraints on the pyramids of life. *Trends in ecology & evolution*, 28(7):423–431.

Trussell, G. C., Ewanchuk, P. J., and Matassa, C. M. (2006). The fear of being eaten reduces energy transfer in a simple food chain. *Ecology*, 87(12):2979–2984.

Vermeij, G. J. (1987). The dispersal barrier in the tropical pacific: implications for molluscan speciation and extinction. *Evolution*, 41(5):1046–1058.

von Elert, E., Martin-Creuzburg, D., and Le Coz, J. R. (2003). Absence of sterols constrains carbon transfer between cyanobacteria and a freshwater herbivore (Daphnia galeata). *Proceedings of the Royal Society B: Biological Sciences*, 270(1520):1209–1214.

Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J., Fromentin, J.-M., Hoegh-Guldberg, O., and Bairlein, F. (2002). Ecological responses to recent climate change. *Nature*, 416(6879):389.

Ware, D. M. (2000). Aquatic ecosystems: properties and models. *Fisheries Oceanography: An Integrative Approach to Fisheries Ecology and Management. Edited by PJ Harrison and TR Parson, Blackwell Science, Oxford*, pages 267–295.

Watson, R., Zeller, D., and Pauly, D. (2014). Primary productivity demands of global fishing fleets. *Fish and Fisheries*, 15(2):231–241.

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12).

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H. (2016). *gtable: Arrange 'Grobs' in Tables*. R package version 0.2.0.

Wickham, H., Francois, R., Henry, L., and Müller, K. (2015). dplyr: A grammar of data manipulation. *R package version 0.4*, 3.

Wu, H., Godfrey, A. J. R., Govindaraju, K., and Pirikahu, S. (2015). *ExtDist: Extending the Range of Functions for Probability Distributions*. R package version 0.6-3.