UNIVERSITY OF CALIFORNIA

SANTA CRUZ

ASSESSMENT IN SALMON AND GROUNDFISH FISHERIES

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

DOCTOR OF PHILOSOPHY

in

Ocean Sciences

by

Andi Stephens

December 2005

The Dissertation of Andi Stephens is approved:

Professor Marc Mangel, Chair

Dr. Alec MacCall

Christopher Edwards

Professor Christopher Edwards

Professor Herbert Lee

L. C. Sloan

Lisa C. Sloan
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

Chapter 1.

Chapter 2.

Chapter 3.

List of Tables

Chapter 1.

Chapter 3.

ASSESSMENT IN SALMON AND GROUNDFISH FISHERIES

Andi Stephens

Management of sustainable fisheries requires the development and application of procedures to assess the abundance and productivity of stocks, as well as risks to the fishery. Such assessment takes place in a setting of environmental variability, limited information, uncertainty, and, as exploitation increases, changes in ecosystem structure. In spite of these difficulties, fishery models must provide adequate advice for managers and policy makers. This document addresses several issues in the assessment and management of marine fisheries in the presence of change, uncertainty, and risk.

I first focus on an investigation of risk: the development of a model for the assessment of risk to wild salmon from aquaculture and from escaping aquaculture salmon. This age-structured population dynamics model was developed with the aim of addressing both risk to the wild population and communication of that risk to policy makers and stakeholders in the fishery. Towards that end, the model is embedded in a menu-driven interface, enabling a hands-on investigation of a variety of possible ecological interactions between the species.

I then turn to a method for assessing fishing effort in a mixed-target fishery. Many recreational and artisanal fisheries can shift between habitats (e.g., from fishing onshore to offshore species), and records of catch do not reflect the shift explicitly. This makes calculation of the catch per unit effort (CPUE, an index of abundance) for a particular

species difficult. In order to assess the amount of effort that should be included in a species CPUE, a logistic regression technique was employed, using the species composition of catch to infer fishing targets in catch histories from the California recreational fishery. The multispecies method was then used on data simulated to suggest conditions that might occur in the actual fishery, in order to evaluate its effectiveness.

Acknowledgements

We go through life deeply grateful that someone has taught us how to read; we are aware of the worlds this has opened up to us, yet this learning occurs before we are able to recognize the enormity of the gift, and we never have the chance to thank those who so profoundly changed our lives. My advisor Marc Mangel has given me new literacy in mathematics, opening up a wide window on the processes of science, and has given me a new set of tools with which to examine the world. Along the way he has impressed on me the significance of the philosophical underpinnings of our work, and has especially highlighted the importance of communicating and teaching science. For these things as well as his generous support and encouragement I am grateful.

Alec MacCall has helped me to re-envision myself as a scientist and a student of science. Few people have ever been so easily able to recognize the roadblocks I’ve faced and taught me to work my way around them. I am grateful for his kindness and support, and for the example he has shown me as a mentor.

Chris Edwards had a tremendously clarifying impact on my thinking. I am sorry I did not have his guidance earlier in my work, but I am grateful for the time he has spent improving my analyses.

Herbie Lee has been a kind teacher and a good friend. It is hard to say how much that matters. I am certain I could not have made it through this process without him.

I have been lucky to have had a number of wonderful collaborators and mentors outside the campus community who have taught, encouraged and inspired me, especially Andy Rosenberg, Andy Cooper, Kai Lorenzen, Mike Bonsall, Ian Fleming, and Simon Levin.

Many others have contributed to my professional development. I'd like to thank the members of the Ocean Sciences Department and the Department of Applied Math and Statistics. Both are warm communities with outstanding faculty and students, and I was blessed to be a part of both. I am especially grateful to Peggy Delaney, Mary Silver, Raphe Kudela, Jon Zehr, Diana Joy Austin, Meyo Lopez, David Draper, Bruno Sansó and Raquel Prado, and Anasthasios Kottas.

For the last three years, the faculty and students in the Center for Informal Learning and Schools have been a tremendous influence on my thinking, as well as a warm and supportive community of colleagues. I especially thank Christie Rowe, Marina Ramon, Dave Cordes, Scott Seagroves, Hoyt Peckham, Sally Duensing and Candice Brown.

I have been blessed with many friends in Santa Cruz, on and off campus: Susan Mangel, Alex Van Zyl, Kate Chabarek, A.J. Tedesco, John Balawejder, Karianne Terry, Alison Gong, John and Wilma Field, Dave Johnston, Sebastian, Maike and Esben Rost, Nick Wolf and Katrina Dlugosch, and Steve and Jackie Rodriguez. I owe a debt of gratitude to you all.

Penultimately, I need to thank the members of the Mangel Lab past and present, but especially Holly Kindsvatter, Kate Siegfried, E.J. Dick, Yasmin Lucero, Leah Johnson, Dan Merl, Chris Simon and Anand Patil.

Then there are the old friends, without whom I wouldn't be me: Mary Ellen Tunney, Carolyn Dare Councill, and Lawrence Llewelyn Butcher.

Thank you all.

Attribution

The work described in Chapter 2 was published in 2004 (Stephens and MacCall, 2004). Alec MacCall was my coauthor, advising me on the details of logistic regression and contributing background information about the California Recreational Fishery.

Introduction

In a review of 200 fish stocks, the Food and Agriculture Organization of the United Nations found that 35 percent of stocks were in decline (FAO 1997). Depletion of popular stocks such as orange roughy has lead to the development of new fisheries for previously unexploited species, and these fisheries too may overestimate the resilience of stocks. If such trends are to be resisted, it is increasingly important that methods be developed for rigorous evaluation of fish populations and the risks posed to those populations. In this thesis I investigate methods for the analysis of stocks, employing complementary statistical and mathematical approaches which may contribute to better-informed management in marine fisheries.

A population of fish – a stock – is a renewable resource, sustained by the reproduction of the adults that have survived predation, disease, fishing, and environmental hazards. If the reproductive rate balances or exceeds the mortality rate, the stock remains viable. If on the other hand the rate of reproduction is lower than the mortality rate, the stock will eventually become extinct. Anthropogenic impacts often affect the balance between viability and extinction; these impacts may be in the form of harvest, removal of water for agriculture or mining, discharge of pollutants, changes to coastal or riverine environments, or, as in the case of salmon aquaculture, the introduction of non-native species (Pikitch, *et al.*, 2004). They impinge upon systems already subject to a great deal of environmental stochasticity due, for example, to seasonal fluctuations in weather pattern, interannual effects such as El Niño ocean conditions, and interdecadal changes in climate regime such as the North Atlantic Oscillation (NAO) or Pacific Decadal Oscillation (PDO).

Fishery management in such an environment has evolved to follow a sustained-yield model, which depends upon targeting a conservative yield (McEvoy, 1986). Historically, fisheries have existed as a common resource, with access open to all. Under these conditions, a rational strategy for each fisher is to take as much of the fish as he can, superseding others, in order to maximize his profit (Iudicello, *et al.*, 1999). This "race for fish" quickly leads to extinction of the stock – a "tragedy of the commons" (Hilborn, *et al.*, 2004; Hilborn, *et al.*, 2005). If a fishery is to survive, the fishing industry must respect the biological limits of the stock's productivity. The success of a sustained-yield fishery relies on setting fishing limits so that fishery yield, the surplus production of the stock beyond what is needed to sustain it, is maintained at a steady rate.

Setting these limits requires (at a minimum) adequate knowledge of the size of the stock, and its reproductive rate. It may also require some sensitivity to potential risks to the fishery from external (e.g., anthropogenic) sources. For all but a few freshwater species, these values can never be known precisely, but must be estimated according to various statistical or mathematical methods.

In the first part of the thesis, I describe a model designed for forecasting the fate of wild Atlantic salmon faced with various impacts from salmon aquaculture. Salmon aquaculture increasingly threatens already dwindling wild populations, and lessons learned in the management of risks to salmon can inform the aquaculture practices in the farming of other marine finfish species (Naylor, *et al.*, 2005). This age-structured model tracks population changes under a variety of ecological scenarios, permitting a comparative analysis of the risks to the wild fish. The model simulates competitive effects between wild fish and escaped aquaculture fish, the physiological growth and reproduction of escaped aquaculture fish, the transfer of disease from the aquaculture facility to wild fish, and the impact of the alternative mating strategy employed by fish in the freshwater stages of life. These are risks that have long been identified for wild salmon (e.g., Hindar, *et al.*, 1994; Hansen, *et al.*, 1997), however their relative impacts are unknown. The results of this model represents the first attempt to quantify those effects.

The latter part of the thesis focuses on problems associated with assessing the state of a particular species when fishery census information is available only as aggregate data about a group of species. For this work, I investigate using a statistical model to relate the ecological environment to the population status.   This is an explanatory model that obviates the predictive relationships between species, and allows us to make inferences about patterns of habitat use by the various species within a mixed fishery.

In the marine environment, direct observational studies are often difficult due to spatial scales and environmental constraints, and experimentation can be costly. When the subject of inquiry is a population of relatively long-lived organisms, the time-scales of interest may be too long for experimental manipulation. Experiments on endangered populations are at best ill-advised. In cases such as these, modeling approaches provide important insight into otherwise intractable questions.

The relationship of a model to the system modeled is often compared to the relationship of a map to a landscape. The map reduces three dimensions (really four) to two. For purposes of clarity, the map contains only those landscape features of special interest, such as interstate highways, represented in miniature. As those of us who have traveled Route 66 know, it is important that the map be current, reflecting those features that exist today, and today’s map must change in order to be useful in the future.

Like the map, a model reduces the dimensionality of the system, incorporating only those features essential to the questions we pursue, and it must faithfully reflect the ways in which the system may change over time. Just as the map uses a set of graphical conventions to symbolize the landscape, the model uses a set of formal mathematical symbols to describe the system it portrays. The process of designing a map is one of deciding what features of the landscape are useful for our purposes, and likewise, the process of designing a model draws our attention to features of a system, forcing us to evaluate their relevance to the questions we wish to ask.

What we can learn from models varies with the choice of model. Dynamical models allow us to examine changes in a system over time, illuminating the mechanisms of change. They permit evaluation of the relative importance of system components in contributing to system dynamics. These models may be deterministic or stochastic. Deterministic models, those that can be solved analytically, allow us to anticipate the eventual state of a system given an initial known state. An example of this is the prediction of the eventual the steady-state population size of a phytoplankton colony in a flask, and the length of time necessary for it to reach that size, given an initial number of cells of a species of known growth rate.

Stochastic models take into account the natural variability in a system. The underlying assumption in these models is that some mechanism to which the system responds is not fixed, but has a value that can be described by a probability distribution. These models cannot provide precise numerical results, but can tell us about the responses of complex systems under different experimental regimes. The salmon fishery assessment in Chapter 1 employs a population dynamics model with a stochastic component in the portion of the model addressing the success and reproduction of escaped aquaculture fish.

Statistical models quantify relationships within a system; however they provide no insight into the mechanisms operating on it. A simple example of this type of model is the prediction of annual family income based on variables such as age, education level, and geographic location of wage-earners. A model like this is constructed with mechanistic assumptions in mind, and one of the things we learn from it is which variables are most relevant to the quantity we wish to predict. For the stock assessment study of Chapter 2, I employ a logistic regression model in order to determine the amount of fishing effort applicable to a certain species of fish. The predictive variables here are other fish species caught, and the underlying mechanism assumed is that certain species prefer particular habitats, and will be found there alongside others with similar habitat preferences.

The practice of developing models is one of abstracting words into symbols, precisely describing the logic of the question at hand, and thus eliminating the complexity of natural language. Unfortunately, this elegant formalism reduces a problem that can be understood universally, such as “How many fish can we catch today without endangering next year’s harvest?” to a series of calculations not readily understood by many of the people most directly affected by their outcome, those who fish or who are responsible for regulating the fishery.

In this way, the use of the model leads to another type of scientific question; that of how to communicate the interpretation of results from *in silico* experiments for the social mileau in which they are relevant. In general, this often occurs through a filtering mechanism whereby the models and their results are first vetted by a technical committee, and then presented to management councils and the public in the form of lengthy reports full of charts and tables. Understandably, the fisherman who bears the burden of restrictive legislation may feel that decision-making on the basis of model results is at best suspect.

Fishery scientists, the press, the public and policy makers made a difference in the direction of fishery management policy by recognizing overfishing and the need for the precautionary approach (Iudicello, et al., 1999), however learning theory suggests that we are better able to understand processes when we are able to investigate them personally, in a hands-on manner (Ash and Klein, 1999; Paris, 1997). The salmon model provides a menu-driven user-interface to facilitate investigation, making the exploration of risks to the fishery available to anyone with an interest in it. Enabling policy-makers and stakeholders to investigate risks for themselves may make model outcomes more understandable, and lead to better-informed management.

Similarly, communication problems arise within the research community. As new quantitative techniques are developed and disseminated, and advances in computer software make these methods generally available, the community using those methods has become more diverse, and represents a broader spectrum of mathematical skills and preparation (Taper and Lele, 2004.). This leads to a greater need for studies that illuminate abstract methods in concrete terms, communicating the appropriate use, the strengths and limitations of mathematical methods to the researcher in terms of his own field. This is the thrust of Chapter 3, in which I relate quantitative model results to characteristics of species and of the fishery.

The work I describe in the three chapters that follow illustrates two complementary approaches to evaluating the state of fish populations and developing prognoses for their future. The first chapter describes a population dynamics model for risk assessment in salmon aquaculture. Salmon are anadromous, with a marine stage and a freshwater stage, and aquaculture impinges on them differently in these different environments. Aquaculture facilities can be quite harmful to wild species. Disease can spread rapidly from fish farms to wild species, and they attract predators in large numbers. When aquaculture fish escape, they represent the introduction of an exotic species into the environment. Aquaculture fish may compete with the wild fish or interbreed with them, disrupting the genetic structure of populations evolved for specific environments. This model represents the first attempt to analyze these interactions quantitatively. This work focuses on the potential of the model for evaluating the mechanisms that drive population change, and explores a novel approach to communicating methods and results.

In Chapter 2, I illustrate the use of a statistical model for determining population trends in a multi-species fishery. The data available in this type of fishery poses serious problems in stock assessment because it aggregates catch information for many species together. Disaggregating the data – determining which records pertain to a single species – has in the past been approached by a variety of ad-hoc methods. These often fail to percieve effort to fish a species when it is absent: the zero catch that is crucial to detecting stock depletion. I use a logistic regression on the species present in each catch to infer whether or not it could have been an effort to catch a particular species. Where it is adopted, this technique will improve the accuracy of assessments, as well as providing much-needed consistency in analysis.

In the third chapter, I analyze the performance of the multi-species method in simulated data. Straightforward statistical methods such as logistic regression are unfortunately easy to misuse, since they rely on a variety of assumptions that may not hold in a particular fishery, such as the assumption that fish habitat remains constant, an assumption that is violated regularly by environmental disturbances, such as El Niño. Regression results can be difficult to interpret, since a regression can successfully predict much of the data under circumstances in which it is malfunctioning. I investigate the strengths and limitations of the multispecies regression in terms of the characteristics of the fishery, explicitly addressing the problems of interpreting results in context.

Interactive Risk Analysis for Management of Escaped Aquacultured Salmon

Abstract

We describe an interactive model that can be used to investigate the risk posed to wild salmon by escaped aquaculture salmon. The Salmon Management Aquaculture Risk-assessment Tool (SMART) simulates a small population of wild salmon based in a particular stream/estuary/ocean system, into which an aquaculture facility is losing fish to escapes. This system is based on features of the Gulf of Maine salmon streams, and we parameterize the survival characteristics of the wild salmon from the U.S. Fish and Wildlife Service Report on Atlantic Salmon Stocks in Maine (U.S. Fish and Wildlife, 1999). The survival and reproductive success of escaping smolts is calculated using a within-year physiological model of growth and maturation. Results from the growth model and parameters from the Atlantic salmon report are used in a between-year model predicting the population trajectories of wild and aquaculture fish for a hundred years into the future. The SMART model, written using MATLAB simulation software (Mathworks, Inc., 1984-2002) presents a menu-driven interface that allows the user to investigate different types of ecological interaction scenarios, and different options for management of the escapes. The interactive nature of the model permits a hands-on sensitivity analysis that represents

an intuitive way to present information about risks to a non-technical audience**.** Results from the model suggest the most important parameters to measure in the field.

1.1. Introduction

Atlantic salmon are farmed worldwide, both within the North Atlantic, their native range, and in Pacific and Southern Hemisphere waters. They are a tremendous commercial success, and production of aquaculture salmon far outnumbers the natural production of wild salmon (Whoriskey, 2000).

Although derived from wild salmon, aquaculture salmon are not the same as the native species. Wild salmon occur as locally adapted populations that generally reproduce in the natal stream in which they originated, and sub-populations of Atlantic salmon differ genetically, reflecting local adaptations for survival (Clegg, *et al*, 2003). The popular brood stocks of aquaculture salmon are hybrids of European and American origin, and have been selected over generations to enhance their value in the market. While some features, such as fast growth, may enhance their ability to survive in the wild, other features may make them unsuited for the range or environmental conditions that natural salmon populations experience (Fleming, *et al*., 2002).

When aquaculture fish escape, they represent the introduction of a non-native species (Sakai, *et al.*, 2001), which may have serious ecological effects. These could include competitive or interference effects with wild salmon. If escaped aquaculture salmon become established, they may drive wild populations, already under pressure from habitat destruction and overfishing, further on towards certain extinction. Aquaculture salmon have been found returning to streams in Norway, Iceland, Ireland, and the Canadian east (New Brunswick) and west (British Columbia) (Whoriskey, 2000, Volpe, 2000, and Lacroix and Stokesbury, 2004).

The long-range consequences of ecological interactions between wild Atlantic salmon and escaped aquaculture salmon provides an important framework within which to conduct an ecological risk analysis. Risk analyses are common in engineering and environmental policy (Anand, 2002) , often addressing concerns around chemical hazards, but rarely used to evaluate ecological risks to a species or ecosystem. A risk analysis requires specifying the potential states of nature, their probabilities, potential management actions (note that this includes taking no action), their effects on the states of nature, and the value of each combination of state and action (Anand, 2002). In this case, the states of nature we are interested in are the kind of interactions that might occur between wild and escaped Atlantic salmon. We have developed an age-structured model of salmon population dynamics to investigate outcomes for a variety of possible interactions. This is the Salmon Management Aquaculture Risk-assessment Tool, or SMART. Integral to this model are a disease transmission model, and a physiological model of survival and reproduction potential for escaped aquaculture salmon. That is, we have modeled both within-year individual dynamics and between-year population dynamics.

We use the model to simulate a small population of wild salmon based in a particular stream/estuary/ocean system, into which an aquaculture facility is losing fish to escapes. Given the number of smolts and adults that escape each year, we calculate the changes in the populations of wild and escaped fish, projecting forward in time from 2000 to year 2100. In addition to investigating a variety of ecological risks, we investigate the impacts of various management decisions in the event of escapes. Management responses to aquaculture escapes include legislative responses, such as mandating containment in the form of secure sea-cages in aquaculture facilities, and responses in the field, such as opening salmon fishing after an escape, to remove the escapees (Goldburg, *et al*., 2001).

Information about risks is often supplied to fisheries managers and stakeholders in the form of results from complex mathematical and statistical analyses in which most of the investigation of the problem at hand has been carried out by the authors (Hilborn, *et al*., 2003). The modeling process (and any implicit decisions about precautionary approaches inherent therein) is far from transparent. However, in contributing scientific advice to discussions of public policy, it important to avoid an "elitist" stance, and to include stakeholders in the discussion (Anderson, et al., 2003).

Our approach provides an interactive risk analysis tool, to permit those involved in the policy process to conduct their own sensitivity analysis, investigating outcomes over a range of scenarios (Carpenter, 2000). Inquiry-based investigation is a development in science education towards allowing the individual to conduct hands-on experiments with a phenomenon in order to gain a better sense of its character (e.g. Ash and Klein, 1999, Paris, 1997). Central to this concept is the notion of guided inquiry: the person is given a range of ideas within which to form his own questions (Minstrell, 1999). Bringing this approach to a risk-assessment framework can provide better public insight into the science that informs policy decisions. To this end, we imbedded our model in a menu-driven user-interface that makes a wide variety of scenarios and management strategies available for investigation.

The model was parameterized as much as possible from the Atlantic Salmon Status Review (U.S. Fish and Wildlife Service, 1999). Survival rates, rates of return, and age at smolt transformation are among the values taken from the report. Other sources for life-history parameters were field studies performed in Canada and elsewhere (Hutchings and Jones, 1998, Whalen, *et al.*, 2000, Garant, *et al.*, 2003). We first describe the details of the model and interface, then show results of simulations and sensitivity analyses, and end with a discussion of implications of the model.

1.2. Methods

The following sections describe the four parts of the model: the age-structured population model for wild and escaped aquaculture fish, the within-year physiological model of survival and reproduction for escaped aquaculture smolts, the between-year model of disease transmission, and the user-interface for scenario investigation (Table 1.1).

1.2.1. Age-structured population model of wild and escaped fish (between-year model).

The age-structured model is based on annual time steps for wild and escaped fish. The wild fish are assumed to undergo smolt transformation after either 1 or 2 freshwater years and return to freshwater after either 1 or 2 ocean years. As many of the parameters as possible were estimated from data in the USFWS report on the status of Maine salmon (http://library.fws.gov/salmon). Values for these parameters and those we estimated are given in Appendix A, Tables A.1 and A.2. We assume no fishing before 1770, fishing between 1771 and 1985, and no fishing after 1985 (as happened). We assume that freshwater habitat destruction begins in 1835 and reduces freshwater habitat to 50% of its original value.

Table 1.1. Some of the assumptions of the model.

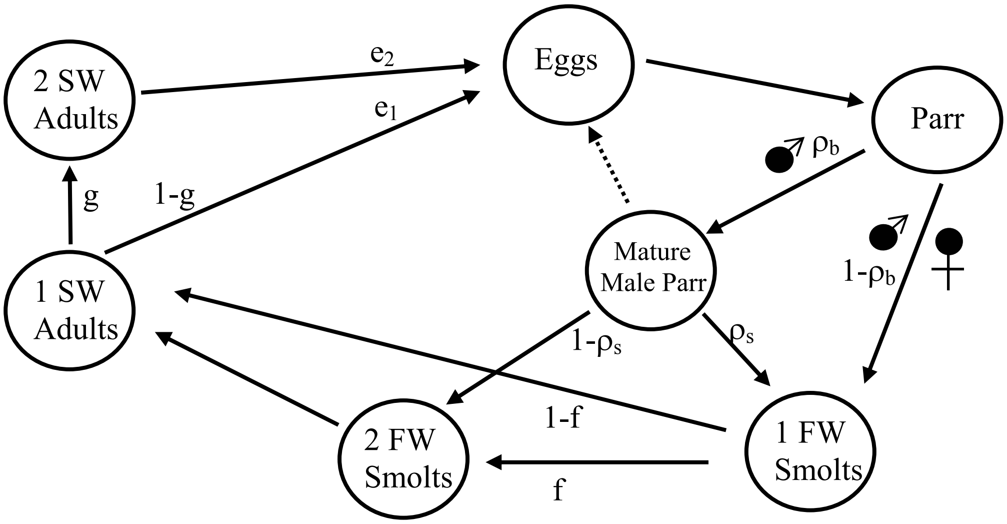
Characteristic	Description	Nominal Value
Habitat improvement	Freshwater only	1% per year
Fishing	occurs from 1771 - 1985	determines the base population of wild fish
Smolting	occurs after one or two years in freshwater	80% of parr smolt after one year
Adult returns	occur after one (grilse) or two years in the ocean	5% of smolts return after one sea-year.
Survival	Rates depend on life stage	range from 8% - 60%
Escapes	Reproduction of escaped smolts is governed by the physiological model. Escapes begin on Julian day 90 (March 31) and continue throughout the year	20 smolts/year 100 adults/year Catastrophic escapes consist of 5,000 adults and 1,000 smolts.
Freshwater Competition	egg and/or parr	choice of intensities
Competition at sea	We assume (for now) that ocean resources are non-limiting	none
Disease	Affects seawater (estuarine) out-migrating smolts	choice of dynamics

We model risks to the wild population in terms of competitive interactions, disease, genetic introgression, and increased predation due to physical proximity of aquaculture facilities to salmon habitat. These risks may be modified by application of either or both of two management strategies, prevention of escapes or recapture of escaped fish. Because salmon use an alternative mating strategy in which males may mature as parr and contribute to reproduction, we model male and female populations independently. Values for parr maturation rates and reproductive success are from recent field studies (Hutchings and Jones, 1998; Whalen, *et al*., 2000; Garant, *et al*., 2003).

Wild fish

Competition may occur within the redds (egg competition) and between parr (parr competition). To avoid chaotic dynamics, the competition terms are all of the form $1/(1+\beta N)$ where N is the number of individuals at the appropriate life history stage and β is a measure of the intensity of competition. This density dependent term reduces survival from one life history stage to the next. Habitat destruction has a similar effect on reducing survival. We assume that oceanic survival is density independent. The life cycle of wild fish is shown in Figure 1.1.

Figure 1.1. Life cycle of wild salmon. Freshwater stages are designated FW, sea-going stages are SW, numbers refer to years spent in that stage. Solid lines indicate transitions between stages. The fraction of fish transitioning between stages are given along transition lines, and e_1 and e_2 are the numbers of eggs contributed by one- and two-sea-winter adults, respectively. The dashed line represents the genetic contribution of mature male parr.



The state variables for females are

E(t) = eggs at the start of year t

$R_{ij}(t)$ = fish (parr) that will spend i seasons resident in freshwater who are in the j^{th} year of residence at the start of year t.

$S_i(t)$ = smolts at the start of year t who spent i (=1, 2) seasons in freshwater.

$A_{ij}(t)$ = fish that spend i seasons in the sea who are in the j^{th} year at sea at the start of year t (post smolts or adults).

These numerical values represent the density of fish (Elliott,1994), not absolute numbers (i.e., the model doesn't represent a particular stream system with a specific carrying capacity).

Dynamics for male fish are the same as for females, except at the parr-to-smolt transition, which male parr may delay if they mature.

Before the introduction of aquacultured fish, the dynamics of the wild stock for eggs is:

$$E(t+1) = e_1 h_f(t) A_{11}(t) \exp(-F(t)) + e_2 h_f(t) A_{22}(t) \exp(-F(t)) \qquad (1)$$

where e_j is the egg production per female for fish who spent j years at sea, $h_f(t)$ is the freshwater habitat in year t, and F(t) is fishing mortality in year t. We assume equal production of male and female eggs (Mathisen and Zheng, 1994).

The parr dynamics common to both males and females are

$$
\begin{aligned}
R_{11}(t+1) &= \sigma_0(E(t)/2)h_f(t)f\Phi_e(E(t)) \\
R_{21}(t+1) &= \sigma_0(E(t)/2)h_f(t)(1-f)\Phi_e(E(t)) \\
R_{22}(t+1) &= \sigma_1\Phi_r(R(t))R_{21}(t)h_f(t)
\end{aligned}
\tag{2}
$$

where new parameters are the maximum per capita survival σ_0 and σ_1, the fraction f of fish that are resident in fresh water for one year, and the density dependent competition term for interactions within the redds

$$\Phi_e(E(t)) = \frac{1}{1+\beta_e E(t)} \tag{3}$$

where β_e represents the intensity of egg competition for resources (e.g., oxygen). The competition term for parr, $\Phi_r(R(t))$, decreases the survival of R_{21} parr to R_{22}, and is defined later. Note that half the eggs become female parr, and the other half become males.

In the case of competition between wild and aquaculture eggs, $E_a(t)$, egg competition is

$$\Phi_e(E(t)) = \frac{1}{1 + \beta_e(E(t) + \eta E_a(t))} \tag{4}$$

Here the impact of the aquaculture eggs is modified by η, which ranges between 1 and 1.5 as a user-settable parameter describing the increased effect of competition due to the presence of aquaculture eggs. Setting $\eta=1$ means that the eggs are equivalent. Larger values for η increase the impact of aquaculture eggs on the survival of all eggs. This term captures the idea that aquaculture fish on the spawning grounds may reduce success for all spawners, for example by overlaying redds.

Male parr may mature in November of their first year, and contribute to spawning. They then either smolt the following spring or remain for a second year as mature parr. At the end of this second year survivors become smolts.

$$\begin{aligned}
R_{11m}(t+1) &= \sigma_0(E(t)/2)h_f(t)f\Phi_e(E(t)) \\
R_{mm}(t+1) &= \sigma_1 R_{11m}(t)h_f(t)\ \rho_b(1-\rho_s) \\
R_{21m}(t+1) &= \sigma_0(E(t)/2)h_f(t)(1-f)\Phi_e(E(t)) \\
R_{22m}(t+1) &= \sigma_1\Phi_r(R(t))R_{21m}(t)h_f(t)
\end{aligned} \tag{5}$$

where R_{11m} is the population of male parr that would generally spend one year in freshwater, ρ_b is the rate of their maturation at the end of the first year, and ρ_s is the fraction of those mature parr that smolt the following spring. Those that mature and do not smolt become R_{mm} parr.

The competition term for all parr is

$$\Phi_r(R(t)) = \frac{1}{1+\beta_{11}[R_{11\bullet}(t)+\gamma R_{a\bullet}(t)]+\beta_{12}R_{12\bullet}(t)+\beta_{22}[R_{22\bullet}(t)+R_{mm}+\gamma R_{amm}(t)]} \tag{6}$$

The β_{ij} terms are competition strength for parr of each stage. $R_{ij\bullet}$ represents the summed males and females of type R_{ij}. Aquaculture parr, $R_{a\bullet}$ and R_{amm} (mature males), may have a competitive advantage over natural parr, thus their competition is modified by γ, analogous to η in the egg-competition term, which ranges between 1 and 3. Enhanced competition of aquaculture parr reflects their rapid growth relative to wild parr. If there is no competition between wild and aquaculture parr, the equation is modified by the removal of the $\gamma R_{a\bullet}(t)$ and $\gamma R_{amm}(t)$ terms. Mature male parr ($R_{mm}(t)$ and mature aquaculture males, $R_{amm}(t)$) are second-year parr with respect to competition.

Female smolts are produced from the resident female parr according to

$$\begin{aligned} S_1(t+1) &= h_f(t)\sigma_1 R_{11}(t)\Phi_r(R(t)) \\ S_2(t+1) &= h_f(t)\sigma_2 R_{22}(t)\Phi_r(R(t)) \end{aligned} \tag{7}$$

where the maximum per capita survival for a second year in the stream is σ_2 .

Male smolts are then produced by

$$\begin{aligned} S_{1m}(t+1) &= h_f(t)\sigma_1 \Phi(R(t))[R_{11m}(t) + \rho_b\rho_s R_{11m}(t)] \\ S_{2m}(t+1) &= h_f(t)\sigma_2 \Phi(R(t))[R_{22m}(t) + R_{mm}(t)] \end{aligned} \tag{8}$$

Finally, if g_j represents the fraction of fish that return after j years at sea, the adult dynamics are

$$\begin{aligned} A_{11}(t+1) &= r_p\xi_w d_f\sigma_3\, h_0(t)[g_1S_1(t) + g_2\, S_2(t)] \\ A_{21}(t+1) &= r_p\xi_w d_f\sigma_3\, h_0(t).[(1\text{-}g_1)S_1(t) + (1\text{-}g_2)S_2(t)] \\ A_{22}(t+1) &= r_p\sigma_3\, h_0(t)A_{21}(t) \end{aligned} \tag{9}$$

where r_p is the fraction of the wild adults escaping impacts of the recapture of escaped fish, ξ_w is the fraction of smolts surviving enhanced predation, d_f is the fraction of smolts surviving disease exposure on out-migration (when applicable), σ_3 is maximum per capita survival at sea, and $h_0(t)$ is the habitat at sea at time t. Disease is assumed to be contracted on out-migration through the estuary, and therefore doesn't affect the A_{22} fish.

For the case of enhanced predation, ξ_w is decreased from 1 to 0.8, which represents the fraction of wild smolts surviving the effects of enhanced predation, assumed to occur as fish are attracted to the vicinity of sea-cages by excess feed in the water. Enhanced predation occurs because predators are similarly attracted to the sea-cages This doesn't affect the A_{22} adults, who are at sea.

In the case of recapture of aquaculture adults, the wild fish are reduced by a recapture penalty of 5% ($1-r_p$), which represents losses to the wild population due to handling, or mis-identification of wild fish as aquaculture fish.

The fraction of random matings that involve two wild fish is

$$r = \left[\frac{E(t+1)}{E(t+1) + E_a(t+1)} \right]^2 \qquad (10)$$

where E(t+1) is given by Eqn 1 and $E_a(t+1)$ is an analogous expression for aquaculture fish. We let $\rho_a = .3$ denote the probability of assortative mating. The number of eggs produced from assortative wild-wild crosses is thus $E' = E(t+1)\,[\rho_a + r(1-\rho_a)]$, and for the case of genetic introgression involving adults only, we replace E(t+1) in Eqn 1 by E′.

If mature parr are contributing to introgression, E′ is further modified according to the ratio of aquaculture male parr to the returning wild females, O(t), and the rate of successful parr fertilization of eggs, ρ_f.

$$O(t) = \min\begin{cases} 3 \\ R_{am}(t) \Big/ \left[(A_{11}(t) + A_{22}(t))\exp(-F(t))hf(t)\right] \end{cases} \tag{11}$$

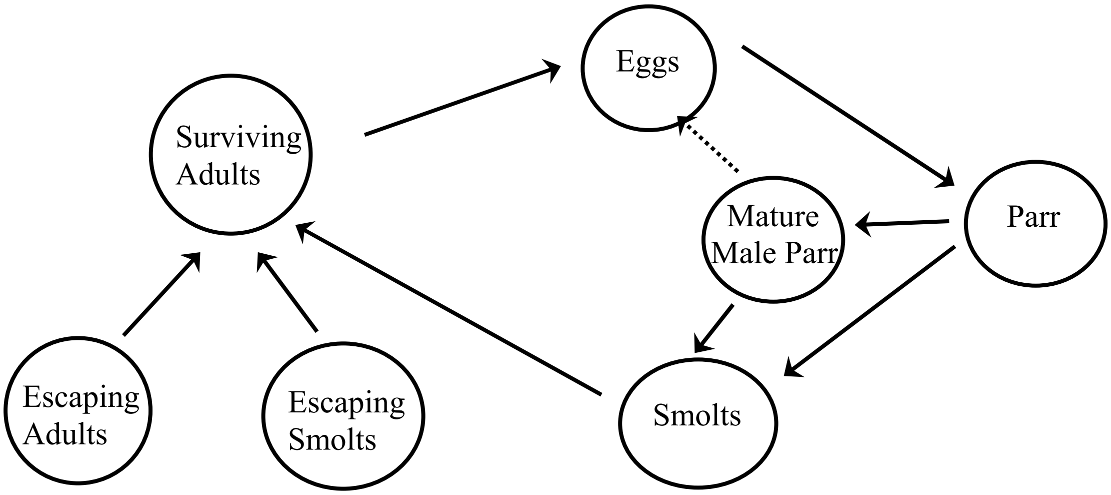
Then $E' = E' - \rho_f\, E'O(t)$, and we replace E(t+1) in Eqn 1 by E′.

Escaped fish

All escaped fish are assumed to be one year smolts and to be grilse. Each year, 20 smolts and 100 grilse escape from farms, leaking out over the course of the year. This number can be modified by exclusion of either smolt or adult escapes, so that the effect of each on the wild population can be examined in isolation. Escaping grilse return to the stream the next year and spawn, while smolts may take several years to mature at sea. Survival, reproduction and return of aquaculture smolts are all calculated in the physiological model, described below.

An optional scenario provides for the catastrophic escape of 5,000 grilse and 1,000 smolts in addition to the usual pattern of constant escapes. In this case, we model three pulses of varying duration: a one-year, two-year and five-year pulse, the latter two simulating yearly-repeating catastrophes. The life-cycle of aquaculture fish is shown in Figure 1.2.

Figure 1.2. Life cycle of aquaculture fish. Freshwater stages are designated FW, sea-going stages are SW, numbers refer to years spent in that stage. Solid lines indicate transitions between stages. The fraction of fish transitioning between stages are given along transition lines, and e_1 and e_2 are the numbers of eggs contributed by one- and two-sea-winter adults, respectively. The dashed line represents the genetic contribution of mature male parr.



Equations for escaped fish are similar to Equations 1-11 for the wild fish, however the escaped fish are assumed to be less-well adapted to local conditions (Fleming, *et al.*, 2000), and are penalized at each stage by a maladaption parameter, ρ_m. In the current model, ρ_m is same for all stages. Eggs are produced according to

$$E_a(t+1) = \rho_m e_1 A_a(t) \tag{12}$$

where ρ_m is a parameter describing the maladaption of aquaculture fish to living in the wild, currently set to 0.3 and applied at each life history stage. In the case of genetic introgression, this is modified so that

$$E_a(t+1) = E_a(t+1)+ \zeta(E(t+1)-E') \tag{13}$$

where E' is as calculated in Eqn. 10 or 11, depending on parr involvement, and ζ is the rate of survival of hybrid eggs, currently set to 0.5. Aquaculture-derived eggs are assumed to produce even sex-ratios in parr.

Dynamics for male and female aquaculture parr are

$$\begin{aligned} R_a(t+1) &= \rho_m\sigma_0(E_a(t)/2)\Phi_e E_a(t)h_f(t) \qquad (14)\\ R_{am}(t+1) &= \rho_m\sigma_0(E_a(t)/2)\Phi_e E_a(t)h_f(t) \end{aligned}$$

where σ_0 is the survival of eggs to parr, and is the same as for wild fish, $h_f(t)$ is the freshwater habitat in year t, and the egg-competition term is

$$\Phi_e(E_a(t)) = \frac{1}{1+\beta_e E_a(t)} \qquad (15)$$

except in the case of wild-aquaculture egg competition, in which case Eqn. 4 applies.

Mature male aquaculture parr arise from the successful reproduction of escaped adults and are those male parr that mature and do not smolt

$$R_{amm}(t) = R_{am}(t)\rho_m h_f(t)\sigma_1\rho_b(1-\rho_s)\ \Phi_r(R_a(t)) \qquad (16)$$

and $\Phi_r(R(t))$, the competition term for aquaculture parr, is described below.

The dynamics for female aquaculture-derived smolts produced in the stream are

$$S_h(t+1) = \rho_m h_f(t)\sigma_1 R_a(t)\Phi_r(R_a(t)) \qquad (17)$$

where σ_1 is the maximum per-capita parr-to-smolt survival rate, the same as for the wild fish.

Male hybrid smolts are the sum of the mature and immature parr

$$S_{hm}(t)= \rho_m h_f(t)\sigma_1[R_{am}(t)(1- \rho_b) + R_{am}(t)\, \rho_b(1-\rho_s) + R_{amm}(t))\Phi_r(R_a(t)] \qquad (18)$$

If there is competition between wild and aquaculture parr, $\Phi_r(R_a)$ is given by Eqn. 6; otherwise, it is

$$\Phi_r\left(R_a\right)=\frac{1}{1+\beta_{11}(R_a(t)+R_{am}(t))+\beta_{22}R_{amm}(t)} \qquad (19)$$

The escaping aquaculture smolts are given by

$$S_a(t+1) = \Xi S_e \qquad (20)$$

which is simply the yearly rate of smolt escape, S_e, modified by the rate of escape reduction, Ξ, which may vary between 0 and 100 percent. They are assumed to be equally divided between males and females.

The aquaculture adults are calculated as

$$A_a(t+1) = \xi_a\rho_m\sigma_3(d_f S_h(t)) + \Xi A_e) \qquad (21)$$

where d_f is the fraction of out-migrating hybrid smolts surviving disease, $S_h(t)$, applied in the case of disease transmission, σ_3 is survival at sea, and ξ_a is the rate of survival of aquaculture–derived smolts from enhanced predation. A_e is the rate of adult escape, which is modified by escape reduction, Ξ. Escaping aquaculture adults are assumed to be equally male and female.

Finally, we calculate the contribution of escaped smolts to the surviving population of aquaculture spawners as

$$A_a(t) = A_a(t)+S_a(t-\upsilon)\Pi \tag{22}$$

where υ is the reproductive lag (or time-to-maturity) in years for escaped smolts from the physiological model, and Π is their expected reproduction (the product of survival and gonadal weight).

These parameters for escaping smolts are calculated at the beginning of the simulation by the physiological model as a series of probability distributions representing different escape dates throughout the year, and values randomly drawn from these distributions govern expected reproduction each year. This is the only source of stochasticity in the model.

1.2.2. Survival and reproductive potential of escaped fish based on physiological parameters (within-year model).

We now turn from consideration of populations in competition to the individual growth and maturation of an escaping smolt. We assume that salmonids function according to developmental switches that control gonadal development (Mangel, 1994a, Mangel, 1994b, Thorpe *et al.*, 1998). In the model, the timing of these switches is based on current knowledge of Atlantic salmon in Scotland; we assume that photoperiod is the external cue that synchronizes maturation. We begin with a review of Thorpe, *et al.* (1998), which summarizes the evidence for this.

To reproduce in November, a fish must initiate physiological changes the previous November at which time an individual responds to a developmental switch that determines the maturation process. This is designated by G1. The response involves comparing a combination of the absolute level of lipid reserves and rate of change of lipid reserves with a genetically determined maturation threshold, designated by M1. The justification for such a threshold is that lipids are required for both somatic function during the year and development of gonads (c.f. Henderson and Wong, 1998, Jonsson and Jonsson, 1998) which takes time.

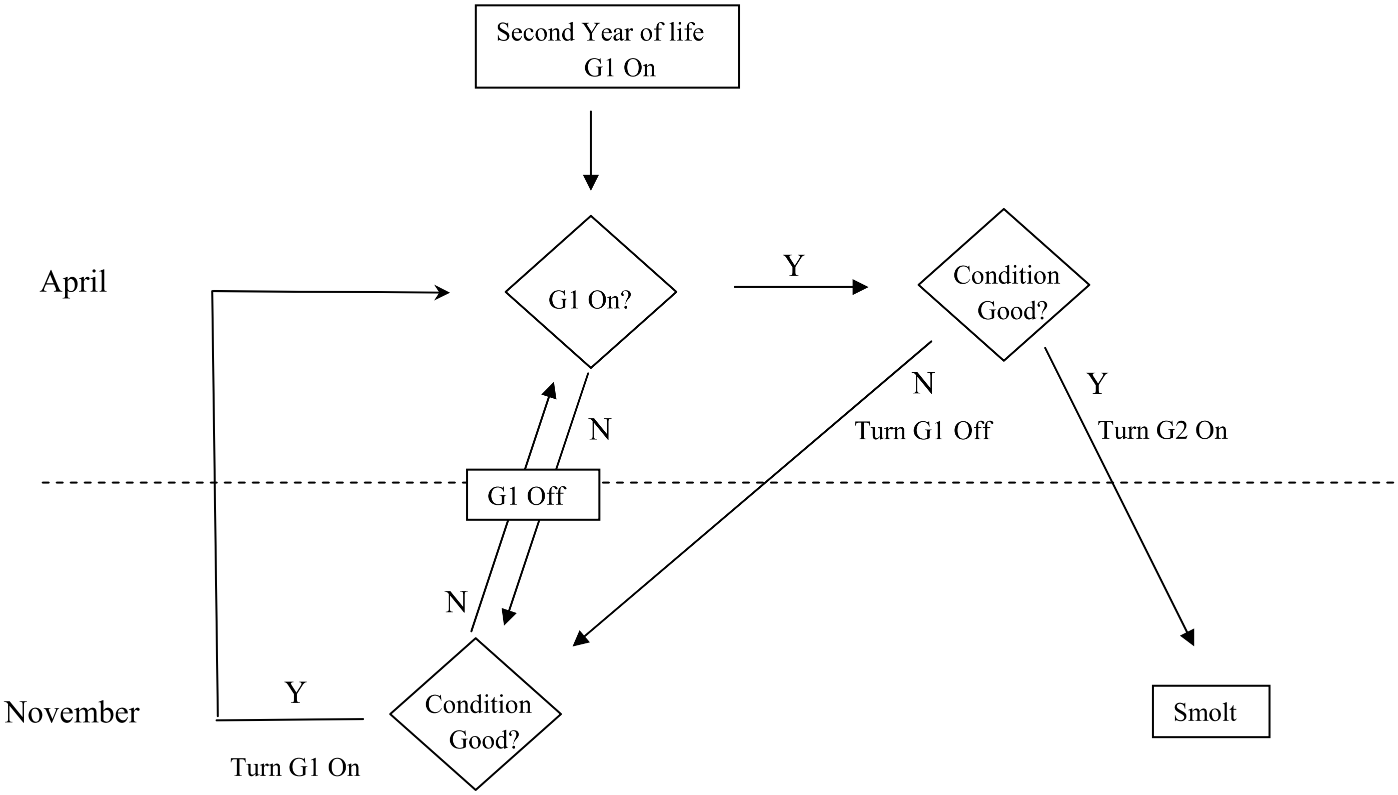
Thus, there is a correlation between the lipid state in the current November and the potential level of reproduction the following November. If the projection based on lipid and rate of change of lipid is less than M1, maturation is inhibited; otherwise gonadal development continues. We assume that the fish assesses current state and rate of change of state and acts on these to the extent that the current values provide information about future ones, using photoperiod as the cue (e.g. Bjornsson *et al*., 1994; Imsland *et al*., 1997; Forsberg, 1995).

In April, a second maturation switch, G2, occurs and a similar comparison is made between the projection of lipids the following November and a second maturation threshold M2. If G1 = 1 when the combination of lipid and rate of change of lipid exceeds the threshold, then a fish that matures in November has followed the path G1 = 1 the previous November and G2 = 1 the previous April. A fish that does not mature could have followed either G1 = 0 (in which case G2 is also 0), or G1 = 1 but G2 = 0 (in which case G1 is reset to 0). The latter case would arise when growth opportunities exceed the threshold M2 associated with G2.

Since it is possible (through photoperiod and temperature manipulations) to produce fish that mature in the first November of their lives, G1 = 1 at the time of fertilization. Then the first developmental switch the fish will encounter is the G2 switch on Julian day 106 (mid-April). The fish monitors its performance between Julian day 85 and Julian day 106, and based on lipid accumulation rate in that interval and total lipid level, the salmon predicts lipid levels for the following November, Julian day 315. If the predicted lipid levels are below the genetically-determined threshold of M2, the switch remains off (G2 = 0), and G1 is reset to zero.

The next switch the fish encounters will be the G1 switch on Julian day 315. Lipid accumulation rate from Julian day 294 – 315, and current lipid levels are used to predict lipid levels for Julian day 471, the following April. If the predicted lipid levels exceed M1, the G1 switch is turned back on (G1 = 1), and the fish proceeds to the G2 switch on Julian day 471. Otherwise, the G1 switch is turned off (G1 = 0), and the fish must wait an entire year before it can re-initiate maturation. If it decides to mature (G2 = 1 on Julian day 106), it will become anorexic on Julian day 197 of that year and begin to lose weight, but not length. The decision tree for the maturation process is illustrated in Fig 1.3.

Fig 1.3. Decision tree for maturation. In the second year of life, parr maturation is determined by physical condition in April, leading to smolting in November for parr in good condition.



Growth Model

The model examines a number of processes in detail, but in the end we are interested only in the survival-to-reproduction of the escaping fish, and the timing of their return to freshwater to spawn. These follow from their physical condition at two points within the year, the spring and fall checkpoints. The following equations govern the genetically-programmed growth pattern of the cultured salmon (weight W, length L, and lipid levels Λ), and determine M1 and M2.

While immature:

$$W(t) = \left[W(t-1)^{1/3} + \frac{1}{3} q a(t) \right]^3 \tag{23}$$

$$\log(L(t)) = 1.613 + 0.312 \log(W(t)) \tag{24}$$

$$\Lambda(t) = -0.82 + 0.00418 W(t) + 0.000199 L(t) \tag{25}$$

Where q is food finding and processing ability, and a(t) is food assimilation as a function of temperature, τ, governed by the equation:

$$a(t) = 1 - \left[\frac{\tau(t) - 6}{12} \right]^2 \tag{26}$$

Temperature oscillates between 8.5° C and 12.5° C, with the peak occurring on Julian day 210.

$$\tau(t) = 10.5 + 2\cos\left(\frac{2\pi(t-210)}{365} \right) \tag{27}$$

The clock starts on Julian day 84 in the second year of life, which is the beginning of the window for evaluation of the first G2 switch. Initial weight of the fish is 47 grams, and q is set to 0.08. From these values, we can calculate the lipid targets for M1 (Λ(471)) and M2 (Λ (680)).

Once salmon begin to mature, (Thorpe, *et al.*, 1998)

$$W(t) = \left[W(t-1)^{1/3} + \frac{1}{3} q c_r a(t) \right]^3 \tag{28}$$

$$\log(L(t)) = 1.556 + 0.323 \log(W(t)) \tag{29}$$

$$\Lambda(t) = -0.82 + 0.00418 W(t) + 0.000199 L(t) \tag{30}$$

where c_r is the cost of reproduction, set to 0.988 (Mangel, 1994a). On Julian day 562, under optimal conditions, the salmon will enter the pre-reproduction anorexic period, at which point it will begin losing weight (but not length), and the weight equation changes to:

$$W(t) = W(t-1)(1 - c_a) \tag{31}$$

where c_a, the cost of anorexia, is 0.001.

The mass of gonads, Γ, at the time of reproduction (t = 680) is calculated as:

$$\Gamma(680) = -0.326 + 0.194 W(680) \tag{32}$$

The expected reproduction equals gonad mass divided by the mean weight of an egg, 0.1145 g (Jonsson, *et al.*, 1996), times expected survival. Expected survival, Ω, from the day of escape, t_e, to day 680 under optimal conditions is

$$\Omega(680) = \exp\left[-\sum_{s=t_e}^{680} \mu_0 + \mu_1 W(s)^{-0.37}\right] \tag{33}$$

where μ_0 is the size-independent and μ_1 the size-dependent component of mortality.

Stochastic Environment

To account for the fact that both the environment and the performance of individual fish are variable, we introduce stochasticity into the model by treating weight and q as variable, and modify the optimal-case equations as follows.

In the weight-gain equations, we separate q into an individual component, q_i and an environmental component that can vary over time, $q_e(t)$, as per Thorpe, *et al.* (1998). This results in equations for immature and maturing fish

$$W(t) = \left[W(t-1)^{1/3} + \frac{1}{3} q_i q_e(t) a(t)\right]^3 \tag{34}$$

$$W(t) = \left[W(t-1)^{1/3} + \frac{1}{3} q_i q_e(t) c_r a(t)\right]^3 \tag{35}$$

When a fish escapes, the environmental component q_e changes. In particular, $q_e = 1$ while in the culture environment, and $q_e = 0.5$ after the escape. During anorexia, q_e does not apply, since the fish is not attempting to feed; however we assume those with higher q_i lose weight at a faster rate, so that the weight equation during anorexia is

$$W(t) = W(t-1)\left[1 - c_a\left(\frac{q_i}{q}\right)\right] \tag{36}$$

Finally, we assume that q_i also affects survival. Higher q_i will dictate lower survival due to gonadal steroids interacting with the immune system (Mangel, 1994), and behaviors such as increased risk-taking to acquire food. The new survival equation is

$$\Omega(680) = \exp\left[-\sum_{s=t_e}^{680}\left(\mu_0 + \mu_1 W(s)^{-0.37}\right)\left(1 + c_g\left(\frac{q_i - q}{q}\right)\right)\left(1 + c_w\left(\frac{q_i - q_w}{q_w}\right)\right)\right] \tag{37}$$

The growth penalty c_g applies throughout the life of the fish, while the wild penalty c_w applies after the escape. Optimal q for life in the wild, q_w, is different than the optimal value in the aquaculture environment. We set $c_g = 0.8$, $c_w = 0.8$, and $q_w = 0.9q$.

Decision points

With a given initial weight and a given q_i, we can predict an individual's maturation date and expected reproductive success. In the 21-day evaluation period before a trigger date T (day 106 for G2 and day 315 for G1), we measure performance as

$$\kappa_i = \left(\frac{1}{21}\right) \sum_{s=(T-21)}^{T} \frac{\psi_i(s)}{\psi(s)} \tag{38}$$

where $\psi_i(s)$ and $\psi(s)$ are the actual and genetically-expected daily specific growth rates, calculated as

$$\psi(s) = W(s-1)^{-2/3}(W(s) - W(s-1)) \tag{39}$$

W(s) is either the genetically-expected or the actual weight, depending on whether ψ (s) or $\psi_i(s)$ is being calculated. On the trigger day T, at the end of the assessment window, weight, length and lipids are predicted for the next trigger date (either 209 days later for G2 or 156 days later for G1). Predicted weight W_p is a function of performance for both immature (Eqn 18) and maturing fish.

$$W_p(t) = \left[W_p(t-1)^{1/3} + \kappa_i \frac{1}{3} qa(t) \right]^3 \tag{40}$$

$$W_p(t) = \left[W_p(t-1)^{1/3} + \frac{1}{3} q c_r a(t) \right]^3 \tag{41}$$

The predicted weight for the first day of the projection period is the actual weight of the fish on that day. For prediction, we use the genetically-encoded q rather than the individual-specific q_i.

From the predicted weight, we calculate predicted lengths and lipids for the trigger dates according to the appropriate equations for the immature/maturing fish. If predicted lipids are greater than the target value M2 for the G2 trigger date, then the *in silico* fish matures, otherwise both G2 and G1 are zero. If predicted lipids are greater than the target value M1 for the G1 trigger date, then G1 = 1, and the fish proceeds until it approaches the G2 trigger day again.

For a given inital weight, q_i, and day of escape, we calculate the day on which the individual will reproduce, as well as its expected reproductive success and whether it survives to reproduction. We assume that initial weights are drawn from a normal distribution, $N(47,2.66)$, and q_i is drawn from $N(0.08,0.09)$ that is truncated at zero and re-normalized. Survival and reproduction calculated from these equations are used to parameterize the age-structured model for escaped smolts, and the day of reproduction is used to determine the size of the lag between escape and reproduction. Parameters for the physiological model are in Appendix A, Table A.3.

1.2.3. Disease

The wild fish are attracted to the scent of food emanating from the farm, and will feed there. While near the farm, they are exposed to waterborne diseases (bacterial or viral) if the farmed fish are diseased (McVicar, 1997). Mortality from disease will depend on the amount of exposure each fish receives. This exposure is modeled as a function of the gradient of food and disease particles from the aquaculture facility, and the distribution of fish along the food gradient as it declines with distance from the farm. We begin by modeling the diffusion of food.

Food concentration, I, decays exponentially with distance x from the source (the farm), and is modeled according to

$$\frac{\partial I}{\partial t} = \frac{D_I}{2} \frac{\partial^2 I}{\partial x^2} - m_I I \tag{42}$$

Where D_I is the diffusion constant for food particles, and m_I is the rate of removal.

The boundary conditions here are that $I(0,t) = I_0$, the initial concentration at the farm, and in the limit as $x \rightarrow \infty$, $I(x,t) = 0$. We are interested in the steady-state concentration gradient, so we set $\partial I/\partial t = 0$.

The concentration of food at a distance x from the farm is then

$$I(x) = I_0 \exp\left(-x\sqrt{2m_I / D_I}\right) \qquad (43)$$

The concentration of disease particles, bacteria or viruses, can also be modeled using the diffusion equation. By analogy,

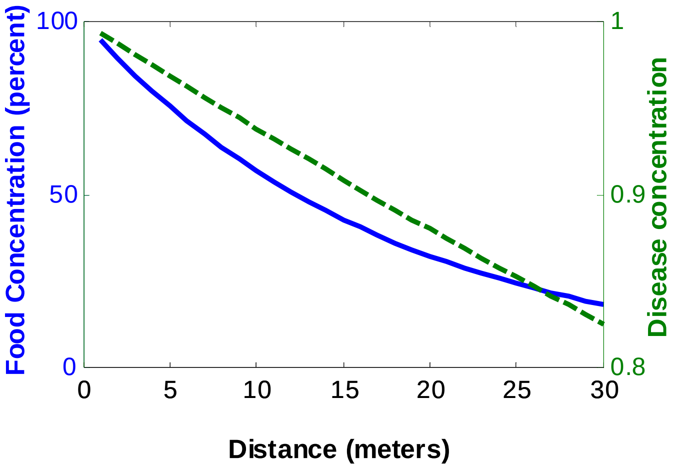
$$B(x) = B_0 \exp\left(-x\sqrt{2m_B / D_B}\right) \qquad (44)$$

where B is the concentration of disease particles, B_0 their initial concentration, m_B is the rate of removal (sinking and cell death), and D_B is their diffusion coefficient.

The rate of sinking of food pellets or bacteria and their diffusion coefficients are calculated using Stokes' Law (Elberizon and Kelly, 1998; Mann and Lazier, 1996). We assume that food particles may be whole pellets or pieces, and have therefore chosen size characteristics that allow for diffusion for some distance away from the aquaculture facility.

Disease particles, being lighter, diffuse further and faster than food particles, and have a longer effective lifetime (the viability of a typical salmon bacterium, *Aeromonas salmonicida* in water is 10 days (Ciprio and Bullock, 2001)), so they remain at higher concentrations throughout the water column. Figure 1.4 shows food and disease gradients conveniently scaled for illustration.

Figure 1.4. Food and Disease gradients. The dashed line is the food concentration; the solid line is the concentration of disease particles. Gradients are scaled for illustration purposes.

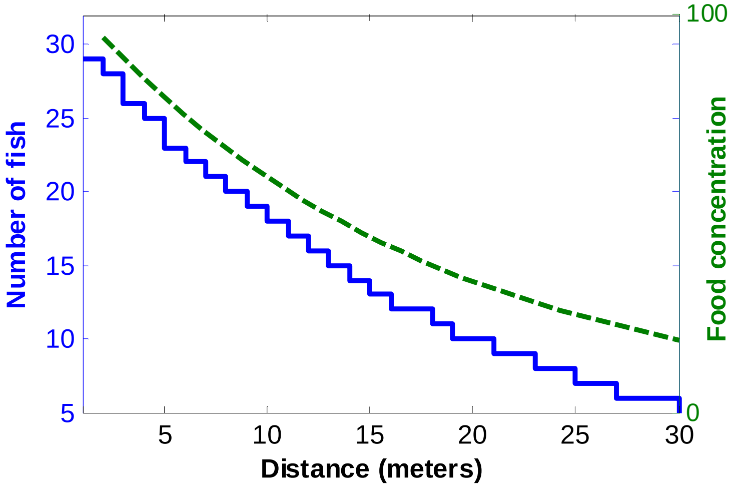


Wild salmon will distribute themselves along the diffusion gradient of the food particles so that at an equilibrium state each animal has access to equal amounts of the food. When the food per fish ratio falls below a threshold such that the food is no longer sufficient to meet their need, any salmon not already sharing the food will leave the area, escaping exposure to disease. We assume that when satiated, fish leave the aquaculture facility and follow their conspecifics out to sea. Those fish that have been exposed to a sufficient concentration of disease particles soon sicken and die.

This fish distribution is modeled so that as each salmon approaches the farm, the food/fish ratio is calculated along the food gradient from the number of fish already at each point with the addition of the new animal. The site with the maximum ratio is chosen, and the new fish joins those already there. This continues until all fish are distributed along the food gradient, or until the maximum food/fish ratio is below the threshold. Figure 1.5 is an example, again scaled for illustration. In MATLAB, the algorithm is

```
fish_distribution = ones(1:1000);
while (fish_arrive)
    i = location(max(food/fish_distribution))
    if (max(food/fish_distribution) > food_threshold)
        fish_distribution(i) = fish_distribution(i) + 1;
    else
        break;
    endif
end
fish_distribution = fish_distribution – 1
```

Figure 1.5. The distribution of food (dashed) and of fish (solid). Because fish occur as discreet individuals, their distribution is a step-function.



We assume that the probability of death from the disease is a function of the concentration of disease particles an animal experiences. Below some threshold concentration, B_t, exposed animals escape disease mortality. Above that threshold, animals contract the disease and die with probability P(B). One model for the probability of transmission is an asymptotic function

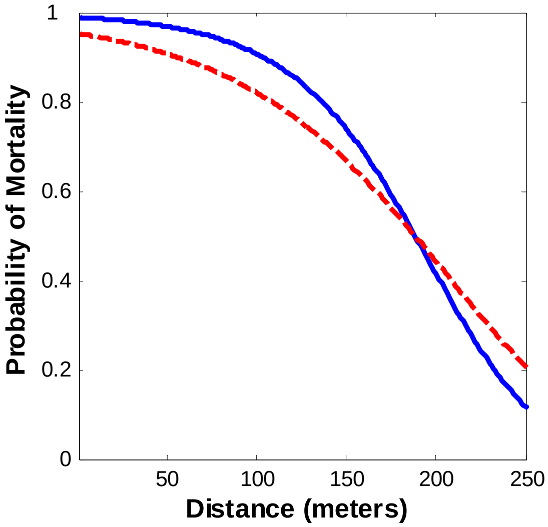
$$P(B) = \frac{(B(x) - B_t)^{\gamma}}{c^{\gamma} + (B(x) - B_t)^{\gamma}} \tag{45}$$

When the exponent $\gamma < 2$, the probability of mortality rises sharply when the concentration of disease particles exceeds a threshold value, B_t before reaching an asymptote at 1. This function is equivalent to a Holling Type II functional response

(MacCallum, *et al*., 2001), or Michaelis-Menten enzyme kinetics. We call this "Type II" disease for the purposes of this model.

At $\gamma > 2$, Eqn. 45 predicts a slow increase in mortality near the disease threshold, and a maximum rate of mortality at high concentrations, a sigmoidal function similar to a Holling Type III functional response, here called "Type III" transmission. In either case, no transmission occurs at bacterial concentrations below the threshold $B_{t.}$ In Figure 1.6, we show the values of these functions with distance from the farm, scaled for illustration.

Figure 1.6. Holling Type II (dashed) and Type III (solid) functional responses. Probabilities are scaled for convenient comparison.



The number of salmon dying (Z) is then the sum of the product of the probability function and the salmon distribution gradient, J_x (Figure 1.5).

$$Z = \sum_x P_x J_x \tag{46}$$

Parameters used in the disease model are given in Appendix A, Table A.5.

Interactions

The ecological scenarios available fall into two categories. Scenarios in which no interactions occur between the wild and farmed fish can be explored with or without aquaculture escapes occurring. Interaction scenarios, such as Egg or Parr Competition or Genetic Introgression, can occur in any combination, or all together. Catastrophic escapes cannot occur without at least one other interaction defined.

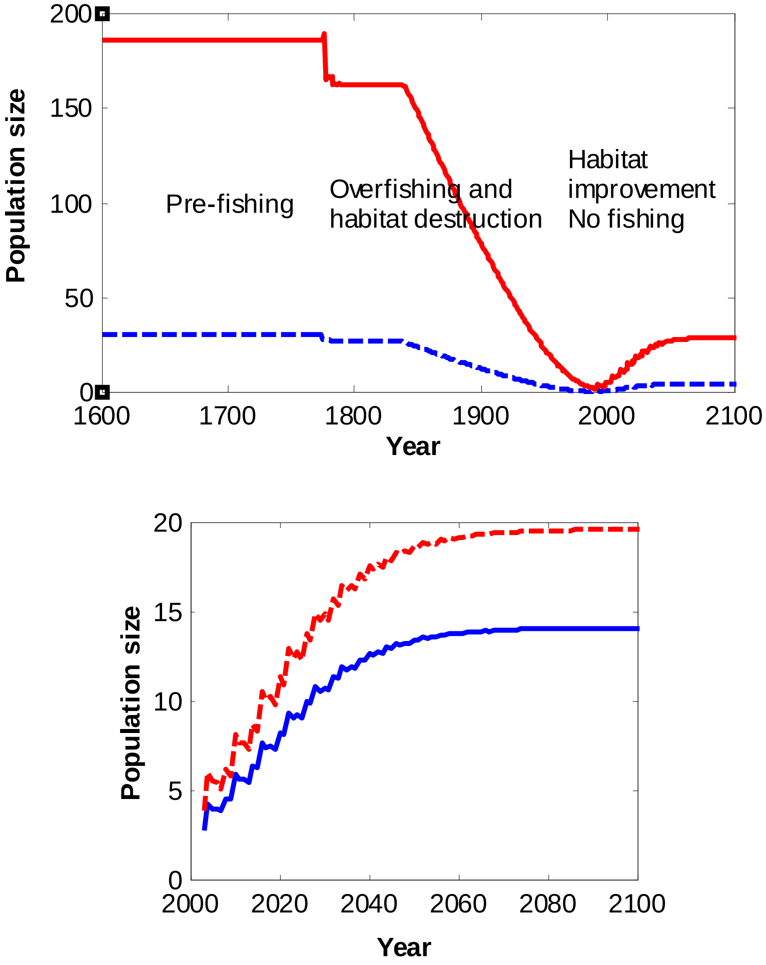
Escapes: The numbers of yearly escapees, 20, 40, 60 or 80. This is the total number of smolts and adults, escaping in equal numbers.

No Aquaculture: The output of the model is the trajectory of wild stock in the absence of aquaculture. In this case, no management actions are considered, and the output of the model is a plot of the populations of adult salmon and grilse from the year 1600 to the year 2100. The plot shows the draw-down of the population from fishing and habitat loss, followed by recovery projected to occur between 2000 and 2100 (Figure 1.8).

No Interactions: The output of the model is the trajectory of wild stock and aquaculture stock given that escapes are occurring, but no specific interactions apply. However, escaped fish survive and reproduce, and their existence dilutes the fraction of the population that are wild salmon.

Egg Competition: The output of the model is the trajectory of wild stock and aquaculture stock given that egg competition is occurring in the redds, leading to fewer hatchings of wild fish. This parameter can be set to any of {none, low, medium and high} values, and can be set whether or not the populations are interacting.

Figure 1.8. Model results for wild salmon in the absence of aquaculture, upper panel. The solid line is the population trajectory of 2-sea-winter adults returning to spawn; the dashed line is the population trajectory of adults returning after 1-sea-winter. Bottom panel: differential returns of female (dashed) and male (solid) fish during the population recovery period.



Parr Competition: The output of the model is the trajectory of wild stock and aquaculture stock given that parr compete for resources in the streams, affecting parr survival. This parameter can be set to any of {none, low, medium and high} values, and can be set whether or not the populations are interacting.

Enhanced Predation: The output of the model is the trajectory of wild stock and aquaculture stock given that there is predator attraction to the mouths of rivers due to aquaculture and thus enhanced predation on both wild and escaped smolts.

Genetic Introgression: The output of the model is the trajectory of wild stock and aquaculture stock given that there is genetic introgression caused by mating between wild and aquaculture fish. The offspring of crosses between wild and aquaculture fish are tracked as aquaculture fish. Only the offspring of two wild parents are considered wild fish. Choices for introgression include no introgression, hybridization between adults, and the additional contribution of mature male parr to introgression.

Disease: The output of the model is the trajectory of wild stock and aquaculture stock given that out-migrating smolts are attracted to aquaculture facilities because of abundant food concentrations in the water, and they contract a disease from proximity to the penned fish, which kills them. Choices include no disease, Type II or Type III disease.

Catastrophic Escapes: This option is only available if an interaction scenario has been chosen. Aquaculture escapes occur as usual in most years. In year 2030, instead of the regular numbers, 500 adults and 500 smolts escape. This large escape is repeated in 2060 and 2061, and then again in each year between 2080 and 2085.

Management Options

Adult recapture: The model predicts the trajectory of wild stock and aquaculture stock under the ecological scenario defined by selected interactions. The rate of recapture may be chosen to be between 10% and 100%. Recapture causes an additional mortality on wild stocks proportional to the rate of recapture, up to 5% for 100% recapture. Note that because recapture is of returning adults, but both smolts and adults escape, that even with 100% recapture there will still be aquaculture fish at the end of the simulation, representing those smolts that have escaped and not yet returned to be captured.

Reduced escapes: The model predicts the trajectory of wild stock and aquaculture stock under the ecological scenario defined by selected interactions. The amount of escape reduction may be selected to be between 10% and 100%.

Results options

The third menu is for selection of output options, shown in Figure 1.7. The three options are:

- Text Only

  A text description of the scenario and outcome are displayed in the Matlab command window

- Plot post-management results

  A text description of the scenario and outcome are displayed in a results window, and trajectories are plotted for the populations under managed conditions. A third window shows the trajectories of wild male and female fish.

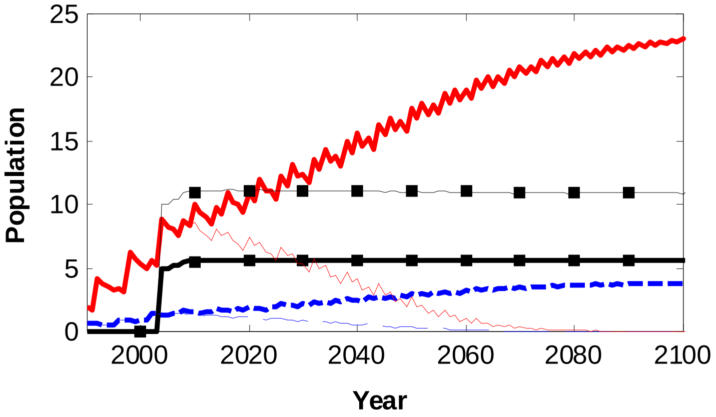
- Plot All results

  A text description of the scenario and outcome are displayed, and trajectories are plotted for the pre- and post-management results.

The results options are only offered when the first simulation is run.

Once a scenario, management and results options have been chosen, the simulation is run and results are displayed as a set of panels, one with a text description of the scenario and results, another a plot of the population trajectories. A third result panel displays the male and female trajectories individually. In Figure 1.9, we show the resultant display from a managed scenario, with all results plotted. The third panel shows the differential survival of males and females, an artifact of the additional year in freshwater experienced by many of the mature male parr.

Figure 1.9. Example results from a managed scenario. The left panel shows the populations of wild grilse (dashed), 2-year adults (solid) and aquaculture adults (squares) returning to spawn. Returns under unmanaged conditions are thin lines; returns under managed conditions are heavy lines. The bottom panel shows the numerical results for these conditions.



```
Simulation conditions:

    Escape rate (smolts : adults)     10 : 10
    Genetic introgression:     Adults Interbreeding

Managment actions:

    Escape reduction                  50 %

Fish stocks in year 2100 (no management):

    Wild                Farmed

    0            10.8667


Wild population extinct in year  2062

Fish stocks in year 2100 (with management):

    Wild                Farmed

    26.8553            5.60323
```

1.3. Results and Discussion

The model results show that the final population of aquaculture salmon in the stream is closely tied to the number of adults escaping. Escaping smolts and reproduction of aquaculture-derived fish in the wild contribute only a small percentage of their final numbers. Table 2 shows the populations of the wild and aquaculture adult spawners in year 2100 under each individual scenario, given that 20 adults and 20 smolts escape each year, first with no management, then with 50% escape reduction.

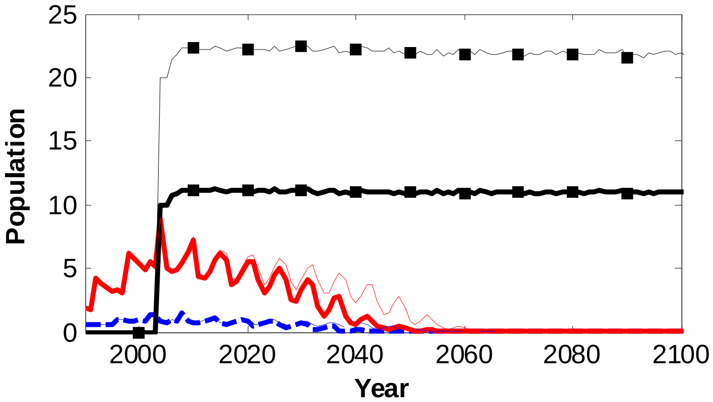
For the parameters used here, genetic introgression (Fig. 1.9) and disease (Fig. 1.10) have the most dramatic affects on the wild population. In the case of genetic introgression involving mature parr, the wild fish persist for only twenty years after escapes begin, and reducing escapes by half only increases their persistence by another three years. Even without the influence of mature parr, introgression decimates the wild population after 30 years. Reducing escapes by 50% increases this to 50 years, half of the management window.

Disease transmission here is characterized by different dynamics, but the same parameters. In Fig. 1.10 and 1.11, we can see that the sigmoidal dynamics of the "Type III" illness has a much greater impact than the "Type II" hyperbolic function, driving the population to extinction in the one case while permitting persistence in the other. Interestingly, the reduction in escapes has the non-intuitive effect of decreasing survival of the wild fish. This is due to the inverse density-dependence of the disease model, which dictates that as numbers decline, a greater percentage of the fish are able to feed near the farm, and become sick. The aquaculture smolts produced in the wild participate in this, so the proportion of wild fish diseased is lowered in their presence.

Table 1.2. Populations of adult spawners returning to freshwater in year 2100.

	Unmanaged			**50% Reduction in Escapes**		
	Wild	**Farmed**	**Year Extinct**	**Wild**	**Farmed**	**Year Extinct**
Genetic Introgression (Adults only)	0	22.5	2035	0	11.3	2056
Genetic Introgression (Adults and Mature Parr)	0	22.5	2023	0	11.3	2026
Egg Competition (medium)	27.8	22.3	—	30.8	11.2	—
Parr Competition (medium)	31.5	22.5	—	32.5	11.3	—
Enhanced Predation	16.5	14.0	—	16.5	7.0	—
Type II Disease	24.4	22.4	—	23.9	11.2	—
Type III Disease	0	22.1	2053	0	11.8	2042
No Interactions	33.7	22.6	—	33.7	11.3	—
No Aquaculture	33.7	0	—	—	—	—

Figure 1.10. Population trajectories under the "Type III" Disease scenario. Populations of wild grilse (dashed), 2-year adults (solid) and aquaculture adults (squares) returning to spawn. Returns under unmanaged conditions are thin lines; returns under managed conditions (50% escape reduction) are heavy lines.



Enhanced predation, which is independent of the number of escapes, also takes a heavy toll, not only on the wild population but on the aquaculture smolts as well.

Egg competition, (Fig. 1.11) is the result of non-hybridizing adult interference in the spawning process – such as redd destruction – and is also quite harmful to the wild population, much more so than is Parr competition for resources in the stream. However, either of these scenarios produces an eventual equilibrium between the two populations.

Figure 1.11. Population trajectories with medium Egg Competition. Populations of wild grilse (dashed), 2-year adults (solid) and aquaculture adults (squares) returning to spawn. Returns under unmanaged conditions are thin lines; returns under managed conditions (50% escape reduction) are heavy lines.

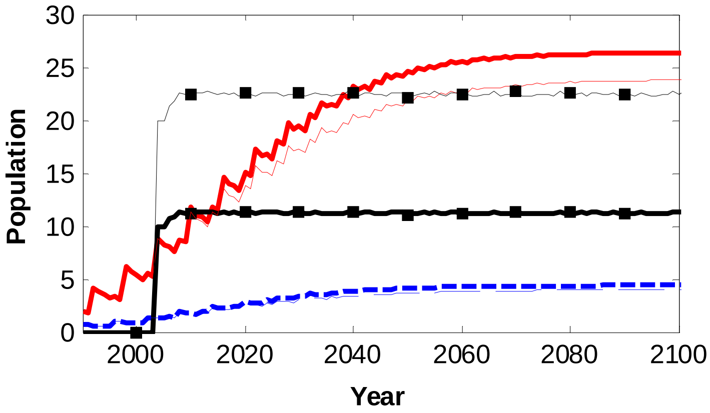
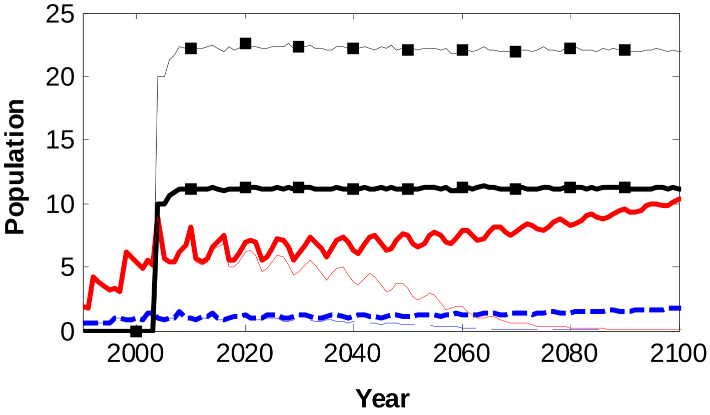


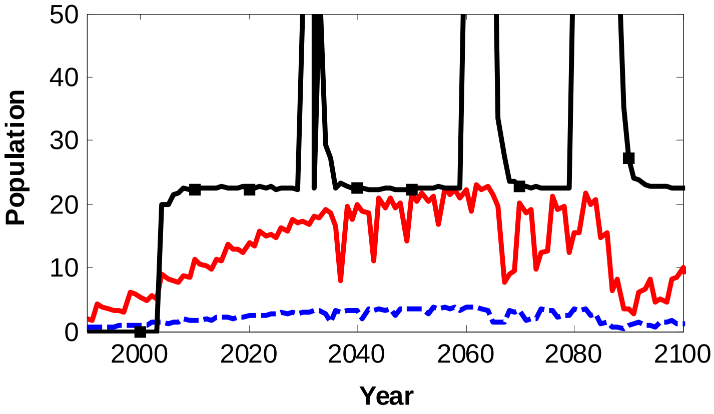
Figure 1.12. "Type II" disease and Egg competition (medium). Populations of wild grilse (blue), 2-year adults (red) and aquaculture adults (black) returning to spawn. Returns under unmanaged conditions are dashed lines; returns under managed conditions (50% escape reduction) are solid lines.



Combining scenarios can also drive the wild fish to extinction. For example, the result of combining "Type II" disease with medium levels of egg competition is shown in Figure 1.12. Independently, these scenarios result in population equilibria. Extinction here occurs around year 2068. With lowered competition, the wild population persists, and eventually begins to recover. The interplay between escape numbers and interaction effects is critical.

Finally, the advantage of a staged life history is shown in the resilience of the wild population in the face of catastrophic escapes (Figure 1.13). Because only the freshwater stages are impacted, the adult salmon at sea are able to replenish the stream population after a crisis.

Figure 1.13. Medium Egg competition with catastrophic escapes of 500 adults and 100 smolts in years 2030, 2060 and 2061, and 2080-2085. Populations of wild grilse (dashed), 2-year adults (solid) and aquaculture adults (squares) returning to spawn.



1.4. Conclusions

This model uniquely brings together four very different concepts: the life-history variation of the physiological model, age-structured vulnerability, a model of disease transmission, and an interface for investigation of different ecological scenarios. The stochastic physiological model expresses our uncertainty about the success of the aquaculture fish. The age-structured model allows us to investigate the differences in vulnerability of wild fish at different ages, and in different environments. The disease model demonstrates the importance of little-understood dynamics in marine systems (McCallum, *et al*., 2004).

Genetic introgression, egg- and parr-competition all represent threats occurring in freshwater, as does habitat loss, which limits the maximum size of wild populations (Fig. 1.8). Of these, genetic introgression provides the most serious threat (Table 1.2). The wild population declines in an exponential fashion (Fig. 1.9), leading to extinction of the wild population within decades. Extinction occurs even more rapidly when the contribution of aquaculture-derived smolts to spawning is considered. This risk to wild salmon is best addressed by strategies such as the production of sterile aquaculture fish (c.f. Cotter, *et al.*, 2000), so that escaped animals are unable to interbreed with the wild stock.

In-stream competition between juvenile fish for resources ("Parr competition") and competition between adults for nest sites ("Egg competition") restricts the size of wild populations, but there is no risk of extinction: in the absence of other effects the populations achieve a stable equilibrium (Fig. 1.12). Since juvenile fish are produced by escaped adults, parr competition is also addressed by the culturing of sterile fish. The intensity of egg competition depends on two factors: the tendency of escaped fish to spawn, and the number escaping. This is best managed by tightening controls at the aquaculture facility, limiting the potential for escapes.

Enhanced predation, disease, and recapture strategies affect post-smolts and adults, ocean-resident fish. Enhanced predation results from the existence of aquaculture facilities in proximity to salmon streams, and can be addressed only by siting these facilities at a distance suitable to limit the exposure of wild fish. Empirical studies on migration patterns should be used to identify sensitive habitat, and contribute to policy on licensing for aquaculture facilities.

The risk of disease transmission from these facilities remains ambiguous. We have modeled two different disease functions, and they present very different results. Type II disease transmission leads to eventual population equilibrium; Type III disease transmission to extinction. In the face of such uncertainty, we can only suggest a precautionary approach to management. Infected facilities should be aggressively treated or be fallowed to prevent transmission to wild populations. Since many common marine diseases are not species-specific, this approach will benefit not only salmon but other species as well.

The model demonstrates that salmon are not more vulnerable in freshwater than in the ocean; it shows instead that risks of varying degrees occur in both environments. If we are to maintain wild salmon populations, management must pursue a two-pronged approach, keeping escaped aquaculture fish out of the freshwater environment, and keeping aquaculture facilities away from those estuaries still frequented by wild salmon.

This model was parameterized as much as was possible from the literature; however we have had to guess at parameters involving competition, disease transmission, and the probability of assortative mating. These are clearly important factors in the population dynamics, yet they are not well studied. For example, the least concentration of A. *salmonicida* sufficient to cause furunculosis is unknown, as is its LD_{50}, the concentration at which 50% of the exposed fish die (Cipriano R., USGS, *pers. comm.*). Such information would be helpful not only in improving the model, but could inform monitoring at aquaculture facilities.

Since realistic fishery models depend on numerous arbitrary choices (Schnute, 2003), models such as this one should serve as tools for thought, and contribute to a frank discussion of the options (Schnute and Richards, 2001). There is an increasing trend towards developing cooperative practices in management by working with stakeholders – the North Atlantic Salmon Conservation Organization (NASCO) is an example of one such (international) effort. It is also increasingly important that we provide full disclosure about scientific uncertainty and unknowns (Stephenson and Lane, 1995), which is rarely fully understood or appropriately used in policy discussions (Rosenberg, 2003). The interface we present here is an attempt to permit non-technical parties a window into analysis of unknowns.

Lackey (2003) and others argue that policymakers and stakeholders need to be informed about the assumptions of the model. Our interface allows the user to make his own assumptions about likely scenarios, and then investigate outcomes over a range of inputs. Precautionary approaches to managmement require an assessment of risk, and this model is an effort to make risk assessment an open process through much-needed user-friendly software (Harwood and Stokes, 2003). We hope to see this approach widely adopted.

Acknowledgements

This work was funded by the National Marine Fisheries Service Northeast Fisheries Science Center. Co-authors for this work are Andrew Rosenberg, Andrew B.Cooper, and Marc Mangel, all of whom contributed substantially to my understanding of marine systems.

A Multispecies Approach to Subsetting Logbook Data for Purposes of Estimating CPUE

Abstract

An initial step in catch and effort analysis is determination of what subset of the data is relevant to the analysis. We propose an objective approach to subsetting trip records of catch and effort data when fishing locations are unknown; the species composition taken on a fishing trip is used to infer if that trip's fishing effort occurred in a habitat where the species of interest (the target species) is likely to occur. We use a logistic regression of multispecies presence-absence information to predict the probability that the target species would be present. A critical value of probability that best predicts target species presence and absence in the data set forms an objective basis for subsetting the trip records. We test this approach by applying it to a data set where individual fishing locations are known, and we show that the method is an effective substitute for information on individual fishing locations.

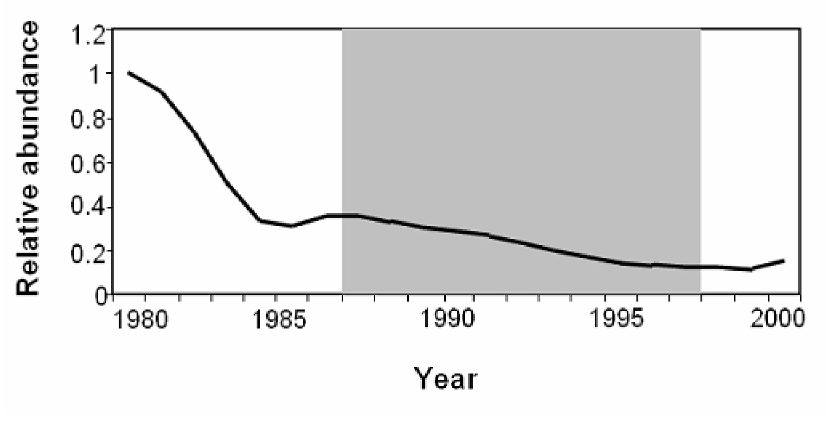
2.1. Introduction

An initial step when analyzing large data sets often involves separating the data into the subset of observations that is considered to be relevant and informative, which is retained for analysis, and the subset of observations that is considered to be uninformative, which is discarded. We refer to this process as ‘subsetting’ the data. In practice, subsetting is often based on *ad hoc* and subjective decision rules, and introduces a source of uncertainty into the analysis that is seldom evaluated. We propose an objective decision rule for subsetting catch and effort data based on the species composition of catches taken on individual fishing trips. Unlike an *ad hoc* decision rule, calculations based on this decision rule are reproducible by independent analysts and the results are amenable to statistical analysis, including the estimation of precision.

Fishery data in the form of landings receipts, logbooks, or catches sampled directly in the field often reflect a variety of alternative species or habitats targeted by the fishermen, even within a single fishing trip. Consequently, some of the records in a data set may not be relevant to calculating catch-per–unit-effort (CPUE) for a particular species (referred to here as the target species). For example, the Marine Recreational Fishery Statistics Survey (MRFSS) (Osborn et al., 2002) provides records of species catch and angler effort since 1980 for recreational fishermen on the west coast of the United States. If these records are to be used as the basis of a CPUE index of abundance for a particular target species, one of the first steps in the analysis is to distinguish which of the catch and effort records are informative for that species and which are not.

Bocaccio (*Sebastes paucispinis*) forms a focus for this study. Boccacio is a mobile species with weak site-fidelity until late maturity, although it is found in close association with similar rockfish species along rocky bottoms (Love et al., 2002). Historic abundance has been estimated by MacCall (2003) based on a number of different abundance indices (Fig. 2.1). The abundance of bocaccio declined severely after the early 1970's, and a current management goal is to rebuild the stock (MacCall, 2003).

Fig. 2.1. Relative abundance of bocaccio over the period covered by the MRFSS and CDF&G surveys (MacCall, 2003). The shaded region denotes the period covered by the CDF&G survey.



A CPUE index of abundance is potentially valuable for assessing boccacio. However, fishing trips that targeted tuna or salmon are unlikely to provide information on the abundance of a groundfish species such as bocaccio, and fishing trips that encountered these pelagic species should clearly be deleted when subsetting a data set such as MRFSS. However, even with this improvement, the data remaining may contain an unknown proportion of fishing trips that did not sample bocaccio habitat, and that proportion may vary substantially from year to year, contributing to imprecision or spurious trends in a CPUE index of bocaccio abundance. Choices of where to fish may be influenced by, for example, environmental conditions, expected catch rates, or changes in fishing regulations. The latter two influences are likely to exhibit long-term changes over time.

If fishing locations were included in the records, it would be possible to restrict the analysis to catch and effort data for only those locations known to be bocaccio habitat. However, information on fishing location may not be available. For example, the MRFSS data were usually collected dockside at the end of the fishing trip, and do not indicate where the actual fishing occurred, nor how many locations were fished. In this paper, we examine an approach to 'subsetting' that uses the species composition from fishing trips to infer whether the fishing occurred in habitat appropriate for use in CPUE calculations.

2.2. Materials and Methods

2.2.1. Data

Partyboats (a.k.a. commercial passenger fishing vessels) are vessels that run regularly scheduled fishing trips for which tickets are sold to the public. Partyboats represent a major segment of the recreational fishery off the west coast of the United States. We believe that partyboat trips sample the species composition at each location visited during a fishing trip better than private boat trips because the catch from a partyboat trip usually represents the fishing effort of many more anglers.

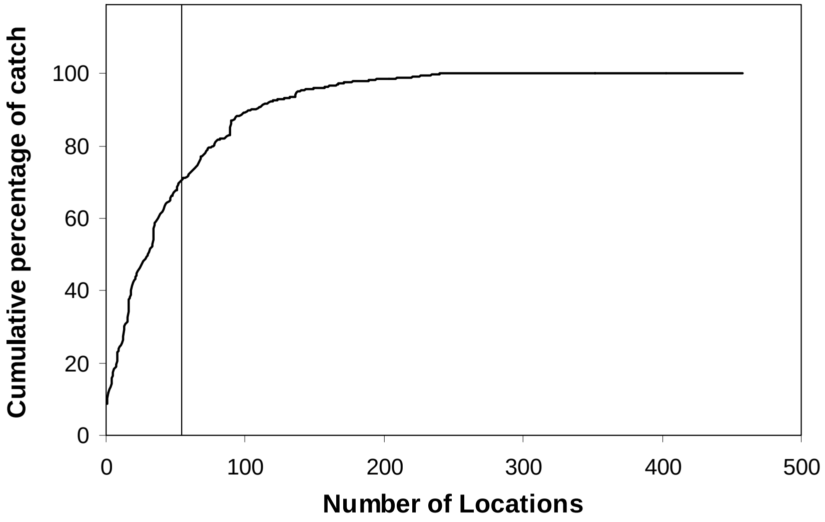
Three data sets for partyboats off northern California are considered in the analyses of this paper: a) catch and effort data sampled by the MRFSS program (1980-89; 1993-99) (MRFSS), b) site-specific catch and effort data sampled onboard fishing vessels by the California Department of Fish and Game (CDF&G) (1987-98) ('CDF&G site-visit'), and c) a version of the second data set created by reorganizing the CDF&G records so that site visits are aggregated into records of (location-blind) trips ('CDF&G aggregate-trip'). After calculating CPUE for bocaccio, all CDF&G and MRFSS catch data were converted from their original values to categorical presence / absence indicators (1/0).

The data from the MRFSS program were obtained from the RecFIN database (VanBuskirk, 2003). The MRFSS data are compiled from post-fishing interviews on the dock. MRFSS aims to obtain the distribution of the catch-per-trip at the species level, the unit of nominal fishing effort is an angler-trip, and fishing locations are not recorded. Many records, especially those from the early years, are incomplete or unclear (e.g. lacking information on date, number of anglers, or species caught). Deletion of such records prior to analysis reduced the data set by 20%. Data after 1999 were available, but were not included in the analyses because of major changes in fishing regulations, including reduced bag limits. The MRFSS / RecFIN data comprise 12905 usable records of catch composition and fishing effort.

The CDF&G data were provided by D. Wilson-Vandenberg (CDF&G, pers. comm.). The CDF&G sampling recorded catches and effort (in angler-minutes) at specific fishing sites. Data recorded by the CDF&G program include the location and duration of fishing at each site, the maximum and minimum depth at the site, and the number of each species of fish caught. We used 4544 per-site fishing observations from this dataset, comprising 458 locations and 106 species, and covering the period Jan 1987 – Dec 1998. The CDF&G program did not actively sample partyboat trips targeting salmon or tuna, and thus represents a subset of the MRFSS sampling frame (although not of its data; the two programs were conducted independently).

Ideally, a set of reference locations would be chosen for estimating CPUE. These are locations known to have good catch rates for the species of interest. This precludes consideration of locations that are rarely visited and locations at which the target species is rarely caught from having undue influence on CPUE. We used only those data pertaining to locations at which bocaccio had been caught ten or more times, comprising 54 reference locations from the 458 locations fished, for comparison of CPUE estimates in the CDF&G data (Fig. 2.2).

Fig. 2.2. The cumulative percentage of bocaccio catch versus the number of locations in the CDF&G data. The vertical line indicates the contribution to the total catch of the 54 reference locations.



The estimated abundance of bocaccio available to the central California recreational fishery declined by two thirds during the 1987-98 period sampled by the CDF&G program and by over 80% during the 1980-99 period sampled by MRFSS (Fig. 2.1).

2.2.2. Catch per Unit Effort

Determining which catch and effort records pertain to a particular target species, involves discriminating between trips that fished in habitat where the target species is found (which will be referred to as target habitat) from trips that fished in non-target habitat, i.e., in which the target species was unlikely to be caught. The latter trips are not informative, and potentially contaminate the calculation of CPUE. Ideally, nominal fishing effort (*E*) and the fishing mortality rate (*F*) for a species are related by a catchability coefficient, *q*:

$$F = qE \tag{1}$$

and average abundance (*B*) is related to the CPUE by:

$$B = (1/q)(C/E) \tag{2}$$

where *C* is the catch.

The actual value of the catchability coefficient may not be known, but, under the assumption that it is constant, CPUE is often used as an index of relative abundance when conducting stock assessments. Ideally, the measure of nominal fishing effort is defined so as to be proportional to the fishing mortality rate that it generates (Ricker, 1975). Thus, the catchability coefficient is equal to the fishing mortality rate generated by one unit of nominal fishing effort. Fishing is unlikely to catch the target species in non-target habitat, so $C \approx 0$ and $q \approx 0$. If the catch and effort records reflect a mixture of fishing activity in both target and non-target habitats, the catchability coefficient reflects the proportions of target and non-target effort in the mixture:

$$B = (1/q_{mixed})(C_{tar}/(E_{tar} + E_{nom})) \tag{3}$$

The subscript *tar* in Eq. (3) indicates records from target habitat, *non* indicates records from non-target habitat, and q_{mixed} refers to the catchability coefficient that applies to the combined data. This may not pose a serious problem under some circumstances. For example, if the data contain a constant proportion of target to total effort, the value of q_{mixed} will be smaller than q_{tar} by the ratio $E_{tar}/(E_{tar} + E_{nom})$, but will still be constant. However it is unlikely that this ratio will be invariant over long periods of time because many of the factors influencing the behavior and preferences of recreational fishermen may change.

Historically, calculation of CPUE involved straightforward ratio estimators, often supplemented by complicated analyses of fishing power used to address systematic differences in the catchability coefficient among different classes of vessels in the fleet (Gulland, 1983). More recently, generalized linear models (GLMs) have been used to derive indices of abundance more directly from catch and effort data (Stefánsson, 1996). A major advantage of the GLM approach is that a wide variety of influences on the catchability coefficient can be accounted for in a relatively simple analysis. For example, the distinction of target and non-target habitats is straightforward if fishing locations are known, and this can be incorporated directly in the analysis. Using the notation in the 'R' computing language (Ihaka and Gentleman, 1996), the CPUE index can then be obtained using a GLM of the form:

$$\log(\text{CPUE}) \sim \text{year} + \text{location} + \text{other} \tag{4}$$

where the exponentiated 'year' effects estimated by the model serve as the CPUE index. The 'location' effects account for systematic differences among fishing locations, and the 'other' effects could include sources of variability such as seasonal patterns in fish abundance or availability. Although, in principle, this approach could be applied to the entire catch and effort data set, it is still advantageous to delete records for locations that rarely or never produce the target species because the GLM treats fluctuations in relative CPUE at all locations as being equally informative. For example, if CPUE declines by half at well-measured target locations, CPUE should also decline by half at locations which rarely produce any catch of the target species, even though that change would scarcely be measurable. Of course, in the case where locations are known, it is rather easy to subset the data to include records only for those locations that consistently produce catches of the target species.

In this paper, we address the problem of how to subset catch and effort data for estimation of CPUE when fishing locations are not known. The proposed method uses the observed species composition to infer whether the fishing effort occurred in a habitat in which the target species would be expected to live. This inference takes the form of a logistic regression (described below) that uses the presence or absence of other common species to estimate the probability that the target species would be encountered. Selection of a critical value allows the catch and effort data to be divided into the records in target and non-target habitat. Once the data have been 'subsetted', the CPUE index can be obtained using a GLM of the form:

$$\log(\text{CPUE}) \sim \text{year} + \text{other} \tag{5}$$

where the exponentiated 'year' effects provide the CPUE index, and 'other' refers to any additional factors. The data include numerous records for which boccaccio CPUE was zero. We used a delta-gamma GLM, where presence-absence is modeled using a logistic regression (binomial family in the R computing package), and the records with non-zero values are modeled using a separate GLM assuming a gamma probability distribution (Stefánsson, 1996; Dick, 2004). Estimates of precision for the annual CPUE indices are obtained using a jackknife procedure (Belsley et al., 1980).

The model we used to calculate CPUE is a main-effects model. We investigated interaction terms and found they were rarely significant and ranged between three and five orders of magnitude smaller than the main effects, justifying their omission (Maunder and Punt, 2004).

2.2.3. Logistic regression

Statistical classification problems, such as the present subsetting problem, are typically addressed using either discriminant function analysis or logistic regression. Press and Wilson (1978) reviewed the properties and performance of these two approaches. Discriminant function analysis (McCullagh and Nelder, 1989) requires that the variables be normal with identical covariance matrices. Logistic regression with maximum likelihood estimation is preferable if the explanatory variables are not multivariate normal, such as in the present case where they are categorical variables.

Although individual fishing locations may not be known, the species composition of a fishing trip provides information that can be used to infer whether the fishing trip included effort expended in target habitat. We use a logistic regression to make this inference. The species compositions from catch records are first used to estimate the parameters of the logistic regression which then used to estimate the probability that the target species would have been encountered on each trip. Those records for which the estimated probability exceeds a chosen critical value are then used in the CPUE analysis with some assurance that many of the records of catch and effort from non-target habitat have been removed.

Let Y_j be a categorical variable describing the presence / absence of the target species for trip *j*:

$$Y_j = \begin{cases} 1 \text{ if the target species is caught} \\ 0 \text{ if the target species is not caught} \end{cases}$$

Similarly, let x_{ij} describe the presence / absence of non-target species *i* in the catch during trip *j*.

We assign a score for each trip *j* as a function of the species $(1, 2, .., k)$ caught during that trip:

$$S_j = \exp \sum_{i=0}^{k} x_{ij} \beta_i \tag{6}$$

The coefficients $\beta_1, \beta_2, .., \beta_k$ quantify the predictive impact of each species while β_0 is the intercept of the regression - the probability that fishing was in the habitat of the target species when none of the others species was present.

This score is then converted into a probability of observing the target species given the vector of presences and absences of the *k* non-target species:

$$\pi_j = \Pr\{Y_j = 1\} = \frac{S_j}{1 + S_j} \tag{7}$$

where π_j is the predicted probability that $Y = 1$ for trip *j*.

Given $\beta_0, \beta_1, .., \beta_k$ and the presence / absence indicators $x_{1j} ... x_{kj}$, the log-likelihood (excluding constants independent of the parameters) is the sum:

$$\boldsymbol{L}\{Y \mid \beta_0 ... \beta_k, x_{1j} ... x_{kj}\} \quad = \sum_{j \in j+} log(\pi_j) + \sum_{j \in j-} log\left(1 - \pi_j\right) \tag{8}$$

where *j*+ denotes records where the target species was caught, and *j*- denotes records where the target species was not caught.

The log-likelihood is maximized using the statistical package R (Ihaka and Gentleman, 1996). The estimated β coefficients reflect the association (positive or negative) between the non-target and the target species, and the π_j is the estimated probability that trip *j* occurred in the habitat of the target species.

The set of trips to be used in the CPUE analysis is defined as those for which π calculated above is less than a critical value. The critical value is selected so the number of incorrect predictions (both false positive – the target species is estimated to be found in the habitat fished during the trip when it doesn't, and false negatives – the target species is estimated not to be found in the habitat fished when it does) is a minimum. This number is quantified by the absolute value of the difference between the number of trips observed to have caught the target species, and the number proposed to be in target habitat. We evaluate this difference as the critical value is increased from zero (all trips are in target habitat) to one (no trips are in target habitat) and identify the value that leads to the smallest absolute difference.

2.2.4. Validation with known locations

The 'CDF&G site visits' data set (for which location is known) was analyzed in two ways as a 'sea truth' to validate the proposed 'subsetting' approach:

a) We fitted the following model, which includes location as a covariate, assuming a binomial error distribution, to estimate the probability of encountering bocaccio at each location:

$$Y \sim \text{location} + \text{year} + \text{season} \tag{9}$$

where Y indicates bocaccio presence / absence and there are 12 years and 4 (trimester) seasons. Interaction terms could be included in Equation (9) but their inclusion was not supported statistically.

b) We applied the proposed 'subsetting' approach to determine probability of encountering bocaccio in each location.

This validation analysis was performed for all catch records for the locations at which bocaccio occurred at least once.

2.3. Results

2.3.1. Validation with known locations

We compared the performance of the proposed method for ‘subsetting’ catch and effort records (Section 2.2.3) with the location-based method (Eq. 9) using the ‘CDF&G site visits’ data set. 106 species are recorded in this data set, but 30 account for 99% of the catch (Fig. 2.3). The two methods were therefore applied to both the full (106 species) and restricted (30 species) data sets. The results are insensitive to the number of species, so the results reported pertain to the 30 species data set only. A backwards stepwise-regression procedure was used to reduce the regressor species used by the proposed method further. Fig. 2.4 shows the regression coefficient for each non-target species retained for the analysis of site-visits. Species that were never caught with bocaccio are lumped into a category of ‘non-coocurring species’.

Fig. 2.3. The cumulative percentage of catch versus number of species in the CDF&G (solid line) and MRFSS (dashed line) data sets. The vertical lines indicate the contributions of the species used in the analyses.

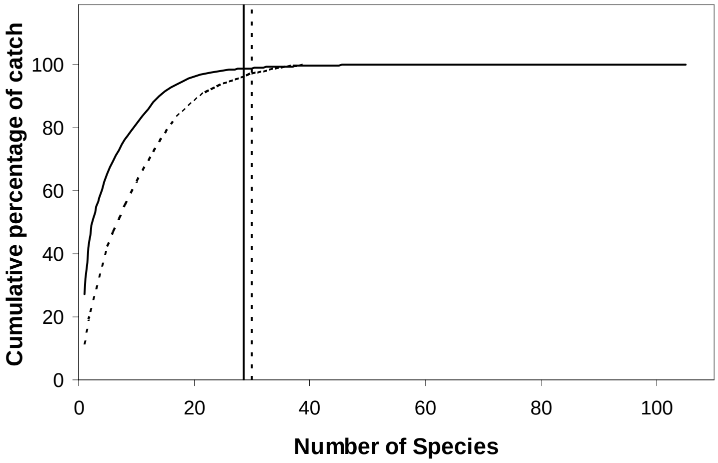


Fig. 2.4. Estimates of species-specific regression coefficients based on the 'CDF&G site-visits' data set.

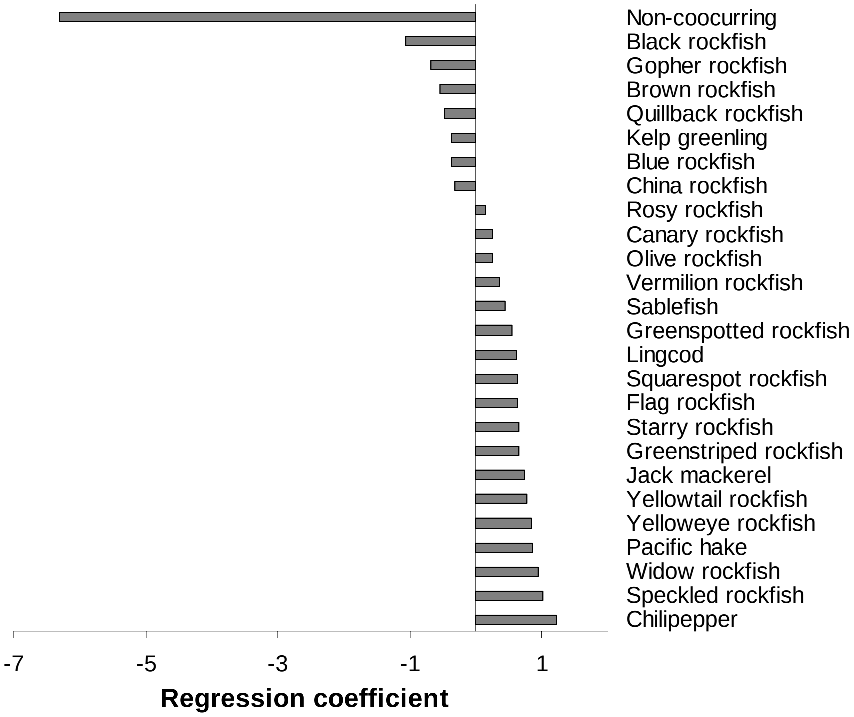


Figure 2.5 compares the estimated probability of encountering boccacio for each location from: a) Equation 9 – x-axis, and b) the proposed method – y-axis. The estimated probability of encountering boccacio is higher for the proposed method than when direct account is taken of location. However, for the purposes of subsetting the data, the important issue is the relative ranking of locations and not the estimated probability of encountering boccacio. Fig. 2.6 therefore plots the locations ranked by the species-based method (x-axis) against the number of locations ranked equally or better by Equation 9 (y-axis).

Fig. 2.5. Per-location probabilities of encountering boccacio based on regressions using location (x-axis) and species composition (y-axis) as predictors.

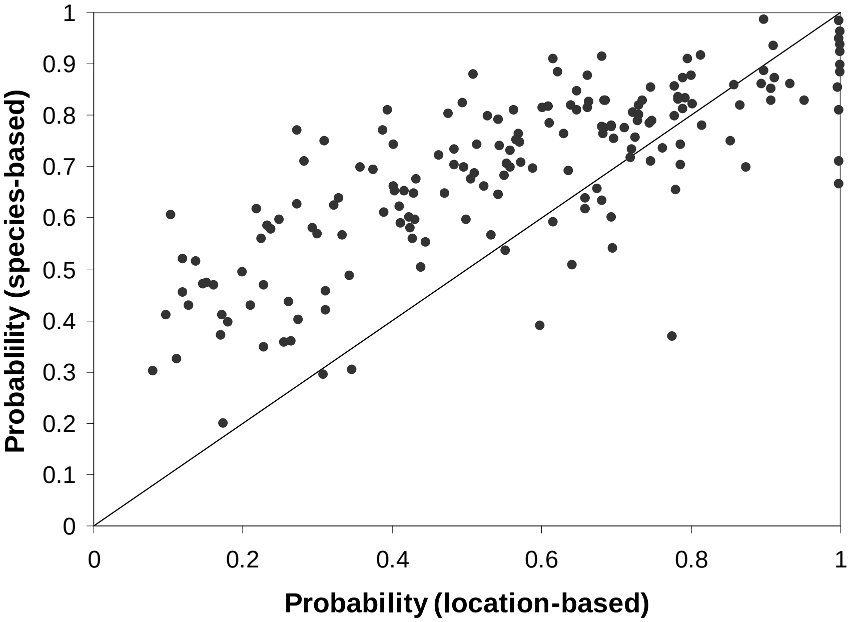


Fig. 2.6. Locations ranked by the species composition method (best to worst) – x-axis, and the number of locations ranked equally or better using Equation 9 – y-axis.

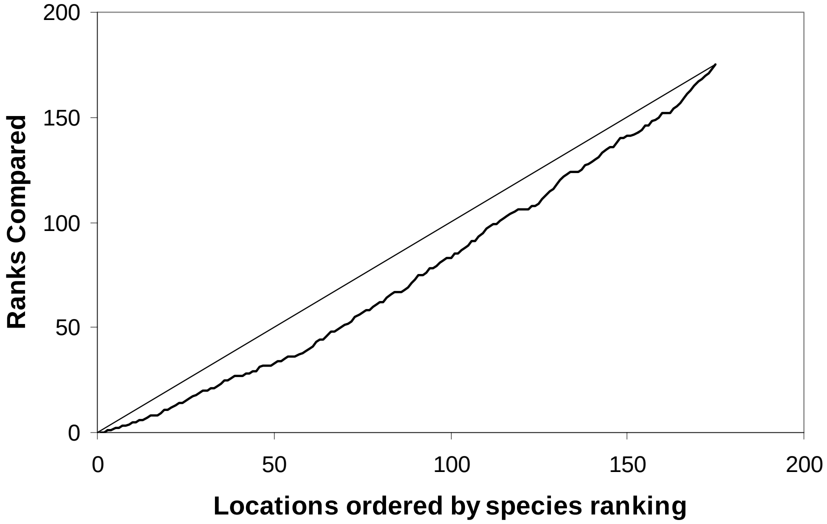
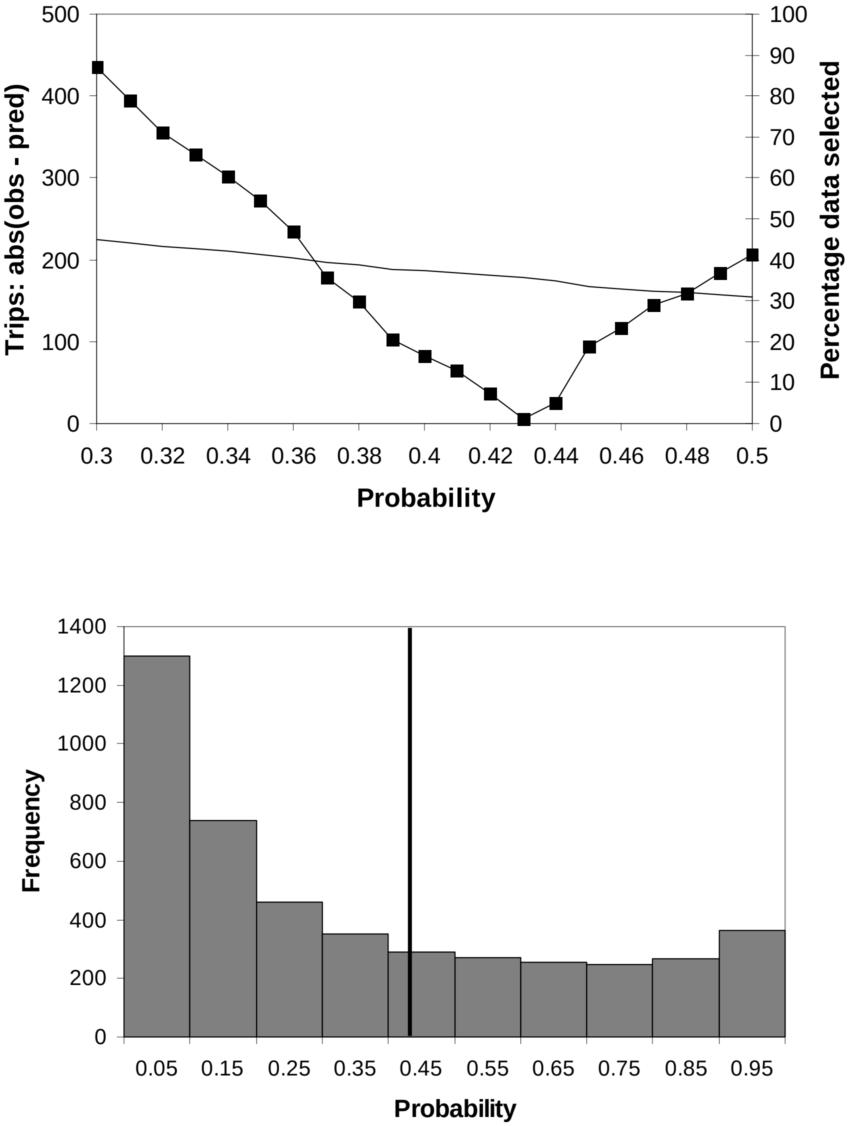


Figure 2.7 provides additional diagnostic statistics for the proposed method. The critical probability at which the difference between the observed and expected number of trips encountering boccacio is minimized is clearly defined and equals 0.43 (Fig. 2.7, upper panel). About one third of the records are selected for use in calculating the CPUE index, although this fraction is not particularly sensitive to the critical value in the range evaluated (Fig. 2.7, solid line). The distribution of the probability of encountering boccacio among sites suggests that many site visits have very little chance of catching bocaccio (Fig.2.7, lower panel). These are the least relevant records for estimating the CPUE index, and are discarded by the subsetting procedure.

Fig. 2.7. Results of the application of the proposed method to the ‘CDF&G site-visit’ data (*n*=4 44). The upper panel plots the difference between the number of records in which bocaccio are observed and the number in which they are predicted to occur (symbols), and percentage of records retained (solid line), as a function of the critical value while the lower panel shows a histogram of the probabilities generated by the species-based regression. The vertical line indicates the critical value for which false prediction is minimized.



2.3.2. Evaluation of aggregate trip data

The critical probability value increases from 0.43 to 0.53 (Fig. 2.8, upper panel), and the distribution of probabilities shifts to larger values (Fig. 2.8, lower panel) when the CDF&G data are aggregated. Actual fishing trips rarely visit only one location, and, in fact, usually visit at least two locations per trip which means that a greater percentage of the aggregate trips encounter bocaccio at some point.

Another change that occurs when the data are aggregated is that fewer explanatory species remain from the original 30 used when analyzing the site-visit data after the stepwise-regression (Fig. 2.9). Since the catch in an aggregate trip includes more species than an individual site-visit catch, species that were only weakly informative for site-specific data become even less informative for aggregate data.

Fig. 2.8. Results of the application of the proposed method to the 'CDF&G aggregate-trip' data (*n*=2 267). The upper panel plots the difference between the number of records in which bocaccio are observed and the number in which they are predicted to occur (symbols), and percentage of records retained (solid line), as a function of the critical value while the lower panel shows a histogram of probabilities generated by the species-based regression. The vertical line indicates the critical value for which false prediction is minimized.

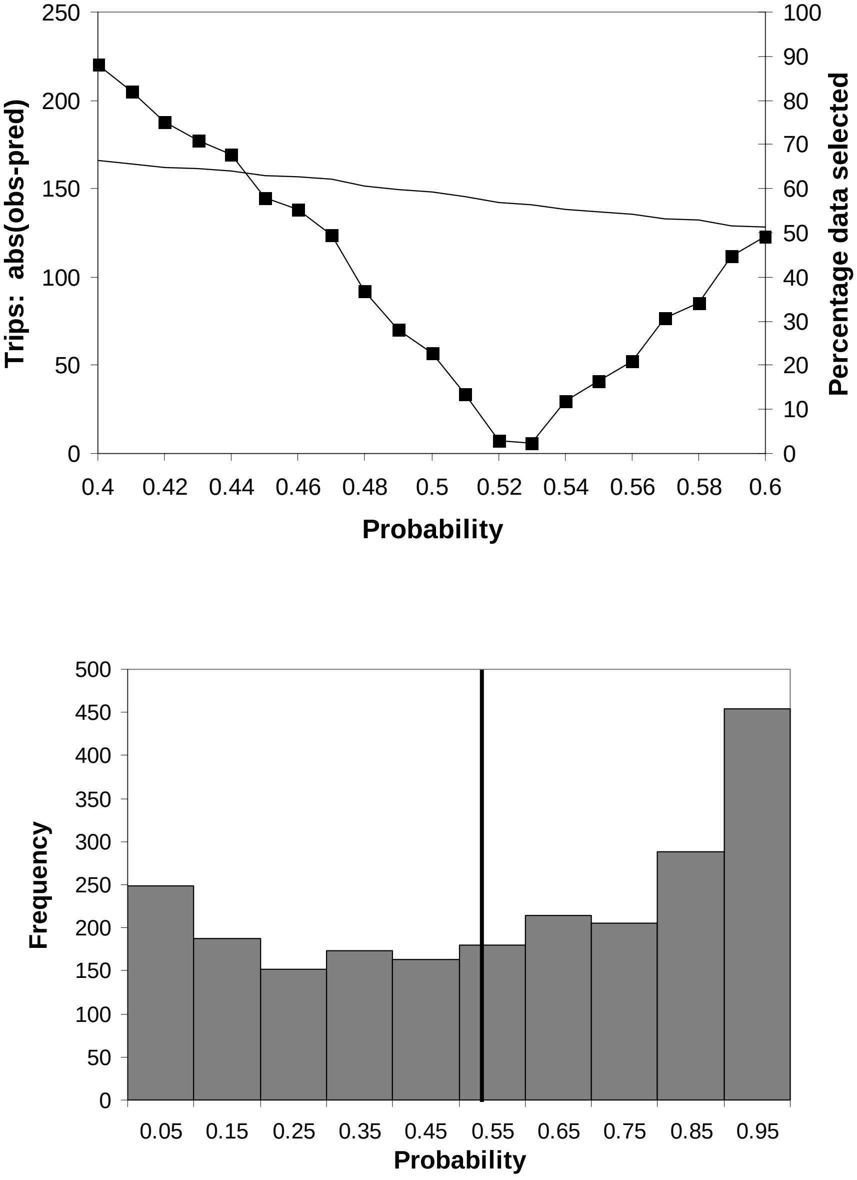
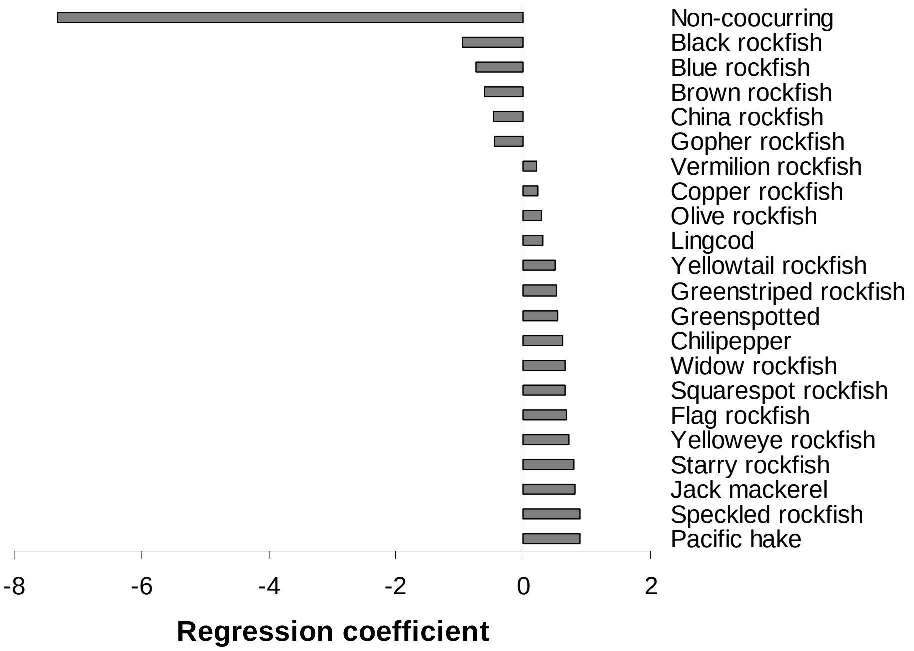


Fig. 2.9. Estimates of species-specific regression coefficients based on the ‘CDF&G aggregate-trip’ data set.



2.3.3. Application to the MRFSS data

We used 30 species when applying the proposed method to the MRFSS data to be consistent with the analysis of the CDF&G data. This amounts to 75% of the species, and 97% of the catch (Fig. 2.3). The critical value analysis (Fig. 2.10, upper panel) and probability histogram (Fig. 2.10, lower panel) suggest that bocaccio are less prevalent in the MRFSS data set than in the CDF&G data set. This reflects a difference in the data collected. For example, the MRFSS data set includes a large number of salmon and tuna trips, which typically do not visit bocaccio habitat. Figs

2.4 and 2.11 show that the relationships among the species are consistent (in terms of both magnitude and sign of their associated coefficients) between the MRFSS and CDF&G data.

Fig. 2.10. Results of the application of the proposed method to the MRFSS data (*n*=12 905). The upper panel plots the difference between the number of records in which bocaccio are observed and the number in which they are predicted to occur (symbols), and percentage of records retained (solid line), as a function of the critical value while the lower panel shows a histogram of probabilities generated by the species-based regression. The vertical line indicates the critical value for which false prediction is minimized.

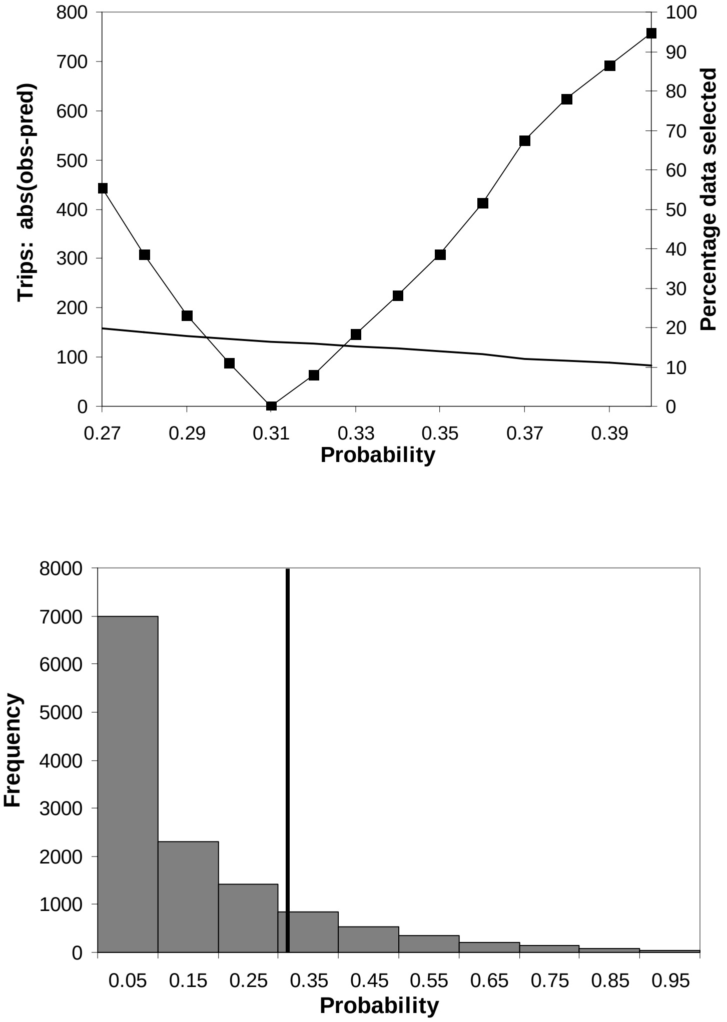
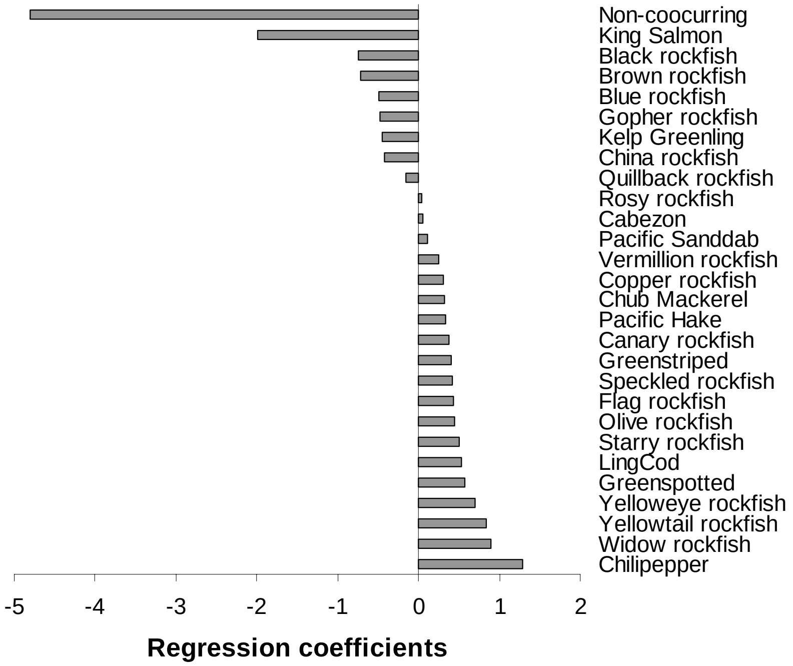


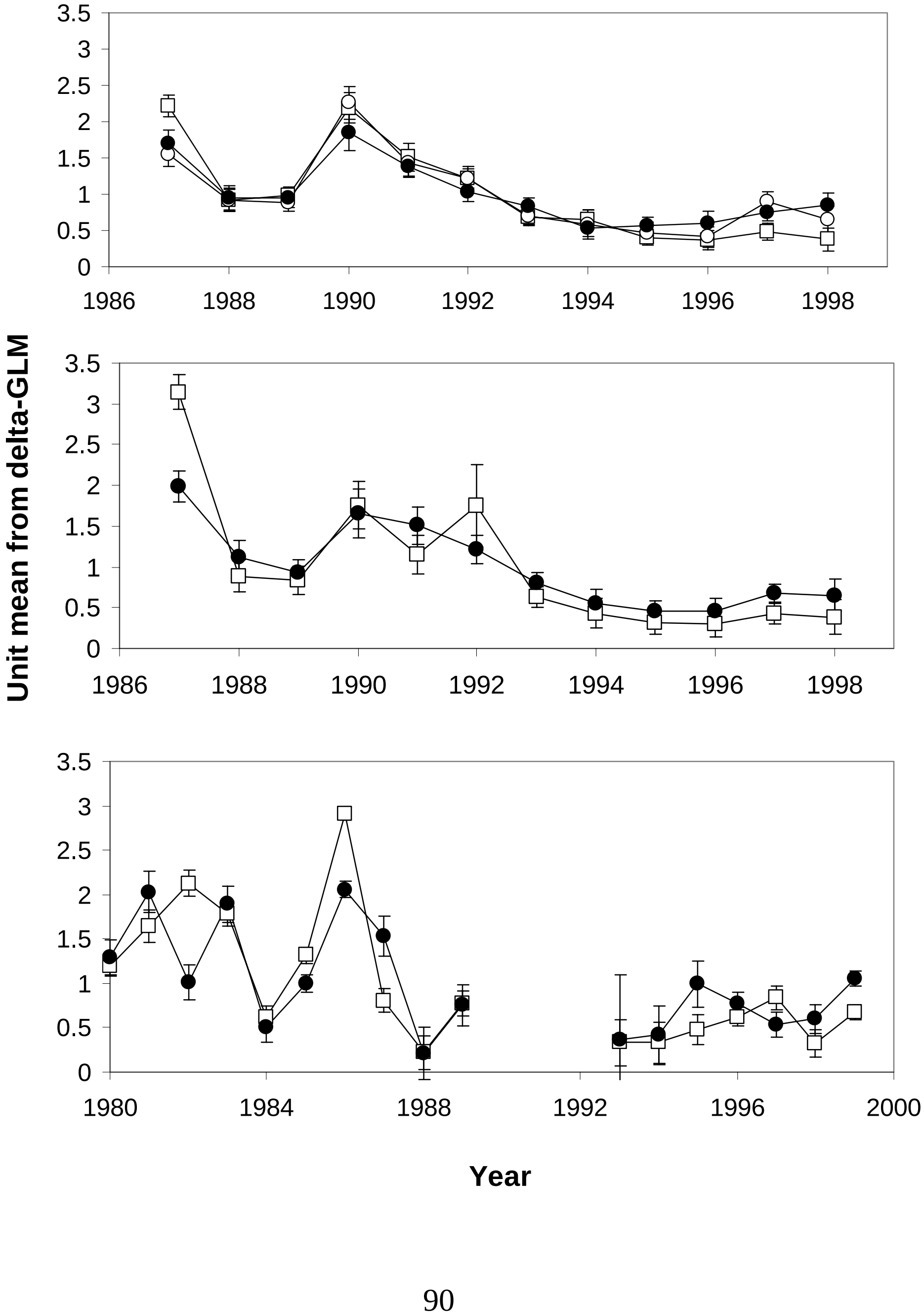
Fig. 2.11. Estimates of species-specific regression coefficients based on the MRFSS data set.



2.3.4. CPUE analysis

The decline of the CPUE indices based on the full (i.e. no exclusions of non-targeted records) ‘CDF&G site-visits’ data (open squares in Fig. 2.12, upper panel) is exaggerated compared to that of the CPUE indices based on data subsetted by location (open circles) or species catch composition (closed circles), particularly after 1995. A similar exaggerated decline in CPUE is apparent for the ‘CDF&G aggregate-trip’ data set (Fig. 2.12, middle panel).

Fig. 2.12. Time-series of CPUE from analyses of CDF&G site-visit data (upper panel), CDF&G aggregate data (center panel), and MFRSS data (lower panel). The CPUE indices based on all records are indicated by open squares, those from records selected using location criteria by open circles and those selected by species regression by closed circles. The errors bars indicate one standard error.



Subsetting the MRFSS data changes some of CPUE indices considerably (Fig. 2.12. lower panel). For example, 1998 was an El Niño year, and a good year for tuna. Many partyboat trips specifically targeted tuna that year. Compared with the abundance trends in Fig. 2.1 (which were based on nine data sets), the CPUE index from the species regression follows the initial decline to 1984 better than the CPUE index from the full data set, and, apart from 1993 and 1994, is relatively constant during the 1990's.

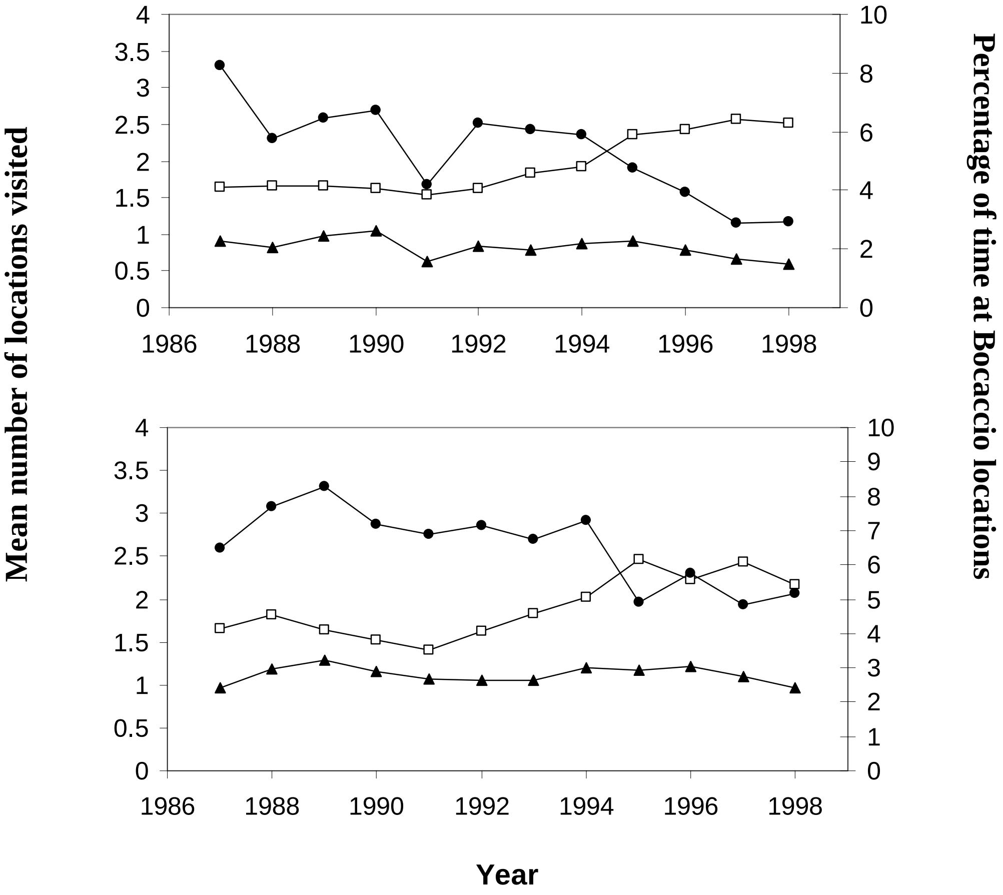
Other discrepancies in these data may be explained in terms of life-history. Bocaccio recruitment is generally low with rare, large recruitments. The years 1980 and 1985 were large recruitment years (MacCall, 2003), providing large numbers of young fish for anglers in 1982 and 1986. The CPUE indices for 1982 and 1986 based on the full data set are much higher than those based on 'subsetted' data presumably because boccacio were being caught outside the usual habitats (trips in such habitats are assigned low probabilities by the proposed method and may be discarded) as well as within them.

2.3.5. Site-specific changes in effort

The number of locations visited per trip in the CDF&G data, and the percentage of fishing time spent at the top 54 bocaccio locations (those at which bocaccio occurred

10 or more times) changed over time. According to the CDF&G ‘site-visit’ data set, the average number of locations visited during a trip rose by 45% from 1987 to 1998, while the number of visits to top bocaccio sites stayed the same, indicating an increasing diversification of fishing sites over time (Fig. 2.13, upper panel). The percentage of the time spent fishing the best bocaccio sites dropped by 64% over 1987-98. In other words, during the period of bocaccio decline, vessels switched targets and progressively targeted habitats where bocaccio were less likely to be present. This target switching could not have been easily detected without the location-based data, and its effect cannot be entirely removed from species-subsetted data (Fig. 2.13, lower panel); the same pattern of target diversification persists, although the trend is less pronounced. There is a 40% increase in the number of sites visited per trip, and a 20% decrease in time spent fishing at the bocaccio sites.

Fig. 2.13. Mean number of locations visited per trip (squares), mean number of visits to top bocaccio sites per trip (triangles), and percentage of time spent at top bocaccio locations (circles). Results are shown in the upper panel for the full 'CDF&G site-visit' data set and in the lower panel for the same data set after subsetting.



2.4. Discussion

The three datasets are similar in terms of CPUE trends, critical value analyses and species selection. The species coefficients for the regressions are satisfying from a biological perspective, with regard to both magnitude and direction of influence. In particular, presence of chilipepper (*S. goodei*) is consistently a strong positive predictor of bocaccio, and the two species are well known to co-occur in fishery landings (Williams and Ralston, 2002). In fact, they were treated as a single species in some assessments until fairly recently (Ralston et al., 1998). Presence of black rockfish (*S. melanops*), a species with a more northerly range than bocaccio (Williams and Ralston, 2002), is a negative predictor in all three datasets.

The tradeoff when selecting the critical value is between choosing more data (data quantity), which increases precision, and including less-relevant data (data quality), which decreases both precision and accuracy; these two aspects are assumed to be approximately equal in the vicinity of the proposed critical value. If issues of data quantity and quality are not of equal concern, a different critical value could be considered. The critical value and probability analyses all show that precise cutoff values can be identified to distinguish data subsets (Figs 2.7, 2.8 and 2.10). Further, the probability distributions themselves identify characteristics of the data sets, such as the increased probability of bocaccio in the aggregate trip data (Fig. 2.8), and the predominance of low-probability trips associated with inclusion of more non-targeted fishing activity in the data collected by the MRFSS survey (Fig. 2.10).

We chose to restrict the subsetting analysis to categorical presence and absence data. Abundance of the explanatory species (i.e. their CPUE) could be used as explanatory variables in a similar approach. We prefer use of presence and absence data, because they should be less influenced by trends in abundance of other species. Using a rather large number of explanatory species also contributes to minimizing the effect of abundance trends.

The CPUE indices based on the ‘subsetted’ data sets (closed circles in Fig. 2.12) follow the abundance trend from the bocaccio stock assessment (Fig. 2.1) better than those based on the full data sets (squares in Fig 2.12). The unique location-based data in the CDF&G dataset allows us to examine some of the sources of bias that can appear in aggregate trip data. Target switching (Fig. 2.13) would not have been directly visible without the location-based data, and its effect cannot be removed entirely from species-subsetted data. Our approach to subsetting trips has removed some, but not all, of the confounding influence of target switching.

Logistic regression of target species occurrence on presence and absence of other species provides a practical method for subsetting recreational fishing catch and effort data, and could be applied to many other types of multispecies abundance data where there is a mixture of relevant and non-relevant records (see, for example, Guisan et al.

(2002), for a discussion of a similar application in terrestrial settings). This method is especially valuable in that it is reproducible by independent analysts. It also reduces the need for *ad hoc* decisions in stock assessments, and should contribute to improved consistency among such assessments. Subsetting the data using a species-based logistic regression also removes, or at least reduces, a common criticism about use of recreational CPUE data: that target switching can result in spurious trends in the abundance index.

Acknowledgements

The authors gratefully acknowledge Deb Wilson-Vandenberg (CDF&G) for providing data on northern California partyboat sampling. We also thank Wade VanBuskirk (Pacific States Marine Fisheries Commission) for providing the detailed RecFIN data for this project. An early version of this work was begun with Teresa Ish. E.J. Dick consulted generously on delta-GLM analyses. We thank Nick Davies and an anonymous reviewer for their helpful comments. We especially thank André Punt and Marc Mangel for numerous discussions and suggestions. Funding for this research was provided by the Center for Stock Assessment Research.

Challenging A Multispecies Logistic Regression for Subsetting Catch-Effort Data by Simulation

Abstract

Many statistical methods can be straightforward to use, but difficult to interpret. One way to develop a better understanding of a method, its metrics, and its limitations is to apply the method to a dataset in which the answers are already known. In this study, I simulate data to resemble fishery records of catch in a multispecies fishery. I then employ a logistic regression method that uses the species present in a region to predict habitat (Stephens and MacCall, 2004). Analysis of the regression results provides insight into the limitations of the method: it fails to perform well when data are limited, when either the target species or the regressor species do not practice site fidelity, or when the regressors are predominantly negative or predominantly positive predictors of the target species. The regression is relatively robust to changes in regressor populations

3.1. Introduction

In the interests of good fisheries management, we need to know the current and historical abundance of a species in order to know whether its population is stable, in decline, or increasing. The goal of fishery management is to maximize the equilibrium harvest rate; the rate at which fish can be caught while maintaining the productive population (Iudicello, *et al.*, 1999). In order to do this, we need to understand the processes that contribute to population change. One simple representation of the population dynamics of a fished species is given by the Schaefer model:

$$\frac{dB}{dt} = rB\left(1 - \frac{B}{K}\right) - C \qquad (1)$$

where B is the abundance (numbers or biomass) of the species, r is its intrinsic growth rate, which encompasses both the birth rate and the rate of natural mortality, K is the carrying capacity of the environment, and C is the catch, or harvest rate. Knowing each of these values, we can predict the way in which the population will change with time, and implement as a management strategy the largest catch rate that sustains a stable population.

However, the only one of these variables that can be known with any certainty is the rate of catch. Then the abundance of a species must be calculated from records of catch, which typically include information about the number or biomass (tonnage) of fish of each species taken, and information about the type of fishing effort (trawling or line-fishing) or number of fishers involved, as well as the number of hours or days expended on a fishing trip. If we make the assumption that the catchability of a species – its susceptibility to fishing effort – is unchanging over time, we can extract species abundance information from fishing records according to the following relationship:

$$C = qEB \tag{2}$$

where C is the catch, E is effort, and B is the abundance of the species; q is the catchability factor. Consequently, C/E provides an index of abundance:

$$\frac{C}{E} = \frac{qEB}{E} = qB \propto B \tag{3}$$

This is the Catch per Unit Effort, or CPUE (Gulland, 1983). Assuming that the catchability is constant, fluctuations in the CPUE reflect fluctuations in the abundance of that species (Ricker, 1975).

If in calculating CPUE we include effort not directed at the species we are attempting to evaluate, we underestimate its abundance. If the extraneous effort were constant in time, CPUE would still be proportional to abundance; however this is not often the case. Effort may change over time due to changes in fishing regulations (e.g., bag limits on the number of fish caught), weather patterns, or consumer preference for market fish. These may not affect the species of interest. In a fishery with multiple targets, such as the California recreational fishery, fishing for major targets such as tuna may completely mask population changes in low-abundance species (Stephens and MacCall, 2004). In analyzing the data for a fishery such as this, the problem is one of teasing out the effort pertaining to a single species.

Many species of fish prefer certain types of habitat (Williams and Ralston, 2002). For example, benthic species may require cobbled or sandy bottoms, and many demersal species are found only at certain depths. If we assume that fishing trips to a particular habitat constitutes effort to catch the group of species there, we need to know the location of fishing trips in order to judge the type of effort they represent. In a commercial fishery, information about the locations fished is available as crew-reported latitude/longitude, or is recorded as a GPS track. However, location information is rarely available for a recreational fishery, since the locations of the great fishing spots are jealously guarded information. The California Department of Fish and Game observer program has collected 12 years of location-specific catch-and-effort data for some participating recreational party-boats: captains agreed to permit data collection on the condition that fishing locations not be published (D. Vandenberg-Wilson, CDF&G, pers. comm.).

Fortunately, the species present in the catch from a fishing trip can tell us about the types of habitats visited during that trip (Stephens and MacCall, 2004). Once we associate habitats with catch records, we can select the subset of the data associated with a particular habitat. CPUE for a species frequenting this habitat can then be calculated from effort expended there, excluding extraneous fishing it places the target species never occurs.

For example, half or more of the fishing effort in the California recreational fishery may be devoted to tuna fishing offshore, but this fluctuates from year-to-year (VanBuskirk, 2003). Fluctuations in CPUE for groundfish targets calculated using all effort would be dominated by fluctuations in the tuna fishery. By removing this extraneous effort from calculations for rarer species, we calculate a CPUE that fluctuates more closely with the population of interest.

The multispecies logistic regression of Stephens and MacCall (2004) calculates the probability that the catch record of an individual fishing trip contains the target species (regardless of whether it actually did) by using the presence or absence of other species in that catch record. Effort for the target species is then derived from the subset of the trip records with high probabilities.

The numerical products of a logistic regression include (a) probabilities that each catch included the target species, (b) regression coefficients that describe the relationship of the various indicator species to the target, and (c) regression measures such as the χ^2 test statistic that can be used to determine the goodness of fit of the model to the data. It can be difficult to interpret these products in terms of the natural world; however scientists in many fields of research are rapidly adopting this type of species-based regression as an analytical technique (Guisan, *et al.*, 2002). How do we determine whether or not a particular situation lends itself to regression analysis?

We might want to ask certain questions about the multispecies regression before we apply it to a particular study. For example: does the regression method work differently for target species with different abundances? What happens to regression metrics if species abundances change? Species may exhibit many different characteristics with respect to habitat use; for example, they may be ubiquitously distributed, or they may be found only in a narrowly defined habitat. Does the regression fail to perform well when the target species is ubiquitously distributed? Ocean weather cycles may drive species onshore or offshore, or to a different region. Does a change in habitat use adversely affect regression performance? In a fisheries context, the records of catch used in the regression may only represent a small percentage of fishing trips. For the California recreational fishery, this is estimated to cover about 10% of recreational fishing trips (A. MacCall, NMFS, pers. comm.). At what point do the data become so sparse that the regression fails? Most importantly, we need to know the characteristics of a system in which the regression seems to perform well, when in fact it does not.

In order to understand under what conditions the multispecies logistic regression is useful and when it may fail to provide insight, I challenged the method of Stephens and MacCall (2004) in a simulation study, applying it to a simulated ocean in which all of the details are known. I developed a model in MATLAB (Mathworks, 2004) to distribute populations of fish in a simulated ocean. Species belonging to different habitat groups are distributed according to habitat in the ocean. I used different distribution schemes to generate fishery records that reflect a variety of physical and biological ocean conditions.

I simulate conditions in which species change their use of the habitat, migrating from one area to another, as might occur with changes due to the Pacific Decadal Oscillation. I also simulate changes in the abundance of species within and outside of the target species consortium. This is the type of change that might occur due to changes in fishing practices, or differential vulnerabilities to fishing. Finally, I simulate a fishery with sparse data. This reflects the actual data collection practices in the California recreational fishery. I apply the multispecies logistic regression to these datasets, investigating the response of regression diagnostics to varying conditions, as well as to different characteristics of target and regressor species.

3.2. Methods

3.2.1. Logistic regression

In order to distinguish catch records pertinent to a particular target from those which are not, I use the species composition of the catch to infer whether the target species could have been among the species caught. The assumption inherent in this approach is that a particular habitat or set of habitats were fished in order to produce the combination of species present in the catch. In order to infer the possibility of the target species, I use a logistic regression of the catch data, which are simply the presence or absence (1 or 0) of each species in the catch. Classification problems such as this are typically addressed using either logistic regression or discriminant function analysis (Press and Wilson, 1978). The basic idea underlying discriminant function analysis is the determination of whether groups differ with regard to the mean of a variable, which is then used to predict group membership of new cases. However, discriminant function analysis requires that the variables be normally distributed, with identical covariance matrices (McCullagh and Nelder, 1989). Since the catch data are discrete, categorical variables rather than continuous, normal variables, logistic regression with maximum likelihood estimation is the preferred approach (Press and Wilson, 1978).

Logistic regression is a generalized linear modeling approach in which data consisting of a dependent or response variable is related to a number of covariate or explanatory variables. Fitting the model is a process of estimating a coefficient describing the predictive strength and the direction of the relationship (positive or negative) of each covariate to the dependent variable. The coefficients are estimated concurrently (which retains some measure of independence among them) rather than sequentially, and the goodness-of-fit of the coefficients is determined by whether or not they maximize the log-likelihood of the model. For binomially distributed data, the log-likelihood is:

$$\mathbf{L}\left\{ T \,\middle|\, \beta_0 \ldots \beta_k, P_{1j} \ldots P_{kj} \right\} = \sum_{j \in j+} log(\pi_j) + \sum_{j \in j-} log\left(1 - \pi_j\right) \tag{4}$$

Where $\beta_{1..}$ β_k are the regression coefficients for the k covariate species, and β_0 is the regression intercept. $P_{1j} \ldots P_{kj}$ are the presence/absence indicators of the covariates in each trip j, and T is the presence/absence vector for the target species. The probability that the target was caught in trip j is π_j if $T_j = 1$ (trips designated j+), and if $T_j = 0$, the probability that the target was caught is $(1 - \pi_j)$ (trips designated j–).

The probability π_j is calculated for each trip as:

$$\pi_j = \frac{exp \sum_{i=0}^{k} P_{ij} \beta_i}{1 + exp \sum_{i=0}^{k} P_{ij} \beta_i} \tag{5}$$

This equation for the log-likelihood is derived from the binomial probability density function, excluding constants independent of the parameters. For trips in which only the target species is caught, this probabililty is dependent only on β_0.

As the model is fitted, the presence/absence data from catch records are used to estimate the parameters of the logistic regression. These are then used to estimate the probability that the target species would have been encountered on each trip. The log-likelihood is maximized using the statistical analysis package R (Ihaka and Gentleman, 1996), which performs an interative gradient-search using partial derivatives, a method based on the Newton-Raphson method (McCullagh and Nelder, 1989).

The set of catches assumed to be in the habitat of the target species, and therefore predicted by the model to include it is those for which π calculated above is greater than a critical threshold value. If I use a critical value of 0, all catches are predicted to include the target. At a critical value of 1, none are. I define the "best" threshold value as one that minimizes the difference between the number of target catches predicted and the number actually occurring in the data. I identify the probability value at which this difference is minimized by iteratively choosing a value, calculating the predicted catch, and comparing it to the observed catch. This choice for the definition of the critical threshold is somewhat arbitrary, but it offers the advantage that the model will usually make similar numbers of false negative and false positive predictions; catches in which the target occurred but was not predicted, and those in which it was predicted but did not occur.

3.2.2. Generating the data

The yearly catch occurs in a 25x25 cell grid representing a small section of coastal ocean (Figure 3.1). The edges of the grid represent the northern and southern extents of the fishing grounds, the shoreline, and the offshore fishing limit. Six areas of the spatial grid are chosen to be habitat centers for species groups, one each for "Northern", "Southern", "Ubiquitous", "Onshore", "Pelagic", and "Rocky-Reef" habitats.

Species in each group are normally distributed around their population center, and the variances associated with the x and y dimensions define the extent of the habitat in each dimension (Table 3.1). The global variance, σ^2, is set at the beginning of the simulation. This parameter determines the extent of each group's habitat along those dimensions for which it is allowed to vary. For example, habitat for the Onshore group extends along the shore from north to south, but as the global variance increases, they can be found further offshore. Similarly, the Rocky Reef fish are found further from the center of the reef as the global variance increases. This regulates the extent of overlap of the habitats. Several species, of differing abundances, may belong to each habitat group (Table 3.2).

Figure 3.1. Ocean habitats. The simulated ocean consists of edge and reef habitats. These are delineated by heavy lines. Species in the "Ubiquitous" habitat group are distributed throughout the ocean.

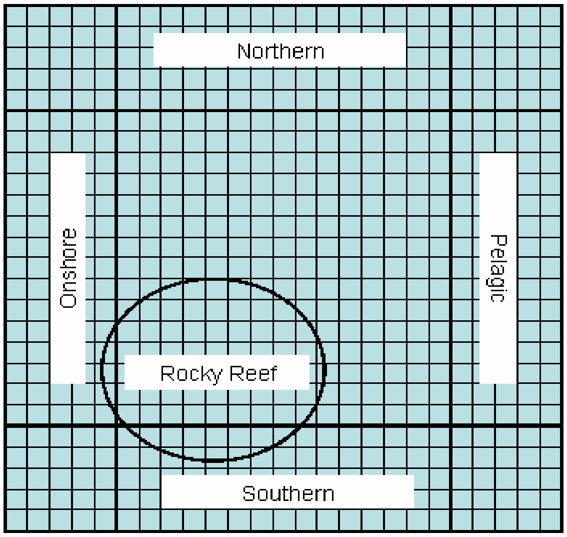


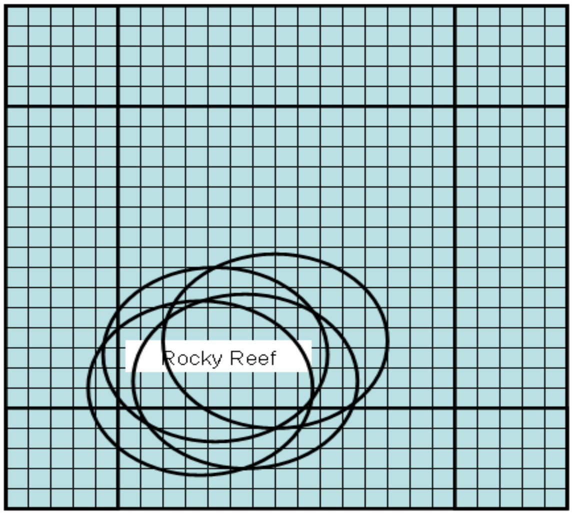
Table 3.1. Habitat characteristics. Coordinates for the center of each habitat group define its location in the ocean grid, and variances determine the extent of population distributions in each dimension. Note that for the Ubiquitous species, variances are always fixed.

Habitat Group	**Population Center (x,y)**	**Variance (σ_x^2, σ_y^2)**
Rocky Reef	15,10	σ^2,σ^2
Ubiquitous	12,12	250,250
Onshore	2,12	σ^2,250
Pelagic	23,12	σ^2,250
Northern	12,2	250,σ^2
Southern	12,23	250,σ^2

Table 3.2. Species populating the ocean. Entries across rows are the number of species of each size-class defined in each habitat group, and species are in size-classes "Rare" through "Superabundant".

Population size →	**Rare**	**Low**	**Average**	**Common**	**Superabundant**
	30	**200**	**300**	**600**	**4000**
Habitat Group ↓					
Rocky Reef (R)		3	1	1	
Ubiquitous (U)		1	1	3	1
Onshore (O)	2	1	2		
Pelagic (P)	1				1
Northern (N)		1	1	1	
Southern (S)	1			2	

Figure 3.2. Differing use of the habitat by individual species within the group. Illustrated for the Rocky-Reef group.



In order to permit species in the same habitat group to exhibit different habitat use, the group-specific habitat center is displaced by a small amount for each species. For each species i (i = 1,2,…,23), the habitat-center coordinates for it's group, x_g and y_g are each modified by ε drawn from a discrete uniform distribution on [-2,2].

$$x_i = x_g + \varepsilon_x; \quad \varepsilon_x \sim U(-2,2)$$
$$y_i = y_g + \varepsilon_y; \quad \varepsilon_y \sim U(-2,2) \tag{6}$$

Each individual of the species, j, is then assigned a grid cell by drawing x and y coordinates from a normal distribution around the species population center,

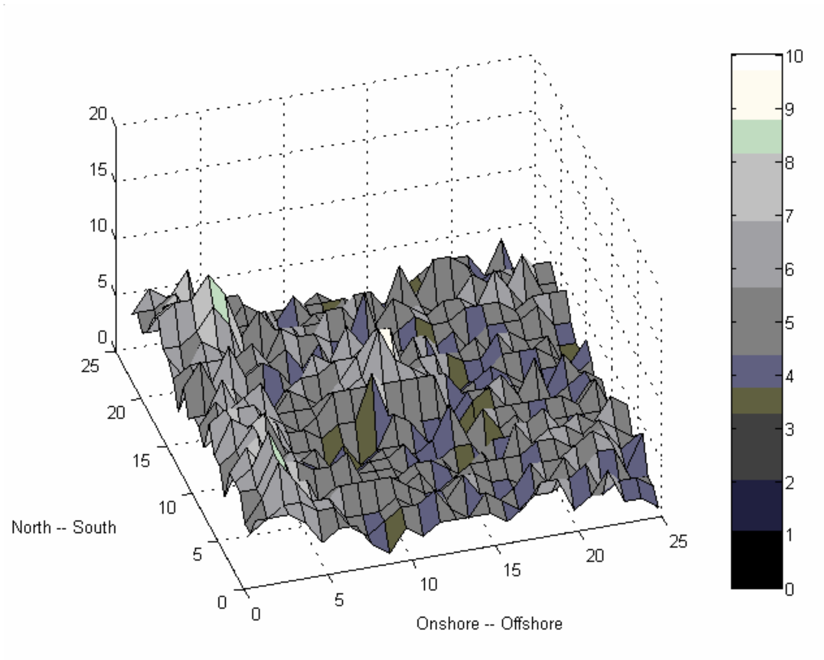
$$x_{ij} \sim N(x_i, \sigma_{xg}^2)$$
$$y_{ij} \sim N(y_i, \sigma_{yg}^2) \tag{7}$$

where the variances are the given group variances (Table 3.1). The same variance is used for all species in the group, and does not change from year to year within a 20-year fishing simulation. Any fish whose coordinates fall outside the edges of the ocean (i.e., coordinates smaller than 1 or larger than 25) is ignored.

Once all the species have been distributed according to their abundance and habitat preference, the list of species present in each cell represents the year's catch in that location. A twenty-year time-series of catch is generated by repeating this procedure.

Note that each species may have a slightly different population center each year because of the random variation introduced. Figure 3.3 depicts a typical distribution of species in the ocean with a small habitat variance. There are more species inshore, which accounts for the peaks on the right side of the graph, and the Rocky Reef is visible towards the center-south of the ocean as a small peak of species distribution.

Figure 3.3. Distribution of species in the simulated ocean, $\sigma^2 = 1$. A typical distribution result for the simulation model. The scale is from low numbers (dark) to high numbers (light.



In order to create an alternative uniform fishing history in which any of the species can occur anywhere within the grid, I draw from a discrete Uniform [0,1] distribution for the occurrence of each species in each grid cell for each year.

3.2.3. Experiment descriptions

3.2.3.1. Variation in overlap of habitats..

In order to examine the differential performance of the logistic regression when habitats are distinct and when they overlap to different degrees, I generated 20-year fishing histories with different values of the global habitat variance, σ^2 = {1,3,5,10,25} and compared the performance of the regression on each catch history, as well as on a uniformly-distributed catch. The target species for this experiment was a low-abundance rocky-reef fish.

3.2.3.2. Variation in target species characteristics.

A catch history with σ^2 = 5 was used for logistic regressions for targets with differing habitat-group characteristics, and different abundances. Targets were a pelagic, superabundant species, and a ubiquitous, average-abundance species.

3.2.3.3. Variation in coverage of the data.

A catch history with σ^2 = 5 was generated, and a percentage of cells were removed from each year's record, so that the regression was run on 100%, 10%, 1% and 0.1% of the same data. The cells removed were randomly re-selected in each year, in order to avoid any geographic bias. This could represent varying degrees of data sparseness in the fishery; in other words, greater or lesser amounts of data collection effort.

3.2.3.4 Variation in regressor characteristics

In order to investigate the dependency of the regression on the predictive variables, I ran the regression first with the full set of regressor species. From those results the two extreme regressors, the most positive and the most negative, were chosen and used for a second regression. I then ran the regression with only one, negatively-associated regressor. Finally, a regression was run with four Ubiquitous species as the only regressors. The same catch history, created with $\sigma^2 = 5$, was used in all cases.

3.2.3.5. Abrupt change in habitat use.

The first 10 years of the catch history reflect an onshore, northern original habitat center for the target species, and in the last 10 years the target species moved to a pelagic habitat in the south. The regression was run on the entire time-series, then a second regression was trained on the first half of the time-series and used for prediction in the second half. I used $\sigma^2 = 3$ to generate the data for this experiment.

3.2.2.6. Populations changing over time.

The populations of three of the indicator species in the same habitat group (Rocky-Reef) as the target increased by 20% per year, while two other species in the group stayed constant. Three indicator species from the Northern and Southern groups declined by 20% per year. Significance of the difference in regression characteristics was determined using the Pearson's χ^2 test for the significance of a proportion (McCall, 2001).

Table 3.3. Summary of experimental conditions. LARR = Low-abundance, Rocky Reef species.

Variable	**Target**	σ^2
Habitat distinctiveness	LARR	σ^2 = 1,3,5,10,25 and uniform distribution
Target characteristics	Pelagic, superabundant Ubiquitous, common	σ^2 = 1,3,5,10,25 and uniform distribution
Amount of data	LARR	$\sigma^2 = 5$
Number and type of regressors	LARR	$\sigma^2 = 5$
Target habitat fidelity	Low-abundance, migratory species	$\sigma^2 = 3$
Changing regressor populations	LARR	$\sigma^2 = 5$

3.2.4. Evaluating regression performance and goodness-of-fit.

In order to evaluate the performance of the regression, I examined the probabilities predicted for the target species, the regression coefficients for different habitat-use groups, the correctness of the predictions, the number of false positive and false negative predictions, and the ratio of the Pearson's χ^2 statistic to the degrees of freedom in the model.

3.2.4.1. Predictions

The fitted values of the regression – the probabilities of the target species in each record – are compared to the records themselves, and the percentage of predictions correct are reported. Histograms of the distributions of probabilities provide a qualitative measure of regression performance, and reflect characteristics of the data. What is desirable in the histogram is an obvious separation of high-probability catches from low-probability catches, with a number of indeterminate records that is small relative to the numbers of positive and negative predictors. If this number is less than 5% of the catch records, it suggests that the other 95% of the records can be predicted with a high level of confidence.

3.2.4.2. Regression coefficients

Regression coefficients for each predictor species are either positive or negative. A positive coefficient indicates that the predictor species occurs with the target in fishing records; a negative coefficient indicates that it does not.

The magnitude of a coefficient reflects the frequency of the co-ocurrence of the predictive species with the target. If they are always found together, the coefficient will be positive and large, if they are found together in half of the records, the coefficient will be close to zero, and if they are infrequently together, the coefficient will be large and negative. If the pattern of co-occurrence with the target is very high or very low, their coefficients may take on extreme values, and in calculating them in the R programming language, they may register as “NaN” or “NA”, indicating computational problems such as the register overflows that occur when a number is too small or too large for the computer to represent.

A strongly predictive regression will produce coefficients of a greater magnitude than the regression coefficients for a weakly predictive regression. In order to evaluate the regression coefficients, I take the mean value of regression coefficients of the habitat group (i.e. the average of the coefficients of all Northern species is the “Northern mean coefficient”).

3.2.4.3. Correct and incorrect predictions

Obviously, it is important that the percentage of the predictions that correctly predict the target be as high as possible, but it is also important to evaluate the incorrect predictions. If the number of false positive predictions is very different than the number of false negative predictions, then the regression is biased either towards over- or under-prediction of the target.

3.2.4.4. Degrees of freedom and the χ^2 statistic: Goodness of fit.

The degrees of freedom (d.f.) in the model are given by the number of data points – in this case, the number of fishing trips – minus the number of parameters β estimated in the model. This is the number of covariate, predictive species plus the intercept β_0, which gives us the probability that the target species was caught when none of the other species were.

$$d.f. = j - (i + 1) \tag{8}$$

The d.f. therefore describes the relationship between the amount of data we are using, and the number of parameters we are estimating. Maximizing this number is a tradeoff between the accuracy of model predictions and our belief in the model's accuracy, since the greater the number of estimated parameters, the more skeptical we should be that they are all close estimates.

Comparing the value of Pearson's χ^2 test statistic with d.f. is typically used to evaluate the goodness-of-fit of the model: how well the model fits the dataset to which we are applying it. This is basically a matter of determining whether or not the data is distributed according to a particular probability distribution. It is called the χ^2 statistic because its distribution is approximately $\chi^2_{(n-k)}$, where n is the number of data (fishing records), and k is the number of parameters estimated in the model (the number of covariate species, plus 1 for the intercept, β_0). Generally, the smaller the value of χ^2, the better the model is predicting the data.

For the binomial data considered here, χ^2 is calculated according to

$$\chi_j = \frac{y_j - \pi_j}{\sqrt{\pi_j(1-\pi_j)}} \tag{9}$$

$$\chi^2 = \sum_j \chi_j^2 \tag{10}$$

Equation 9 describes the Pearson's residuals in the model. The numerator is the deviance of a predicted probability from the actual observation of the target. This deviance is smaller for better predictions. This is where the accuracy of the model comes in. The deviance is weighted by dividing by a term describing the extremity of the prediction, i.e., the divisor in Eqn. 9 is small when π_j is very close to 0 or 1 – when the model is highly predictive – and large for values of π_j near 0.5, where the model exhibits less certainty. This gives extra weighting to the trips for which the model is highly predictive, but incorrect.

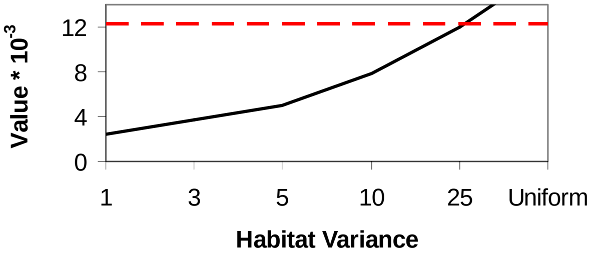
For each individual prediction, χ^2 is exactly 1 for an incorrect prediction with $\pi = 0.5$, larger than 1 for an incorrect prediction with $\pi \neq 0.5$, and smaller than 1 for correct predictions. Then an overall χ^2 = d.f. suggests some combination of incorrect predictions and large uncertainty. As a rule of thumb, χ^2 should be smaller than the model degrees of freedom (McCullagh and Nelder, 1989).

3.3. Results and Discussion

3.3.1. Variation in overlap of habitats.

This experiment illuminates the changes in the predictive ability of the regression under different conditions of habitat overlap. The target for this regression is a low-abundance fish in the Rocky Reef habitat group. First, we see that the logistic model is a good fit for this data as long as habitats are distinct, since the χ^2 statistic is smaller than the 12,477 degrees of freedom until $\sigma^2 = 25$ (Figure 3.4).

Figure 3.4. Model goodness-of-fit for varying habitat overlap. The regression χ^2 value for 12,477 (solid line) is smaller than the d.f. (dashed line) for $\sigma^2 < 25$.



The statistics in Table 3.4 were generated using the critical threshold derived by minimizing the difference between the number of target catches observed and those higher than the threshold. We can see that the number of false positives is similar to the number of false negatives for all cases – the regression is neither underpredicting nor overpredicting the target. The percentage of correct predictions varies inversely with habitat variance; when the habitats are distinct the regression performs well, even though there are fewer locations in which the target occurs. The maximum probability of the target predicted by the model doesn't correlate exactly with the percentage correct, but the lowest values for the maximum probability correlate with the lowest percentage correct. The range of coefficients, the maximum minus the minimum value, is large when the regression is correctly predicting the target, and smaller when it is not.

Table 3.4. Regression characteristics for changing habitat variances. The target is a low-abundance, Rocky Reef species.

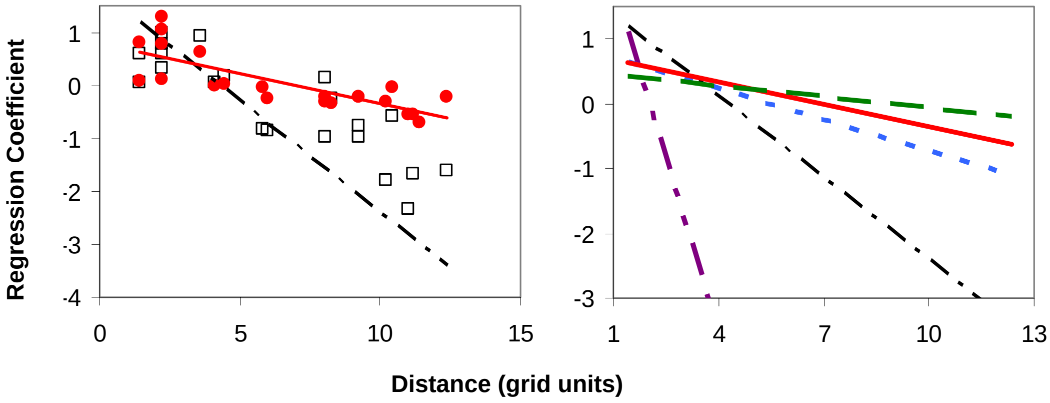
	σ^2 = **1**	σ^2 = **3**	σ^2 = **5**	σ^2 = **10**	σ^2 = **25**	**Uniform**
Observations	545	1101	1539	2194	2900	6260
Significant Regressors	10	13	15	14	8	2
% Correct	95.66	93.87	91.59	86.42	75.48	51.70
% False Negatives	2.18	3.07	4.192	6.84	12.50	23.38
% False Positives	2.15	3.06	4.216	6.74	12.02	24.92
Threshold	0.27	0.33	0.36	0.35	0.33	0.5
Coefficient Range	20.13	14.41	3.46	3.73	3.16	0.14
Maximum Probability	0.86	0.91	0.90	0.89	0.75	0.58

There is an interesting pattern in the number of regression coefficients reported as significant: as the habitat variance increases from 1 to 5, the number of significant regressors increases, but it then decreases again with increasing variance. This is intuitively satisfying. In the first case, few species are found with the target, and in the latter, many are in at least some of the records, leading to decreased significance. The ideally predictive regressor species are those that always or never co-occur with the target.

The coefficients estimated for the predictor species vary with distance between each species' habitat center and that of the target (Fig. 3.5). The coefficient estimated for a species whose habitat center is close to the target's is positive, and the coefficient for a more distant species is negative. The slopes of the trend lines flatten as σ^2 increases (Fig. 3.5, right panel).

Figure 3.5. Relationship of distance to regression coefficients. Left: coefficients and trend lines for $\sigma^2 = 10$ (closed circles, solid line), and $\sigma^2 = 3$ (open squares, broken line).

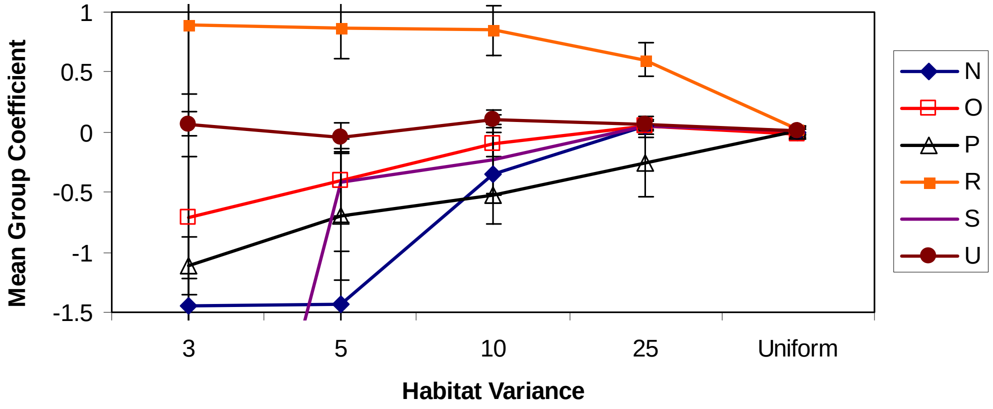
Bottom panel: trend lines for coefficients from regressions with $\sigma^2 = 1$ (– • •), $\sigma^2 = 3$ (– • –), $\sigma^2 = 5$ (•••), $\sigma^2 = 10$ (solid), and $\sigma^2 = 25$ (– –).



The regression coefficients are positive as expected for co-occurring Rocky Reef fish and negative for the groups that rarely co-occur with the target, such as the Offshore and Northern species (Figure 3.6). They retain their original sign as σ^2 increases, but decrease in magnitude.

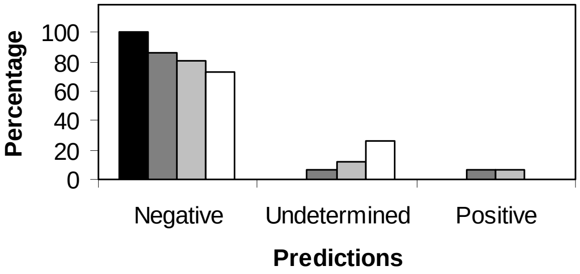
The regression coefficients are positive as expected for co-occurring Rocky Reef fish and negative for the groups that rarely co-occur with the target, such as the Offshore and Northern species (Figure 3.6). They retain their original sign as σ^2 increases, but decrease in magnitude. The coefficients for the Ubiquitous group are quite small, indicating low predictive power. For the case of the uniform distribution, all coefficients are quite small.

Figure 3.6. Regression coefficients for varying habitat overlap. These are averages for predictor species, grouped by habitat. Rocky Reef fish are represented by filled squares, the Ubiquitous group by open circles, Onshore fish by open squares, Southern species are the unmarked line, Northern species are represented by filled diamonds, and Pelagic species by open triangles. Error bars are one



The fitted probabilities can be interpreted as predicting a positive (presence) or negative (absence) catch of the target species. I define a negative catch as having probability $\pi < .3$, a positive catch as those with $\pi \geq .7$, and the rest as catches in which the presence of the target cannot be determined. This choice of cutoff values permits us to see the very few positive predictions of the low-abundance target species. The ideal distribution of predictions would have no "undetermined" catch. When the probabilities are viewed in this way, we see that there is a trend to greater indeterminacy as σ^2 increases (Figure 3.7). At $\sigma^2 = 25$, the regression fails to detect the target, which is present in 2900 records (Table 3.4) and 26 % of the catch records are in the "undetermined" category. For the case in which all species are uniformly distributed, the predictions are 100% negative, in spite of the fact that the target occurs in 50% of the catch records (Table 3.4).

Figure 3.7. Predictions for varying habitat overlap. The percentage of negative catches ($\pi < .3$), positive catches ($\pi \geq .7$), and undetermined catches ($.3 \leq \pi \leq .7$) of the target species predicted by the regression for the uniformly distributed catch (black) and for $\sigma^2 = 5,10$, and 25 (dark grey, light grey and white bars, respectively). The target is a low-abundance, Rocky Reef fish. At $\sigma^2 = 25$, the regression fails to detect the target.

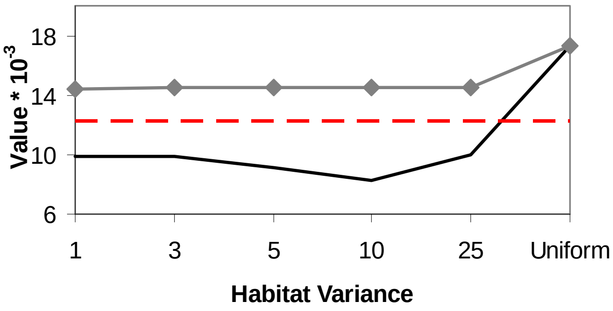


We can see that only a small percentage of the data can be attributed to effort to catch the target species. A CPUE index generated using all of the data would seriously overestimate the abundance of the species

3.3.2. Variation in target species characteristics

When used with different targets, the regression behaves somewhat differently than in the prediction of a Rocky-Reef fish. Regressions were run in the same dataset as for the Rocky-Reef target. The two targets used are a Pelagic, superabundant species, and a Ubiquitous, average-abundance species. The model goodness-of-fit measure demonstrates the appropriateness of logistic regression for species with these characteristics (Figure 8). The regression deviance for the Pelagic target is less than χ^2 for $\sigma^2 < 25$. The deviance for the Ubiquitous target is $>$ χ^2 for all cases, indicating that the logistic regression fails to model the data in this case.

Figure 3.8. Regression goodness-of-fit for varying target characteristics. The line marked with diamonds is the deviance for regressions on a Ubiquitous target, the solid line is the deviance for the Pelagic target, and the dashed line is χ^2 for 12477 d.f.



All incorrect predictions for the Pelagic target are false positives for cases of $\sigma^2 < 25$ (Table 3.5), because almost all of the regression coefficients are negative (Figure 3.9, top panel), but the target is very abundant, so the regression overpredicts the target. Note that the critical threshold value for these cases is 0. When habitat use broadens, the regressors are less strongly negative, and false negative predictions occur. For the Ubiquitous target, the incorrect predictions are evenly divided between false positive and false negative predictions, except when $\sigma^2 = 10$, where the target is overpredicted (Table 3.6). This is also the regression with the greatest percentage of correct predictions (91%). In general, the percentage of correct predictions for each of these targets is less than for the Rocky Reef fish (Table 3.4).

The number of significant coefficients for the Pelagic target is slightly larger than for the Rocky Reef fish, since they co-occur with more species, but the number of significant regressors for the Ubiquitous species is much lower, because they tend to co-occur with all species.

The coefficient range for the Pelagic target is similar to that for the Rocky Reef target, with the same trend towards smaller values as the percentage of correct predictions falls. The maximum probability predicted also shows a downward trend, but in this case the maximum stays quite high, probably due to the abundance of the target. For the Ubiquitous target, the coefficient range is < 1 throughout, dropping to 0.15 in the simulation with uniformly distributed species (Table 3.6). The maximum probabilities are low, but highest in the case of the uniform species distribution.

Table 3.5. Regression characteristics for a superabundant, Pelagic target.

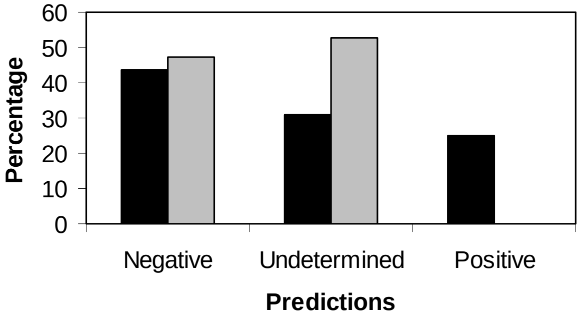
	σ^2 = 1	σ^2 = 3	σ^2 = 5	σ^2 = 10	σ^2 = 25	Uniform
Observations	10763	10923	10901	10875	10888	6303
Significant Regressors	12	14	13	15	8	1
% Correct	86.104	87.384	87.208	87	76.888	58.536
% False Negatives	0	0	0	0	12.168	23.192
% False Positives	13.896	12.616	12.792	13	10.944	18.272
Threshold	0	0	0	0	0.05	0.5
Coefficient Range	20.14	19.28	21.24	6.95	3.75	0.12
Maximum Probability	0.93	0.96	0.99	0.99	0.98	0.55

3.6. Regression characteristics for a Ubiquitous target. For Ubiquitous species, σ^2 is fixed, however changing σ^2 affects the non-Ubiquitous regressor species.

	σ^2 = 1	σ^2 = 3	σ^2 = 5	σ^2 = 10	σ^2 = 25	Uniform
Observations	10763	10923	10901	10875	10888	6303
Significant Regressors	6	3	1	4	5	2
% Correct	76.4	77.296	86.16	90.936	77.44	55.52
% False Negatives	12.2	12.152	7.248	2.896	13.664	22
% False Positives	11.4	10.552	6.592	6.168	8.896	22.48
Threshold	0.24	0.24	0.24	0.24	0.24	0.5
Coefficient Range	0.47	0.57	0.346	0.49	0.41	0.15
Maximum Probability	0.41	0.37	0.40	0.41	0.40	0.57

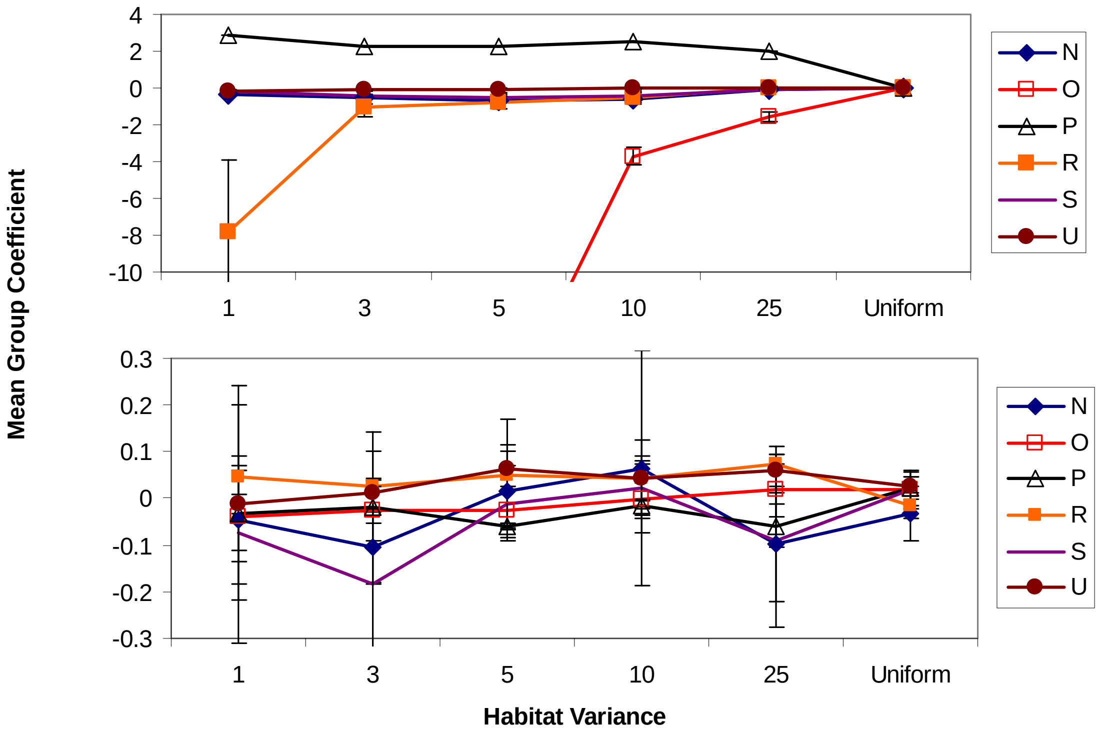
If we again put aside the critical threshold criterion and consider the model predictions as positive, negative or indeterminate, we see that for the majority of the catch records, the presence or absence of the Ubiquitous target cannot be determined, while the Pelagic target is only predicted in 25% of the catch (Fig. 3.9). Again, the CPUE measures generated from these data would be significantly different than a CPUE generated from the full catch history.

Figure 3.9. Predictions for varying target characteristics. The percentage of negative catches ($\pi < .3$), positive catches ($\pi \geq .7$), and undetermined catches ($.3 \leq \pi \leq .7$) of the target predicted by the regression for the $\sigma^2 = 5$, for the Pelagic target (black) and the Ubiquitous target (grey).



Finally, the pattern of mean group coefficients estimated in the regressions is completely different for these two targets (Figure 3.10). The regression coefficients for the Pelagic species (Figure 3.10, top panel) show a pattern similar to that of the Rocky Reef regressions (Figure 3.6). The habitat groups have either negative or positive trends, with the other Pelagic species the positive regressor. The magnitude of the regressors is a little larger than the for the Rocky Reef target. In contrast, the coefficients for the Ubiquitous target show no discernable trend, change sign, and are quite small in every case (Figure 3.10, bottom panel).

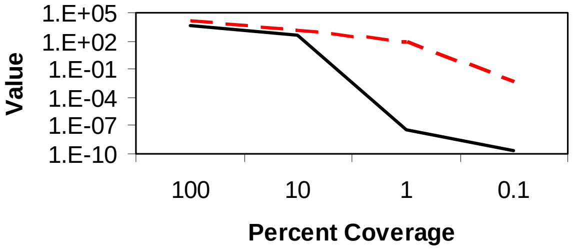
Figure 3.10. Regression coefficients for varying target characteristics. These are averages for predictor species, grouped by habitat. The top panel shows the coefficients from a regression with a Pelagic target, the bottom panel shows coefficients for a Ubiquitous target. Rocky Reef fish are represented by filled squares, the Ubiquitous group by open circles, Onshore fish by open squares, Southern species are the unmarked line, Northern species are represented by filled diamonds, and Pelagic species by open triangles. Error bars are one standard deviation. Note that there is only one Pelagic regressor.



3.3.3. Variation in coverage of the catch records

For this experiment I returned to using a low-abundance Rocky-Reef species as the target species, and using only the data generated for $\sigma^2 = 5$, I changed the number of records in the catch history. As the number of records declines, the fit of the regression appears to improve (Figure 3.11). However, the only way for the χ^2 statistic to be smaller than one is for all incorrect predictions to be false negatives, with a critical threshold much smaller than one.

Figure 3.11. Model goodness-of-fit for varying data coverage. The regression χ^2 value (solid line) is smaller than the d.f. (dashed line) at all levels of data coverage. The target is a low-abundance Rocky Reef species, and $\sigma^2 = 5$.



The number of false positive and false negative predictions is similar, and the percentage of correct predictions increases to 100% (Table 3.7). This may be an artifact of the reduction in data: when there are more records the possibility of anomalous catch records increases. If we look at the number of regressors significant in each case, the number declines, and is 0 at 1% coverage. The 9 regressors listed for 0.1% coverage are those for which the regression was unable to calculate a value, and returned “NA”, which signals computational problems related to register over- or under-flow; in other words, the values were too small or too large for the computer’s internal representation. No regressors were significant in the 0.1% case.

Table 3.7. Regression characteristics for varying data coverage. At 0.1% coverage no regressors are significant, and 9 have the value “NA”, which signals computational problems, meaning that the regression was unable to calculate a coefficient value. The target is a low-abundance Rocky Reef species, and $\sigma^2 = 5$.

	Full dataset	**10%**	**1%**	**0.1%**
Observations	1539	166	16	3
Significant Regressors	15	7	0	0 (9 NA)
% Correct	91.9	92	100	100
% False Negatives	4.056	3.96	0	0
% False Positives	4.016	4.1	0	0
Threshold	0.36	0.36	0.01	0.01
Coefficient Range	16.57	17.57	779.45	357.92
Maximum Probability	0.92	0.96	1	1

Coefficients for the least-data regressions (Figure 3.13, Table 3.7) took on very large values, or were anomalously signed (0.1%). The target predictions in these regressions were ideal, however; either positive or negative, and 100% correct (Table 3.7, Figure 3.12). Maximum probabilities for these cases was 1. It is very difficult to

Figure 3.12. Predictions for varying data coverage. The percentage of negative catches ($\pi < .3$), positive catches ($\pi \geq .7$), and undetermined catches ($.3 \leq \pi \leq .7$) of the target species predicted by the regression for the $\sigma^2 = 5$, for the full dataset (black), 10% (dark grey) 1% (light grey) and 0.1% (white) of the data. The target is a low-abundance Rocky Reef species, and $\sigma^2 = 5$.

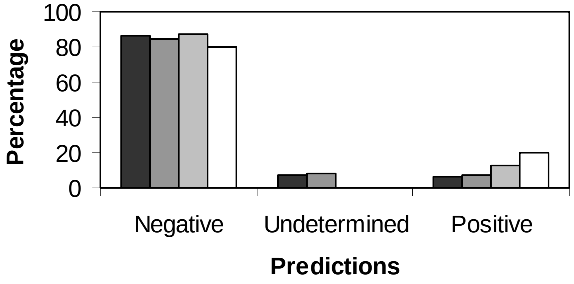
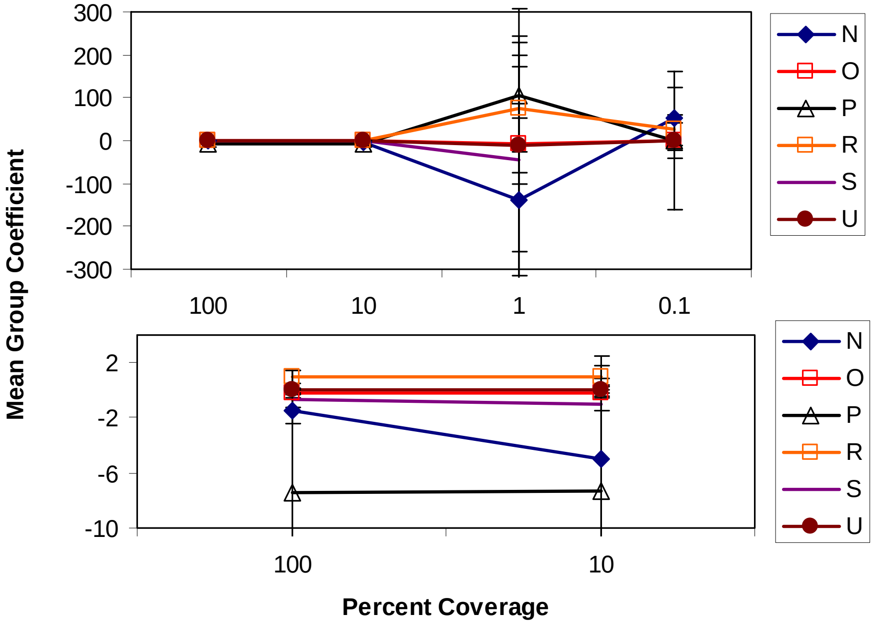


Figure 3.13. Regression coefficients for varying data coverage. These are averages for predictor species, grouped by habitat. Rocky Reef fish are represented by filled squares, the Ubiquitous group by open circles, Onshore fish by open squares, Southern species are the unmarked line, Northern species are represented by filled diamonds, and Pelagic species by open triangles. Error bars are one standard deviation. The target is a low-abundance Rocky Reef species, and $\sigma^2 = 5$. The bottom panel is scaled to show the values at 100% and 10% coverage.

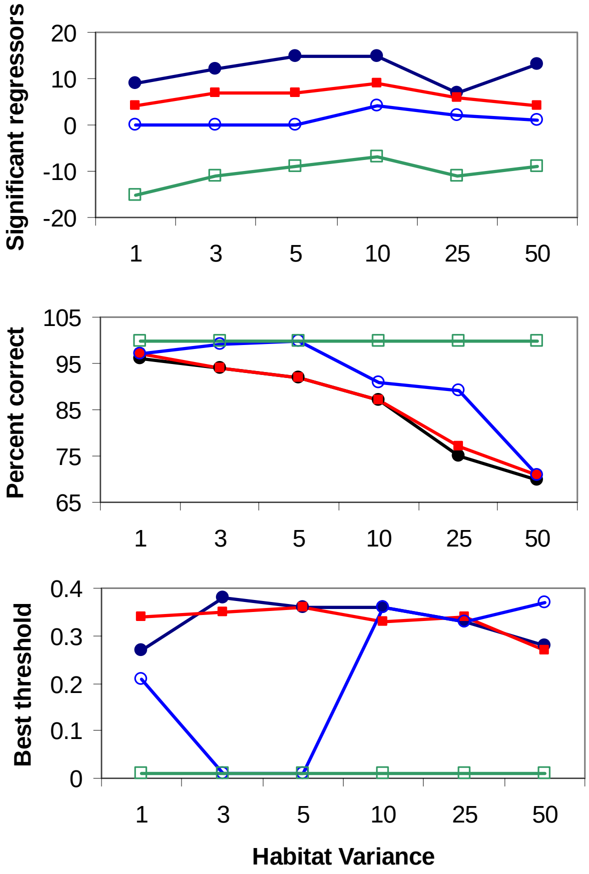


interpret results like those in the 1% case, which seems to be working perfectly by most measures, but has an unusual range of coefficients.

The interacting effects of the variance and the size of the dataset is quite interesting. The regression performs similarly on the 100% dataset and the 10% dataset, although slightly better in the former case. The regression on 0.1% of the data fails consistently. The failure markers are the size of the best threshold for data discrimination, the lack of any significant regressors, and the failure of the regression to provide coefficient estimates for many of the regressors. Predicting 100% of the catches correctly may mean that the model is overfitting the data in these cases, which would occur if the number of parameters it is estimating is greater than the number of observations of the target in which the co-occurrence of predictor species differs. The regression did not generate any intermediate probabilities, at any variance.

If we use the size of the best threshold value and the failure to predict intermediate values as indicators of regression failure, then at 1% data coverage there are two cases in which the regression fails. These occur at $\sigma^2 = 3$ and $\sigma^2 = 5$. At $\sigma^2 = 5$, the regression predicts no intermediate probabilities, and at $\sigma^2 = 3$ less than 2% of the probabilities generated are intermediate. Again, this signals overfitting the model, and may mean that the data provide few patterns of species co-occurrence with the target.

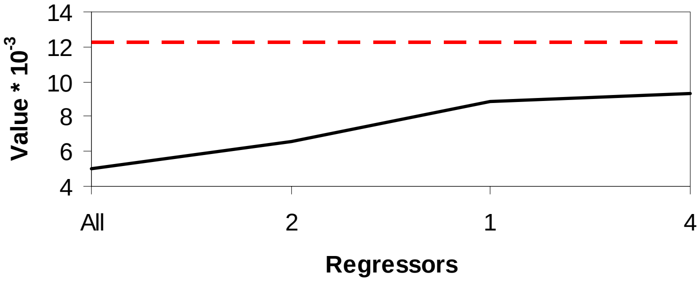
Figure 3.14. Regression performance across a range of habitat variances (σ^2), and dataset sizes. Top panel: Number of significant regressors. Negative values reflect coefficients the regression was unable to estimate (NA). Closed circles represent the 100% dataset, closed squares are the 10% dataset, open circles are the 1% dataset, and open squares are the 0.1% dataset. Middle panel: percentage of correct predictions. Bottom panel: the best threshold for minimizing the difference between observed and predicted target catches.



3.3.4. Variation in regressor characteristics

The target species for this regression is a low-abundance Rocky-Reef fish, in data generated with $\sigma^2 = 5$. I first ran the regression with all non-target species as regressors. Using the correlation coefficients, I chose the most positive and most negative significant predictors, and reran the regression with just those two. I then ran the regression using only the single negative predictor species. I compare these to a regression run with four Ubiquitous species as regressors. For all cases, the regression fit the data well (Figure 3.15), although the trend is towards greater χ^2 with fewer regressors, and is greatest for the case of 4 Ubiquitous regressors.

Figure 3.15. Regression goodness-of-fit for varying regressor characteristics. The χ^2 statistic (solid) is less than d.f. (dashed) for regressions using all regressors, 2, 1, and 4 Ubiquitous regressors.



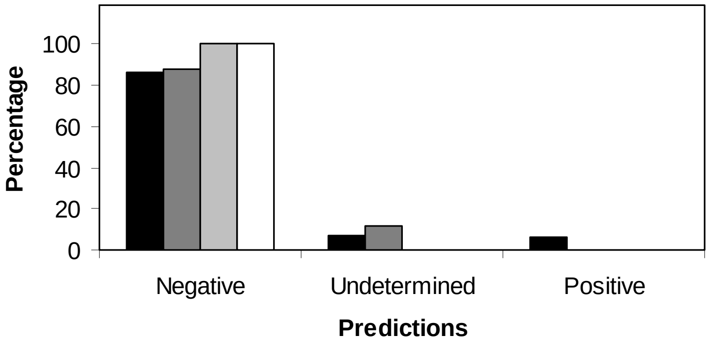
The number of false positives equals the number of false negatives for all cases except the regression run with 1 negative predictive species, which underpredicts the target (Table 3.8). The regression run with four Ubiquitous regressors produced the lowest percentage of correct predictions, and found 2 regressors to be significant, though with coefficients < 0.21. Both that regression and the 1-regressor case generated extremely low maximum probabilities (Table 3.8). The coefficient range is small for the case of the four Ubiquitous regressors, and of course is 0 when there is only one regressor.

Table 3.8. Regression characteristics for varying regressors.

	All regressors	**2**	**1**	**4 Ubiquitous regressors**
Observations	1539	1539	1539	1539
Significant Regressors	13	2	1	2
% Correct	91.59	90.25	87.69	78.86
% False Negatives	4.192	4.87	12.31	10.38
% False Positives	4.216	4.88	0	10.76
Threshold	0.36	0.07	0.16	0.14
Coefficient Range	3.14	5.10	0	0.15
Maximum Probablility	0.90	0.61	0.15	0.17

Only the regression using all regressors predicted the target species at all (Figure 3.16). The regression with 2 regressors was also able to predict a small number of undetermined catches, while the other two regressions simply return 100% negative predictions.

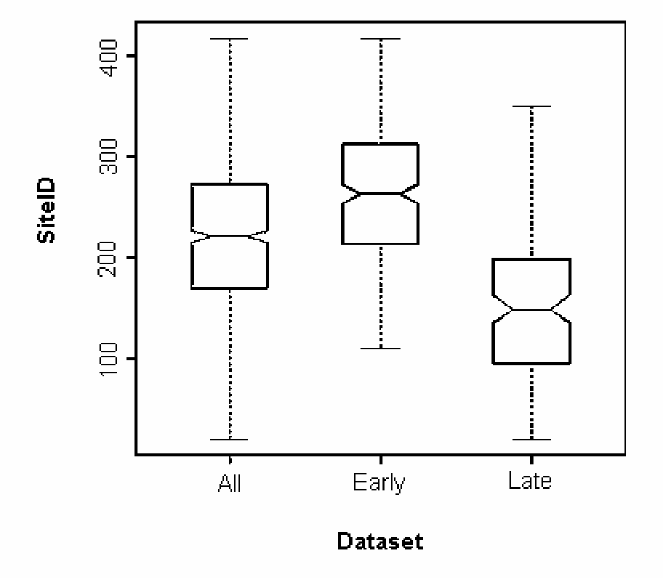
Figure 3.16. Predictions for varying regressor characteristics. The percentage of negative catches ($\pi < .3$), positive catches ($\pi \geq .7$), and undetermined catches ($.3 \leq \pi \leq .7$) of the target species predicted by the regression for $\sigma^2 = 5$, for all regressors (black column), 2 regressors (dark grey), 1 regressor (light grey) and for the regression run with 4 Ubiquitous regressors (white).



3.3.3. Abrupt Change in habitat use.

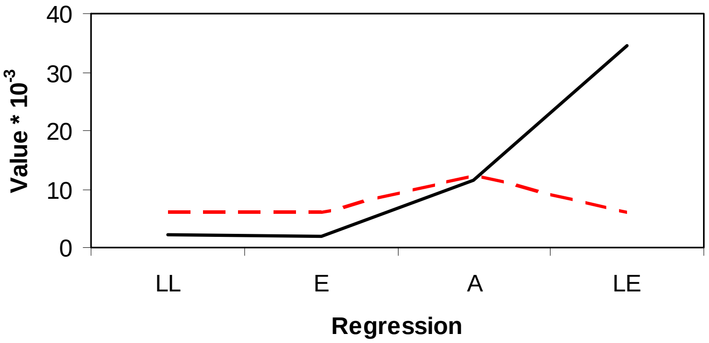
For this experiment, I generated data with $\sigma^2 = 3$, for a migratory target species that moved from the onshore, northern corner of the ocean to the offshore, southern corner abruptly after ten years. I ran the regression on the first ten years of the data, the “Early” data set, and then used the coefficients estimated in the Early regression for prediction in the last ten years of the data, the “Late” dataset. I compare the results with those generated by running the regression on each dataset, Early, Late and the Full dataset. Figure 3.17 shows that there is little overlap between the two habitats.

Figure 3.17. Sites where the target species occurred in each dataset as habitat use changed. Sites are identified numerically by grid location.



The goodness-of-fit criterion (χ^2 > regression deviance) is met for the regressions run in all three datasets; in the Full dataset the values are quite close (Figure 3.18), suggesting that the logistic regression provides a poorer fit to the data. However, deviance for the Late data predicted by the model built on the Early data clearly shows the mis-fit of the model. Numbers of false positives and false negatives are similar for all cases (Table 3.9). The percentage of predictions correct using the critical threshold criterion is low in both the Full dataset and the Late data predicted by the Early model.

Figure 3.18. Model goodness-of-fit for changes in habitat use. χ^2 (solid) and d.f. (dashed) for the four regressions. From left to right, the Late regression (LL), the Early regression (E), the regression on all the data (A) and the Late data predicted by the model built for the Early dataset (LE). The target is a low-abundance, migratory species, and $\sigma^2 = 3$.

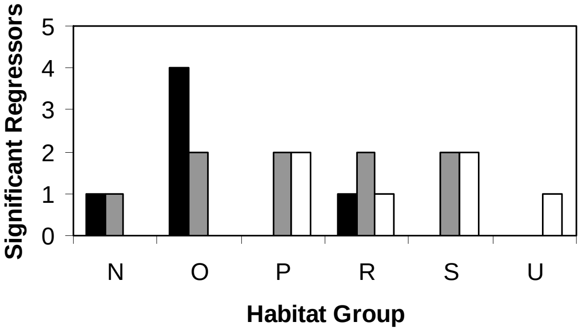


3.9. Regression Characteristics for a Migratory target. The target species is a low-abundance species that resides in the onshore, northern corner of the simulated ocean for the first ten years, and the southern, pelagic corner for the last ten.

	Early	Late	Full	Late/Early
Observations	481	520	1001	520
% Correct	93.12	92.85	86.66	84.05
% False negatives	3.47	3.58	6.5	8.32
% False positives	3.41	3.57	6.83	7.63
Coefficient Range	20.47	21.43	2.70	20.47
Maximum Probability	0.81	0.82	0.24	0.81

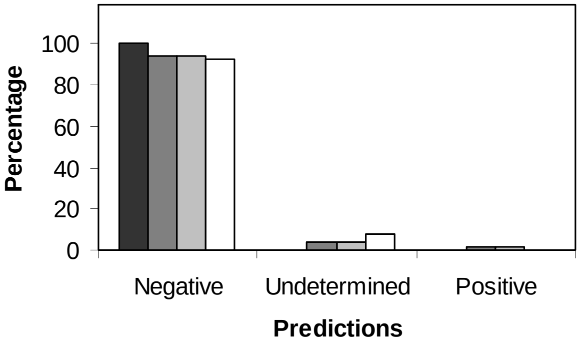
The coefficients significant for each regression reflect the movement of the target species (Figure 3.19). For the Late data, the Northern and Onshore species are absent from the list of significant regressors. For the Early data, the Pelagic and Southern regressors are not significant. In the Full datset, significant regressors come from all of these groups. The only group in common among all three is the Rocky Reef group.

Figure 3.19. Significant regressors for changes in habitat use. The Early dataset is in black, the Late dataset is represented by the white bars, and the Full dataset is in grey. The target is a low-abundance, migratory species, and $\sigma^2 = 3$.



Interestingly, the only regressions to predict positive catch of the target species are the Early data, and the Late data predicted using the Early regression (Figure 3.20). The regression in the Full dataset predicts no catch.

Figure 3.20. Predictions for changes in habitat use. The regression in the Full dataset (black), Early dataset (dark grey), the Late data predicted by the Early regression (light grey) and the Late dataset (white). The target is a low-abundance, migratory species, and $\sigma^2 = 3$.



3.3.6. Populations changing over time.

This experiment investigates the response of the regression to changes in regressor populations. The target species is again a low-abundance, Rocky Reef species, in a dataset generated with σ^2 = 5. Three co-occurring species declined in numbers by 20% each year, while 2 species that do not co-occur with the target increase by 20% each year for the 20 years of data. I compare regression results in this data with results in data with static populations.

3.10. Regression Characteristics for changing and stable regressor populations, σ^2 = 5, when the target is a low-abundance Rocky Reef species.

	Changing	**Stable**
Observations	1546	1539
Significant Regressors	16	15
% Correct	91.2	91.6
% False negatives	4.38	4.19
% False Positives	4.41	4.22
Threshold	0.31	0.36
Deviance	5467	5023
χ^2	12218	12218
Coefficient Range	2.57	3.45
Maximum Probability	0.93	0.90

The regression seems very robust to changes in population size. This is to be expected, since we are working with presence/absence data. The regression deviance for the stable populations is slightly lower (better) than for the changing populations, and both are much smaller than χ^2 (Table 3.10), indicating a good fit between the model and the data in each case. None of the regression characteristics are statistically different. The regression coefficients have the same sign and similar magnitudes (Figure 3.21). Finally, we can see that the regressions generate nearly identical predictions (Figure 3.22).

Figure 3.21. Mean Regression Coefficients for population changes. Regression coefficients averaged by habitat group for regressions with stable (grey) and changing (white) populations. The target is a low-abundance Rocky Reef species, and $\sigma^2 = 5$. Error bars represent one standard deviation.

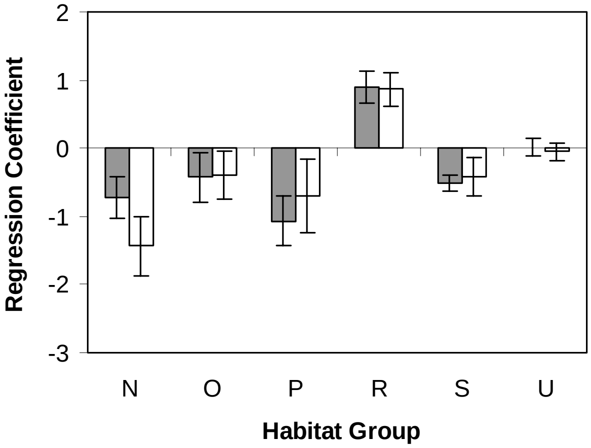
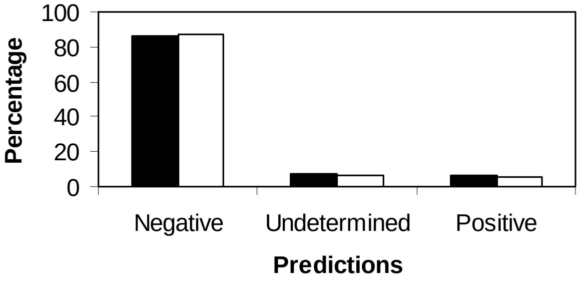


Figure 3.22. Predictions for population changes. The percentage of negative catches ($\pi < .3$), positive catches ($\pi \geq .7$), and undetermined catches ($.3 \leq \pi \leq .7$) of the target species predicted by the regression for stable (black) and changing (white) regressor populations. The target is a low-abundance migratory species, and $\sigma^2 = 5$.



3.4 Rules-of-thumb for regression diagnostics

The results for these experiments lend themselves to developing rules-of-thumb for interpreting regression diagnostics. First, note that regression deviance < X^2 indicates only that the model used is a good model for the data, but doesn't correlate well with model performance.

Data:

- The data suited to this type of analysis must be presence/absence data, and should present more than a few regressor species. Species with positive and negative co-ocurrence with the target mush be represented. For this study, 22 regressors were used in most experiments.
- A classic problem in regressions of this type is colinearity among the regressors. In a logistic regression, this occurs when two regressors have the same pattern of occurrence. Data should be checked for colinearity (McCullagh and Nelder, 1989), however many software packages such as R will report colinearity in the data.
- Related to colinearity is the problem that may occur if there are few species that co-occur with the target species, or if there are few patterns of co-occurrence. In this case, the model may overfit the data, and reducing the number of parameters used may improve model performance.

- Habitats within the fishery grounds should be in an appropriate size range. The variance, or standard deviation (σ^2) in these experiments defines their sizes. Because fish locations are drawn from a normal distribution, 68% of the fish will be within one standard deviation from the habitat center. For example, at $\sigma^2 = 10$, most Rocky Reef fish will be found in an area 20 grid units on a side within the ocean, which is 25 units square. This is a ratio of 400:625, or roughly 2/3 of the fishing ground, and the regression performed well under these conditions. However, when $\sigma^2 = 25$, the habitat covered the entire fishing ground, and the regression failed to predict the catch.

Probabilities:

- The highest probability predicted by the model is a measure of how much faith we can put in the model's positive predictions. The lowest probability predicted is a measure of the certainty we have in predictions of absence of the target.
- The number of indeterminate probabilities is a measure of how well the model is able to distinguish the target's habitat.
- A lack of intermediate probabilities is a signal that either the regression is optimally segregating positive and negative predictions, or that the data may be too sparse, leading to overfitting the model. Other indicators of this type of failure are low numbers of significant regressors, and extreme values of regression coefficients.

Regressors:

- As the number of significant regressors increases, the accuracy of predictions increases, however the regression may still perform well with few regressors.
- A preponderance of either positive or negative regressors will tend to bias the model towards over- or under-predicting the target.
- The range of regression coefficients should be greater than one, but not extremely large. For this data, poor regressions had coefficient ranges either less than one, or greater than a hundred.
- Species that exhibit low habitat fidelity are poor regressors. Coefficients for these species will be close to zero.

Predictions:

- A regression that predicts all-negatives or all-positives is failing to predict the less-common case – the occurrence of a rare species or absence of an abundant one.
- The numbers of false positive and false negative predictions should be roughly the same. A predominance of either one indicates bias, and suggests that the predictive species are largely ones that cooccur with the target, or species that are seldom found with it. Removal of some regressors may balance the predictions.

Target characteristics:

- Species that exhibit low habitat fidelity or are migratory under different environmental regimes are poor targets.

Finally, when the data are collected over a periods of time long enough for habitat associations to change, or when habitat use is not well understood, it is important to examine regression performance in temporal subsets of the data.

3.5. Conclusions

It is very important to be able to relate the results of the method to a biological situation. First, the physical habitat must be one that lends itself to this type of analysis. The system should be one in which different species adhere to different habitats, and the target species in particular must exhibit a fair amount of habitat fidelity. Analyses should be confined to periods in which these conditions hold true. Second, when we understand the response of regression metrics to physical parameters, then even if the regression fails it can provide information about the system that might otherwise be difficult to observe. I have highlighted here several effective performance measures, and described their behavior in the light of different plausible ecological conditions.

The species-based regression is quite robust except when the main assumption is violated: that species remain true to habitat. Thus the regression fails when habitats are indistinct, when either the target species or the regressor species use all habitats equally, and when the species change habitat use patterns, which is typical of many marine species under changing environmental regimes.

Regression deviance < X^2 indicates only that the model used is a good model for the data, and doesn't correlate well with model performance. While this is useful for catching extreme misfits, model performance is best judged using the criteria listed in the previous section, based on the probabilities and regression coefficients generated by the regression.

Regression coefficients should reflect biological facts in the fishery: species found rarely with the target species should have negative coefficients, and species often found with it should have positive ones. Ideally, one would prefer a variety of significant regressors, both positive and negative. Coefficients falling in the range [1..25] seem to be associated with good regressions, those much larger or smaller signify problems.

Probabilities should not be narrowly distributed, nor should they all be close to 0.5. Although it is intuitively satisfying to desire a probability distribution that perfectly distinguishes presence and absence of the target, this may indicate that the model is overfitting the data. False predictions should be examined for hints of bias in the regression.

Although the basic assumptions underlying the use of CPUE as an abundance index are often questioned (e.g., Peterman and Steer, 1981; Swain and Sinclair, 1994; Harley, Myers, and Dunn, 2005), it continues to be widely used. Other methods of subsetting exist, but have their own shortcomings, including dependence on assumptions that may not hold.

The most commonly used method of subsetting data historically has been to include only those catches that contain the target; however this misses the empty catches in-habitat, and it is changes in the proportion of these that reflect abundance changes. A common work-around for this has been to assume that the catch represents a constant proportion of the population, which doesn't take changes in fishing trends into account. However, fishing practices do change in ways that impact individual species (see, for example, Stephens and MacCall, 2004). Another historically employed method for subsetting recreational data relies on what the captain or fisher reports as the "target" species, but often this is the species most-frequently caught, designated in retrospect as the target.

In subsetting commercial data, some have suggested using only records from fishers whose data implies a reliable targeting of the species. However, the criteria for defining what this means is evolved ad-hoc (Taylor, 2003; Punt *et al.*, 2001b). Commercial data are often subset based on vessel-type (Jimenez, *et al.*, 2004); Bishop, et al offer an interesting comparison of several of these in subsetting data in the Australian Northern Prawn fishing industry, using vessel and gear types (which are assumed constant in the recreational fishery) (Bishop, *et al.*, 2004).

Other GLM-based methods have involved the use of zero-inflated distributions, such as the zero-inflated Poisson (ZIP) that rely on a mixed probability distribution that accounts for zero-catches based on some presumption of a systematic reason for zeros, one that can be modeled with a known probability distribution (Maunder and Punt, 2004). Related to this is the concept of the delta-GLM, modeling the probability of obtaining a zero catch binomially, and the catch-rate for positive catches (of numbers or biomass of fish) separately, often using a log-normal distribution (Stefansson, 1996). These are sometimes called “hurdle models”. Finally, there are computationally intensive approaches such as Bayesian hierarchical analysis (another mixed-probability technique) (c.f. Argaez, *et al.*, 2005).

These methods may not be well-suited to recreational datasets, either because of underlying assumptions or because they may be difficult to implement properly. For these reasons, the logistic regression will be the analytical tool of choice for many fisheries researchers. This study will provide those analysts with added insight into the use of this straightforward method in the very important problem of filtering effort data in multispecies fisheries.

Acknowledgements

This work was supported in part by the Center for Stock Assessment Research, a partnership between The Southwest Fisheries Science Center Santa Cruz Laboratory and UC Santa Cruz. I am grateful to Chris Edwards, Alec MacCall, Herbie Lee, Anand Patil, Athanasios Kottas and expecially Marc Mangel for thoughtful discussion and valuable insights.

Concluding Remarks

The purpose of this thesis is to address problems of estimation and projection in two very different fisheries, the New England salmon fishery, and the California recreational fishery. A variety of techniques have been developed to assess marine populations, but many tend to be inadequate for the analysis of multiple-species systems, or systems in which the animals in question develop through different life history stages (Quinn and Deriso, 1999, Schnute, 2003). These two fisheries exemplify those complex cases. In the first case, population status is known, but the wild salmon are under increased risk of extinction due to aquaculture. Here it is the risk to the fishery I am attempting to characterize, assessing the scope and magnitude of a variety of potential risks and evaluating outcomes for the fishery. In the second case, population information for all species in the fishery is confounded, and the problem is one of isolating the data pertinent to a single species in order to assess its state, and create management strategies to insure its continued existence as an economically viable resource. These problems call for complementary approaches. I employed a dynamic population simulation for risk analysis in the salmon fishery, and a bio-statistical analysis for the California recreational fishery.

Inherent in addressing conservation ecology issues in fisheries is the additional dimension of communicating the scientific approach, its results, and the limits and uncertainties inherent in these approaches. There are several audiences to target: the community of scientists developing methods for such analyses, the community of scientists who are users or consumers of such methods research, and the managers and stakeholders of the fishery. Much science proceeds under the presumption that communication with at least one of these audiences is the purview of others, and put such considerations in the background. My work has attempted to keep issues of communication in the foreground.

In Chapter 1, I addressed risk analysis in the Atlantic salmon fishery, showing that wild salmon are threatened in a variety of different ways by aquaculture. The age-structured population model demonstrates that although the risks to the anadromous fish are due to different mechanisms at different stages in their development, salmon are in fact not more vulnerable in freshwater than in the ocean. In freshwater, extinction can occur because of genetic introgression and competitive effects. In the ocean/estuary, extinction can be caused by the existence of the aquaculture facility, the predators it attracts, and especially the potential for disease magnified in the farmed fish population and transferred to wild fish. Management of these issues is paramount to the continued existence of wild populations.   The problems in saltwater can be addressed by more effective regulation of aquaculture facilities; restricting their proximity to the streams in which wild salmon persist, and aggressively treating disease outbreaks. Freshwater threats involve the reproduction of escaped aquaculture fish, and these are best addressed by the use of sterile (triploid) fish in culture.

Another important result of the model is what became obvious during its construction. Many empirical studies are performed on salmon and on aquaculture facilities, yet no studies to date characterize the magnitude of competitive pressure or the potential for interbreeding between wild and farmed fish in ways that could inform a modeling effort. Further, there is much straightforward work to be done in developing safety standards for aquaculture facilities with respect to disease. Water monitoring downstream of aquaculture facilities could probe for concentrations of disease organisms in the environment, but metrics such as LD_{50} for many of these diseases (e.g. vibriosis, furunculosis and the like) are not known.

The most important audience for the results of this study are the fishery managers, aquaculturists, and stakeholders. These are the people who need to be well informed about the risks to the fishery and options for aquaculture operations, but they are not an audience for a modeling paper such as Chapter 1 of this document. For this reason, I imbedded the model in a menu-driven interface, making it possible for non-modelers to investigate risks for themselves. Much education research shows that lessons learned by hands-on exploration are better understood and retained, and I hope that this risk assessment tool can contribute substantially to opening a window into the assessment process.

In Chapter 2 I focused on a method for improving estimates of abundance in a mixed-species fishery. The California recreational fishery targets over 50 species, and many more are caught incidentally. Many of the species usually caught by recreational fishers are suffering population declines, and others may follow as fishing effort shifts away from restricted species to those without management plans. Abundance estimates for fish populations are typically constructed from catch and effort data, but few records of recreational catch represent effort spent on one target species. Recognizing that certain habitats are preferred by certain species, I used the species present in a catch in a logistic regression to infer the habitat that was fished. This allowed me to restrict estimates of fishing effort for a species to the records of fishing that occurred in its habitat.

This method proved suitable for analysis of the California recreational fishery data. It provided insight into population trends in the three datasets analyzed, as well as information about trends in fishing effort. Because the method uses only presence/absence catch data, it is robust to changes in catch produced by changing regulations, such as bag limits. The value of such a method lies in its reproducibility. Many stock assessment techniques depend on arbitrary or *ad hoc* decision-making by analysts, so that the same data may be used to produce divergent assessments. The multispecies method provides a means of standardizing approaches to data analysis.

Publication of the multispecies method (Stephens and MacCall, 2004) generated interest and inquiries from fisheries analysts in the NMFS Northwest Fisheries Science Center, the Southeast Fisheries Science Center, researchers at UC Davis and at the California Department of Fish and Game. Questions raised by these inquiries led to the work in Chapter 3, in which I generated simulated fishery data, and analyzed performance of the multispecies method under a variety of plausible scenarios representing changing ocean conditions, changing population sizes and consortia, and data collection regimes.

The multispecies regression depends to some extent on balanced information in the catch records: species used to predict the target must represent both co-ocurring and non-coocurring species. The regression may at times appear to succeed under circumstances that warrant some skepticism, as when paucity of data lead to apparently perfect predictions, an instance of the model overfitting the data. It is important to evaluate all indicators of regression performance together: the probabilities and coefficients it generates, the model degrees of freedom and χ^2 statistic. The regression is robust in many ways, but it fails whenever the underlying assumptions are violated: that species exhibit fidelity to habitats that are distinct within the fishing grounds. As long as the method is properly applied, it will provide much-needed consistency in stock assessments, and will vastly improve assessment accuracy.

Appendix A. Parameters for the SMART Model.

The parameters we used for the between-year model of salmon populations are shown in Tables A.1 and A.2. Those in Table A.1 are from the U.S. Fish and Wildlife Atlantic Salmon Stocks report (U.S. Fish and Wildlife, 1999). Values in TableA.2 were estimated. Critical points in the timeline for the physiological model are shown in Table A.3. Parameters used in the physiological growth model are given in Table A.4. Parameters for disease are given in Table A.5.

Table A.1. Parameters from Atlantic Salmon report and published literature.

Survival parameters	
$\sigma_0 = .08$	Average survival to fry * Average survival to parr
$\sigma_1 = .5$	Survival to 1st-year smolt
$\sigma_2 = .6$	Survival a second year in-stream
$\sigma_3 = \sqrt{(.1)}$	Ocean year survival rate
Life history parameters	
$f = .8$	Percent of parr that smolt after 2 years
$g_1 = 0.05$	One-year smolts that return as grilse
$g_2 = 0.05$	Two-year smolts that return as grilse
$e_1 = 5400$	Egg production of a 1-SW female
$e_2 = 7200$	production of a 2-SW female
$\rho_b = .5$	rate of maturing first-year parr
$\rho_f = .125$	success rate of mature parr
$\rho_s = .21$	rate at which mature parr smolt

Table A.2. Estimated parameters (default values for "low" competition).

Competition coefficients	NB: Coefficients apply to wild and aquaculture fish
β_e = 0.000005	Egg competition
β_{11} = 0.0001	One-freshwater year smolts
β_{21} = 0.0001	Two-freshwater year smolts in their first year
β_{22} = 0.0001	Two-freshwater year smolts in their second year
Aquaculture parameters	
ζ = .5	Cross-survival of hybrid eggs
ρ_m = .3	Penalty for maladaptation
ρ_a = .3	Probability of assortative mating
Scenario-dependent	
r_p = 0.05	Mortality of wild fish due to recapture
ξ_w = 0.8	Survival of wild fish from enhanced predation
ξ_a = 0.6	Survival of aquaculture fish from enhanced predation
Initial wild populations	
A_{11} = 50	Grilse returning to spawn
A_{22} = 200	Two-seawinter adults returning to spawn

Table A.3. Timeline for maturing smolts

Julian day	
84	Day 84 in second year of life; first day of the simulation
106	Evaluation point for maturation the following November
197	Onset of anorexia for a fish maturing in November
315	Evaluation point for maturation in November of the third year
680	Day of reproduction

Table A.4. Parameters for the physiological model

Parameter	
$q = 0.08$	Ability to find and process food
$q_e = 1$ or 0.5	Environmental component of q; changes with escape
$q_i =$	Individual component of q; drawn from a Normal(0.08,0.09)
$q_w = 0.9q$	Optimal value of q in the wild
$c_a = 0.001$	Cost of anorexia
$c_r = 0.988$	Cost of reproduction
$c_w = 0.8$	Penalty for living in the wild after escape
Initial weight	Drawn from a Normal(47, 2.66)

Table A.5. Disease parameters

Disease parameters	
D_I/m_I = 1.45e-2	Diffusion for food particles
D_B/m_B = 2.3E-6	Diffusion for disease particles
$I_t = B_t$= .1	Threshhold values for food and disease
c = .8 , γ = 1	“Type II” transmission, Holling Type II
c = .8 , γ = 3	“Type III” transmission, Holling Type III

Appendix B. Sample code for the multispecies regression in R

```
# Set up CDFG regressions on species and location

# Read in data sorted by visit to each site

fish = read.table('Sample.data', header=T, sep=',')
fish=data.frame(fish)
attach(fish)

# Get Regressor species and set up formula for GLM

f1 = 'Bocaccio ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10'
f2 = '+ X11 + X12 + X13 + X14 + X15 + X16 + X17 + X19 + X20'

fish.form=as.formula(paste(f1,f2))

# Regress on all species

my.lm=glm(formula=fish.form,family=binomial)


obs = sum(Bocaccio)
thresh=seq(0,1,by=0.01)
thresh.effect=thresh
thresh.count=thresh

for ( i in 1:length(thresh) ) {
    thresh.effect[i] = abs(obs - sum(fitted.values(my.lm) > thresh[i]))
    thresh.count[i] = sum(fitted.values(my.lm) > thresh[i])
}
mythresh=cbind(thresh,thresh.effect,thresh.count)
best = min(thresh.effect)
best.thresh = thresh[thresh.effect == best]

print('Regressing on all species')

Btrips.pred=ifelse(fitted.values(my.lm) > best.thresh,1,0)
False.neg = sum(ifelse(Bocaccio > Btrips.pred,1,0))
False.pos = sum(ifelse(Bocaccio < Btrips.pred,1,0))
Correct.pred = sum(ifelse(Bocaccio == Btrips.pred,1,0))
Trips = length(fish[,1])
Pct.correct = Correct.pred/Trips*100
Pct.correct
False.neg/Trips*100
False.pos/Trips*100

years = seq(min(YEAR),max(YEAR))
foo = hist(fitted.values(my.lm),plot=F, breaks=9)
myhist = data.frame(cbind(foo$mids,foo$count))

foo = coefficients(my.lm)
mycoeffs = data.frame(cbind(names(foo), foo))


x11()
```

```
plot(thresh, thresh.effect, main='Northern California MRFSS Survey Data',
ylab = 'Difference between actual and predicted trips', xlab = 'Probablility
Threshold', pch=16)
lines(thresh, thresh.effect)

yr.byspcs=matrix(rep(0,20*3), ncol=3)
yr.byspcs=data.frame(yr.byspcs)
names(yr.byspcs) = c('Year', 'Actual', 'Predicted')

for (i in 1:length(years)) {
    yr.byspcs[i,1] = years[i]
    yr.byspcs[i,2] = sum(Bocaccio[YEAR == years[i]])
    yr.byspcs[i,3] = sum(fitted.values(my.lm)[YEAR == years[i]])
}

x11()

leg.txt = c("Observed", "Predicted")
plot(yr.byspcs[,1], yr.byspcs[,2], xlab='Year', ylab='Bocaccio Trips (Actual
and Predicted)', main='Northern California MRFSS Survey Data', pch=15)
lines(yr.byspcs[,1], yr.byspcs[,3], col=2)
lines(yr.byspcs[,1], yr.byspcs[,2])
points(yr.byspcs[,1], yr.byspcs[,3], col=2, pch=16)
legend(1990,35, legend=leg.txt, col=1:2, pch=15:16)

#  Probability histograms

x11()
hist(fitted.values(my.lm), xlab='Probability', ylab='Frequency',
main='Northern California MRFSS Survey Data')


write.table(yr.byspcs, quote=F,row=F,sep=',', file ='out.MRFSS.yr.byspcs')
write.table(myhist, quote=F,row=F,sep=',', file ='out.MRFSS.hist')
write.table(mycoeffs, quote=F,row=F,sep=',', file ='out.MRFSS.coeffs')
write.table(mythresh, quote=F,row=F,sep=',', file ='out.MRFSS.thresh')


#  Selected trips

fish.select = fish[fitted.values(my.lm)> best.thresh,]
write.table(fish.select, quote=F,row=F,sep=',', file='Selected_Data')
```

Bibliography

Anand, P., 2002. Decision-Making When Science is Ambiguous. Science 295:1893.

Anderson, J.L., Hilborn, R.W., Lackey, R.T., and Ludwig, D., 2003. Watershed restoration – adaptive decision making in the face of uncertainty. pp. 203-232, In: Strategies for Restoring River Ecosystems: Sources of Variability and Uncertainty in Natural and Managed Systems. Wissman, R.C., and Bisson, P.A., Eds., American Fisheries Society, Bethesda, Maryland, 276 pp.

Argaez, J.A., Christen, J.A., Nakamura, M., Soberon, J., 2005. Prediction of potential areas of species distributions based on presence-only data. Environ. and Ecol. Statist., 12. 27-44.

Ash, D., and Klein, K., 1999. Inquiry in the informal learning environment. pp. 216-240. In: Teaching and Learning in an inquiry-based classroom, J. Minstrell and E. Van Zee, Eds., AAAS Press.

Belsley, D.A., Kuh, E., and Welsch, R.E. 1980. Regression Diagnostis: Identifying Influential Data and Sources of Collinearity. Wiley and Sons, New York.

Bishop, J., Venables, W.N., and Wang, Y.-G., 2004. Analysing commercial catch and effort data from a Penaeid trawl fishery: A comparison of linear models, mixed models, and generalised estimating equations approaches. Fish. Res. 70:179-193.

Bjornsson B.T., Taranger G.L., Hansen T., et al. 1994. The Interrelation Between Photoperiod, Growth-Hormone, And Sexual-Maturation Of Adult Atlantic Salmon (Salmo-salar). Gen. Comp. Endocr. 93 (1): 70-81

Carpenter, S.R., 2000. Ecological futures: building an ecology of the long now. Ecology 83:2069-2085.

Cipriano, R.C., and Bullock, G.L., 2001. Furunculosis And Other Diseases Caused By *Aeromonas salmonicida*, Disease Leaflet 66, USGS/Leetown Science Center, National Fish Health Research Laboratory, Kearyneysvill, West Virginia.

Clegg, M.T., Barten, P.K., Fleming, I.A., Gross, M.R., *et al*. 2001. Genetic Status of Atlantic Salmon in Maine: Interim Report from the Committee on Atlantic Salmon in Maine.

Cotter, D., O'Donovan, V., O'Maoileidigh, N., Rogan, G., et al., 2000. An evaluation of the use of *triploid* Atlantic *salmon* (Salmo salar L.) in minimising the impact of escaped farmed *salmon* on wild populations. Aquaculture 186: 61-75.

Dick, E.J., 2004. Beyond ’lognormal vs. gamma’: discrimination among error distributions for generalized linear models. Fish. Res. 70, 347–362.

Elberizon, I.R., and Kelly, L.A., 1998. Empirical Measurements of Parameters Critical to Modelling Benthic Impacts of Freshwater Salmonid Cage Aquaculture. Aqua. Res. 29:669-677.

FAO (Food and Agriculture Organization of the United Nations), 1997. Review of the State of World Fishery Resources: Marine Fisheries. Rome: FAO.

Fleming, I.A., Hindar, K., Mjolnerod, I.B., Jonsson, G., Balstad, T., and Lamberg, A., 2000. Lifetime success and interactions of farm salmon invading a native population, Proc. R. Soc. Lond. B 267:1517-1523.

Fleming I.A., Agustsson T. , Finstad B. , Johnsson J.I., Björnsson B.T., 2002. Effects of domestication on growth physiology and endocrinology of Atlantic salmon (*Salmo salar*). Can. J. Fish. Aquatic Sci. 59:1323-1330.

Forsberg O.I., 1995. E mpirical Investigations On Growth Of Post-Smolt Atlantic Salmon (Salmo-Salar L) In Land-Based Farms - Evidence Of A Photoperiodic Influence. Aquaculture 133 (3-4): 235-248.

Garant, D., Fleming, I.A., Eimun, S., and Bernatchez, L., 2003. Alternative male life-history tactics as potential vehicles for speeding introgression of farm salmon traits into wild populations. Ecol. Letters 6: 541-549.

Goldburg, R.J., Elliott, M.S., and Naylor, R.L., 2001. Marine Aquaculture in the United States: Environmental Impacts and Policy Options. Pew Oceans Commission, Arlington, Virginia.

Guisan, A., Edwards, T.C. Jr., Hastie, T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Model. 157, 89-100.

Gulland, J.A. 1983. Fish Stock Assessment: A Manual of Basic Methods. Wiley and Sons, New York.

Ihaka, R. & Gentleman, R. 1996, R: A Language for Data Analysis and Graphics, Journal of Computational and Graphical Statistics, 5, 299-314.

Hansen, L.P., Windsor, M.L., and Youngson, A.F., 1997. Interactions Between Salmon Culture and Wild Stocks of Atlantic Salmon: The Scientific and Management Issues. ICES J. Mar. Sci. 54:963-1225.

Harley S.J., Myers, R.A., and Dunn. A. 2001. Is catch-per-unit-effort proportional to abundance? Can. J. Fish. Aquat. Sci. 58: 1760–1772.

Harwood, J., and Stokes, K., 2003. Coping with uncertainty in ecological advice: lessons from fisheries. TREE 18(12): 617-622.

Henderson B.A., Wong J.L., 1998. Control Of Lake Trout Reproduction: Role Of Lipids.
J Fish Biol 52 (5):1078-1082.

Hilborn, R., Branch, T.A., Ernst, B., *et al*., 2003. State of the World's Fisheries. Annu. Rev. Environ. Resour. 28:15.1-15.40.

Hilborn, R., Punt, A.E., and Orensanz, J.M. 2004. Beyond Band-aids in Fisheries Management: Fixing World Fisheries. Bull. Mar. Sci., 74(3): 493-507.

Hilborn, R., Orensanz, J.M., and Parma, A.M. 2005. Institutions, Incentives and the Future of Fisheries. Phil. Trans. R. Soc. B 360:47-57.

Hindar, K., 2001. Interactions of cultured and wild species, pp. 102-131, In: Marine Aquaculture and the Environment. M.Tlusty, D. Bengtson, H.O. Halvorson, S. Oktay, J. Pearce and R.B. Rheault , Eds. Cape Cod Press, Falmouth, MA.

Hindar, K. and Balstad, T., 1994. Salmonid Culture and Interspecific Hybridization. Conservation Biology 8:881-882.

Hutchings, J.A., and Jones, M.E.B., 1998. Life history variation and growth rate thresholds for maturity in Atlantic salmon, *Salmo salar*. Can.J.Fish.Aquat. Sci. 55 (Suppl. 1): 22-47.

Ihaka, R., Gentleman, R., 1996. R: a language for data analysis and graphics. J. Comput. Graph. Statist. 5, 299–314.

Imsland A.K., Folkvord A., Jonsdottir O.D.B., et al. 1997. Effects of exposure to extended photoperiods during the first winter on long-term growth and age at first maturity in turbot (Scophthalmus maximus). Aquaculture 159 (1-2):125-141.

Iudicello, S., Weber, M., and Wieland, R., 1999. Fish, Markets, and Fishermen: The Economics of Overfishing. Island Press, Washington, D.C.

Jeffers, J.N.R., 1978. An Introduction to Systems Analysis: with ecological applications. Edward Arnold Ltd., London.

Jennings, S., Kaiser, M.J., and Reynolds, John D., 2001. Marine Fisheries Ecology. Blackwell
Science Ltd., Oxford, UK.

Jimenez, M.P., Sobrino, I., and Ramos, F., 2004. Objective methods for defining mixed-species trawl fisheries in Spanish waters of the Gulf of Cadiz. Fish. Res. 67: 195-206.

Jonsson N, Jonsson B. 1998. Body Composition And Energy Allocation In Life-History Stages Of Brown Trout. J Fish Biol 53 (6):1306-1316

Lackey, R.T. 2003. Pacific Northwest Salmon: Forecasting their status in 2100. Reviews in Fisheries Science. 11(1):35-88.

Lacroix, G.L., and Stokesbury, M.J.W., 2004. Adult Return of Farmed Atlantic Salmon Escaped as Juveniles into Freshwater. Trans. Amer. Fish. Soc., 133: 484-490.

Love, M.S., Yokalovich, M., and Thorsteinson, L., 2002. The Rockfishes of the Northeast Pacific. University of California Press, Berkeley.

MacCall, A.D. 2003. Status of bocaccio off California in 2003. In: Status of the Pacific coast groundfish fishery through 2003 stock assessment and fishery evaluation Vol. 1. Pacific Fishery Management Council, 7700 NE Ambassador Place, Suite 200, Portland, OR 97220-1384.

Mangel, M. 1994a. Climate change and salmonid life history variation. Deep Sea Research, II (Topical Studies in Oceanography) 41:75-106.

Mangel, M. 1994b. Life history variation and salmonid conservation. Cons. Biol. 8(3): 879-880.

Mann, K.H., and Lazier, J.R.N., 1996. Dynamics of Marine Ecosystems: Biological-Physical Interactions in the Oceans, 2nd Ed., Blackwell Science, Malden, MA. p. 23.

Mathisen, O.A., and Zheng, J., 1994. Changing ***sex*** ratios during spawning migration of pink salmon in southeast Alaska. NORTHEAST PACIFIC PINK AND CHUM SALMON WORKSHOP., 1994, pp. 137-146. http://www.psc.org

Mathworks, Inc. 2004. http://www.mathworks.com

Maunder, M.N., Punt, A.E., 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res. 70. 141-159.

McCall, R.B. 2001. Fundamental Statistics for the Behavioral Sciences, 8th Ed. Wadsworth/Thompson Learning, Belmont, CA.

McCallum, H., Barlow, N., and Hone, J., 2001. How should pathogen transmission be modelled? Trends in Ecology & Evolution, 16(6):295-300.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Chapman & Hall, New York.

McEvoy, A.F., 1986. The Fisherman's Problem: Ecology and Law in the California Fisheries, 1850-1980. Cambridge University Press.

McVicar, A.H., 1997. Disease and parasite implications of the coexistence of wild and cultured Atlantic salmon populations, ICES J. Mar. Sci., 54: 1083-1103.

Minstrell, J. 1999. Implications for teaching and learning inquiry: A summary. pp: 471-496. In: Teaching and Learning in an inquiry-based classroom, J. Minstrell and E. Van Zee, Eds. AAAS Press.

Naylor, R.L., Williams, S.L., Strong, D.R., 2001. Aquaculture – A Gateway for Exotic Species. Science 294: 1655-1656.

Naylor, R., Hindar, K., Fleming, I.A., Goldburg, R., et al. 2005. Fugitive Salmon: Assessing the Risks of Escaped Fish from Net-Pen Aquaculture. BioScience 55(5): 427-437.

Osborn, M.F., Van Voorhees, D.A., Gray, G., Salz, R., Pritchard, E., Holliday, M.C., 1996. Marine Recreational Fishery Statistics Survey, National Marine Fisheries Service, NOAA, U.S. Dept. of Commerce. http://www.psmfc.org/recfin/data.htm.

Paris, S., 1997. Situated motivation and informal learning. J. Museum Educ. 22(213): 22-26.

Peterman, R.M., and Steer, G.J. 1981. Relation between sportfishing catchability coefficients and salmon abundance. Trans. Am. Fish. Soc. 110: 585–593.

Pickitch, E.K., Santora, C., Babcock, E.A., Bakum, A., *et al.*, 2004. Ecosystem-Based Fishery Management. Bioscience 305:346-347.

Press, S. J., and S. Wilson. 1978. Choosing between logistic regression and discriminant analysis. J. Amer. Stat. Assn. 73:699-705.

Punt, A.E., Pribac, R., Walker, T.I., Taylor, B.L., 2001. Population modeling and harvest strategy evaluation for school and gummy shark. Report of FRDC Project No. 99/102. Cited in: Maunder and Punt, 2004.

Quinn, T.J., and Deriso, R.B., 1999. Quantitative Fish Dynamics. Oxford University Press, New York.

Rosenberg, A.A., 2003. The Precautionary Approach in Application from a Manager's Perspective. Bull. Mar. Sci. 70(2):577-588.

Ralston, S., Pearson, D., and Reynolds, J. 1998. Status of the Chilipepper Rockfish Stock in 1998. In Pacific Fishery Management Council, 1998. Appendix: Status of the Pacific Coast Groundfish Fishery Through 1998 and Recommended Acceptable Biological Catches for 1999: Stock Assessment and Fisherry Evaluation. Pacific Fishery Management Council, Portland, Oregon.

Ricker W.E. 1975. Computation and Interpretation of Biological Statistics of Fish Populations. Fisheries Research Board of Canada, Bulletin 191.

Sakai, A.K., Allendorf, F.W., Holt, J.S., Lodge, D.M., et al., 2001. The Population Biology of Invasive Species. Annu. Rev. Ecol. Syst. 32: 305-332.

Schnute, J.T., and Richards, L.J., 2001. Use and Abuse of Fishery Models. Can.J. Fish. Aquat. Sci. 58: 10-17.

Schnute, J.T., 2003. Designing Fishery Models: A Personal Adventure. Nat.Res.Modeling 16(4): 393-413.

Stefansson, G. 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES J. Mar. Sci. 53, 577-588.

Stephens, A., and MacCall, A. 2004. A Multispecies Method for Subsetting Logbook Data for Purposes of Estimating CPUE. J. Fish. Research 70, 299-310.

Stephenson, R.L., and Lane, D.E., 1995. Fisheries management science: a plea for conceptual change. Can.J. Fish. Aquat. Sci. 52: 2051-2056.

Swain, D.P., and Sinclair, A.F. 1994. Fish distribution and catchability: what is the appropriate measure of distribution? Can. J. Fish. Aquat. Sci. 51: 1046–1054.

Taper, M.L., and Lele, S.R., Eds. 2004. The Nature of Scientific Evidence: Statistical, Philosophical and Empirical Considerations. The University of Chicago Press, Chicago, IL.

Taylor, P.R., 2003. Standardized CPUE for the northwest Chatham Rise orange roughy fishery. NZ Fisheries Association Report No. 2003/32. Cited in: Maunder and Punt, 2004.

Thorpe, J.E., Mangel, M., Metcalfe, N.B., and Huntingford, F.A., 1998. Modelling the proximate basis of salmonid life-history variation, with application to Atlantic salmon, *Salmo salar* L. Evol. Ecol., 12: 581-599.

U.S. Fish and Wildlife Service, 1999. Atlantic Salmon Stocks Status Report, http://library.fws.gov/salmon

VanBuskirk, W., Ed., 2003. RecFIN Database, National Marine Fisheries Service, NOAA, U.S. Dept. of Commerce, http://www.psmfc.org/recfin/data.htm

Volpe, J.P., Taylor, E.B., Rimmer, D.W., Glickman, B.W., 2000. Evidence of Natural Reproduction of Aquaculture-Escaped Atlantic Salmon in a Coastal British Columbia River. Cons. Biol. 14(3): 899-9023.

Whalen, K.G., Parrish, D.L., Mather, M.E., and McMenemy, J.R., 2000. Cross-tributary analysis of parr to smolt recruitment of Atlantic salmon (*Salmo salar*). Can. J. Fish. Aquat. Sci. 57: 1607-1616.

Whoriskey, F.G.; Carr, J.W., 2001. Returns of transplanted adult, escaped, cultured Atlantic salmon to the Magaguadavic River, New Brunswick. ICES J. Mar. Sci. 58(2) 504-509.

Williams, E.H., and Ralston, S. 2002. Distribution and co-occurrence of rockfishes (family: Sebastidae) over trawlable shelf and slope habitats of California and southern Oregon. Fish.Bull. 100, 836-855.