META-ANALYTICAL APPROACHES TO THE STUDY OF FISH POPULATION DYNAMICS

By

Shelton J. Harley

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT
DALHOUSIE UNIVERSITY
HALIFAX, NOVA SCOTIA
AUGUST 2002

0-612-75725-0

DALHOUSIE UNIVERSITY

FACULTY OF GRADUATE STUDIES

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "Meta-Analytical Approaches to the Study of Fish Population Dynamics", by Shelton James Harley, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dated: August 1, 2002

External Examiner:

Research Supervisor:

Examining Committee:

[redacted]

DALHOUSIE UNIVERSITY

Date: **August 2002**

Author: **Shelton J. Harley**

Title: **Meta-analytical approaches to the study of fish population dynamics**

Department: **Biology**

Degree: **Ph.D.** Convocation: **October** Year: **2002**

Permission is herewith granted to Dalhousie University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.

[redacted]

Signature of Author

Contents

To my wife, Mei Lin

List of Tables

List of Figures

Acknowledgements

In the process of this research I have received considerable support, encouragement, and assistance from many people and organizations. I thank the Killam Foundation, the Department of Fisheries and Oceans, the New Zealand Ministry of Fisheries, and the Inter-American Tropical Tuna Commission for their financial support of various parts of this thesis.

I thank my supervisor Ransom Myers for his seemingly endless supply of ideas and for not letting me be satisfied with a job half-done. I would also like to thank my committee, Chris Field, Jeff Hutchings, and Bob Mohn for their discussions and guidance throughout this thesis. Also, I would like to thank Ken Frank for valuable reviews of the original thesis proposal and preliminary exam. I would like to thank Jeremy Collie for agreeing to be the external examiner for this work and the many useful comments he provided. Stephen Smith provided invaluable advice on using LaTeX.

Some of the models described in this thesis were developed in collaboration with various colleagues. I thank Ransom Myers for assistance deriving and interpreting the observation error and state-space models described in Chapters 3 and 4; Ransom Myers and Chris Field for discussion on implementing the maximum simulated likelihood approach; Mark Maunder for advice on the simulation studies in Chapters 6 and 7; Alistair Dunn for the early work that motivated the analysis in Chapter 3; and Patrick Cordue for some useful interpretations of the models that I developed in Chapter 6.

All of the data used in this thesis have been collected by various research agencies around the world. I would like to acknowledge the scientists from: International Council for the Exploration of the Sea (ICES); National Marine Fisheries Service

(NMFS); Department of Fisheries and Oceans (DFO); and the National Institute of Water and Atmospheric Research (NIWA) for collecting and making these data available.

I would like to thank all members of the Myers' lab for their support during this thesis. In particular I would like to thank the following, for reading of far too many long manuscripts: Nick Barrowman, Penny Doherty, Julia Baum, Susanna Fuller, Leah Gerber, Jamie Gibson, Ian Jonsen, Dan Kehler, Kristjan Kristjanson, Heike Lotze, Kurtis Trzcinski, and Boris Worm.

Finally, I would like to thank my wife, Mei Lin, for her incredible support and allowing me to come half way across the world to pursue my ambitions.

Abstract

The aim of this thesis is to reduce uncertainties in our understanding of fish population dynamics by: (a) developing statistical methods that allow information to be shared or combined; and (b) applying these methods to real fisheries problems. Meta-analysis is a statistical framework for making inferences by combining results across studies or experiments. In this thesis I demonstrated that meta-analysis is particularly useful for studies of fisheries population dynamics because: (1) there is often considerable observation error in fisheries data that creates difficulties in making useful inferences; and (2) there are potentially relevant data for many fish populations or years. I showed, using data for many populations, that catch-per-unit-effort abundance indices generally underestimate the declines in the population being fished – a finding critical to fisheries stock assessment. I developed a generalized framework to test hypotheses relating to the regulation of fish abundance during the larval and juvenile stages – thus allowing for future meta-analyses of this problem. I extended the traditional meta-analytical methods to combine estimates of the relationship between the size of a fish and its catchability to trawl surveys – this allowed for predictions to be made for species where no data were available. Using a method from the field of econometrics, I developed an approach to integrate random effects into a complex nonlinear model – this allowed for improved modelling and interpretation of million dollar acoustic surveys of a New Zealand fish population. I demonstrated how inferences from meta-analysis can be used in the assessment of stocks with little information (e.g., developing fisheries) and to obtain reliable estimates required for management decisions. To conclude, I provided a discussion of how the further extension of methods and fisheries databases would be critical to the future of meta-analysis, and highlighted important directions for future research.

Chapter 1

Introduction

An improved understanding of fish population dynamics is not only of general ecological importance, but critical to the sustainable exploitation of fish populations. Without this understanding, human activities can lead to the over-exploitation and "collapse" of fish populations (Hutchings and Myers 1994b; Walters and Maguire 1996; Myers et al. 1997). The purpose of this thesis is to develop and apply statistical methods that will reduce the uncertainty in our understanding of fish population dynamics. I will be focussing on methods to improve estimates of relative abundance, absolute abundance, predicting abundance from juvenile surveys, and finally target levels of abundance for management.

There are two major impediments to our improved understanding of fish population dynamics: (1) poor quality data; and (2) lack of adequate quantitative methods to analyze these data. The time series data collected from fish populations (e.g., surveys or catch-per-unit-effort data) have three attributes that make it difficult to derive useful inferences from them. First, the time series are generally short – by the standards of statisticians and with respect to the life history of the species surveyed. Second, our observations are often indirect (e.g., research surveys) and are plagued with considerable observation error. Finally, there are many poorly understood abiotic (e.g., environmental conditions) and biotic (e.g., predator-prey dynamics) factors that add additional (process) noise to our data. While these limitations make inferences about fish populations difficult, fisheries scientists are still required to provide information for fishery managers to make decisions. Thus, the emphasis is on using the best available information and making the most of all relevant data (Hilborn and Liermann 1998).

In this thesis I will argue that we can begin to overcome some of the problems caused by inadequate data by using methods that combine and/or share information across populations, species, or years. I will describe some of the more commonly used methods in this thesis. However, many aspects of fisheries data do not meet the assumptions required for these analyses. We need methods that allow for both observation error, process noise, and nonlinear dynamics. In particular we require methods that can be integrated into the complex and computationally intensive population dynamic models that are often used in fisheries.

The early work of Beverton and Holt (1959), and many other fisheries ecologists since (e.g., Hilborn and Liermann 1998 and Myers and Mertz 1998), emphasized the importance of looking at many data sets simultaneously to improve our understanding of fish population dynamics. Myers and Mertz (1998) argue that fishery data is of such poor quality that without looking at many data sets simultaneously one cannot understand the true underlying dynamics. Pauly (1998) noted that the single-species focus of many fisheries scientists hinders us from seeing the 'big picture' and understanding the processes that lead to the variation that we observe. Hilborn and Liermann (1998) stressed that fisheries science cannot move forward unless we learn from past experience. Collie and Walters (1991) found that stock assessment methods that recognized similarities among substocks outperformed methods that did not. These arguments are all consistent with the goals of a methodological framework known as 'meta-analysis'.

Meta-analysis is a set of statistical procedures for combining information from independent studies or experiments. The term meta-analysis was first used in the field of education but this approach has been popularized in the social sciences and is now commonly used in the synthesis of results from medical clinical trials (Hedges and Olkin 1985; Cooper and Hedges 1994). In meta-analysis, estimates of a 'metric' from a number of studies are combined. Traditionally, this metric is a quantitative measure of the outcome of an experiment. In fisheries many researchers have adapted these traditional approaches to combine important biological and ecological parameters across populations and species (Myers and Mertz 1998).

An approach that is used with many meta-analytic approaches is that of *hierarchical modelling.* Hierarchical models have multiple levels that have uncertainty at each level. For example, consider estimates of natural mortality for some fish species. At the population level there will be estimation error in the individual estimates. Next, at the species level there may be variation among populations in the true value of natural mortality, i.e., there may be a distribution for a species. Perhaps, there may even be considerable overlap in these distributions for species within a genera or family. This would be a perfect example of how we could combine estimates of such a parameter (e.g., natural mortality) in a meta-analysis using a hierarchical model to estimate how much the parameter varies within a family. The terms meta-analysis and hierarchical modelling will both be used in this thesis.

The following is a simple example to illustrate some features of a meta-analysis.

Example: Do barbless hooks reduce hooking mortality in trout?

Here is a re-analysis of the data presented by Schill and Scarpella (1997). They compared 11 side-by-side trials of barbed and barbless hooks that compared the number of trout that died after being caught on barbed or barbless hooks. I will perform a meta-analysis of the odds ratio, i.e., the odds of dying after being caught on a barbless hook divided by the odds of dying after being caught on a barbed hook. I consider the barbed hook as the control as we are testing the effect of some new treatment (the barbless hook) on mortality. The data are in the form

	died	survived
barbed	a	b
barbless	c	d

where the probabilities of death from barbed hooks and barbless hooks are $A = a/(a+b)$ and $C = c/(c+d)$ respectively. The odds ratio is a common metric used to describe these types of data, i.e., clinical trials of drugs or medical procedures. A value of 1 for the odds ratio indicates no difference between the hook types in the probability of mortality . The analysis is usually performed using the log odds ratio because it has better sampling qualities. The log odds ratio is given by

$$\theta = \log\left(\frac{A(1-C)}{C(1-A)}\right),$$

with standard error

$$\mathrm{SE}(\theta) = \frac{1}{(a+b) \times A(1-A)} + \frac{1}{(c+d) \times C(1-C)}.$$

As there was one trial where there were no deaths for one of the hooks, I added 0.5 to the number of deaths in both the barbed and barbless groups for all studies (as recommended by Shadish and Haddock 1994). The log odds ratios were combined using a random effects model but the results and data are shown in Figure 1.1 in terms of the odds ratio.

The trials have been ranked, from top to bottom, in order of increased weight in the analysis, i.e., the influence each trial's results had on the overall results. There were three important results: (1) the overall mean odds ratio favors barbless hooks for survival but the difference in mortality is not statistically significant; (2) studies with greater total observed deaths had more influence; and (3) the homogeneity test (described in Chapter 2) indicated that the true odds ratio did not differ across studies, i.e., there was no significant variation across trials in the effect of barbless hooks.

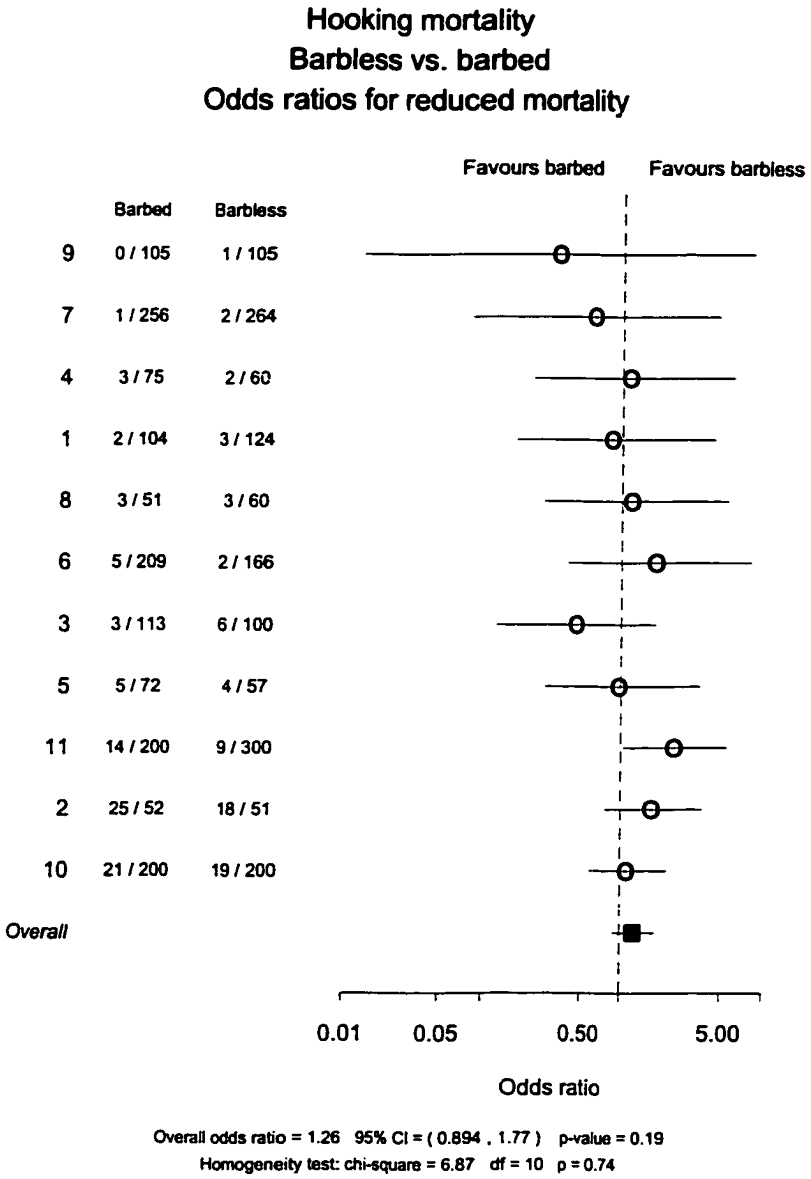


Figure 1.1: Results of the random effect analysis of the log odds ratios from experiments comparing barbed and barbless hooks. Note that as the results are presented in terms of the odds ratio, the axes are asymmetric.

1.1 Aim and purpose

The aim of this thesis is to reduce uncertainties in our understanding of fish population dynamics, particularly problems faced in fishery management. This thesis will achieve this aim in two ways:

- by developing methods that allow information to be shared or combined, and
- by applying these methods to real fisheries problems.

The specific purpose of this thesis is to make the following contributions:

1. Provide a comprehensive and critical review of meta-analysis applications in fisheries, identifying key approaches, issues and potential pitfalls;
2. Increase the potential use of meta-analysis in addressing fisheries problems by developing and demonstrating useful meta-analysis methods for:
    (a) testing hypotheses using data from many populations,
    (b) combining information in the form of a functional relationship rather than a single quantity or value,
    (c) improving estimates of key parameters in complex dynamic models by sharing information across studies,
    (d) incorporating external information in stock assessment to improve estimates of key quantities required for fisheries management.

As I will be extending existing methods there will be a strong focus on models and approaches in this thesis. These models are (a) necessary (evil) to answer the important questions considered in the following chapters.

1.2 Overview

In Chapter 2 I will provide a comprehensive and critical review of meta-analysis applications in fisheries. Key components of meta-analysis will be reviewed including the choice of a metric (to combine) and potential biases that can affect the results of a meta-analysis. These biases will also be discussed in each of the subsequent chapters. I will differentiate three levels of meta-analysis and indicate where the examples described in this thesis fit into these categories.

A fundamental question in fisheries stock assessment is the relationship between catch-per-unit-effort (CPUE) indices of abundance and the true population abundance (Hilborn and Walters 1992; Quinn and Deriso 1999). This problem will be considered in Chapter 3. As it is generally difficult with any single data set to determine the relationship between the abundance indices and true abundance, it is often assumed that CPUE indices are proportional to true abundance. I will describe the analysis of 297 CPUE data sets using linear mixed effect models that incorporate observation error. In the second part of that chapter I will use a simple random effects model to combine estimates of a parameter describing the relationship between CPUE and abundance.

Another important fisheries question relates to determining when the year class strength of a cohort is determined. The practical implication is: how soon before a cohort enters the fishery can we predict its strength? This question is related to the timing and variation in density-independent mortality, and the magnitude of any density-dependent mortality. The original goal of Chapter 4 was to develop models to test a number of hypotheses using meta-analysis. However, adequate models for addressing this problem were not available and were not trivial to develop. Previous studies used linear mixed effects models with observation error (like those used in Chapter 3). The important aspects of this problem require methods that allow nonlinear dynamics and the incorporation of process noise (i.e., extra true variation not explicitly modelled). Therefore rather than a perform a full meta-analysis of this problem, I will describe the development of models (and methods to estimate the parameters of the models) for testing hypotheses about the timing and role of density-independent and density-dependent mortality in the regulation of year class strength. The models will utilize some meta-analytical principles as they share information across cohorts to allow improved estimation of model parameters. I will apply these methods to data from two Atlantic cod (*Gadus morhua*) populations.

Obtaining estimates of absolute abundance of a species is critical to the traditional single-species management regime. However, the apparent ecosystem effects of fisheries (Pauly et al. 1998) has led to a move towards "ecosystem-based" approaches to fisheries management (Walters et al. 1997) which requires estimates of abundance for both commercially exploited species and other associated species. In Chapter 5, I will estimate length-specific scaling factors, known as catchability coefficients, for a range of species groups. These scalars are necessary to correct trawl survey estimates of abundance because not all species and sizes of fish are caught equally well by the survey gear. In this analysis I use a three-parameter curve to describe the relationship between trawl survey catchability and the size of a fish.

In Chapter 6 I will use meta-analytical approaches to estimate residence time and biomass of New Zealand hoki (*Macruronus novaezelandiae*) on spawning grounds. I develop a number of complex dynamic models that describe the biological process. As the data for any one year is seldom sufficient to obtain reliable estimates for the key parameters, I simultaneously model data for all years using a hierarchical approach. This is possible if we are willing to assume that spawning dynamics are likely to be similar (i.e., not independent) from year to year.

The analysis described in Chapter 3 will produce probability distributions for one of the key parameters of interest in fisheries stock assessment (the relationship between CPUE and abundance). In a Bayesian context, this information can be interpreted as a *prior* as it provides us with information about likely values of a parameter *before* the analysis of the (local) data. When we are modelling a single exploited population this information can be used as auxiliary information if the local data are insufficient to allow us to estimate important quantities (i.e., sustainable yields and biomass target levels). In Chapter 7 I describe how this prior information can be used to improve the performance of fisheries assessment models when local data are poor. This can be critically important in either developing fisheries or fisheries of such low value that detailed research programs are not economically viable.

In Chapter 8 I will discuss how my research has extended the application of meta-analytical techniques in fisheries and contributed to our understanding of fish population dynamics. Future research areas that may be amenable to meta-analytic enquiry will be identified and I will discuss the key technical issues that this work has raised.

Chapter 2

Review of meta-analytical approaches in fisheries

This chapter is divided into two parts. In the first I discuss some meta-analytical approaches and review their application to the study of fish population dynamics. I will identify the strengths and weaknesses of each approach. In the second I review the process of a meta-analysis, from question definition to the inferences or generalizations obtained, highlighting critical issues and potential pitfalls.

Overall there have been relatively few researchers and studies that have applied these methods in fisheries. I consider studies of important biological parameters (e.g., those describing growth, natural mortality, and maturity), important ecological processes (e.g., spawner recruitment (SR) dynamics), anthropogenic impacts on these populations (e.g., discard mortality), and our ability to detect such impacts (e.g., the interpretation of abundance indices). I focus on analyses that have combined estimates of some quantity with estimates of their relative reliability.

Throughout this thesis, the term 'study' refers to the unit over which things are combined. If it is a population, year, or cohort, I will also use those specific terms. The terms 'metric', or 'parameter', will be used to describe the quantity combined in a meta-analysis. A collection of these will be referred to as 'meta-data'. When referring to some value from an experiment or population, I will use k to denote the number of studies with a subscript i referring to the ith study.

2.1 Meta-analytical approaches

Statistical approaches for meta-analysis are often different from those used in standard statistical problems. In this section I will describe a range of methods and provide examples of where they have been applied to fisheries problems. I will devote considerable time to describing quantities that have been combined using meta-analytical methods. I hope to introduce the use of these techniques as a way to answer questions beyond the reach of non-meta-analytic methods.

Fisheries scientists are often unaware of the formal meta-analytical methods that can be used address many of their research questions. Researchers unaware of meta-analytical methods have often attempted to perform a 'meta-analysis' using standard statistical techniques, e.g., general linear models. Unfortunately, meta-data often violate the statistical assumptions of homogeneity of variances and the heteroscedasticity is rarely overcome by transformation (Gurevitch and Hedges 1999).

Formal meta-analysis has been used to show that:

- Barbless hooks do not greatly reduce catch-and-release mortality in trout (Schill and Scarpella 1997).
- Increased juvenile mortality, not reduced recruitment, was an important factor in the wide-spread collapse of Atlantic cod (*Gadus morhua*) stocks in the early 1990s (Myers et al. 1997)
- Environmental effects on Atlantic cod recruitment are greater at the northern and southern distribution limits (Myers 1998; Planque and Frédou 1999).

Analysis of p-values

There are two main methods for analyzing p-values from significance tests. If one wishes to know how often some particular effect occurs the simplest method is vote counting (Bushman 1994) which involves counting the number of studies that found a significant effect of the process (e.g., $p < 0.05$). Myers et al. (1995) counted the number of populations for which a parameter describing depensation was statistically significant. Similarly, Brodziak et al. (2001) counted the number of times particular spawner recruitment models provided the best fit to data for a population.

Researchers examining environmental effects on recruitment have often used vote counting before using one of the more powerful techniques (Myers and Drinkwater 1988).

For biologists the most well-known meta-analytical method is that of combining p-values (Fisher 1954; Sokal and Rohlf 1995). Myers et al. (1997), and Myers and Drinkwater (1989) combined p-values from one-sided significance tests in addition to vote-counts.

These approaches are often used where more complex methods (i.e., those that combine the values of the effect rather than its significance) are not possible due to the lack of estimates of reliability of the metric. Vote counting is generally vulnerable to publication bias, since non-significant results are less likely to be published (available for counting).

In the fisheries applications, often the researchers performed the individual analyses themselves to overcome this bias. Perhaps a greater problem is that when a large number of studies are considered, the proportion of significant results is related to the power of the individual tests (Myers and Mertz 1998). This is of particular concern in fisheries where individual studies are often low in power, which will lead to increased Type II error (i.e., failure to reject the null hypothesis). Futhermore, these approaches can only provide evidence for the wide-spread occurrence of some process and tell us nothing of its size and range of possible values.

Non-parametric methods

The sign-test is a common non-parametric method used to test for an effect when the reliability of different metrics is unknown. It involves using a binomial test of the number of times an effect (i.e., correlation or regression coefficient) falls in a particular direction (Bushman 1994). As this technique does not utilize information on the reliability of each effect, it is commonly used with non-parametric regression or correlation techniques (Drinkwater and Myers 1987; Myers and Drinkwater 1989). Myers et al. (1997) used the sign test to examine correlations among Atlantic cod stocks for the relationship between adult and juvenile mortality. The sign test has also been used to show that trout catch-and-release mortality associated with barbless hooks is not significantly different to that of barbed hooks (Schill and Scarpella 1997) and that the addition of zooplanktivorous fish leads to reduced zooplankton abundance and increased phytoplankton abundance (Brett and Goldman 1996).

Some researchers have used simple non-parametric measures to test for evidence of spawner-recruitment relationships. Myers and Barrowman (1996) looked at mean recruitment at high and low levels of spawner abundance, while Gilbert (1997) looked at the ratio of first differences from spawner-recruitment time series.

While the sign-test can test only for wide-spread evidence for some effect or process it may be more appropriate for fisheries applications as it is not limited by the power of individual analyses. Other non-parametric methods must be examined by simulation to determine their power and implicit assumptions. For example, Myers (1997) performed simulations to show that Gilberts' 'derivative-test' was low in power which may have lead to his rejection of the spawner-recruitment hypothesis for marine fishes. Also, any population exhibiting strong overcompensation (i.e., reduced recruitment at high levels of spawner abundance as described by a Ricker curve) would reduce the power to detect evidence for recruitment overfishing in the analysis of Myers and Barrowman (1996).

Conventional approaches to meta-analysis

Generally, researchers have more information than just p-values and are interested in combining estimates of some quantity. Here I describe the more conventional approaches to meta-analysis, where some metric (with an estimate of reliability) are combined for a number of studies. There are two conceptually different approaches: (1) fixed effects models; and (2) random effects models (with both Bayesian and frequentist variations).

The following text and the schematic provided in Figure 2.1 illustrates the difference between the two approaches. I use estimates of the parameter describing the relationship between catch-per-unit-effort and abundance (β) for flatfish; an analysis that will be the focus of Chapter 3.

Here we have k estimates of β_i and associated estimation error variances v_i =

$\mathrm{SE}(\hat{\beta}_i)^2$. In the *fixed* effects model there is best estimate $\bar{\beta}$ which is given by

$$\bar{\beta} = \frac{\sum_{i=1}^{k} w_i \hat{\beta}_i}{\sum_{i=1}^{k} w_i}$$

where $w_i = 1/v_i$. This is shown in the left side of Figure 2.1. It is assumed that all of the variation observed in the individual β_i is due to estimation error. The mean effect is simply a weighted average of the individual estimates with a standard error that also incorporates these weights (Shadish and Haddock 1994).

The *random* effects model is more complicated as we assume that there is an underlying distribution from which the true values were 'drawn'. This distribution describes the between-study variation and is often referred to as the random effects distribution, σ_β^2. The mean is estimated by

$$\bar{\beta} = \frac{\sum_{i=1}^{k} w_i^* \hat{\beta}_i}{\sum_{i=1}^{k} w_i^*}$$

where $w_i^* = 1/(v_i + \sigma_\beta^2)$. The estimate of between-sample variation, σ_β^2, is derived from the statistic Q,

$$\sigma_\beta^2 = \frac{Q - (k-1)}{\sum_{i=1}^{k} w_i - \frac{\sum_{i=1}^{k} w_i^2}{\sum_{i=1}^{k} w_i}},$$

where

$$Q = \sum_{i=1}^{k} w_i (\hat{\beta}_i)^2 - \frac{\left(\sum_{i=1}^{k} w_i (\hat{\beta}_i)^2\right)^2}{\sum_{i=1}^{k} w_i},$$

and $w_i = 1/v_i$.

In the random effects model it is assumed that the variation observed in the individual $\hat{\beta}_i$ is due to both estimation error and true variation, $\mathrm{var}(\hat{\beta}) = v_i + \sigma_\beta^2$. This is shown on the right side of Figure 2.1. The bottom right panel overlays the histogram of individual estimates with the estimated random effects distribution. The significance of the between sample variation, σ_β^2 can be tested using the Q-statistic; rejecting the null hypothesis whenever Q exceeds the $100(1-\alpha)$ percentage point for the chi-squared distribution with $k-1$ degrees of freedom.

It is also possible to have *mixed* effects models, this occurs when a model has multiple parameters and some are treated as fixed effects and others are treated as random effect (Zar 1996; Pinheiro and Bates 2000).

As the variance of the random effects distribution represents true variation it should have some interpretation. For experimental data, significant between-study variation could reflect uncontrolled factors in each study or differences in protocols (Hedges et al. 1999). Exploited populations provide the best example of poor control of uncontrolled factors (e.g., environmental conditions) and differences in protocols (e.g., differing exploitation histories). For this reason, when combining metrics across populations, the random effects model is often more valid than fixed effect models. When a random effects model is used to analyze population metrics, the estimated between-study variance is much less than that observed among individual estimates as estimation error is accounted for (Myers and Mertz 1998; Myers et al. 1999).

The random effects distributions have different interpretations under Bayesian and traditional frequentist approaches, though the analyses themselves are often very similar (Cooper and Hedges 1994). The frequentist would consider the β_i to represent a sample from a population of all β. Thus, the interpretation and inference from the random effects distribution is related towards some defined population, e.g., the relationship between CPUE and abundance for all flatfish fisheries. The Bayesian interpretation explicitly excludes the notion of a sample from a population. For a Bayesian, the random effects distribution simply represents our (the analysts) uncertainty in the process under consideration, i.e., the variance is a measure of our ignorance. But when we wish to use our random effect distribution as a prior for a subsequent analysis (Chapter 7) we have to decide whether a particular prior is relevant for a particular situation. At this stage, the Bayesian must think like a frequentist and decide (implicitly) if the studies combined in the meta-analysis are similar enough (i.e., are from the same population) to the study under consideration. So while there are a number of approaches for estimating random effects models - method of moments, analytical, maximum likelihood (Chapter 3), maximum simulated likelihood (Chapters 4 and 6), Bayesian (Chapter 5), the key factor for inference will be the population which the random effects distribution represents.

We can use random effect models to obtain 'improved' individual estimates for each data set included in a meta-analysis. It is possible to estimate the realized values

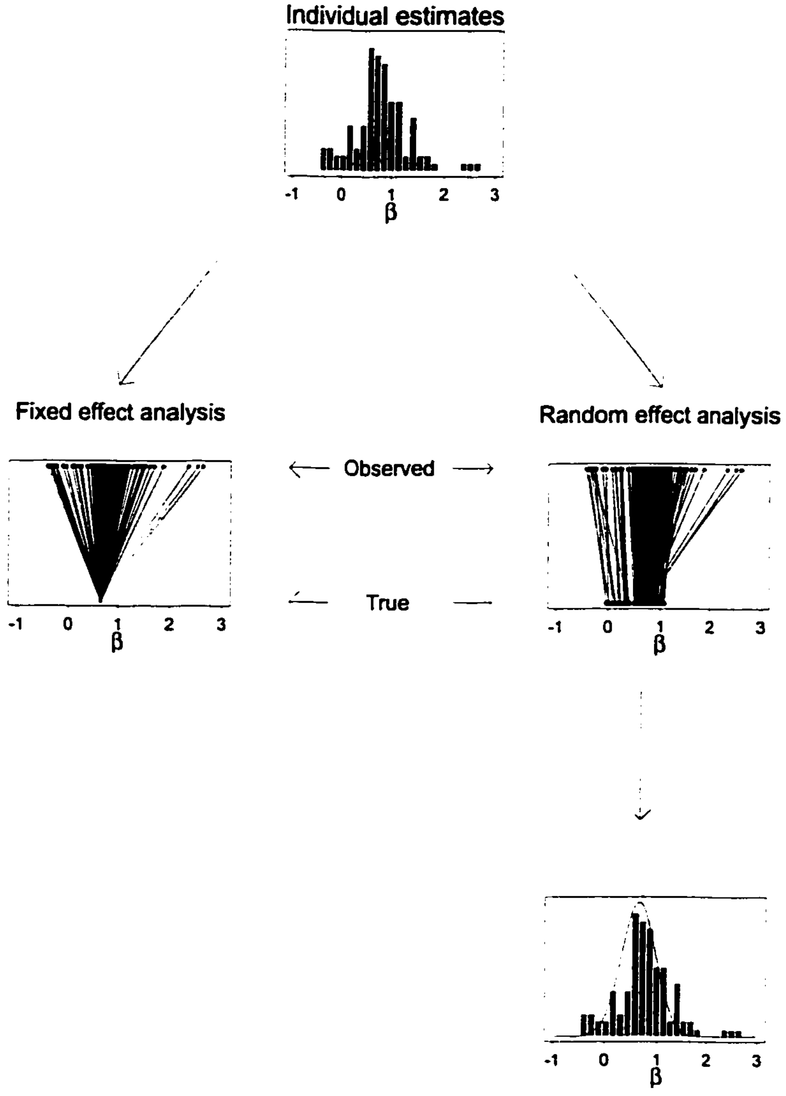


Figure 2.1: Comparison of fixed and random effects analysis of the relationship between CPUE and abundance for flatfish (from Chapter 3).

for each random effect where each study in the analysis is optimally combined with all other studies (Robinson 1991; Myers and Mertz 1998). More simply stated, the realized value is the best estimate for each study considering how well we know the original estimate for that study and borrowing information from the other studies. These new estimates are referred to as the best linear unbiased predictions (BLUPs) (Robinson 1991) and Empirical Bayes estimates (Carlin and Louis 1996) in different modelling frameworks.

In summary, with random effects models we can estimate two useful quantities: (1) the between-study variance which can be used for future analyses; and (2) individual estimates for each study, where the information for each stock and all others are combined.

To use these models we must have accurate estimates of the reliability of each estimate. These models will be meaningless unless the estimate (with uncertainty) is an accurate summary of the study (Hedges et al. 1999). If this assumption is not correct, it might be overcome if we analyze all of the studies simultaneously. This approach will be discussed in a later section, but first I will describe the different metrics that have been combined using meta-analysis.

Standardized metrics for meta-analysis

In this section I describe a range of different fisheries-metrics that have been combined using fixed or random effects models.

Experimental metrics

The metrics most commonly used in traditional meta-analysis are those derived from experimental studies. In some fisheries examples these have been applied to observational data. A common metric, the standardized difference, d, is $(\mu_1 - \mu_2)/\sigma$, where the μs are the means for each group and σ is often the within-group standard deviation. Myers et al. (1997) used this metric to test whether recruitment was lower during the period proceeding the collapse of cod stocks in the northwest Atlantic, and Riessen (1999) used it to compare the response of *Daphnia* to chemical cues released by predators.

An alternative metric is the natural logarithm of response ratio of the form $L_i = \log(R_i) = \log(\bar{X}_{i,E}) - \log(\bar{X}_{i,C})$, where the subscripts E and C refer to the experimental and control values respectively (Hedges et al. 1999). The natural logarithm is used to linearize the metric and provide a more normal sampling distribution (Hedges et al. 1999). It was used in a fixed effects framework by Micheli (1999) to indicate weak coupling between phytoplankton and herbivores in marine pelagic ecosystems.

A common metric in medical studies is the log odds ratio (DerSimonian and Laird 1986) used for the analysis of two-way tables (i.e., treatment/control and alive/dead). This metric is amenable to analyses of catch-and-release hooking mortality (e.g., Schill and Scarpella 1997) as shown in Chapter 1. Osenberg et al. (1999) describe some other metrics for ecological experiments.

Correlation coefficients

The correlation coefficient (r) is a common metric from both experimental and observational studies and is amenable to meta-analysis. To ensure that the sampling distribution of r is approximately normal, it is often transformed with Fishers' Z-transformation (Rosenthal 1994). Micheli (1999) combined correlation coefficients between nutrients and primary productivity with abundance of phytoplankton, mesozooplankton, and zooplanktivorous fish biomass. Worm and Myers (2002) used a random effects model to combine correlation coefficients (Z-transformed) between pink shrimp (*Pandalus borealis*) and Atlantic cod biomass for many regions in the North Atlantic. They inferred top-down control of shrimp abundance from the observed negative correlations.

Dimensionless numbers

Dimensionless numbers are important to meta-analysis and are often either the ratio of two values with the same units, or a power term in an equation (Myers and Mertz 1998). These have been used extensively in comparative analyses of life history characteristics (Beverton and Holt 1959; Beverton 1992; Charnov 1993; Pauly 1998) and are important in many ecological models. The basis for using these dimensionless numbers is that they are assumed to be comparable across studies or populations. A number of examples are described below.

Myers et al. (1995) examined depensation (i.e., reduced per-capita production at low population sizes) using the shape parameter, δ, in the three-parameter Beverton-Holt. Liermann and Hilborn (1997) felt that δ had no simple biological interpretation and could not be compared across populations so they used an alternative parameterization of the Beverton-Holt to provide a bounded dimensionless metric, q. They approximated the likelihood profile for q using a beta distribution and then combined these using hierarchical Bayesian methods. Both studies found no evidence for the wide-spread occurrence of depensation but acknowledged that many populations have little information on this process.

The shape parameter (β) of the power curve, $Y = aX^{\beta}$ has been frequently used in meta-analytical contexts. It has been used to measure the extent of density-dependent mortality in juvenile fishes (Myers and Cadigan 1993a) and examine density-dependence in the relationship between spawner abundance and recruitment (Cushing 1971). I will use it in Chapter 3 to examine the relationship between CPUE and abundance and in Chapter 4 to extend the models of density-dependent mortality used by Myers and Cadigan (1993a).

A quantity related to the slope at the origin of the spawner recruitment curve (described in the next section) is steepness, z, which is the proportional decline in mean recruitment in the unfished state, associated with a 80% reduction in spawner abundance (Mace and Doonan 1988). Myers et al. (1999) provided bounds for z for a range of species and Dorn (2002) combined all available estimates of steepness for rockfish (Genus *Sebastes*) in the North Pacific.

Myers et al. (1997) analyzed coefficients of delayed density-dependent mortality in sockeye salmon (*Oncorhynchus nerka*). They used an extended Ricker model with lags of density dependent mortality at years $j-1$ and $j-2$ and found little evidence of delayed density-dependent mortality.

Ecological metrics with units

Some parameters with units are amenable to meta-analysis but often must be standardized to ensure comparability across populations. In spawner recruitment data, the units of spawners and recruits vary across data sets, so the basic slope at the origin is not comparable (Myers and Mertz 1998). Myers et al. (1999) used estimates of survival and maturity to standardize the slope at the origin and provide two quantities, the average lifetime ($\hat{\alpha}$) and annual ($\tilde{\alpha}$) number of replacements spawners produced per spawner at low abundance. They showed that $\tilde{\alpha}$ was relatively constant within species and that the true variability was much less than that observed among the individual estimates.

It is possible to standardize carrying-capacity of many population models by dividing it by the area inhabited by a population (Myers et al. 2001). Other population parameters such as natural mortality and the intrinsic rate of increase are comparable across populations, but have not yet been analyzed with estimates of reliability.

Modelling the original data within a hierarchical model

There have been several attempts to simultaneously model the data for each study rather than combine estimates from individual (independent) analyses. The reliability of each data set is estimated at the same time as the variation among data sets. This has be done for spawner-recruitment data (Myers et al. 1999; Myers et al. 2001) and is described in Chapter 5 for the analysis of length-specific trawl survey catchability.

Combining the 'raw' data has several advantages. First it allows us to overcome problems when it is not possible to obtain individual estimates. It is not uncommon for algorithms determining maximum likelihood estimates to converge at meaningless values in some instances (Kehler 2001, and Chapter 3). When combining individual estimates data sets with failed convergence would be excluded. Kehler (2001) used simulations to show that this can introduce a conditional bias into the combined estimate (or random effects distribution). Second, if correlations exist between the metric of interest and other model parameters this correlation structure should be considered. Simultaneous modelling is perhaps an easier approach than combining variance-covariance matrices. Third, we need not assume that the individual estimate and its sampling variance contain all information from the data set.

One important consideration of data-based analysis is that it is not always clear how the data have been weighted in the analysis. When we combine individual estimates we input the weightings, but in the data-based analysis they can be hidden from the user because fitting methods and software are complex. It is strongly recommended that users examine carefully the weightings assigned to the different data sets.

Fully integrated approaches to meta-analysis

The final step in meta-analytical methods is to integrate fixed and random effects approaches into population dynamics models. Su et al. (2001) modelled the temporal spawning distribution for pink salmon (*Oncorhynchus orbuscha*) from Kadashan Creek and a similar approach will be described for modelling the residence time of spawning hoki (*Macruronus novaezelandiae*) in Chapter 6. In both studies the parameters describing the temporal arrival distribution (mean and variance) were treated as random effects across years. Both found that if each year was treated independently, estimates in some years were unreliable. In each case, inferences about residence time were of secondary importance to the estimation of absolute abundance, but the estimation procedure was improved though the use of the random effects framework.

This type of approach might be powerful but requires considerable data and is computationally demanding. There will be a great need for simulation studies to test these methods if complex nonlinear models are used.

Summary

While there have been many applications of meta-analysis in fisheries to date, there has been a bias toward recruitment-based analyses. This does not represent the limits of how these methods can be applied, as can be seen by the more recent applications. It is quite clear these methods have considerably increased our understanding of fish population dynamics and provided a framework for analyzing important fisheries problems.

By deriving ecologically important metrics, we have adapted standard meta-analytical approaches for experimental studies into powerful tools for combining data from different populations to attempt to separate error from true variation. Estimates of reliability for metrics are critical to the efficient use of meta-analytical methods. There has also been a progression over time from simple vote-counting procedures to many hierarchical models in which the data themselves are combined. These higher-level methods have allowed for more powerful inferences.

2.2 Important issues in meta-analysis

In this section I highlight important issues - based on my critical review of the literature and analyses presented in the following chapters - that researchers must consider when applying meta-analytic approaches to the study of fish population dynamics. Many of these issues are based upon experiences from other fields but I have used fisheries examples to help clarify these points. Also, many of the issues discussed will parallel those of single experiments but others will be more specific to meta-analysis.

Question definition

A clear and concise definition of the problem to be addressed is a critical part of any meta-analysis and this should be stated before the meta-data are collected (Gurevitch et al. 2001). Clear definitions of key terms should be provided, as slight differences in questions asked can lead to differences in conclusions (Englund et al. 1999). The question should logically lead to the data considered for the analysis.

Meta-data

The meta-data compiled for any meta-analysis is dependent on three factors: (1) the question asked; (2) the methods by which the analysts search for studies; and (3) the selection criteria used once candidate studies have been located. In the next two sections I describe important aspects of data selection. Differences between collating experimental results and fisheries data are highlighted.

Searching for data

When searching for studies, the process should be detailed to enable other researchers to locate the same studies. Protocols for this have been developed in other fields (White 1994; Cooper and Hedges 1994). Reference database search engines are ideal for meta-analysis as it is simple to provide details of the search (Reed and Baxter 1994). However, researchers generally obtain their meta-data through a number of sources (White 1994).

A variety of stock assessment model output or input has been used in meta-analysis (Myers et al. 1995; Liermann and Hilborn 1997). Collecting these types of data is fraught with three main problems: (1) stock assessments are seldom published in the primary literature and therefore often limited to personal contacts of the researchers; (2) the quality of the assessment data will vary as peer review of assessments is limited; (3) data may be available, but is not always presented in a useable form.

There are two large-scale databases that are of general use in meta-analytical approaches to fisheries, FishBase[1] (Froese and Pauly 2000), and Ransom Myers' Stock Recruitment database[2]. FishBase contains biological, ecological, and landings information for over 25,000 species. Myers' database contains over 700 multivariate time series of spawner abundance, recruitment, catches, and fishing mortality. It also contains other biological parameters (e.g., natural mortality and age-at-maturity) required to allow construction of standardized metrics (e.g., the maximum reproductive rate). Unfortunately, FishBase does not provide estimates of the reliability for the biological parameters and the Myers database lacks the necessary data to standardize the spawner-recruitment data for many of the populations.

Selection

Once candidate studies or populations have been located it must be decided if they will be retained for the analysis. This can be based upon a re-evaluation of the relevance or study quality (e.g., sample size) (Englund et al. 1999; Gurevitch et al. 2001). Below I provide some examples of exclusion criteria that have been used.

In Chapter 5 I describe the analysis of trawl survey catchability. Initially, 60 catchability-at-length data sets were found to meet the initial selection criterion but subsequently 13 were excluded from the analysis. Data sets were excluded for three reasons: (1) data were not considered representative for inferences for the Scotian Shelf; (2) estimates were considered unreasonable, i.e., extreme outliers; and (3)

---

[1] Available in two forms, either on CD-ROM, or online at www.fishbase.org

[2] Available online at http://fish.dal.ca/~myers/welcome.html

some catchability-at-length series were dome-shaped and inconsistent with the logistic model. The first criterion was justified based on the intended inferences, the second was more subjective but post-hoc analyses indicated that these extreme values were well above the upper 90% confidence limits of the remaining data, while the third may have limited the acceptance of the results.

Myers et al. (1999) excluded SR data for coho salmon because they were inconsistent with the Ricker SR curve that was used in the analysis. Hinrichsen (2001) excluded 100 (of 300) potential spawner recruitment data sets because of high levels of autocorrelation.

A common selection criterion is the duration of an experiment (for experimental data) or the length of the time series of the data (for population data). In the analysis of experimental studies, it is important to ensure that the time frame for each experiment is consistent with the ecological process that one is examining (Brett and Goldman 1997). In a non-fisheries application, Hassell et al. (1989) showed that many studies that failed to detect density-dependence were run over a shorter time period than the 'successful' studies. Gilbert (1997) used the minimum number of data points required to fit his model as a cutoff. For the analysis in Chapter 3, only time series with more than 9 paired CPUE / survey observations were used, based on empirical evidence (Dunn et al. 2000) and simulation studies (Myers and Cadigan 1993a) that suggested that shorter series may not produce reliable results. If criteria can be determined *a priori*, conclusions may be more convincing. Alternatively, statistical modelling of temporal effects is likely to be complicated (due to threshold effects etc.).

Exclusion of data based upon subjective considerations of the data should be avoided (Englund et al. 1999). Large-scale discarding of data could potentially reduce the generality of the results. It is recommended to include all available data but explicitly consider the reliability of each (Cooper and Hedges 1994; Turek and Brett 1997; Gurevitch et al. 2001). Otherwise, selection criteria must be explicitly stated to allow researchers to evaluate potential biases of the meta-analysis.

Representativeness

A common reason for the findings of a meta-analysis to be ignored is that the studies included were not considered to be representative by other researchers (Francis 1997). If the focus is too narrow the results will be only of limited interest and may lead to inappropriate extrapolation, but if the focus is too broad the results could simply become uninteresting (Cooper and Hedges 1994; Gurevitch et al. 2001). There are statistical guidelines for whether studies should be combined or not. If the between-study variation is many times greater than the within-study variation, combination of the data sets is questionable (Gurevitch et al. 2001). For the analysis of trawl survey catchability in Chapter 5 inferences were limited to seasonal estimates of catchability at length for a region with a particular seafloor structure. This narrow focus is ideal for the application to the Scotian Shelf but may be not useful for other situations.

The issue of representativeness is particularly important when combining metrics across species and populations. Before choosing representative stocks for an analysis we must carefully consider taxonomy, life history characteristics, and ecology (Hilborn and Liermann 1998). This problem may be compounded if those researchers with sufficient quantitative skills to use the appropriate methods for meta-analysis, are not those with sufficient understanding of the biology and ecology of the species of interest. For example, within the tunas there is considerable variation among life history characteristics (Vetter 1988; Kearney 1991) but often tunas and mackerels are analyzed together at the family level (Scombridae) (Beverton 1992; Myers et al. 1999). Dorn (2002) combined all available spawner-recruitment data for a diverse group of rockfish (Genus *Sebastes*) in the North Pacific.

Meta-data biases

By the time studies have been selected a number of biases may already be present in the meta-data. Three of these are described below.

Publication bias

Publication bias is one of the most well known meta-analytical biases and occurs because non-statistically significant results are published less frequently than those with statistically significant findings (Greenhouse and Iyengar 1994; Gurevitch and Hedges 1999). Below I provide three fisheries examples of publication bias.

Myers (1998) recognized that publication bias significantly limited his examination of environmental variables for the prediction of recruitment. To be considered, a previously published environmental recruitment relationship had to be retested using updated data (Myers 1998). Researchers who retested an existing relationship and found it significant (or not), might not publish the results or have the paper rejected (for simply confirming a previously published finding or trying to publish a non-significant finding). In Chapter 3 I discuss how potential data on the relationship between CPUE and abundance are not presented or used in the assessments. Beverton (1992) noted that maximum age information may suffer from a form of publication bias (though not identified as such) whereby the data for older ages are accumulated into a plus group. This may hinder the use of some commonly used empirical relationships (e.g., Hoenig 1983) that use this information to predict natural mortality.

Several graphical and statistical methods exist to detect publication bias and assess the sensitivity of results to it (Greenhouse and Iyengar 1994). By plotting the effect size versus the sample size, publication bias is often indicated by the lack of studies presenting negligible effects with low sample sizes (Gurevitch et al. 2001). Commonly in fisheries applications the raw data is available (Drinkwater and Myers 1987). In such circumstances, one could examine the *potential* consequences of publication bias by examining the sensitivity of the original results to those of different meta-data sets, where alternative selection criteria were used.

Population bias

A consequence of what Gurevitch and Hedges (1999) termed 'research bias' is *population bias* (Hilborn and Liermann 1998). This occurs because data on fish populations are not collected randomly. Often data collection is biased toward: (1) exploited populations; (2) populations of sufficient economic interest to warrant research; and (3) populations managed sufficiently well (or are resilient enough) to allow enough data to be collected and reported. For example, Myers' spawner-recruitment database is limited to species where age-structured data is available. The database lacks information on small tropical fishes that make up a large number of the world's fish species.

The consequences of this bias can range from negligible to severe when the results of a meta-analysis are applied to a particular population. Ricker (1973) showed that if substocks among salmon population complexes have variable resilience to overfishing it is possible for some stocks to be overfished. If the substocks become extinct before sufficient data have been collected, we would overestimate the general range of levels of resilience. Given this, we must be cautious in applying the results of meta-analyses of stocks with long exploitation histories, to newly exploited populations. In general we must be sure that the species or populations included in an analysis are truly representative of any population to which we would like to apply our results.

Conditional bias

Obtaining sensible parameter estimates can be difficult for nonlinear and complex linear models (Barrowman and Myers 2000; Kehler 2001). Kehler (2001) used data- and simulation-based analyses to show that a conditional bias can influence a meta-analysis if these estimates are ignored or not considered appropriately. In simulation studies, when the raw SR data were modelled without those data sets with unreasonable estimates, he found that the overall mean slope at the origin was underestimated (Kehler 2001). In this example, the modelling of all the data partly overcame these problems.

This bias is likely to have influenced the quantitative conclusions made in Chapter 3. It was found that one-third of the individual analyses gave inadmissible variance components estimates (negative) so the estimates of the shape parameter were not included in the random effects analysis.

The direction of conditional bias will depend on the analysis and in particular the reason for the non-convergence. This problem will generally occur only when nonlinear models are being used and it might be overcome by modelling the individual data sets together.

The analysis

In this section I focus on two important aspects of the analysis: first, the consideration of the reliability of different studies; and second, the importance of recognizing the different forms of non-independence that can exist among studies.

Weighting data sets

In order to use meta-analytical methods such as fixed and random effects models we require estimates of the reliability of the metrics we wish to combine. Researchers who combine studies without weightings will overestimate the true variation and may come to incorrect conclusions (Myers and Mertz 1998). Without this information we are restricted to simple analyses such as the sign test or may inappropriately use standard statistical methods (e.g., GLM's). When time series data are being used for the combination of correlation coefficients it is critical that the autocorrelation be accounted for (Micheli 1999; Worm and Myers 2002). This autocorrelation results in an over-estimation of the true degrees of freedom and can lead to improper weightings by underestimating the uncertainty of the individual correlation coefficients.

Often estimates of reliability are the estimated standard error for the metric but in some instances the entire likelihood profile is used to represent the uncertainty (Liermann and Hilborn 1997). Whichever is used it should provide a good estimate of the reliability of the individual estimates. For example, if a researcher feels that the standard error does not accurately reflect reliability, a further weighting (Shadish and Haddock 1994) or categorical label (Orwin 1994; Stock 1994) can be used. This latter approach has been used widely in fisheries applications. Turek and Brett (1997) used further reliability scalars for catch-and-release mortality studies that did not provide full details of the experimental procedures. In the analysis of the relationship between CPUE and abundance I consider separately (for sensitivity) CPUE series associated with either single or multiple trawl surveys. Both Myers (1997) and Gilbert (1997) repeated analyses including SR data sets of potentially differing quality (lengths of SR series). Such an approach is strongly recommended as it illustrates the robustness of the findings of the meta-analysis (Englund et al. 1999).

When simultaneously modelling the individual data for each study, weightings are often estimated 'internally' (Myers et al. 1999). Post-hoc analyses of weightings will lead to improved understanding of meta-analysis.

The weighting of different data sets is critical when carrying out meta-analysis. The initial selection criteria cannot be overlooked, because this is when many data sets are given a weight of zero. Multiple analyses using different weightings can illustrate the robustness of the analysis as can identification of features of the individual data sets related to the weightings.

Non-independence

Studies must be similar before they warrant combining, but there could be problems if some are more similar than others. This is especially the case when inferences are to be made to a larger group. Non-independence among studies or populations will generally lead to underestimation of the standard errors for the mean effect (Gurevitch and Hedges 1999) and the estimate of between-study variation. Below I discuss some common causes of non-independence.

Within-study selection

The most obvious form of non-independence among meta-data is multiple results from a single analysis or experiment. General protocols for this are lacking and different decisions can influence the conclusions of a meta-analysis (Turek and Brett 1997; Englund et al. 1999).

It does not appear reasonable to assume that a single publication represents a single data point. Some researchers 'lump', presenting many results in a single publication, others 'split', having a separate publication for each result, and some sets of results can be published in more than one publication (Gurevitch et al. 2001). In experimental studies, researchers using cross-treatment designs can produce multiple estimates of an effect size (Gurevitch et al. 2001). In stock assessment sometimes more than one analysis is performed for a single population (Horn et al. 2000).

Methods and researchers

When considering experimental results one must consider non-independence such as multiple studies from the same laboratory or analyses using the same methods

(Gurevitch and Hedges 1999). This lack of independence might apply to metrics obtained through particular statistical or stock assessment methodologies.

Spatial correlation

Spatial autocorrelation is often important when examining population metrics. If important processes act at a scale greater than the individual population or study (e.g., El Nino or a mixed fishery), populations closer together could be influenced by the same factors. Myers et al. (1997) included an empirically based weighting function which considered the distance between rivers to account for correlation in their analysis of sockeye salmon dynamics.

Taxonomic considerations

If an analysis involves combining results across species, one must consider non-independence that may be introduced if differences are smaller among more closely related species (Harvey and Page 1991; Charnov 1993; Gurevitch and Hedges 1999). In any meta-analysis for species grouped based on taxonomy (e.g., at the level of order or family), it will be important to consider the effects of similarities based on phylogenetic closeness. As species or populations are never selected at random, the analyst should consider the consequences of this lack of representativeness on the generality of the results.

There are a number of levels at which non-independence exist among studies. While it is recommended that statistical approaches be used to model the between-study correlation structure (Myers et al. 1997; Gurevitch and Hedges 1999), there have been few examples where this has been done. Not recognizing non-independence can lead to underestimation of the true uncertainty and increased Type II errors (rejecting the null hypothesis too frequently) (Gurevitch and Hedges 1999).

Inferences and generalizations

The inferences from a meta-analysis should be based on two factors: (1) the studies included in the analysis; and (2) the information available from each study. There are also other more technical issues related to our inferences.

If the reliability of the different estimates is not known our potential inferences are very limited. However, if the reliability is known and can be modelled we can make inferences about the combined estimates and whether there is variation between-studies. Earlier I discussed the importance of data selection criteria and decisions on representative studies, both limit how widely the results can be generalized.

Often a meta-analysis will report results at different taxonomic levels and it is not always clear which results should be used. Obviously if there is sufficient data at lower taxonomic levels, then these should be used, but determining what is sufficient is difficult. Myers et al. (1999) provided estimates of the maximum reproductive rate at the species and family levels but cautioned against the application of family level estimates. In some instances these appeared to clearly overestimate the range of values for a species (e.g., Silver hake, *Merluccius bilinearis*). In Chapter 5 I model trawl survey catchability for Atlantic cod and haddock both separately and together. While I find clear differences between the two species, the combined analysis allow inferences for other demersal gadoid-like species for which no other information was available.

When examining experimental results it is important to be aware of what inferences are possible. Taylor and White (1992) combined estimates of catch-and-release mortality for barbed and barbless hooks from independent studies. Though it may be appropriate to make inferences about each (e.g., the probability of mortality for each hook type), they inappropriately obtained an estimate of the ratio of mortality caused by each (Turek and Brett 1997; Taylor and White 1997; Schill and Scarpella 1997). A more appropriate comparison was made by Schill and Scarpella (1997) and used as an example in Chapter 1.

Estimates of between-study variation from meta-analysis are commonly recommended for use as prior distributions for Bayesian population models (Hilborn and Liermann 1998; Harley et al. 2001). If prior distributions are overly precise (i.e., too narrow) they will overwhelm the information contained in the data (Punt and Hilborn 1997). The methods used by Myers et al. (1999) and in Chapter 3 provide maximum likelihood estimates of the prior distribution. Such prior distributions might be too narrow and restrictive. A more appropriate approach would be to simulate points using the variance covariance matrix of the mean and the variance of the random effects distribution, and then use the distribution of simulated points as a prior (Chapter 5).

This explicitly accounts for the uncertainty in estimating the parameters of the random effects distribution. A potentially serious problem with prior distributions is the use of independent priors for correlated variables. Myers et al. (1999) provided priors for the slope at the origin while Liermann and Hilborn (1997) provided priors for the level of depensation. These two parameters are highly correlated in any SR data set and the use of independent priors will lead to serious biases.

The inferences obtained from meta-analyses must be carefully considered. It is important to recognize the limitations of analyses in which estimates of reliability are unavailable. If multiple estimates (or priors) are provided at different levels it is recommended that all levels be considered to assess sensitivity.

Chapter 3

The relationship between CPUE and abundance

3.1 Introduction

Despite its well-documented shortcomings, the catch-per-unit-effort (CPUE) from many commercial and recreational fisheries is commonly used in the assessment of fish populations (Hilborn and Walters 1992; Gillis and Peterman 1998). It is generally assumed that the CPUE abundance indices are proportional to the true abundance of the population. However, it has long been recognized that this may be untrue (Beverton and Holt 1957). If CPUE indices are interpreted (and modelled) incorrectly, estimates of fishing mortality and abundance could be biased. As it is seldom possible to accurately determine the relationship between CPUE and abundance from a single data set (Collie and Walters 1991), meta-analysis is required. Using a meta-analytical approach I will show that CPUE generally declines more slowly than the abundance of the population being fished.

Paloheimo and Dickie (1964) reviewed the early use of CPUE data in fisheries and stressed the importance of understanding both the spatial distribution of fish and the allocation of fishing effort to interpret CPUE data. Possibly the first use of CPUE data in a fisheries setting is attributed to Baranov (1918) who incorporated catch data in the analysis of herring (Seber 1982). Reasons why CPUE might not be proportional to abundance have been investigated by simulation (Sampson 1990;

Gillis and Peterman 1998), and through examination of empirical data (Peterman and Steer 1981; Crecco and Overholtz 1990; Rose and Leggett 1991; Swain and Sinclair 1994). Quinn (1985) and Quinn and Collie (1990) both assumed the CPUE was not proportional to abundance, assuming a square-root relationship.

The model of proportionality between CPUE (U) and abundance N at time t is,

$$U_t = qN_t, \tag{3.1}$$

where q is the catchability coefficient. A number of non-linear models have been proposed, the simplest being the power curve,

$$U_t = qN_t^{\beta}, \tag{3.2}$$

where if $\beta = 1$ the model reduces to Eq. 3.1, and if $\beta \neq 1$ then catchability changes with abundance (Figure 3.1). When $\beta > 1$, U declines faster than N in a situation known as *hyperdepletion*. Conversely if $\beta < 1$, U declines slower than N which results in *hyperstability*. This model does have the property that when $\beta \neq 1$, catchability is no longer given by q, but rather a function of q and β (catchability $= qN^{\beta-1}$). Other non-linear models have been proposed. Bannerot and Austin (1983) allowed non-linearity in both effort and abundance, while Richards and Schnute (1986) proposed a very general and flexible range of three-parameter models.

Here I am only interested in the shape of the relationship between CPUE, so the traditional catchability parameter, q, is not considered. The shape parameter of the power curve, β, is the focus of this analysis as this non-dimensional parameter describes the shape of the relationship between CPUE and abundance. As β is a dimensionless number, it is more amenable to meta-analysis. I will only consider the power curve in this analysis, as it requires the estimation of only one additional parameter and has the additional property that it can be written as a linear model by simply taking the logarithm of both sides (shown later). Linear models have better understood properties than nonlinear models and the parameters of linear models are generally easier to estimate. By considering only the power model, I recognize that I am not considering alternative functional forms for the relationship between CPUE and abundance.

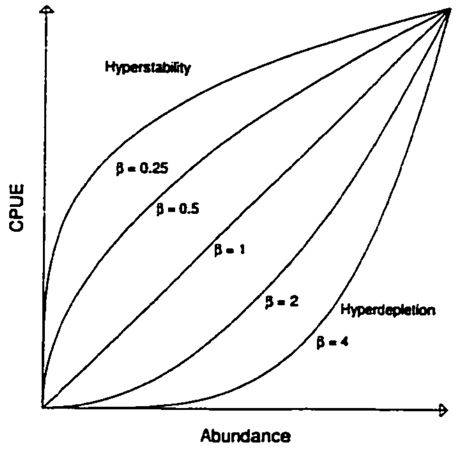


Figure 3.1: Relationship between catch-per-unit-effort (CPUE) and abundance for the power model with different values of the shape parameter β.

One important component of Eq. 3.2 is true abundance, N_t. Without large-scale depletion experiments it is very unlikely that we will ever know (without considerable error) absolute abundance. We must use some proxy or relative abundance index that is proportional to true abundance. It is important to note that often CPUE indices are used in such cases, but this is the assumption that is being tested! Research survey abundance indices are generally used to test the proportionality assumption of CPUE. I will show later that uncertainty in the research survey abundance indices requires more complex statistical models.

While many studies have focused on CPUE for a single stock or fish community, there have been few attempts to compile data across fisheries and examine the problem in a broader context. Dunn et al. (2000) examined the relationship between CPUE and abundance for all available New Zealand data sets as well as some from stocks from Atlantic Canada. In general, it was found that the small numbers of paired CPUE/survey observations lead to considerable difficulties in estimating β. They also compared CPUE indices with model estimates of abundance (that had not used CPUE data in the tuning process) as this led to longer time series. Simulations showed that with few paired observations (4–8), a very large (> 50%) reduction in biomass was required to accurately estimate β. When $\beta = 1$, a reduction of up to 75% was required to distinguish between proportional, hyperstable and hyperdepleted relationships. When the reduction in abundance was low (12.5–25%), the estimated values, $\hat{\beta}$, tended to 1, regardless of the true β.

I attempted to compile many CPUE/survey data sets from around the world and then estimate the shape parameter of the power curve. These individual estimates could then be combined using random effect meta-analysis.

3.2 Methods

There are three parts to this meta-analysis that I will describe here: (1) the data requirements for the analysis; (2) the linear mixed effects models with observation error necessary to estimate the relationship between CPUE and abundance; and (3) the random effects meta-analysis used to combine estimates across the many populations.

3.2.1 CPUE and abundance data

For this study I require estimates of fishery CPUE and fishery-independent estimates of true abundance. I use research survey abundance estimates as the fishery independent estimate of abundance. It could also be possible to use biomass estimates from a stock-assessment model that did not use CPUE data, but I have not considered this potential data source. With the procedures used here (described in sections below), I require a CPUE time series for a population for which there is also available either: (1) one research survey time series with a reliable estimate of its sampling error, (2) multiple research surveys, or (3) a single "reliable" survey series and multiple CPUE series. I restricted this analysis to data of types (2) and (3) as this did not require assumptions about the observation error of the survey series.

International Council for the Exploration of the Sea (ICES) stock assessment reports offer a standard format and data from a number of fisheries. The large number of trawl surveys conducted in the ICES region and the length of these series made it possible to obtain a larger number of data sets for a wider range of species than considered by Dunn et al. (2000) (Table 3.1). The CPUE and survey indices were obtained from the data input files for the stock assessments conducted in 1999. In each case, the CPUE indices were chosen by the experts in each ICES assessment committee as a series representative of the stock status. In some cases the CPUE indices were corrected for fishing power, e.g., horsepower, and sometimes the assessment scientists did not believe it was necessary. The ICES CPUE indices represent data that are typical of that used by experienced assessment biologists. Details are given in each of the assessment reports available from ICES and data files can be obtained directly from ICES website (http://www.ices.dk/committe/acfm/acfm.htm).

Table 3.1: Details of catch-per-unit-effort (CPUE) data, by species and stock, used in the meta-analysis: number of CPUE series per stock, ages covered, length of the time series, and whether more than one trawl survey was available. ICES = International Council for the Exploration of the Sea; RV = Research vessel.

Species/Stock	**CPUE series**	**Ages**	**Years**	**Multiple RV surveys**
Blue whiting (*Micromesistius poutassou*)				
Northeast Atlantic	1	1–5	12	Yes
Cod (*Gadus morhua*)				
ICES IV & IIIa (North Sea and Skaggerak)	5	1–5	16	Yes
ICES VIIe-k (Western Channel)	3	1–5	10	No
ICES IIIa (Kattegat)	3	2–5	12	No
Four spot megrim (*Lepidorhombus boscii*)				
ICES VIIIc & IXa (West Iberian Shelf)	2	1–7	11	No
Haddock (*Melanogrammus aeglefinus*)				
ICES IV (North Sea)	2	0-5	10	Yes
ICES VIb (Rockall Bank)	3	2–7	10	No
Hake (*Merluccius merluccius*)				
ICES IIIa, IV, VI, VII, & VIIIab (North)	4	2–5	12	No
Megrim (*Lepidorhombus whiffiagonis*)				
ICES VII & VIIIab (Bay of Biscay)	1	2–9	9	Yes
ICES VIIIc & IXa (West Iberian Shelf)	2	1–7	10	No
Plaice (*Pleuronectes platessa*)				

Table 3.1: CPUE data, continued.

Species/Stock	CPUE series	Ages	Years	Multiple RV surveys
ICES VIIe (Western Channel)	3	3–5	10	No
ICES VIIfg (Celtic Sea)	3	2–5	9	No
ICES VIIa (Irish Sea)	2	2–4	10	No
ICES VIId (English Channel)	3	2–5	9–10	Yes
Saithe (*Pollachius virens*)				
ICES IV & IIIa (North Sea and Skaggerak)	3	3–7	9	No
Sole (*Solea vulgaris*)				
ICES VIIa (Irish Sea)	2	3–9	10–11	No
ICES VIId (English Channel)	2	2–4	11	No
ICES IV (North Sea)	2	2–4	13	Yes
ICES VIIfg (Celtic Sea)	2	2–5	11	No
ICES VIIIab (Bay of Biscay)	2	1–8	11	No
ICES VIIe (Western Channel)	2	3–6	15	No
Whiting (*Merlangius merlangus*)				
ICES IV (North Sea)	3	2–5	16	Yes
ICES VIa (West of Scotland)	3	2–6	11	No
ICES VIIe-k (Celtic Sea)	2	1–6	12	No

Table 3.2: Gear types of the catch-per-unit-effort series.

Code	Gear type	Data sets
BeTrawl	Beam trawl	55
DSeine	Danish seine	32
LTrawl	Light trawl	32
NTrawl	Shrimp trawls	20
OTrawl	Otter trawl	12
PTrawl	Pair trawl	5
Trawl	Unspecified trawl	142

ICES data are usually presented in an age-disaggregated form resulting in the large number of data sets available. One limitation of using published data is that in many cases data were not presented for all ages, e.g., survey data for ages 1–6 and CPUE data for ages 4–9. In such cases, I only used the ages where there was overlap. Also, there were instances where there was no overlap, when survey data were used for younger ages and CPUE for the older ages. Thus, to a certain extent I was limited by which data were readily available. I restricted myself to data sets for which there were at least nine paired observations of CPUE and survey abundance because shorter series are unlikely to produce reliable estimates (Myers and Cadigan 1993a; Dunn et al. 2000). This did not result in the exclusion of many data sets as shorter time series were seldom used in the ICES assessments.

The CPUE series represented a range of different gear types or fleets (Table 3.2). For the majority of the trawl data, the specific trawl type was not given. Most gears involve a single vessel towing a net of some type. The Danish seine gear is the only gear that is not towed, rather it is set and then hauled, and the pair trawl gear represents a net jointly towed by two vessels.

3.2.2 Observation error models

PROC CALIS (Covariance Analysis and LInear Structural equations) estimates parameters of linear structural equations using covariance structure analysis (SAS Insitute Inc. 1990). This is a general framework that can include observation error models as a special case. To obtain individual population estimates of β I use the power curve,

$$U_t = qN_t^\beta \exp(\epsilon_t).$$

I take the logarithm of the different components such that $n_t = \log(N_t)$ and $u_t = \log(U_t)$. An important assumption that I use is normality of the independent variable, so we have $n_t \sim \mathrm{N}(\mu, \phi)$, which states that the true abundances, N_t, are independently lognormally distributed.

As we do not actually observe the true abundances, rather estimates from research trawl surveys, the observation error in n_t can be incorporated into the model by

$$o_t = n_t + \zeta_t, \quad \text{and}$$

$$o_t \sim \mathrm{N}(\mu, \phi + \sigma_\zeta^2)$$

where o_t is the observed value of n_t from the survey, and the observation error, is assumed to be normal, i.e., $\zeta_t \sim \mathrm{N}(0, \sigma_\zeta^2)$. For CPUE we can write

$$u_t = \log(q) + \beta n_t + \epsilon_t,$$

$$u_t \sim \mathrm{N}(\log(q) + \beta\mu, \beta^2\phi + \sigma^2),$$

where β is the slope of the relationship between the logarithm of CPUE and the logarithm of the abundance, and σ^2 is the estimation error variance. From these assumptions, the theoretical variance-covariance matrix of the joint distribution of the CPUE series (u_t) and the observed survey series (o_t) is,

$$\Sigma = \begin{bmatrix} \phi + \sigma_\zeta^2 & \beta\phi \\ & \beta^2\phi + \sigma^2 \end{bmatrix}, \tag{3.3}$$

where the parameters to be estimated are β, ϕ, σ_ζ^2, and σ^2. This problem is not identified, i.e., unique solutions do not exist, unless one of the four parameters can be estimated independently. This is because there are four unknown parameters but only three elements of $\mathbf{\Sigma}$. This example represents a standard structural regression problem.

The identification problem is overcome if we have multiple surveys or multiple CPUE series. If we have two surveys, and with the same assumptions as before, the survey estimates are,

$$o_{t,1} = n_t + \zeta_{t,1} \text{ where } \zeta_{t,1} \sim \mathrm{N}(0, \sigma^2_{\zeta_1}),$$

$$o_{t,2} = n_t + \zeta_{t,2} \text{ where } \zeta_{t,2} \sim \mathrm{N}(0, \sigma^2_{\zeta_2}),$$

and the theoretical variance-covariance matrix of the CPUE and two survey series is,

$$\mathbf{\Sigma} = \begin{bmatrix} \phi + \sigma^2_{\zeta_1} & \phi & \beta\phi \\ & \phi + \sigma^2_{\zeta_2} & \beta\phi \\ & & \beta^2\phi + \sigma^2 \end{bmatrix}$$

where the parameters to be estimated are $\sigma^2_{\zeta_1}$, $\sigma^2_{\zeta_2}$, σ^2, β, and ϕ. As there are less free parameters than unique elements of the observed variance-covariance matrix, the model is identified (more details on rules for determining whether a model is identified are discussed by Bollen 1989). It is simple to extend $\mathbf{\Sigma}$ for the case of more than two surveys, and the case of multiple CPUE series and a single survey series. The case of two CPUE series and one survey series is exactly identified. I used all trawl surveys that were available (and used in the stock assessment) for each stock (Table 3.1).

I used maximum likelihood methods to estimate the parameters of $\mathbf{\Sigma}$ (details of the methods can be found in Bollen 1989). The basic estimation procedure is to fit the observed variance-covariance matrix with the theoretical one. The fitting function that is minimized in order to find the maximum likelihood estimates is

$$\log|\mathbf{\Sigma}| + tr(\mathbf{S}\mathbf{\Sigma}^{-1}) - \log(|\mathbf{S}|) - d, \tag{3.4}$$

where $\mathbf{S}$ is the observed variance-covariance matrix, tr is the matrix trace and d is the dimension of $\mathbf{S}$ (Bollen 1989). Maximum likelihood does not always produce sensible results; i.e., the parameters describing the variance can be estimated to be less than zero if they are not constrained. This often happens with short time series in which the sample variance is not an adequate approximation to the true variance. Any model in which at least one of the estimates of the variance parameters had to be constrained to be greater than zero, or whose estimates were very small was rejected. The robustness of the results to this assumption is described later as are alternative methodologies to overcome this problem. Standard errors are estimated using the inverse of the Hessian matrix.

3.2.3 Meta-analysis of individual parameter estimates

In Chapter 2 I described different techniques for combining estimates of effect size. Here I will use a random effects model to combine estimates of the shape parameter. I prefer the random effects model over the fixed effects model because I do not expect the relationship between CPUE and abundance to be the same for all fisheries; rather there will be other factors relating to fish and fisher behaviour.

Considering the shape parameter β, I assume that each of the true unobserved β_is are normally distributed, that is $\beta_i \sim \mathrm{N}(\mu_\beta, \sigma_\beta^2)$. I also assume each of the true values for β_i are estimated with error. I assume normal observation error, $\hat{\beta}_i \sim \mathrm{N}(\beta_i, v_i)$, where v_i is the estimated variance, $(\widehat{\mathrm{SE}(\beta_i)})^2$. Finally, the variability observed in $\hat{\beta}_i$ (i.e., all of the individual estimates) can be written as a sum of two variances,

$$\mathrm{var}(\hat{\beta}_i) = \sigma_\beta^2 + v_i.$$

I follow the maximum likelihood approach of Hardy and Thompson (1996) for combining the estimates of β_i. This approach allows the estimation of both the underlying distribution from which the β_i's are drawn and the profile log likelihood which can be used to calculate confidence limits. The form of the joint log likelihood in terms of μ_β and σ_β^2 is,

$$l(\mu_\beta, \sigma_\beta^2) = -\sum_{i=1}^{k} \frac{1}{2} \log 2\pi(\sigma_\beta^2 + v_i) - \sum_{i=1}^{k} \left(\frac{(\hat{\beta}_i - \mu_\beta)^2}{2(\sigma_\beta^2 + v_i)} \right). \tag{3.5}$$

I also used this method for the estimation error variance of the relationship between CPUE and abundance, (σ_i^2). To meet the assumption of normality, I assumed that $\log \sigma_i$ was a normal random variable (i.e., $\log \sigma_i \sim \mathrm{N}(\mu_{\log \sigma}, \sigma^2_{\log \sigma})$). I also assumed that the errors were normal so that $\widehat{\log \sigma_i} \sim \mathrm{N}(\log \sigma_i, \omega_i)$. I derived approximate estimates of ω_i from $(\widehat{\mathrm{SE}(\sigma_i^2)})$ using the delta method (Kendall et al. 1987). An advantage of using $\log \sigma$ is that σ is approximately equal to the coefficient of variation of the series under reasonable conditions. For this reason I will discuss the results in terms of σ.

3.3 Results

Large fluctuations in the CPUE and survey indices were observed. The median relative range (maximum / minimum CPUE index) in CPUE was approximately 12-fold, and the 95th percentile was 0.69. This suggests that there is enough contrast in the data to estimate β. It was also found that the lowest point in a CPUE series was just as likely to be the first or last year, although overall the median decline in log CPUE was 5% per year.

While it is not possible to display all data used in the analysis, nine randomly selected data sets from the 297 considered in the analysis (using a random number table), are presented in Figure 3.2. Seven of the nine series had estimates of $\beta < 1$ from the maximum likelihood structural models. Series (d) exhibited hyperdepletion, and series (g) did not provide feasible results (i.e., negative variance estimates as described below). From an examination of the data, there is no evidence for more complex models though for any data set a different model may fit better.

Of the 297 data sets that were analyzed, 209 were retained for the subsequent meta-analysis. That is, data sets were excluded if the maximum likelihood estimates for σ^2 were negative, or if the estimates for ϕ were either negative or very small ($<$ 0.06). In the case where ϕ was very small, there was very little (estimated) variation in true abundance, and thus it was not possible to obtain accurate estimates of how CPUE changed with abundance (i.e., regression with estimates of the dependent variable at only a single value of the independent variable) (Figure 3.3). Sensitivity to this cut-off criterion is discussed at the end of the results section and was investigated

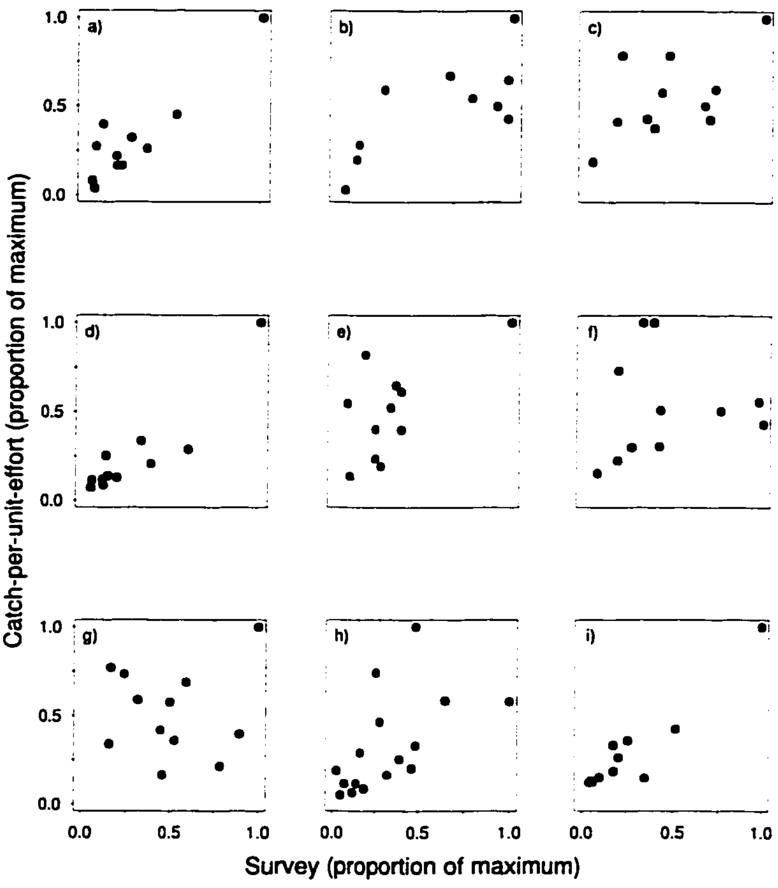


Figure 3.2: Nine randomly selected series of catch-per-unit-effort (CPUE) and associated series of trawl survey abundance. Both series are scaled to a maximum of 1. Estimates of $\beta < 1$, implying hyperstability, are associated with data sets with the majority of points above the one to one line (dashed) that is provided for reference. (a) Age four sole, Irish Sea, trawl, $\hat{\beta} = 0.73$; (b) Age three megrim, ICES VIIIc and IX, trawl, $\hat{\beta} = 0.87$, (c) Age four whiting, ICES VIa, Danish seine, $\hat{\beta} = 0.38$, (d) Age six sole, Irish Sea, beam trawl, $\hat{\beta} = 1.20$, (e) Age five sole, ICES VIID, beam trawl, $\hat{\beta} = 0.66$, (f) Age three whiting, ICES VIa, Danish seine, $\hat{\beta} = 0.80$, (g) Age four blue whiting, Northeast Atlantic, pair trawl, $\hat{\beta} = -21.70$, (h) Age three plaice, ICES VIID, beam trawl, $\hat{\beta} = 0.53$, (i) Age four sole, Irish Sea, beam trawl, $\hat{\beta} = 0.62$.

by the simulations of Myers and Cadigan (1993a).

I compared the data series for which reasonable estimates of β were obtained and those that were excluded. No differences were found in the length of series or contrast in the CPUE or survey indices. It was found, however, that the availability of multiple surveys was different between the two groups. For data sets where multiple surveys were available 87% of the series gave acceptable results (positive variances) versus 67% for those without. This finding is not surprising. I then compared the results for those data sets that gave acceptable results on the basis of whether there were multiple surveys or not. There was evidence (but not statistically significant by an analysis of medians) that the individual estimates of β were greater for those series without multiple surveys (median of 0.84 versus 0.67). However, the standard errors were also higher indicating that these estimates were less precise. Overall, the estimation error variance estimates were lower and less variable for the series with multiple surveys.

Except for data for the single hake stock, the distribution of the shape parameter, β, did not differ significantly across the species (Figure 3.4) and most of the $\hat{\beta}_i$ were less than 1. Estimates for hake were mostly greater than 1, and considerably different from the other species considered. This might be a consequence of having survey indices which are not proportional to abundance. The area inhabited by this stock was greater than most others included in the analysis. The survey covered only a small part of the entire stock area. For species other than hake, the upper quartile of the distribution of β was generally less than one and the upper 95% confidence limit for the median was also less than 1. Estimates for the youngest (0) and oldest ages (8 and 9) were variable but there were few observations for these ages. There is some evidence that the upper quartile increases with age, but any trend in median $\hat{\beta}$ at age is unclear (Figure 3.4). There appear to be few differences that can be attributed to gear type. Aside from the median $\hat{\beta}$ for the otter trawl, which is close to 1, the median for all of the other gear is about 0.7. The 95% confidence limits for the medians for both beam trawl and trawl are below 1. Only one obvious outlier was present among the estimates of β_i; this was an estimate of -3.23 for age one whiting in ICES area VIIek, for a nephrops trawl. This estimate had a high standard error (2.31) and had very little influence on the subsequent random effects meta-analysis.

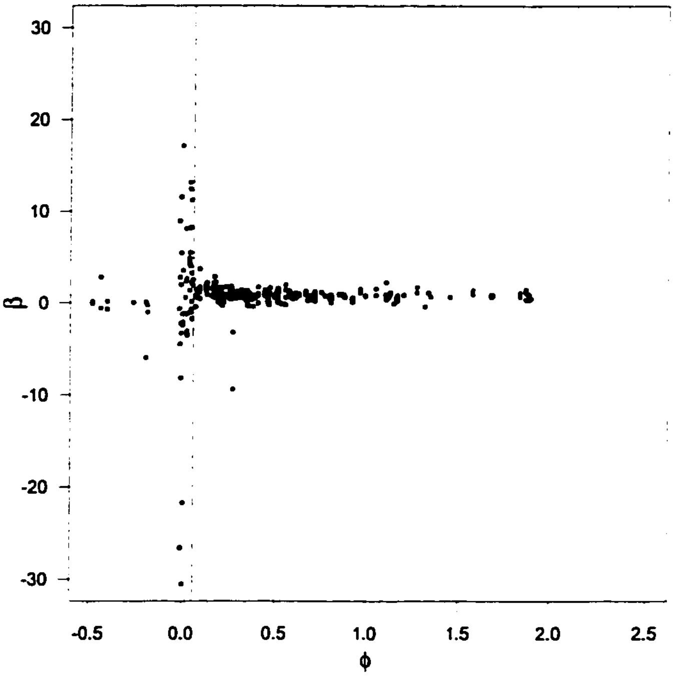


Figure 3.3: Scatterplot of estimated true variation in abundance, ϕ, versus the shape parameter describing the relationship between CPUE and abundance, β. The y-limits of the plots has been reduced for clarity, a further 20 estimates of β up to ± 100 were associated with small or negative estimates of ϕ. The dashed vertical line at 0.06 represents the cut-off used in the meta-analysis with data to the left excluded.

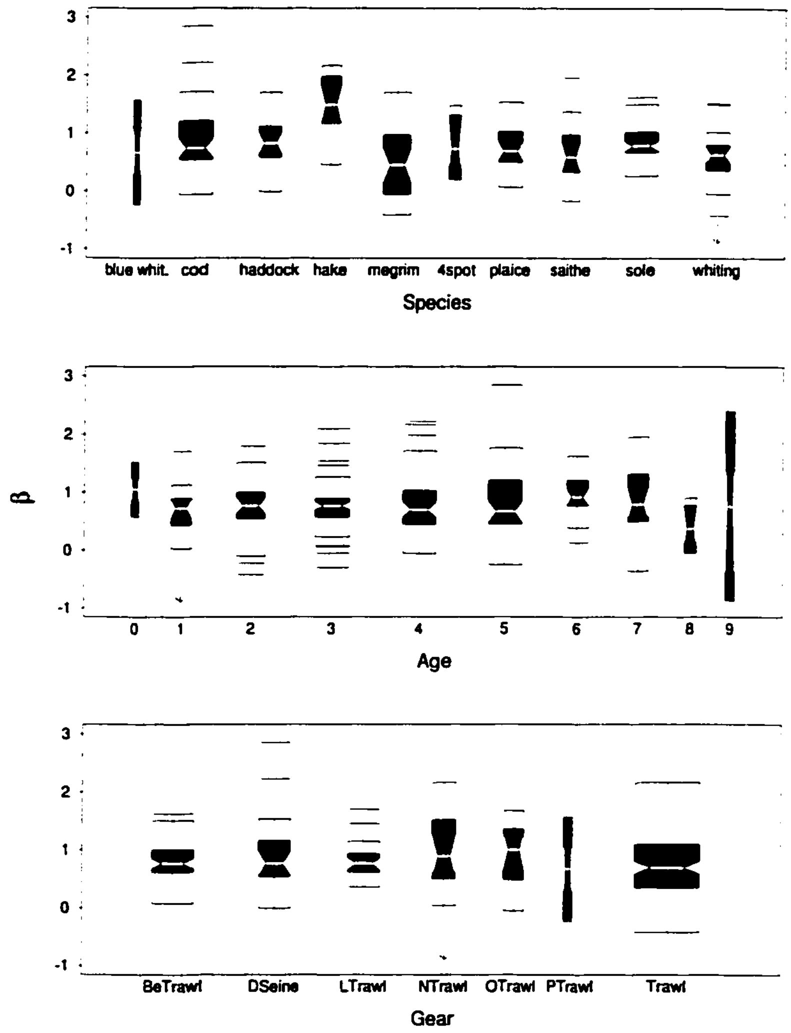


Figure 3.4: Estimates of the shape parameter β by species, age, and gear type. The upper quartile and lower quartile provide the outline of the box (the line inside the box represents the median). The width of the boxes are proportional to the square root of the number of data points. The notches are the approximate 95% confidence intervals of the median. Whiskers are drawn to the nearest value not beyond 1.5*(inter-quartile range) from the quartiles; points beyond are drawn individually as outliers. The gear types are denoted with codes provided in Table 3.2. The species codes "blue whit." and "4spot" refer to blue whiting and four-spot megrim respectively. The arrows indicate the outlier with $\hat{\beta} = -3.23$ (details given in text).

Sole and whiting had the lowest median $\hat{\sigma}$ (≈ 0.4) (Figure 3.5). Estimates for the flatfish species were generally lower but not significantly different from those of the Gadiformes. The highest estimates of σ were for haddock, with the upper quartile greater than 0. (i.e., coefficient of variation (c.v.) > 1). The youngest ages had the highest estimates of σ, and the distributions of σ were very similar for the other ages. Beam trawl estimates of σ showed little variation and had the lowest median of the gear types. Trawl estimates were also tightly distributed.

Estimated parameters of the random effect distributions for both β and $\log \sigma$ for cod were the least certain as indicated by the wide 95% confidence regions for the parameters (Figure 3.6). The random effects variances were also largest for cod, indicating higher inherent variability in the true distributions. However, the sample size for cod was less than the other groups. The confidence region for the Gadiformes was within the cod regions. The variance of the random effect distribution for $\log \sigma$ for flatfish was estimated to be zero. For none of the groups does the 95% confidence region for μ_β include values greater than 1.

For all groups most of the $\hat{\beta}_i$ were between 0.6-0.9 (Figure 3.7). The mean of the random effect distribution for both cod and Gadiformes were similar (0.75 and 0.73 respectively), and higher than that for flatfish (0.64). The variances of the random effect distributions were similar for all groups. The mean of the random effect distribution for $\log \sigma$ was the same for both cod and the Gadiformes (-0.65) which represents a median c.v. of 0.52 (Figure 3.8). This was lower than for flatfish, (-0.86, representing a c.v. of 0.42). This suggests that flatfish CPUE provides more precise abundance indices (i.e., lower noise) than those of Gadiformes.

Robustness

The previous results used data sets selected based on the estimated value for ϕ. To assess the robustness of my conclusions to this cut-off I repeated the random effects analysis for β using different criteria (Table 3.3). The original criterion was $\phi < 0.06$, so cut-offs of 0 and 0.1 were examined for sensitivity. Except for the Gadiformes, none of the estimates of μ_β or σ_β^2 varied by more than 5% from the base case. For the Gadiformes, using all data, the estimation procedure could not converge.

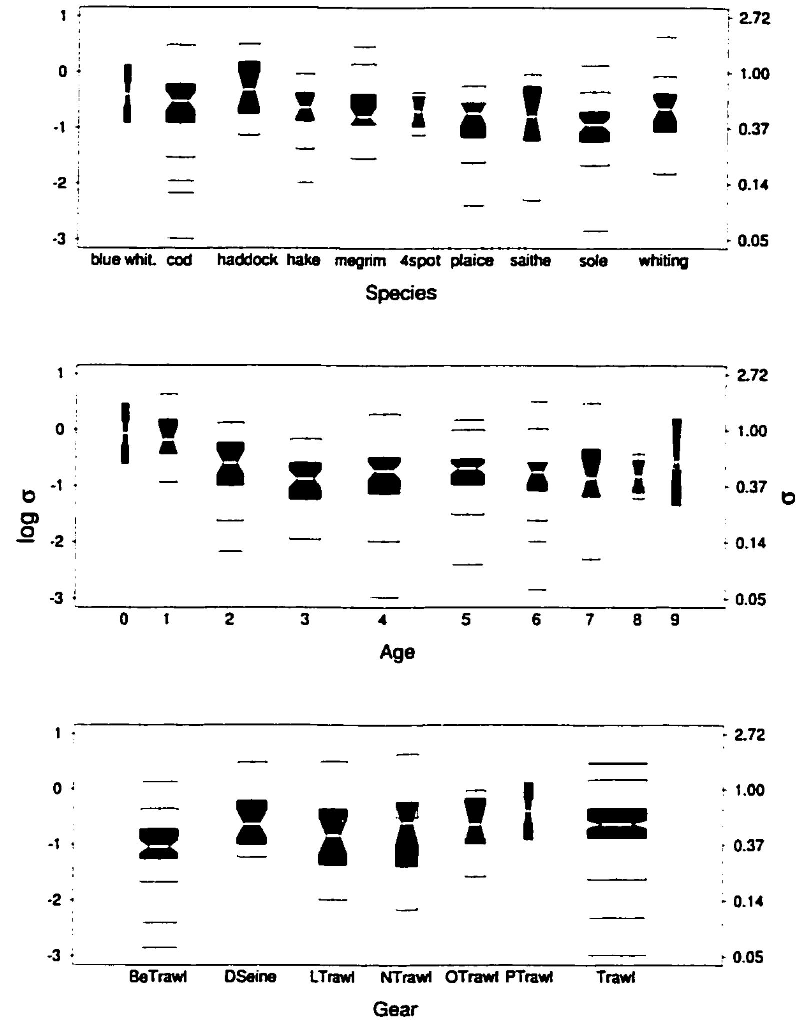


Figure 3.5: Estimates of the logarithm of the estimation error standard deviation ($\log \sigma$) by species, age, and gear type. See Figure 3.4 for details of how to interpret the boxplot.

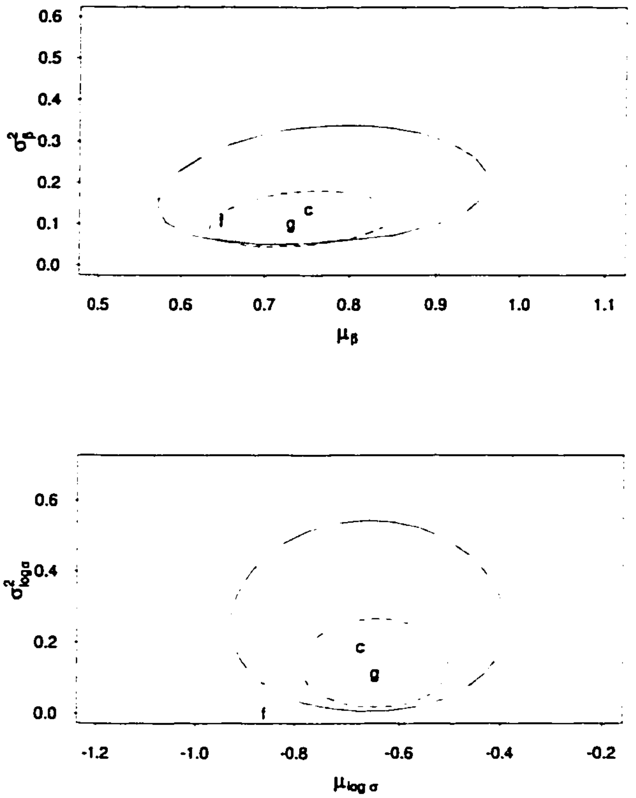


Figure 3.6: Log likelihood contours for μ_β and σ_β^2 (top), and $\mu_{\log\sigma}$ and $\sigma^2_{\log\sigma}$ (bottom) for cod (solid line), flatfish (dotted line), and Gadiformes (dashed line). The contours represent the approximate 95% joint confidence region ($\chi^2_{0.05,2}/2 \approx 2.991$). The maximum likelihood estimates for each group are indicated by lowercase letters, i.e., c for cod, f for flatfish, and g for Gadiformes.

Table 3.3: Estimates of μ_β (with σ_β^2 in parentheses) for analyses based on different selection criterion for ϕ.

Group	all	$\phi > 0.0$	$\phi > 0.06$	$\phi > 0.1$
Cod	0.757 (0.179)	0.714 (0.179)	0.753 (0.134)	0.757 (0.136)
Flatfish	0.500 (0.250)	0.625 (0.121)	0.647 (0.108)	0.650 (0.105)
Gadiformes	NA	0.729 (0.105)	0.732 (0.096)	0.729 (0.089)

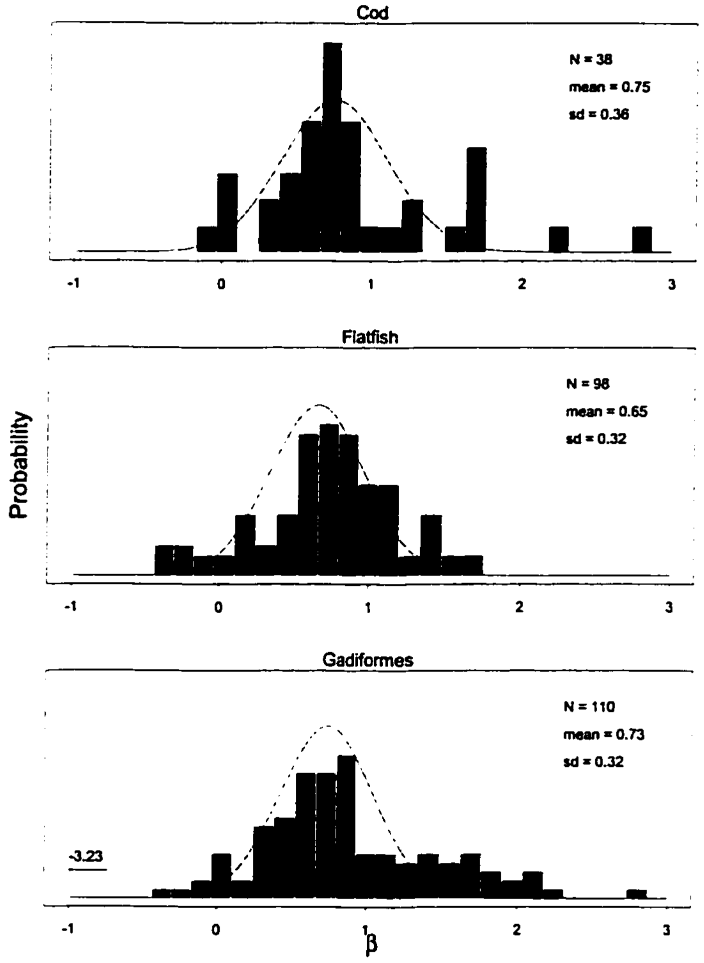


Figure 3.7: Histogram of individual estimates of β_i with the estimated random effects distribution overlaid. The histogram has been scaled to sum to 1 so that it is directly comparable to the distribution. (a) Cod: based on 38 data sets, the random effect distribution for β was $\beta \sim \mathrm{N}(0.75, 0.13)$. (b) Flatfish: based on 98 data sets, $\beta \sim \mathrm{N}(0.65, 0.11)$. (c) Gadiformes: based on 110 data sets, $\beta \sim \mathrm{N}(0.73, 0.10)$. The arrow indicates the outlier described in the text.

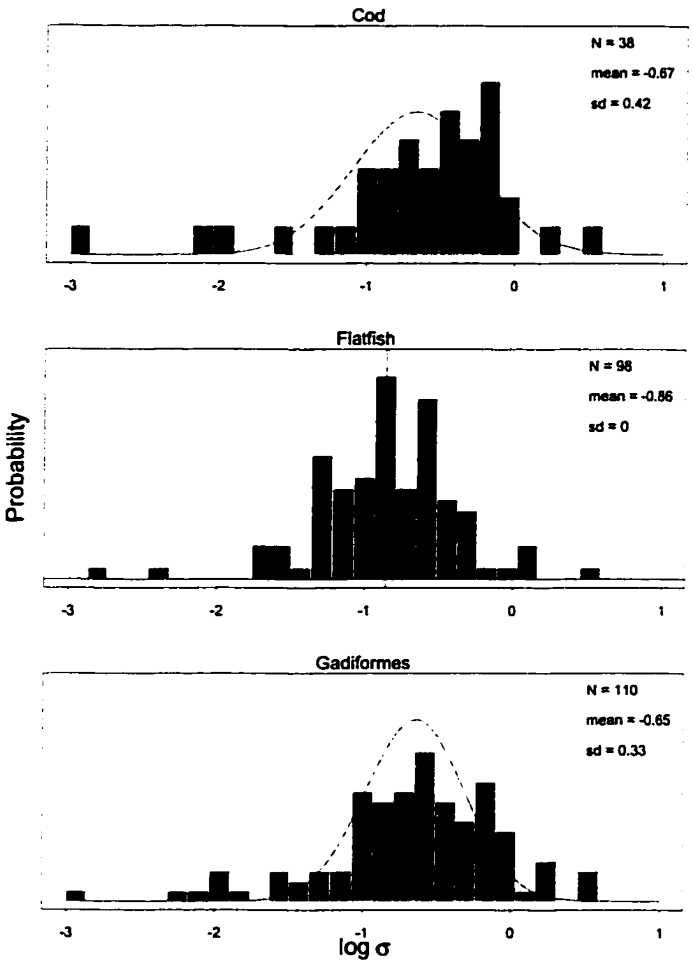


Figure 3.8: Histogram of individual estimates of the log of the estimation error standard deviation ($\log \sigma_i$) with the estimated random effects distribution overlaid. The histogram has been scaled to sum to 1 so that it is directly comparable to the distribution. (a) Cod: based on 38 data sets, the random effect distribution for $\log \sigma$ was $\log \sigma \sim \mathrm{N}(-0.67, 0.18)$. (b) Flatfish: based on 98 data sets, $\log \sigma \sim \mathrm{N}(-0.86, 0.0)$. (c) Gadiformes: based on 110 data sets, $\log \sigma \sim \mathrm{N}(-0.65, 0.11)$.

3.4 Discussion

While the validity of CPUE as a proportional index of abundance has long been questioned, the analysis described here provides the first large-scale attempt to test the assumption of proportionality. The findings here, while based upon data from a single area, must lead to concern among those who model CPUE data assuming proportionality. I used a power curve where the shape parameter, β, describes the shape of the relationship between CPUE and abundance. Focusing on data from the ICES region, I compiled 297 data sets, from which I was able to obtain reasonable estimates of the relationship between CPUE and abundance for 209.

It was found that the availability of multiple survey series led to more reasonable estimates of β and the other variance components of the model. Of the series where reasonable estimates of β were obtained, only 62 (30%) were greater than 1. It is not surprising that most of the estimates were less than one; in fact many researchers have found evidence of hyperstability in the relationship between CPUE and abundance, e.g., Peterman and Steer (1981), Crecco and Overholtz (1990), and Swain and Sinclair (1994). Quinn (1985) and Quinn and Collie (1990) both assumed the CPUE was not proportional to abundance, assuming a square-root relationship. This relates to a power term of 0.5 and is slightly lower to those estimated here.

The random effect distribution for β had most of its weight below 1 for all three groups. For cod, 76% of the mass was below 1, and this was lower than that for flatfish (86%) and gadiformes (81%). This represents the first large-scale examination of the relationship between CPUE and abundance, and as such, provides the strongest evidence against the assumption of proportionality. However, here I have only examined a limited number of gear types, mainly trawl-type gears so it is important to consider the generality of the findings. I have not considered CPUE series from purse seine fleets, although the fishing process of these types of vessels have been studied previously, with evidence of hyperstability (Clark and Mangel 1979; Allen and Punsly 1984). It will be important to examine other data, especially from other gear types, to determine if hyperstability is a property of other fishing methods (e.g., longlining or recreational hook and line fisheries). If possible the analysis should also be extended to consider invertebrate fisheries and the different gear types used in these fisheries (e.g., dredges and pots).

Some stock assessment methodologies do consider non-linearity in the relationship between CPUE and abundance, e.g., Darby and Flatman (1994). However, this assessment package does not currently have the ability to incorporate external information on β, such as that provided by the random effect distributions presented here. Furthermore, in many ICES assessments the power curve is more commonly used for only the youngest age classes. Here I have demonstrated that the power curve is the appropriate model at all ages.

The random effect distributions here could be used as prior distributions in a Bayesian assessment (Walters and Ludwig 1994; Punt and Hilborn 1997; Meyer and Millar 1999a). In Bayesian analysis, the prior distributions are combined with the the data to give posterior probability distributions that are used for inference. The variance of the random effect distributions may under-estimate the true variance as it does not consider the uncertainty in the estimation of μ_β and σ^2_β (Efron 1996). For this reason it may be important to "increase" the variance by some factor before using these distributions as Bayesian priors or as additional likelihood components.

In some fisheries, research survey abundance indices may not be available, or the survey series may be quite recent and not provide information on historical stock sizes. If CPUE data exist, it may be possible to get some indication of historical stock sizes using the priors for β that have been derived here. In Chapter 7 I will provide a framework for the inclusion of the priors for β constructed here.

I have also obtained estimates of the observation error variance of a CPUE series, $\log\sigma$. Since σ represents the square root of the observation error variance of a CPUE series, if the distribution of the ϵs is approximately lognormal, then σ is an approximation of the c.v. of the CPUE data source. This c.v. gives an indication of the precision of CPUE indices, and is extremely important for situations where alternative data types (e.g., CPUE and trawl survey) are available for a stock assessment. In such cases the weighting of the different data sources can be important (Quinn and Deriso 1999). This distribution for $\log\sigma$ will also be useful for simulation studies. It was found that $\log\sigma$ was generally lower for flatfish than the Gadiformes. This difference implies that CPUE data from flatfish will be more useful for "tuning" stock assessment models.

3.4.1 Robustness and alternative methods

The analysis presented here relies on the assumption that research survey indices are proportional to true abundance. If the surveys were carried out within some random sampling framework then one would hope this assumption would hold. However, in some cases, the surveys may not cover the whole stock. In such circumstances, proportionality may still hold but non-proportionality is also possible. By using the data used by the assessment scientists, I have made the same assumptions that they had made. Given this assumption, I am more confident in estimates for stocks for which multiple surveys were available or the surveys provided good coverage of the stock area.

The model used to estimate density-dependent catchability was used in a slightly different formulation to estimate density-dependent mortality by Myers and Cadigan (1993a). Cadigan (1993) and Myers and Cadigan (1993a) carried out extensive simulations to determine the effect of small sample size and violations of the model assumptions on bias of the parameter mathematically equivalent to β. Both studies found that reasonable violations of the model assumptions (alternative distributions for the error distributions, autocorrelation, and time trends in the series), had relatively little effect on bias, but that small sample size, generally led to overestimation of the parameter mathematically equivalent to β by $\approx$ 10%, and underestimation of the estimation error variance by $\approx$ 10%. Cadigan (1993) developed two methods to correct for bias in the individual estimates, but found that they did not work in practice. It is recommended that the estimates of β presented here be reduced by 10% to account for the general trends in bias observed in these simulation studies.

I have only considered one model for the relationship between CPUE and abundance (Eq. 3.2). This power curve is the simplest model allowing nonlinearity between CPUE and abundance. There was no evidence that more complicated models would provide more parsimonious fits to the data. In fact, even with a large number of data sets, it was difficult to obtain reliable estimates of the shape parameter of the power curve.

I have used a structural rather than functional regression approach for estimation in the presence of measurement errors (Fuller 1987). In the structural approach, we assume that the independent variable, i.e., the true abundance, is a random variable that has a common distribution. The functional approach, assumes that the independent variables are fixed effects and must be estimated. The functional approach has the advantage that changes over time can be easily modelled, but has the considerable disadvantage that a separate unknown abundance would have to be estimated for each year. This leads to an estimator which is not guaranteed to be consistent because more parameters are estimated as more data are added (Fuller 1987).

I used maximum likelihood to estimate the parameters of the random effects distributions, but other approaches are possible. A moment estimator for the mean of the random effects distribution was provided by DerSimonian and Laird (1986). This approach requires that the variance of the random effects distribution (σ_β^2) is known without error and thus no estimate of its uncertainty is possible. The likelihood approach presented by Hardy and Thompson (1996) enables the analyst to obtain profile-based confidence limits for both the mean and variance of the random effects distribution. It would also be possible to use hierarchical Bayesian methods to estimate the random effects distribution (Carlin and Louis 1996; Liermann and Hilborn 1997). When using Bayesian methods, it is necessary to specify prior distributions for the mean and variance of the random effects distribution. In instances where the maximum likelihood estimates for the variance are significantly different from zero, it is unlikely that the prior specification for variance in a Bayesian model would lead to different estimates than those from the likelihood approach. However, in cases where the data are poor, it is common for the maximum likelihood estimate of random effects variance to be zero (Searle et al. 1992). In such circumstances, the choice of prior for the Bayesian analysis is likely to be important.

It is also possible to conduct a fully Bayesian analysis of these data. The individual modelling of the relationship between CPUE and abundance for each series could be performed simultaneously, where the β_is are estimated, while considering the information provided by other data sets. In such a case I would estimate the prior distribution within the analysis rather than separately as has been done here. I could also estimate the random effects distribution for $\log \sigma$ at the same time. As I will show in Chapter 6 treating some parameters as random effects often leads to more stable estimates of other model quantities. This might help overcome the problems with negative variance components that I encountered here. This type of approach has been used in the estimation of the maximum reproductive rate from spawner-recruitment data (Myers et al. 1999). While this approach is theoretically possible, it would be computationally intensive but will be the focus of future research.

The data used for this analysis were series of age-specific CPUE and research survey abundance indices as used in ICES stock assessments in 1999. As ICES assessment scientists often have access to a number of CPUE and survey series, the data used in the assessment models may be of better quality than that used by assessment scientists with less available data (i.e. they have enough data to consider not using less reliable series). Thus it is possible that the data used in this analysis may be better than that used by assessment scientists elsewhere. Data series that exhibited extreme hyperstability or large observation error may have been excluded by the assessment scientists. If the data included here are less noisy (regardless of the value of β), the estimates of the random effect distribution for $\log \sigma$ will over-estimate the usefulness of CPUE as a predictor of abundance (on the whole).

By assuming that estimates for the different ages are independent (in the random effects analysis), I have likely over-estimated the effective number of data sets. This would not bias the means but could cause the confidence regions in Figure 3.6 to be under-estimated. There is a great need for methods to incorporate non-independence into simple random effect models as I have used here.

Dunn et al. (2000) used both research survey (acoustic and trawl) and model-based (i.e., predicted abundance from statistical catch-at-age models) estimates of abundance to examine the relationship between CPUE and abundance. I have chosen not to use model-based abundance estimates for four reasons. (Aside from problems that might occur due to biases in the assessment methods.) First, there is seldom one single model that is used in any one assessment. A number of structurally different models may be compared with different results. Second, for the model-based abundance indices to be independent of the CPUE series it would require that the CPUE data not be included in the model. In many cases, such a model run may not have been considered. Third, it is difficult to obtain estimates of the observation error variance for model-based estimates. Finally, the model-based abundance indices are unlikely to be independent from one year to the next as an underlying population dynamics model generates them. This would complicate any analysis, as it would require the incorporation of autocorrelation in the estimation procedure. Thus, while model-based abundance indices have the advantage that they synthesize a range of information on abundance; the disadvantages of model-based indices outweigh these.

Also I have used age-disaggregated abundance indices instead of the aggregated indices used by Dunn et al. (2000). While aggregated indices may be less noisy (due to averaging of error over a number of ages), any differences in the age-specific selection of the survey and fishing gears will lead to biases in the estimation of β. It is also easier to get good contrast with age-structured indices due to variability in recruitment. Any ageing error would most likely act to reduce the interannual variability in abundance of the older age-classes. The median relative range (maximum / minimum CPUE index) in CPUE for a series was 12-fold. Based upon the simulations performed by Dunn et al. (2000), the variability observed here is sufficient to allow accurate estimation of β.

The relationship between age-aggregated CPUE and abundance may be different to that for age-disaggregated indices. Therefore, careful consideration must be given before applying the results of this analysis. The relationship between CPUE and abundance is most likely a complex function of the abundance of individual cohorts. Simple simulations were conducted in which a single β was used for the entire population, and then separate βs were estimated for each age class. It was found that the age-specific estimates were always closer to one suggesting that the age-based estimates may over-estimate the extent to which CPUE is proportional to abundance.

Changes in catchability, i.e., efficiency of the fishing fleet over time, have the potential to cause biases in the estimates of β. For our CPUE data, the lowest value in a series was just as likely to be the first or last year, but the median time trend was a 5% decline per year. It is likely that the findings here are robust to this potential bias.

3.4.2 Potential causes of hyperstability

The analysis presented here is purely descriptive, i.e., I have estimated a parameter describing the relationship between a CPUE series and abundance as estimated by a single or multiple research survey time series. The analysis of fish distribution patterns has long been recognized as critical to the interpretation of CPUE indices (Paloheimo and Dickie 1964; Clark 1982; Prince and Hilborn 1998). Without additional data on the spatial distribution of fishing effort and fish abundance the two are confounded. However, it is possible to speculate as to some possible causes of the predominance of hyperstability that was observed.

One hypothesis is that the efficiency of fishers is increasing over time and that they can keep their catch rates high regardless of the distributional pattern of fish. This may partly explain the patterns although most often assessment scientists will attempt to correct or standardize effort for changes in efficiency. It is more likely that spatial factors are important.

The critical requirement for hyperstability (in the absence of increasing efficiency) is that fish aggregate and become more concentrated as abundance declines. If fish were distributed uniformly, fishers could not allocate effort spatially to induce hyperstability. If fish are aggregated and fishers fish at random, CPUE can still be proportional, but if they do not fish at random, both hyperstability and depletion are also possible. If fishers allocate effort to these hotspots then hyperstability will occur. In some cases, fishers may not identify hotspots or may not fish these regions for economic or reasons of safety (i.e., if hotspots are far offshore). In such cases hyperdepletion could occur.

Myers and Cadigan (1995) found that 'northern cod' (Atlantic cod in areas 2J+3KL) exhibited a substantial contraction in range during the period up to the closure of the fishery, and Hutchings (1996) showed that the proportion of trawl survey sets with high catch rates of cod remained constant throughout this time. Density-dependent habitat selection theory (Fretwell and Lucas 1970; MacCall 1990) suggests that range expansion and contraction could occur with changes in abundance.

Evidence of density-dependent habitat selection will clarify the role of fish behaviour. As an example, if we look at the cumulative distribution of abundance versus area for northern cod we see that over time abundance became more concentrated (Figure 3.9). An index of concentration is given by the Gini coefficient (Myers and Cadigan 1995) measuring the difference between the observed distribution and a uniform distribution. Another potential index is the area inhabited by x% of stock (Swain and Sinclair 1994). These both support the conclusion that the stock become more concentrated over time and that there is some relationship between these indices and stock abundance (Figure 3.10). A stock that becomes more concentrated as abundance declines is more likely to be vulnerable to hyperstability of CPUE.

It might be anticipated that the amount of hyperstability would be related to the extent to which a species concentrates. Unfortunately I have not detected any significant differences between the species examined here because of the large variation within species. This is likely due to the fact that different fleets (with different behaviour) are fishing different populations of the same species and this may mask the species' effects.

While I have demonstrated here that CPUE indices are likely to decline slower than the abundance of fish under exploitation, I have not provided any evidence for possible causes (for stocks included in the analysis). I have illustrated, using data for northern cod, the types of analyses that should be undertaken in the future to further examine this critical process.

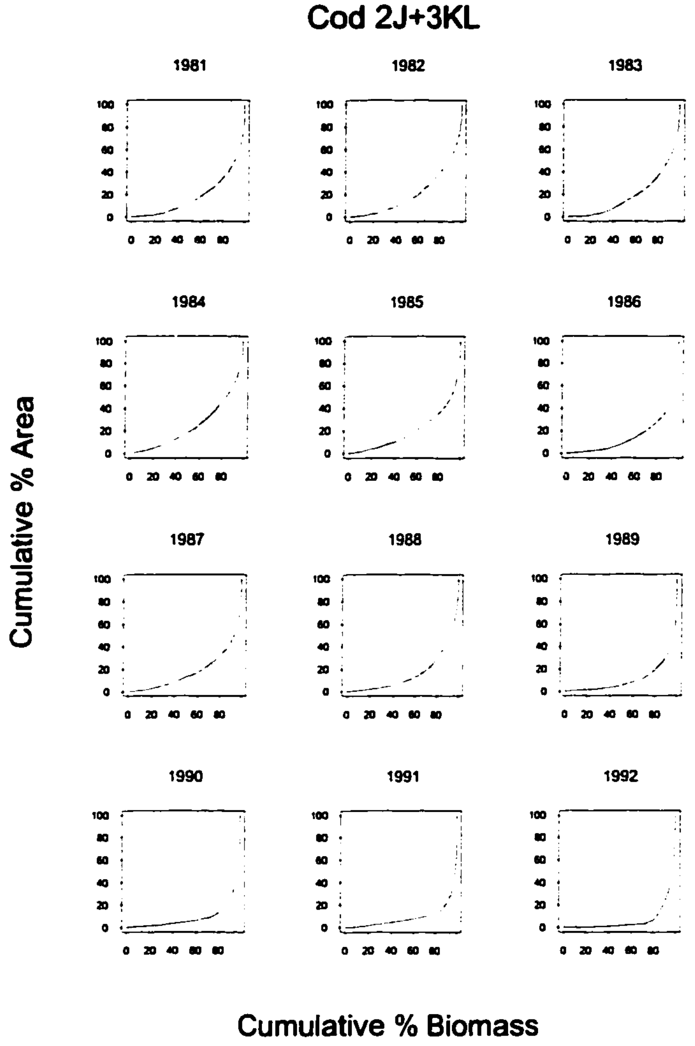
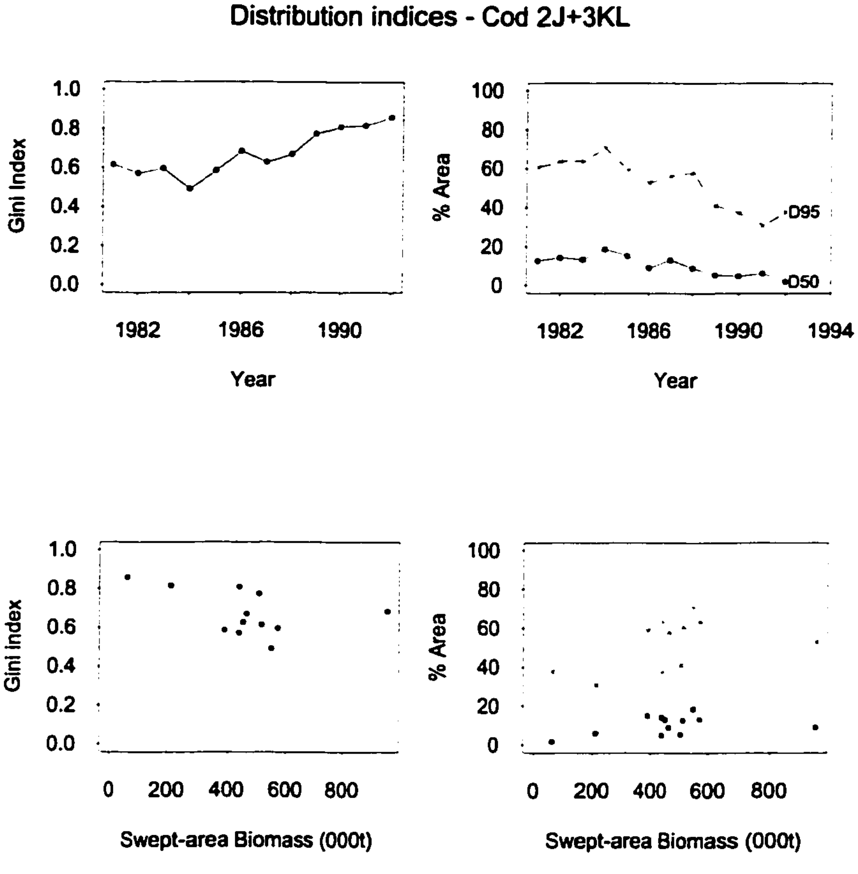


Figure 3.9: Cumulative abundance versus cumulative area for Atlantic cod in areas 2J+3KL. The greater the curvature (of the solid line), the greater the concentration of fish in a small part of the survey area. The dashed line represents the pattern that would be expected if fish were uniformly distributed across the survey area. (Adapted from Myers and Cadigan (1995)).



Distribution indices - Cod 2J+3KL

Figure 3.10: Spatial distribution measures for 2J+3KL cod. The top two graphs show the time series of estimates for the Gini coefficient and two population range indices. The bottom graphs show the relationship between the indices and estimate of swept-area biomass from the survey.

Chapter 4

Methods for examining density-dependent mortality in fish

4.1 Introduction

An important question regarding the regulation of a fish population relates to the age at which the year class strength (YCS) or abundance of a cohort is determined. Also, the early estimation of YCS is critical to our management of exploited fish populations. Our ability to accurately predict future catches (and abundance) from spawner abundance (Mertz and Myers 1995) or juvenile surveys (Shepherd 1997) relies on an understanding of these processes. As we often observe large variation in the abundance of cohorts that 'recruit' to a fishery, it is critical for fisheries management to be able to accurately predict the strength of a cohort before it reaches the fishery (Shepherd 1997). If we can determine this (critical) age, then we can best direct research to factors that may be responsible for the large variation observed in YCS.

To make general conclusions about this problem it is necessary to build models that describe the population dynamics of fish, from the early life stages (e.g., eggs and larvae) to the time at which they enter a fishery. Results from these models for many populations should then be combined using meta-analytical approaches (e.g., such as those used in Chapter 3) to test hypotheses. In this chapter, I have only considered this first part in addressing this problem. I will demonstrate that while it is possible to construct the models necessary to describe the dynamics, this is not a trivial task.

In the following sections I will outline the biological basis for this problem and the many technical difficulties that the biology and data impose. I will describe the typical sorts of data that are available for such analyses and the development of new and innovative methods to address these problems. These methods will rely heavily on sharing information across cohorts, a principle consistent with meta-analysis. I will illustrate the approach through two examples but will not perform a full-scale meta-analysis using data from many populations.

4.1.1 Biological framework

The motivation for this work is based upon Hjort's (1914) critical period hypothesis which states

> "the numerical value of a year class is apparently stated at a very early age, and continues in approximately the same relation to that of other year classes throughout the life of the individuals."

This hypothesis does not allow for stochastic variation in density-independent mortality that could change the order of YCS of multiple cohorts over time (after this critical period). There is no explicit discussion of density-dependent mortality in this hypothesis. It may be possible to interpret the second part of this quote two ways (R. A. Myers, Dalhousie University, pers. comm.). First, a 'strong' interpretation whereby no density-dependent mortality occurs after the critical period, or a 'weak' interpretation whereby density-dependent mortality after the critical period might reduce variation in YCS. These both satisfy the requirement that the ordering of YCS (after the critical stage) does not change, and imply that any density-dependent mortality after the critical period must follow a Cushing (1971) or Beverton-Holt (1957) model (or some similar monotonic model). Ricker-like (1954) density-dependent dynamics, i.e., that allow for over-compensation, could produce patterns not consistent with Hjort's hypothesis in that the ordering of the abundance of cohorts could change after the critical period.

Several features of the abundance of cohorts over time, consistent with Hjort's hypothesis, are illustrated in Figure 4.1. This represents an example of the weak version, whereby variability in YCS has increased from eggs to the age where the juveniles were first observed (0.75 years in this example) then decreased over time while the ordering of year class remains the same during the juvenile stage, i.e., the lines do not overlap after the first age.

With this description it is possible to make a number of testable hypotheses. However, until recently there have been no methods appropriate for such investigations. To fully test Hjort's hypothesis, we require methods that allow us to test for variable density-independent mortality at different stages and allow a range of dynamic models to be considered.

4.1.2 Technical implications

Many ecologically important problems involve data and models that seldom satisfy the assumptions required for traditional statistical procedures. This is the case for the problem considered here. There are many aspects of such an analysis that have in the past posed significant technical obstacles. To make progress on this problem, it is crucial to come to terms with several issues. First, the true abundance of a cohort is never measured without error, and this must be explicitly included in the model. Second, population dynamics are inherently nonlinear, and models must be formulated accordingly. Third, we must be able to accommodate extra variation not explicitly described by parameters in our models of population dynamics (i.e., process noise), perhaps due to variation in density-independent mortality. Fourth, there is seldom sufficient information in the data on any one cohort to obtain accurate estimates of density-dependent mortality, and thus information must be shared among cohorts. Lastly, even with excellent data, it is rare that definitive conclusions can be reached from any one data set, i.e., it is necessary to combine data across many populations. This is necessary to determine the generality of any single stock finding. These final two issues motivate the analysis presented here because to attack this problem properly, meta-analytical approaches must be used.

These aspects should (theoretically) be considered in every analysis of this problem but technical limitations have restricted the types of models and approaches that have been considered. Myers and Cadigan (1993b) used linear mixed effects models with observation error to examine this question. de Valpine and Hastings (2002) consider

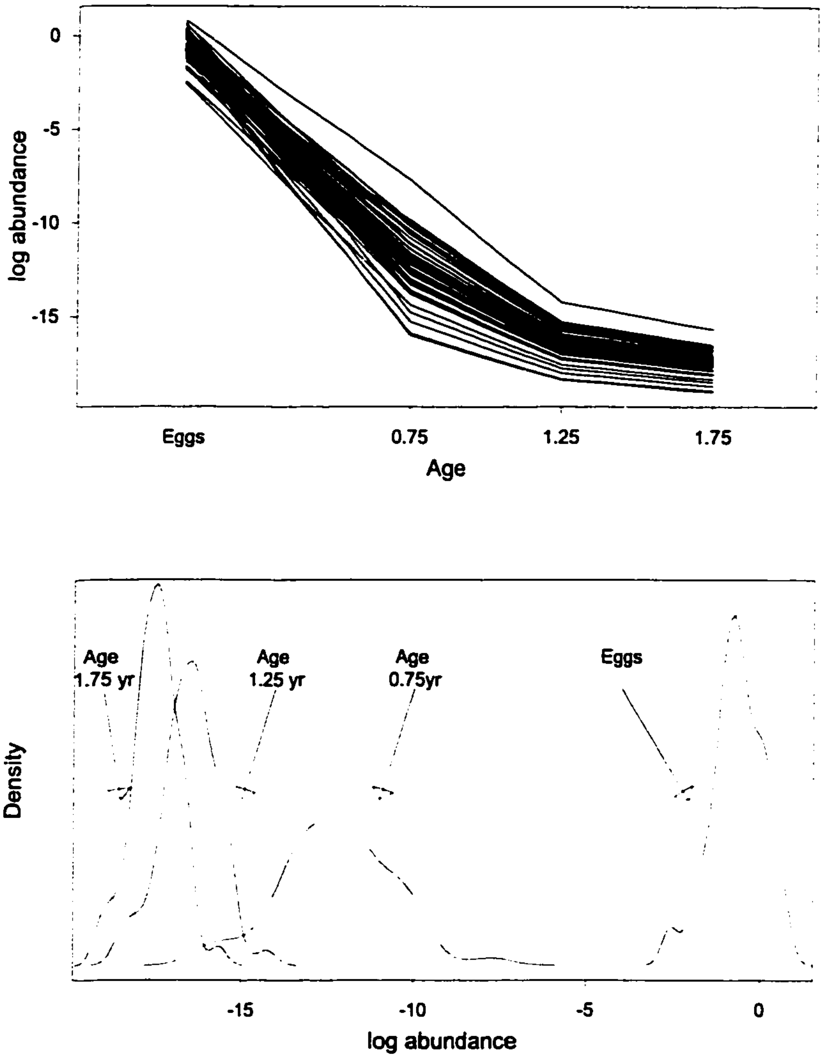


Figure 4.1: Schematic illustrating some components of Hjort's critical period hypothesis. (Top) Abundance trajectories of 50 cohorts projected through time assuming stochastic variation occurs some time before age 0.75 and that some density-dependent mortality occurs after this time. (Bottom) Density distributions showing variation in abundance at age across the 50 cohorts.

the Ricker model and the nonlinear Beverton-Holt model in a state-space framework. Their approach explicitly considered points 1–3 above but they did not apply their approach to real data sets. Here I will introduce an approach that integrates the state-space approach of de Valpine and Hastings (2002) into a mixed effect analysis where information is shared across cohorts to improve estimation of the model parameters. This approach will be implemented using maximum simulated likelihood (Mariano et al. 2000).

4.2 Methods

In the following sections I will describe the data used for the examples and the development of models and modelling approaches to overcome the inherent problems with the data and nonlinear dynamics.

4.2.1 Data

The general data requirements for this problem are estimates of abundance of a cohort over a number of time periods. In order to estimate variation in density-independent mortality, it is necessary to have these observations for many cohorts.

As examples demonstrating the application of the methods, I consider data for two Atlantic cod populations: North Sea (as used by Myers and Cadigan (1993b) and in Chapter 3) and Georges Bank (O'Brien and Cadrin 1999). The time series of estimates of abundance of a cohort are from bottom trawl surveys conducted in each region. These two areas were chosen because of their long time series and because two surveys were generally conducted annually in different seasons giving two observations of each cohort annually. However, these data are not sufficient to explicitly test Hjort's hypothesis as they do not sample cod during the larval stages.

It has been hypothesized that much of the variation in year class strength is determined in the early larval stage (Hjort 1914). Unfortunately, long time series of egg or larval cod abundance are not available so I use estimated spawner biomass (from surveys or stock assessment models) as a proxy for egg production. I acknowledge that there are reasons why the assumption that SSB is proportional to egg production may not hold (Trippel et al. 1997). Spawning stock biomass (SSB) was estimated from numbers-at-age, weight-at-age, and proportion mature-at-age. Because the growth and maturity of Atlantic cod in these stocks has varied over time, annual estimates of weight-at-age and maturity were used where available.

Data for cod from Georges Bank consisted of two survey series; standardized (for vessel and door changes) mean catch per tow at age (numbers) from the U.S. Northeast Fisheries Science Center (NEFSC) offshore spring (1968–2000) and autumn (1963–2000) bottom trawl surveys. For the spring survey I used data for ages 1.25, 2.25, and 3.25 years. Observations of age 0.25 year fish were not used because of the very large number of observed zeros. For the autumn survey I used data for ages 0.75, 1.75, 2.75, and 3.75 years. Estimated SSB was based on numbers-at-age from the autumn survey prior to spawning. This gave eight observations for each cohort, seven from surveys and one based on SSB. Fewer observations were available for cohorts at the start and end of the survey series.

The data for North Sea cod was an updated version of the data considered by Myers and Cadigan (1993a). The surveys used were the International Bottom Trawl Survey (1971–1998) for ages 1.25, 2.25, and 3.25 years, and the English Groundfish Survey (1977–1998) for ages 0.75, 1.75, and 2.75 years. Estimated SSB was from the stock assessment as reported in Myers stock-recruitment database (http://fish.dal.ca/ myers/welcome.html). This gave seven observations for each cohort, six from surveys and one based on SSB.

The trawl survey data used in this analysis have one further complication. The trawl survey does not sample fish in proportion to their abundance, i.e., older ages are often more vulnerable to the survey gear than the younger ages. I will show that the parameter for catchability (q_a), required to accommodate this, is confounded with other parameters in the dynamics models described in later sections.

4.2.2 Formulation of models and approaches

The modelling approach that I have taken to address this problem is quite complex. First, I describe the state-space approach that is used because of its ability to consider the important sources of variation (e.g., variation in year class strength and interannual variation in survival). Second, I describe a range of models for describing the dynamics of juvenile fish abundance, incorporating both density-dependent and density-independent mortality. Third, I describe how it is necessary to share information across cohorts to estimate parameters. Finally, I introduce the method of maximum simulated likelihood (MSL) as a means to estimate the parameters of these models.

Throughout this section I will follow the development and notation used by Kitigawa (1987) and later de Valpine and Hastings (2002). The linear dynamic models and treatment of random effects and observation error follows the development of Myers and Cadigan (1993a; 1993b).

State space model

State-space models and the Kalman filter represent the standard approach to model structural time series (Harvey 1989). In the following section I hope to provide a simple explanation that motivates the use of this approach here.

The time series that we have here is the abundance of a single cohort over time; these represent the 'states'. It would seem reasonable that the abundance of a cohort at some future time would be related to the abundance now, and what it was in previous times. We might have a population model to predict future abundance. As models will always be a simplification of reality, our predictions will not be accurate as there will be other processes (not modelled) that will result in some deviation from our model predictions. In the state-space model, this is referred to as the 'transition equation' where the prediction error is included as process noise.

There is a second important feature of a state-space model. It is important to recognize that we never actually know what the true abundance of a cohort is. The 'measurement equation' takes into account how we observe abundance (e.g., perhaps with a survey that does not sample all ages of a cohort equally), and that we observe it with error. These two equations, the measurement and transition equations, are what make a state-space model (Harvey 1989; de Valpine and Hastings 2002). I will now provide a more detailed description.

The general model can be represented as two sets of equations, one to represent the measurement process and one to represent the transition from one state (age) to the next. Here, the number of fish (or the logarithm of fish) in cohort t estimated at time a is given by $y_{t,a}$. While we observe $y_{t,a}$ from the trawl survey, the true abundance, which we denote as $n_{t,a}$, is an unobservable random variable known as a state variable[1]. The observable is related to the state in the measurement equation:

$$y_{t,a} = G(n_{t,a}, \epsilon_{t,a}; \theta) \tag{4.1}$$

for $t = 1, 2, \cdots, T$, where T denotes the number of cohorts modelled, for $a = 1, 2, \cdots, A$, where A denotes the number time periods at which a cohort is observed, and θ is a vector of parameters to be estimated from the data. The parameters of θ in the measurement equations are the survey catchability coefficients and the observation error variances. The estimation errors, $\epsilon_{t,a}$, are typically assumed to be mutually independent and normally distributed (if the log of the abundance is used). It is useful to define $Y_{t,b}$ to be the information on cohort t up to age b, i.e., $Y_{t,b} = \{y_{t,1}, y_{t,2}, \cdots, y_{t,b}\}$.

The second key equation of a state-space model is the transition equation which describes the population dynamics between two time periods. The transition equation has the form

$$n_{t,a} = F(n_{t,a-1}, \nu_{t,a-1}; \theta), \tag{4.2}$$

where $\nu_{t,a}$ describes the random process noise at time a, and θ includes the parameters of the models that describe the transition from state to state. It is assumed that $\epsilon_{t,a}$ and $\nu_{t,a}$ are mutually independently distributed from one another and normally distributed, i.e., $\epsilon_{t,a} \sim \mathrm{N}(0, \sigma^2_{\epsilon_a})$ and $\nu_{t,a} \sim \mathrm{N}(0, \sigma^2_{\nu_a})$, though the general approach allows for these assumptions to be relaxed.

Likelihood

In order to estimate the unknown parameters in equations 4.1 and 4.2, the likelihood function is maximized. In what follows, I use the generic notation $P(\cdot \mid \cdot)$ to indicate the probability density of the first argument given the second. For example,

---

[1] In general descriptions of state-space models, $\alpha_{t,a}$ would often be used. This is because Greek letters are generally used for unobservable random variables. I have not done this here so to be consistent with the population dynamic models described later.

let $P(y_{t,a}|n_{t,a})$ be the conditional probability density function derived from Eq. 4.1 and let $P(n_{t,a}|n_{t,a-1})$ be the conditional density function derived from Eq. 4.2. Because the true abundance is never observed without error, it is necessary to use a Kalman filtering approach. Specifically, I first calculate the probability of $n_{t,a}$ given past observations of the cohort. This is known as a one step ahead prediction and is given by

$$P(n_{t,a}|Y_{t,a-1};\theta) = \int P(n_{t,a}|n_{t,a-1};\theta)P(n_{t,a-1}|Y_{t,a-1};\theta)\mathrm{d}n_{t,a-1}$$

where the above equation is calculated recursively. The next step is to combine the new sample data, $y_{t,a}$ with the past information $Y_{t,a-1}$. This is done using Bayes rule (though it is not formally a Bayesian procedure (Harvey 1989)),

$$P(n_{t,a}|Y_{t,a};\theta) = \frac{P(y_{t,a}|n_{t,a};\theta)P(n_{t,a}|Y_{t,a-1};\theta)}{\int P(y_{t,a}|n_{t,a};\theta)P(n_{t,a}|Y_{t,a-1};\theta)\mathrm{d}n_{t,a}}$$

The first equation is usually known as the prediction equation, whereas the second is known as the update equation.

There are two approaches typically used to determine the initial condition in a state space model: (1) it is fixed at some value; or (2) it is described by a prior distribution. Later I will introduce an alternative approach where the initial abundance of cohorts (eggs), $n_{t,0}$, are treated as random effects, i.e., $n_{t,0} \sim \mathrm{N}(\mu_0, \sigma_0^2)$. Using such an initial condition, it is possible to calculate each of the above quantities recursively.

The likelihood is written in terms of the probability of observing each new data point, $y_{t,a}$, given the previous data, i.e. $Y_{t,a-1}$, for that age and cohort. This is the innovation form of the likelihood and is given by

$$L(\theta|Y_{t,A}) = \prod_{t=1}^{T}\prod_{a=0}^{A} P(Y_{t,a}|\theta) = \prod_{t=1}^{T}\prod_{a=0}^{A}\int P(y_{t,a}|n_{t,a};\theta)P(n_{t,a}|Y_{t,a-1};\theta)\mathrm{d}n_{t,a}, \qquad (4.3)$$

Note that the integral is the denominator in the update equation.

Measurement equation

The probability $P(y_{t,a}|n_{t,a})$, is the basis for the measurement equation. This describes the difference between the true state and our observation of it. If we assume normal errors, we have

$$y_{t,a} \sim \mathrm{N}(q_a n_{t,a}, \sigma_a^2) \tag{4.4}$$

where the observation error variances, σ_a^2, are estimated from the data, and q_a describes the catchability coefficient for fish of age a.

The advantage of the approach developed here is that other error structures can be accommodated. In particular, in some instances there are many observed zeros that require more complex error structures.

Density-dependent mortality in fish

Three models were considered here to describe the transition dynamics that were described generically in Eq. 4.2. I have used the power or Cushing (1971) model as was used by (Myers and Cadigan 1993a). This model has the advantage that it can be linearized by taking the logarithm of both sides (demonstrated below). The other two models, the Ricker (1954) and Beverton-Holt (1957), represent the two most commonly used models to describe patterns of density-dependent mortality in fish (Hilborn and Walters 1992).

The Cushing model represents a case in which density-dependent mortality in a cohort is proportional to the logarithm of abundance at the previous age (Myers and Cadigan 1993b),

$$N_{t,a} = N_{t,a-1} e^{-m_{a-1}-(1-\lambda_{a-1}) \log N_{t,0}} e^{\nu_{t,a-1}}$$

where m_{a-1}, λ_{a-1}, and $\nu_{t,a-1}$ describe the density-independent and density-dependent mortality, and process noise occurring between periods $a-1$ and a. Using the natural logarithm of such models is useful for estimation (Myers and Cadigan 1993b; de Valpine and Hastings 2002) (and Chapter 3).

While the terms q_a and m_{a-1} are in separate equations (i.e., the measurement and transition equations respectively) they are confounded in my analysis given the data that are available. We can not tell the contribution of each to an observed reduction in abundance of a cohort, i.e., is observed abundance lower because of density-independent mortality, or lower catchability to the trawl survey. In such instances it is only possible to estimate their sum $\gamma_a = \log(q_a) + m_{a-1}$. Including this, and writing $\log N_{t,a} = n_{t,a}$, we get

$$n_{t,a} = F_{\text{CUSH}}(n_{t,a-1}, \nu_{t,a-1}) = \lambda_{a-1} n_{t,a-1} + \gamma_a + \nu_{t,a-1}. \tag{4.5}$$

For the Ricker (1954) model it is assumed that density-dependent mortality is proportional to abundance (rather than the logarithm of abundance),

$$n_{t,a} = F_{\text{RK}}(n_{t,a-1}, \nu_{t,a-1}) = n_{t,a-1} - \lambda_{a-1} N_{t,a-1} + \gamma_a + \nu_{t,a-1}. \tag{4.6}$$

The third model is the Beverton Holt given by

$$n_{t,a} = F_{\text{BH}}(n_{t,a-1}, \nu_{t,a-1}) = n_{t,a-1} - \log(1 + (N_{t,a-1}/\lambda_{a-1})) + \gamma_a + \nu_{t,a-1}. \tag{4.7}$$

The parameters describing density-dependent mortality have different interpretations in each model. For the Cushing model it is the slope between the logarithm of abundance at adjacent ages, for the Ricker it is in units of $1/N_{t,a}$ representing the abundance of age a fish that results in the maximum number of age $a+1$ fish, and for the Beverton-Holt the units are $N_{t,a}$ and it is the abundance at age a for which subsequent abundance at age $a+1$ is half of the maximum.

The density-dependent mortality parameter from the Cushing is dimensionless and amenable to meta-analysis while the other models are problematic. Myers et al. (2001) showed that it is possible to standardize the data (by the area of the stock) so that the parameters from the Beverton-Holt are directly comparable across populations.

Mixed effects model

A mixed effects model is one which treats some parameters as fixed effects and others as random effects. This approach is required for this problem for two reasons. First, it is not possible to estimate all of the parameters with data for a single cohort so we must share information across cohorts. I can treat the density-dependent and density-independent (and catchability) terms as age-specific fixed effects across cohorts,

$$\lambda_a = \{\lambda_{1,a}, \lambda_{2,a}, \cdots, \lambda_{T,a}\}$$

and

$$\gamma_a = \{\gamma_{1,a}, \gamma_{2,a}, \cdots, \gamma_{T,a}\}.$$

Second, for the state-space approach it is necessary to integrate over the distribution for all of the states, $n_{t,a}$. An extension to the typical state-space treatments of this problem (i.e., treating the initial state as fixed or described with a prior), is to treat the initial condition as a random effect. This is accomplished by treating the distribution (across cohorts) in the abundance of eggs, $n_{t,0}$, as a normal random variable,

$$n_{t,0} \sim \mathrm{N}(\mu_0, \sigma_0^2).$$

Combining this with the maximum simulated likelihood approach described below allows us to easily integrate over all the other states.

Maximum simulated likelihood

Ignoring first the state-space component, a number of techniques exist for estimating the parameters of mixed effect models, but most of these techniques do not allow for observation error in the independent variable (e.g., $n_{t,a-1}$). In Chapter 3, variance component models were used because the model was linear and normal observation errors were assumed. Here I use a more general approach that allows these assumptions to be relaxed if necessary.

The integrals in the state-space models can only be solved analytically in the linear, normal case (i.e., the Kalman filter); otherwise the integrals are intractable. Nonlinear problems can be handled using either linear approximations to nonlinear models or the "extended" Kalman filter using the Taylor series expansion. The recent advent of easy to access Markov Chain Monte Carlo (MCMC) methods such as BUGS (Bayesian Integration Using Gibbs Sampling) has made it possible to analyze nonlinear state-space problems (Meyer and Millar 1999b; Meyer and Millar 1999a).

Here I introduce a technique, commonly used in econometrics, that provides good accuracy, can be applied in virtually any situation, and is relatively easy to program. The method of maximum simulated likelihood (MSL) approximates the integrals by a series of draws from the distributions over which integration is required (Gourieroux and Monfort 1991). In the example here I have to integrate over the distributions of the true abundance at each stage. The simulated likelihood approach has the same flexibility of Bayesian techniques but without some of the problems (e.g., specification of priors for the estimated parameters).

Remembering the likelihood from Eq. 4.3

$$L(\theta|Y_{t,A}) = \prod_{t=1}^{T}\prod_{a=0}^{A} P(Y_{t,a}|\theta) = \prod_{t=1}^{T}\prod_{a=0}^{A}\int P(y_{t,a}|n_{t,a},\theta)P(n_{t,a}|Y_{t,a-1},\theta)\mathrm{d}n_{t,a},$$

We approximate this by averaging the likelihood over a series of H draws

$$L(\theta|Y_{t,A}) = \prod_{t=1}^{T}\prod_{a=0}^{A} P(Y_{t,a}|\theta) = \prod_{t=1}^{T}\prod_{a=0}^{A}\frac{1}{H}\sum_{r=1}^{H} P(y_{t,a}|n^h_{t,a};\theta)P(n^h_{t,a}|Y_{t,a-1};\theta), \qquad (4.8)$$

where the $n^h_{t,a}$ are a random sample (sampled relative to the probability) from $P(n_{t,a})$.

The MSL approach is implemented as follows:

1. Generate TH random normal deviates of mean zero and variance of unity that are kept throughout the estimation procedure, u^h_t (one for each cohort/simulation combination). These will used to generate samples from $P(n_{t,0})$, the random effects distribution for $n_{t,0}$.

2. Generate THA random normal deviates of mean zero and variance of unity that are kept throughout the estimation procedure, $w^h_{t,a}$ (one for each cohort/simulation/age combination). These are used for the process noise.

3. Provide starting values for μ_0, σ^2_0, and σ^2_ν.

4. Generate the time series of $n^h_{t,a}$ based on the current estimates of the model parameters and the random deviates from 1 and 2; i.e., $n^h_{t,1} = F(n^h_{t,0}) + \nu^h_{t,0}$.

5. Calculate $L(\theta|Y_{t,A})$ by averaging the likelihood over the multiple draws (Eq. 4.8).

6. Repeat the steps 4–5 until convergence.

There are a number of methods for obtaining samples from $P(n_{t,0})$ the simplest is based on the Crude Frequency Estimator (Arias and Cox 1999) which is described above. Here a simple random sample is taken of size TH where the simulation values are drawn relative to their probability. A number of researchers have examined various ways to improve the performance of the simulator (Arias and Cox 1999; Hajivassiliou 1999). The method that I have examined here is the that of 'antithetic-acceleration' which is one of the most common methods used to reduce the variance of the estimator (see Hajivassiliou (1999) for technical details of the improvements possible using antithetic-acceleration). The approach involves changing the way the initial random deviates are generated. For antithetic-acceleration, I first generate $0.5TH$ random normal deviates using the Crude Frequency Estimator. A second $0.5TH$ are obtained by multiplying the first half by -1. The improvement occurs because of the more even sampling of the likelihood surface. This improves the performance of the integration because less function evaluations are required.

As this method involves random sampling to approximate the integral we must be aware of simulation error that occurs with the model converging at different solutions with different seeds to generate the random deviates. To assess this, I ran the model for different seeds for the random number generator and different values for H. The model estimates for the early ages were very unstable for even the largest H that could be used given computational restrictions (e.g., memory). Model estimates for parameters describing the dynamics at other ages were quite stable for as few as 50 simulations. The results presented here were obtained using a sample size of 300, this was the largest H that could be used given limited memory (256MB).

Because we are dealing with likelihoods which are small and are multiplied across the A observations, there can be numerical/computational problems. I encountered situations where some of the simulated $n_{t,a}^{h}$ were very unlikely for some of the data; i.e., a simulated strong year class when the data are for a very weak year class. This led to very low likelihoods, which when multiplied together, led to underflow (computer zero) and the program would crash. This is a very important problem that is often associated with MSL – particularly when a normal likelihood function is used – that is seldom discussed (Hajivassiliou 1999). To overcome this problem I added a small constant to the likelihood function, essentially putting a limit on how low a likelihood could be. This fattens the tails of the distribution but it is no longer a true likelihood. Simulation studies using a range of simple models indicated that the smaller the constant the less the bias in the estimates of the parameters. Here I have used a value of 10^{-9}.

4.3 Results

Two data-specific problems that I encountered led to slightly different model formulations being used with the two example data sets. First, for the fully state-space model there were too many parameters that need to be estimated given the data available for the analysis. For example, with the eight ages observed for Georges Bank, there are 31 parameters (a mean and variance for initial abundance, eight $\sigma^2_{\epsilon_a}$s, and seven each of the $\sigma^2_{\nu_a}$s, λ_as, and γ_as) to be estimated and about half of them are variance terms. Also, the process noise and observation errors are generally highly confounded. Without auxiliary information or many cohorts to share information over, it is not feasible. Here, I have only estimated process noise for the first time period, i.e., eggs to age 0.75 years for both examples. This did not allow me to test if there was significant density-independent mortality after the critical stage (a test of Hjort's hypothesis). In Section 4.4, I will address the data requirements for a more thorough analysis.

Second, as I will show later, there is considerable variation in the relationship between abundance of eggs and the first age for the survey (0.75 years). The model has four parameters to fit to this pattern: (1) observation error in the estimates of eggs; (2) observation error in the estimates at age 0.75; (3) process noise; and (4) density-dependent mortality. There is considerable confounding between the first three and there is such a large amount of total variation (process noise plus observation error) that makes it difficult to reliably detect and estimate the fourth. Estimates of the parameters regarding dynamics at the first stage were uncertain and very unstable across different model runs (while the other parameters behaved well). I have fixed the observation error variance for the spawner biomass at 0.1. This is because it is expected that the abundance of spawners (over a range of ages classes) is much more reliably estimated than juveniles.

These models essentially look at the complete covariance matrix for all of the observations of log abundance. This is shown in the pairwise scatter plots of the data (Figures 4.2 and 4.3). When examining these plots it is important to remember that

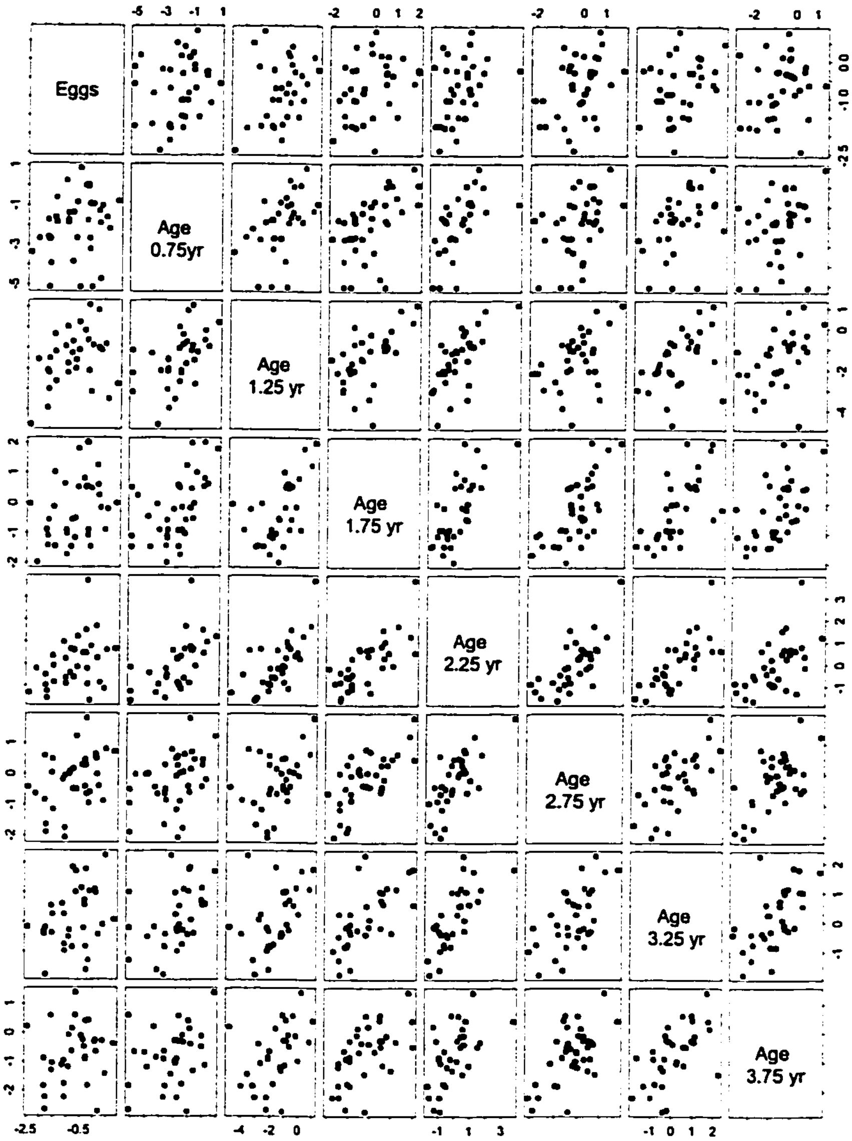


Figure 4.2: Pairwise scatter plots of the logarithm of abundance for cod from Georges Bank.

the points are not assumed to be without error; the observation error variances are estimated in the model. The important feature of the Georges Bank data is that there is considerable scatter between all the juvenile ages and the spawner index. As it is consistent across the ages and that juvenile abundance is quite well correlated (with other juveniles), it suggests large process noise during the first stage (for a small observation error variance). Generally there is less noise in the relationships as you move from left to right (Figure 4.2). A very similar pattern is seen for the North Sea although there are less observations (Figure 4.3).

Because of difficulties estimating the parameters for the first stage (even with one observation error variance fixed) and the simulation noise associated with a random sampling process, four runs (with different seeds) were made for each model. For the Cushing model I present all four runs to illustrate the problem, but only consider the best run (lowest negative log likelihood) for comparison with the other models.

For Georges Bank, estimates of the model parameters for the ages after the first observed (i.e., 0.75 years), were very stable across runs with generally varying less than 10% (but mostly less than 5%) (Table 4.1). The mean describing the distribution of initial abundance was quite variable but the variance was not. The estimated process noise term was relatively constant and indicated considerable process noise during the first stage. The density-dependent term for the first stage was the most unstable but this is not surprising given the high estimated process noise. This led to instability in the estimates of the cumulative density-dependence term. The observation errors were very similar for the last five ages suggesting that it may be possible to estimate a single term for these.

For the North Sea, the general patterns were similar but the inter-run variability was much larger for the first two stages and slightly larger for the juvenile stages than was observed for Georges Bank (Table 4.2). Process noise was estimated to be much higher and this was associated with greater instability in the estimates of the density-dependent term for the first stage. The parameters describing the initial distribution were more consistent across runs than for Georges Bank. The observation errors were again similar for the last four ages (except some problems with $\sigma^2_{\epsilon_6}$).

An important finding is whether there is significant density-dependence during the juvenile stage. This allows us to compare between the weak and strong interpretations

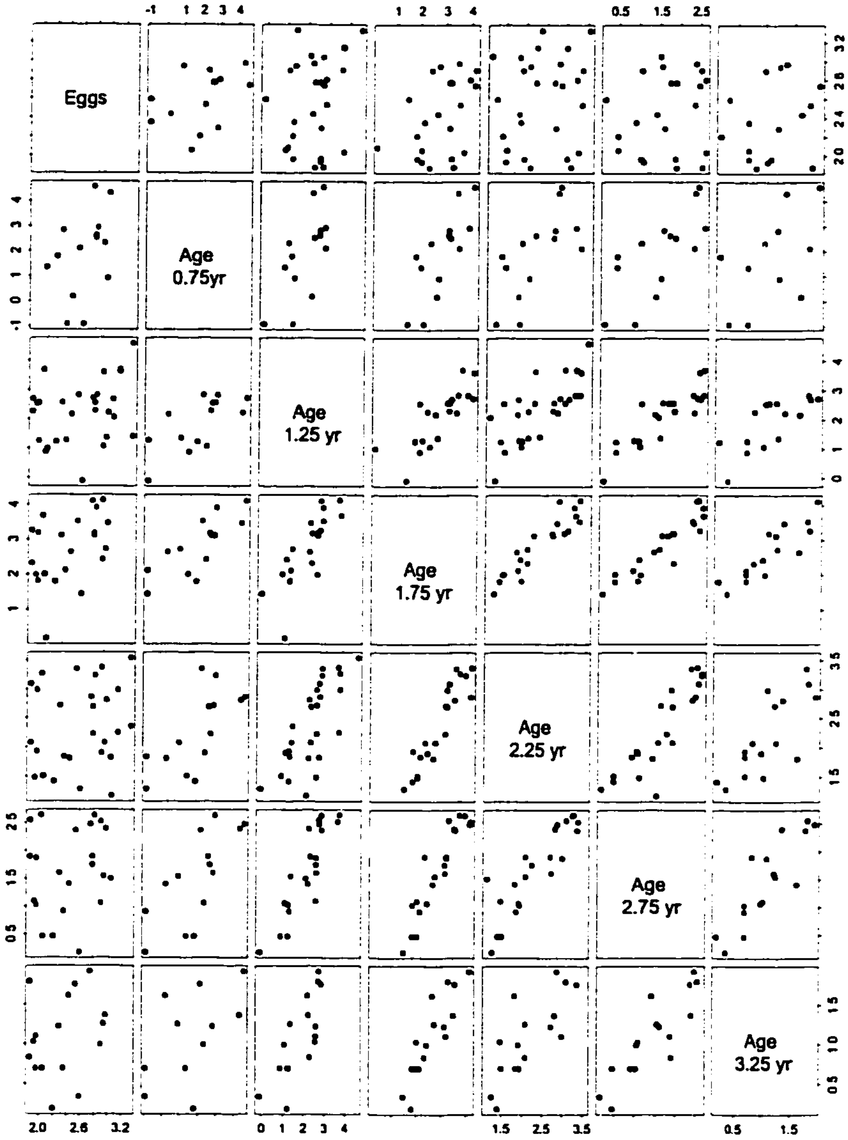


Figure 4.3: Pairwise scatter plots of the logarithm of abundance for cod from the North Sea.

Table 4.1: Estimated parameters for Georges Bank for four model runs of the Cushing model with $H = 300$ and different seeds used for the random number generation. DD and DI are density-dependent and density-independent mortality respectively. Cum. all and Cum. juv. are the cumulative density-dependence over the entire period and juvenile phase respectively. $\mathcal{L}$ is the total negative log likelihood (smaller is better). The assumed base case model results are in the first column. Values in bold were fixed in the estimation.

Parameter	run			
Spawner abundance				
μ_0	-0.841	-1.156	-0.801	-0.582
σ_0^2	0.805	0.857	0.622	0.813
Process noise				
σ_ν^2	0.938	1.019	0.899	0.906
Observation errors				
$\sigma_{\epsilon_1}^2$	**0.100**	**0.100**	**0.100**	**0.100**
$\sigma_{\epsilon_2}^2$	1.234	1.222	1.247	1.241
$\sigma_{\epsilon_3}^2$	0.974	0.960	0.937	0.951
$\sigma_{\epsilon_4}^2$	0.511	0.553	0.580	0.563
$\sigma_{\epsilon_5}^2$	0.535	0.520	0.463	0.476
$\sigma_{\epsilon_6}^2$	0.499	0.500	0.516	0.497
$\sigma_{\epsilon_7}^2$	0.603	0.608	0.597	0.621
$\sigma_{\epsilon_8}^2$	0.669	0.683	0.676	0.687
DD mortality				
λ_2	0.769	0.152	0.334	0.538
λ_3	0.932	0.920	0.970	0.968
λ_4	1.097	1.050	1.001	1.026
λ_5	1.020	1.057	1.112	1.092
λ_6	0.722	0.715	0.681	0.701
λ_7	1.157	1.146	1.177	1.126
λ_8	0.828	0.815	0.816	0.817
Cum. all	0.555	0.104	0.236	0.376
Cum. juv.	0.721	0.683	0.707	0.699
DI mortality + catchability				
γ_2	1.383	2.019	1.852	1.652
γ_3	-0.409	-0.387	-0.491	-0.484
γ_4	-1.405	-1.337	-1.269	-1.297
γ_5	-0.233	-0.245	-0.254	-0.253
γ_6	0.205	0.205	0.202	0.206
γ_7	-0.351	-0.351	-0.360	-0.352
γ_8	0.827	0.821	0.820	0.826
Objective function				
$\mathcal{L}$	345.7	352.7	350.0	348.0

Table 4.2: Estimated parameters for the North Sea for four model runs of the Cushing model with $H = 300$ and different seeds used for the random number generation. A description of other table features is given in Table 4.1.

Parameter	run			
Spawner abundance				
μ_0	2.952	2.874	2.774	2.933
σ_0^2	0.225	0.421	0.302	0.215
Process noise				
σ_ν^2	1.114	1.039	0.640	1.009
Observation errors				
$\sigma_{\epsilon_1}^2$	**0.100**	**0.100**	**0.100**	**0.100**
$\sigma_{\epsilon_2}^2$	1.108	1.105	0.996	1.109
$\sigma_{\epsilon_3}^2$	0.528	0.535	0.648	0.500
$\sigma_{\epsilon_4}^2$	0.317	0.325	0.007	0.341
$\sigma_{\epsilon_5}^2$	0.348	0.340	0.288	0.353
$\sigma_{\epsilon_6}^2$	0.004	0.012	0.291	0.004
$\sigma_{\epsilon_7}^2$	0.250	0.249	0.235	0.250
DD mortality				
λ_2	2.367	1.085	3.500	2.131
λ_3	0.693	0.708	0.544	0.725
λ_4	1.015	0.997	1.146	0.995
λ_5	0.755	0.775	0.820	0.733
λ_6	1.261	1.220	1.059	1.268
λ_7	0.583	0.584	0.689	0.584
Cum. all	0.924	0.423	1.302	0.834
Cum. juv.	0.390	0.390	0.372	0.391
DI mortality + catchability				
γ_2	4.970	0.888	7.692	3.593
γ_3	-0.793	-0.762	-1.255	-0.701
γ_4	-0.549	-0.584	-0.215	-0.605
γ_5	-0.160	-0.106	0.014	-0.223
γ_6	1.248	1.156	0.794	1.252
γ_7	-0.240	-0.239	-0.077	-0.239
Objective function				
$\mathcal{L}$	309.0	320.2	313.5	316.2

described earlier. Estimates for the basecase (lowest negative log likelihood) runs for both regions are presented in Figure 4.4. For Georges Bank there is good evidence that some density-dependent mortality occurs before age 0.75 years, but I can not determine (from the data) if this is during the larval stage or after settlement. There is evidence of some density-dependence during the juvenile stage though the confidence limits do not exclude $\lambda = 1$. The results for the North Sea are quite different. The high estimate of λ_2 implies that the difference between weak and strong cohorts is increased during the early stage. This is difficult to interpret biologically and the high uncertainty in this estimate must be considered. There is strong evidence for density-dependent mortality during the juvenile stage, with the 95% confidence interval excluding 1.

The Ricker and Beverton-Holt models, as parameterized here, lead to some problems when estimating density-dependence. When there is no density-dependence or evidence of the opposite effect (e.g., $\lambda > 1$ for the Cushing) the parameter for the Ricker goes to zero while the parameter for the Beverton-Holt goes to infinity. While it is not simple to constrain either of these models to stop this behaviour, it is possible to constrain the Cushing model by setting an upper bound of 1 for λ. When this was done the inter-run variability for both stocks (but mainly for the North Sea) was further reduced. In both stocks, density-dependence during the first stage was increased (λ_2 was smaller), the decrease for the North Sea was such that it reversed the previous result with density-dependence now stronger over the entire period than the juvenile ages (Figure 4.5).

Results for the three models and the constrained Cushing are provided in Tables 4.3 and 4.4. For Georges Bank the parameter estimates were remarkably similar across models. In particular, the observation error variances were very similar. In instances where the Cushing model estimated $\lambda > 1$, the constrained Cushing model hit the upper bound as did the Beverton-Holt, while the Ricker hit the lower bound. These results are consistent with expectations based on the results of the unconstrained Cushing.

For the North Sea there was less agreement across models, but the qualitative conclusions were similar regarding the relative process noise and the occurrence of density-dependent mortality at a particular stage. The high estimate of λ_2 for the

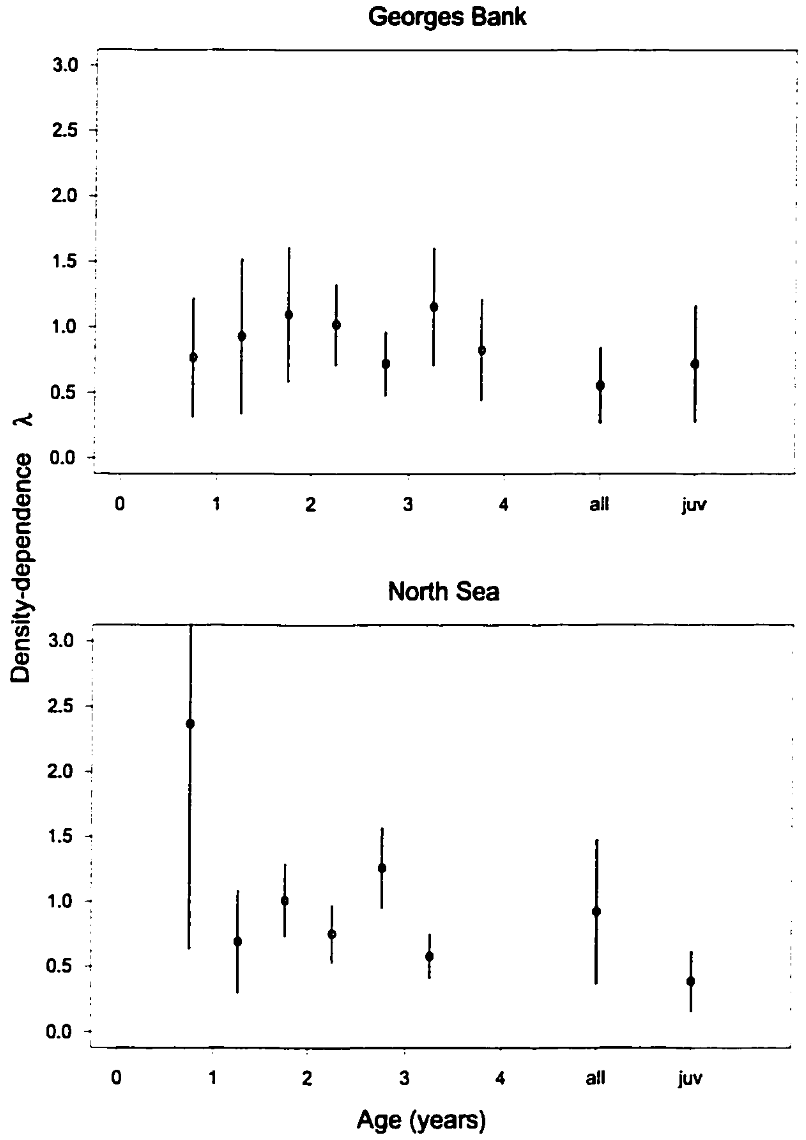


Figure 4.4: Estimates of the density-dependent mortality parameter λ for the Cushing model for each time period, and cumulative over the entire time period and the juvenile ages. Error bars represent the 95% confidence intervals.

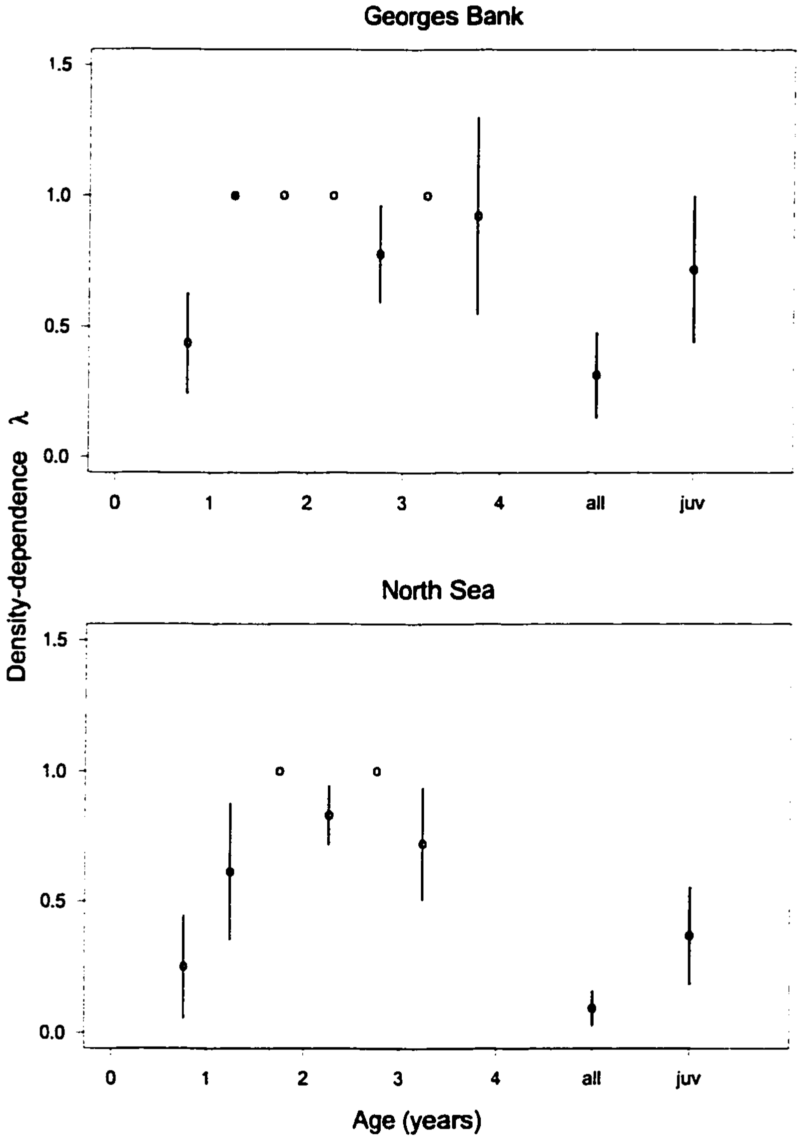


Figure 4.5: Estimates of the density-dependent mortality parameter λ for the Cushing model where an upper bound of 1 was used. Estimates are for each time period, and cumulative over the entire time period and the juvenile ages. Error bars represent the 95% confidence intervals.

Table 4.3: Comparison of parameter estimates for the four models compared for Georges Bank. Note that the density-dependent mortality parameters, λ_a, have different interpretations in each model. Values in bold were fixed in the estimation and values in italics were estimated to be at either the lower or upper bound.

Parameter	Cushing	Cushing ($\lambda \leq 1$)	Ricker	Beverton-Holt
Spawner abundance				
μ_0	-0.841	-0.737	-0.637	-0.736
σ_0^2	0.805	0.701	0.839	0.699
Process noise				
σ_ν^2	0.938	0.833	0.742	0.825
Observation errors				
$\sigma_{\epsilon_1}^2$	**0.100**	**0.100**	**0.100**	**0.100**
$\sigma_{\epsilon_2}^2$	1.234	1.240	1.244	1.244
$\sigma_{\epsilon_3}^2$	0.974	0.966	0.972	0.948
$\sigma_{\epsilon_4}^2$	0.511	0.532	0.542	0.522
$\sigma_{\epsilon_5}^2$	0.535	0.516	0.533	0.511
$\sigma_{\epsilon_6}^2$	0.499	0.500	0.485	0.510
$\sigma_{\epsilon_7}^2$	0.603	0.619	0.594	0.610
$\sigma_{\epsilon_8}^2$	0.669	0.661	0.642	0.655
DD mortality				
λ_2	0.769	0.436	1.267	0.322
λ_3	0.932	*1.000*	*0.000*	*300*
λ_4	1.097	*1.000*	*0.000*	*300*
λ_5	1.020	*1.000*	*0.000*	*300*
λ_6	0.722	0.778	0.080	5.357
λ_7	1.157	*1.000*	*0.000*	*300*
λ_8	0.828	0.924	0.082	12.734
Cum. all	0.555	0.314	-	-
Cum. juv.	0.721	0.719	-	-
DI mortality + catchability				
γ_2	1.383	1.716	0.468	0.247
γ_3	-0.409	-0.565	-0.563	-0.570
γ_4	-1.405	-1.253	-1.257	-1.254
γ_5	-0.233	-0.236	-0.231	-0.247
γ_6	0.205	0.211	0.093	-0.014
γ_7	-0.351	-0.325	-0.328	-0.337
γ_8	0.827	0.843	0.734	0.755
Objective function				
$\mathcal{L}$	345.7	347.7	346.1	346.8

Cushing model is not consistent with the results for the other three models. However, it should be noted that the estimates for the Cushing were both uncertain and unstable across runs.

It is possible to compare the negative log likelihood of each model to see which is most consistent with the data. For both stocks the Cushing model provides the best fit to the data. The Ricker generally provides a poor fit to the data. This suggest that there is little evidence of over-compensation in these two stocks.

4.4 Discussion

Since the pioneering work of Hjort (1914) there has been interest in determining the (critical) period at which the year class strength of a cohort is determined. Here I have described a modelling framework that allows this question to be addressed in ways not previously possible. While I have not presented a full meta-analysis of this problem, the development of these models represents a critical contribution to this problem. In this section I will focus on four aspects of this study: (1) the analysis of the two cod data sets; (2) the dynamic models using to describe the 'transition' part of the state-space approach; (3) the state-space approach and the implementation using maximum simulated likelihood; and (4) how the 'pilot' analysis described here provides insights into how to best perform a full meta-analysis.

Implications of findings

The data for the two examples used here were not ideal for testing the core critical period hypothesis of Hjort, but did provide interesting findings and confirmations of previous work. I found that there was a considerable increase in variation in abundance between eggs and the age at which cod were first adequately sampled by the bottom trawl surveys. These observations are consistent with Hjort's hypothesis, but I did not have data at early ages (e.g., larval stages) to allow determination of the critical age. My analyses focussed on variation in density-dependent mortality at ages where cod were vulnerable to the trawl survey. The estimation of significant density-dependent mortality was consistent with the 'weak' interpretation of Hjort's hypothesis and the findings of early studies (Myers and Cadigan 1993b; Bradford

Table 4.4: Comparison of parameter estimates for the four models compared for the North Sea. A description of other table features is given in Table 4.3.

Parameter	Cushing	Cushing ($\lambda \leq 1$)	Ricker	Beverton-Holt
Spawner abundance				
μ_0	2.952	3.059	2.749	2.744
σ_0^2	0.225	0.282	0.459	0.520
Process noise				
σ_ν^2	1.114	1.609	1.247	1.520
Observation errors				
$\sigma_{\epsilon_1}^2$	**0.100**	**0.100**	**0.100**	**0.100**
$\sigma_{\epsilon_2}^2$	1.108	0.997	1.101	1.018
$\sigma_{\epsilon_3}^2$	0.528	0.646	0.637	0.663
$\sigma_{\epsilon_4}^2$	0.317	0.005	0.004	0.004
$\sigma_{\epsilon_5}^2$	0.348	0.292	0.288	0.284
$\sigma_{\epsilon_6}^2$	0.004	0.294	0.286	0.288
$\sigma_{\epsilon_7}^2$	0.250	0.233	0.233	0.229
DD mortality				
λ_2	2.367	0.250	0.033	7.905
λ_3	0.693	0.615	0.005	28.065
λ_4	1.015	*1.000*	*0.000*	*300*
λ_5	0.755	0.830	0.008	117.745
λ_6	1.261	*1.000*	*0.000*	*300*
λ_7	0.583	0.722	0.055	11.978
Cum. all	0.924	0.092		
Cum. juv.	0.390	0.369		
DI mortality + catchability				
γ_2	4.970	-1.142	-0.178	-0.097
γ_3	-0.793	-1.120	-0.438	-0.704
γ_4	-0.549	-0.548	-0.578	-0.582
γ_5	-0.160	0.048	0.353	0.349
γ_6	1.248	0.655	0.649	0.624
γ_7	-0.240	-0.020	0.098	0.044
Objective function				
$\mathcal{L}$	309.0	314.0	316.6	313.7

and Cabana 1997). The three population dynamics models compared, almost always agreed in terms of the presence of density-dependent mortality in the juvenile stages and all agreed that there was substantial variation in density-independent mortality.

Improvement of dynamic models

Though the models described here are quite complex and cover a range of linear and nonlinear dynamics there are many ways in which they can be improved. First, I have assumed that adjacent cohorts are independent though Myers and Cadigan (1993a) reported strong density-dependent mortality between adjacent cohorts for cod and a variety of other groundfish. There may also be trends over time which should be modelled. Second, I have ignored potential "year-effects" in the observation error terms for the surveys. I assumed that the errors were independent for estimates of abundance of different ages from a survey (in a year). Myers and Cadigan (1993a) demonstrated how to incorporate other types of measurement error, for example if errors among ages in the same survey years are correlated. Francis et al. (2001) found that year effects are common in trawl surveys. A more appropriate model would attempt to separate observation errors into age- and year-specific components. Third, density-independent mortality and trawl survey catchability were confounded in the analysis. Though catchability may vary from year to year (Francis et al. 2001) there may be advantages from including some estimates of relative catchability at age/size in the analysis (e.g., from the analysis described in Chapter 5). Fourth, it appears that there is a decreasing trend with age in the observation errors. It might be advantageous to either share errors across some ages or include a simple model to describe this trend.

Modelling approaches

Modelling the population dynamics of young fish presents some important problems reflecting a number of nasty features often found in fisheries data. The most important feature, though often ignored, is the large amount of observation error in our estimates of abundance (Myers and Cadigan 1993b). I have addressed this problem in two ways, first, by explicitly accounting for this in the modelling and second, by sharing information across multiple cohorts using meta-analytical principles. The maximum simulated likelihood approach used here offers the opportunity to develop models that consider all the important features of the data and dynamics. Previous researchers have had considerable difficulty estimating the parameters of these models. The approach here has allowed the development of models not previously possible.

The ability to include nonlinear models, such as the Beverton-Holt, including mixed effects and observation error represents a critical advance. Seldom are important questions in ecology linear, so this has represented a considerable restriction. Important dynamics such as depensation and over compensation (Myers et al. 1995) generally require nonlinear models. The approach described here will accommodate such models.

The method of maximum simulated likelihood (MSL), commonly used in econometrics, approximates integrals by drawing multiple samples from the probability distribution of interest. Here I used this approach to integrate over both the distribution of initial abundance and the process noises. While this approach is extremely flexible and can be adapted to almost any situation, it does have some potential weaknesses. Presently there are no custom-built packages for implementing this approach so users must program problems themselves in the modelling package of their choice. Here I used AD Model builder because of its powerful nonlinear minimizer and its computational speed.

Also, as MSL is a simulation-based or computer-intensive method there are technical issues relating to the implementation of the approach, e.g., how many samples to take. There are no set rules for this and I found that different models (and different data sets) required different numbers of simulations to overcome simulation noise (i.e., getting different results with different random seeds). In such circumstances it may be appropriate to combine the estimates of the parameters (including the covariance structure) across the multiple model runs.

Though I have not demonstrated this here, the assumption of normal observation errors can be relaxed using the approach developed in this chapter. This will be particularly useful for fisheries applications and there will be two obvious extensions that will be possible: (1) the use of robustified likelihood functions to overcome the often high levels of contamination in fisheries data (Fournier et al. 1990; Chen et al. 2000); and (2) modelling data with many observed zeros (Lo et al. 1992) as is often the case with survey data. This will allow for an extremely generalized framework for state-space models.

The state-space implementation can be memory intensive when a relatively large number of simulations are required. The application here is particularly troublesome due to the type of model (i.e., trying to separate process noise from observation error). While the models did not take long to converge (about 3 minutes on a Pentium III 650) the amount of memory required can be substantial, potentially limiting 'desktop' computer implementations. For the Georges Bank example with $H = 300$, about 100MB of memory was required.

Meta-analysis of mortality patterns in fishes

The general aim of this chapter was to see if it was possible to develop an approach to enable analytical examination of density-dependent mortality in fishes. The method here relies heavily on sharing information across cohorts, this is of course entirely consistent with the principles of meta-analysis. The one meta-analytical step not addressed here was the combination of information across stocks (only two were considered). This type of approach has led to many interesting findings (Myers and Cadigan 1993b; Bradford and Cabana 1997) and is the obvious focus of future research.

This pilot study has identified a number of important issues for any future meta-analysis. First, it is very difficult to estimate process noise and observation error simultaneously. This is critical to test for variation in density-dependent mortality at different life stages. To overcome this it might be necessary to obtain independent estimates of observations errors. If estimates of abundance come from standardized surveys, it should be possible to obtain reasonable estimates from the survey data. Second, it will be necessary to examine alternative parameterizations of population models used to describe density-dependent mortality. If it is possible to standardize the data, it should allow for simpler interpretation and combination of model parameters across populations (Myers et al. 2001).

Chapter 5

Length-specific research trawl survey catchability

5.1 Introduction

In order to assess and understand changes in ecosystems, scientific surveys are often carried out to estimate abundance of species. However, before we can best use these survey results we must translate the estimates to absolute abundance. Two fundamental difficulties with most scientific surveys are: (1) not all sizes and species are sampled equally efficiently by the survey gear; and (2) it is difficult to translate from an index to an estimate of absolute abundance. Both of these problems occur for fisheries research trawl surveys, which are the most common approach used to obtain fishery independent estimates of abundance. It is sometimes possible to estimate the efficiency of research trawl surveys either from experimental studies (Somerton et al. 1999) or by comparing survey estimates to absolute abundance obtained from catch-age-analyses (Sparholt 1990). Unfortunately, only a few reliable estimates from experiments are available and catch-age-analyses are usually carried out only for the most valuable commercial species. Thus, we have no estimates for the majority of species caught in research surveys.

I will describe a meta-analysis of length-specific catchability for a range of species to marine bottom trawl research surveys. I will demonstrate how meta-analysis can used used to combine estimates of a functional relationship, i.e., the relationship between the size of a fish and its catchability for a bottom trawl survey. This extends traditional meta-analysis where either a single quantity is combined or a single parameter from a curve, e.g., the slope at the origin for the spawner recruitment curve. The rationale for using meta-analysis for this problem is two-fold: first, a single estimate of catchability (derived independent of all other data on the problem) can be extremely unreliable and using it could lead to substantial biases in reconstructed biomass; and second, we need some way to combine estimates for similar groups of species to obtain estimates that can be applied to the majority of species for which no direct data are available.

5.1.1 Trawl survey catchability

In general, catchability describes how the abundance and size composition of a species differs between the population and the survey catch. Catchability is the value required to scale the survey estimate of abundance to absolute abundance. We can define absolute abundance N_l of length class l,

$$N_l = \frac{c_l}{q_l} \cdot \frac{A}{a}, \tag{5.1}$$

where q is the catchability scalar, c is the mean catch per tow, a is the area swept by one trawl tow, and A is the total survey area. When abundance is measured as the biomass, this leads to $c\frac{A}{a}$ being referred to as the "swept-area biomass estimate". It is also often know as "trawlable biomass" or as "minimum trawlable biomass".

The interpretation of an estimate of catchability is simplified by the calculation of expected density D of fish,

$$D = N/A, \tag{5.2}$$

which, when scaled to account for the area-swept by the tow (a), leads to an expected catch rate c^*. Thus catchability is the observed catch rate divided by the expected,

$$q = c/c^*. \tag{5.3}$$

Catchability less than 1 indicates that less fish were caught than predicted in the path (assumed here to be the width of the net opening multiplied by the horizontal distance travelled) of the trawl, conversely catchability greater than 1 indicates that more fish were caught than predicted to occupy the area trawled. Catchability could be greater than 1 if the trawl gear was able to herd fish into the path of the net or if the calculation of area-swept abundance involved extrapolations to untrawlable regions where the density of fish was lower (than the surveyed region). Understanding these processes has important consequences for the interpretation of survey indices and subsequent management decisions (Godø and Walsh 1992; Somerton et al. 1999).

Most previous attempts to scale research surveys involved transforming relative abundance to absolute abundance using a single species-specific scalar. This is done by scaling the total abundance of the species caught in the survey to that in the survey area. This will be referred to as a "bulk estimate" of catchability (Harley et al. 2001). Bulk estimates do not utilize information on the length/age-structure of the population or the length/age-specific selectivity of the survey gear.

The catchability process can be described by three components as suggested by Edwards (1968): (1) availability to the gear (the vertical distribution of the species); (2) vulnerability (including herding effects, net avoidance, and mesh selectivity); and (3) spatial/seasonal factors (the spatial distribution of the fish in the trawlable and untrawlable regions). The vertical distribution of a number of species has been found to vary with time of day (Casey and Myers 1998). Fish behaviour in and around trawl gear is known to influence vulnerability (Wardle 1983), as well as mesh size and the size of the catch (Suuronen and Millar 1992; Godø et al. 1999). The second component encompasses the more general field of gear selectivity which includes the properties of both commercial and survey gears. While the approach in the field of gear selectivity has used direct observational studies and experiments, the approach here is indirect where all three components of catchability are considered together.

Edwards (1968) considered each of the three components of catchability described above to assign catchability estimates for a number of species found off the New England coast. Sparholt (1990) assigned estimates of catchability from ten species from a multispecies virtual population analysis (MVPA) of the North Sea to a range of other species found in the area. The assignment was based on morphological and behavioural similarities between species. Casey (2000) used the estimates from

Edwards (1968) and applied a correction for diel variability in catchability to allow the reconstruction of absolute abundance for different fish groups on the Southern Grand Bank. These studies used a single scalar for catchability and thus ignored the length-specific processes that are known to influence gear selectivity.

The length-specific estimates developed here have three main advantages over the bulk estimates. First, they allow separate consideration of adults and juveniles. A single bulk estimate would scale up both adult and juvenile survey abundance the same. This leads to difficulties in estimating natural or total mortality. Second, in instances where there are large changes in the size composition of the population, a single bulk catchability will introduce extra bias into estimates of absolute abundance. Third, it is simpler to assign estimates of catchability to other species. For example, with a single bulk estimate it is difficult to assign estimates of catchability across species from a family if there are large differences in size (such as large and small gadids). Estimates of length-specific catchability can help overcome this difficulty.

I collected estimates of research trawl survey catchability estimated from stock assessments worldwide. I used hierarchical models that incorporate information from other data sets to provide improved estimates for a single data set, as well as the standard non-hierarchical techniques that assume that the only information on catchability-at-length for a given data set, is contained within the data set. Multiple estimates for the same or similar species were combined using hierarchical Bayesian techniques to obtain meta-analytical estimates of catchability-at-length (and associated uncertainty) for a number of species groups. I propose that results of this analysis could be used both as auxiliary information (or "priors") in stock assessment models and to assist in the reconstruction of fish community structure from trawl survey time series, e.g., (Sparholt 1990; Casey 2000).

5.2 Methods

This approach to this problem has three components that are described in the sections below: (1) identifying and collating relevant data sets; (2) developing a range of models to describe the relationship between length and catchability; and (3) integrating the models and data in a meta-analytical framework to allow the combination of estimates from similar species.

5.2.1 Data

I attempted to compile all the data on research trawl catchability in the world. I focused on four regions: New Zealand, the North Sea, and the east and west coasts of North America as these represent most of the regions where trawl surveys and complex assessments (i.e., age and length-structured) are used. I compiled 60 data sets with age/length specific catchability estimates for 19 species where the following criteria were met: (1) an age-structured (or length-structured) stock assessment was performed using research trawl survey data; (2) the survey covered all, or most, of the stock assumed for modelling purposes; and (3) there were sufficient age/length classes to allow the fitting of a logistic curve to the data. A further nine data sets with insufficient observations (of ages or sizes) for curve fitting were used to estimate bulk catchability (Harley et al. 2001). Estimates of catchability were obtained from published and unpublished stock assessment reports produced by fisheries management agencies. Stock assessment scientists were contacted directly when extra details or clarification were required. The sample unit or data set consisted of length-specific catchability estimates for a population from a survey. Where multiple research surveys had been conducted, several data sets were available for a single population. All trawl surveys were designed incorporating a random stratified sampling strategy, although for some surveys (e.g., English Groundfish Survey, EGFS) the stations are not re-randomized each year. For each data set I collected the catchability estimates at age (or size), estimates of uncertainty (where available), biological parameters to allow the transformation of age to length, and details of the survey.

Table 5.1: Catchability data used in the meta-analysis. NAFO = Northwest Atlantic Fisheries Organization; ICES = International Council for the Exploration of the Sea; EGFS = English Groundfish Survey; IBTS = International Bottom Trawl Survey; NZ = New Zealand.

Species/Stock	**Survey**	**Source**
Atlantic cod *Gadus morhua*		
NAFO 2J+3KL (Labrador and N. Grand Bank)	fall	Shelton et al. (1996)
NAFO 3NO (Southern Grand Bank)	fall	Davis et al. (1996)
NAFO 3NO (Southern Grand Bank)	spring (a) - Canadian	Davis et al. (1996)
NAFO 3NO (Southern Grand Bank)	spring (b) - USSR	Davis et al. (1996)
NAFO 3Ps (St. Pierre Bank)	spring	Brattey et al. (1999)
NAFO 3Ps (St. Pierre Bank)	winter - French	Brattey et al. (1999)
NAFO 4T (St. Lawrence)	fall	Chouinard et al. (1999)
NAFO 4Vn (Sydney Bight)	summer	Mohn et al. (1998)
NAFO 4VsW (E. Scotian Shelf)	spring	Mohn et al. (1998)
NAFO 4VsW (E. Scotian Shelf)	summer	Mohn et al. (1998)
NAFO 5Z (Georges Bank)	fall	O'Brien and Cadrin (1999)
NAFO 5Z (Georges Bank)	spring (a) - Yankee 36	O'Brien and Cadrin (1999)
NAFO 5Z (Georges Bank)	spring (b) - Yankee 41	O'Brien and Cadrin (1999)
NAFO 5Y (Gulf of Maine)	fall	Mayo (1998)
NAFO 5Y (Gulf of Maine)	spring	Mayo (1998)
ICES IV & IIIa (North Sea and Skaggerak)	fall (EGFS)	ICES (1999)
ICES IV & IIIa (North Sea and Skaggerak)	winter (IBTS)	ICES (1999)

Table 5.1: Catchability data, continued.

Species/Stock	Survey	Source
Haddock *Melanogrammus aeglefinus*		
NAFO 4TVW (St. Lawrence and Scotian Shelf)	summer	Frank et al. (1997)
NAFO 4X (Browns Bank)	summer	Hurley et al. (1999)
ICES IV & IIIa (North Sea and Skaggerak)	fall (EGFS)	ICES (1999)
ICES IV & IIIa (North Sea and Skaggerak)	winter (IBTS)	ICES (1999)
NAFO 5Z (Georges Bank)	fall	Gavaris and Van Eeckhaute (1998)
NAFO 5Z (Georges Bank)	spring (a) - Yankee 36	Gavaris and Van Eeckhaute (1998)
NAFO 5Z (Georges Bank)	spring (b) - Yankee 41	Gavaris and Van Eeckhaute (1998)
NAFO 5Z (Georges Bank)	spring (c) - Canadian	Gavaris and Van Eeckhaute (1998)
Whiting *Merlangius merlangus*		
ICES IV & IIIa (North Sea and Skaggerak)	fall (EGFS)	ICES (1999)
ICES IV & IIIa (North Sea and Skaggerak)	winter (IBTS)	ICES (1999)
Hake *Merluccius australis*		
NZ HAK1 (Sub Antarctic)	fall	Dunn et al. (2000)
NZ HAK1 (Sub Antarctic)	spring	Dunn et al. (2000)
NZ HAK4 (Chatham Rise)	summer	Dunn et al. (2000)
Hoki *Macruronus novaezelandiae*		
NZ HOK1	summer (a) - Chatham Rise	Cordue (2000)
NZ HOK1	fall (Sub Antarctic)	Cordue (2000)

Table 5.1: Catchability data, continued.

Species/Stock	**Survey**	**Source**
NZ HOK1	summer (b) - Southland	Cordue (2000)
Silver hake *Merluccius bilinearis*		
NAFO 4VWX (Scotian Shelf)	summer	Showell (1998)
Pollock / saithe *Pollachius virens*		
ICES IV & IIIa (North Sea and Skaggerak)	fall (EGFS)	ICES (1999)
Walleye pollock *Theragra chalcogramma*		
Eastern Bering Sea	fall (AFSC)	J Ianelli (NMFS, 7600 Sand Point Way NE, Seattle, WA)
American plaice *Hippoglossoides platessoides*		
NAFO 5Y (Gulf of Maine)	fall	O'Brien et al. (1999)
NAFO 5Y (Gulf of Maine)	spring	O'Brien et al. (1999)
Yellowtail flounder *Limanda ferruginea*		
Cape Cod (CCod)	fall	Cadrin (2000a)
Cape Cod (CCod)	spring	Cadrin (2000a)
NAFO 5Z (E. Georges Bank)	fall	Cadrin (2000b)
NAFO 5Z (E. Georges Bank)	spring	Cadrin (2000b)
Southern New England (SNE)	fall	Cadrin (2000c)
Southern New England (SNE)	spring	Cadrin (2000c)
Ling *Genypterus blacodes*		

Table 5.1: Catchability data, continued.

Species/Stock	**Survey**	**Source**
NZ LIN3&4 (Chatham Rise)	summer	Horn et al. (2000)
NZ LIN5&6 (Southern Plateau)	fall	Horn et al. (2000)
NZ LIN5&6 (Southern Plateau)	spring	Horn et al. (2000)

After careful consideration, 47 estimates of length-specific catchability from 11 species were used in the hierarchical model (Table 5.1). I excluded 13 data sets for the following reasons:

- Two data sets had length-specific catchability curves with a strong descending right-hand limb, a pattern inconsistent with the majority of the data and the logistic curve used in the meta-analysis (atka mackerel *Pleurogrammus monopteryguis* and Pacific cod *Gadus macrocephalis* from the Bering Sea and the Gulf of Alaska respectively).
- Five flatfish data sets from Bering Sea surveys were excluded because the sea floor characteristics and gear types were not comparable to that encountered on the Scotia Shelf surveys. The flatter bottom in the Bering Sea results in reduced probability of escape below the ground gear and thus higher estimates of catchability (D. Somerton, National Marine Fisheries Service (NMFS), 7600 Sand Point Way NE, Seattle, WA, pers. comm.).
- One data set for yellowtail flounder (*Limanda ferruginea*) from Southern New England was excluded as the survey gear was specifically set up to target flatfish (S. Cadrin, NMFS, 166 Water Street, Woods Hole, MA, pers. comm.) and was not comparable with the other yellowtail flounder data sets.
- Five data sets were excluded because of unrealistic catchability estimates (many greater than five). Three of these were for the New Zealand red cod (*Pseudophycis bachus*). The assessments for these stocks were very uncertain and highly sensitive to alternative model formulations (Annala et al. 2000). The other two data sets that were excluded were for American plaice (*Hippoglossoides platessoides*) in Northwest Atlantic Fisheries Organization (NAFO) areas 3LNO (Morgan et al. 1999) and 4T (Marin et al. 1998) where the catchability estimates were around 4.

In all cases, it would be possible to include these unusual data sets in the analysis and the meta-analytical framework could be used to obtain perhaps better estimates for these data, but it was felt that including them would only serve to unrealistically increase the uncertainty in the meta-analytic estimates of catchability.

The remaining data sets were divided into seven groups for the hierarchical modeling (Table 5.2). The two gadoid groups include species from the order Gadiformes. Atlantic cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*) were modelled both individually and combined as demersal gadoids. The combined analysis was performed to allow the allocation of catchability estimates from this group to other species. The demersal gadoid group also included whiting (*Merlangius merlangus*). The pelagic gadoid group included data from two New Zealand fisheries, hoki (*Macruronus novaezelandiae*) and hake (*Merluccius australis*), saithe (*Pollachius virens*) from the North Sea, walleye pollock (*Theragra chalcogramma*) from the Bering Sea, and silver hake (*Merluccius bilinearis*) from the Scotian Shelf.

New Zealand ling (*Genypterus blacodes*) was considered as a separate group. Ling is an elongated eel-shaped fish that generally feeds at the bottom on crustaceans and fish (Annala et al. 1999). It lives to about 30 years with females obtaining lengths of 160 cm. The Chatham Rise trawl survey catches ling of all sizes (personal observation). This species was included as it might provide estimates of catchability for species such as wolffish and eelpout that are common on the Scotian Shelf.

It was difficult to determine which species to include in the flatfish group. All the Bering Sea data sets were excluded as noted above, and many of the assessments with very high catchability estimates were for flatfish species. The data used in the analysis comprised six data sets for yellowtail flounder and two for American plaice.

Length-specific catchability

The estimates of length-specific catchability used in this analysis were derived from catch-at-age analyses. There are two approaches to catch-at-age analyses: (1) traditional virtual population analysis (VPA) methods that assume catch-at-age is known exactly and estimate catchability-at-age for each year, and (2) methods that assume error in the catch-at-age data and estimate a single age-specific selectivity curve for each gear/research vessel type. Estimates derived from both modeling approaches were used here.

The traditional VPA (and its extensions) assumes that the catch-at-age is known without error and uses research survey data to "tune" the analysis to estimate the number of survivors in the last year and the fishing mortality. In order to make these

Table 5.2: Groupings and numbers of data sets used for the hierarchical meta-analysis.

Group/Species	**Common name**	**Data sets**
Atlantic cod		
Gadus morhua	Atlantic cod	17
Haddock		
Melanogrammus aeglefinus	haddock	8
Demersal gadoids		
Gadus morhua	Atlantic cod	17
Melanogrammus aeglefinus	haddock	8
Merlangius merlangus	whiting	2
Pelagic gadoids		
Macruronus novaezelandiae	New Zealand hoki	3
Merluccius australis	New Zealand hake	3
Merluccius bilinearis	silver hake	1
Pollachius virens	pollock or saithe	1
Theragra chalcogramma	walleye pollock	1
Ling		
Genypterus blacodes	New Zealand ling	4
Flatfish		
Hippoglossoides platessoides	American plaice	2
Limanda ferruginea	yellowtail flounder	6

estimates it is also necessary to estimate catchability-at-age. It is usually assumed that estimation errors are lognormal, i.e.,

$$S_{ta} = q_a N_{t,a} \exp(\epsilon_{ta}),$$

where $\epsilon_{ta} \sim \mathrm{N}(0, \sigma^2)$, $N_{t,a}$ is the estimated number of fish at age a in year t from the VPA, q_a is the catchability at age a for the survey, S_{ta} is the survey estimate of the numbers, and ϵ_{ta} is the estimation error. The log catchability Q_a is estimated using a simple average

$$\hat{Q}_a = \sum_{t=1}^{T} (\log(N_{t,a}) - \log(S_{t,a}))/T,$$

with the usual standard error.

Other stock assessment methods separate fishing mortality into age- and year-specific components and use a different approach to estimate catchability (Quinn and Deriso 1999). Now we estimate a single age or length-specific selectivity curve for each survey series. In this case, a more complex objective function is minimized. Here we need to consider two parts of the relationship. First, the total biomass for a species is estimated from the survey in each year, B_t, and this is compared to the biomass predicted by the model, again usually assuming a lognormal error term. That is,

$$B_t = \left(\sum_a N_{t,a} q w_a s_a \right) \exp(\epsilon_t),$$

where $\epsilon_t \sim \mathrm{N}(0, \sigma^2)$, w_a is average weight-at-age, and the survey catchability, q, is assumed to be constant for each survey.

In a separate likelihood, the proportion of fish caught in a survey of age a in a given year is used to estimate the selectivity at age s_a (often standardized to a maximum of 1), typically by assuming the proportion at age follows a multinomial, or related distribution. In this case the estimate of catchability-at-age is

$$\hat{q}_a = \hat{q} \hat{s}_a.$$

Estimates of catchability-at-age were converted to catchability-at-length using a von Bertalanffy growth function. A single curve was used for each population.

5.2.2 Catchability models

Studies of gear selectivity provide a useful starting point for potential functional forms for the relationship between length and trawl survey catchability. Millar (1995) described seven different functional forms for modelling hook and gillnet selectivity that were considered. I compared a gamma density curve and a logistic curve. For each of these curves a scaling parameter was required as the density under the gamma curve was not required to be 1, and the logistic curve was not necessarily asymptotic at 1. These two models were chosen as they provide two differing hypotheses. The gamma probability curve is a flexible curve that allows for descending limbs on both sides while the logistic model is monotonic.

The gamma probability density function with unit scale and shape parameter (α), location parameter (m), and vertical scaling parameter (γ), leads to the gamma model for catchability-at-length $q_i(l)$ for data set i

$$q_i(l) = \gamma'_i \frac{(l_l - m)^{\alpha-1} e^{-(l_l-m)}}{\Gamma(\alpha)} \exp(\varepsilon_{i,l}). \tag{5.4}$$

Assuming lognormal observational errors, the three parameter logistic model has the form

$$q_i(l) = \gamma'_i \frac{\exp(\alpha_i + \beta_i l)}{1 + \exp(\alpha_i + \beta_i l)} \exp(\varepsilon_{i,l}), \tag{5.5}$$

where γ_i is the logarithm of maximum catchability, α_i and β_i are the traditional parameters describing the two parameter logistic model, and $\varepsilon_{i,l} \sim \mathrm{N}(0, \sigma_i^2)$.

A comparison of the fit of these two functions to the data is provided by Harley et al. (2001) and not repeated here. Overall the logistic model generally provided a better fit to the data so this was the only model considered in the meta-analysis. To constrain the estimates of γ'_i to be positive we estimate $\gamma_i = \log(\gamma'_i)$. This was also found to increase the speed of convergence and improve the mixing of the Bayesian model described in the sections below.

5.2.3 Modeling approaches

Both independent (non-hierarchical) and mixed effect (hierarchical) approaches were used to model catchability-at-length. Standard non-hierarchical modelling methods only utilize information from the data set in question, while hierarchical models have the ability to combine information from the data set with extra information from other data sets (Liermann and Hilborn 1997). Both approaches and their implementation are described below.

Non-hierarchical models

Model 1 (Individual data sets model):

In this case, I treat the parameters as fixed and survey-specific for each data set. It is based on the unreasonable assumption that there is no relationship among estimates of catchability for a species from different surveys. This is the model shown in Equation 5.5 where separate γ, α, β, and σ^2 are estimated for each data set. This model was fit using maximum likelihood rather than within a Bayesian framework. However, a Bayesian analysis with uninformative or reference priors should give the same results. Model 1 was estimated using the NLMINB function in S-PLUS (S-PLUS 1999). NLMINB is a simple nonlinear function minimizer which allows the user to specify bounds for the estimated parameter (though this was not done here). The output from this model are the values for the parameters that maximize the likelihood function given the data.

Hierarchical models

Hierarchical Bayesian modelling is a technique that allows the combination of data from independent sources (Gilks et al. 1994; Carlin and Louis 1996). When different data sets represent individual populations, such models assume that there is something in common about members of a taxonomic group, and attempt to model that similarity in some way (Myers and Mertz 1998). Recently, the hierarchical Bayesian approach has enjoyed considerable popularity, because it permits estimation of a very broad class of models and may produce more realistic assessments of the uncertainty of parameter estimates (Carlin and Louis 1996).

Bayesian estimation

Simply stated, the Bayesian estimation procedure combines information from two sources. The first is the statistical model of the data, $\mathbf{y}$, and the unknown parameter(s), θ, in the form of a probability distribution $f(\mathbf{y}|\theta)$. The second is the prior probability of the parameters $\pi(\theta|\eta)$, where η are the parameters of the prior distribution, known as the "hyper-parameters" (Carlin and Louis 1996). Bayesian inference is based on the posterior distribution $p(\theta|\mathbf{y},\eta)$ derived using Bayes Theorem,

$$p(\theta|\mathbf{y},\eta) = \frac{f(\mathbf{y}|\theta)\pi(\theta|\eta)}{\int f(\mathbf{y}|\mathbf{u})\pi(\mathbf{u}|\eta)d\mathbf{u}}. \tag{5.6}$$

This posterior distribution is used for Bayesian statistical inference as opposed to the frequentist approach which uses $f(\mathbf{y}|\theta)$ for inference. When θ consists of a vector of unknown parameters, this posterior distribution represents the "joint" posterior distribution. By integrating out the other parameters it is possible to obtain individual or "marginal" posterior distributions for each parameter of interest. In our case here, the data, $q_i(l)$, are model estimates of catchability at length from a species and survey and the vector θ contains the parameters to be estimated,

$$\theta = \{\alpha, \beta, \gamma_1, \ldots, \gamma_k, \sigma_1^2, \ldots, \sigma_k^2\}.$$

where k is the number of data sets in each group.

Model 2 (Random γ's, single α and β model):

This hierarchical model assumes a single α and β for all of the data sets within each meta-analysis group but allows variation in the height parameter, γ_i. Thus, it is assumed that there is one single general shape for the logistic curve for catchability across the group but the height of the curve varies across surveys,

$$q_i(l) = \gamma_i' \frac{\exp(\alpha + \beta l)}{1 + \exp(\alpha + \beta l)} \exp(\varepsilon_{i,l}), \tag{5.7}$$

where $\varepsilon_{i,l} \sim \mathrm{N}(0, \sigma_i^2)$. The assumed hierarchical structure for the γ_i is shown in Figure 5.1. The γ_i are assumed to come from a normal prior distribution which is parameterized with a mean μ, and variance σ^2 (the γ_i' were lognormally distributed),

$$\gamma_i \sim \mathrm{N}(\mu, \sigma^2).$$

The hyper-parameters are assumed to come from their own prior (hyper-prior) distributions. A diffuse normal prior is used for the mean, μ (Figure 5.1). This represents a general state of ignorance about the relative probability of different values of μ. (Note, this does imply that γ_i' is non-negative). The conjugate prior for the variance of a normal distribution is the inverse gamma (Carlin and Louis 1996), here I assume that the inverse of the variance, or precision, has a prior probability described by a gamma distribution. I parameterized the gamma distribution with two parameters, (r, ω), such that

$$f(1/\sigma^2 | r, \omega) = \omega^r (1/\sigma^2)^{r-1} e^{-\omega(1/\sigma^2)} / \Gamma(r), \quad x > 0$$

where r is the "shape" parameter, ω is the inverse of the "scale" parameter, and $\Gamma(r)$ is the complete gamma function of r (Spiegelhalter et al. 1995). In the base case models it was assumed that $1/\sigma^2 \sim \mathrm{gamma}(0.001, 0.001)$ (Figure 5.1).

In contrast to the hierarchical structure of the estimates of γ_i, the parameters of the prior distributions for α and β were fixed rather than estimated. Very broad normal priors were assumed for α and β and the prior distributions were constrained using upper and lower bounds ($\mathrm{I}_{(\mathrm{lower,\ upper})}$). These bounds are used to both constrain the parameters within very broad range of reasonable estimates and to assist the Hastings-Metropolis algorithm used in the estimation. Another important consideration was the relative reliability of the different data sets. I have assumed that each data set had its own estimation error variance, σ_i^2, which was sampled from a known gamma distribution. Priors for these other parameters are given below:

- $\beta_g \sim \mathrm{N}(0.13, 10000)\ \mathrm{I}_{(0,0.5)}$
- $\alpha_g \sim \mathrm{N}(-5, 10000)\ \mathrm{I}_{(-20,-2.5)}$
- $\sigma_i^2 \sim \mathrm{gamma}(1 \times 10^{-6}, 1 \times 10^{-6})$

The estimated posterior distributions were examined for their sensitivity to alternative

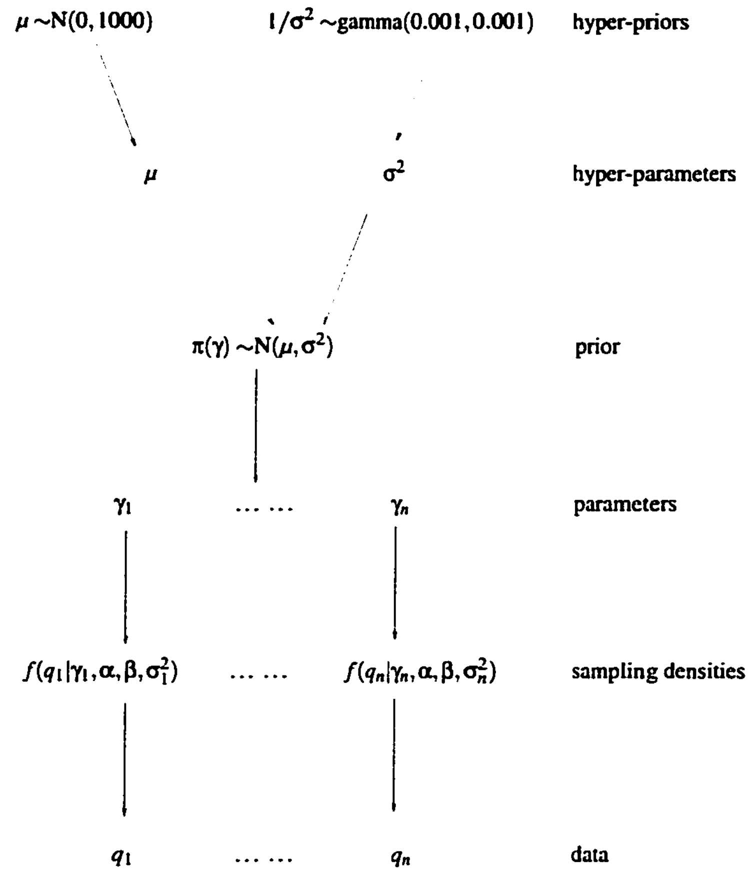


Figure 5.1: The general hierarchical structure of the model for the height of the logistic curve, $\gamma_i, i = 1, \ldots n$. The estimation error for the ith data set is given by σ_i^2 and σ^2 is the variance for the prior distribution for γ_i. (Following Efron (1996)).

priors (including the bounds). As little influence of these constraints and priors was detected, here I only present results for these priors.

More complex models are possible in which all three parameters are assumed to come from their own common distribution. Such a model was not considered here due to the small numbers of data sets in each group.

For all groups except ling, a seasonal categorical variable (summer/fall versus spring/winter) was added to help better model the difference in the mean maximum catchability μ within each group. In this case I estimated a separate μ for each season (μ_s) but assumed that σ^2 was the same for each season. Thus I allow the mean of the distribution from which the γ_is was drawn to vary with season

$$\gamma_{i,s} \sim \mathrm{N}(\mu_s, \sigma^2). \tag{5.8}$$

A Bayesian approach can be used to obtain estimates of individual parameters, e.g. mean or median of the marginal posterior distribution. It is the parameters derived from the marginal distributions that I will compare to the maximum likelihood estimates. While sampling from the joint posterior distribution, it is possible to derive distributions for values which are functions of the estimated parameters. This has the advantage of conserving any correlation structure that may exist between the estimated parameters and is critical to the meta-analysis of a relationship between two quantities as I have done here. I used this approach to generate "predictive distributions" for catchability-at-length. If $q_{k+1}(\cdot)$ represents a new data set of observations of catchability-at-length, independent of the data used in the estimation of $\mathbf{q}(\cdot)$, given the underlying unknown parameters, θ, then the predictive distribution for $q_{k+1}(\cdot)$ is given by

$$p(q_{k+1}(\cdot)|\mathbf{q}(\cdot)) = \int f(q_{k+1}(\cdot)|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{q}(\cdot))d\boldsymbol{\theta}, \tag{5.9}$$

where $\theta = (\sigma_i^2, \ldots, \sigma_k^2, \mu_s, \sigma^2, \alpha, \beta)$ (after Carlin and Louis 1996). This distribution summarizes all information concerning the new observations of catchability-at-length, given the likelihood, priors, and data observed (Carlin and Louis 1996).

The hierarchical Bayesian approach was conducted using BUGS (Bayesian Inference Using Gibbs Sampling) (Gilks et al. 1994). This is freely available software that can use both the Gibbs sampler and the Metropolis-Hastings algorithm to sample the joint posterior distribution for the parameters of interest. We can obtain estimates of the marginal posterior distributions for any function of the estimated parameters, e.g., predicted catchability-at-length as a function of the parameters of the scaled logistic at each step.

Model runs

A hierarchical Bayesian analysis was conducted for each group and predictive intervals for length-specific catchability were determined. Model 1 was run using the NLMINB function in S-PLUS for each data set individually and Model 2 was run for each of the meta-analysis groups described in Table 5.2 using BUGS.

When Markov Chain Monte Carlo methods are used to sample from the joint posterior distribution it is often important to have a burn-in period to allow the algorithm to get from the starting values to sampling the joint distribution. In Chapter 7 I use an alternative implementation of Bayesian integration that does not require user-supplied starting values. Here, a "burn-in" of 5000 iterations was allowed and following this the Metropolis-Hastings algorithm was run for 200 000 iterations with every 100th sample retained. The implementation of Bayesian integration in BUGS requires the user to provide starting values. This reduced the autocorrelation in the chains. We retained 2000 samples of each parameter of interest and the predictive distributions of catchability at length.

5.3 Results

Maximum likelihood estimates were obtained from Model 1 for all data sets individually, and Bayesian estimates were obtained using Model 2 (Figure 5.2). The point estimates from Model 2 (which I will refer to as Bayesian estimates) were the median of the posterior distribution for the parameters of the scaled logistic, γ_i', α, and β. For some data sets the maximum likelihood estimates were not sensible (e.g., cod 4VsW, haddock 4TVsW, and saithe IV and IIIa). Also, for many of the cod data sets, the data suggested very high catchability for the largest one or two sizes (e.g., 2J3KL fall, 3NO surveys, and IV and IIIa). I can not determine if this effect is real or an artefact of the stock assessment model used to generate the catchability estimates.

The advantage of the Bayesian estimates in this situation is that they use information from other similar stocks to obtain improved estimates. In most cases where the maximum likelihood estimates were questionable, the Bayesian estimates appear far more reasonable. However, the Bayesian estimates for saithe and whiting were considerably different from the individual fits (Figure 5.2). As the saithe and whiting data are from the North Sea surveys there may be other factors leading to this difference (e.g., survey gear and substrate type) that are not being accounted for.

One motivation for using a Bayesian framework was the construction of predictive distributions for catchability-at-length. These could be used as auxiliary information for the analysis of a new data set for a group. The predictive distributions summarized by quantiles of the posterior distribution are given in Figure 5.3 and Table 5.3. The uncertainty of these distributions reflects the variability across the data sets within each group. Separate distributions were generated for each season.

In the groups where season was modelled, catchability was higher for summer/fall surveys compared to those conducted in spring and winter. However, for the pelagic gadoid group all but one of the surveys was from summer and fall. For the largest size class modelled, catchability for summer/fall surveys for cod was about 20% higher than estimated for spring/winter surveys while for haddock this difference was 60%.

For flatfish, catchability was also shown to be higher for the summer/fall surveys. Catchability at length was lower for flatfish than any of the other groups. Catchability for cod of a given length was less than that for haddock. Not surprisingly, catchability was higher for the demersal gadoids than those that exhibit more pelagic behaviour. Catchability-at-length for ling was very low for small individuals but increases very rapidly between 60–80 cm. It is not certain what causes this dramatic increase but it could be due to behavioural changes in adult ling.

Only in a very limited number of situations were the Bayesian point estimates of catchability near to, or higher than, 1. Catchability of large haddock (> 60 cm) for summer/fall surveys and catchability of large ling (> 100 cm) were above 1. Maximum catchability of flatfish was below 0.5 for both seasons.

To determine what the predictive distributions actually represent with regard to the data, Figures 5.4 and 5.5 show the 90% percentiles of the predictive distributions

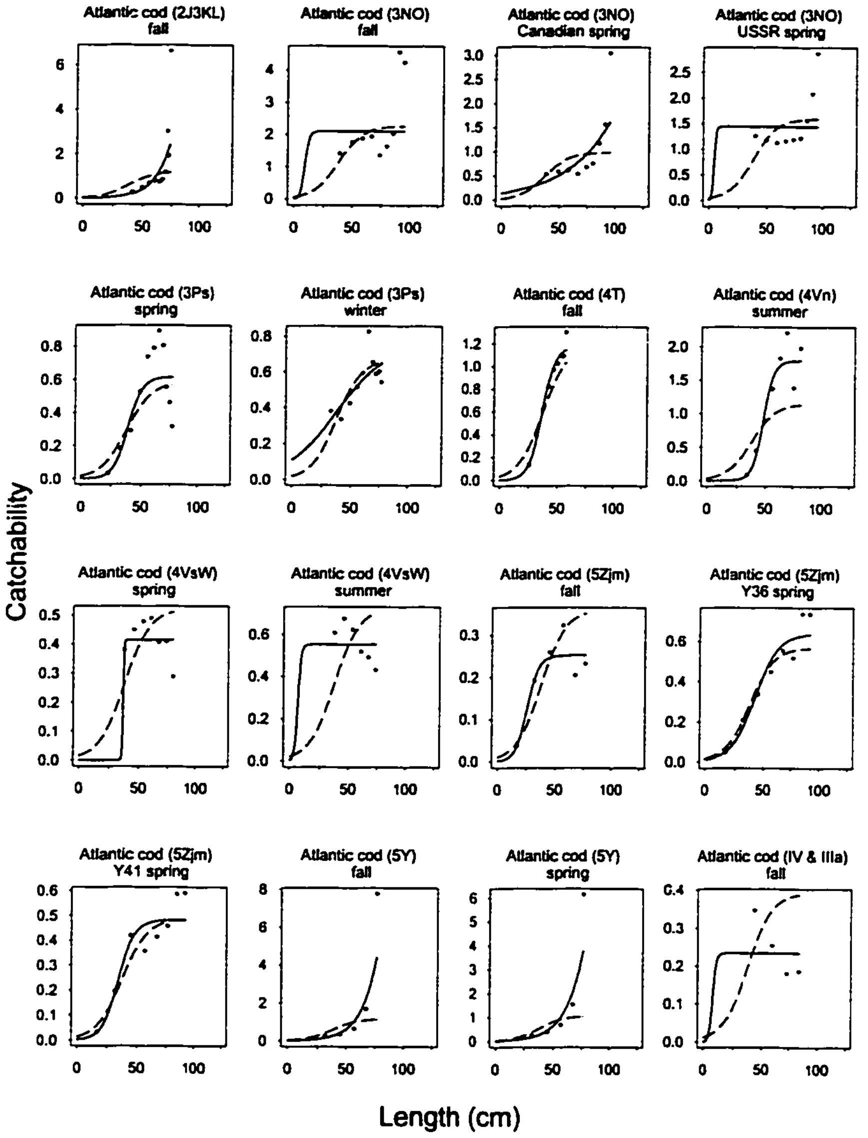


Figure 5.2: Individual maximum likelihood estimates for Model 1 (solid lines), and Bayesian estimates based upon the median of the marginal posterior for each the three parameters of the logistic curve Model 2 (dashed lines) for each of the data sets used. Estimates for cod and haddock are from the demersal gadoid model rather than the individual cod and haddock analyses. (See Table 5.1 for details of each data set; the plots are in the same order.)

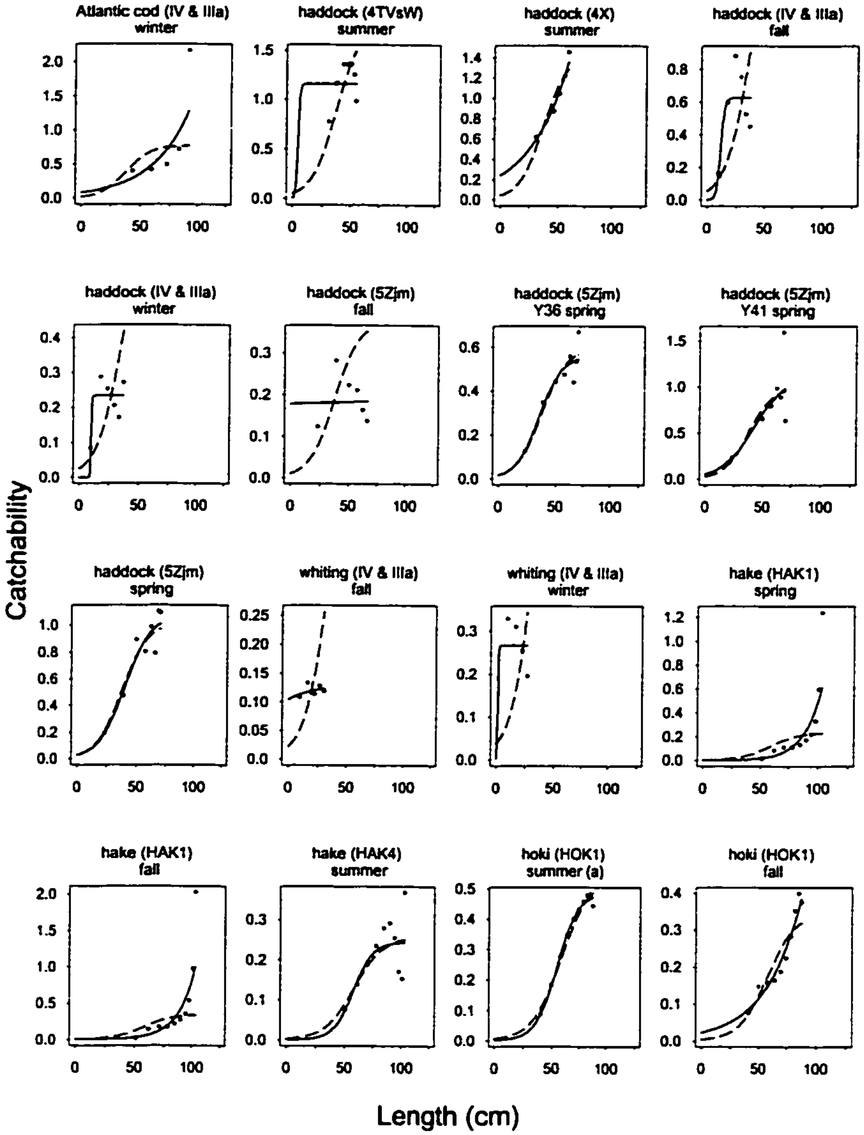


Figure 5.2: Model fits, continued

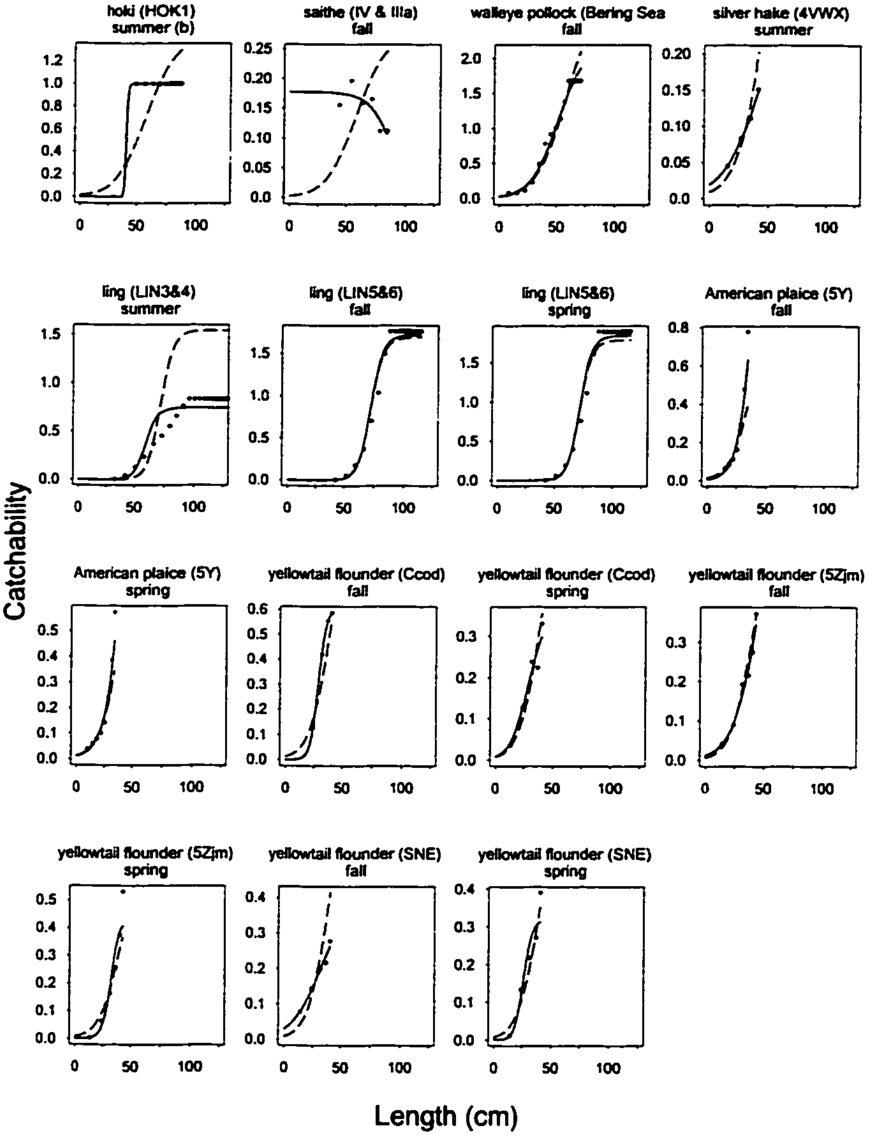


Figure 5.2: Model fits, continued

Table 5.3: Parameters of the scaled logistic curve used to approximate the predictive distributions of catchability at length from Figure 5.3. (Note, while the same α and β was assumed for each season, there are slight differences in the values below as these values are estimated by fitting to the percentiles of the samples from the predictive distributions.).

Atlantic cod

	summer/fall			spring/winter		
	α	β	γ	α	β	γ
mean	−5.06	0.1390	0.949	−5.02	0.1370	0.769
$P_{0.05}$	−5.31	0.1490	0.422	−5.39	0.1470	0.396
$P_{0.25}$	−5.21	0.1440	0.651	−5.17	0.1410	0.562
$P_{0.5}$	−5.14	0.1410	0.870	−5.05	0.1380	0.719
$P_{0.75}$	−5.03	0.1380	1.150	−4.98	0.1360	0.917
$P_{0.95}$	−4.85	0.1310	1.760	−4.87	0.1330	1.310

Haddock

	summer/fall			spring/winter		
	α	β	γ	α	β	γ
mean	−2.77	0.0646	1.590	−2.78	0.0650	1.020
$P_{0.05}$	−2.92	0.0707	0.679	−2.90	0.0688	0.507
$P_{0.25}$	−2.84	0.0694	1.080	−2.86	0.0684	0.736
$P_{0.5}$	−2.80	0.0661	1.500	−2.81	0.0663	0.931
$P_{0.75}$	−2.76	0.0628	1.990	−2.76	0.0645	1.180
$P_{0.95}$	−2.68	0.0588	2.860	−2.70	0.0633	1.740

Demersal gadoids

	summer/fall			spring/winter		
	α	β	γ	α	β	γ
mean	−3.47	0.0914	1.040	−3.47	0.0917	0.843
$P_{0.05}$	−3.65	0.0969	0.522	−3.65	0.0981	0.484
$P_{0.25}$	−3.58	0.0943	0.758	−3.55	0.0950	0.650
$P_{0.5}$	−3.50	0.0925	0.968	−3.52	0.0938	0.797
$P_{0.75}$	−3.43	0.0908	1.240	−3.44	0.0908	0.988
$P_{0.95}$	−3.37	0.0871	1.830	−3.35	0.0868	1.350

Table 5.3: Parameters of the scaled logistic curve, continued

Pelagic gadoids

	summer/fall			spring/winter		
	α	β	γ	α	β	γ
mean	−4.58	0.0785	0.640	−4.57	0.0781	0.733
$P_{0.05}$	−4.70	0.0814	0.276	−4.67	0.0806	0.045
$P_{0.25}$	−4.63	0.0798	0.431	−4.62	0.0795	0.153
$P_{0.5}$	−4.61	0.0789	0.580	−4.58	0.0788	0.335
$P_{0.75}$	−4.57	0.0780	0.769	−4.59	0.0778	0.707
$P_{0.95}$	−4.53	0.0775	1.190	−4.60	0.0798	2.300

Flatfish

	summer/fall			spring/winter		
	α	β	γ	α	β	γ
mean	−4.43	0.109	0.986	−4.39	0.1090	0.856
$P_{0.05}$	−4.43	0.1240	0.398	−4.36	0.1190	0.275
$P_{0.25}$	−4.35	0.1160	0.609	−4.4	0.1170	0.485
$P_{0.5}$	−4.35	0.1110	0.831	−4.4	0.1120	0.716
$P_{0.75}$	−4.4	0.1060	1.190	−4.39	0.1070	1.060
$P_{0.95}$	−4.65	0.1020	2.220	−4.44	0.1010	2.010

Ling

	α	β	γ
mean	−13.70	0.1910	1.840
$P_{0.05}$	−13.80	0.1920	0.996
$P_{0.25}$	−13.90	0.1930	1.360
$P_{0.5}$	−13.90	0.1930	1.660
$P_{0.75}$	−13.70	0.1900	2.010
$P_{0.95}$	−13.60	0.1890	2.720

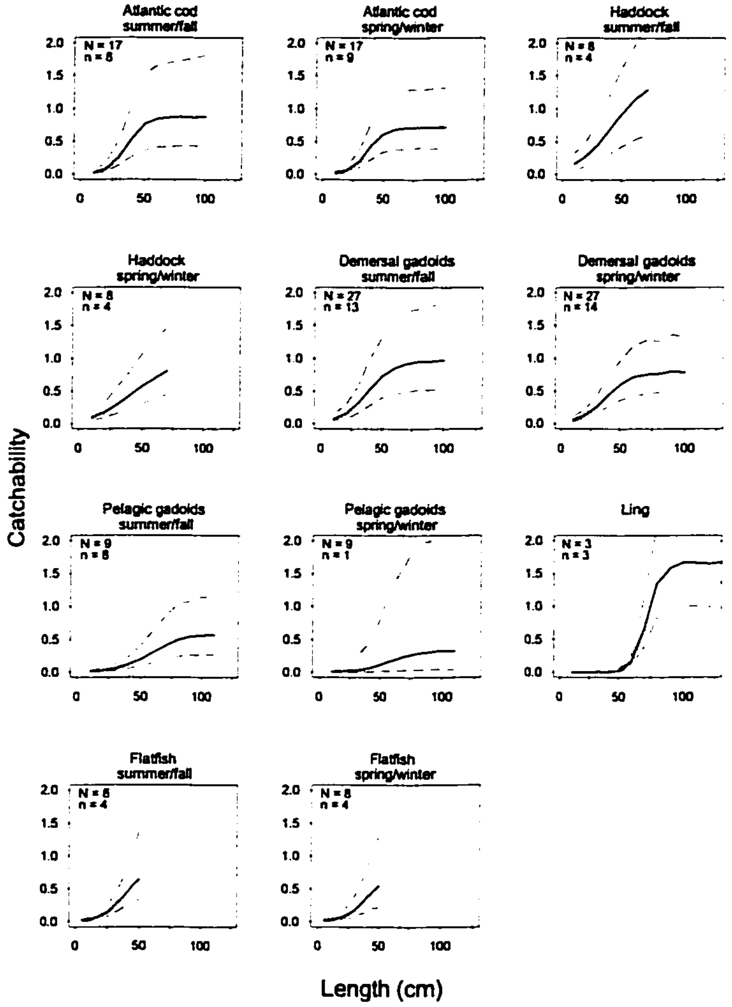


Figure 5.3: Predicted length-specific catchability for each meta-analytical group. Where a seasonal variable was included, estimates for both seasons are given. The solid line is the median of the posterior, the short dashed lines are the 25th and 75th percentiles of the posterior predictive distribution, and the long dashed lines are the 5th and 95th percentiles. The number of data sets used in each meta-analytic group is N and n is the number of data sets for the particular season. The range of values of length correspond to that observed for the group.

with the individual fits and data overlaid.

5.3.1 Robustness

One criticism of hierarchical Bayesian methods is the reliance on priors for the parameters describing the distribution of the random effects (hyper-priors) (Carlin and Louis 1996). For sensitivity, I used $1/\sigma^2 \sim \text{gamma}(1 \times 10^{-6}, 1 \times 10^{-6})$. This hyper-prior gives more weight to higher precisions and thus lower random effect variances. This sensitivity analysis was run for both demersal gadids and flatfish. The results were almost identical indicating the models were not sensitive to the choice of prior. While this is reassuring here, this result is not always the case when only a small number of data sets are available.

The upper and lower bounds for α and β were varied to test for sensitivity to the flat but bounded priors that were used. The results were insensitive to these changes.

5.3.2 Fish community reconstruction

The key goal of this work was to provide estimates of catchability that could be applied to a range of species to allow reconstruction of the fish community on the eastern Scotian Shelf off Nova Scotia, Canada. This work was part of the Comparative Dynamics of Exploited Ecosystems in the Northwest Atlantic (CDEENA) project. Since the 1960s there have been dramatic changes, both physical and biological, in the ocean environment from Labrador to George's Bank. These changes have been particularly pronounced on the eastern Scotian Shelf (Harley et al. 2001).

Reconstruction of fish community structure allows investigations of how changes occurred over time. Of particular interest are potential interactions among species, e.g., the increase of one species group following the decrease of another. As we recognize that differential catchability exists, the first insight obtained from the estimates of catchability is what the fish community structure looks like at any given time. This allows us to determine the major (in terms of biomass) components of the system. Next we can see how changes occur over time – whether biomass is conserved? and how much "total" biomass varies from year to year.

For the eastern Scotian Shelf, much of this work was undertaken by using estimates

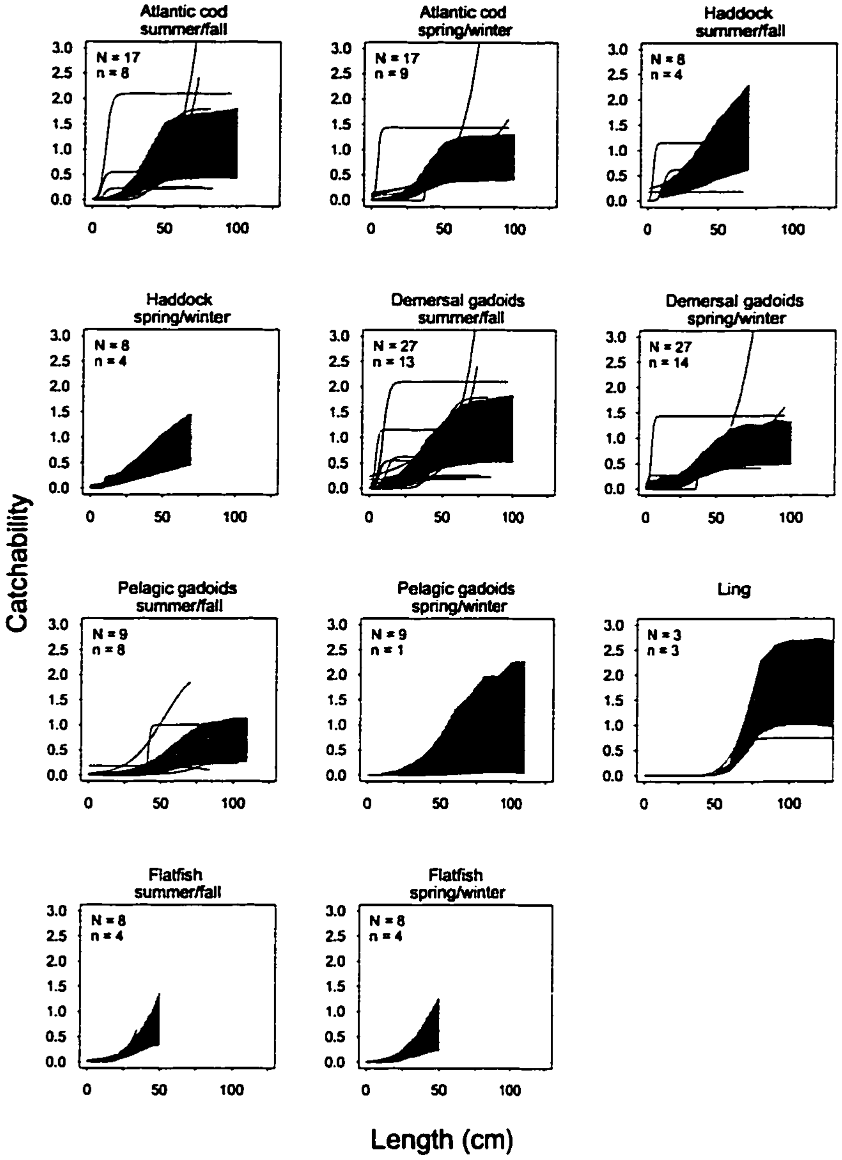


Figure 5.4: Predicted length-specific catchability for each meta-analytical group with the individual fits (maximum likelihood) for each data set overlaid. The shaded region indicates the 90% percentiles of the predictive distributions.

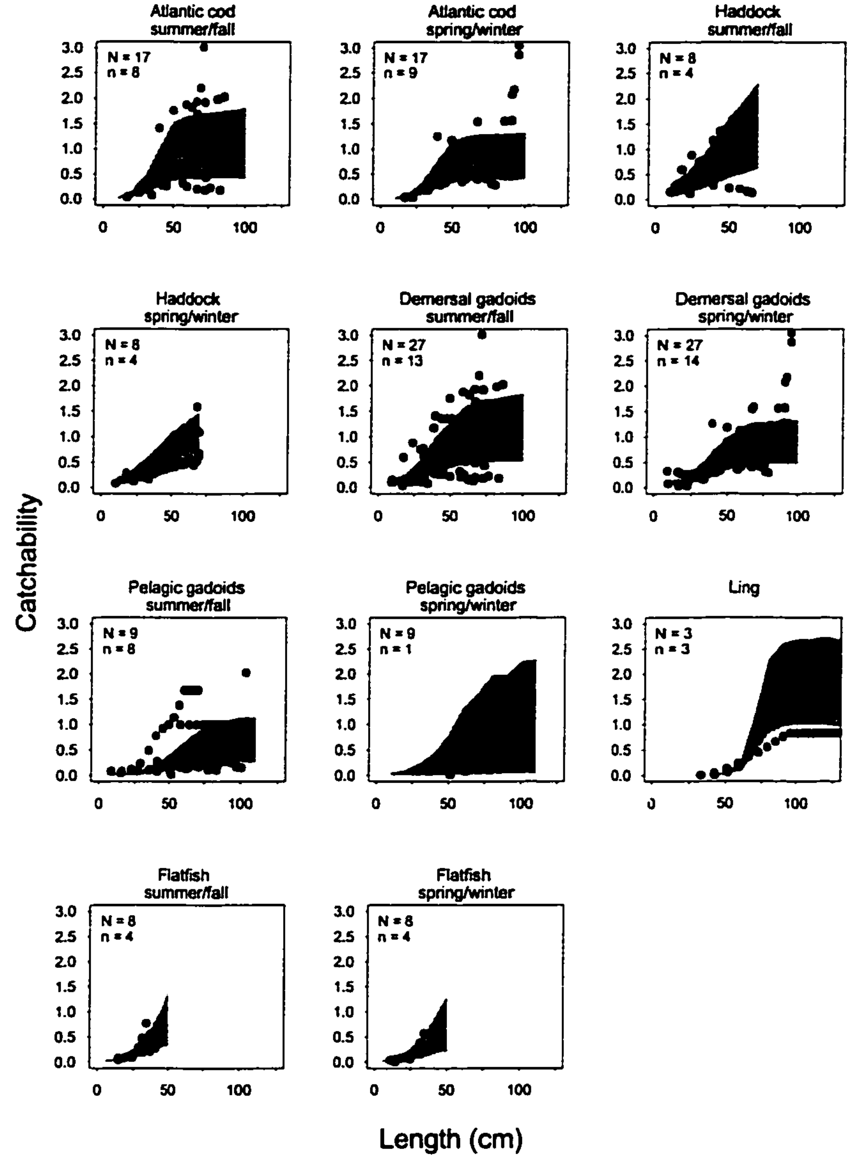


Figure 5.5: Predicted length-specific catchability for each meta-analytical group with the original data overlaid. The shaded region indicates the 90% percentiles of the predictive distributions.

of catchability derived in this chapter (O'Boyle 2000). In the following sections I provide some details of the types of insights that can be made from reconstructing fish community structure when length-specific estimates of catchability are available.

Cod

Once the most abundant commercial species on the Scotian Shelf, Atlantic cod are now estimated to be at very low levels. Using survey estimated numbers at length and a length-weight relationship I can attempt to reconstruct cod biomass for different size classes (Figure 5.6). Here I will use three classes, below 30cm, 30–50cm, and above 50cm.

The relative biomass trajectory for total cod biomass is similar under all scenarios although the decline since 1983 is less when length-specific catchability is applied (Figure 5.6). Total biomass is highest for the single catchability scalar. The biomass of the middle group is greatest under all scenarios. Interestingly, biomass in the most recent years is similar for the length-specific and single scalar scenarios though the historical reconstructions are quite different.

Multiple species

As mentioned earlier in this chapter, the work here was motivated by the need to obtain absolute biomass estimates for different species groups. These estimates could then be used to consider the community structure in an ecosystem framework using such models as ECOPATH with ECOSIM (Walters et al. 1997). In Figure 5.7 I show biomass for different functional groups used in an ECOPATH model (A. Bundy, Department of Fisheries and Oceans, Canada, Dartmouth, Nova Scotia, pers. comm.). I have pooled a number of groups together to provide a simple plot.

Total biomass of these species is much higher (300%) when catchability is taken into consideration. While the biomass of the large Gadiforms is only slightly higher when catchability is considered, the contribution of flatfish and small Gadiforms is much higher. Our interpretation of a large gadoid dominated community is changed when catchability is considered. During the early 1980s biomass for all groups except flatfish increased. The increase in large Gadiforms lagged behind the increase in small Gadiforms (recruit followed by growth). Interestingly total fish community biomass

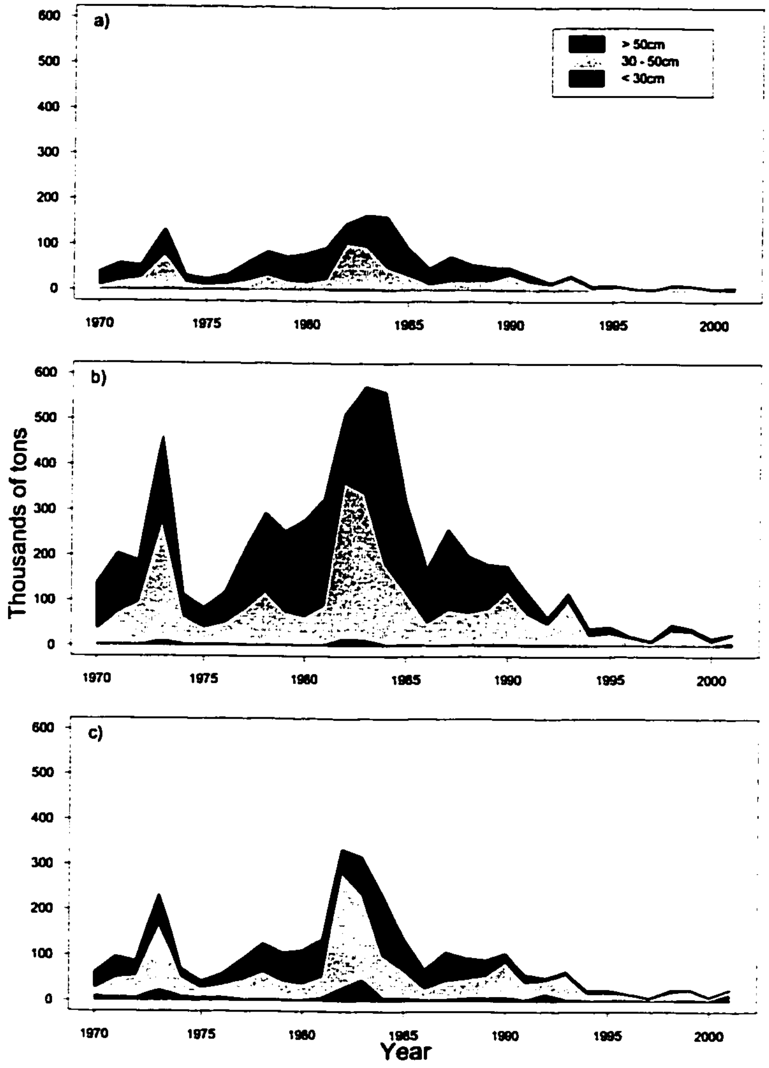


Figure 5.6: Cumulative biomass of different size classes of cod for the eastern Scotian Shelf from 1970 to 2001: (a) assuming catchability is unity for all sizes; (b) assuming a single bulk catchability estimate of 0.29 for all sizes (from Edwards 1968); and (c) assuming length-specific catchability (median of predictive distributions) as estimated in this study. Length frequency data and annual length-weight relationships were provided by P. Fanning (Department of Fisheries and Oceans, Canada, Dartmouth, Nova Scotia).

(for these groups at least) is similar for 1970 and 1998, though there are slightly more large gadoid and other demersal biomass in 1998. Over this time period there has not been any dramatic changes in fish community structure though total biomass appears to have been higher during the 1980s compared to the 1970s or 1990s. Unfortunately it is very difficult to determine the biomass of small pelagic species to see if there were any changes for this group over this period (e.g., to see if it was lower during the 1980s). As small pelagic species are very poorly caught by bottom trawls, both catchability and reconstructed biomass trajectories will be very uncertain (potentially as much as an order of magnitude Harley et al. (2001)).

5.4 Discussion

Estimation of the catchability of research trawl surveys is essential to both the management of single species, and more importantly, the reconstruction of fish communities and interpretation of their dynamics. Here I have constructed length-specific catchability estimates for a number of species groups that will allow for increased understanding of fish community dynamics. I have used meta-analysis to combine estimates of the relationship between two quantities in a way not previously done.

First, I discuss the use of hierarchical Bayesian analysis over more traditional fixed effects analysis. Second, I discuss the potential uses for the catchability estimates and the usefulness of this process as a potential diagnostic for stock assessments. Third, I compare these results to those of previous researchers, and finally I discuss the robustness of the results and discuss future extensions of this analysis.

Hierarchical modelling approach

In this analysis I treat each survey/species combination as a random effect in a hierarchical model, rather than a more traditional fixed effects model. This provides a more reliable estimator under a wide range of conditions (Robinson 1991). Kiefer and Wolfowitz (1956) noted that in estimation situations where the number of parameters increases to infinity, maximum likelihood parameter estimates are not consistent, but by treating parameters as coming from a distribution—that is, as random effects—consistency could be obtained. Thus, if we assume that the surveys have something

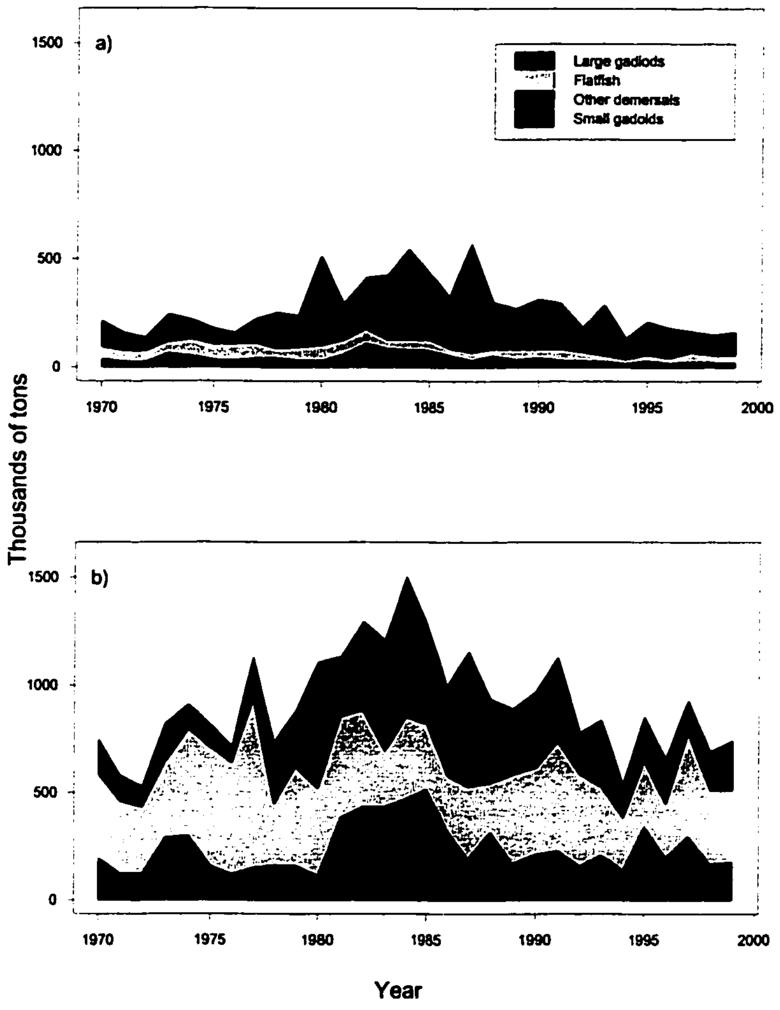


Figure 5.7: Cumulative biomass of various functional groups (some combined) as used in the ECOPATH model for the eastern Scotian Shelf (A. Bundy, Department of Fisheries and Oceans, Canada, Dartmouth, Nova Scotia, pers. comm.) from 1970 to 2001; (a) without correcting for catchability; and (b) corrected for catchability scalars including those from this study and some presented by Harley et. al (2001).

in common, we can obtain better estimates than if we assume that the surveys are unrelated. This can clearly be seen by the comparison of the individual and Bayesian estimates of the catchability curves.

Because the same survey gear was not used for all surveys, this will lead to extra variability. This variability enters the model through the random effect variance (σ^2) which accounts for for all variability that cannot be accounted for by observation error alone. The use of separate means for the different seasons would act to reduce the the random effect variance. The random effect approach does have an additional advantage over fixed effects for each survey gear as we can provide some estimates of catchability for a new survey gear type (for a given species) in the form of the predictive distributions.

Potential uses for estimates of length-specific catchability

Catchability estimates are generally only available for commercially important species that are formally assessed. Here I estimate catchability for a number of species groups. As an application of the results of this study a number of species from the Scotian Shelf were assigned to the different groups used in the meta-analysis. In this instance, length-specific catchability estimates have a considerable advantage over bulk estimates. For example, rather than assume that the catchability of a red hake (*Urophycis chuss*) (size $\approx$ 50 cm) is the same as cod (size $\approx$ 100 cm), we can assume that it has the same catchability as a 50 cm Atlantic cod. Details of the assignment of catchability estimates to 78 fish species is given in Harley et al. (2001).

Hilborn and Liermann (1998) discuss the importance of the synthesis of existing information for the improved assessment of exploited fish populations. The results from a meta-analysis such as the one undertaken here could be incorporated into stock assessments in the form of auxiliary information. This could be done through Bayesian prior distributions (Punt and Hilborn 1997) or as a penalty function. For a stock assessment that estimates both survey selectivity and survey catchability (e.g., a common west coast U.S. assessment) it might be appropriate to set the maximum selectivity at 1 and then construct a prior distribution for catchability based on μ_s. This approach was used in the assessment of the Pacific ocean perch off the coast of Washington and Oregon (Ianelli et al. 2000).

It is not clear how the results of a meta-analysis could be incorporated into a traditional VPA framework other than diagnostic check for the model estimates. A less powerful way to use these results would be by informally comparing stock assessment results with those from a meta-analysis. Another approach would be to conduct fully Bayesian assessments within a hierarchical framework. This would involve the simultaneous modelling of all the populations and would involve some parameters (e.g., catchability or spawner-recruitment parameters) being estimated in a hierarchical manner. At present, this represents a daunting prospect in terms of both programming complexities and computational demands. Millar and Methot (2002) made one of the first attempts to perform multiple assessments simultaneously treating catchability in a hierarchical manner. Their analysis was only partly integrated as it relied on single-species stock assessment models to provide relative recruitment indices and selectivity that were treated as known without error in their analysis. Nevertheless, this does represent an important step in a direction that I believe will lead to improved stock assessments.

In many assessments survey indices are modelled as a mean catch rate rather than as area-swept abundance. In these circumstances, model estimates of "relative" catchability have no simple interpretation. Calculation of absolute catchability from catch rates is done using Equation 5.1 incorporating estimates of the area swept by a tow and the total survey area. Upon doing this calculation, I found that, for some data sets, estimates of catchability were greater than 4 (these data were excluded here). Very high estimates of catchability (e.g., greater than 2 or 3) imply that the survey catches considerably more than fish than would be predicted to be in the path of the tow (Equations 5.2 and 5.3). Conversely, low estimates of catchability suggest that there is considerably more fish out there than are being caught by the survey. This could suggest that abundance is being overestimated by the model. I recommend that area-swept estimates of catchability be derived as a part of any stock assessment as these values can be more easily interpreted and used as a diagnostic. In cases where catchability is either very high, or low, this may suggest a problem with the assessment.

Comparisons with previous work

As this is the first study to estimate length-specific catchability estimates for a number of species, there are no other estimates that I can compare these results with directly. However, it is possible to make some comparisons with the bulk catchability estimates of Edwards (1968) and Sparholt (1990). Sparholt provided estimates for both the English Groundfish survey (EGFS) and the International Young Fish survey but only the former will be discussed here. Both Edwards and Sparholt estimated haddock catchability to be higher than that of cod but both estimates are lower than that found here for larger individuals. Edwards' estimates for a given species were generally lower than that found by Sparholt except for flatfish where catchability for the EGFS was estimated at 0.096 compared to the range of 0.3–0.57 suggested for flatfish species by Edwards. My estimates for flatfish are more consistent to those of Edwards. Overall, the qualitative conclusions regarding the relative catchability of various species are similar to those of previous researchers.

Though cod and haddock share a number of morphological similarities it appears that the catchability of haddock is higher than that of cod. This implies behavioural differences in vertical availability to the trawl and/or reaction to the trawl gear. Main and Sangster (1981) found that cod and flatfish generally stayed low as the trawl passes while haddock and whiting tend to swim upwards. This suggests that cod are better at escaping under the ground gear of a the bottom trawl gear used in the surveys analyzed, than haddock are at escaping over the top.

Robustness and further work

A number of different sources of uncertainty have been included into my analysis but there are additional sources of uncertainty that warrant consideration. First, I have ignored variability in length at age by using a single von Bertalanffy growth function for each population. Second, I only allowed the height of the logistic curve to vary within each meta-analytic group. In doing so I ignored any variability in the shape. However, this was necessary because of the small number of data sets that were in each group. Third, I assumed that any seasonal influences on catchability were the same for summer/fall and spring/winter surveys. Once again small sample sizes imposed this restriction. Fourth, I assumed a functional form for the catchability curve that did not allow decreasing catchability at larger sizes. Finally, that diel variability in catchability may result in differences between surveys conducted around the clock and those conducted during daylight hours only. These final two sources of uncertainty are discussed in greater detail below.

Thompson (1994) used a four-parameter logistic curve that allowed for a descending right-hand limb which in some cases may provide a better fit to the data than the three-parameter model assumed here. While there may be behavioural factors responsible for a descending right-hand limb in some cases, this descending limb is confounded (in estimation) with increased mortality at older ages (Thompson 1994). I believe that the logistic curve used here provides a good starting point but that curves allowing a descending right-hand limb should be considered in cases where observational studies support decreased catchability at larger sizes. It is possible to extend future analysis to include other models such as the four-parameter logistic curve used by Thompson (1994) but estimates from more complex models may be less reliable as many data sets had only 5–8 length/catchability observations.

Most research vessel surveys outside of the Northwest Atlantic take place only during daylight hours. While this appears to matter relatively little for some species it makes a great deal of difference for others, e.g., sandlance (Casey and Myers 1998). As Casey and Myers (1998) found that the diel correction factor for cod and haddock was not significantly different from 1, I have combined both day and day/night survey estimates for cod, haddock, and the demersal gadoids. Although Casey and Myers (1998) provide diel corrections for many species on the Scotian shelf, there are several important species, e.g. silver hake and pollock, for which diel corrections were not well estimated. When applying the meta-analytical estimates derived here to other species, it will be important to consider diel variability.

The seasonable variable used to model differences in the maximum catchability within a group suggested that trawl survey catchability is higher in summer/fall surveys than spring/winter surveys. Due to the small number of data sets available, it is possible that season may be confounded with other factors or that the results are simply due to one or two extreme data sets. However, it is interesting that the same directional effect was observed for all four of the meta-analytic groups (not including the demersal gadoids as the haddock and cod data was considered separately). It is recommended that seasonal differences be investigated in more detail to attempt to confirm this finding and determine if this is also observed in other groups (e.g., herring and mackerel).

I considered Bayesian hierarchical models, which are only one of a family of hierarchical modelling techniques that could be used in this problem. The nonlinear mixed-effect model represents a frequentist approach that uses maximum likelihood type methods for the estimation of the parameters of the random effects distribution (Lindstrom and Bates 1990). Neither of these approaches is without its weaknesses. While Bayesian methods often rely heavily on prior specifications and require convergence of the Metropolis Hastings algorithm to obtain accurate estimates of the posterior distributions, it is possible to estimate the random effect in mixed-effects models to be 0 (Searle et al. 1992).

Several extensions to this approach are possible, which would probably give improved results. It could be feasible to model morphology and behaviour simultaneously in a meta-analysis, i.e., include species like ling with the demersal gadoids but include a parameter that described the effects of morphology on catchability. It may also be possible to estimate gear effects and bottom type in more detail; however, I was primarily limited by data.

Chapter 6

Estimation of residence time and spawning biomass of hoki

6.1 Introduction

Stage-structured data and migration dynamics are often important for modelling population dynamics. With spawning migrations, the analysis of stage-structured data is often crucial for estimating population abundance and dynamics. The key issue to be addressed in this chapter is how best to estimate population size when not all individuals are present in an area at the same time and when data are available for multiple seasons or years.

I construct complex models to describe the spawning dynamics of New Zealand hoki (*Macruronus novaezelandiae*) and use these to estimate residence time and spawning biomass. The incorporation of the meta-analytic approach is in how the parameters describing spawning dynamics are treated. I will demonstrate how it is now possible to integrate the treatment of parameters as random effects into a complex nonlinear dynamic model. The models developed here take advantage of data collected over a number of years. This is critical as I will demonstrate that data for a single year are generally insufficient to reliably estimate the important parameters of these models.

This general type of problem has been reported for estimating the abundance of other species. For example, seals often give birth over an extended period of time, so that not all pups are present in a survey at the same time (Myers and Bowen 1989; Myers et al. 1997). It is possible to use stage composition of seal pups to estimate the distribution of births over time so that aerial surveys or mark-recapture data can be calibrated to estimate total abundance. Similar problems occur for stage composition of insects (Vogt and Morton 1991), and a variety of other taxa (Manly 1990). Su et al. (2001) developed models of spawning pink salmon (*Oncorhynchus gorbuscha*) returns to Kadashan Creek in southeast Alaska where a hierarchical framework was used to share information across years. So while the specific example is directed towards acoustic surveys of a spawning concentration of fish, the approach and problem is much more general.

Although stage composition data are very common in ecology, they are not always used effectively. Often we may have data for a number of time periods. Here I develop methods that allow these parameters that describe composition data to vary with time or among populations. However, they are still constrained in a biologically reasonable manner. I will compare the traditional approach of independent analysis of data for each time period to hierarchical model structures that allow improved estimation of population abundance and dynamics parameters.

I will first construct a stage-based model that predicts the proportion of hoki that are on the spawning grounds at any one time. This allows for the acoustic surveys to be calibrated. Often the only feasible time to carry out an acoustic survey is during spawning, when the fish are concentrated into a relatively small area. This technique is used for many of the worlds largest commercial fisheries, e.g., walleye pollock in Alaska, and the large herring and cod stocks off of Norway. Often a critical assumption of such surveys is that all spawning fish are present at the time of the survey; an assumption that is often neither met nor tested. For hoki there is good evidence that not all hoki are on the spawning ground at any one time during the spawning season (Coombs and Cordue 1995; Bulman et al. 1999).

My development differs from previous studies in two ways. First, in this example it has not been possible to obtain independent estimates of stage duration for the hoki, as it possible for other species. Tagging of hoki to estimate stage duration is not feasible. Second, I use hierarchical modelling to obtain estimates for the stage transitions by combining data over many years. This allows for much more stable estimates, and reduces some of the well known problems (Wood 1997) with fitting stage-structured population models to data.

In the following sections I provide background on the hoki population considered here. Next I review the spawning behaviour of hoki and other related species. It is this biological information that provides the basis for the complex dynamic models that are then developed. I then describe how the analysis can be incorporated into a meta-analytical framework through a range of hierarchical models.

Harley (2002) described preliminary results for the estimation of residence time and spawning biomass for the two major hoki spawning grounds in New Zealand. Here I will only present results for one of these, the Cook Strait.

6.1.1 New Zealand hoki

Hoki form New Zealand's largest commercial fishery with landings of about 250kt per year since 1996–97 (Annala et al. 2000). The stock structure and population dynamics of hoki are complex and were first described by Livingston (1990b). Most hoki are found around the South Island with two distinct substocks defined by their spawning grounds (Figure 6.1). The larger western stock spawns of the west coast of the South Island, and the smaller eastern stock spawns in Cook Strait. The nursery ground for both stocks, and the feeding ground for the eastern stock, is the Chatham Rise. The current stock assessment for hoki uses a complex dynamic model that incorporates many of the ecologically important processes, e.g., a two-stock model with separate feeding and spawning grounds (Ballara et al. 2002). A variety of data are fitted by the model, such as catch-per-unit-effort indices, research trawl survey numbers-at-age, commercial catch-at-age, and acoustic biomass estimates for the spawning grounds.

If the abundance indices from the acoustic surveys could be used as absolute abundance indices, it could significantly reduce uncertainty in the current assessment. The two greatest problems that need to be overcome to enable the use of acoustic indices as estimates of absolute abundance for spawning hoki are: (1) accurate determination of target strength for hoki; and (2) estimation of residence time for hoki on the spawning ground (Coombs and Cordue 1995). It is the second of these problems that I will address here using a meta-analytic approach.

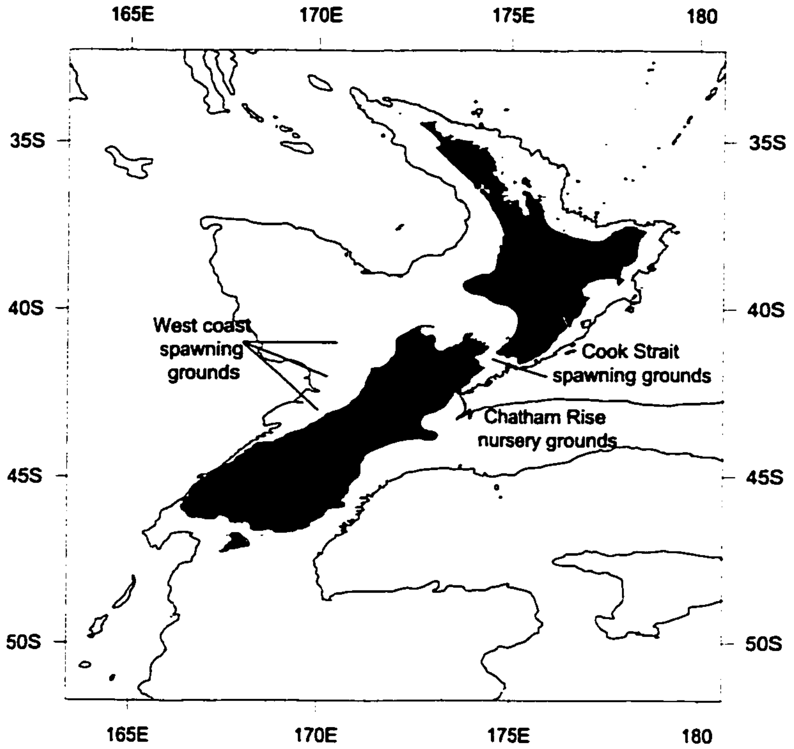


Figure 6.1: Some areas of critical importance to New Zealand hoki. The solid line represents the 1000m depth contour. (Adapted from Schofield and Livingston 1998.)

6.1.2 Spawning of hoki and related species

Hoki are a member of the order Gadiformes which includes hake, the grenadiers (family Macrouridae), and Atlantic cod (*Gadus morhua*). The family Merlucciidae includes hoki and the true hakes (e.g., *Merluccius australis*).

Many Gadiformes are summer or spring spawners (Hutchings and Myers 1993; Alkseyev 1995), whereas hoki are winter spawners (Livingston 1990a). For Atlantic cod, the onset of spawning appears to be associated with changes in water temperature, but the direction of the correlation is not consistent across areas (Hutchings and Myers 1993; Marteinsdottir and Bjornsson 1999). Gunn et al. (1989) suggested that water column conditions, in particular water temperature in autumn and early winter, could be important in determining the onset of spawning in hoki (also known as blue grenadier) off southern Australia. They also suggested that lunar conditions might influence spawning behaviour.

Hoki appear to be determinant spawners, producing all eggs for a spawning season at once (Livingston et al. 1997; Schofield and Livingston 1998). However, it appears that the eggs are released in a number of batches, with each batch of eggs being hydrated shortly before release. Thus, individual hoki must undergo the final steps of maturation several times during a spawning season (M Livingston, NIWA, Greta Point, Wellington, pers. comm.).

Hoki undertake long migrations to reach the spawning grounds (Gunn et al. 1989; Livingston 1990b) as do other Gadiformes, e.g., northeast Arctic cod (T. Marshall, Institute for Marine Research, Norway, pers. comm.). While the fisheries for hoki are based on the spawning grounds, it is unusual to find many spent fish in commercial landings (Langley 1993). This could occur if spent fish no longer aggregate, or leave the ground immediately to resume feeding. The consequences of the absence of spent fish from the landings and the multi-batch spawning discussed above are important to the models developed here.

There has been more interest in determining spawning duration (Hutchings and Myers 1993; Marteinsdottir and Bjornsson 1999) rather than the length of time a fish reside on the spawning grounds. Residence time should be longer than spawning duration if it takes spent fish some period of time to leave the spawning grounds. The general approach to the estimation of spawning duration uses probit regression, where the proportion of fish that are maturing, spawning, or spent, are modelled over the spawning season (Hutchings and Myers 1994a). For hoki there is the complication that they are only observed on the spawning ground in commercial catches rather than through a widespread survey encompassing all of the population (e.g., Hutchings and Myers 1993). Studies of captive fish have also been used to estimate spawning duration in cod (Kjesbu et al. 1996), but no such studies have been attempted for hoki.

The onset and duration of spawning has been shown to vary with age and size (Langley 1993; Hutchings and Myers 1994a; Kjesbu et al. 1996; Marteinsdottir and Bjornsson 1999). Larger and older females tend to spawn earlier in the season and for longer than smaller individuals. Most estimates of spawning duration have been for females, but there is some evidence that spawning duration could be longer for males than females (Hutchings and Myers 1994a). There is evidence for both Atlantic cod (Morgan and Trippel 1996) and hoki (Livingston et al. 2000) that males arrive at the spawning grounds before females.

6.2 Methods

The approach taken here is similar to that used in Chapter 5. I will first describe the data that are available to help estimate residence time, next I will describe the dynamic model which is required to incorporate the important biological processes, and finally I describe how it is possible to integrate the dynamic model into a meta-analytical framework to improve estimates of the important biological parameters.

6.2.1 Data

Data used for this analysis consist mainly of biological data collected through port sampling. Random samples (within a landing) of fish were taken, aged, and measured, with females assigned to one of five gonad-stage categories.

1. Immature / resting

2. Maturing

3. Mature / partially spent
4. Running ripe
5. Spent

For modelling, stages 2 and 3 were combined and stage 1 fish excluded as there were very few of them and it was not possible to determine how long immature or resting fish are likely to stay on the grounds. Length and age-structured data were available but not used here as the addition of age-structure to the models increased the number of parameters to be estimated without providing sufficient data to estimate them (Harley 2002). An estimate of the proportion of females in the catch was also available as all fish that were measured were also sexed. All data were aggregated into two week strata. Data were in the form of proportions at stage and sex rather than absolute numbers so there were no estimates of the true sample sizes. As fish are not from a truly random sample, rather sampling of fish within trips, the total number of fish sampled may over-estimate the effective sample size. The number of trips sampled in a two-week period was used as an index of relative sample size and the sensitivity to scaling these samples sizes (4–16 trips) by a constant was examined. Reported commercial landings were also available for these periods.

The final source of data was abundance estimates from the acoustic surveys of the spawning grounds. During the acoustic survey a number of snapshots (instantaneous biomass estimates) are obtained for the spawning ground over the season. Each snapshot was usually completed within 24 hours (Cordue et al. 1999). If target strength is determined correctly, each snapshot represents an estimate of the total abundance of hoki on the spawning ground at that time. Each snapshot estimate had a variance calculated using stratified random sampling estimators. The c.v.s for the acoustic biomass estimates ranged from 20% to 80%. These c.v.s likely over-estimate the true precision of the acoustic surveys for two main reasons. First, there is uncertainty in target strength as target strength changes with the size of the individuals and their orientation in the water column. Second, uncertainty in species composition of mixed species aggregations as different species can have very different target strength.

Data were available for seven years (1993–1999). The gonad-stage, sex ratio, and catch data were available for 5–6 two week time strata for each year while the acoustic snapshots were modelled using the day on which they were obtained.

6.2.2 Dynamic model

The spawning dynamics of hoki are complex due to a number of factors. The important assumptions that I have modelled here are: (1) all fish do not have to be on the grounds at once; (2) fish arrive and depart following a cumulative normal distribution; and (3) male residence is a function of the female dynamics. As gonad stage data are available only for females, it is not possible to estimate male residence time directly. Female arrival and residence time are estimated parameters, while male residence is derived from the female dynamics.

It is important to distinguish the two groups that are modelled through the season. The first is the *total spawning population* that will migrate to the grounds in a given year. This is reduced throughout the year by fishing mortality (no natural mortality was assumed during the spawning season). The second is the numbers of *fish on the spawning grounds* at a given day of the season. This is a function of the immigration and emigration of fish with respect to the grounds in addition to fishing mortality.

Considering a model for a single year (I will suppress the notation for year here), six parameters are estimated: the total number of fish in the spawning population (N^*), mean arrival day on the grounds of females (α), variability in arrival day (σ^2), and the three components of female residence time (θ). The three parameters for residence time are the duration of the mature (β) and running ripe stages (γ), and time that spent fish spend on the grounds before leaving (ζ), i.e., $\theta = \beta + \gamma + \zeta$. Harley (2002) described a non-stratified model which did not divide females into the different gonad stages.

The number of males (M^*) and females (F^*) in the total spawning population is

$$F^* = N^* \pi \quad \text{and} \quad M^* = N^*(1 - \pi),$$

where π is the proportion of females in the total spawning population, i.e., $F^*/(M^* + F^*)$. Using the estimated numbers and the approximate mean weight of an individual, the total spawning biomass (B^*) is given by,

$$B^* = F^* w_f + M^* w_m,$$

where w_f and w_m are the mean weight of females (1.36 kg) and male fish (1.14 kg) respectively based upon ages 4 and above (Harley (2002), Appendix 2).

The proportion of the total male spawning population that is on the grounds is assumed to be constant, (ϕ). The proportion of the total spawning female population that is on the grounds, (g_t), varies over the season (day t) as a function of the cumulative arrival proportion (a_t) and cumulative departure proportion (d_t) for females to the grounds (Figure 6.2)

$$a_t = \Phi\left(\frac{t-\alpha}{\sigma^2}\right), \quad d_t = \Phi\left(\frac{t-(\alpha+\theta)}{\sigma^2}\right), \quad \text{and} \quad g_t = a_t - d_t,$$

where the mean day of departure is $\alpha + \theta$, σ^2 is the variance for the arrival curve, and $\Phi()$ is the cumulative normal distribution function. The residence time of a female fish on the ground (θ), is assumed to be the same for all females. When female fish arrive on the grounds, they are assumed to be in the mature stage. The cumulative proportion of females that have reached the running ripe stage (λ_t), and the spent stage (φ_t) is given by

$$\lambda_t = \Phi\left(\frac{t-(\alpha+\beta)}{\sigma^2}\right) \quad \text{and} \quad \varphi_t = \Phi\left(\frac{t-(\alpha+\beta+\gamma)}{\sigma^2}\right).$$

Using the arrival and departure curves we have the proportion of the total female spawning population that is on the grounds (g_t), we now use the other curves to distribute the females across the different stages. The proportion of females that are mature (m_t), running ripe (r_t), and spent (s_t) is simply

$$m_t = \lambda_t - a_t, \quad r_t = \varphi_t - \lambda_t, \text{ and } \quad s_t = d_t - \varphi_t.$$

One major flaw with this model is that it ignores the cycling of hoki through the mature and running ripe gonad-stages during the season. An ad-hoc fix to this problem is to assume that some proportion of fish that have reached the running ripe

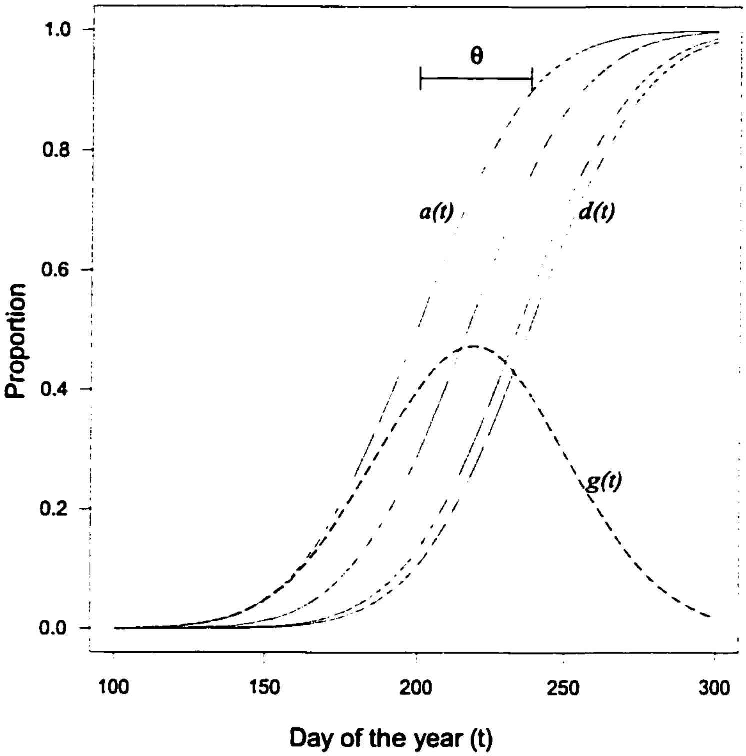


Figure 6.2: Schematic of the curves used to structure the dynamic model. For females, the arrival cumulative proportion curve ($a(t)$) plus the estimate of residence time (θ), is used to obtain the departure cumulative proportion curve ($d(t)$). The difference between the arrival and departure curves gives the proportion of the total population of spawning females that are on the grounds throughout the season ($g(t)$). The additional curves between the arrival and departure curves indication the "arrival" to the running ripe and spent stages.

stage at least once will be staged as maturing fish at day t. This parameter ρ might be expected to change throughout the season but has been fixed here. I investigated whether there was evidence that that ρ might change in linear fashion through the season by treating it as a regression problem, estimating a slope and an intercept. This did not improve the fit of the model and the sign of the estimated slope was not consistent from year to year. The sensitivity of the parameter estimates to the assumed value was assessed. With the addition of this parameter we get

$$m_t = \frac{m'_t + (\rho r'_t)}{m'_t + r'_t + s'_t}, \quad r_t = \frac{r'_t - (\rho r'_t)}{m'_t + r'_t + s'_t}, \text{ and } \quad s_t = \frac{s'_t}{m'_t + r'_t + s'_t}$$

where the $'$ denotes the predicted value before the cycling parameter ρ was included, and the denominator ensures that the proportions sum to 1.

The number of females and males on the grounds on the first day of the model ($t = 100$) is

$$F_{100} = F^* g_{100} \quad \text{and} \quad M_{100} = M^* \phi$$

where the sum of these gives the total number of fish on the grounds (N_{100}). The total spawning population is reduced over the season by exploitation (u_t)

$$F^*_{t+1} = F^*_t - (F_t \times u_t) \quad \text{and} \quad M^*_{t+1} = M^*_t - (M_t \times u_t)$$

where $u_t = C_t/B_t$, C_t is the catch in weight at day t, and B_t is the biomass of fish on the grounds on day t calculated from the numbers of fish and using the mean weights described above

$$B_t = F_t w_f + M_t w_m.$$

As previously mentioned, it is not possible to directly estimate the residence time of male hoki as no gonad stage data is available for them. In the model described above, female residence time is an estimated parameter and a fixed proportion of males (ϕ) are assumed to be on the grounds at all times. Given some estimate of the total spawning season, S_{length}, male residence time, ϑ, is given by

$$\vartheta = S_{\text{length}} \phi.$$

It would be useful to define the total spawning season as some function of the time that female hoki are on the spawning ground. Here I used the day when 5% of the females had arrived on the grounds as the start of the season and the day when 95% of the females had left the grounds as the end of the season. As arrival and departure of females are described by normal curves I can calculate the start, end, and season duration using the estimates of the arrival standard deviation, mean arrival day, and female residence,

$$S_{\text{start}} = \alpha - (1.645\sigma) \qquad S_{\text{end}} = \alpha + \theta + (1.645\sigma).$$

Season length is then given by

$$S_{\text{length}} = \theta + (3.29\sigma), \tag{6.1}$$

with male residence and female residence previously defined, the average residence of all individuals is simply a weighted average of male and female residence,

$$\Theta = \theta\pi + \vartheta(1 - \pi). \tag{6.2}$$

6.2.3 Likelihood functions

The goal of maximum likelihood is to find the parameters that maximize the likelihood function given the data. Here I use different likelihood functions for each data source so to more appropriately model the sampling errors (Eq 6.3). The likelihood function for a single year y is,

$$\begin{aligned} l_y(\text{data}|\alpha, \sigma_\alpha^2, \theta) &= \prod_{i \in t} \frac{S_i!}{T_i!(S_i - T_i)!}(p_i)^{T_i}(1 - p_i)^{S_i - T_i} \times \\ & \prod_{i \in t} \frac{S_i!}{U_i!V_i!W_i!}(m_i)^{U_i}(r_i)^{V_i}(s_i)^{W_i} \times \\ & \prod_{j \in t} \frac{1}{\sqrt{2\pi\sigma_{y,j}^2}} \exp\left(-\frac{(\log(B_{y,j}) - \log(B_{y,j}^{\text{obs}}))^2}{2\sigma_{y,j}^2}\right) \end{aligned} \tag{6.3}$$

where i and j denote days of the season where gonad stage / sex ratio and biomass estimates are available respectively, S_i is the assumed sample size for sample i, $T_i = p_i^{\mathrm{obs}} S_i$, $U_i = m_i^{\mathrm{obs}} S_i, V_i = r_i^{\mathrm{obs}} S_i$, $W_i = s_i^{\mathrm{obs}} S_i$.

The first row of Eq. 6.3 is the binomial likelihood that was assumed for the proportion of females on the spawning grounds (i.e., F_t/N_t), the second is the multinomial likelihood assumed for the gonad-stage data. The sample sizes used for the proportion of females and gonad-stage data were the number of trips that were sampled over each two-week period scaled by a constant. The last row of Eq. 6.3 is the lognormal likelihood used for the acoustic biomass estimates. A comparison with the gamma likelihood is described in the robustness section.

The model was implemented with parameters for all years estimated together – though in the case where years are treated independently (see below) the results obtained are the same as those obtained from running the model with only data for that year. The objective function was the total likelihood over all of the k years (i.e., the product of the individual likelihoods).

6.2.4 Model structure

I use a variety of models which make different assumptions about the relationship among years in spawning dynamics. In all models I treat annual spawning biomass as an independent variable (across years) as there is no reason to believe that it will be closely related because recruitment, fishing mortality, and the proportion of mature fish that spawn each year varies annually.

Independent estimates

The standard modelling approach for a problem such as this would be to estimate the residence time and spawning dynamics separately for each year; i.e., treat the parameters as fixed effects from independent analyses. This represents a very weak and naive analysis as it ignores the obvious fact that knowledge of the residence time in one year provides information about other years. This model will be referred to as Model 1.

We can take the individual estimates of the key biological parameters (with estimated standard errors), mean arrival day, arrival standard deviation, and residence time for each year and combine them using the random effects model previously used in Chapter 3.

Using average residence time (males and females combined) Θ as an example, if we assume that the Θ_k come from a normal distribution,

$$\Theta_k \sim \mathrm{N}(\mu_\Theta, \sigma_\Theta^2),$$

the form of the joint log likelihood in terms of μ_Θ and σ_Θ^2 is,

$$l(\mu_\Theta, \sigma_\Theta^2 | \widehat{\Theta_1}, \ldots, \widehat{\Theta_k}, v_1, \ldots, v_k) = -\sum_{i=1}^{k} \frac{1}{2} \log 2\pi(\sigma_\Theta^2 + v_i) - \sum_{i=1}^{k} \frac{(\widehat{\Theta}_i - \mu_\Theta)^2}{2(\sigma_\Theta^2 + v_i)}$$

where the v_i are the estimation error variances for the Θ_i.

The interesting result is whether $\sigma_\Theta^2 > 0$. When the estimated variance is greater than zero this suggests that there are some variables, not included in the model (e.g., covariates), that are partially responsible for the observed differences. It is likely that environmental conditions and the age composition of the population both influence the onset of spawning and the length of the season. If this occurs, it is reasonable to assume that there will be interannual variation in the quantities describing spawning dynamics.

Fixed effects models

On a biological basis it could be assumed that residence time, and even arrival time should be similar from year to year. Therefore, it suggests that some benefits could be obtained by sharing information across years as they can be considered as observations of the same phenomena. We can summarize different hypotheses about the similarity in residence time across years with two models

- Model 2 (Arrival independent): Single residence time but separate (independent) arrival means for each year. Here I assume that residence time is always the same but the timing of spawning varies from year to year.

- Model 3 (Years constant): Single residence time and arrival for all years; Here we assume that spawning dynamics are the same across all years.

Mixed effects models

The estimated parameters from Model 1 were treated as random effects in a post-hoc analysis. This indicated the mean arrival day on the grounds (α_y) was similar among years but significant interannual variation existed. While such post-hoc analysis is useful for indicating the heterogeneity, it does not allow us to use this information to improve our estimates of other model parameters (e.g., spawning biomass).

A far better approach is to use an integrated analysis that incorporates this information in the estimation process; an obvious candidate is a mixed effects model. A mixed effect model is one in which some parameters are treated as fixed effects (same for all years) while others are treated as random effects (individual estimates assumed to come from a common distribution).

In Model 4, I will allow mean arrival day α to be a random variable, i.e., $\alpha_k \sim \mathrm{N}(\mu_\alpha, \sigma^2_\alpha)$ [1] and treat other parameters as fixed effects (but biomass will still be estimated independently).

In the case where a random effect is included, the likelihood becomes complicated due because we need to integrate over the random effects,

$$L = \int_{\alpha_i} \prod_k l_i(\text{data}|\alpha_i, \sigma^2_{\alpha_i}, \theta_i) g(\alpha_i) d\alpha_i. \tag{6.4}$$

The problem here is that it is not possible to analytically solve the integral in this nonlinear model. It is not feasible to use a Bayesian approach such as BUGS for this problem because the model is complex and BUGS is not efficient for such models. Here I will rely of the maximum simulated likelihood (MSL) approach introduced in Chapter 4. This involves approximating this integration though H random draws

---

[1] Note: do not confuse σ^2_α with σ^2. The former is the variability in the mean arrival day across years while the later is the variance of the arrival distribution within a year.

from $g(\alpha_i)$. The likelihood function for an individual year is now

$$l_y = \frac{1}{H}\sum_{h=1}^{H} l_i^h(\text{data}|\alpha_i^h, \sigma_{\alpha_i}^2, \theta_i). \tag{6.5}$$

The procedure for implementing the MSL approach is as follows:

1. Generate kH random normal deviates of mean zero and variance of unity that are kept throughout the estimation procedure, ϵ_i^h.
2. Provide starting values for μ_α and σ_α^2.
3. Get α_i^h based on the current estimates of μ_α and σ_α^2 using the random numbers from 1, i.e., $\alpha_i^h = \epsilon_i^h \times \sigma_\alpha^2 + \mu_\alpha$ and then calculate l_i^h.
4. Repeat 3 for all h.
5. Calculate the objective function.
6. Repeat the steps 3–5 until convergence.

Similar to the implementation in Chapter 4, I have used the method of antithetic-acceleration to improve the performance of the MSL approach. For antithetic-acceleration, we first generate $0.5kH$ random normal deviates. The second $0.5kH$ are simply the first half multiplied by -1. The improvement occurs because of the more even sampling of the likelihood surface. This improves the accuracy of the integration.

As this method involves random sampling to approximate the integral we must be aware of simulation error that occurs with the model converging at different solutions with different ϵ_y^h. To assess this, I ran the model for different seeds for the random number generator and different values for H. The final H was chosen to be one where the estimated random effect variance σ_α^2 did not vary greatly across runs with different seeds.

I can compare the fit of all of the models developed here using Akaike's Information Criteria (AIC) (Akaike 1974) where the model with the lowest AIC is the most parsimonious.

Table 6.1: Summary of the assumptions about interannual variation in spawning dynamics for each model. I – parameter was estimated independently for each year; F – parameter was estimated to be a fixed effect across years; and R – parameter was estimated to be a random effect across years from an estimated distribution. All four models were developed and implemented using AD Model Builder (ADMB) (Otter Research Ltd. 2000).

Parameter	Model 1	Model 2	Model 3	Model 4
arrival mean (α_y)	I	I	F	R
arrival sd (σ_y^2)	I	F	F	F
female residence (θ_y)	I	F	F	F
population size (N_y)	I	I	I	I

6.3 Results

I compared parameter estimates across the four model runs. I am interested in how both the estimated interannual variability in model parameters and the precision of the estimates change with different assumptions relating interannual variation in spawning dynamics. Estimates (with standard errors) for the parameters describing the arrival distribution and residence time are provided in Figure 6.3 and Table 6.3.

Parameters describing the migration to the spawning grounds were generally within reasonable bounds for all models. Estimates for Model 1 were the most uncertain but this is not surprising of course as 42 parameters were estimated in Model 1 as opposed to 18,12 and 13 parameters for the other models (Table 6.2). Mean arrival day ranged from day 187 (in 1995) to day 219 (1999). When the residence time parameters and the arrival distribution standard deviation were fixed (Model 2), the interannual variability in mean arrival day was much less and the estimates were more precise. Mean arrival day for Model 3 and the estimated mean of the random effect distribution for Model 4 were almost identical. The two-step random effects analysis of mean arrival day using the results from Model 1 yielded a positive (but not significant) variance term. This result is reinforced by the results from Model 4. This suggests that the timing of the spawning season can range by up to a month across years (95% limits for the random effects distribution for the arrival mean).

Combined residence time as modelled here is a direct function of five model parameters but I only consider the final parameter here. Only Model 1 treated residence time as a parameter that can vary across years. The variation in the individual estimates was not large though (38.4–49.3 days) and the random effects analysis indicated no significant interannual variation. The other three models had estimates (of the single residence time) within 0.2 days, 42.2–42.4 days. There is no evidence in the data (as modelled here) that residence time varies significantly across years.

The annual spawning biomass trajectories are similar for each of the four models, though Model 1 displays the greatest reduction in spawning biomass over the time series (Figure 6.4). The biomass estimates were associated with quite small standard errors, reflecting coefficients of variation (c.v.s) from 8–20% across all models. The current assessment model indices are much lower as they assume that all of the spawning population is on the grounds during the survey. All models show a greater relative decline than that described by the indices used in the current assessment.

The model fit to the observed data was generally similar for all models. Models 3 and 4 gave identical fits so only Model 3 is discussed here. Model 3 did not fit well to the proportion of females in the catches for 1995 (Figure 6.5) but fits were similar for all other years. All models predicted similar trends in proportions in each of the three gonad stages through the spawning season (Figure 6.6). The fit to the 1995 data was more variable and none of the models could follow the steep decline in the proportion of mature (stage 1) fish in the catches. The predicted biomass on the spawning grounds always tracked through the acoustic estimates (Figure 6.7). Interestingly the assumption of sampling the plateau appears to be met years 1995–98, but there is fairly strong evidence that the peak spawning biomass was missed in years 1993, 1994, and 1999.

It is useful to use the calculated Akaike's Information Criteria (AIC) for each model to determine the best (most parsimonious) model. The AIC for Models 2, 3, and 4 were very similar and much lower than that for Model 1. As it appears that the peak of the arrival migration does vary from year to year, Model 4 probably provides the best description of the dynamics (of the models examined).

Table 6.2: Maximum likelihood estimates of mean arrival day α, arrival standard deviation σ, male and female residence combined Θ, and total spawning biomass B^* by year and model. Standard errors are in parentheses and estimated random effect (R.E.) distributions are given where they were estimated. Arrival day for Model 4 is the mean of the random effects. A description of the models is provided in Table 6.1.

Year	Model 1		Model 2		Model 3		Model 4	
Mean arrival day								
1993	207.5	(7.4)	208.7	(5.4)	204.0	(2.8)	204.1	(3.4)
1994	210.8	(11.1)	207.7	(9.2)	204.0	(2.8)	204.1	(3.4)
1995	187.2	(6.2)	188.7	(5.9)	204.0	(2.8)	204.1	(3.4)
1996	202.0	(8.2)	204.3	(6.0)	204.0	(2.8)	204.1	(3.4)
1997	201.8	(7.2)	202.8	(6.1)	204.0	(2.8)	204.1	(3.4)
1998	201.1	(7.4)	200.5	(5.9)	204.0	(2.8)	204.1	(3.4)
1999	219.4	(12.5)	214.6	(5.6)	204.0	(2.8)	204.1	(3.4)
R.E. mean	201.7						204.1	
R.E. sd	4.7						5.4	
Arrival sd σ								
1993	33.9	(7.7)	28.6	(2.1)	29.1	(2.2)	28.8	(2.1)
1994	31.8	(10.7)	28.6	(2.1)	29.1	(2.2)	28.8	(2.1)
1995	22.3	(2.8)	28.6	(2.1)	29.1	(2.2)	28.8	(2.1)
1996	29.7	(5.8)	28.6	(2.1)	29.1	(2.2)	28.8	(2.1)
1997	27.9	(5.4)	28.6	(2.1)	29.1	(2.2)	28.8	(2.1)
1998	27.9	(4.8)	28.6	(2.1)	29.1	(2.2)	28.8	(2.1)
1999	32.2	(10.3)	28.6	(2.1)	29.1	(2.2)	28.8	(2.1)
R.E. mean	26.2							
R.E. sd	0.6							
Combined residence time, Θ								
1993	42.7	(8.2)	42.4	(2.7)	42.2	(2.7)	42.2	(2.7)
1994	38.4	(10.1)	42.4	(2.7)	42.2	(2.7)	42.2	(2.7)
1995	44.6	(5.7)	42.4	(2.7)	42.2	(2.7)	42.2	(2.7)
1996	46.9	(8.8)	42.4	(2.7)	42.2	(2.7)	42.2	(2.7)
1997	42.2	(6.8)	42.4	(2.7)	42.2	(2.7)	42.2	(2.7)
1998	45.7	(7.7)	42.4	(2.7)	42.2	(2.7)	42.2	(2.7)
1999	49.3	(19.7)	42.4	(2.7)	42.2	(2.7)	42.2	(2.7)
R.E. mean	43.9							
R.E. sd	0.0							
Total spawning biomass B								
1993	712.5	(123.1)	657.1	(100.4)	634.0	(90.9)	647.0	(95.7)
1994	739.7	(150.2)	639.0	(116.4)	614.9	(78.5)	615.8	(83.0)
1995	290.7	(44.9)	368.5	(42.6)	404.6	(47.3)	378.1	(44.7)
1996	202.0	(37.1)	213.9	(27.5)	216.2	(25.3)	215.4	(25.9)
1997	297.8	(52.4)	303.8	(28.3)	310.5	(28.6)	308.9	(29.0)
1998	151.3	(23.3)	162.7	(15.6)	167.1	(16.2)	165.6	(16.1)
1999	277.7	(42.6)	270.7	(25.6)	258.5	(21.2)	262.5	(22.7)

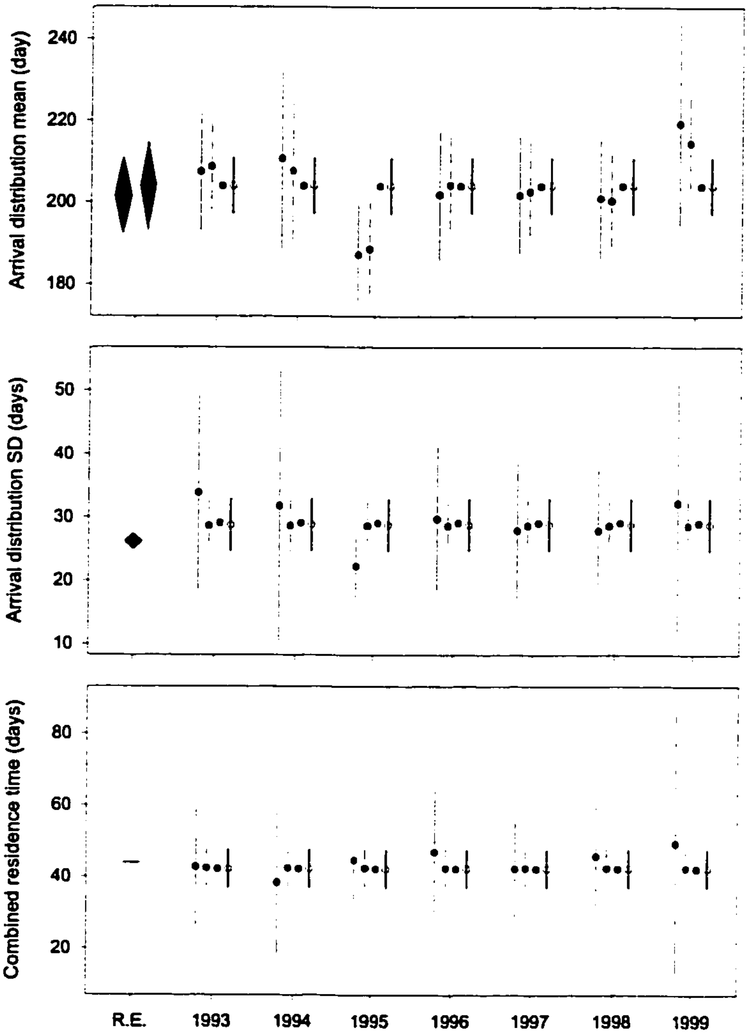


Figure 6.3: Estimates of residence time (combined across both sexes, Θ), mean arrival day of females (α), and variance of the arrival distribution (σ^2) for each year and model: Model 1 (solid line); Model 2 (long dash); Model 3 (dotted line); and Model 4 (solid line open circle). Error bars indicate ± 2 standard deviations. The estimated random effects distribution (R.E.) for Model 1 (dark) and Model 4 (light) are given where they were estimated.

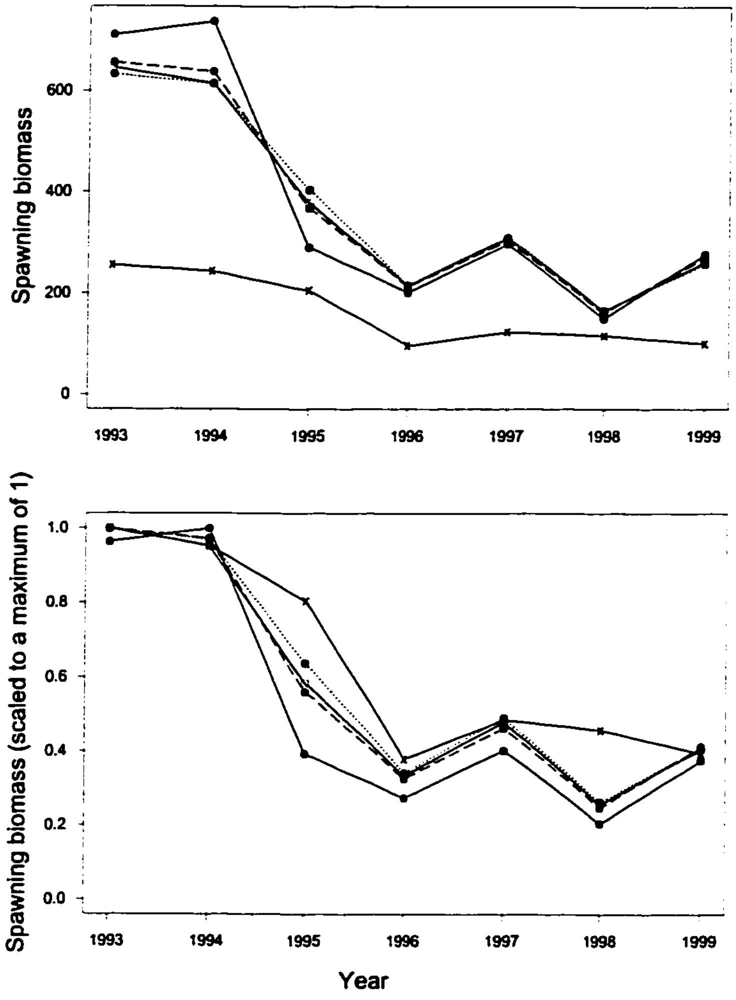


Figure 6.4: Estimates of total spawning biomass (top) and relative total spawning biomass (bottom) (scaled to a maximum of one) on the Cook Strait spawning ground for each model: Model 1 (solid line); Model 2 (long dash); Model 3 (dotted line); and Model 4 (solid line hollow circle). The biomass estimates used in the current stock assessment models are presented as an "x" with the light solid line. The c.v.s for the biomass estimates were 0.15–0.20, 0.09–0.18, and 0.08–0.14 (for Models 3 and 4), for each model respectively.

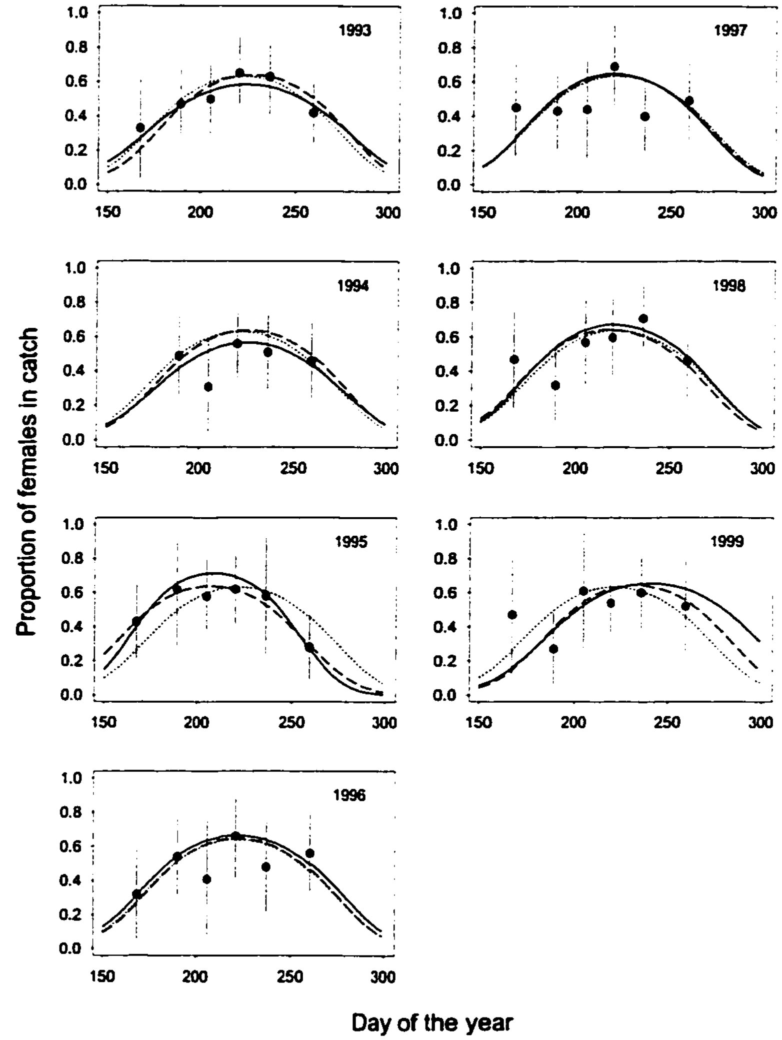


Figure 6.5: Fit to the proportion of females on the Cook Strait spawning ground for Models 1 (solid line), 2 (long dash), and 3 (dotted line). The error bars represent the 95% likelihood-based confidence limits based on the assumed sample size and the estimated proportion.

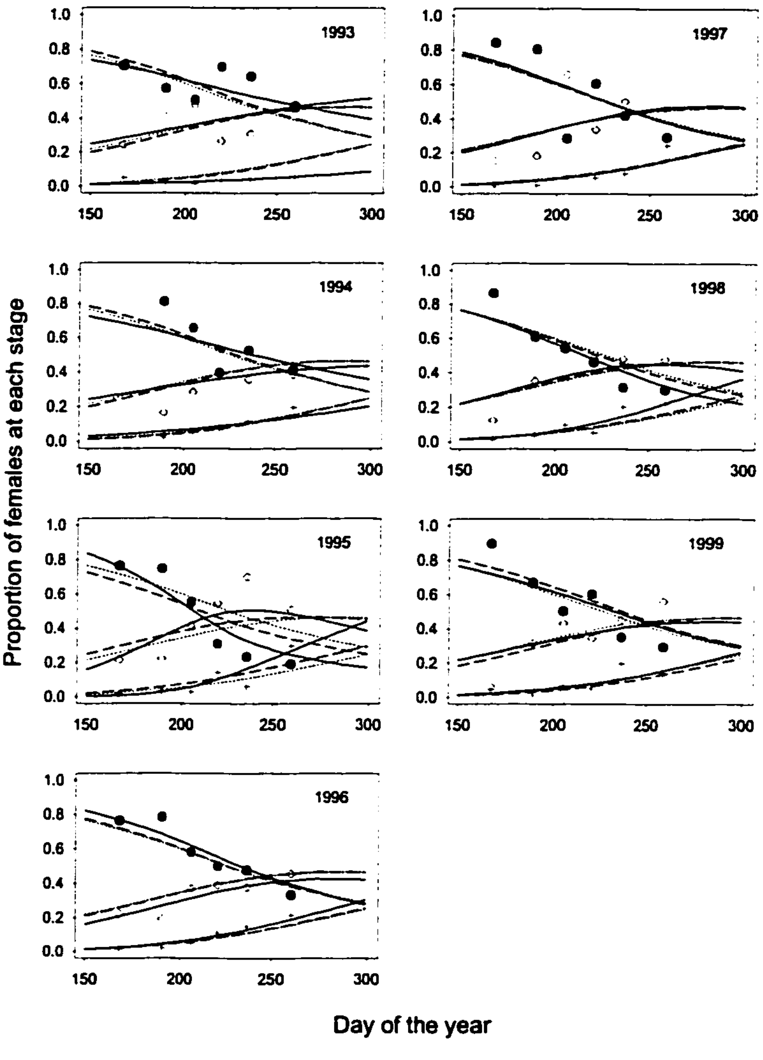


Figure 6.6: Fit to the proportions at each gonad stage on the Cook Strait spawning ground for Models 1 (solid line), 2 (long dash), and 3 (dotted line). Mature/maturing - solid squares; running ripe - hollow diamonds; and spent - (+). To simplify this figure I have not included error bars. The sample sizes are the same as those assumed for the sex-ratio data.

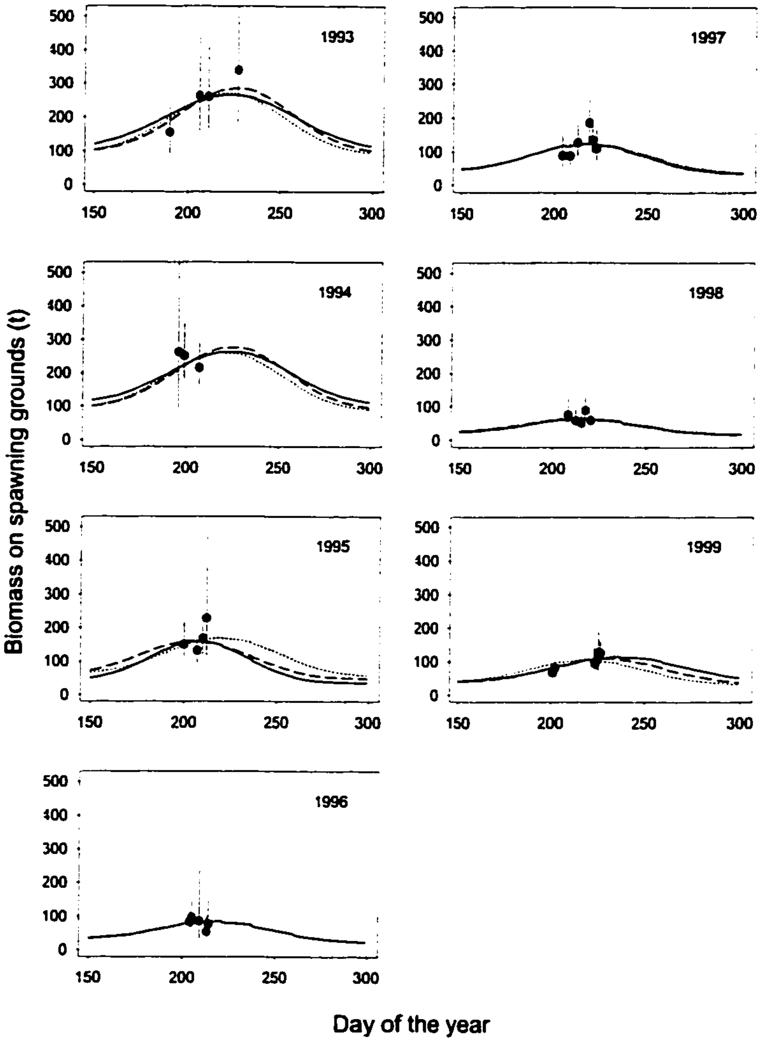


Figure 6.7: Fit to the acoustic snapshot estimates on the Cook Strait spawning ground for Models 1 (solid line), 2 (long dash), and 3 (dotted line). The error bars represent the 95% confidence limits for the snapshot estimate assuming lognormal errors.

Table 6.3: Overall fit for each model. n is the number of parameters estimated, $\mathcal{L}$ is the total negative log likelihood, and AIC is Akaike's Information Criterion (smaller is better), AIC $= 2\mathcal{L} - 2n$.

	Model 1	Model 2	Model 3	Model 4
n	42	18	12	13
$\mathcal{L}$	1084	1092	1098	1097
AIC	2252	2219	2220	2219

6.3.1 Robustness and simulation testing

The robustness of the approach was tested by two methods. First, alternative models were fit by modifying the model structure, e.g., gamma errors were assumed instead of lognormal errors. Second, simulated data were generated from models with known parameter values, and I tested if I could recover these values. This approach was only considered for Models 1–3.

As the MSL approach is based on simulations using random numbers there is a chance for extra variability in the results – i.e., obtaining different results with a different seed for the random number generator or a different number of simulations. For 20 simulations (as used for the results presented) a range of different seeds did not lead to differences of more than 10% in the estimate of the random effects variance. This did not reduce when 40 samples were used. As the 10% is less than the estimates of precision for the original estimates the sample size of 20 appears adequate for this problem. When a sample size of 10 is used, performance is quite variable with 50% of seeds leading to estimates of close to zero for the random effects variance. Inter-seed variation was reduced greatly by the use of antithetic acceleration.

Assuming gamma errors instead of lognormal errors made very little difference to the estimates (estimated spawner abundance changed less than 3%), probably because the assumed c.v.s for acoustic estimates were relatively small. The relative weighting of the sex ratio and the gonad stage composition data to the acoustic estimates depends upon the effective sample size estimated from the sampling protocols. Estimating the effective sample size is inherently difficult because of the difficulty of sampling at random from a commercial operation. When I halve the sample sizes, the model becomes unstable as this data is critical to the estimation of the duration of the gonad stages. When the assumed effective sample size was doubled, then the estimates of spawning biomass changed by less than 2%.

Estimates of spawner biomass and females residence time were most sensitive to the proportion of the total population of spawning males assumed to be on the grounds through the season (ϕ). A 33% change in ϕ resulted in around -15–20% changes in estimates of spawning biomass and residence time with a higher ϕ giving lower biomass. Estimates of biomass were quite insensitive to the cycle parameter ρ. This is most likely because it is confounded with the estimated duration of the mature and running ripe stages (β and γ). While the ratio, β/γ was sensitive to changes in ρ the sum was not.

Simulation tests were used to confirm that I could recover the model parameters from the model, assuming the same model and sampling distributions as assumed to generate the simulated data. The simulations are also useful for testing model parsimony. Results of the simulations were assessed by the relative bias and relative root mean squared error (RMSE) in the estimates of spawning biomass and average female residence time. I define bias and RMSE in terms of the true value, Z, and the estimate, $\hat{Z}$ (Kendall et al. 1987).

$$\%\text{ Bias} = \left(\mathrm{E}\left[(\hat{Z}|Z)\right] - Z\right)\frac{100}{Z} \tag{6.6}$$

$$\%\text{ RMSE} = \left(\left[b(\hat{Z})\right]^2 + \mathrm{Var}(\hat{Z}|Z)\right)^{\frac{1}{2}}\frac{100}{Z} \tag{6.7}$$

where $b(\hat{Z})$ is the bias in the estimates.

The parameter values used in the simulation were those estimated with each model. Here I report only the results for spawning biomass and residence time (Table 6.4) though the results for other parameters showed that the models were good at estimating mean arrival day (less than 2% bias for all simulations).

Spawning biomass

When the real model has independent parameters (i.e., Model 1) the estimation model with this assumption performed worst while Model 2 had the lowest % bias (+1.5%) and Model 3 the lowest % RMSE. In all simulations there was a positive bias in the estimates of spawning biomass although this was generally small (never greater than 1.5% for Model 2).

Average female residence

The models performed quite poorly in estimating average residence time in terms of both bias and RMSE. Estimates of bias were five times those for biomass and RMSE was generally doubled. The high RMSE is not surprising as average residence is the function of most of the parameters in the model but the high (5.8–42.2%) and positive biases were. Estimates of the spent duration contributed most of the bias to average residence time.

Model parsimony

To determine overall model performance I ranked each model in terms of its performance for each set of simulations. In terms of bias in the estimates of spawning biomass, Model 2 performed best regardless of the true model, while Model 3 was slightly better than Model 2 in terms of RMSE. Overall, Model 3 most often had the lowest bias and RMSE for a given simulation and parameter. The superior performance of simpler models (even when they are not the true model) is not unusual (Hilborn and Mangel 1997).

6.4 Discussion

An often common problem when modelling population dynamics in fish is that the dynamics (and therefore models) are complex but there are often insufficient data to reliably estimate the parameters of these models. This is particularly the case when modelling stage-structured data (Wood 1997) as was done here. Here I have

Table 6.4: Median percent bias (top) and percent root mean squared error (RMSE) (bottom) for the estimates of spawning biomass and average female residence time (across years) for 200 simulated data sets. Bias and RMSE were divided by the true simulated value (and multiplied by 100) to give a relative measure comparable across models.

Spawning biomass

	Simulation		
Estimation	Model 1	Model 2	Model 3
Bias			
Model 1	4.4	10.3	10.3
Model 2	1.5	0.5	1.0
Model 3	1.9	1.5	3.6
RMSE			
Model 1	25.7	76.3	49.6
Model 2	19.8	17.9	26.2
Model 3	18.0	18.8	23.3

Average female residence (θ)

	Simulation		
Estimation	Model 1	Model 2	Model 3
Bias			
Model 1	34.7	33.8	42.2
Model 2	34.8	20.0	23.5
Model 3	5.8	12.2	11.4
RMSE			
Model 1	72.7	81.6	86.6
Model 2	61.2	34.0	44.0
Model 3	38.0	43.0	39.7

demonstrated how it is possible to integrate a random effects meta-analysis into a complex nonlinear dynamic model to overcome some of these problems.

A key feature of this model was that it integrated a variety of data. I have used both gonad-stage and biomass estimates together to estimate the timing of spawning, residence time, and total spawning biomass simultaneously. Previous models used to estimate hoki spawning abundance have assumed that the acoustic snapshots provided an estimate of biomass at the peak of the spawning season (Coombs and Cordue 1995). In years where the peak spawning was considered to be well after the acoustic surveys, ad-hoc corrections were made to biomass estimates (Annala et al. 2000).

In this section I will discuss a number of the important implications of this work. First, I will discuss the biological and stock assessment implications of the results. Second, I will compare my approach to the way others have analyzed similar types of data with different species. Third, I review the models that were used and how they could be improved. Last, I outline a number of research recommendations for hoki that came out as a result of this analysis.

Comparison of results

Here I estimated residence time of females on the spawning ground. By accounting for the time that spent fish remain on the spawning grounds, I can compare my estimates of spawning duration to previous studies. The estimates of 32 days are higher than the range of 20–27 days found by Langley (1993) for hoki at the other major spawning ground. The estimates are slightly higher than the estimates for wild Atlantic cod (25.2 days) (Hutchings and Myers 1993) but in the middle of the range of 22–48 days obtained for captive Atlantic cod (Kjesbu et al. 1996). Accounting for the positive bias found in the simulation study would make my estimates more consistent with those from previous studies.

There is strong evidence for variability in the timing of spawning across years. This is consistent with other observations of spawning timing for hoki (Gunn et al. 1989) or other Gadiformes, e.g.,Atlantic cod (Hutchings and Myers 1993).

I compared estimates from all models with the biomass estimates used as indices of relative abundance in the current hoki stock assessment (Annala et al. 2000). The indices used in the assessment were based on the assumption that the acoustic snapshots were taken at the peak of the season, except in the last two years where the biomass estimate was scaled due to concerns that the survey missed the peak of the spawning season (Annala et al. 2000). Trends in relative abundance are again quite similar across models but there is evidence of a greater decline in abundance over time for the models presented here than that suggested by the biomass estimates used in the present stock assessment. Also, there was less variation in estimated biomass over the years for my analysis. The models developed here allow for a more objective approach to determining if the acoustic survey did in fact cover the peak of the spawning season.

Comparisons to analyses of other taxa

The stage data used here allows for a much more precise model than would have been possible by using the acoustic survey alone. Vogt and Morton (1991) found that by incorporating age data to estimate survival and age-dependent trapability they were able to greatly improve the reliability of their estimates of population size in a mark-recapture model. Similarly, Myers et al. (1997) were able to use stage composition data to improve estimates of a grey seal *Halichoerus grypus* population from a mark-recapture study. Although these models are different from the one used here, the general principle of using stage data, along with population estimates, can significantly improve estimates. Although my models have some of the advantages of these models, the hierarchical models allow stage data to be optimally incorporated over several years or over several populations.

The models developed here are very similar to those developed for estimating escapement in salmon (English et al. 1992; Hilborn et al. 1999; Su et al. 2001). In fact it is the area-under-the-curve approach (English et al. 1992; Hilborn et al. 1999) to estimating abundance that was applied to hoki acoustic surveys (Coombs and Cordue 1995).

Modelling approach

I have taken advantage of methods that allow sharing of information across years (or studies). In this analysis I have treated interannual variation in the mean arrival day three ways: (1) independent from year to year; (2) fixed (constant) across years; and (3) as a random variable from an estimated distribution. I found that the models that shared information performed much better than those that did not, even when there was considerable true variation across years. Adkison and Su (2001) also reported that their hierarchical estimators performed much better than the individual model estimates. A logical further step in this analysis would be to treat multiple parameters as random effects. Given the results of the two-step random effects analysis of the results from Model 1, this approach was not necessary here but may be important for other applications.

Here I used a maximum simulated likelihood approach to implement a random effects analysis. This approach provides the most feasible way to implement computationally intensive nonlinear random effects models. While it does not require approximations to the integrals as is required for the NLME routine in S-PLUS, its simulation approach requires that multiple runs be conducted to check for simulation noise.

By removing the likelihood component for the proportion of females from the models it was possible to determine the reliance of the models on these data. Interannual variability in all model parameters increased greatly with a number of "unlikely" results. With the current model and data structures, it does not appear possible to estimate residence time with gonad stage data alone. A number of different combinations of arrival and residence time can produce equally likely fits to the data.

There are a number of possible extensions for the sex- and stage-structured model presented here. First, it would be useful to introduce age structure (but not necessarily a category for each age) into the model. This was considered in the development of the current model but was not developed with due to problems in estimating parameters with the highly noisy age-disaggregated data. Using a hierarchical model, it could be possible to overcome many of the problems associated with the addition of age-structure to the models. While the number of parameters being estimated increases, the use of constraints (i.e., treatment of parameters as fixed and random effects) assists parameter estimation. A further advantage of the hierarchical approach is that it may be possible to obtain better estimates of age-specific process (e.g., larger fish arriving on the grounds earlier) by sharing the information and constraining estimates of the relationship across years. I have used a similar approach here by constraining the arrival standard deviation to be the same across all years.

Research recommendations

One major concern for the reliability of the analysis here is the reliance on fishery data for estimates of the sex- and stage-composition of the population on the spawning ground. This does not imply that the data are wrong in any sense. Rather the issue is whether samples from commercial catches are representative of fish on the grounds, i.e., those fish sampled by the acoustic survey. The acoustic survey is undertaken using a stratified random approach (Coombs and Cordue 1995), but the distribution of fishing effort is far from random with respect to the distribution of the fish. Fishers are likely to be targeting areas where they can achieve the highest catch rates and/or catches of hoki of a particular size. This strategy could vary across the fleet and change through the season (Langley 1993). The affect of this on the representativeness of the samples is not clear. To best obtain representative samples of the spawning population, commercial samples should be either supplemented or replaced by fishery-independent sampling during the acoustic survey. The merging of the two sources would be difficult, but the use of a commercial "catching" vessel associated with the acoustic survey might provide representative samples (from both dense aggregations, and less dense background "scatter" of hoki).

The data used for this analysis were aggregated across two-week strata. This is necessary because of the high inter-haul variability in the sex and gonad stage composition often associated with catches of spawning Gadiformes (Langley 1993; Morgan and Trippel 1996). However, much of the spawning behaviour occurs at shorter time frames. With female residence time of 30 days, a fish might be "represented" within only two observations of the spawning population from the grounds. The use of survey data could allow for the analysis of the gonad stage data on a finer scale.

Model-based estimation of residence time could be improved by laboratory-based studies of spawning behaviour of hoki. Laboratory based approaches have proved useful in improving our understanding of the spawning dynamics of Atlantic cod (Kjesbu et al. 1996). Although estimates of residence time from such studies might not necessarily reflect what might occur in the wild, it could provide useful estimates of the number of spawning batches and the length of time between batches. There might be considerable technical difficulties in doing this though, as hoki live and spawn at considerable depths (200-600m).

Another potentially useful piece of biological data that could be used in models of residence time is gonad weight through the season. It might be expected that this would decline as hoki spawned each batch of eggs. However, the hydration process prior to releasing each batch may complicate this. This information would probably be best used in an age-structured approach.

The model results suggest that the peak of the migration to the spawning grounds varies across years. Therefore, a short acoustic survey conducted during the same period every year could sample different stages of the spawning season each year. Gunn et al. (1989) proposed a number environmental cues that may be important for the dynamics of hoki populations that spawn off southern Australia. It may be useful to investigate if any of these factors (or others) can be used to predict the timing of spawning in a given year.

When estimates of spawning biomass from acoustic surveys are used in assessment models there is a further consideration. There can be a complication if not all mature fish spawn in a given year. This appears to be the case for hoki (Livingston et al. 1997; Bulman et al. 1999). Livingston et al. (1997) used trawl survey data from the feeding grounds for the western hoki stock and determined that 62–82% of females spawned in each year. A similar pattern was observed for *Coryphaenoides rupestris* (Macrouridae) from the North Atlantic where, it was found that on average females spawned every two years and males every year (Alkseyev 1995). Knowledge of spawning frequency is necessary for the appropriate interpretation of estimates of spawning biomass.

Finally, it would be possible to use the model and parameter estimates obtained here to investigate how the sampling strategy, e.g., the length and timing of the acoustic surveys, could be improved to allow for more reliable estimates. This was outside the scope of this thesis, so all simulations were undertaken using based on the historical sampling strategy. Adkison and Su (2001) used their model of pink salmon to investigate how run size could be estimated better by making additional counts. They found that the model estimates could be improved dramatically if a single salmon count was conducted after the peak of the run. This type of analysis will be an important future step for those wishing to optimize future acoustic surveys.

Chapter 7

Improving stock assessment with empirically-derived priors

7.1 Introduction

Possibly the greatest challenge in fisheries management is making decisions in the face of uncertainty. Often data for the stock under consideration can be of such poor quality (i.e., short and/or noisy time series) that it is not possible to reliably estimate the key quantities required for management. This deficiency is most often the case in developing fisheries (McAllister and Kirkwood 1998). In some circumstances models can be used that make predictions about these quantities that are inconsistent with all other data for that species. For these reasons it is critical that all 'relevant' data be used when assessing an exploited stock. The aim of this chapter, is to show how results from meta-analysis can help to partly overcome some of these problems.

For meta-analysis to be accepted it must be shown to have practical applications. Together with Bayesian methods of stock assessment, meta-analysis can play an important role in estimating key parameters and thus reducing the uncertainty that plagues fisheries management. The Bayesian approach provides a useful framework for fisheries modelling as it allows for external information to be explicitly incorporated into the estimation procedure (Punt and Hilborn 1997; McAllister and Kirkwood 1998). It is also useful for incorporating uncertainty into decision analysis (Punt and Hilborn 1997; McAllister and Kirkwood 1998; Meyer and Millar 1999b). A random effects distribution from a meta-analysis can be used to represent the 'prior' information about a parameter. A prior distribution is a quantitative summary of our knowledge about the likely values of a parameter *before* considering the data for the stock.

While these Bayesian methods seem intuitive – especially the ability to explicitly incorporate external information – there has been considerable debate among both statisticians (Carlin and Louis 1996) and assessment scientists (Walters and Ludwig 1994; Millar 2001) about the exact form that prior distributions should take. Many arguments are based around the use of subjective priors (chosen by a panel of "experts") or the use of prior distributions chosen to represent a complete lack of knowledge (Walters and Ludwig 1994; Millar 2001). I argue that empirically derived priors, based on a thorough examination of existing data, should be used whenever possible. If we are not willing to do this we can be mislead by our single-stock data. In many instances, information can be obtained from other populations of the same or ecologically-similar species (Hilborn and Liermann 1998; Myers and Mertz 1998).

In this chapter I combine empirically-derived priors from the analysis of the relationship between catch-per-unit-effort (CPUE) and abundance in Chapter 3, and analysis of the lifetime maximum reproductive rate of Myers et al. (1999), with a simple population dynamics model. I call these empirically-derived or empirical priors as they have been derived from the analysis of many data sets. They are different from the more commonly used subjective priors based on expert opinion. I illustrate how these priors can be used to improve our assessment of exploited fish populations where there is limited data available for the stock of interest. For simple population dynamics models, as would likely be used in developing fisheries, I show that empirical priors are critical.

My focus differs from that of previous studies (e.g., McAllister and Kirkwood 1998) in that I will examine how priors can improve estimates of important parameters, rather than investigate how priors influence the performance of different harvest strategies. I compare the reliability of estimates for key model parameters with and without priors derived from the analysis of many data sets (i.e., meta-analysis).

I use a two-fold approach to demonstrate the influence of priors: (1) an analysis of real data for the hake fishery (a mixed fishery for *Merluccius capensis* and *M. paradoxus*) off the coast of Namibia (presented by Hilborn and Mangel 1997); and (2) a simulation study in which the values of the parameters and biases in the priors are known.

7.2 Methods

In the following sections I describe how the results from meta-analysis and a simple stock assessment model can be combined to provide a framework to improve stock assessment and subsequently fisheries management (Figure 7.1). First I describe the prior distributions derived from meta-analysis. Second, I describe the population dynamics model and the key model parameters for which I have prior information. Finally, I discuss the practical aspects of the implementation of a Bayesian stock assessment.

7.2.1 Empirically derived priors

Prior information for three parameters will be used here (Table 7.1). I use the priors developed for the order Gadiformes in Chapter 3 for β, the parameter describing the relationship between CPUE abundance indices and true abundance, and σ the assumed coefficient of variation (c.v.) of the observation errors of the CPUE series. The final prior is for the lifetime maximum reproductive rate, a critical parameter describing populations dynamics at low population sizes and deriving many commonly used biological reference points (Myers et al. 1999). The prior information for the lifetime maximum reproductive rate is for hake (Family Merlucciidae) and comes from an updated analysis (R. A. Myers, unpublished analysis) of that presented by Myers et al. (1999).

In Chapter 3, where the priors for β were derived, I warned that the estimates were likely to be positively biased. Biased priors, especially those that are also too precise, can cause problems when they are used in stock assessment (McAllister and Kirkwood 1998). Here, I have taken my own recommendation and reduced the mean of the prior distributions by 10% but have kept the variance of the priors the same

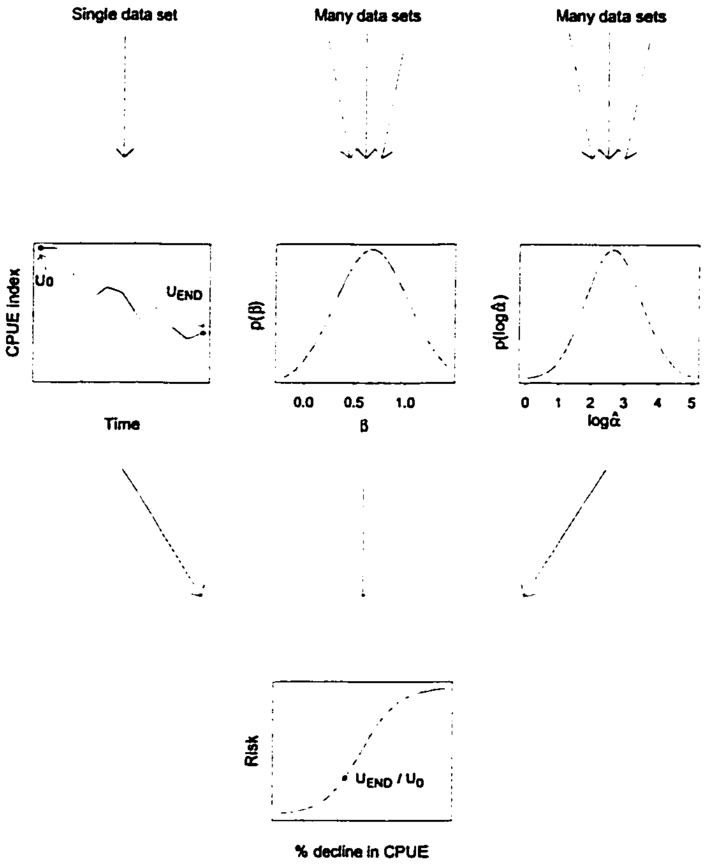


Figure 7.1: Framework for using empirical priors in the assessment of populations using CPUE data. The final risk can be determined using only the priors and CPUE decline or with other information (e.g., catches inside a dynamic model).

(i.e., I have also increased the coefficient of variation).

The lifetime maximum reproductive rate is calculated from a standardization of the slope at the origin of the spawner recruitment curve. Here I summarize how this is done (fuller details are provided by Myers et al. (1999)). To estimate the slope at the origin, α, the Ricker curve is often preferred over the Beverton-Holt as it provides more conservative estimates (Myers et al. 1999). Using the Ricker spawner-recruitment model we have

$$R_t = \alpha B_t e^{-\beta B_t}. \tag{7.1}$$

Density-dependent mortality is assumed to be the product of β times the spawning biomass (B_t). From this we can derive $\hat{\alpha}$ which represents the lifetime number of spawners produced by each spawner at very low abundance,

$$\hat{\alpha} = \alpha \cdot \mathrm{SPR}_{F=0}$$

where $\mathrm{SPR}_{F=0}$ is the spawning biomass resulting from each recruit in the limit of no fishing mortality ($F = 0$), based on estimates of natural mortality, maturity schedules and growth.

For the models described in the next section, I prefer to use steepness (z) rather than the lifetime maximum reproductive rate. The term steepness was first defined by Mace and Doonan (1988) and represents the recruitment, relative to recruitment at equilibrium in the absence of fishing, that occurs when spawner abundance has been reduced to 20% of its virgin level. This can be given in terms of $\hat{\alpha}$,

$$z = \frac{\hat{\alpha}}{4 + \hat{\alpha}}. \tag{7.2}$$

In meta-analyses of the lifetime maximum reproductive rate, $\log \hat{\alpha}$ is treated as a normal random variable rather than $\hat{\alpha}$. Therefore the prior for $\hat{\alpha}$ that is used is actually a normal prior on $\log \hat{\alpha}$. I assume that the prior distributions for β,σ, and $\log \hat{\alpha}$ are described by normal distributions and are mutually independent.

In addition to the empirical priors presented in Table 7.1, prior distributions were required for all of the other model parameters. All were assumed to have uniform

Table 7.1: Parameters describing the empirical prior distributions for the shape of the relationship between catch-per-unit-effort and abundance, β; the logarithm of the standard deviation of the observation errors, $\log\sigma$; and the logarithm of the lifetime maximum reproductive rate, $\log\tilde{\alpha}$.

μ_β	σ^2_β	$\mu_{\log\tilde{\alpha}}$	$\sigma^2_{\log\tilde{\alpha}}$	$\mu_{\log\sigma}$	$\sigma^2_{\log\sigma}$
0.657	0.10	2.667	0.693	-0.65	0.11

priors with upper and lower bounds that were wide enough to include all possible values. For the catchability coefficient, the uniform prior was on $\log q$ as this is more appropriate for a scale parameter such as catchability (Meyer and Millar 1999b).

7.2.2 Population dynamics model

The population dynamics assumed here are based on the "lagged recruitment, survival, and growth (LRSG)" model described by Hilborn and Mangel (1997). When compared to the logistic model, the LRSG model has the added realism of the lag in recruitment (i.e., the number of years taken for a fish to reach maturity). Also for the logistic model, the population biomass that produces the greatest production or yield, B_{MSY}, is defined to be 50% of the carrying capacity (here referred to as B_0). This assumption can be relaxed in the LRSG model by incorporating a spawner-recruitment curve. The dynamics of the LRSG model are given by:

$$B_{t+1} = sB_t + R_t - C_t \tag{7.3}$$

where R_t is recruitment that resulted from B_{t-L} (following equation), L is the lag between spawning and subsequent recruitment to the mature population, and s accounts for changes in biomass due to natural survival and growth of individuals within the population. The constant L leads to two assumptions: (1) there is a knife edge maturity ogive; and (2) that only mature fish are vulnerable to fishing. Following the example of Hilborn and Mangel (1997), recruitment is described by the Beverton-Holt spawner-recruitment relationship

$$R_t = \frac{B_{t-L}}{a + bB_{t-L}}$$

where the recruitment at the equilibrium unfished state is

$$R_0 = B_0(1-s).$$

It would also be possible to use a Ricker model instead of the Beverton-Holt for recruitment. However, there are no closed forms for the reference points when a Ricker is assumed so the analysis is more complicated.

The spawner-recruitment relationship parameters a and b are defined by

$$a = \frac{B_0}{R_0}\left(1 - \frac{z-0.2}{0.8z}\right),$$

$$b = \frac{z-0.2}{0.8zR_0}.$$

From this we can obtain estimates of the fishery management targets B_{MSY}

$$B_{\mathrm{MSY}} = \frac{1}{b}\left(\sqrt{\frac{a}{1-s}} - a\right), \tag{7.4}$$

and the maximum yield that can be taken at that biomass, MSY,

$$\mathrm{MSY} = B_{\mathrm{MSY}}\left(s - 1 + \frac{1}{a + bB_{\mathrm{MSY}}}\right). \tag{7.5}$$

We can write the ratio of B_{MSY} to B0 as,

$$\frac{B_{\mathrm{MSY}}}{B_0} = \frac{\sqrt{1 - \frac{z-0.2}{0.8z}} - \left(1 - \frac{z-0.2}{0.8z}\right)}{\frac{z-0.2}{0.8z}}, \tag{7.6}$$

and can see that this is only dependent upon z.

The prior for $\log\tilde{\alpha}$ can now be used with Eqs. 7.2 and 7.6 to calculate the implied prior (i.e., implied based on the priors for the other parameters) for B_{MSY}/B_0. The ratio B_{MSY}/B_0 is a commonly used target reference point (Francis 1999). The prior for $\log \hat{\alpha}$ implies that hake are likely to be moderately resilient to overfishing with most of the weight on the prior distribution for B_{MSY}/B_0 between 0.20 and 0.45 (Figure 7.2).

Implementation

Model-based assessment using Bayes theorem

We can use the LRSG model to estimate the unknown parameters (R_0, s, $\log \hat{\alpha}$, σ, and q) using time series of catch and CPUE data. I assume that the catch data is known without error and that CPUE series (U_t) has lognormal errors and is related to true abundance through the power model (Eq. 7.7).

$$U_t = qB_t^{\beta} \exp(\epsilon_t). \tag{7.7}$$

Assuming lognormal errors for the CPUE indices, the parameter estimates are found by minimizing the negative log likelihood of the observed and predicted CPUE indices

$$\mathcal{L}(R_0, s, \log \hat{\alpha}, \sigma, q|U_t) = \sum_{t=1}^{T} \log \sigma + 0.5 \log(2\pi) + \frac{(\log(qB_t^{\beta}) - \log(U_t))^2}{2\sigma^2}. \tag{7.8}$$

I compare the risk of the stock being below B_{MSY} for three model scenarios:

- Model 1: No empirical priors and CPUE is assumed to be proportional to abundance; all parameters (except β) estimated.
- Model 2: No empirical priors, all parameters estimated.
- Model 3: Empirical priors from Table 7.1, all parameters estimated.

These scenarios were chosen to illustrate a number of points: Model 1 represents perhaps the most traditional approach to the problem; Model 2 shows that the data are not good enough to support the estimation of extra parameter without additional information; and Model 3 indicates the optimal use of prior information.

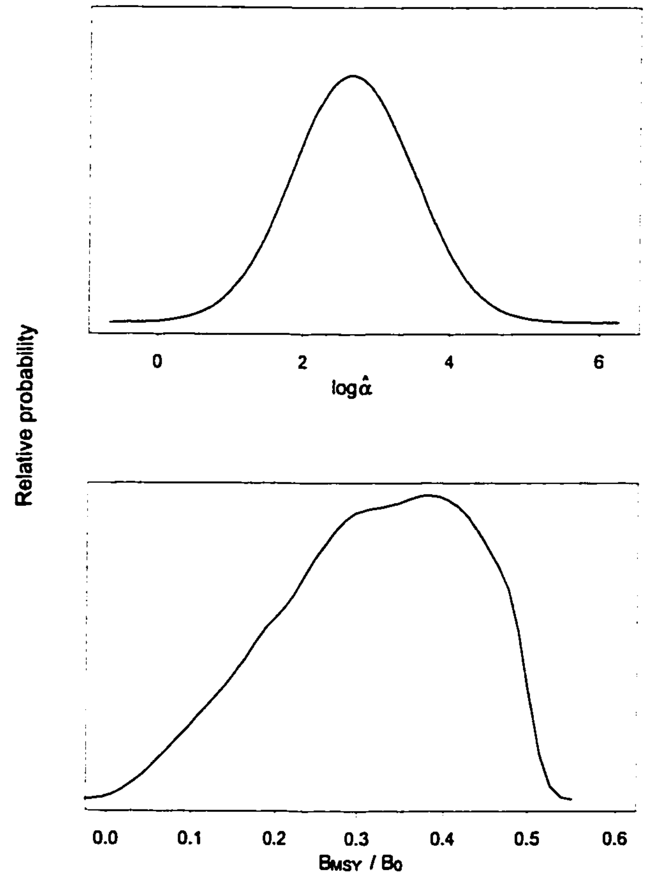


Figure 7.2: Prior probability distribution for $\log\hat{\alpha}$ and the implied prior for B_{MSY}/B_0.

To integrate the empirical priors with the data in the model it is necessary to use Bayes theorem. The Bayesian approach to inference is important for integrating empirical priors from meta-analysis. I will introduce this approach below using general notation.

The Bayesian estimation procedure combines information from two sources. The first is the statistical model of the data, $\mathbf{y}$, and the unknown parameter(s), θ, in the form of a probability distribution $f(\mathbf{y}|\theta)$. The second is the prior probability of the parameters $\pi(\theta|\eta)$, where η are the parameters of the prior distribution, and are known as "hyper-parameters" (Carlin and Louis 1996) (our hyper-parameters are fixed so they are suppressed in Eq. 7.9 below). Bayesian inference is based on the posterior distribution $p(\theta|\mathbf{y})$ derived using Bayes Theorem,

$$p(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{\int f(\mathbf{y}|\mathbf{u})\pi(\mathbf{u})d\mathbf{u}}. \tag{7.9}$$

This posterior distribution is used for Bayesian statistical inference as opposed to the likelihood approach which uses $f(\mathbf{y}|\theta)$ for inference. When θ consists of a vector of unknown parameters (R_0, s, $\log\hat{\alpha}$, σ, q, and β in my example) this posterior distribution represents the "joint" posterior distribution. By integrating over the other parameters it is possible to obtain individual or "marginal" posterior distributions for each parameter of interest. For even simple models, there are no closed forms for the integrals. Modern Bayesian methods rely on Monte Carlo Markov Chain (MCMC) sampling for the integration.

The LSRG model was implemented in AD model builder (ADMB) (Otter Research Ltd. 2000). For Bayesian inference a MCMC algorithm known as the Hastings-Metropolis (Hastings 1970) algorithm was used. This algorithm jumps from point to point in n-dimensional space (where n is the number of parameters being estimated) using a proposal function to determine in what direction and how far it should jump. If parameter values at the new point fit the data better, i.e., higher posterior probability, it accepts the new point (in n-dimensional space), if the new parameter values do not fit better, the new point is accepted with probability proportional to the fit (i.e., if the fit is almost as good the point is accepted more often than if the fit was much worse). The algorithm is initialized with starting values and can take a number of steps before it begins sampling adequately (i.e., new points are accepted with a reasonable probability). This procedure is continued for many steps, with the goal of providing a representative sample of the joint posterior distribution. The output is an autocorrelated time series of values for each parameter that is often called a MCMC chain.

The number of samples taken before the algorithm begins sampling properly is referred to as the 'burn-in' period and these samples are usually discarded.

For my analysis the algorithm was run with an initial burn-in of 500 samples and then the algorithm was run for 2,500,000 steps with every 500th sample retained for inference (total of 2500). The burn-in period is probably not as critical for the Hastings-Metropolis algorithm implemented in ADMB because ADMB uses the maximum likelihood estimates for model parameters as starting values. The sub-sampling of every 500th value is known as thinning and reduces autocorrelation. The sampling strategy employed was sufficient to provide good samples of the marginal posterior distributions for Model 3.

There are a number of tests to assess if the MCMC chain is a representative sample from the posterior distribution (i.e., the MCMC algorithm has converged) (Cowles and Carlin 1996). However, even if a chain satisfies a suite of tests, this does not guarantee that convergence has occurred. The three main tests that can be applied to single chains test for autocorrelation, stationarity, and stability. Test statistics can be calculated for these tests to indicate convergence, but in general two very basic visual comparisons can more easily detect lack of convergence. I have used two approaches to assess convergence, the first is a plot of successive samples of the chain (with thinning of every 500th sample). The second is the posterior distributions for the chain split into three even parts (start, middle, and end) (Figure 7.6). I will describe how these can be used when the results are presented.

Prior-based assessment

It is also possible to use an even simpler approach to assess stock risk using only the empirical priors and the observed decline in CPUE. If we believe that the first observed CPUE is from the start of the fishery (i.e., when the population was first exploited), we can use the priors and the equations from above to estimate the probability that a stock is below B_{MSY} at the end of the time period.

By modifying Eq. 7.7 we can estimate the current biomass ratio, B_{END}/B_0 , from the ratio of CPUE (end/initial) and an estimate of β. First we divide each side by the initial abundance,

$$\frac{U_{\text{END}}}{U_0} = \left(\frac{B_{\text{END}}}{B_0}\right)^{\beta},$$

and then re-arrange to get

$$\frac{B_{\text{END}}}{B_0} = \left(\frac{U_{\text{END}}}{U_0}\right)^{1/\beta}. \tag{7.10}$$

This prior-based approach is implemented as follows:

1. Simulate a large number (e.g., 10000) of random values from the prior distributions for β and $\log \hat{\alpha}$.
2. Transform the values of $\log \hat{\alpha}$ to steepness (Eq. 7.2) and then to B_{MSY}/B_0 (Eq. 7.6) (Figure 7.2).
3. Calculate the estimated true decline in abundance for an observed decline in CPUE (Eq. 7.10).
4. Calculate the probability that true abundance is below B_{MSY}, i.e., the number of draws for which the simulated B_{END}/B_0 was less than the simulated B_{MSY}/B_0.

This is possible because the ratio B_{MSY}/B_0 only depends on $\log \hat{\alpha}$ (through steepness). Of course it is not possible to estimate MSY or absolute biomass using this simplistic approach.

7.3 Results

None of the models predicted any real chance (probability < 0.005) of the stock being below B_{MSY} at the end of the time period. Alternatively it is possible to calculate the probability of being below B_{MSY} based on the empirical priors for β and $\log \hat{\alpha}$ (using Eq. 7.6) and the 65% decline in CPUE over the time period. This gives a probability of 51% (Figure 7.3). If it is assumed that CPUE is proportional to abundance, the probability is only 4%. This large difference between the model-based and prior-based estimates is partly explained by the fit of the model to the CPUE data (Figure 7.4).

Table 7.2: Estimated posterior mode and 95% credible region (in parentheses) for important model parameter from each model run: Model 1 (no empirical priors and CPUE assumed proportional); Model 2 (no empirical priors with the relationship between CPUE and abundance estimated); and Model 3 (empirical priors with the relationship between CPUE and abundance estimated).

	$\log\hat{\alpha}$	MSY	B_{MSY}/B_0	$B_{\mathrm{END}}/B_{\mathrm{MSY}}$
Model 1	7.0 (1.9–6.9)	402 (251–431)	0.03 (0.03–0.28)	16.4 (1.6–15.5)
Model 2	7.0 (1.9–6.9)	541 (272–584)	0.03 (0.03–0.28)	21.0 (1.9–19.2)
Model 3	2.9 (1.8–4.4)	277 (235–404)	0.19 (0.10–0.29)	2.2 (1.4– 5.1)

The model does not fit to the high CPUE indices at the start of the model which are important for estimating the decline in CPUE.

While only the fit for Model 3 is presented, the fit for the other two models were visually indistinguishable from that for Model 3. This is quite surprising as I will show that the different models can give quite different results.

Aside from stock risk, stock status (i.e., the abundance of the stock) and sustainable yields are the critical outputs from a stock assessment. I have presented estimates of uncertainty for some of the estimated and derived parameters based on the 95% credible regions from the marginal posterior distributions (Table 7.2).

Models 1 and 2 both had point estimates for $\log\hat{\alpha}$ at the upper bound. There was no advantage in having a higher upper bound as a value of 7 already implies a value of steepness very near 1 (i.e., there is no decline in recruitment at low spawner abundance). The high estimates for $\log\hat{\alpha}$ imply that the target biomass, B_{MSY}, is very low, 3% of the unfished population size. Because of this low target biomass, the stock status (in relation to the target) is very good. When the empirical priors were used, the results were quite different, although the stock status was still good (i.e., $B_{\mathrm{END}} > B_{\mathrm{MSY}}$). Estimates of the productivity parameters (e.g., $\log\hat{\alpha}$ and MSY) were much lower and were less uncertain (tighter credible intervals). The values for all parameters were closer to the mean of their priors.

The reliability of the uncertainty estimates discussed above is based on the model

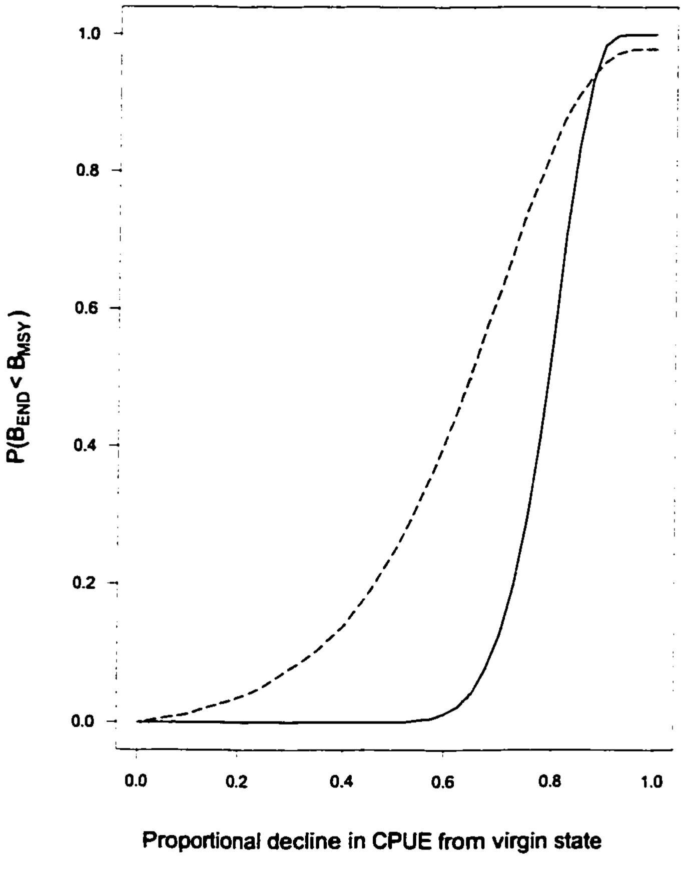


Figure 7.3: Probability that the population is below B_{MSY} for a given decline in CPUE based solely on the empirical priors. The two lines represent the probability assuming that CPUE is proportional to abundance (solid) or integrating over the empirical prior for β (dashed). The vertical dotted line represents the 65% decline observed for hake.

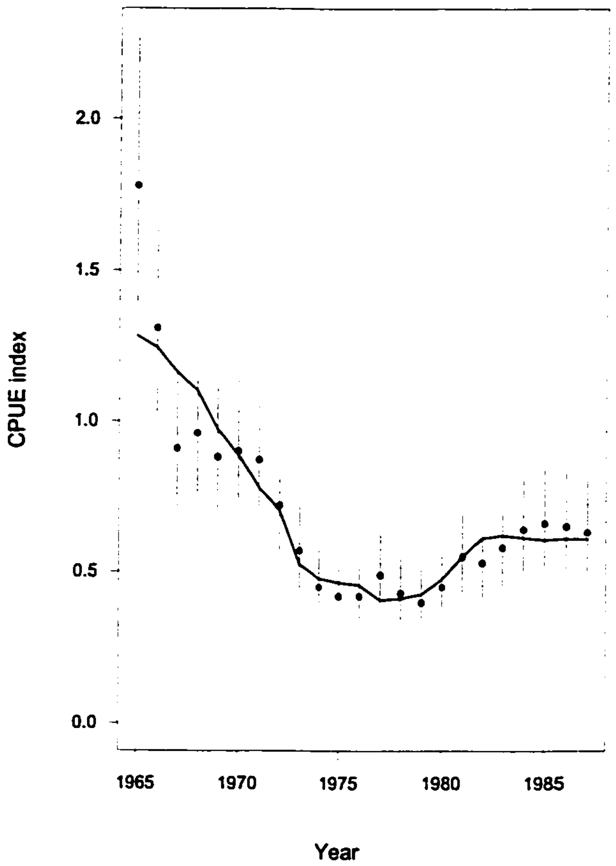


Figure 7.4: Observed (points) and predicted (solid line) CPUE indices from Model 3 (empirical priors). 95% confidence intervals are provided for the observations based on the observation error estimated within the model.

being correct (it is most certainly too simplistic) and the MCMC algorithm converging. I only present diagnostics for $\log \hat{\alpha}$ as it is the most important model parameter (for determining productivity) and the patterns in its samples were representative of those for other parameters.

The traces of MCMC chains for Models 1 and 3 suggest that the MCMC algorithm converged as the times series appear to be stationary with little autocorrelation at the first lag (0.14 and 0.07 respectively) (Figure 7.5). However, there were serious problems (autocorrelation of 0.96 at lag 1 including thinning of 500) for Model 2. The lack of convergence is also illustrated in the split posterior distributions (Figure 7.6). For a stationary MCMC chain with adequate sampling all three density curves should be identical. Overall the density curves were reasonable though only those for Model 3 were near identical.

Another important examination to make for Bayesian analysis is a comparison of the prior and estimated marginal posterior distributions for the parameters of interest. I compared the priors and posteriors for the three parameters for which we had empirically-derived priors, $\log \hat{\alpha}$, β, and σ (Figure 7.7). The priors and posteriors for both $\log \hat{\alpha}$ and β were similar indicating that the data contain little information for these parameters. Conversely, for σ the posterior distribution was very precise compared to the priors suggesting that there was much information for this in the data.

Not surprisingly Model 3 outperformed Model 2, i.e., when the relationship between CPUE and abundance was estimated with and without priors. In particular, the MCMC algorithm had considerable difficulties without the empirical priors. Overall, Model 3 was more conservative in terms of yields, target biomass, and stock status. This is because the limited data suggested that hake were highly resilient to overfishing, something not consistent with the prior information.

Simulation study

While it is clear that the empirical priors improved the performance of the MCMC sampling when all parameters were estimated, my inferences on the improvements in estimating key parameters are based on the assumption that the prior distributions were correct. To confirm my conclusions I used a simulation study. The estimated

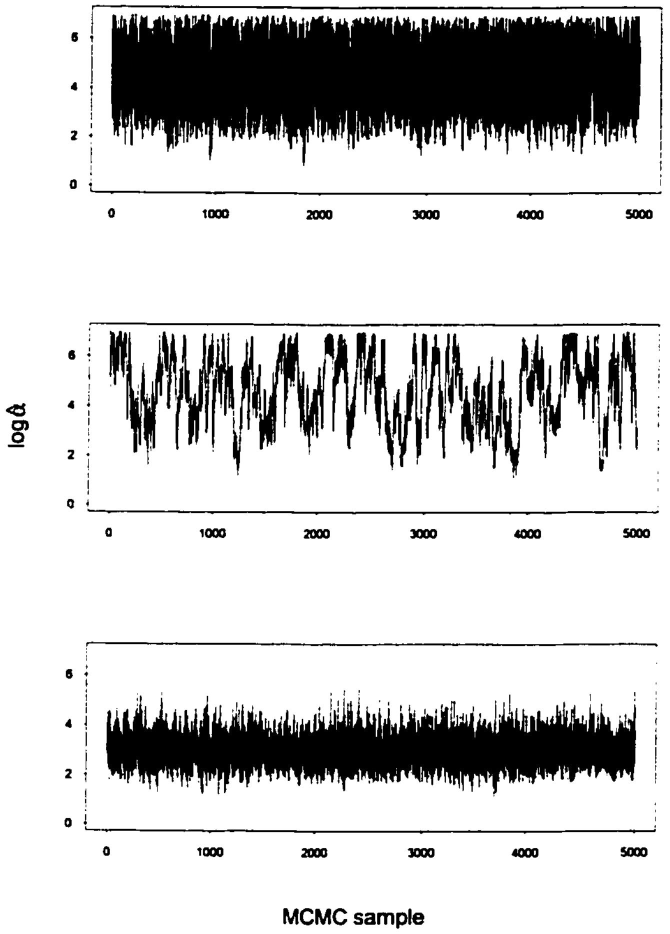


Figure 7.5: Time series of MCMC samples from the marginal posterior distribution for $\log\hat{\alpha}$ for Model 1 (top), 2 (middle), and 3 (bottom).

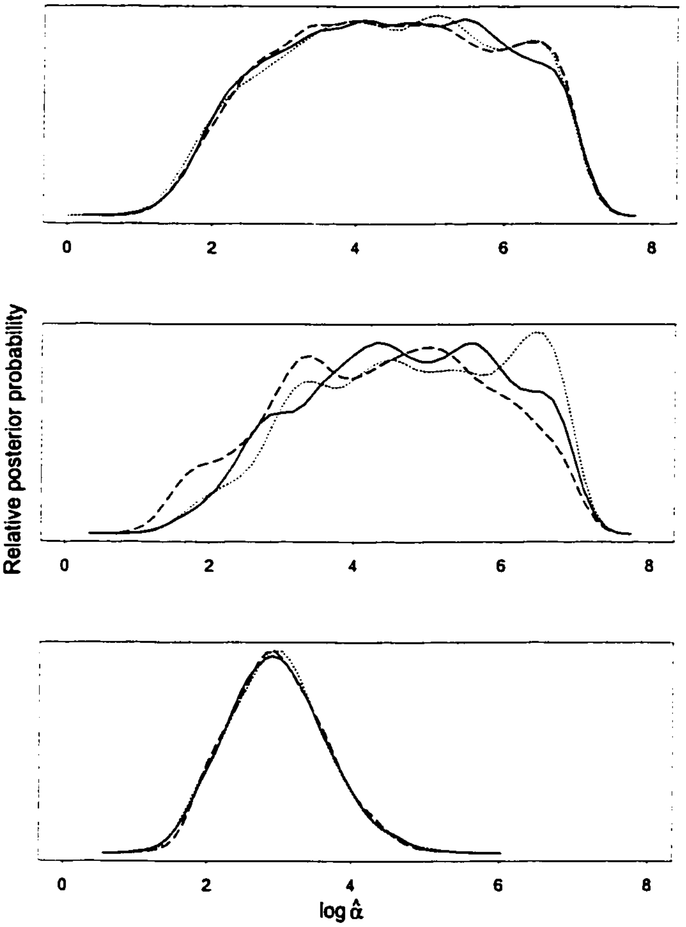


Figure 7.6: Relative posterior probability for $\log\hat{\alpha}$ for the MCMC samples split into three parts (start, middle, and end) for Models 1 (top), 2 (middle), and 3 (bottom) for each species. Convergence is implied if all three density curves overlap.

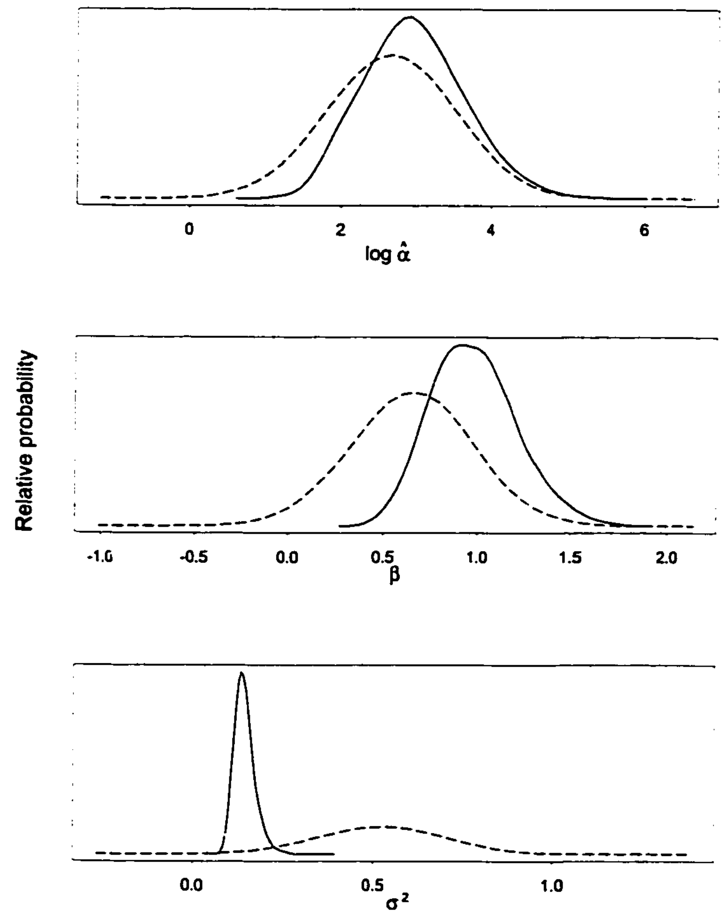


Figure 7.7: Comparison of the empirically-derived prior distributions (dashed line) and the marginal posterior distributions (solid line) for $\log \hat{\alpha}$ (top), β (middle), and the observation error variance σ (bottom) from Model 3 (empirical priors).

observation error variance was used to generate 500 CPUE data sets using the LRSG model with the observed catches and the parameter estimates for Model 3 (where the empirical priors had been used). I estimated the model parameters with these 500 data sets for four scenarios:

- Model S1: No empirical priors and CPUE is assumed to be proportional to abundance (though not in the simulated data).
- Model S2: No empirical priors, β estimated
- Model S3a: Empirical priors from Table 7.1; all parameters estimated.
- Model S3b: Priors with the same standard deviation as the empirical priors and a mean equal to the values used to simulate the data; all parameters estimated.

As there is a difference between the mean of the empirical priors and the point estimates; i.e., the mode of the posterior is different from the mean of the prior distribution (Figure 7.7), the empirical priors used in Model 3a are not consistent with the true (simulated) values. These priors should result in biased estimates. For comparison, the priors used in Model 3b are consistent with the data.

It was clear that without the priors, Models S1 and S2 had considerable difficulty accurately estimating $\log\tilde{\alpha}$, the parameter critical to the estimation of the biomass reference point, B_{MSY} (Figure 7.8 and Table 7.3). There was considerable variation in the estimates of $\log\tilde{\alpha}$ for the simulated data sets, with a large average positive bias (54% and 49% for Models S1 and S2 respectively). This resulted in a two-fold problem in the estimates of yields. There was a large negative bias (around -30%) in B_{MSY} and a moderately large positive bias (around 15%) in MSY. Management decisions based on such biased results would permit catches that are too high, resulting in population sizes that are too low. Interestingly, both B_0 and s were still estimated quite accurately.

The addition of empirical priors reduced the bias in $\log\tilde{\alpha}$ and associated parameter considerably. The 'biased' priors (Model 3a) resulted in moderate negative biases in estimates of β (-23%) and B_0 (-19%), but quite small biases in $\log\tilde{\alpha}$ (2%) and MSY (8%). Using priors that had the same mean and the values used in the simulation

Table 7.3: Average % bias in the estimates of model parameters and derived parameters over the 500 simulated data sets for the models: Model S1 (no empirical priors and CPUE assumed proportional); Model S2 (no empirical priors with the relationship between CPUE and abundance estimated); Model S3a (empirical priors with biased means); and Model S3b (empirical priors).

Model	B_0	$\log\tilde{\alpha}$	s	β	B_{MSY}	MSY	B_{MSY}/B_0
S1	1.5	53.7	-4.3	16.1	-30.5	16.5	-21.2
S2	-0.8	48.9	-3.7	3.9	-31.1	14.7	-24.2
S3a	-18.9	-2.8	-4.9	-23.7	-16.0	-7.7	4.2
S3b	-8.6	1.9	-3.0	-9.5	-9.9	15.5	-0.7

reduced the bias all parameters except MSY. It is not clear why this occurred, though MSY is a complex function of almost all other estimated parameters.

7.4 Discussion

Two of the greatest advantages of meta-analysis are the ability to test for the large-scale generality of processes and to derive estimates of true variability (random effects distributions) that can be used as priors in Bayesian analyses. In this chapter I have focused on this second aspect to show how empirically derived priors can be used in data poor situations. Priors can be critical for stock assessment because decisions always need to be made, but often local data have insufficient information to accurately estimate the important parameters (McAllister and Kirkwood 1998).

Here I have focused on a key parameter for interpreting trends in CPUE and a parameter describing the resilience of a stock to over-fishing. I have demonstrated that these priors improve both the estimation of key parameters and the performance of the Hastings-Metropolis algorithm used in the MCMC sampling.

I have demonstrated that it was possible to get estimates of stock risk, i.e., the probability that the stock was below some target level, using only empirical priors and the observed decline in CPUE. For the example presented here, the results did not agree with those from the models but this does not necessarily imply that it will

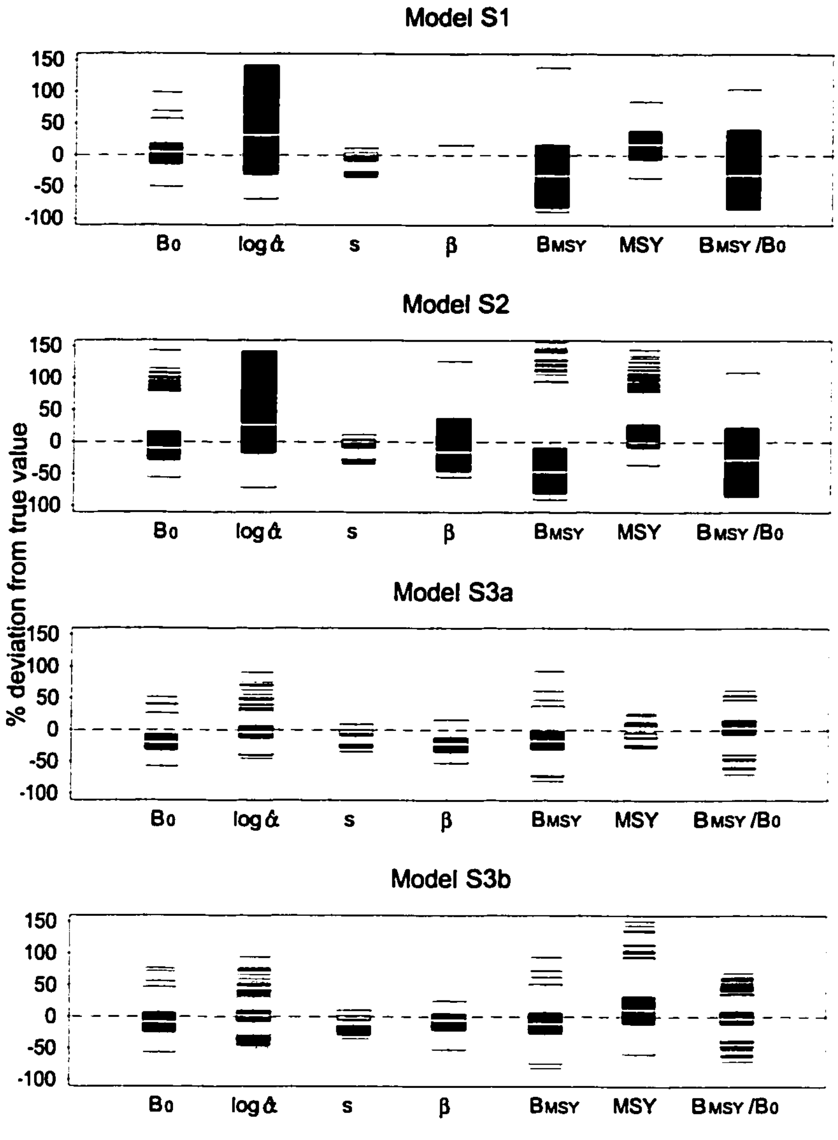


Figure 7.8: Distribution of relative (%) deviations from the true value in the simulations for some of the estimated and derived parameters for each model: Model S1 (no empirical priors and CPUE assumed proportional); Model S2 (no empirical priors with the relationship between CPUE and abundance estimated); Model S3a (empirical priors with biased means); and Model S3b (empirical priors). The upper quartile and lower quartile provide the outline of the box (the line inside the box represents the median). Whiskers are drawn to the nearest value not beyond 1.5*(inter-quartile range) from the quartiles; points beyond are drawn individually as outliers. A dashed horizontal line at zero is given for reference.

work poorly in all situations.

Previous researchers have demonstrated how 'biased' or overly precise priors adversely affect management decisions (Adkison and Peterman 1996; McAllister and Kirkwood 1998). Hopefully, this will not be used a rationale for not using priors (and Bayesian assessment methods). Analysis of both real and simulated data showed that it was very difficult to accurately estimate the critical parameters required for fisheries management without informative prior distributions. The use of simple models and only basic data (i.e., catches and CPUE) can result in parameter estimates that are not reasonable. Also, without auxiliary information on stock size, it is not possible to accurately estimate the relationship between CPUE and abundance. In the example presented here, models without priors tended to greatly over-estimate the productivity of the stock. This is unlikely to not a general finding though – rather a function of the data set being analyzed.

While the Bayesian approach provides the ideal framework for incorporating this prior information there are many difficulties regarding the application of these methods. Firstly, the accuracy of the estimates of uncertainty are dependent on obtaining representative samples from the joint posterior distribution (and of course the assumption that the model is correct). When the full model was estimated (i.e., β plus the other parameters) without the use of empirical priors, the MCMC algorithm performed very poorly. The empirical priors also reduced the uncertainty in the parameter estimates.

There are two other important issues regarding the use of priors in stock assessment: (1) the precision of priors; and (2) independence of priors. McAllister and Kirkwood (1998) showed that overly precise priors can have a negative impact on assessment and management even if the mode of the priors is consistent with the true parameter values. Overly precise priors often lead to overly precise posterior distributions when there is little information in the data for the parameters of interest. This can reduce the ability of assessment scientists to learn about the system (McAllister and Kirkwood 1998). They recommend that priors should have a c.v. of at least 0.5 as to not overly dominate the data. The c.v. of the prior that I used for $\log \hat{\alpha}$ was only 0.31 and since the results appear reasonable, I must argue for a more subjective approach for deciding if a prior is overly precise. When the limited data and model gives estimates which suggest very high productivity, as was the case with hake, a less precise prior would lead to less conservative management.

A second important issue when using prior distributions for multiple parameters is whether the priors are independent. In Chapter 3, I estimated the observation error variance for the CPUE series and the shape parameter simultaneously. Priors for these two parameters should be used together. If I assumed that CPUE was proportional to abundance ($\beta = 1$), I might expect the estimated observation errors to be higher. Including depensation (Myers et al. 1995; Liermann and Hilborn 1997) in a spawner-recruitment curve provides another useful example. I would need to use a bivariate prior distribution for the slope at the origin ($\log \tilde{\alpha}$) and the parameter describing the degree of depensation. Because these two parameters are highly correlated in most data sets, using independently derived priors (e.g., from analyses of Myers et al. 1999 and Liermann and Hilborn 1997) would be inappropriate.

There is no question that the Bayesian approach offers many advantages to the assessment of developing or data-poor fisheries (McAllister and Kirkwood 1998) and empirical priors will reduce many of the arguments surrounding the selection of priors (Hilborn and Liermann 1998). However, there are still many potential problems with empirical priors. In Chapter 2 I highlighted a number of potential areas in which a meta-analysis can go wrong, e.g., biased data or inappropriate inferences. If there is limited data for the species under consideration, data for other species are often used to construct empirical priors. Different selection criteria of 'relevant' species, can lead to different priors (Myers et al. 2002). However, it is still clear that meta-analysis will play an important role in the acceptance and future development of Bayesian approaches to stock assessment.

Chapter 8

Conclusions

The aim of this thesis is to reduce the uncertainty in our understanding of fish population dynamics. In this thesis I have developed meta-analytical methods which, if applied to particular problems, should help reduce this uncertainty by better modelling fisheries data. I have applied these methods here to a number of examples to both illustrate their application and ability answer important ecological and applied questions. Many of the methods used here are very new to the field of fisheries, extending the recent work of Barrowman (2001) and the simple methods generally used for meta-analysis (Cooper and Hedges 1994). In this final chapter I will discuss how the work here has contributed to both our understanding of fish population dynamics and the way in which we do meta-analysis. In Section 8.1 I will review the different meta-analytical methods used (and developed) in this thesis and discuss how these methods may be further extended. In Section 8.2 I will review how the work in each of the preceding chapters has contributed to our understanding of fish population dynamics and how this work should be extended in future studies. Finally, in Section 8.3 I identify key issues and directions for meta-analysis in fisheries.

8.1 Methods for meta-analysis in fisheries

There are two main applications for meta-analytical methods as developed in this thesis: (1) to test large-scale ecological hypotheses or combine results across many studies; and (2) to improve our ability to estimate key parameters by sharing information across studies. The two applications are related and have been used together often such as in the analysis of spawning dynamics of hoki.

In Chapter 2, I provided a comprehensive critical evaluation of meta-analysis in fisheries. I described some of the simple methods of meta-analysis such as vote-counting and the sign-test and the extension of these approaches to more advanced methods of fixed and random effect models. I demonstrated these advanced methods in three contexts: (1) analysis of model output (Chapter 3); (2) simultaneous analysis of multiple data series (Chapters 4 and 5); and (3) integrating meta-analysis into population dynamics models (Chapter 6).

The simple maximum likelihood estimation of the random effects distribution used in the analysis of the relationship between CPUE and abundance and in the hoki analysis provides a basic introduction to the hierarchical models used in meta-analysis. This is a useful approach for hypothesis testing but is very inefficient in terms of helping us improve estimates of key parameter. For example, in the hoki analysis the real quantities of interest were the estimates of spawning biomass for each year rather than residence time or the timing of the arrival to the spawning grounds. The two-step analysis indicated significant interannual variation in arrival time, but this information was not used to obtain better estimates of spawning biomass.

An area where the simple random effects models could be improved, is the ability to explicitly model non-independence among studies, e.g., in the analysis of CPUE in Chapter 3, data on multiple ages for a single stock were used. This does not appear to be a simple problem, but if solved, it will greatly improve the utility of this approach to meta-analysis.

While parameters of linear mixed effects models are simple to estimate, nonlinear mixed effects model provide many problems. Most commonly used methods rely on linear approximations (Barrowman 2001). Here I have used two flexible approaches that approximate the integrals by a series of draws from the distribution of interest. In Chapter 5 I used a Bayesian approach in the analysis of length-specific trawl survey catchability. The analysis was performed using a software package called BUGS (Bayesian Integration Using Gibbs Sampling). This is widely used by applied statisticians and is more commonly being used in fisheries applications (Meyer and

Millar 1999b). While it is widely used, it is very much a 'black-box' where the technical aspects are hidden from the user. As this black-box factor is of great concern, some researchers (e.g., Su et al. (2001)) have developed their own software using the same methods. Also, the version of BUGS that I have used (Classic Bugs) was particularly slow – a factor that may confine it to simple fisheries applications.

An exciting achievement in this thesis was the use of the method of maximum simulated likelihood (MSL). This method, commonly used in econometrics (Gourieroux and Monfort 1991; Mariano et al. 2000) but not previously used in fisheries, opens up significant opportunities in fisheries science. The performance of this method was a little different for the two applications presented here. The analysis of density-dependent mortality (Chapter 4) represents a very challenging statistical problem as it requires estimating observation errors and process noise simultaneously. Although results were robust to different starting values, different random seeds (for the random deviates used in the sampling) provided quite different results for some highly correlated variables. It should be recognized that traditional variance component models also have considerable problems (e.g., variance components estimated to be negative) with such data (Myers and Cadigan 1993b). The hoki analysis using MSL was quites different as all of the observation error variances were assumed known. With as little as 20 simulations the model consistently converged within 2% for different starting values and random deviates. This flexible method offers considerable promise in fisheries applications of meta-analysis.

One limitation of the meta-analytical methods used here was that they may not be robust to outliers (Fournier et al. 1990; Hilborn and Mangel 1997; Chen et al. 2000) and may not produce results that are robust to deviations from model assumptions (Huber 1981). Barrowman (2001) developed robustified algorithms for estimating parameters of nonlinear mixed effects models but they have not been considered for the more complex methods presented here. However, I have used simulation methods to test the methods. Robustification of these methods is not a simple statistical problem and is likely beyond the capabilities of most fisheries scientists. It will be important to work closely with statisticians in the future development of these methods.

8.2 Findings and implications

While the focus has been on developing methods for meta-analysis in this thesis, the examples used have provided interesting findings, some with implications for the way in which we model fish populations.

In Chapter 2 I critically reviewed the use of meta-analysis in fisheries up to and including much of the work described here. I discussed potential problems in the application of meta-analytical techniques, e.g., data biases. I concluded that there are three key aspects to a successful meta-analysis: (1) a well-defined question to be addressed; (2) a suitable metric or parameter to be combined; and (3) inferences that are consistent with data used.

In Chapter 3 I used meta-analysis to test one of the most important assumptions in fisheries, that catch-per-unit-effort (CPUE) indices are proportional to abundance. I demonstrated that CPUE predominantly declines slower than the abundance of the population being fished. If not recognized it will lead to over-estimation of abundance and under-estimation of fishing mortality rates (Crecco and Overholtz 1990). This has serious implications for fisheries management. Meta-analysis was necessary to test this assumption because there was often insufficient data for a single data set to make strong conclusions.

As the models used in the analysis were purely descriptive, it is not possible to determine whether the patterns observed are due to fish or fisher behaviour. However, I did take the opportunity to speculate in Chapter 3. Important future work should be directed at the distribution patterns of fish (Clark 1982; Prince and Hilborn 1998). A meta-analysis of the relationship between concentration indices and total abundance would be very useful.

The original purpose of Chapter 4 was to investigate the time at which the year class strength of a cohort is determined. Unfortunately, I was only able to demonstrate that it is possible to construct the complex model required for this problem, and that methods, borrowed from econometrics, can be used to estimate the parameters of these models. It is not a trivial task to develop models for this problem. Myers and Cadigan (1993a; 1993b) used linear mixed effects models with observation error to examine this problem. The models developed in Chapter 4 extend upon their work in three significant ways: (1) enabling density-dependent mortality and variation in density-independent mortality to be estimated simultaneously for multiple time periods; (2) allowing any type of nonlinear dynamics to be included; and (3) allowing alternative error structures (than gaussian errors) to be used. While the estimation of the key parameters required the use of meta-analytical principles, i.e., sharing information across cohorts, I have not performed a full meta-analysis of data from many populations. This will be an important future step. The two examples shown did confirm early studies of the presence of considerable variation in density-independent mortality before the age that cod are first vulnerable to bottom trawl survey gear, and that density-dependent mortality reduced year class strength during the juvenile stage.

The analysis of fish communities relies heavily on estimates of absolute abundance. This is fundamental to properly evaluate ecosystem impacts of fishing such as the observed 'fishing down the food-webs' (Pauly et al. 1998). In Chapter 5 I demonstrated how meta-analysis could be extended from combining single quantities, to combining information on the shape of the relationship between two variables, i.e., the size of a fish and its trawl survey catchability in my example. I derived a series of length-specific catchability scalars that could be used to translate swept-area abundance estimates from trawl surveys into absolute abundance.

Meta-analysis was useful for this problem for two main reasons: (1) meta-analytical estimates, combined across many studies, are preferable to a single estimate from an assessment that may be unreliable or biased; and (2) it was necessary to get estimates for species for which no data was available – data for similar species were combined for this purpose.

Two interesting findings in Chapter 5 were the seasonal differences in catchability and the difference in catchability between cod and haddock. Future work should examine these more closely as well as investigate alternative models for describing the relationship between catchability and length.

The most important development of the potential for meta-analysis in fisheries was shown in the analysis of spawning dynamics of hoki (*Macruronus novaezelandiae*) in Chapter 6. I demonstrated, using maximum simulated likelihood, that it was possible to integrate a random effects approach into a complex and highly nonlinear population dynamics model. This approach allowed improved estimates of spawning biomass and contributed to our understanding of hoki spawning dynamics and interpretation of the million dollar acoustic surveys. Traditional methods that analyze data for each year independently gave very poor and uncertain estimates compared to the fixed and random effect models. This analysis represents a growing trend toward integrating meta-analytical methods into population dynamics models (Su et al. 2001; Millar and Methot 2002).

The analysis in Chapter 7 was motivated by the recommendations of many researchers to use results from meta-analysis in Bayesian implementations of stock assessment. I demonstrated how the results from meta-analysis could be combined with simple population dynamics models to provide biomass reference points and estimates of stock risk. I showed how simple models, like those likely to be used in developing fisheries, could give extremely biased results of important parameters used to determine sustainable yields and biomass reference points. The use of empirically-derived priors greatly reduced bias and uncertainty in the estimates.

8.3 Future work

Based on the analysis and findings of this thesis, I have identified the following key issues and directions for future work.

Databases

The development and maintenance of large databases is of critical importance to the advancement in fisheries ecology (Hilborn and Liermann 1998). For the meta-analysis of fish population dynamics, FishBase (Froese and Pauly 2000) and Myers' stock recruitment database are probably the two most commonly used databases. Considerable effort will be required to compile estimates of reliability for the biological parameters provided in FishBase and there are two potential limitations to the usefulness of Myers' stock recruitment database in future research. First, much effort will be required to ensure the database is continually maintained as stock assessments are often conducted every 1-2 years. Second, stock recruitment time series from the many assessments that constrain the stock recruitment relationship (perhaps using priors based on analyses using the database) cannot be included.

More traditional meta-analysis of effect sizes from experiments uses the published article as it source. The greatest problem will still be incomplete reporting (Gurevitch et al. 2001), e.g., missing sample sizes and standard errors. This can only be overcome by increased vigilance of journal editors and reviewers. Articles produced by the Ecological Society of America (e.g., Ecology) often have appendices stating how and where the data used in analyses can be obtained (either from the authors or from the ESA's Electronic Data Archive). This will allow for more advanced models to be used for meta-analysis.

Efforts must be made to address those problems with the two most important fisheries databases. Without adequate data the methods that can be applied (and subsequent inferences) will be limited.

Derivation of new metrics for meta-analysis

Hilborn and Liermann (1998) argued that for meta-analysis to be successful, researchers must be on the lookout for additional parameters that can be combined. In Chapter 2 I describe a number of ecologically relevant metrics. Many potentially useful metrics have been combined without estimates of reliability, e.g., mean trophic level (Pauly et al. 1998) and the intrinsic rate of increase (Myers et al. 1997). Unfortunately the statistical properties of many metrics have not been provided and determining these are beyond the capabilities of many ecologists (Gurevitch and Hedges 1999). It is best to standardize metrics so that they can adequately be described by the estimated standard error. For example, both the correlation coefficient and the response ratio must be transformed to achieve this. While we must be on the lookout for new metrics, we must be also prepared to transform or standardize these if required.

Meta-analytical methods for the analysis of life history characteristics

I have not considered the numerous studies of life history characteristics in this thesis but will briefly discuss them here. These studies include Beverton and Holts' (1959) investigation of growth, mortality, and maturation, the analyses of natural mortality by Pauly (1980) and Hoenig (1983), Charnovs' (1993) study of numerous life history metrics, and Froese and Binohlans' (2000) discussion of empirical relationships derived from FishBase.

These are well known to almost all fisheries scientists and all (except for the first) have one common flaw: individual estimates were combined and analyzed while ignoring the reliability of each study. Only Beverton and Holt (1959) acknowledged this problem and made no quantitative predictions. Most researchers have focussed on the relative uncertainty between the x and y-variables used in regression analyses (Charnov 1993; Froese and Binohlan 2000) and ignored the reliability of estimates from the different studies. I believe that this is the reason for the very wide prediction errors (up to one order of magnitude) for many empirical relationships (Roff 1984). Without estimates of reliability it is not possible to separate the true variation from the estimation error. A second problem in some analyses is that multiple estimates from the same population are given the same weight when combined with other estimates (Pauly 1980; Froese and Binohlan 2000).

The problems faced in the analysis of life-history characteristics are similar to those faced in meta-analysis but meta-analytical principles have seldom been applied to the analysis of life-history characters. Further efforts are needed to make researchers more aware of meta-analytical methods such as random effects models and the importance of independence and data weighting.

Areas for potential research

Evidence suggesting slow recovery of overexploited stocks (Hutchings 2000) necessitates a re-evaluation of our views on population dynamics at low population sizes (Myers et al. 1995). Various models have been suggested to explain reduced survival at low population sizes, e.g., meta-population structure (Frank and Brickman 2000), predation forces (Swain and Sinclair 2000), and increased vulnerability to environmental factors (Gilbert 1997; Planque and Frédou 1999). We need to develop models to test these complex hypotheses and meta-analytic approaches offer a potentially powerful tool, but we could be hampered by the lack of quantitative assessments for populations at low levels.

Our knowledge of natural mortality in fish populations is still limited. All analyses of spawner-recruitment data and the subsequent standardization depend upon estimates of natural mortality (Myers et al. 1999). Some inferences may be biased if estimates of natural mortality are incorrect. Future meta-analyses of natural mortality must be conducted carefully to ensure that estimates are derived independent of empirical relationships (Beverton 1992). Careful consideration will also have to be given to within-population variation and possible compensatory changes in natural mortality induced by fishing (Vetter 1988).

A very recent development in fisheries science has been the use of electronic tags to collect information on the habits of tuna and other large pelagics (Schaefer and Fuller 2002). The data are in the form of long times series of depth, location (often with considerable error), and temperature observations. Determination of movements and behaviour will be critical to our management of these species. State-space models used in Chapter 4 are ideal for modelling movement (Sibert et al. 2002) and meta-analytical approaches will be crucial for combining information across multiple tracks for different individuals. This will be an exciting area for future research that should benefit greatly from meta-analysis.

The motivation for the methods developed in Chapter 6 was to allow for the assessment of multiple populations simultaneously. Technical limitations had previously made this difficult, but it has long been recognized that by recognizing similarities among neighbouring populations, it will be possible to reduce uncertainty in key population parameters (Collie and Walters 1991). Now, with the availability of methods (e.g., fully Bayesian or maximum simulated likelihood) and fast and efficient nonlinear modelling software (Otter Research Ltd. 2000), there is considerable room for growth in these types of applications. Harley and Maunder (unpublished manuscript) described an age-structured model that can simultaneously estimate parameters for multiple populations in a meta-analytical framework. Parameters can be either: (1) estimated separately (assumed independent) for each population; (2) fixed effects; or (3) random effects. Parameters that have previously been the subject of two-step meta-analysis (i.e., meta-analysis of individual model output), could now be estimated recognizing possible similarities across populations. Candidate parameters include: the maximum reproductive rate (Myers et al. 1999), carrying capacity per unit area (Myers et al. 2001), trawl survey catchability (Harley and Myers 2001), and depensation (Liermann and Hilborn 1997).

Final remarks

Meta-analysis has the potential to improve our understanding of fisheries ecology and identify patterns not apparent from the examination of single studies (Myers and Mertz 1998; Gurevitch et al. 2001). I have demonstrated this here with several examples using different methods and problems. The methods are particularly useful in fisheries, where data can often be of poor quality, because they share information across studies.

It is important that methods associated with meta-analysis be treated with the same rigor as methods associated with experimental design. Failure to do so will reduce the acceptance of both the methods and the results of the particular study. The results should be replicable based upon the published description of the methods. Selection criteria must also be explicit. Researchers must attempt to model both the uncertainties and the structure (e.g., non-independence) in their data and ensure that their inferences and generalizations are soundly based.

There is still a great need for more data and experiments, without this our methods will be weak. We should also strive to develop methods to directly apply these meta-analytical insights to the assessment of populations where local data are poor (Froese and Binohlan 2000).

Meta-analysis will continue to serve an invaluable role in our quest to understand fish population dynamics. When we combine studies, we generally see that the true variation is less than that observed from individual studies. Once we can identify the true variation we will understand more about the dynamics of exploited fish populations and how to manage them.

These methods allow us to answer the important large-scale questions often asked in ecology and the use of hierarchical models should soon become the standard method for the analysis of population dynamics data.

Bibliography

Adkison, M. D., and Peterman, R. M. 1996. Results of Bayesian methods depend on details of implementation: an example of estimating salmon escapement goals. Fisheries Research **25**: 155–170.

Adkison, M. D., and Su, Z. 2001. A comparison of salmon escapement estimates using a hierarchical Bayesian approach versus separate maximum likelihood estimation of each year's return. Canadian Journal of Fisheries and Aquatic Sciences **58**: 1663–1671.

Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control **AC-19**: 716–723.

Alkseyev, F. 1995. Reproductive cycle of roundnose grenadier, *Coryphaenoides rupestris* (Macrouridae), from the North Atlantic. Journal of Applied Ichthyology **35**: 665–672.

Allen, R., and Punsly, R. 1984. Catch rates as indices of abundance of yellowfin tuna *Thunnus albacares* in the eastern Pacific Ocean. Inter-American Tropical Tuna Commission Bulletin **18**: 301–379.

Annala, J. H., Sullivan, K. J., and O'Brien, C. J. 1999. Report from the Fishery Assessment Plenary, April 1999: stock assessments and yield estimates. Plenary Report, Ministry of Fisheries.

Annala, J. H., Sullivan, K. J., and O'Brien, C. J. 2000. Report from the Fishery Assessment Plenary, May 2000: stock assessments and yield estimates. Plenary Report, Ministry of Fisheries.

Arias, C., and Cox, T. L. 1999. Maximum simulated likelihood: a brief introduction for practitioners. Agriculture & Applied Economics: Staff paper series No. 421, University of Wisconsin-Madison.

Ballara, S., Cordue, P., Livingston, M., O'Driscoll, R., and Kendrick, T. 2002. A review of the 1999–2000 hoki fishery and assessment of hoki stocks for 2001. New Zealand Fisheries Assessment Report 2002/4, Ministry of Fisheries, Wellington, New Zealand.

Bannerot, S. P., and Austin, C. B. 1983. Using frequency distributions of catch per unit effort to measure fish-stock abundance. Transactions of the American Fisheries Society **112**: 608–617.

Baranov, F. I. 1918. On the question of the biological basis of fisheries. Nauchn. Issled. Ikhtiologicheskii Inst. Izv. **1**: 81–128.

Barrowman, N. J. 2001. Nonlinear Mixed Effects Models for Meta-Analysis. Ph. D. thesis, Dalhousie University, Halifax, Nova Scotia, Canada.

Barrowman, N. J., and Myers, R. A. 2000. Still more spawner-recruitment curves: the hockey stick and its generalizations. Canadian Journal of Fisheries and Aquatic Sciences **57**: 665–676.

Beverton, R. J., and Holt, S. J. 1957. On the dynamics of exploited fish populations. MAFF Fisheries Investigations Series II **19**: 533.

Beverton, R. J. H. 1992. Patterns of reproductive strategy parameters in some marine teleost fishes. Journal of Fish Biology **41**(Supplement B): 137–160.

Beverton, R. J. H., and Holt, S. J. 1959. A review of the lifespans and mortality rates of fish in nature, and their relation to growth and other physiological characteristics. *In* CIBA Foundation Colloquia on Ageing 5. *Edited by* G. E. W. Wolstenholme and M. O'Connor. CIBA Foundation, London, pp. 142–174.

Bollen, K. A. 1989. Structural equations with latent variables. John Wiley & Sons, New York.

Bradford, M. J., and Cabana, G. 1997. Interannual variability in survival rates and the causes of recruitment variation. *In* Early Life History and Recruitment in Fish Populations. *Edited by* R. C. Chambers and E. A. Trippel. Chapman and Hall, pp. 469–493.

Brattey, J., Cadigan, N. G., Lilly, G. R., Murphy, E. F., Shelton, P. A., and Stansbury, D. E. 1999. An assessment of the cod stock in NAFO Subdivision 3Ps. Canadian Stock Assessment Secretariat Research Document 99/36, Fisheries and Oceans Canada.

Brett, M. T., and Goldman, C. R. 1996. A meta-analysis of the freshwater trophic cascade. Proceedings of the National Academy of Sciences of the United States of America **93**: 7723–7726.

Brett, M. T., and Goldman, C. R. 1997. Consumer versus resource control in freshwater pelagic food webs. Science **275**: 384–386.

Brodziak, J. K. T., Overholtz, W. J., and Rago, P. J. 2001. Does spawning stock biomass affect recruitment of New England groundfish? Canadian Journal of Fisheries and Aquatic Sciences **58**: 306–318.

Bulman, C. M., Koslow, J. A., and Haskard, K. A. 1999. Estimation of the spawning stock biomass of blue grenadier (*Macruronus novaezelandiae*) (teleostei: Merlucciidae) off western Tasmania based upon the annual egg production method. Marine and Freshwater Research **50**: 197–207.

Bushman, B. J. 1994. Vote-counting procedures in meta-analysis. *In* The Handbook of Research Synthesis. *Edited by* H. Cooper and L. V. Hedges. Russell Sage Foundation, New York, pp. 193–213.

Cadigan, N. 1993. Structural equation models and small sample bias reduction with application to fishery data. Unpublished MSc Thesis, Memorial University of Newfoundland, Newfoundland, Canada.

Cadrin, S. X. 2000a. Cape Cod yellowtail flounder. *In* Assessment of 11 Northeast Groundfish Stocks through 1999. Northeast Fisheries Science Center Reference Document 00-05, pp. 83–98.

Cadrin, S. X. 2000b. Georges Bank yellowtail flounder. *In* Assessment of 11 Northeast Groundfish Stocks through 1999. Northeast Fisheries Science Center Reference Document 00-05, pp. 45–64.

Cadrin, S. X. 2000c. Southern New England yellowtail flounder. *In* Assessment of 11 Northeast Groundfish Stocks through 1999. Northeast Fisheries Science Center Reference Document 00-05, pp. 65–82.

Carlin, B. P., and Louis, T. A. 1996. Bayes and Empirical Bayes Methods for Data Analysis, Volume 69 of *Monographs on Statistics and Applied Probability.* Chapman and Hall, London.

Casey, J. M. 2000. Fish assemblages on the Grand Banks of Newfoundland. Unpublished MSc Thesis, Memorial University of Newfoundland, Newfoundland, Canada.

Casey, J. M., and Myers, R. A. 1998. Diel variation in trawl catchability: Is it as clear as day and night? Canadian Journal of Fisheries and Aquatic Sciences **55**: 2329–2340.

Charnov, E. L. 1993. Life History Invariants: Some Explorations of Symmetry in Evolutionary Ecology. Oxford University Press, Oxford.

Chen, Y., Breen, P. A., and Andrew, N. L. 2000. Impacts of outliers and misspecification of priors on Bayesian fisheries-stock assessment. Canadian Journal of Fisheries and Aquatic Sciences **57**: 2293–2305.

Chouinard, G. A., Sinclair, A., Currie, L., Poirier, G., and Swain, D. 1999. Assessment of cod in the Southern Gulf of St. Lawrence, March 1999. Canadian Stock Assessment Secretariat Research Document 99/23, Fisheries and Oceans Canada.

Clark, C. W. 1982. Concentration profiles and the production and management of marine fisheries. *In* Economic Theory of Natural Resources. *Edited by* W. Eichorn, R. Henn, K. Neumann, and R. W. Shepard. Physica-Verlag, Wurzburg-Wien, pp. 97–112.

Clark, C. W., and Mangel, M. 1979. Aggregation and fishery dynamics: a theoretical study of schooling and the purse seine tuna fisheries. U.S. Fish. Bull. **77**: 317–337.

Collie, J. S., and Walters, C. J. 1991. Adaptive management of spatially replicated groundfish populations. Canadian Journal of Fisheries and Aquatic Sciences **48**: 1273–1284.

Coombs, R. F., and Cordue, P. L. 1995. Evolution of a stock assessment tool: Acoustic surveys of spawning hoki (*Macruronus novaezelandiae*) off the west coast of South Island, New Zealand, 1985–91. New Zealand Journal of Marine and Freshwater Research **29**: 175–194.

Cooper, H., and Hedges, L. V. 1994. The Handbook of Research Synthesis. Russell Sage Foundation, New York.

Cordue, P. L. 2000. MIAEL estimation of biomass and fishery indicators for the 1999 assessment of hoki stocks. New Zealand Fisheries Assessment Report 2000/10, Ministry of Fisheries.

Cordue, P. L., Ballara, S. L., and Harley, S. J. 1999. Acoustic survey of spawning hoki in Cook Strait during winter 1998. New Zealand Fisheries Assessment Research Document 99/21, Ministry of Fisheries.

Cowles, M. K., and Carlin, B. P. 1996. Markov Chain Monte Carlo convergence diagnostics: a comparative review. Journal of the American Statistical Association **91**: 883–904.

Crecco, V., and Overholtz, W. J. 1990. Causes of density-dependent catchability for Georges Bank haddock *Melanogrammus aeglefinus*. Canadian Journal of Fisheries and Aquatic Sciences **47**: 385–394.

Cushing, D. H. 1971. The dependence of recruitment on parent stock in different groups of fishes. Journal du Conseil International pour l'Exploration de la Mer **33**: 340–362.

Darby, C. D., and Flatman, S. 1994. Virtual Population Analysis: Version 3.1 Users Guide. Information Technology Series no, 1, Directorate of Fisheries research, Lowestoft, UK.

Davis, M. B., Murphy, E. F., Stansbury, D., and Shelton, P. A. 1996. An assessment of the cod stock in NAFO Divisions 3NO. Northwest Atlantic Fisheries Organization Scientific Council Report 96/80, Department of Fisheries and Oceans.

de Valpine, P., and Hastings, A. 2002. Fitting population models incorporating process noise and observation error. Ecological Monographs **72**: 57–76.

DerSimonian, R., and Laird, N. 1986. Meta-analysis in clinical trials. Controlled Clinical Trials **7**: 177–188.

Dorn, M. W. 2002. Advice on west coast rockfish harvest rates from Bayesian meta-analysis of stock-recruit relationships. North American Journal of Fisheries Management **22**: 280–300.

Drinkwater, K. F., and Myers, R. A. 1987. Testing predictions of marine fish and shellfish landings from environmental variables. Canadian Journal of Fisheries and Aquatic Sciences **44**: 1568–1573.

Dunn, A., Harley, S. J., Doonan, I. J., and Bull, B. 2000. Calculation and interpretation of catch-per-unit-effort (CPUE) indices. New Zealand Fisheries Assessment Report 2000/1, Ministry of Fisheries, Wellington, New Zealand.

Dunn, A., Horn, P. L., Cordue, P. L., and Kendrick, T. H. 2000. Stock assessment of hake (*Merluccius australis*) for the 1999–2000 fishing year. New Zealand Fisheries Assessment Report 2000/50, Ministry of Fisheries, Wellington, New Zealand.

Edwards, R. L. 1968. Fishery resources of the North Atlantic area. *In* The Future of the Fishing Industry of the United States. *Edited by* D. W. Gilbert. University of Washington, pp. 52–60.

Efron, B. 1996. Empirical Bayes methods for combining likelihoods. Journal of the American Statistical Association **91**: 538–565.

English, K. K., Bocking, R. C., and Irvine, J. R. 1992. A robust procedure for estimating salmon escapement based on the area-under-the-curve method. Canadian Journal of Fisheries and Aquatic Sciences **49**: 1982–1989.

Englund, G., Sarnelle, O., and Cooper, S. D. 1999. The importance of data-selection criteria: meta-analyses of stream predation experiments. Ecology **80**(4): 1132–1141.

Fisher, R. A. 1954. Statistical Methods for Research Workers (12th ed.). Oliver and Boyd, Edinburgh.

Fournier, D. A., Sibert, J. R., Majkowski, J., and Hampton, J. 1990. MULTIFAN a likelihood-based method for estimating growth parameters and age composition from multiple length frequency data setd illustrated using data for southern bluefin tuna (*Thunnus maccoyii*). Canadian Journal of Fisheries and Aquatic Sciences **47**: 301–307.

Francis, R. I. C. C. 1997. Comment: How should fisheries scientists and managers react to uncertainty about stock-recruit relationships. Canadian Journal of Fisheries and Aquatic Sciences **54**: 982–983.

Francis, R. I. C. C. 1999. Moving towards bmsy. New Zealand Fisheries Assessment Research Document 99/2, Ministry of Fisheries.

Francis, R. I. C. C., Hurst, R. J., and Renwick, J. A. 2001. An evaluation of catchability assumptions in New Zealand stock assessments. New Zealand Fisheries Assessment Report 2001/1, Ministry of Fisheries, Wellington (New Zealand).

Frank, K. T., and Brickman, D. 2000. Allee effects and compensatory dynamics within a stock complex. Canadian Journal of Fisheries and Aquatic Sciences **57**: 513–517.

Frank, K. T., Mohn, R. K., and Simon, J. E. 1997. Assessment of 4TVW haddock in 1996. Canadian Stock Assessment Secretariat Research Document 97/107, Fisheries and Oceans Canada.

Fretwell, S., and Lucas, H. 1970. On the territorial behavior and other factors influencing the distribution in birds. I. Theoretical development. Acta Biotheoretica **19**: 16–36.

Froese, R., and Binohlan, C. 2000. Empirical relationships to estimate asymptotic length, length at first maturity and length at maximum yield per recruit in fishes, with a simple method to evaluate length frequency data. Journal of Fish Biology **56**: 758–773.

Froese, R., and Pauly, D. 2000. FishBase 2000: concepts, design and data sources. ICLARM, Los Baños, Languna, Phillipines.

Fuller, W. A. 1987. Measurement Error Models. John Wiley & Sons, New York.

Gavaris, S., and Van Eeckhaute, L. 1998. Assessment of haddock on Eastern Georges Bank. Canadian Stock Assessment Secretariat Research Document 99/72, Fisheries and Oceans Canada.

Gilbert, D. J. 1997. Towards a new recruitment paradigm for fish stocks. Canadian Journal of Fisheries and Aquatic Sciences **54**: 969–977.

Gilks, W. R., Thomas, A., and Spiegelhalter, D. J. 1994. A language and program for complex Bayesian modelling. The Statistician **43**: 169–178.

Gillis, D. M., and Peterman, R. M. 1998. Implications of interference among fishing vessels and the ideal free distribution to the interpretation of CPUE. Canadian Journal of Fisheries and Aquatic Sciences **55**: 37–46.

Godø, O. R., and Walsh, S. J. 1992. Escapement of fish during bottom trawl sampling — implications for resource assessment. Fisheries Research **13**: 281–292.

Godø, O. R., Walsh, S. J., and Engås, A. 1999. Investigating density-dependent catchability in bottom-trawl surveys. ICES Journal of Marine Science **56**: 292–298.

Gourieroux, C., and Monfort, A. 1991. Simulation-based inference in models with heterogeneity. Annales D'Economie et de Statistique N 20/21.

Greenhouse, J. B., and Iyengar, S. 1994. Sensitivity analysis and diagnostics. *In* The Handbook of Research Synthesis. *Edited by* H. Cooper and L. V. Hedges. Russell Sage Foundation, New York, pp. 41–55.

Gunn, J. S., Bruce, B. D., Furlani, D. M., Thresher, R. E., and Blaber, S. J. M. 1989. Timing and location of spawning blue grenadier (*Macruronus novaezelandiae*) (teleostei: Merlucciidae) in Australian coastal waters. Australian Journal of Marine and Freshwater Research **40**: 97–112.

Gurevitch, J., Curtis, P. S., and Jones, M. H. 2001. Meta-analysis in ecology. *In* Advances in Ecological Research. Volume 32. *Edited by* H. Caswell. Academic press, pp. 199–247.

Gurevitch, J., and Hedges, L. V. 1999. Statistical issues in ecological meta-analyses. Ecology **80**(4): 1142–1149.

Hajivassiliou, V. 1999. Some practical issues in maximum simulated likelihood. *In* Simulation-based inference in econometrics. *Edited by* R. Mariano, T. Schuermann, and M. J. Weeks. Cambridge University Press, pp. 9–37.

Hardy, R. J., and Thompson, S. G. 1996. A likelihood approach to meta-analysis with random effects. Statistics in Medicine **15**: 619–629.

Harley, S. J. 2002. Estimation of residence time for hoki (*Macruronus novaezelandiae*) on the Cook Strait and west coast South Island spawning grounds. New Zealand Fisheries Assessment Report 2002/1, Ministry of Fisheries, Wellington, New Zealand.

Harley, S. J., and Maunder, M. N. A hierarchical multi-stock age-structured model for meta-analysis. Unpublished.

Harley, S. J., and Myers, R. A. 2001. Hierarchical Bayesian models of length specific catchability of research trawl surveys. Canadian Journal of Fisheries and Aquatic Sciences **58**: 1569–1584.

Harley, S. J., Myers, R. A., Barrowman, N. J., Bowen, K. G., and Amiro, R. E. 2001. Estimation of research trawl survey catchability for biomass reconstruction of the eastern Scotian Shelf. Canadian Stock Assessment Secretariat Research Document 2001/084, Fisheries and Oceans Canada.

Harley, S. J., Myers, R. A., and Dunn, A. 2001. A meta-analysis of the relationship between catch-per-unit-effort and abundance. Canadian Journal of Fisheries and Aquatic Sciences **58**: 1705–1772.

Harvey, A. C. 1989. Forecasting, structural time series models and the Kalman filter. Cambridge University Press, New York.

Harvey, P. H., and Page, M. D. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford.

Hassell, M. P., Latto, J., and May, J. M. 1989. Seeing the wood for the trees: detecting density dependence for existing life-table studies. Journal of Animal Ecology **58**: 883–892.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**: 97–109.

Hedges, L. V., Gurevitch, J., and Curtis, P. S. 1999. The meta-analysis of response ratios in experimental ecology. Ecology **80**(4): 1150–1156.

Hedges, L. V., and Olkin, I. 1985. Statistical Methods for Meta-analysis. Academic Press, San Diego.

Hilborn, R., Bue, B. G., and Sharr, S. 1999. Estimating spawning escapements from periodic counts: a comparison of methods. Canadian Journal of Fisheries and Aquatic Sciences **56**: 888–896.

Hilborn, R., and Liermann, M. 1998. Standing on the shoulders of giants: learning from experience in fisheries. Reviews in Fish Biology and Fisheries **8**: 1–11.

Hilborn, R., and Mangel, M. 1997. The Ecological Detective. Princeton University Press, Princeton, New Jersey.

Hilborn, R., and Walters, C. J. 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty. Chapman and Hall, New York.

Hinrichsen, R. A. 2001. High variability in spawner-recruitment data hampers learning. Canadian Journal of Fisheries and Aquatic Sciences **58**: 769–776.

Hjort, J. 1914. Fluctuations in the great fisheries of Northern Europe. Rapports et Procès-Verbaux des Réunions du Conseil International pour l'Exploration de la Mer **20**: 1–228.

Hoenig, J. M. 1983. Empirical use of longevity data to estimate mortality rates. Fishery Bulletin **82**: 898–903.

Horn, P. L., Harley, S. J., Ballara, S., and Dean, H. 2000. Stock assessment of ling (*Genypterus blacodes*) around the South Island (fishstocks LIN 3, 4, 5, 6, and 7). New Zealand Fisheries Assessment Report 2000/37, Ministry of Fisheries, Wellington, New Zealand.

Huber, P. J. 1981. Robust statistics. Wiley, New York.

Hurley, P., Black, G., Comeau, P., and Mohn, R. K. 1999. Assessment of 4X haddock in 1998 and the first half of 1999. Canadian Stock Assessment Secretariat Research Document 99/147, Fisheries and Oceans Canada.

Hutchings, J. A. 1996. Spatial and temporal variation in the density of northern cod and a review of hypotheses for the stock's collapse. Canadian Journal of Fisheries and Aquatic Sciences **53**: 943–962.

Hutchings, J. A. 2000. Collapse and recovery of marine fishes. Nature **406**: 882–885.

Hutchings, J. A., and Myers, R. A. 1993. Effect of age on the seasonality of maturation and spawning of Atlantic cod, *Gadus morhua*, in the Northwest Atlantic. Canadian Journal of Fisheries and Aquatic Sciences **50**: 2468–2474.

Hutchings, J. A., and Myers, R. A. 1994a. Timing of cod reproduction: interanual variability and the influence of temperature. Marine Ecology Progress Series **108**: 21–31.

Hutchings, J. A., and Myers, R. A. 1994b. What can be learned from the collapse of a renewable resource? Atlantic cod, *Gadus morhua*, of Newfoundland and Labrador. Canadian Journal of Fisheries and Aquatic Sciences **51**: 2126–2146.

Ianelli, J., Zimmerman, M., and Harley, S. J. 2000. Status and future prospects for the pacific ocean perch resource in waters off Washington and Oregon as assessed in 2000. *In* Status of the Pacific coast groundfish fishery through 2000. Appendix D. Pacific Fishery Management Council, 2130 SW Fifth Ave., suite 224,Portland, Oregon, 97201.

ICES 1999. Report to the working group on the assessment of demersal stocks in the North Sea and Skagerrak. ICES CM 7, ICES.

Kearney, R. E. 1991. Extremes in fish biology, population dynamics, and fisheries management: Pacific skipjack and southern bluefin tuna. Reviews in Aquatic Science **4**(2-3): 289–298.

Kehler, D. G. 2001. Bias in spawner-recruitment models. Unpublished MSc Thesis, Dalhousie University, Halifax, Nova Scotia, Canada.

Kendall, M., Stuart, A., and Ord, J. K. 1987. The advanced theory of statistics. Vol. 1 (Fifth ed.). Macmillan Inc., New York.

Kiefer, J., and Wolfowitz, J. 1956. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Ann. Math. Statist. **27**: 887–906.

Kitigawa, G. 1987. Non-Gaussian state-space modeling of nonstationary time series (with discussion). Journal of the American Statistical Association **82**: 1031–1063.

Kjesbu, O. S., Solemdal, P., Bratland, P., and Fonn, M. 1996. Variation in annual egg production in individual captive Atlantic cod (*Gadus morhua*). Canadian Journal of Fisheries and Aquatic Sciences **53**: 610–620.

Langley, A. 1993. Spawning dynamics of hoki in the Hokitika Canyon. N.Z. Fish. Tech. Rep. 34, Ministry of Agriculture and Fisheries, Wellington, New Zealand.

Liermann, M., and Hilborn, R. 1997. Depensation in fish stocks: A hierarchic Bayesian meta-analysis. Canadian Journal of Fisheries and Aquatic Sciences **54**: 1976–1985.

Lindstrom, M. J., and Bates, D. M. 1990. Nonlinear mixed effects models for repeated measures data. Biometrics **46**: 673–687.

Livingston, M. 1990a. Spawning hoki (*Macruronus novaezelandiae*) concentrations in Cook Strait and off the east coast of the South Island, New Zealand, August–September 1987. New Zealand Journal of Marine and Freshwater Research **24**: 503–517.

Livingston, M., Bull, B., and Gilbert, D. 2000. The observed sex ratios of hoki (*Macruronus novaezelandiae*) in New Zealand, 1983–99. New Zealand Fisheries Assessment Report 2000/24, Ministry of Fisheries, Wellington, New Zealand.

Livingston, M., Vignaux, M., and Schofield, K. 1997. Estimating the annual proportion of nonspawning adults in New Zealand hoki, *Macruronus novaezelandiae*. Fishery Bulletin **95**: 99–113.

Livingston, M. E. 1990b. Stock structure of New Zealand hoki, *Macruronus novaezelandiae*. New Zealand Fisheries Assessment Research Document 90/8, Ministry of Agriculture and Fisheries, Wellington, New Zealand.

Lo, N., Jacobson, L., and Squire, J. 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Canadian Journal of Fisheries and Aquatic Sciences **49**: 2515–2526.

MacCall, A. D. 1990. Dynamic geography of marine fish populations. University of Washington Press, Seattle.

Mace, P. M., and Doonan, I. J. 1988. A generalized bioeconomic simulation model for fish population dynamics. New Zealand Fisheries Assessment Research Document **88/4**.

Main, J., and Sangster, G. I. 1981. A study of the fish capture process in a bottom trawl by direct observations from a towed underwater vehicle. Scottish Fisheries Research Report 23, Depart of Agriculture and Fisheries for Scotland.

Manly, B. F. 1990. Stage-structured Populations. Chapman and Hall, London.

Mariano, R., Schuermann, T., and Weeks, M. J. 2000. Simulation-based inference in econometrics. Cambridge University Press, Cambridge. Methods and applications.

Marin, R., Chouinard, G. A., Forest-Gallant, I., and Piorier, G. A. 1998. Assessment of NAFO Division 4T Americain plaice in 1996 and 1997. Canadian Stock Assessment Secretariat Research Document 98/06, Fisheries and Oceans Canada.

Marteinsdottir, G., and Bjornsson, H. 1999. Time and duration of spawning cod in Icelandic waters. ICES CM.

Mayo, R. 1998. Assessment of the Gulf of Maine Atlantic cod stock for 1997. NEFSC Ref. Document 98/08, NMFS.

McAllister, M. K., and Kirkwood, G. P. 1998. Using Bayesian decision analysis to help achieve a precautionary approach for managing developing fisheries. Canadian Journal of Fisheries and Aquatic Sciences **55**: 2642–2661.

Mertz, G., and Myers, R. A. 1995. Estimating the predictability of recruitment. Fishery Bulletin **93**: 657–665.

Meyer, R., and Millar, R. B. 1999a. Bayesian stock assessment using a state-space implementation of the delay difference model. Canadian Journal of Fisheries and Aquatic Sciences **56**: 37–52.

Meyer, R., and Millar, R. B. 1999b. BUGS in Bayesian stock assessments. Canadian Journal of Fisheries and Aquatic Sciences **56**: 1078–1086.

Micheli, F. 1999. Eutrophication, fisheries, and consumer-resource dynamics in marine pelagic ecosystems. Science **285**: 1396–1398.

Millar, R. B. 1995. The functional form of hook and gillnet selection curves cannot be determined from comparative catch data alone. Canadian Journal of Fisheries and Aquatic Sciences **52**: 883–889.

Millar, R. B. 2001. Reference priors for Bayesian fisheries models. Canadian Journal of Fisheries and Aquatic Sciences.

Millar, R. B., and Methot, R. D. 2002. Age-structured meta-analysis of U.S. west coast rockfish (Scorpaenidae) populations and hierarchical modeling of trawl survey catchabilities. Canadian Journal of Fisheries and Aquatic Sciences **59**: 383–392.

Mohn, R. K., Fanning, C. P., and MacEachern, W. J. 1998. Assessment of 4VsW cod in 1997 incorporating additional sources of mortality. Canadian Stock Assessment Secretariat Research Document 98/78, Fisheries and Oceans Canada.

Mohn, R. K., Lambert, T., Wilson, S., and Black, G. 1998. Update of the status of 4Vn cod: 1996. Canadian Stock Assessment Secretariat Research Document 97/67, Fisheries and Oceans Canada.

Morgan, M. J., Brodie, W. B., and Bowering, W. R. 1999. An assessment American plaice in NAFO Divisions 3LNO. Northwest Atlantic Fisheries Organization Scientific Council Report 99/40, Department of Fisheries and Oceans.

Morgan, M. J., and Trippel, E. A. 1996. Skewed sex ratios in spawning shoals of Atlantic cod (*Gadus morhua*). ICES Journal of Marine Science **53**: 820–826.

Myers, R. A. 1997. Comment and reanalysis: Paradigms for recruitment studies. Canadian Journal of Fisheries and Aquatic Sciences **54**: 978–981.

Myers, R. A. 1998. When do environment-recruit correlations work? Reviews in Fish Biology and Fisheries **8**: 285–305.

Myers, R. A., and Barrowman, N. J. 1996. Is fish recruitment related to spawner abundance? Fishery Bulletin **94**: 707–724.

Myers, R. A., Barrowman, N. J., Hilborn, R., and Kehler, D. G. 2002. Inferring the Bayes priors with limited direct data with applications for risk analysis and reference points. North American Journal of Fisheries Management **22**: 351–364.

Myers, R. A., Barrowman, N. J., Hutchings, J. A., and Rosenberg, A. A. 1995. Population dynamics of exploited fish stocks at low population levels. Science **269**: 1106–1108.

Myers, R. A., Bowen, K. G., and Barrowman, N. J. 1999. The maximum reproductive rate of fish at low population sizes. Canadian Journal of Fisheries and Aquatic Sciences **56**: 2404–2419.

Myers, R. A., and Bowen, W. D. 1989. Estimating bias in aerial surveys of harp seal pup production. J. Wildl. Manage. **53**: 361–372.

Myers, R. A., Bradford, M. J., Bridson, J. M., and Mertz, G. 1997. Estimating delayed density-dependent mortality in sockeye salmon, (*Oncorhynchus nerka*); a meta-analytic approach. Canadian Journal of Fisheries and Aquatic Sciences **54**: 2449–2463.

Myers, R. A., and Cadigan, N. G. 1993a. Density-dependent juvenile mortality in marine demersal fish. Canadian Journal of Fisheries and Aquatic Sciences **50**: 1576–1590.

Myers, R. A., and Cadigan, N. G. 1993b. Is juvenile natural mortality in marine demersal fish variable? Canadian Journal of Fisheries and Aquatic Sciences **50**: 1591–1598.

Myers, R. A., and Cadigan, N. G. 1995. Was an increase in natural mortality responsible for the collapse of northern cod? Canadian Journal of Fisheries and Aquatic Sciences **52**: 1274–1285.

Myers, R. A., and Drinkwater, K. 1988. Ekman transport and larval fish survival in the Northwest Atlantic. Biol. Oceanography **6**: 45–64.

Myers, R. A., and Drinkwater, K. F. 1989. The influence of Gulf Stream warm core rings on recruitment of fish in the Northwest Atlantic. J. Mar. Res. **47**: 635–656.

Myers, R. A., Hammill, M. O., and Stenson, G. B. 1997. Using mark-recapture to estimate the numbers of a migrating stage-structured population. Canadian Journal of Fisheries and Aquatic Sciences **54**: 2097–2105.

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

Myers, R. A., MacKenzie, B. R., Bowen, K. G., and Barrowman, N. J. 2001. What is the carrying capacity for fish in the ocean? a meta-analysis of population dynamics of North Atlantic cod. Canadian Journal of Fisheries and Aquatic Sciences **58**: 1464–1476.

Myers, R. A., and Mertz, G. 1998. Reducing uncertainty in the biological basis of fisheries management by meta-analysis of data from many populations; A synthesis. Fisheries Research **37**: 51–61.

Myers, R. A., Mertz, G., and Fowlow, P. S. 1997. Maximum population growth rates and recovery times for Atlantic cod, *Gadus morhua*. Fishery Bulletin **95**: 762–772.

O'Boyle, R. 2000. Proceedings of a workshop on the Ecosystem Considerations for the Eastern Scotian Shelf Integrated Management (ESSIM) area. Canadian Stock Assessment Secretariat Proceedings Document 2000/014, Fisheries and Oceans Canada.

O'Brien, L., and Cadrin, S. X. 1999. Assessment of the Georges Bank Atlantic cod stock for 1998. NEFSC Ref. Document 99/03, NMFS.

O'Brien, L., Mayo, R. K., and Esteves, C. 1999. Assessment of Amercian plaice in the Gulf of Maine - Georges Bank region for 1998. NEFSC Ref. Document 99/09, NMFS.

Orwin, R. G. 1994. Systematic coding. *In* The Handbook of Research Synthesis. *Edited by* H. Cooper and L. V. Hedges. Russell Sage Foundation, New York, pp. 139–162.

Osenberg, C. W., Sarnelle, O., Cooper, S. D., and Holt, R. D. 1999. Resolving ecological questions through meta-analysis: goals, metrics and models. Ecology **80**(4): 1105–1117.

Otter Research Ltd. 2000. An introduction to AD Model builder Version 4: for use in nonlinear modeling and statistics. Otter Research Ltd.

Paloheimo, J. E., and Dickie, L. M. 1964. Abundance and fishing success. Journal du Conseil International pour l'Exploration de la Mer **155**: 152–163.

Pauly, D. 1980. On the interrelationships between natural mortality, growth parameters and mean environmental temperature in 175 fish stocks. Journal du Conseil International pour l'Exploration de la Mer **39**: 175–193.

Pauly, D. 1998. Tropical fishes: patterns and propensities. Journal of Fish Biology **53**(Supplement A): 1–17.

Pauly, D., Christensen, V., Dalsgaard, J., Forese, R., and Torres, F. 1998. Fishing down marine food webs. Science **279**: 860–862.

Peterman, R. M., and Steer, G. J. 1981. Relation between sport-fishing catchability coefficients and salmon abundance. Transactions of the American Fisheries Society **110**: 585–593.

Pinheiro, J. C., and Bates, D. M. 2000. Mixed-effects models in S and S-PLUS. Springer-Verlag, New York.

Planque, B., and Frédou, T. 1999. Temperature and recruitment of Atlantic cod (*Gadus morhua*). Canadian Journal of Fisheries and Aquatic Sciences **56**: 2069–2077.

Prince, J., and Hilborn, R. 1998. Concentration profiles and invertebrate fisheries management. Canadian Special Publication of Fisheries and Aquatic Sciences **125**: 187–196.

Punt, A., and Hilborn, R. 1997. Fisheries stock assessment and decision analysis: The Bayesian approach. Reviews in Fish Biology and Fisheries **7**: 35–65.

Quinn, T. J. 1985. Catch-per-unit-effort: a statistical model for Pacific halibut (*Hippoglossus stenolepsis*). Canadian Journal of Fisheries and Aquatic Sciences **42**: 1423–1429.

Quinn, T. J., and Collie, J. S. 1990. Alternative population models for eastern Bering Sea pollock. International North Pacific Fisheries Commission Bulletin **50**: 243–257.

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

Reed, J. G., and Baxter, P. M. 1994. Scientific communication and literature retrieval. *In* The Handbook of Research Synthesis. *Edited by* H. Cooper and L. V. Hedges. Russell Sage Foundation, New York, pp. 57–70.

Richards, L. J., and Schnute, J. T. 1986. An experimental and statistical approach to the question: Is CPUE an index of abundance? Canadian Journal of Fisheries and Aquatic Sciences **43**: 1214–1227.

Ricker, W. E. 1954. Stock and recruitment. Journal of the Fisheries Resource Board of Canada **11**: 559–623.

Ricker, W. E. 1973. Two mechanisms that make it impossible to maintain peak-period yields from stocks of Pacific salmon and other fishes. Journal of the Fisheries Resource Board of Canada **30**: 1275–1286.

Riessen, H. P. 1999. Predator-induced life history shifts in *Daphnia*: a synthesis of studies using meta-analysis. Canadian Journal of Fisheries and Aquatic Sciences **56**: 2487–2494.

Robinson, G. K. 1991. That BLUP is a good thing: the estimation of random effects. Statist. Sci. **6**(1): 15–51. With comments and a rejoinder by the author.

Roff, D. A. 1984. The evolution of life history parameters in teleosts. Canadian Journal of Fisheries and Aquatic Sciences **41**: 989–1000.

Rose, G. A., and Leggett, W. C. 1991. Effects of biomass-range interactions on catchability of migratory demersal fish by mobile fisheries: an example of Atlantic cod (*Gadus morhua*). Canadian Journal of Fisheries and Aquatic Sciences **48**: 843–848.

Rosenthal, R. 1994. Parametric measures of effect size. *In* The Handbook of Research Synthesis. *Edited by* H. Cooper and L. V. Hedges. Russell Sage Foundation, New York, pp. 231–244.

S-PLUS 1999. S-Plus 2000 programmers guide. Seattle, Data analysis products division, MathSoft.

Sampson, D. B. 1990. Local catch per unit effort as an index of global abundance. *In* Catch effort sampling strategies: their application in freshwater fisheries management. *Edited by* I. G. Cowx. Fishing News Books, pp. 275–284.

SAS Insitute Inc. 1990. SAS Technical Report P-200, *SAS/STAT* Software: CALIS and LOGISTIC Procedures, Release 6.04. Cary, NC, SAS Insitute Inc.

Schaefer, K. M., and Fuller, D. W. 2002. Movements, behavior, and habitat selection of bigeye tuna (*Thunnus obesus*) in the eastern equatorial Pacific, ascertained through archival tags. Fishery Bulletin **100**: in press.

Schill, D. J., and Scarpella, R. L. 1997. Barbed hook restrictions in catch-and-release trout fisheries: a social issue. North American Journal of Fisheries Management **17**: 873–881.

Schofield, K., and Livingston, M. 1998. Ovarian development and the potential annual fecundity of western stock hoki (*Macruronus novaezelandiae*). New Zealand Journal of Marine and Freshwater Research **32**: 147–157.

Searle, S. R., Casella, G., and McCulloch, C. E. 1992. Variance Components. John Wiley and Sons, New York, NY.

Seber, G. A. F. 1982. The estimation of animal abundance and related parameters. MacMillan Publishing, New York.

Shadish, W. R., and Haddock, C. K. 1994. Combining estimates of effect size. *In* The Handbook of Research Synthesis. *Edited by* H. Cooper and L. V. Hedges. Russell Sage Foundation, New York, pp. 193–213.

Shelton, P. A., Stansbury, D. E., Murphy, E. F., Lilly, G. R., and Brattey, J. 1996. An assessment of the cod stock in NAFO Divisions 2J+3KL. Northwest Atlantic Fisheries Organization Scientific Council Report 96/62, Department of Fisheries and Oceans.

Shepherd, J. G. 1997. Prediction of year-class strength by calibration regression analysis of multiple recruit index series. ICES Journal of Marine Science **54**: 741–752.

Showell, M. 1998. Assessment of the Scotian Shelf silver hake population in 1997, with projection of yield to 1999. Canadian Stock Assessment Secretariat Research Document 98/141, Fisheries and Oceans Canada.

Sibert, J. R., Musyl, M. K., and Brill, R. W. 2002. Horizontal movements of bigeye tuna near Hawaii determined from Kalman filter analysis of archival tagging data. Fisheries Oceanography: in press.

Sokal, R. R., and Rohlf, F. J. 1995. Biometry (3rd ed.). Freeman, New York.

Somerton, D., Ianelli, J., Walsh, S., Smith, S., Godø, O. R., and Ramm, D. 1999. Incorporating experimentally dervied estimates of survey trawl efficiency into the stock assessment process: a discussion. ICES Journal of Marine Science **56**: 299–302.

Sparholt, H. 1990. An estimate of the total biomass of fish in the North Sea. Journal du Conseil International pour l'Exploration de la Mer **46**: 200–210.

Spiegelhalter, D. J., Thomas, A., Best, N. G., and Gilks, W. R. 1995. B.U.G.S.: Bayesian inference using Gibbs sampling, Version 0.5. MRC Biostatistics Unit.

Stock, W. A. 1994. Systematic coding. *In* The Handbook of Research Synthesis. *Edited by* H. Cooper and L. V. Hedges. Russell Sage Foundation, New York, pp. 125–138.

Su, Z., Adkison, M. D., and Van Alen, B. W. 2001. A hierarchical Bayesian model for estimating historical salmon escapement. Canadian Journal of Fisheries and Aquatic Sciences **58**: 1648–1662.

Suuronen, P., and Millar, R. B. 1992. Size selectivity of diamond and square mesh codends in pelagic herring trawls: only small herrings will notice the difference. Canadian Journal of Fisheries and Aquatic Sciences **49**: 2104–2117.

Swain, D. P., and Sinclair, A. F. 1994. Fish distribution and catchability: What is the appropriate measure of distribution? Canadian Journal of Fisheries and Aquatic Sciences **51**: 1046–1054.

Swain, D. P., and Sinclair, A. F. 2000. Pelagic fishes and the cod recruitment dilemma in the Northwest Atlantic. Canadian Journal of Fisheries and Aquatic Sciences **57**: 1321–1325.

Taylor, M. J., and White, K. R. 1992. A meta-analysis of hooking mortality of nonanadromous trout. North American Journal of Fisheries Management **12**: 760–767.

Taylor, M. J., and White, K. R. 1997. Response: trout mortality from baited barbed hook and barbless hooks. North American Journal of Fisheries Management **17**: 808–809.

Thompson, G. G. 1994. Confounding of gear selectivity and the natural mortality rate in cases where the former is a nonmonotone function of age. Canadian Journal of Fisheries and Aquatic Sciences **51**: 2654–2664.

Trippel, E. A., Kjesbu, O. S., and Solemial, P. 1997. Effects of adult age and size structure on reproductive output in marine fishes. *In* Early Life History and Recruitment in Fish Populations. *Edited by* R. C. Chambers and E. A. Trippel. Chapman and Hall, pp. 29–62.

Turek, S. M., and Brett, M. T. 1997. Comment: trout mortality from baited barbed hook and barbless hooks. North American Journal of Fisheries Management **17**: 807.

Vetter, E. F. 1988. Estimation of natural mortality in fish stocks: a review. Fishery Bulletin **86**: 25–43.

Vogt, W. G., and Morton, R. 1991. Estimation of population size and survival of sheep blowfly, *Lucilia cuprina*, in the field from serial recoveries of marked flies affected by weather dispersal and age-dependent trappability. Researches on Population Ecology **33**: 141–163.

Walters, C., Christensen, V., and Pauly, D. 1997. Structuring dynamic models of exploited ecosystems from trophic mass-balance assessments. Reviews in Fish Biology and Fisheries **7**: 39–172.

Walters, C., and Ludwig, D. 1994. Calculation of Bayes posterior probability distributions for key population parameters. Canadian Journal of Fisheries and Aquatic Sciences **51**: 713–722.

Walters, C., and Maguire, J. J. 1996. Lessons for stock assessment from the Northern cod collapse. Reviews in Fish Biology and Fisheries **6**: 125–137.

Wardle, C. S. 1983. Fish reactions to towed fishing gears. *In* Experimental biology at sea. *Edited by* A. MacDonald and I. G. Priede. Academic Press, New York, pp. 167–195.

White, H. D. 1994. Scientific communication and literature retrieval. *In* The Handbook of Research Synthesis. *Edited by* H. Cooper and L. V. Hedges. Russell Sage Foundation, New York, pp. 41–55.

Wood, S. N. 1997. Inverse problems in structured-population dynamics. *In* Structured-population models in marine, terrestiral, and freshwater systesm. *Edited by* S. Tuljapurkar and H. Caswell. Chapman and Hall, New York, pp. 555–586.

Worm, B., and Myers, R. A. 2002. Strong top-down effects in marine food webs: a meta-analysis of cod-shrimp interactions in the North Atlantic Ocean. Ecology.

Zar, J. H. 1996. Biostatistical analysis (3rd ed.). Prentice-Hall Inc, New Jersey.