

BAYESIAN MODELLING OF THE
SIZE-SELECTIVITY OF FISHING TRAWLS

Tom Mitchell Elliott

A thesis submitted in partial fulfilment of the requirements for the degree of
Master of Science in Statistics, The University of Auckland, 2015.

Abstract

The size-selectivity of fishing gear is of particular interest to fisheries managers, aiding in the reduction of by-catch and increase of profits. Size-selectivity models have typically been fitted using generalised linear models (GLMs) and mixed-effects models implemented using software such as SAS. We investigate a Bayesian approach to fitting these models, with special focus on the comparability of frequentist and Bayesian models. We look at several case studies, which are used to establish the validity of our models by comparing our results to those published in previous analyses. Markov chain Monte Carlo (MCMC) diagnostic tests and posterior predictive checks (PPCs) for overdispersion are explored, and random-effects are shown to be the preferred method of modelling overdispersion in the data. Having formulated a general model, several extensions are investigated. The Poisson distribution is used as an alternative likelihood function, allowing the models to make use of any population size distribution information, and multiple random-effects models are implemented that were previously found to be too complex. Semiparametric selection curves—notably basis splines—are used with the constraint that they be continuous, non-decreasing functions of length. A new R package, BSM, is introduced as a simple tool for fitting many of the Bayesian selectivity models discussed in this thesis.

Acknowledgements

First, I thank my supervisor, Associate Professor Russell Millar, who not only provided many useful insights, but also laid many of the foundations on which this work is built, with his many contributions to the field.

Of course, this work would not have been possible without any data, so thank you to Paulo Fonseca of the Portuguese Institute for the Ocean and Atmosphere, Lisbon, for the rose shrimp data set, and to Matt Broadhurst of the Fisheries Conservation Technology Unit, NSW Department of Primary Industries, Coffs Harbour, for both of the school prawn data sets.

I also thank the donor of the Jane Marsden Postgraduate Scholarships, which provided me with considerable financial support during my degree, and to the Scholarships Office for selecting me as one of the two recipients of the award.

Thank you (immensely) to those friends who were supportive throughout the year, providing me with both the emotional support and social escape that contributed significantly to the maintenance of my sanity, especially as the due date approached. I needn't name names—you know who you are. Thank you.

And, finally, I thank my parents, Robin and Judy. You have supported me unconditionally throughout my many years of study, and I commend you for putting up with a twenty-something year old hanging around home far more than he should. You have been the most supportive, loving parents any child could ask for; that support allowed me to follow postgraduate study, that love fed my passion to do so. For that, I dedicate this thesis to you.

Contents

ABSTRACT	iii
ACKNOWLEDGEMENTS	v
CONTENTS	vii
LIST OF TABLES	xi
LIST OF FIGURES	xv
GLOSSARY	xix
ACRONYMS	xxi
LIST OF SYMBOLS	xxiii
1 INTRODUCTION TO TRAWL FISHERIES	1
1.1 Trawl Fisheries	2
1.2 Types of Size-selectivity Experiments	3
1.3 Outline of Thesis	5
2 SIZE-SELECTIVITY MODELS FOR FISHING TRAWLS	7
2.1 Modelling the Retention Properties of Trawl Gear	8
2.1.1 Generalised Linear Models for Retention Probabilities	8
2.1.2 An Application to Count Data from Trawls	11

2.1.3	Dealing with Multiple Hauls	15
2.1.4	Paired Trawl Experiments	17
2.2	A (Brief) History of Size-selectivity Modelling	19
3	BAYESIAN SIZE-SELECTIVITY MODELLING	23
3.1	Overview of Bayesian Inference	24
3.1.1	Bayes and MCMC	24
3.1.2	MCMC Diagnostic Checks	25
3.1.3	Hierarchical Bayesian Models	27
3.2	A Single-gear Covered-codend Trawl	28
3.2.1	Combined-hauls Model	30
3.2.2	Hierarchical Model	33
3.3	A Covered-codend Trawl with Covariates	34
3.4	A Twin-trawl	41
3.4.1	Combined-hauls Model	42
3.4.2	Hierarchical Model	43
3.5	Integrating Modelling Approaches	46
3.5.1	Weighting Hauls in a Hierarchical Analysis	47
3.5.2	Fitting a Weighted Hierarchical Model	47
3.6	Chapter Summary	48
4	MODEL CHECKING, COMPARISON AND SELECTION	51
4.1	Checking and Correcting for Overdispersion	52
4.1.1	Alternative Distributions	53
4.1.2	Random Effects	54
4.1.3	Covered-codend School Prawn Example	54
4.2	Model Comparison and Selection	56
4.2.1	Covered-codend School Prawn Models	58
4.2.2	Covered-codend Rose Shrimp Models	59

4.2.3	Twin-trawl School Prawn Models	60
5	EXTENSIONS OF BAYESIAN MODELS	63
5.1	The Poisson Distribution	63
5.1.1	Examples of the Poisson Model	65
5.1.2	Weighting the Hierarchical Poisson Model	68
5.2	Selection Curves	69
5.2.1	Covered-codend School Prawn	69
5.2.2	Covered-codend Rose shrimp	71
5.2.3	Twin-trawl School Prawn	71
5.3	Additional Random Effects	72
5.3.1	Covered-codend School Prawn Data	73
5.3.2	Covered-codend Rose Shrimp Data	73
6	SEMIPARAMETRIC SELECTION CURVES	75
6.1	Piecewise Linear Model	76
6.2	Shape-constrained Regression Splines	79
6.3	Basis Spline Model	83
6.4	How Do We Estimate the Parameters of Interest?	86
6.4.1	Examples Using a B-spline Selection Curve	87
6.5	Hierarchical Semiparametric Model	94
6.6	Chapter Summary	95
7	BSM: AN R PACKAGE FOR BAYESIAN SELECTIVITY MODELLING	97
7.1	Loading and Displaying Data	97
7.2	Fitting a Combined-hauls Selectivity Model	99
7.3	Fitting a Hierarchical Model	103
8	SUMMARY OF BAYESIAN SIZE-SELECTIVITY MODELLING	105
8.1	Future Work	107

APPENDIX

A DERIVATION OF L_{50} AND SR FOR THE RICHARDS' CURVE	109
B JAGS MODELS	113
B.1 Combined-hauls Models	114
B.1.1 Covered-codend	114
B.1.2 Paired trawl	115
B.2 Hierarchical Models	116
B.2.1 Covered-codend	116
B.2.2 Covered-codend with covariates	117
B.2.3 Paired trawl	118
B.3 Other Parametric Models	120
B.3.1 Weighted Hierarchical Model	120
B.3.2 Overdispersion Random Effects Model	122
B.3.3 Poisson-likelihood Model	124
B.4 Semiparametric Models	126
B.4.1 Piecewise Linear Model	126
B.4.2 Truncated Multivariate Normal Model	127
B.4.3 B-spline Model	128

List of Tables

3.1	Summary of the posterior distributions of the selectivity parameters of the combined-hauls model fitted to the covered-codend school prawn data.	31
3.2	Summary of the posterior distributions of the selectivity parameters of the hierarchical model fitted to the covered-codend school prawn data.	34
3.3	Summary of the posterior distributions of the parameters of the hierarchical model fitted to the covered-coded rose shrimp data.	38
3.4	Summary of the posterior distributions of L_{50} for each gear used in the covered-codend rose shrimp experiment.	38
3.5	Summary of the posterior distributions of the selectivity parameters for each of the three gears in the twin-trawl school prawn data.	43
3.6	Summary of the posterior distributions of the selectivity parameters of the hierarchical model fitted to the twin-trawl school prawn data.	44
3.7	Summary of the posterior distributions of the parameters used in the weighted hierarchical model fitted to the covered-codend school prawn data.	47
4.1	Comparison of two models with and without overdispersion random effects fitted to the uncombined covered-codend school prawn data.	55
4.2	Comparison of Richards' and Logistic selection curves fitted using a combined-hauls model to the covered-codend school prawn data.	58
4.3	Comparison of the models fitted to the covered-codend rose shrimp data. .	59

4.4	Comparison of the models fitted to the twin-trawl school prawn data by forward selection.	61
5.1	Summary of the posterior distributions of the selectivity parameters of the combined-hauls Poisson model fitted to the covered-codend school prawn data.	66
5.2	Summary of the posterior distributions of the selectivity parameters of the hierarchical Poisson model fitted to the covered-codend school prawn data.	67
5.3	Summary of the posterior distributions for the parameters of the hierarchical Poisson model fitted to the covered-coded rose shrimp data.	67
5.4	Summary of the posterior distributions of the selectivity parameters of the combined-hauls Poisson model fitted to each of the three gears in the twin-trawl school prawn data.	68
5.5	Summary of the posterior distributions of the selectivity parameters of the hierarchical Poisson model fitted to the twin-trawl school prawn data.	68
5.6	Comparison of the Richards' and logistic selection curves fitted using a hierarchical model to the covered-codend school prawn data.	69
5.7	Comparison of the Richards' and logistic selection curves fitted using a hierarchical model to the covered-codend rose shrimp data.	71
5.8	Comparison of Richards' and logistic selection curves fitted using a hierarchical model to the twin-trawl school prawn data.	72
5.9	Comparison of the hierarchical models fitted to the covered-codend school prawn data, with and without a hierarchical prior distribution on SR	73
5.10	Comparison of the hierarchical models fitted to the covered-codend rose shrimp data, with and without a hierarchical prior distribution on SR	74
6.1	Summary of the posterior distributions of the parameters of the piecewise linear model fitted to the stacked covered-codend school prawn data.	78
6.2	Summary of the posterior distributions of the parameters of the shape-constrained semiparametric model fitted to the stacked covered-codend school prawn data. .	82

List of Figures

2.1	A (logistic) selection curve, labelled with the selectivity parameters often used to quantify gear selectivity	8
2.2	The Richards' selection curve.	10
3.1	The scaled covered-codend school prawn data	29
3.2	Trace plots (left) and density plots (right) of the posterior samples of the selectivity parameters of the combined-hauls model fitted to the covered-codend school prawn data	32
3.3	Posterior correlations of two alternative parameterisations of the logistic selection curve	32
3.4	Trace plots (left) and density plots (right) of the posterior samples of the selectivity parameters of the hierarchical model fitted to the covered-codend school prawn data	35
3.5	The scaled covered-codend school prawn data, with the combined-hauls and hierarchical models' estimated (mean) selection curves.	35
3.6	The scaled covered-codend rose shrimp data	36
3.7	Trace plots (left) and density plots (right) of the posterior samples of the parameters used in the hierarchical model fitted to the covered-codend rose shrimp data	39
3.8	Scaled covered-codend rose shrimp data with the (mean) estimated logistic selection curves from the hierarchical model	40

3.9	The scaled twin-trawl school prawn data	42
3.10	The scaled twin-trawl school prawn data with the estimated logistic selection curves fitted by the combined-hauls model	45
3.11	The scaled twin-trawl school prawn data with the estimated logistic selection curves fitted by the hierarchical model	45
4.1	The scaled covered-codend school prawn data with the estimated Richards' selection curve fitted using the combined-hauls model, with and without random effects for overdispersion.	55
5.1	The posterior distribution of length class counts (with 95% credible intervals) for the covered codend school prawn data, estimated by the Poisson model .	66
5.2	The scaled covered-codend school prawn data, with the estimated logistic and Richards' selection curves fitted using Bayesian hierarchical models	70
6.1	The scaled covered-codend school prawn data, with the estimated piecewise linear selection curve	78
6.2	The scaled covered-codend school prawn data, with the estimated shape-constrained regression spline	82
6.3	Trace plots for the parameters used in the basis spline model fitted to the stacked covered-codend school prawn data	85
6.4	The scaled covered-codend school prawn data, with the estimated B-spline selection curves	86
6.5	The scaled covered-codend school prawn data with the logistic, Richards' and B-spline selection curves fitted using combined-hauls models.	89
6.6	The scaled covered-codend rose shrimp data, subsets by mesh and cruise, with the estimated B-spline selection curves.	91
6.7	The scaled twin-trawl school prawn data, subset by gear, with the estimated B-spline selection curves	93

6.8 The scaled covered-codend school prawn data, with the estimated B-spline selection curves fitted using a hierarchical model	95
---	----

Glossary

by-catch Non-target animals (different species or undersized) caught unintentionally.

codend (trawl gear) The narrow, closed-off end of a trawl net.

cover (trawl gear) A fine mesh net placed over the codend in a covered-codend experimental trawl to capture any animals that escape from the codend.

exploited (fully, over) A *fully exploited* fishery is one that is operating at or close to an optimal yield level, with no expected room for further expansion. An *over exploited* stock is exploited above a level that is believed to be sustainable in the long term, evident from the steady decline of the stock (FAO, 2012).

haul A single deployment of a trawl gear.

isotonic Monotone increasing, or non-decreasing, function.

length class continuous animals lengths are counted into discrete class intervals.

stock (fisheries) Characteristics of semi-discrete groups of fish with some definable attributes which are of interest to fishery managers (Begg, Friedland, & Pearce, 1999).

Acronyms

AIC Akaike information criterion.

BIC Bayesian information criterion.

CDF cumulative distribution function.

DIC deviance information criterion.

GLM generalised linear model.

GLMM generalised linear mixed model.

GNLM generalised non-linear model.

GNLMM generalised non-linear mixed model.

MCMC Markov chain Monte Carlo.

MH Metropolis Hastings.

MLE maximum likelihood estimate.

MLS minimum legal size.

MPSRF multivariate potential scale reduction factor.

PPC posterior predictive check.

PSRF potential scale reduction factor.

QMS Quota Management System.

REP replicate estimate of overdispersion.

List of Symbols

L_{50}	the length of 50% retention
SR	the selection range; difference in the lengths of 25% and 75% retention
δ	asymmetry parameter in the Richards' selection curve
ϕ	the relative fishing efficiency parameter used in paired trawls
μ_Z, σ_Z^2	the hierarchical mean and variance for parameter $Z = \{L_{50}, SR, \phi\}$
λ	the <i>estimated</i> total number of animals caught
$\chi_{\text{exp}}^2, \chi_{\text{obs}}^2$	expected and observed Pearson's Chi-square statistics, respectively
p_{od}	posterior probability that $\chi_{\text{obs}}^2 > \chi_{\text{exp}}^2$
σ_{od}^2	the overdispersion variation of observation random effects
ℓ	the length class variable
$y, y_{\text{cod}}, y_{\text{expt}}$	the number of animals counted in the codend/experimental gear
$y_{\text{cov}}, y_{\text{ctrl}}$	the number of animals counted in the cover/control gear
q_1, q_2	the codend/experimental and cover/control sampling fractions
n	the total number of animals counted
p	the (expected) binomial probability, y/n
η	the predictor in the GLM framework
H, M	the number of hauls in the experiment (JAGS models M)
N	the number of length classes

Other symbols, such as α , β , and γ , have various definitions (in an attempt to use the same notation as previous papers). Therefore, their definitions are given in the relevant sections.

Chapter 1

Introduction to Trawl Fisheries

Commercial fisheries are of global importance, with a total value of US\$217.5 billion (FAO, 2012). In New Zealand, our \$3.5 billion commercial fishing industry makes up about 94% of the total marine catch, with recreational and customary fisheries making up the remainder (Ministry for Primary Industries, 2010). Nearly 60% of the world's fish come from wild stocks, many of which are at risk from overfishing, and about 70% of global fish stocks were either fully or over-exploited in 2009 (FAO, 2012). In order to manage these fisheries, fisheries managers enforce strict quotas and size restrictions in an attempt to ensure the sustainability of the world's natural fish stocks.

One defining feature of commercial fisheries is the choice of gear, which can have economical and ecological consequences. Commercial fisheries, as with most other commercial operations, are predominantly focused on monetary profit, so the cost of deploying and maintaining a gear must balance with the profits made from catches. However, the ecological impacts of fishing must also be examined. Many global fish stocks are vulnerable to overexploitation, meaning they could collapse if managed incorrectly (FAO, 2012). Some fisheries, such as the invasive bottom trawl, can substantially impact the local habitat (Collie, Hall, Kaiser, & Piner, 2000; Thrush & Dayton, 2002), hence fisheries managers need to explore how the ecological footprint of fishing gear deployed in the fishery can be reduced.

Selecting gear that is “environmentally friendly” is a highly explored and debated topic.

Fisheries managers are interested in stock attributes such as the rates of by-catch (non-target species), the quantities of sub-legal animals caught, and the incidental mortality rate of discards (by-catch and sub-legal animals caught in the trawl) (Pope, Margetts, Hamley, & Akyuz, 1975; Suuronen, 1995; Chopin, Inoue, & Arimoto, 1996; Begg et al., 1999; Hall, Alverson, & Metuzals, 2000; Cook, 2003).

1.1 Trawl Fisheries

The fishing gear with the most potential to both overexploit the fishery and to damage the habitat is arguably the fishing trawl. This is a non-passive gear that is forcefully towed through the water and can cover huge areas in a single deployment. New gears are frequently being designed with the aim of increasing performance, while reducing the amount of by-catch, which decreases the net value of each catch (Broadhurst, 2000; Hall et al., 2000; Tscherñij, Suuronen, & Journela, 2004). While doing so, it is important that catch rates of the targeted animals are not negatively affected (Tscherñij et al., 2004). One example we are interested in is the *size-selectivity* of fishing trawl gear, where the goal is to deploy a gear that reduces the quantity of animals caught under the minimum legal size (MLS), reducing the proportion of each catch that is then discarded as by-catch (Pope et al., 1975; Millar & Walsh, 1992; Suuronen, 1995; Wileman, Ferro, Fonteyne, & Millar, 1996; Millar & Fryer, 1999; Hall et al., 2000; Cook, 2003). This also helps to reduce the number of undersized animals subject to accidental mortality, which can have major consequences on fisheries that depend on these smaller individuals for reproduction and recruitment (Chopin et al., 1996; Hall et al., 2000; Cook, 2003), providing a strong ecological reason to explore new gears.

Commercial fishing trawl gear consists of a large net made of various types of mesh. This net, the front of which is held open by weights and floats, is dragged through the water, or over the sea floor in the case of bottom-trawls, by the fishing vessel (Pope et al., 1975). Animals are herded by the mouth (front) of the net, which tapers down into a closed *codend*—it is here that the animals actively make most escape attempts. Therefore, the

mesh used in this part of the gear can have a great influence on the overall size-selectivity of the gear—smaller animals can escape from the trawl, while larger ones will be caught (Pope et al., 1975; Wileman et al., 1996; Millar & Fryer, 1999; Fonseca, Campos, & Millar, 2007; Williams, Punt, Wilson, & Horne, 2011). Fisheries managers are therefore interested in the selectivity properties of the codend mesh of gear used in commercial trawl operations.

The size-selectivity of fishing gear is of interest to fisheries managers for both economical and ecological reasons to maintain the stability of fish stocks, which often involves studies of the impacts of fishing on the local ecology (Collie et al., 2000; Hall et al., 2000; Thrush & Dayton, 2002). In both cases, a quantifiable estimate of gear selectivity is desired, which can be used to estimate loss or gain in profits, or for use in stock assessment models (Suuronen, 1995; Chopin et al., 1996). Additionally, researchers need a model for formally comparing the performance of different gear, as fisheries managers, and—more importantly—fishermen, require convincing evidence that a new gear has better selectivity properties (and higher profits) than the one currently being used before it will be adopted into commercial operations.

1.2 Types of Size-selectivity Experiments

Quantification of the size-selectivity of a chosen gear is obtained from experimental trawls. In these, the experimental gear is used alongside a control gear (a fine mesh net that allows no animals to escape) (Pope et al., 1975). There are several different experimental designs commonly used, though in this thesis we focus on the *covered-codend* and *paired* trawls. In a covered-codend trawl, the experimental gear is modified by placing the control mesh over the experimental codend. In these experiments, the *codend* refers to the experimental mesh codend, while the *cover* refers to the control mesh covering it (Pope et al., 1975; Wileman et al., 1996). This method allows researchers to directly quantify the proportions of animals of various lengths caught in the codend, hence getting a direct estimate of how the selectivity of the gear changes over the range of animal lengths. Alternatively, researchers can use a paired trawl, in which both gears (experimental and control) are used, either

towed in parallel by one or more vessels (twin-trawl), fished alternatively by a single vessel, or in a modification known as the trouser-trawl (Pope et al., 1975; Millar & Walsh, 1992). The trouser-trawl involves a gear with two codends—the experimental and the control—separated by a vertical divider. In either of these paired trawls, the relative proportions of animals of various lengths are compared between the experimental and control codends. There are inevitably consequences to each method, which are discussed by Pope et al. (1975) and Millar, Broadhurst, and Macbeth (2004).

In any experimental trawl, the gear is deployed several times, and the counts of animals in the codend and cover—or experimental and control in the case of a paired trawl—are recorded for a range of length classes (Pope et al., 1975). Each deployment of the gear is referred to as a *haul*. For each haul, additional variables may be measured, such as the total catch weight, trawl depth, and in the case of studies that take place over a period of time, the day or season in which the haul took place. Any models used to analyse the data must therefore be able to incorporate these additional variables. Additionally, they need to account for between-haul variability, a significant feature of fishing trawls, attributed to many unknown factors (Fryer, 1999; Millar et al., 2004).

Another feature of experimental trawl data is *subsampling*. This occurs when catch sizes are too large for the fishermen to measure the size of every individual animal, so instead only a subsample is used. An extreme example of this is the rose shrimp experimental data from Fonseca et al. (2007), in which catches of rose shrimp are so large that often less than 5% of the total catch was processed. This value is known as the *sampling fraction*, and is estimated for the codend and cover separately (or experimental and control codends in the case of a paired trawl) (Millar, 1994). Models that do not account for subsampling are likely to give biased results when the sampling fractions from both codend and cover are not equal (discussed in § 2.1.2).

1.3 Outline of Thesis

In this thesis, we are primarily interested in the size-selectivity of fishing gear, with our main focus on the relationship between animal length and retention probability (the probability of being caught). While we expect this relationship to be isotonic (non-decreasing), the specific shape is unknown, and requires estimation. Models will also need to account for other design covariates, such as mesh size or shape, when comparing the selectivity of different gears.

In Chapter 2, we explore how fisheries managers quantify gear selectivity, and see how the logistic curve is used to achieve this by presenting the model as a simple binomial generalised linear model (GLM). We also show how the predictor can be reparameterised to be more useful. This can be further generalised to account for between-haul variability (a mixed model), or to analyse data from a paired trawl experiment. A generalisation of the logistic curve commonly used in size-selectivity studies is also introduced, and we briefly explore the models currently being used in size-selectivity analyses, and define two separate approaches to modelling the data: combined-hauls and mixed-effects models.

In Chapter 3, we focus on fitting the models from several case studies using Bayesian methods and comparing the results. We give a brief overview of Bayesian analysis by Markov chain Monte Carlo (MCMC), diagnostic tests, and hierarchical models. Our main aim for this chapter will be to show that we can fit the same models as the frequentist methods—and get the same results—to show that Bayesian models can be used and trusted for the analysis of experimental trawl data. A novel weighted approach is also proposed that could allow the integration of the combined-hauls and mixed-effects approaches.

Model checking and model comparison are the topics of Chapter 4. This includes posterior predictive checks (PPCs) for overdispersion, common to trawl data, and how this is modelled in a Bayesian framework. Comparison of competing models using deviance information criterion (DIC) will also be discussed.

The next two chapters focus on exploring the flexibility of Bayesian models. Chapter 5 will involve fitting some alternative types of models. These include a Poisson distribution (in

place of a binomial), the use of a generalised logistic curve in a hierarchical model, and the implementation of multiple random effects. In Chapter 6, we explore several semiparametric approaches to quantify gear selectivity, with special focus on the use of basis splines.

In many areas, Bayesian models can be difficult to implement because there is no specific software for many problems, so we developed an R package that allows researchers to easily fit the Bayesian models discussed in this thesis to their own experimental trawl data. Chapter 7 provides a brief introduction to and overview of this software.

Chapter 2

Size-selectivity Models for Fishing Trawls

The main aim of size-selectivity experiments is to investigate the *contact-selection curve* of the chosen gear, defined by Millar and Fryer (1999) as “the (relative) probability that a fish of length ℓ is captured given that it contacted the gear” (p. 92). This is commonly quantified by two parameters that are of particular interest to fisheries managers: L_{50} is the length of an animal with a 50% retention probability, and SR , the selection range, is the difference between the lengths of animals with 25% and 75% retention probabilities. These parameters are described graphically in Figure 2.1.

In this chapter, we first review the general structure of a binomial generalised linear model (GLM), and see how it is applied to data from trawl selectivity experiments. We discuss how subsampling can be incorporated into the model, as well as give an outline of two approaches used to model the data from multiple hauls. The generalised linear mixed model (GLMM) will be presented as a generalisation to the GLM that allows modelling of between-haul variation. We also describe how a generalised non-linear model (GNLM) is used to model data from paired trawl experiments. Finally, we briefly review some of the literature regarding size-selectivity models.

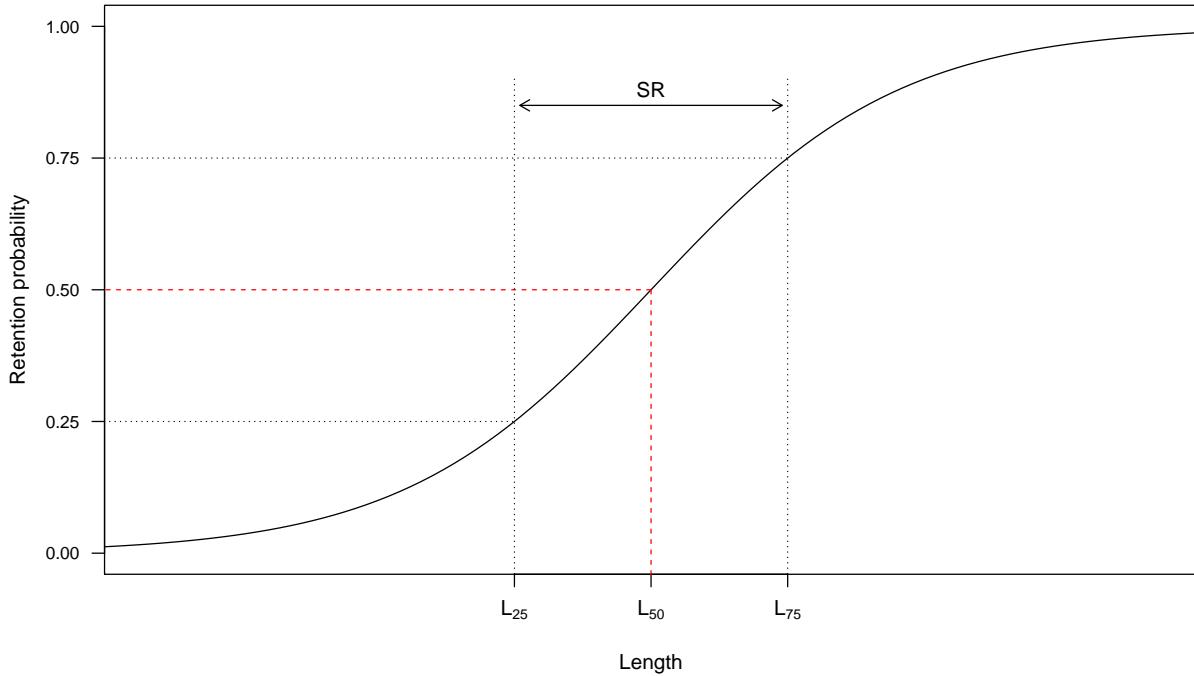


Figure 2.1: A (logistic) selection curve, labelled with the selectivity parameters often used to quantify gear selectivity: L_{50} , the length of 50% retention, and SR , the selection range.

2.1 Modelling the Retention Properties of Trawl Gear

Trawl selectivity studies aim to quantify the size-selectivity of a fishing gear in the chosen fishery (Pope et al., 1975; Millar & Fryer, 1999). The data are recorded as separate counts for a range of length classes, for each of the codend and cover in a covered-codend experiment, or experimental and control codends in a paired trawl. By comparing the counts, the retention probability for an animal of a given length can be estimated. Analyses of these data require a model that relates the main covariate, length, to the observed counts. We must therefore understand the basic structure of the GLMs used to analyse count data.

2.1.1 Generalised Linear Models for Retention Probabilities

GLMs consist of three parts: a *predictor*, a function of the explanatory variables; a *link function* that transforms the predictor to the appropriate domain, and a *distribution function* that describes the error structure of the observations. We must also consider overdispersion,

a feature common to trawl data, which occurs when the observed variance is larger than predicted under the distribution function.

The predictor is a function of the measured variables (in our case length, ℓ) that relates them to the response,

$$\eta = m(\ell). \quad (2.1)$$

A commonly used form is the *linear predictor*, although $m(\ell)$ can be a function of any desired form, and not necessarily parametric (see Chapter 6). For now, we use the simple linear predictor

$$\eta = \beta_0 + \beta_1 \ell \quad (2.2)$$

for simplicity, although this can be extended easily to include additional covariates. In our application, ℓ is the length variable, and β_0 and β_1 the intercept and slope, respectively.

We will (for now) be working with probabilities, which by definition lie between 0 and 1. Therefore, we need to transform the predictor, $\eta \in \mathbb{R}$, to the appropriate domain, $p \in (0, 1)$. In a GLM, this is accomplished by the link function, defined as $g(p) = \eta$. A common example of this is the logit, or log-odds, link function, defined as

$$\eta = \text{logit}(p) = \log\left(\frac{p}{1-p}\right). \quad (2.3)$$

In GLMs, we use the inverse link function to predict the response from the covariates, $p = g^{-1}(\eta)$. The inverse logit link is

$$p = \text{logit}^{-1}(\eta) = \frac{e^\eta}{1 + e^\eta}, \quad (2.4)$$

which is also the cumulative distribution function (CDF) of the *logistic function*, or *logistic curve*, used in logistic regression (McCullagh & Nelder, 1989).

The logistic curve is a popular function for modelling selectivity, as it intuitively relates the length covariate to the retention probability. However, it does make the assumption that this relationship is symmetric, which may not be the case. To allow for asymmetrical

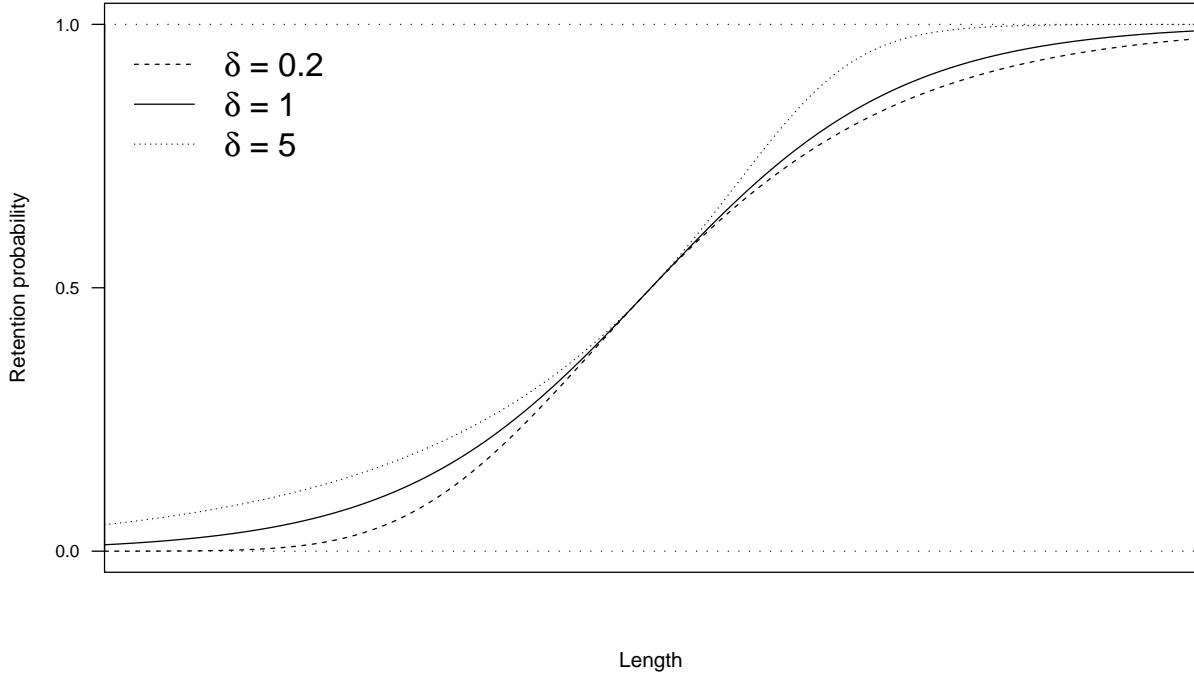


Figure 2.2: The Richards’ selection curve for varying values of δ , while retaining the same selectivity values for L_{50} and SR .

relationships, “generalised” logistic curves, such as the Richards’ curve proposed by Richards (1959), can be used instead. This selection curve has been used in previous trawl analyses, such as in Millar et al. (2004), who found it gave a better fit to their data than the logistic curve.

The Richards’ generalisation we will be working with modifies the logistic curve by adding a single parameter, δ , to allow for asymmetry in the relationship between length and retention probability. The inverse link using the Richards’ selection curve is

$$p = \left(\frac{e^\eta}{1 + e^\eta} \right)^{1/\delta}. \quad (2.5)$$

Note that when $\delta = 1$, the Richards’ curve is equivalent to the logistic in (2.4), which will be useful when comparing models in Chapter 4. The asymmetry due to δ is demonstrated in Figure 2.2. Further generalisations are possible, however only the logistic and Richards’ curves are considered in this thesis.

The final component to fitting a GLM is the choice of distribution function. We must make an assumption about the error structure of the data, which usually depends on the form of the data. In the case of a covered-codend experimental trawl, the binomial is an obvious choice. If y_ℓ is the number of length ℓ animals caught in the codend, n_ℓ is the total number of length ℓ animals caught in both codend and cover, and p_ℓ is the retention probability of a length ℓ animal, then the distribution function for the data is

$$y_\ell \sim B(p_\ell; n_\ell). \quad (2.6)$$

Assuming that the observed data are binomial observations, then the expected number of length ℓ animals caught in the codend, given that n_ℓ were caught in total, is

$$\mathbb{E}[Y_\ell | \ell; n_\ell] = n_\ell p_\ell. \quad (2.7)$$

Similarly, the variance is

$$\text{Var}_f[Y_\ell | \ell; n_\ell] = n_\ell p_\ell (1 - p_\ell). \quad (2.8)$$

However, if the *observed* variation in y is greater than expected, we have what is known as *overdispersion*. In frequentist models, this is often accounted for *after* fitting the GLM. An overdispersion parameter is estimated and used to correct the standard errors of the parameter estimates, but does not affect the estimates themselves. This is known as quasi-likelihood (McCullagh & Nelder, 1989). In Bayesian models, however, we cannot use this same approach. We will discuss how overdispersion can be dealt with in a Bayesian framework in Chapter 3.

2.1.2 An Application to Count Data from Trawls

In this section, the binomial GLM is applied to the simple case of a single-haul covered-codend experiment. This is done first without subsampling, after which we show how the sampling fractions are easily incorporated into the model. In § 2.1.3 and § 2.1.4, we see how

this model can be modified for multiple hauls and paired trawl experiments.

A simple covered-codend experiment

The simplest experimental trawl is a single-haul covered-codend trawl using a single gear with no additional covariates or subsampling. In this case, we obtain two vectors of counts: \mathbf{y}_{cod} containing the codend counts for each length class, and \mathbf{y}_{cov} containing the cover counts for each length class. The corresponding length classes are stored in the vector ℓ .

To apply the binomial GLM to the data, the codend count $\mathbf{y} = \mathbf{y}_{\text{cod}}$ is used as the *response* variable (number of successes), and the total count, which is the sum of the codend and cover counts, $\mathbf{n} = \mathbf{y}_{\text{cod}} + \mathbf{y}_{\text{cov}}$, is used for the number of trials. The lengths ℓ are used to predict the retention probabilities, $r(\ell)$, which is defined as a function of length:

$$r(\ell) = \text{logit}^{-1} (\eta_\ell). \quad (2.9)$$

These are used in the binomial probabilities in (2.6), which, with no subsampling, are $p_\ell = r(\ell)$ (i.e., the probability of “success”—retention in the codend—is $r(\ell)$). This model is easily fit using frequentist and Bayesian methods, and the maximum likelihood or posterior estimate of $r(\ell)$ is the estimate of the contact-selection curve. The next step involves quantifying this relationship in a way that is useful to fisheries managers.

As shown in Figure 2.1, the shape of the contact-selection curve can be summarised in terms of L_{50} and SR . L_{50} is the length that corresponds to a retention probability of $r(\ell) = 0.5$, so by using the definitions in (2.9) and (2.4), this parameter can be expressed in closed-form as a function of the model parameters

$$L_{50} = -\frac{\beta_0}{\beta_1}, \quad (2.10)$$

and the standard error of L_{50} can be obtained by the delta method (Wileman et al., 1996). Similarly for SR , the difference between animal lengths with 25% retention probability, L_{25} ,

and 75% retention probability, L_{75} , it can be expressed as

$$SR = \frac{2 \log(3)}{\beta_1}. \quad (2.11)$$

These formulae allow the quantification of the size-selectivity of a fishing gear in terms that are meaningful to fisheries managers.

In Millar (1993), the logistic curve was parameterised in terms of L_{50} and SR instead of β_0 and β_1 , by expressing the predictor as

$$\eta_\ell = \frac{2 \log(3)}{SR} (\ell - L_{50}). \quad (2.12)$$

This parameterisation is no longer linear, and hence we are working with a GNLM; however, it does have several useful properties. First, a direct estimate of the parameters of interest and their associated standard errors are obtained without requiring the use of the delta method. Secondly, and perhaps most importantly for the work we have been doing, is that this parameterisation removes most of the correlation between the model parameters. It is well known that the correlation between β_0 and β_1 in a standard GLM can be strong, which can cause simulation problems when using Markov chain Monte Carlo (MCMC) methods, which Gelfand, Sahu, and Carlin (1995) suggest can be reduced by centering the covariates. Zuur, Garthwaite, and Fryer (2002) used MCMC to fit hierarchical models to covered-codend data, much like we did in our work (Chapter 3). They showed that centering the length covariate significantly reduced the correlation between β_0 and β_1 . However, we found this did not reduce the correlation as much as using (2.12) did, which we demonstrate in Chapter 3.

The L_{50} and SR parameterisation has been used in frequentist analyses, for example Millar (1993) and Millar et al. (2004), who implemented (2.12) using SAS (SAS Institute Inc., 1999). Williams et al. (2011) also use the same parameterisation in their paper, using a Bayesian approach to model gear selectivity (see § 2.2). Unfortunately, the parameterisation in (2.12) is only applicable when using a logistic selection curve.

In some cases, such as Millar and Fryer (1999), the logistic curve does not give an adequate

fit, so the Richards' curve was implemented instead. In this case, no mention of how the model was parameterised was made, although it is likely that β_0 and β_1 were used, and L_{50} and SR estimated from the resulting fit. We, however, wanted to continue using L_{50} and SR as the model parameters, and found we could express the predictor, with the additional parameter δ , as

$$\eta_\ell = \frac{\delta \log(3) - \log(4^\delta - 3^\delta) + \log(4^\delta - 1)}{SR} (\ell - L_{50}) - \log(2^\delta - 1), \quad (2.13)$$

the derivation of which is given in Appendix A. Using this parameterisation, we can continue to express the model in terms of L_{50} and SR , allowing us to retain the advantages mentioned previously.

Incorporating Subsampling

In many situations, only a subset of the total catch is processed, which we refer to as subsampling. In most cases, the sampling fraction of the codend, q_1 , differs from the sampling fraction of the cover, q_2 (values for q_1 and q_2 are provided with the data set). Analyses must account for this if valid inferences about the selectivity of the gear are to be made.

One method for incorporating sampling fractions is to “scale up” the data *before* fitting the model. That is, the codend and cover counts are divided by their respective sampling fractions, providing an estimate of the total number of animals in the entire haul. The selection curve is then fit to the scaled data $\tilde{y}_\ell = \frac{y_\ell}{q_1}$ and $\tilde{n}_\ell = \frac{y_{\text{cod},\ell}}{q_1} + \frac{y_{\text{cov},\ell}}{q_2}$, without need to modify the model.

An alternative method is to incorporate the sampling fractions directly into the model, specifically into the linear predictor. In a frequentist GLM, this is known as an *offset* term (McCullagh & Nelder, 1989). Millar (1994) show that the necessary offset term is $\log(Q)$, where $Q = \frac{q_1}{q_2}$. Note that when the sampling fractions are equal, the offset is 0. Again, however, this only holds when the selection curve is of logistic form. One of our main aims was to make flexible models that are easy to generalise, for example to use a Richards'

selection curve in place of the logistic. Therefore, we need to use an alternative, more general approach to incorporate the sampling fractions.

Millar (1994) show that, for any selection curve $r(\ell)$, the proportion of animals captured in the codend, accounting for subsampling, is estimated by

$$p_\ell = \frac{Qr(\ell)}{1 - (1 - Q)r(\ell)}, \quad (2.14)$$

where again $Q = \frac{q_1}{q_2}$. The derivation of (2.14), explained by Millar (1994), requires the use of a Poisson approximation, which we see in Chapter 5. Using (2.14) sampling fractions can be incorporated into any model, regardless of the form of $r(\ell)$.

2.1.3 Dealing with Multiple Hauls

In § 2.1.2, only data from a single-haul was considered; however, this is rarely the case. A typical trawl experiment consists of multiple hauls, each with its own count vectors, sampling fractions, and any other covariates measured in the study. Millar and Fryer (1999) discuss two alternative approaches to modelling data from multiple hauls: a *combined-hauls* approach, and a *mixed-effects* approach.

Combined-hauls Fit

The combined-hauls approach is the simpler of the two methods, and provides an estimate of the contact-selection curve used by fisheries researchers. This approach uses the *scaled* counts ($\tilde{\mathbf{y}}$ and $\tilde{\mathbf{n}}$, to incorporate sampling fractions) *stacked* across all of the hauls (Millar & Fryer, 1999). That is, the combined codend and total counts for all H hauls for length ℓ are

$$\mathbf{y}_\ell^+ = \sum_{h=1}^H \tilde{y}_\ell^h \quad \text{and} \quad \mathbf{n}_\ell^+ = \sum_{h=1}^H \tilde{n}_\ell^h. \quad (2.15)$$

This gives an estimate of the total number of animals in each length class caught in the codend and cover over *all* hauls. Scaling ensures that larger hauls contribute more to the fit than smaller hauls (it is often the case that the same number of animals are sampled in

each haul). A selection curve can then be fitted to the combined-hauls data using the same model from § 2.1.2, but using y_ℓ^+ and n_ℓ^+ as the observed variables.

The main complication of the combined-hauls approach is that scaling can inflate the sample size, leading to underestimates of standard errors, while stacking may obscure any overdispersion present in the data (Millar & Fryer, 1999). The replicate estimate of overdispersion (REP) can be used to deal with both of these issues:

$$\text{REP} = \frac{1}{d} \sum_{\ell} \sum_h \frac{(\tilde{y}_\ell^h - \tilde{n}_\ell^h \hat{p}_\ell)^2}{\tilde{n}_\ell^h \hat{p}_\ell (1 - \hat{p}_\ell)}, \quad (2.16)$$

where d is the degrees of freedom (the number of observations minus the number of parameters in the model) (McCullagh & Nelder, 1989; Millar & Fryer, 1999). This is used to adjust the standard errors of the estimated parameters by multiplying them by $\sqrt{\text{REP}}$, which affects confidence intervals and significance tests. While we cannot sensibly employ this technique *directly* into our Bayesian models, we can estimate REP after fitting the models (i.e., by using \hat{p}_ℓ from the estimated retention curve), providing us with comparable standard errors. Formal Bayesian approaches to modelling overdispersion are discussed in Chapter 4. Another problem with the combined-hauls approach is that it can only be used to estimate the “mean” contact-selection curve, but not for comparison of multiple gears (Millar & Fryer, 1999).

Mixed-effects Models

The alternative approach to modelling data from multiple hauls involves adding random effects to the parameters, providing a direct estimate of the between-haul variability in the size-selectivity of the gear. This approach is especially useful when comparing multiple gears, as it provides a formal hypothesis-testing model (Millar et al., 2004; Fonseca et al., 2007).

To fit a random-effects model, we generalise the model seen in § 2.1.2 to a generalised non-linear mixed model (GNLMM). Millar et al. (2004) do this by adding random effects to the *mean* selectivity parameters $\mu_{L_{50}}$ and μ_{SR} , allowing them to vary between hauls. For

haul h , the selection curve is defined by the parameters

$$L_{50}^h = \mu_{L_{50}} + \epsilon_{L_{50}}^h \quad (2.17)$$

and

$$SR^h = \mu_{SR} + \epsilon_{SR}^h, \quad (2.18)$$

where $\epsilon_{L_{50}}^h \sim N(0, \sigma_{L_{50}}^2)$ and $\epsilon_{SR}^h \sim N(0, \sigma_{SR}^2)$. We are especially interested in this approach, as it can be implemented using a Bayesian hierarchical model (Hill, 1965; Tiao & Tan, 1965).

Millar et al. (2004) compared the combined-hauls and mixed-effects approaches to modelling data from multiple hauls, demonstrating the difference between them. The combined-hauls approach is a *marginal* model, in which the goal is to estimate population selectivity. Conversely, the mixed-effects approach is a *conditional* model (i.e., conditional on the individual random effects for each haul). Hence the combined-hauls approach gives weight proportional to the (scaled) total number of animals caught in each individual haul to obtain the overall selectivity curve, while the mixed-effects approach weights each haul equally. Consequently, the results, and therefore the application of each approach can be quite different: the former estimates the selection curve for the gear marginalised over all hauls, providing an estimate of the population contact-selection curve; the latter yields a “mean” selection curve conditional on the observed hauls, focusing primarily on estimating between-haul variability, model comparison, and hypothesis testing for a difference in the selectivity of multiple gears.

2.1.4 Paired Trawl Experiments

So far, only models for covered-codend experiments have been considered. To extend the model to apply to paired trawls (twin and trouser trawls, § 1.2), the model is further generalised to incorporate an additional parameter: the relative fishing efficiency, ϕ , which is the *probability that a fish enters the experimental gear, given that it enters the trawl* (Millar & Walsh, 1992). While it is feasible to assume that $\phi = 0.5$, studies have shown that assuming

so can introduce bias to the selectivity estimates (Millar & Walsh, 1992).

By defining the binomial success probability p_ℓ as the *probability that a length ℓ animal is caught in the experimental codend, given that it is caught in the trawl*, we can derive the necessary probabilities for the observed data: \mathbf{y}_{expt} , the vector of counts for the experimental codend, and \mathbf{y}_{ctrl} , the vector of counts for the control. In the following derivation, we define the sample space as all animals that come into contact with the gear, and either escape through the experimental mesh, or are caught (in experimental or control codend).

To begin, we note that ϕ is the probability that an animal enters the experimental gear, given that it enters the trawl (note that we assume ϕ is consistent over length classes, $\phi_\ell = \phi$, which is consistent with the literature), and $r(\ell)$ is a selection curve of any form. First, for ease of notation, we will define $\mathbb{P}_\ell(\cdot) = \mathbb{P}(\cdot | \text{animal has length } \ell)$. Now, following the logic from Millar and Walsh (1992), we can derive the probability that an animal of length ℓ is caught in the control gear, given that it enters the trawl:

$$\begin{aligned} & \mathbb{P}_\ell(\text{caught in control gear}) \\ &= \mathbb{P}_\ell(\text{enters control gear}) \times \mathbb{P}_\ell(\text{caught in control gear} | \text{enters control gear}) \\ &= 1 - \phi, \end{aligned} \tag{2.19}$$

remembering that the control gear has a retention probability of 1. Similarly for the experimental gear,

$$\begin{aligned} & \mathbb{P}_\ell(\text{caught in experimental gear}) \\ &= \mathbb{P}_\ell(\text{enters experimental gear}) \times \mathbb{P}_\ell(\text{caught in experimental gear} | \text{enters experimental gear}) \\ &= \phi \times r(\ell). \end{aligned} \tag{2.20}$$

We now have the necessary components to derive the probability p_ℓ for the binomial likeli-

hood:

$$\begin{aligned}
 p_\ell &= \mathbb{P}_\ell(\text{caught in experimental gear} \mid \text{caught}) \\
 &= \frac{\mathbb{P}(\text{caught} \mid \text{caught in experimental gear}) \mathbb{P}_\ell(\text{caught in experimental gear})}{\mathbb{P}_\ell(\text{caught})} \\
 &= \frac{\mathbb{P}_\ell(\text{caught in experimental gear})}{\mathbb{P}_\ell(\text{caught in experimental gear}) + \mathbb{P}_\ell(\text{caught in control gear})} \\
 p_\ell &= \frac{\phi r(\ell)}{\phi r(\ell) + 1 - \phi}.
 \end{aligned} \tag{2.21}$$

The observed data is then modelled using (2.21). That is, the number of length ℓ animals caught in the experimental codend, $y_{\text{expt},\ell}$, is a binomial random variable with $n_\ell = y_{\text{expt},\ell} + y_{\text{ctrl},\ell}$ trials, and success probability p_ℓ as defined in (2.21). However, we need to also account for sampling fractions. To do this, the probabilities in (2.19) and (2.20) are multiplied by by q_2 and q_1 , respectively. This gives the required binomial probability as

$$p_\ell = \frac{q_1 \phi r(\ell)}{q_1 \phi r(\ell) + q_2 (1 - \phi)} = \frac{Q \phi r(\ell)}{1 - (1 - Q r(\ell)) \phi}, \tag{2.22}$$

where $Q = \frac{q_1}{q_2}$. The same parameters are estimated, using either the combined-hauls or mixed-effects approach exactly the same as we saw in § 2.1.3, with the additional estimation of ϕ . Millar et al. (2004) implemented random effects on the relative efficiency parameter, ϕ , when using a mixed-effects model, allowing it to vary between hauls. They used a logit parameterisation to ensure the probabilities are between 0 and 1:

$$\text{logit}(\phi^h) = \mu_\phi + \varepsilon_\phi^h, \tag{2.23}$$

where $\varepsilon_\phi^h \sim N(0, \sigma_\phi^2)$.

2.2 A (Brief) History of Size-selectivity Modelling

Modelling the size-selectivity of trawl gear has been the focus of many studies, most of which have used frequentist models, often fit using SAS. A popular method is SELECT, which

implements the combined-hauls approach discussed in § 2.1.3 (Millar & Fryer, 1999). We will briefly review some of these papers, as well as some Bayesian hierarchical implementations of size-selectivity models.

The SELECT (Share Each LEnthclass's Catch Total) method, proposed by Millar (1991), is the same as the binomial method described earlier. This general method reduces the number of parameters in the model significantly by removing the need to estimate Poisson parameters (Chapter 5). Millar and Walsh (1992) show how this method is applied to data from a trouser trawl experiment, and in Millar (1994) the inclusion of sampling fractions is discussed with application using SELECT. Overall, this has proven to be a popular technique for analysing combined-hauls data.

Millar and Fryer (1999) give a detailed overview into the modelling of size-selectivity experiments. They explore several aspects of the models discussed in § 2.1, their application using SAS, and the modelling of between-haul variability using mixed-effects. The modelling of random effects to account for between-haul variability was focused on by Fryer (1999), who implemented random effects on the regression parameters β_0 and β_1 . They discuss the implications of modelling size-selectivity data, including the importance of modelling the variation of L_{50} between hauls.

In a later paper, Millar et al. (2004) investigate the modelling of between-haul variability, with a focus on comparison of the combined-hauls and mixed-effects approaches. They apply each of these to a covered-codend and a twin-trawl experiment, and discuss how the combined-hauls approach quantifies the overall “fisheries” selectivity of the gear, and relate this to the contact-selection curve. In their mixed-effects model, unlike Fryer (1999), they parameterised the selection curve in terms of L_{50} and SR , and hence were able to model the variability of the parameters directly. The models were fitted using PROC NLMIXED in SAS, mostly without issues, however they do note that in situations where they tried to fit a high number of parameters—modelling three gears simultaneously or fitting covariance parameters between L_{50} and SR —they ran into problems and these models could not be fit. We explore this further in Chapter 5.

The combined-hauls approach is used to estimate the contact-selection curve, which has been used in studies of the impacts of altering mesh sizes on the yield of trawls and to aid in the conservation of small, juvenile fish (Suuronen, 1995; Kuikka, Suuronen, & Parmanne, 1996). These papers emphasise the need for size-selectivity experiments to be used in complex fisheries models to ensure the sustainability of fish stocks. Tschernij et al. (2004) also investigate the effects of mesh size changes, however their focus is on the change in profits for fishermen.

Having models that can estimate the contact-selection curve are of critical importance to fisheries. However, models are also needed that can compare the size-selectivity of multiple gears. The combined-hauls approach does not allow for this, so instead alternative approaches are used. Broadhurst et al. (2004) compare the selectivity of three gears, though they used SELECT and simple hypothesis testing using the standard errors of L_{50} and SR estimated for each gear. Millar et al. (2004) analyse these same data, but instead employ the mixed-effects model. They are able to model the variability between hauls and compare the gears using this model. We used this data set, along with several others, as case studies for the models formulated in our work. A similar study was undertaken by Fonseca et al. (2007), in which the effect of mesh size and other covariates were modelled using mixed-effects models implemented using SAS. However, they experienced some convergence issues when trying to model random effects for both L_{50} and SR , an area we investigate further in Chapter 5.

Although frequentist models seem to be popular among size-selectivity models, there have been some uses of Bayesian methods, particularly the use of hierarchical models for between-haul variability. Zuur et al. (2002) give an overview of Bayesian modelling applied to trawl experiments, although they focus on the techniques used to fit, assess and compare models using only a logistic selection curve. They use a hierarchical approach applied to the parameters β_0 and β_1 , which were each allowed to vary between hauls. They found significant correlation between these parameters, and showed how centering the covariates could be used to reduce this.

A more recent Bayesian analysis by Williams et al. (2011) explored size-selectivity models for trawl gear experiments, although their specific study included the analysis of animal escape through the sides of the net, not just the codend. Their models used the L_{50} and SR parameterisation of the predictor, and gave each a hierarchical prior for between-haul variability. Additionally, they use light intensity as a covariate for predicting L_{50} . They used ADMB to generate posterior samples (Fournier, 2001), and several model-checking techniques were used to assess them, which we discuss in Chapter 4.

Chapter 3

A Bayesian Approach to Size-selectivity Modelling

One of the main aims of our work was to develop flexible Bayesian models that were applicable to a range of data sets, and that could easily be modified to account for various experimental designs. However, before we could explore the flexibility of Bayesian models, it was necessary to replicate traditional methods and show that we could obtain concurring results. For this, the models described in Millar et al. (2004) and Fonseca et al. (2007) were used as case studies. This chapter focuses on the formulation of these models and comparison of the results; model checking (specifically for overdispersion) and model comparison will be discussed in Chapter 4.

As with most Bayesian problems, MCMC methods were used to sample from the posterior distribution, so we give a brief overview of Bayesian inference using MCMC in § 3.1. From here, we will look at three case studies: a single-gear covered-codend experiment (§ 3.2), a multiple-gear covered-codend experiment (§ 3.3), and a twin-trawl experiment (§ 3.4). In § 3.5, we present a novel approach to integrating the combined-hauls and mixed-effects analyses.

3.1 A Brief Overview of Bayesian Inference

In most Bayesian analyses, numerical techniques are used to generate results. The reason for this is that Bayesian models typically involve high-dimensional integration, which may be difficult or impossible to find analytically (Gelman et al., 2014). To deal with this, MCMC methods are used to generate samples that can be used to make inferences about model parameters, such as expectation and variance. This section outlines the basic Bayesian model, and gives an overview of the MCMC methods available. MCMC diagnostics will also be discussed, as these were used to assess the samples we obtained during our research. Finally, we introduce the hierarchical Bayesian model, used extensively in our work.

3.1.1 Bayes and MCMC

Bayesian models require the evaluation of the *posterior distribution* of the parameter θ , given the data x . By Bayes' Theorem, this is

$$\pi(\theta | x) = \frac{\pi(\theta) f(x | \theta)}{\int_{\Theta} \pi(\theta) f(x | \theta) d\theta}, \quad (3.1)$$

where $\pi(\theta)$ is the *prior distribution* of the parameter, and $f(x | \theta)$ is the *likelihood* (Gelman et al., 2014). The denominator in (3.1) is often referred to as the *normalising constant*, and does not depend on θ ; however, it is often unknown or difficult to compute.

To evaluate (3.1) when $\int_{\Theta} \pi(\theta) f(x | \theta) d\theta$ is unknown, MCMC algorithms are often used. These algorithms generate dependent samples using

$$\pi(\theta | x) \propto \pi(\theta) f(x | \theta), \quad (3.2)$$

which differs from (3.1) only by a constant. Hence, only the prior and likelihood need to be defined, which is fairly straightforward. Many algorithms exist to obtain samples from (3.2), including the Metropolis Hastings (MH) algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953), the Gibbs sampler (a special case of MH proposed by S. Geman and

D. Geman (1984)), and slice sampling (Neal, 2003).

There are many software packages that implement various MCMC algorithms, such as BUGS (Bayesian inference Under Gibbs Sampling), which led to the popular WinBUGS and OpenBUGS projects (Lunn, Thomas, Best, & Spiegelhalter, 2000; Spiegelhalter, Thomas, Best, & Lunn, 2014). Another program, which uses a very similar model description language as BUGS, is JAGS (Just Another Gibbs Sampler), written by Martyn Plummer (Plummer, 2003). Stan (Stan Development Team, 2014) is a program that implements the No-U-Turn sampler, proposed by Hoffman and Gelman (2014), which is an extension of Hybrid or Hamiltonian Monte Carlo (Duane, Kennedy, Pendleton, & Roweth, 1987; Neal, 1993).

We explored both Stan and JAGS, however decided to use JAGS as it performed at least as well as Stan in most of our initial models, and the modelling language is more familiar and simpler than Stan's. This will hopefully make the models presented in Appendix B easier to understand, and more importantly easier to modify for other applications. We used JAGS version 3.4.0 to fit the models in this thesis. R, a statistical computing language and environment, was used for data manipulation and presentation, a back-end to JAGS (using additional packages), and for post-processing the posterior samples (R Core Team, 2014). Additionally, we started development of the R package BSM, which acts as a wrapper for the majority of the data preparation, fitting, and summary steps (Chapter 7).

3.1.2 MCMC Diagnostic Checks

MCMC algorithms iteratively generate dependent samples from the target distribution, producing a *chain* of samples. These chains must be exposed to diagnostic checks to ensure that they are samples from the posterior distribution. To do this, various properties of the samples, such as *stationarity* and *convergence*, are assessed (Geweke, 1992; Raftery & Lewis, 1992; Gelman & Rubin, 1992). Stationarity implies that the distribution of the chain is constant over time (iterations), while convergence indicates that the observed stationary distribution is likely that of the target (posterior) distribution.

The algorithms used to generate MCMC samples, discussed in § 3.1.1, ensure that the

chain has the required stationary distribution defined in (3.1), and will converge to this distribution with probability tending to 1 (Tierney, 1994). This is known as the *ergodic theorem*, and allows the inference of properties of the posterior distribution using the MCMC samples, such as the expected value:

$$\mathbb{E}[\theta] = \int_{\theta} \theta \pi(\theta | X) \approx \frac{1}{n} \sum_{i=1}^n \theta_i. \quad (3.3)$$

Here, we briefly explore some of the diagnostic tools available to assess stationarity and converge.

As we used R to process the MCMC samples, we had access to the `coda` package, which provides a suite of functions for MCMC diagnostic checks (Plummer, Best, Cowles, & Vines, 2006). We discuss the Gelman and Rubin test statistics for checking convergence, and briefly mention a few other diagnostic tests implemented in this package.

Gelman and Rubin (1992) proposed a diagnostic check to examine the *convergence* of the MCMC chains. MCMC algorithms take a number of iterations to reach a stationary distribution, at which point they are said to have converged. This initial period is known as the *burn in* phase, and can bias the results unless it is discarded (Gelman et al., 2014). The Gelman and Rubin test statistic is used to assess the convergence of MCMC chains (Gelman & Rubin, 1992). This statistic is based on the potential scale reduction factor (PSRF), an estimate of how much the scale of the (conservative) posterior distribution will shrink if the chain is left to run for infinitely many steps. A chain is considered to have converged if the estimate of PSRF, \hat{R} , is close to 1, where \hat{R} is estimated by comparing the variance of $m \geq 2$ independent chains to assess whether or not they have the same distribution. Hence $m = 3$ chains were used in all of the models fitted during our work (which is also the default in JAGS).

From the CODA package, the `gelman.diag()` function is used to compute the PSRF for the parameters. The output of `gelman.diag()` also includes an additional 95% upper credible interval estimate of PSRF, as well as the multivariate potential scale reduction factor

(MPSRF), discussed in Brooks and Gelman (1998), which assesses the convergence of the entire chain, rather than individual parameters.

Other diagnostic tests for convergence and stationarity include those presented by Heidelberger and Welch (1983) and Geweke (1992), which compare the samples from various parts of a single chain to determine if the chains have converged and are therefore stationary. In R, using the `coda` package, these can be accessed by the `heidel.test()` and `geweke.test()` functions respectively.

3.1.3 Hierarchical Bayesian Models

One type of model used extensively in our work was the *hierarchical* Bayesian model, similar to the mixed-effects models used in frequentist statistics (Hill, 1965; Tiao & Tan, 1965). In a hierarchical model, instead of assuming common parameters for each group (i.e., haul), the parameters are allowed to vary according to a defined distribution—this is often a normal distribution with unknown mean, μ , and variance, σ^2 (Gelman et al., 2014). This is known as a *hierarchical prior*.

A hierarchical prior on a parameter, for example L_{50} , allows it to vary between hauls according to a population distribution of L_{50} 's. That is, the hierarchical prior for the L_{50} parameter of haul h is

$$L_{50}^h \sim N(\mu_{L_{50}}, \sigma_{L_{50}}^2), \quad (3.4)$$

where $\mu_{L_{50}}$ and $\sigma_{L_{50}}^2$ are the population mean and variance of L_{50} , respectively, that describe the population distribution of L_{50} for all possible hauls (not just those observed in the study). $\mu_{L_{50}}$ and $\sigma_{L_{50}}^2$ are given their own *hyperprior* distributions. These can contain prior information if any is available, otherwise noninformative hyperpriors are typically used. Fisheries researchers are especially interested in $\sigma_{L_{50}}^2$, which represents between-haul variability of L_{50} .

3.2 A Single-gear Covered-codend Trawl

The simplest experimental trawl to analyse is a covered-codend experiment using a single gear, often used to estimate selectivity for use in other studies. One such data set, used in the analysis presented by Millar et al. (2004), comes from the school prawn (*Metapenaeus macleayi*) fishery. Millar et al. (2004) apply the two approaches to modelling the size-selectivity of trawl gear presented in § 2.1.3—combined-hauls and mixed-effects—to these data. We will be making comparisons of the results obtained from our models to those presented in their paper. The first few rows of the data set are shown below, after having been read into R:

```
head(school.cc)
```

	codend	cover	WtSch	q1	q2	total	Lenclass	Haul
## 1	0	0	2.02	0.585	0.34	0	6	1
## 2	0	0	2.02	0.585	0.34	0	7	1
## 3	0	0	2.02	0.585	0.34	0	8	1
## 4	0	0	2.02	0.585	0.34	0	9	1
## 5	0	0	2.02	0.585	0.34	0	10	1
## 6	0	2	2.02	0.585	0.34	2	11	1

The variables used were: `codend`, the codend count; `cover`, the cover count; `WtSch`, the catch weight (kg) of the haul; `q1` and `q2`, the codend and cover sampling fractions, respectively; `total`, the total count for codend and cover; `Lenclass`, the prawn length (mm); and `Haul`, the haul identification variable. A plot of the scaled data, \mathbf{p}^* , is shown in Figure 3.1, where *scaled* refers to adjusting the counts for subsampling (see § 2.1.2). The plot shows a clear increasing trend in retention probability with length, with most of the change occurring between 10 mm and 20 mm.

The combined-hauls analysis of these data by Millar et al. (2004) estimated L_{50} and SR to be 16.17 mm and 5.17 mm, respectively, with the standard errors (adjusted by REP; see (2.16) on page 16) of 0.20 and 0.85, respectively. They used the SELECT method and a Richards' selection curve, although no value for δ was presented with their results. For the mixed-effects model, random effects on L_{50} alone was preferred (by log-likelihood) over random effects on both L_{50} and SR . The estimated parameters and standard errors were

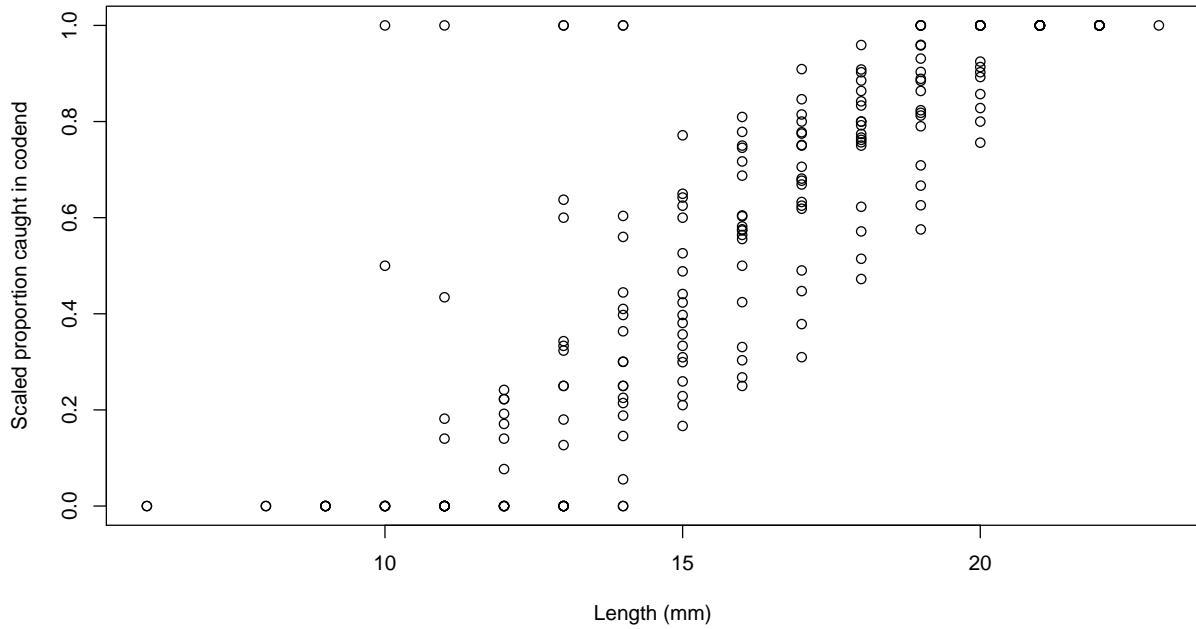


Figure 3.1: The scaled covered-codend school prawn data. The plotted points represent the estimated proportions of shrimp caught in the codend, $\hat{p}_\ell^h = \frac{\hat{y}_\ell^h}{\hat{n}_\ell^h}$.

$\hat{\mu}_{L_{50}} = 15.52$ (0.31), $\hat{\mu}_{SR} = 4.02$ (0.16), and $\hat{\sigma}_{L_{50}}^2 = 1.66$ (0.57). Millar et al. (2004) found that the logistic curve provided an adequate fit to the individual haul data, so the Richards' curve was not explored in their mixed-effects approach. The notable difference between the results of the two approaches is that L_{50} and SR are smaller in the mixed-effects model than in the combined-hauls model. As discussed in § 2.1.3, this is because the combined-hauls approach is a marginal model, while the mixed-effects model is conditional on the observed hauls, and thus, each haul gives equal weight to $\mu_{L_{50}}$ and μ_{SR} .

In this section, we explore both the combined-hauls and a hierarchical approach to modelling the covered-codend school prawn data. Overdispersion is not detectable in the combined data; however, in § 4.1 we see how it can be assessed and modelled when the data are not stacked.

3.2.1 Combined-hauls Model

To formulate a Bayesian model, we needed to define the *prior* and *likelihood*, as discussed in § 3.1. For the likelihood, we used a binomial distribution, defined in (2.6). The binomial probability was calculated by (2.14), using a Richards' selection curve for $r(\ell)$. Note that the sampling fractions were incorporated by scaling the counts, so $Q = 1$ is used in (2.14). The benefit of retaining the sampling fractions in the model shown in JAGS Model B.1.1 of the Appendix on page 114, is that it allows the model to be applied easily to unscaled data, which we see in § 4.1.

For the prior distributions, various alternatives were explored, but had negligible influence on the posterior unless they were highly informative. Hence, we used a noninformative normal distribution, truncated to \mathbb{R}^+ , for the L_{50} parameter, and a noninformative lognormal distribution for SR . The variance parameters for the distributions are parameterised in JAGS as precision, $\tau = \frac{1}{\sigma^2}$. For δ , a lognormal distribution was also used. These priors can be seen in JAGS Model B.1.1 on page 114.

The data were manipulated into matrix form, with individual hauls in columns, and length classes in rows. Therefore, the codend count for length ℓ in haul h is y_{lh} , and n_{lh} for the total count (codend + cover). In JAGS, subscripts i and j (the traditional notation for matrices) are used to denote length and haul, respectively; for the combined-hauls approach, the matrices are column-vectors of the scaled, stacked counts. Because scaling the counts typically gives non-integer values, they were rounded before passing them to JAGS, otherwise the binomial is undefined.

The data were passed from R to JAGS as a list of variables. The counts for codend and total were passed as \mathbf{Y} and \mathbf{N} , which were both $N \times M$ matrices; N is the number of length classes, and M is the number of hauls (in the combined-hauls approach, $M = 1$). N and M were also included in the data list passed to JAGS, along with \mathbf{x} , the length- N vector of length classes. Although redundant, the sampling fractions were included as length- M vectors containing the sampling fractions for each haul, $q1 = q2 = \mathbf{1}_M$.

Table 3.1: Summary of the posterior distributions of the selectivity parameters of the combined-hauls model fitted to the covered-codend school prawn data. For comparability of standard errors, SD_{adj} is the standard deviation multiplied by $\sqrt{REP} = 3.14$.

Parameter	Mean (SD)	SD_{adj}	95% Credible Interval	Millar et al. (2004)
L_{50}	16.16 (0.07)	0.21	16.03 – 16.29	16.17 (0.20)
SR	5.23 (0.30)	0.96	4.70 – 5.89	5.17 (0.85)
δ	3.62 (1.31)	4.12	1.80 – 6.78	

Results and comparison

JAGS Model B.1.1 was fitted using the `BSM` package (Chapter 7), which uses the `jags()` function from the `R2jags` package (Su & Yajima, 2014). 3 chains were used, each with 100,000 iterations consisting of a burn in period of 50,000 and thinning interval of 10, yielding a total posterior sample size of 15,000. The Gelman & Rubin and Heidelberger & Welch tests both passed, indicating convergence of the MCMC chains (see § 3.1.2). Trace plots and density plots of the posterior samples shown in Figure 3.2 show fast mixing and convergence of the chains.

The correlation between the model parameters is weak, $\text{Cor}(L_{50}, SR) = -0.14$, which is demonstrated in Figure 3.3. On the other hand, the correlation between β_0 and β_1 is extreme, $\text{Cor}(\beta_0, \beta_1) = -0.998$. Such strong correlation would cause severe problems for the MCMC sampler if it had to sample these parameters; a long burn in period and large thinning interval would be required to pass the MCMC diagnostic tests.

The posterior summary of the three parameters, L_{50} , SR and δ , is shown in Table 3.1. The posterior mean estimate of L_{50} is almost identical to the estimate from Millar et al. (2004), while there is a small, but not significant difference in SR . For the standard errors, the sample standard deviations are reported, although we also computed REP using (2.16) and the posterior mean estimates to obtain adjusted errors $SD_{adj} = SD\sqrt{REP}$. These adjusted errors are comparable to those from Millar et al. (2004), although there is a slight difference in the errors reported for SR . Figure 3.5 on page 35 shows the data overlaid with the retention curve estimated using the posterior means of the selectivity parameters.

The combined-hauls approach is fairly intuitive and easy to implement. The main com-

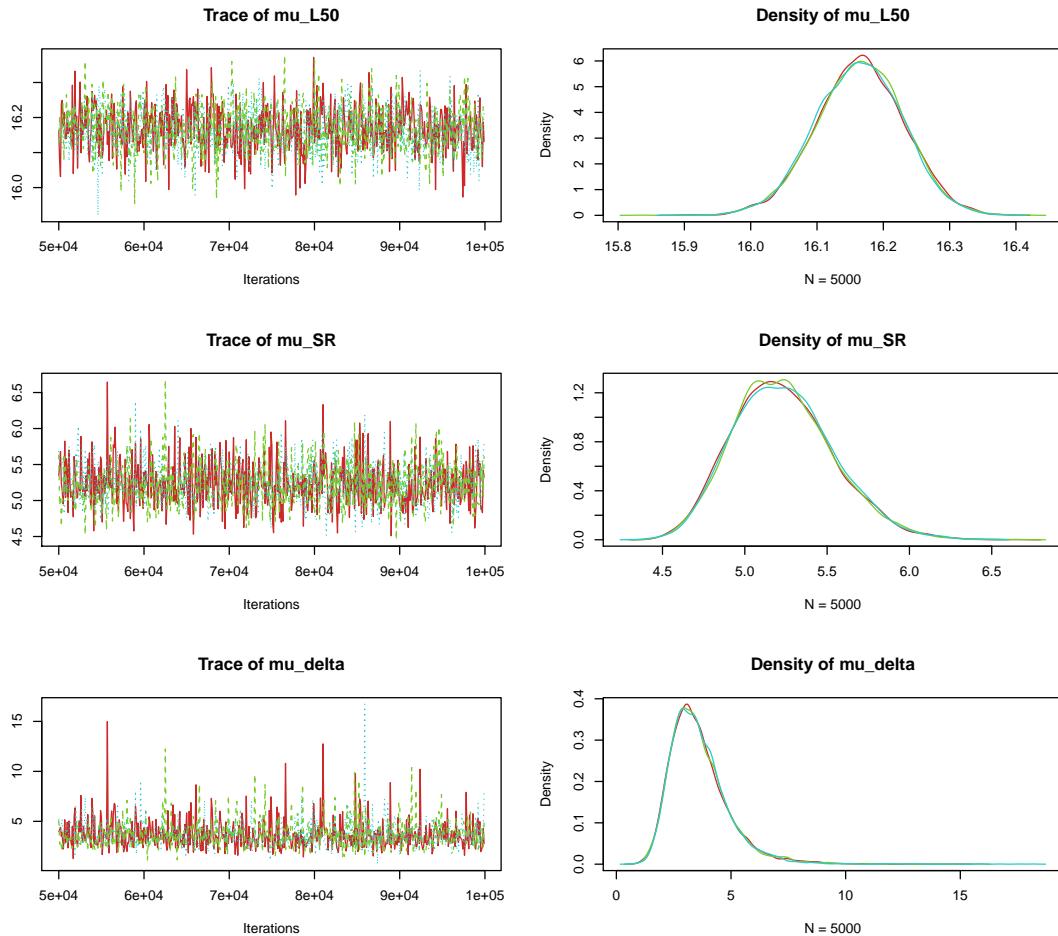


Figure 3.2: Trace plots (left) and density plots (right) of the posterior samples of the selectivity parameters of the combined-hauls model fitted to the covered-codend school prawn data.

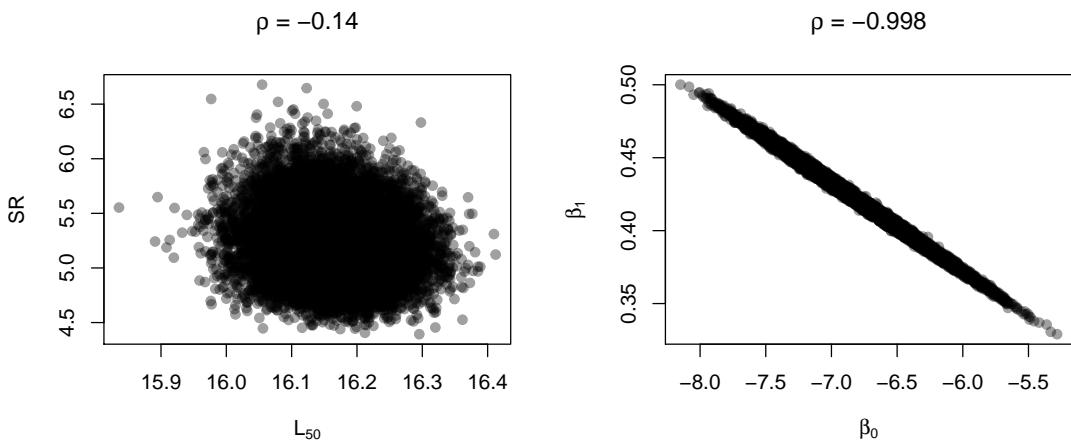


Figure 3.3: Posterior correlations of two alternative parameterisations of the logistic selection curve. There is almost no correlation between L_{50} and SR (left), while the correlation between β_0 and β_1 is strongly negative (right).

plication is the method for dealing with overdispersion. While REP offers a post-fit catch-all adjustment, it is typically not the preferred method in a Bayesian analysis. We would prefer to directly model the overdispersion in the unstacked data. We explore various possibilities for doing this in Chapter 4.

3.2.2 Hierarchical Model

We have seen that a Bayesian model can be used to estimate the same selection curve as a standard maximum-likelihood analysis by using a combined-hauls approach. We now explore the alternate approach to modelling these data: a random-effects model to directly incorporate between-haul variability, which will be especially useful when comparing the selectivity of different gears. This is done using a hierarchical Bayesian model (§ 3.1).

In a hierarchical model, instead of using a single selection curve for all hauls, it is allowed to vary between hauls. The retention probability for an animal of length ℓ is therefore defined for haul h as

$$r^h(\ell) = \text{logit}^{-1}(\eta_l^h), \quad \eta_l^h = \frac{2 \log(3)}{SR^h} (\ell - L_{50}^h). \quad (3.5)$$

Here, a hierarchical prior is given to both L_{50} and SR . Millar et al. (2004) used a bivariate normal distribution for L_{50} and SR , but we used independent priors; posterior correlation between the parameters could be estimated from the samples (Figure 3.3). Millar et al. (2004) used random effects only on L_{50} , as this gave a better predictive model than the model with random effects for both L_{50} and SR . Hence we present the model with a hierarchical prior on L_{50} , and retain a common SR . In Chapter 5 we look at models including random effects on SR , and decide for ourselves if it is necessary. We also need a prior distribution on $\sigma_{L_{50}}^2$, which was the noninformative uniform prior on the standard deviation proposed by Gelman (2006):

$$\sigma_{L_{50}} \sim U[0, 100]. \quad (3.6)$$

Table 3.2: Summary of the posterior distributions of the selectivity parameters of the hierarchical model fitted to the covered-codend school prawn data.

Parameter	Mean (SD)	Median	95% Credible Interval	Millar et al. (2004)
$\mu_{L_{50}}$	15.51 (0.35)	15.52	14.81 – 16.20	15.52 (0.31)
μ_{SR}	4.04 (0.16)	4.03	3.75 – 4.35	4.02 (0.16)
$\sigma_{L_{50}}^2$	2.14 (0.90)	1.94	1.02 – 4.38	1.66 (0.57)

Results and comparison

JAGS Model B.2.1 was fitted using 100,000 iterations consisting of a burn in period of 50,000 and thinning interval of 10, yielding a total posterior sample size of 15,000. Trace plots and density plots of the posterior samples are displayed in Figure 3.4; again, we see the MCMC chains are mixing well and appear stationary. MCMC diagnostic checks for convergence and stationarity passed. The results are displayed in Table 3.2, and are very similar to those published by Millar et al. (2004). Both the means and errors are the same for $\mu_{L_{50}}$ and μ_{SR} . The estimates of $\sigma_{L_{50}}^2$ are different, though the posterior median is closer to the frequentist estimate than the posterior mean; however, the errors associated with $\sigma_{L_{50}}^2$ are large and the difference between the estimates is not significantly different. The fitted curve using the mean parameter estimates is shown in Figure 3.5, along with the combined-hauls fit from § 3.2.1.

In this section, we demonstrated that the combined-hauls and mixed-effects approaches can be fit using Bayesian models and give similar estimates of selectivity as obtained from the traditional methods. While formulating the models used in this section, our goal was for flexibility and extensibility, which we exploited throughout the rest of our work.

3.3 A Covered-codend Trawl with Covariates

In § 3.2, we modelled data from a covered-codend experimental trawl that used only a single gear. In many situations, however, these types of experiments are carried out with a range of different gears in order to compare their selectivity properties. An example of this is the

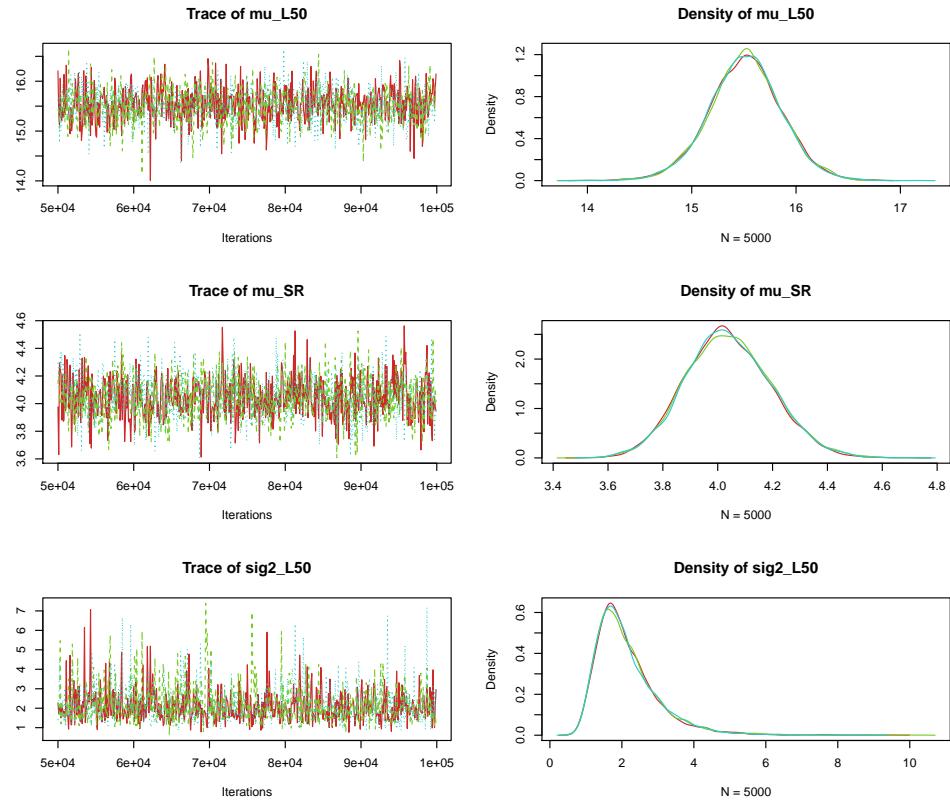


Figure 3.4: Trace plots (left) and density plots (right) of the posterior samples of the selectivity parameters of the hierarchical model fitted to the covered-codend school prawn data.

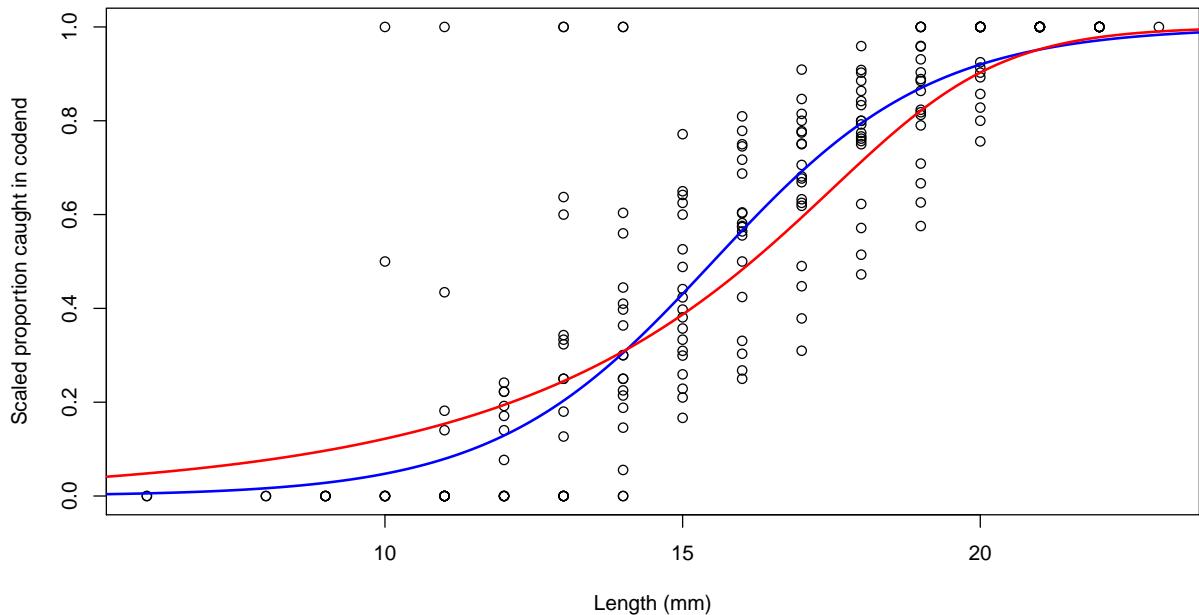


Figure 3.5: The scaled covered-codend school prawn data, with the combined-hauls (red) and hierarchical (blue) models' estimated (mean) selection curves.

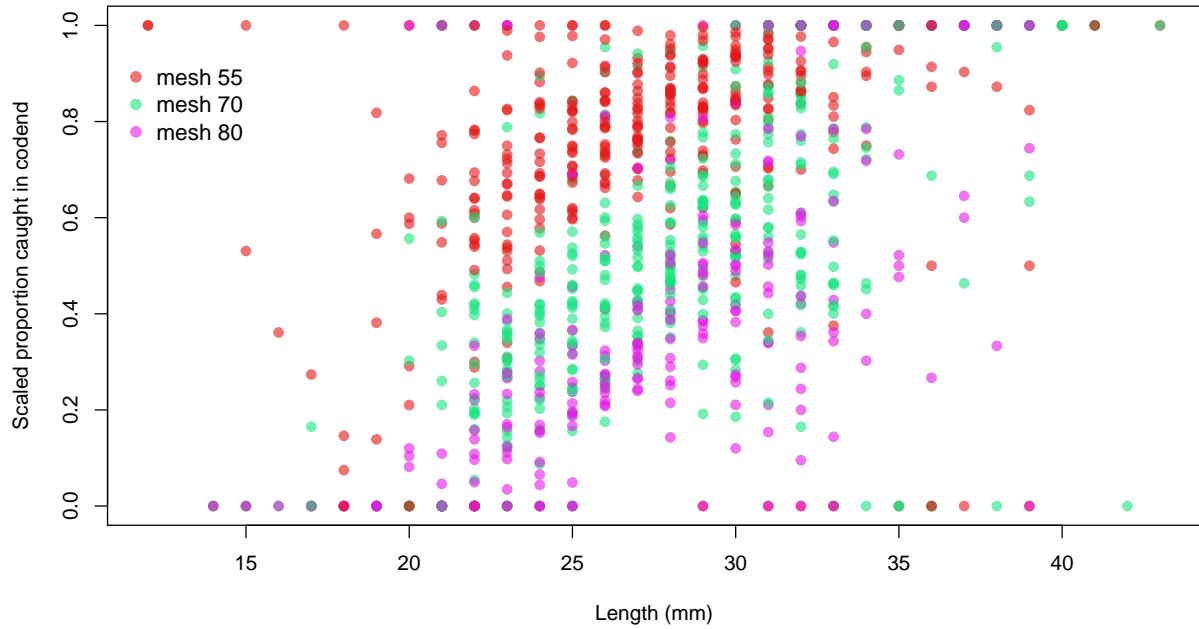


Figure 3.6: The scaled covered-codend rose shrimp data. Points have been coloured according to mesh size.

covered-codend experiment in the Portuguese rose shrimp fishery, analysed by Fonseca et al. (2007). The experiment consisted of three separate cruises in autumn 1998, spring 1999 and summer 1999; the autumn cruise used a different mesh material than the other two. In all of the cruises, mesh sizes of 55, 70 and 80 mm were used.

The variables collected from the rose shrimp study, defined in Fonseca et al. (2007), were: **Cruise**: the “cruise” identifier (1 = autumn 1998, 2 = spring 1999, 3 = summer 1999); **Mesh**: the mesh size (55, 70, and 80 mm); **Haul**: the haul identification variable; **Length**: the rose shrimp carapace length (mm); **CodCatch**: the codend count; **TotCatch**: the total count (codend + cover); **q1** and **q2**: the sampling fractions for codend and cover, respectively; **Depth**: the depth (m) of the trawl (note that these were bottom-trawls); **CodTwine**: the type of twine used (0 = 3.0 mm single braided polyamide twine used in cruise 1, or 1 = 4.5 mm single braided polyethylene twine used in cruises 2 and 3). The scaled data are displayed in Figure 3.6; the colour coding by mesh size suggests a relationship between mesh size and retention probability.

Fonseca et al. (2007) used mixed-effects models, fitted using PROC NLMIXED in SAS, and modelled L_{50} with random haul effects, although convergence issues were experienced when both L_{50} and SR were modelled with random effects. They fitted several different models, and found the best predictive model used cruise and mesh size to model L_{50} ; coefficients for the variables depth and catch weight were found to be non-significant. The linear form of the L_{50} parameter for haul h in cruise $i = \{1, 2, 3\}$ using mesh $j = \{1, 2, 3\}$ (for a mesh size, m_j , of 55, 70 and 80 mm, respectively), was given as

$$L_{50ij}^h = \alpha_i + \beta m_j + \epsilon_{ij}^h, \quad \epsilon_{ij}^h \sim N(0, \sigma_{L_{50}}^2). \quad (3.7)$$

Fonseca et al. (2007) found that α_2 and α_3 were not significantly different from each other, so the two cruises (spring and summer, 1999) were modelled with a common effect, $\alpha_{2&3}$. α_1 was not found not to be significantly different from 0, implying that L_{50} is proportional to mesh size for the Autumn cruise, and hence they fixed $\alpha_1 = 0$. This effect is known as geometric similarity (Millar & Fryer, 1999; Fonseca et al., 2007). Mesh size was not found to have a significant effect on SR , so it was modelled with a common value, μ_{SR} , for all gears.

In this section, the Bayesian implementation of the model fitted by Fonseca et al. (2007) is examined and compared to the results reported in their paper. In Chapter 4, we explore various other models originally proposed in their paper, which included weight and depth as covariates for L_{50} , and modelling of both L_{50} and SR with random effects.

Formulating the model

The model fitted to the rose shrimp data was very similar to the hierarchical model used in § 3.2.2 (JAGS Model B.1.1), which was used to model the covered-codend school prawn data. The main difference is that there is no longer a $\mu_{L_{50}}$; instead, we implemented (3.7) at the node `L50_bar`, as shown in JAGS Model B.2.2. The other adjustment to the model was the implementation of appropriate priors on the coefficients $\alpha_{2&3}$ and β ; these were given noninformative normal distributions. The list of data passed from R to JAGS was the

Table 3.3: Summary of the posterior distributions of the parameters of the hierarchical model fitted to the covered-coded rose shrimp data.

Parameter	Mean (SD)	Median	95% Credible Interval	Fonseca et al. (2007)
μ_{SR}	11.37 (0.32)	11.36	10.77 – 12.03	11.32 (0.32)
β	0.41 (0.01)	0.41	0.39 – 0.43	0.41 (0.01)
$\alpha_{2\&3}$	-3.05 (0.88)	-3.05	-4.82 – -1.34	-3.03 (0.85)
$\sigma_{L_{50}}^2$	12.77 (2.44)	12.52	8.79 – 18.28	11.72 (2.17)

Table 3.4: Summary of the posterior distributions of L_{50} for each gear used in the covered-codend rose shrimp experiment. The model assumed a common μ_{SR} , and a linear effect of mesh size on L_{50} .

Cruise	Mesh (mm)	L_{50ij} (SD)	95% Credible Interval	Fonseca et al. (2007)
Autumn	55	22.38 (0.59)	21.24 – 23.53	22.38 (0.57)
	70	28.48 (0.75)	27.03 – 29.95	28.48 (0.73)
	80	32.55 (0.86)	30.89 – 34.23	32.55 (0.83)
Summer/Spring	55	19.33 (0.54)	18.28 – 20.37	19.34 (0.52)
	70	25.43 (0.53)	24.41 – 26.46	25.44 (0.50)
	80	29.50 (0.55)	28.42 – 30.57	29.51 (0.52)

same as described in § 3.2, with the additional length- M vectors `cruise`, a binary vector identifying the spring and summer hauls, and `mesh`, the mesh size for each haul.

Results and comparisons

The model was fitted using 3 chains of 50,000 iterations, comprising a burn in period of 25,000, and thinning interval of 10. This gave a total posterior sample size of 7,500. Figure 3.7 shows the chains mixing quickly, and each parameter appears to have converged to the same stationary distribution. MCMC diagnostic tests for stationarity and convergence passed.

Table 3.3 summarises the posterior distributions of the model parameters, which match very closely with the results reported by Fonseca et al. (2007). Again, however, the Bayesian estimate of $\sigma_{L_{50}}^2$ is slightly larger with more variability, although this difference is not significant. Also, the distribution of $\sigma_{L_{50}}^2$ is not symmetrical; the maximum likelihood estimate (MLE) obtained in a frequentist analysis would be closer to the posterior *mode*, which is approximately 12.10.

Having estimated the model parameters, the selectivity of the different gears can be

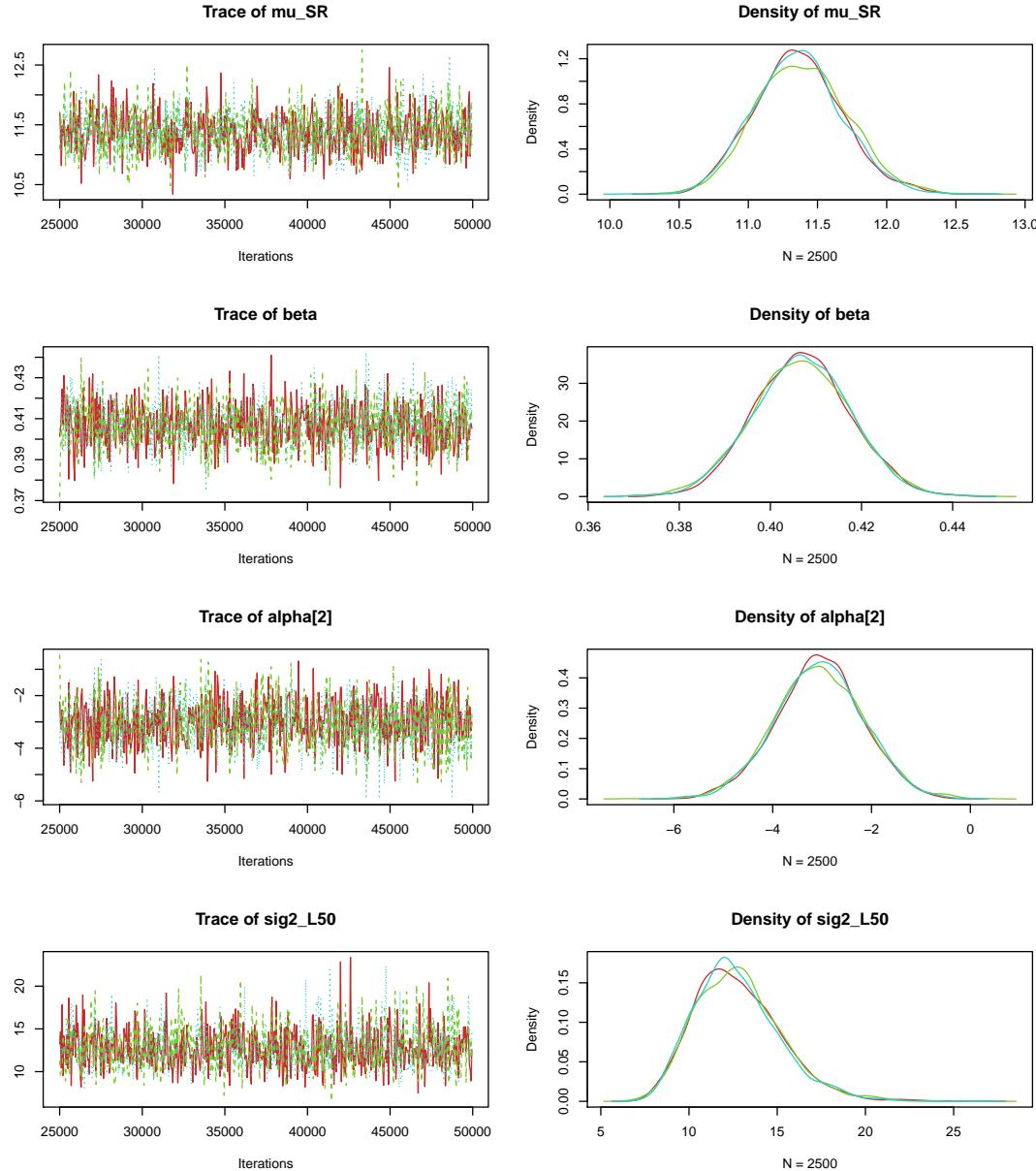


Figure 3.7: Trace plots (left) and density plots (right) of the posterior samples of the parameters used in the hierarchical model fitted to the covered-codend rose shrimp data.

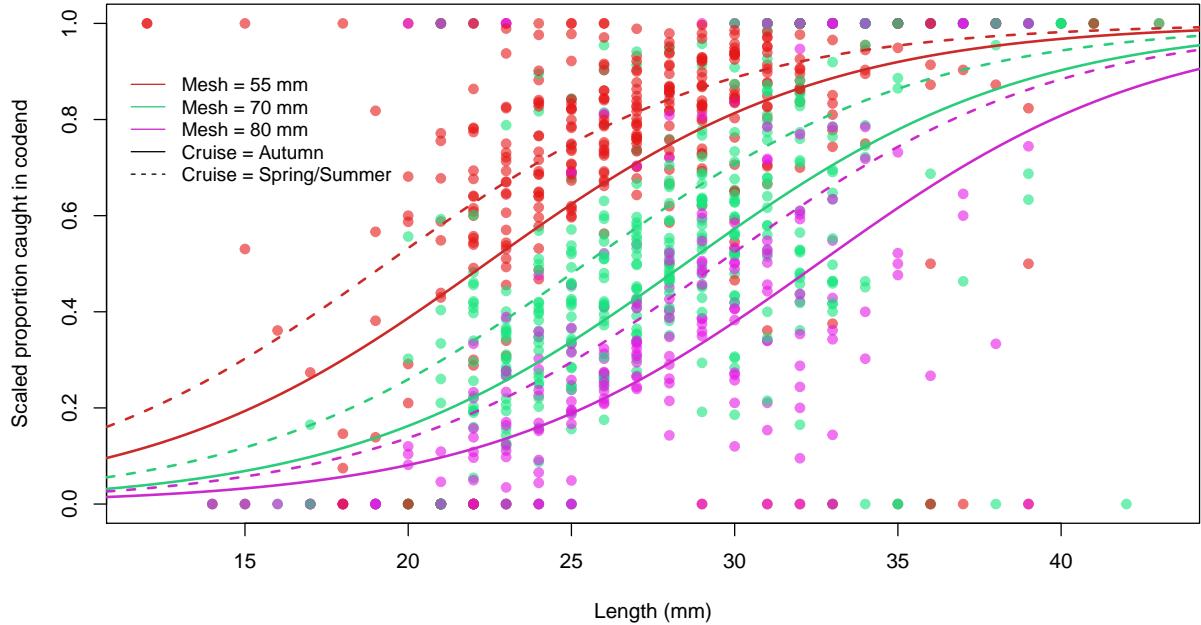


Figure 3.8: Scaled covered-codend rose shrimp data with the (mean) estimated logistic selection curves from the hierarchical model.

quantified. The major co-variate of interest is mesh size, with gears using 55, 70 and 80 mm meshes in the experimental hauls. They were also carried out over three seasons: autumn, spring and summer. However, as discussed in Fonseca et al. (2007), confounding occurs because the autumn cruise used polyamide twine, while the others used polyethylene. Because Fonseca et al. (2007) found no significant difference between the spring and summer cruises, they were modelled with a common effect, $\alpha_{2\&3}$. Hence there are six combinations of gear. Using the MCMC samples, posterior distributions of L_{50} were computed for each gear using (3.7), from which the results are shown in Table 3.4. The Bayesian estimates are all the same as those given by Fonseca et al. (2007), with any differences being of the order 0.01—negligible in comparison to the estimates' standard errors.

The retention curves drawn in Figure 3.8 were computed from the posterior mean estimates of L_{50} (Table 3.4) and μ_{SR} . They appear to fit the bulk of the data, although it is worth noting that there is considerable variation around these estimated curves, evident from the large estimate of between-haul variability in L_{50} , $\sigma_{L_{50}}^2$. Another cause of this variability

may be the small sampling fractions, many of which are less than 5%, a consequence of the large quantity of rose shrimp in each catch (Fonseca et al., 2007). Despite the variability, it is evident from the results that there is a significant effect of mesh size on the size-selectivity of the gear, and an effect due to season, twine type, or a combination of both. Again, the Bayesian hierarchical model fitted to these data obtained very similar results to those from the frequentist GNLM used by Fonseca et al. (2007).

3.4 A Twin-trawl

We have now covered several types of models used for analysing covered-codend experiments, including both the combined-hauls and mixed-effects approaches. However, some experimental trawls use a paired-trawls approach. While there are several variants of this, we specifically look at data from a twin-trawl experiment in the school prawn fishery by Broadhurst et al. (2004), which took place in Lake Woolooweyah, New South Wales. The study involved three experimental gears: 40 mm knotted diamond-mesh, and 20 mm square mesh with either a tapered or untapered codend. Millar et al. (2004) applied the combined-hauls and mixed-effects approaches to model these data.

The variables in the data set used by Millar et al. (2004) were: `Haul`, the haul identification variable; `Length`, the length class variable (mm); `Expt` and `Control`, the experimental and control codend counts, respectively; `q1` and `q2`; the sampling fractions for the experimental and control codends, respectively; `Mesh`, the mesh used ($1 = 40$ mm knotted diamond-mesh, $2 = 20$ mm tapered square mesh, $3 = 20$ mm untapered square mesh). Figure 3.9 displays the scaled data. In this case, the proportion of school prawn caught in the experimental codend, out of the total number of school prawn caught in the haul, is plotted on the y -axis.

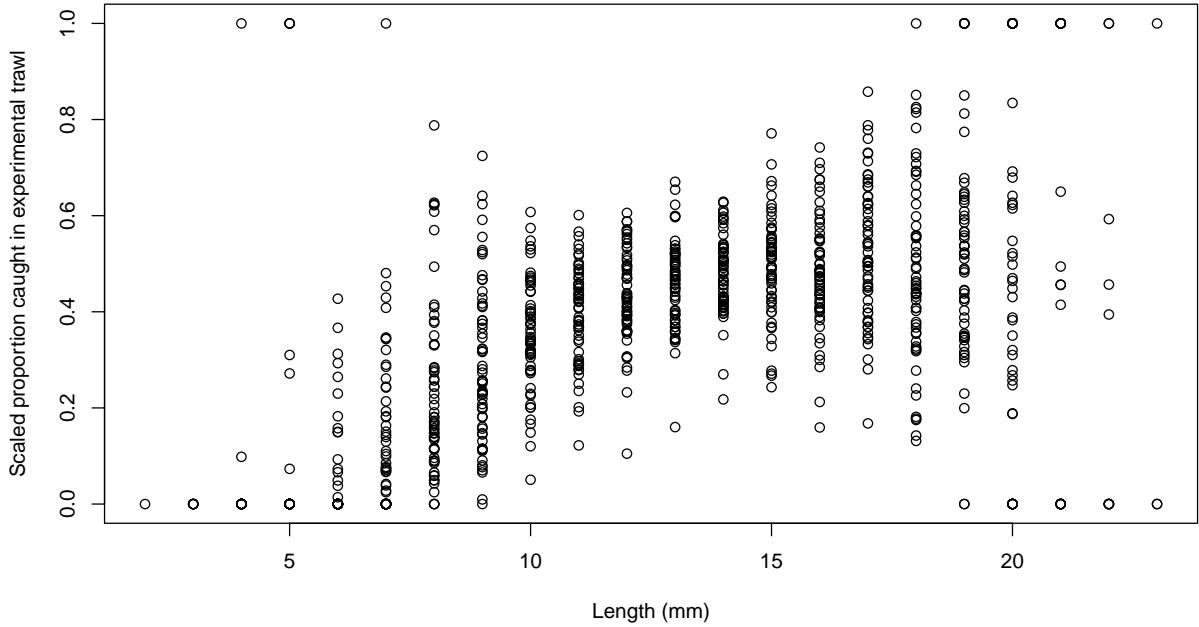


Figure 3.9: The scaled twin-trawl school prawn data. Note that the proportions are now the proportion caught in the experimental codend, which we expect to approach approximately 0.5 as retention probability approaches 1.

3.4.1 Combined-hauls Model

In the combined-hauls approach, Millar et al. (2004) fitted three separate selection curves to the data—one for each of the three gears—obtaining a unique estimate of both L_{50} and SR for each gear. They used the combined data for each gear, and describe how REP was calculated to adjust the standard errors to account for overdispersion and sample size inflation caused by scaling the counts. The Richards' curve was not found to provide a significantly better fit to the data, so a logistic selection curve was assumed.

In our Bayesian approach, each gear was modelled separately using a binomial likelihood with probability defined by (2.22), and assuming $r(\ell)$ to be logistic. The model is otherwise identical to JAGS Model B.1.1 used for the covered-codend data (§ 3.2.1), with the additional estimation of the parameter ϕ . For this, we used a noninformative normal density on $\text{logit}(\phi)$, ensuring that the relative fishing efficiency parameter is a valid probability $\phi \in (0, 1)$. The model is displayed in JAGS Model B.1.2.

Table 3.5: Summary of the posterior distributions of the selectivity parameters for each of the three gears in the twin-trawl school prawn data. For comparability to the frequentist model, the standard deviations were adjusted, SD_{adj} , by multiplying by \sqrt{REP} .

Gear	Parameter	Mean (SD)	SD_{adj}	95% Credible Interval	Millar et al. (2004)
40 mm diamond (REP = 8.38)	L_{50}	8.55 (0.09)	0.26	8.37 – 8.73	8.56 (0.26)
	SR	3.88 (0.21)	0.60	3.51 – 4.33	3.88 (0.59)
20 mm tapered square (REP = 10.52)	L_{50}	10.10 (0.07)	0.22	9.97 – 10.23	10.10 (0.21)
	SR	3.16 (0.08)	0.27	3.00 – 3.33	3.15 (0.27)
20 mm uptapered square (REP = 8.47)	L_{50}	10.27 (0.08)	0.24	10.12 – 10.44	10.28 (0.24)
	SR	3.49 (0.12)	0.35	3.27 – 3.74	3.49 (0.34)

Results and comparison

The model was fitted to each gear separately, each using 3 chains of 20,000 iterations, comprising a burn in period of 10,000 and thinning interval of 5, yielding a total of 6,000 samples. Stationarity and convergence tests of the MCMC samples passed the diagnostic tests.

The posterior sample is summarised in Table 3.5, from which we can see that the mean estimates of L_{50} and SR are almost identical to those obtained by Millar et al. (2004) for each of the three gears. The standard deviations of the parameters were adjusted by REP, using the formulae given in Millar et al. (2004), giving us errors comparable to those presented in their paper. From the estimated curves drawn in Figure 3.10, we can distinctly see that the selectivity of the diamond mesh gear is different from the selectivity of the square mesh gears, which are themselves fairly similar.

3.4.2 Hierarchical Model

The mixed-effects model fitted to the twin-trawl data by Millar et al. (2004) modelled L_{50} and ϕ with random haul effects. ϕ was modelled on the logit scale using (2.23) on page 19 to ensure valid probabilities $\phi^h \in (0, 1)$ were obtained. A common effect for the two 20 mm square meshes was used, as they were not found to be significantly different in previous studies (Broadhurst et al., 2004; Millar et al., 2004). In the Bayesian model, a baseline-factor approach was used to model the gear effect. That is, we define m_i to be a binary

Table 3.6: Summary of the posterior distributions of the selectivity parameters of the hierarchical model fitted to the twin-trawl school prawn data.

Gear	Parameters	Mean (SD)	Median	95% Credible Interval	Millar et al. (2004)
40 mm diamond	$\mu_{L_{50}}$	8.05 (0.43)	8.05	7.18 – 8.87	8.08 (0.40)
Both 20 mm square	$\mu_{L_{50}}$	10.23 (0.28)	10.23	9.67 – 10.78	10.23 (0.27)
All gears	μ_{SR}	3.51 (0.17)	3.50	3.18 – 3.86	3.48 (0.17)
	μ_ϕ	0.0044 (0.0356)	0.0047	-0.0659 – 0.0735	0.0025 (0.034)
	$\sigma_{L_{50}}^2$	2.36 (0.66)	2.26	1.36 – 3.94	2.02 (0.53)
	σ_ϕ^2	0.04 (0.01)	0.04	0.02 – 0.07	0.038 (0.01)

variable to indicate the use of either of the square mesh gears (tapered or untapered), thus the expected L_{50} for a haul using gear i is

$$L_{50i} = \mu_{L_{50}} + \beta m_i. \quad (3.8)$$

where β is the effect of the square mesh on L_{50} .

Results and comparison

A Bayesian hierarchical implementation (JAGS Model B.2.3) of the model presented by Millar et al. (2004) was fitted using 3 chains of 30,000, comprising a burn in of 15,000 and thinning interval of 5, giving a total of 9,000 samples. Again, MCMC diagnostic tests for stationarity and convergence passed.

Table 3.6 summarises the posterior samples from the fitted model. The Bayesian mean estimates of L_{50} are fairly similar to those from Millar et al. (2004), with the same associated errors. The Bayesian estimate of μ_{SR} is slightly larger, but the difference is not significant based on the standard error. As we have seen in the previous models, the Bayesian estimate of $\sigma_{L_{50}}^2$ is slightly larger than the frequentist one, although it remains within one standard error of the estimate from Millar et al. (2004). For the relative fishing efficiency parameter, the mean estimate, μ_ϕ , and associated standard deviation are not significantly different from those originally obtained by SAS. Figure 3.11 displays the curves estimated for each of the gears (using the posterior means).

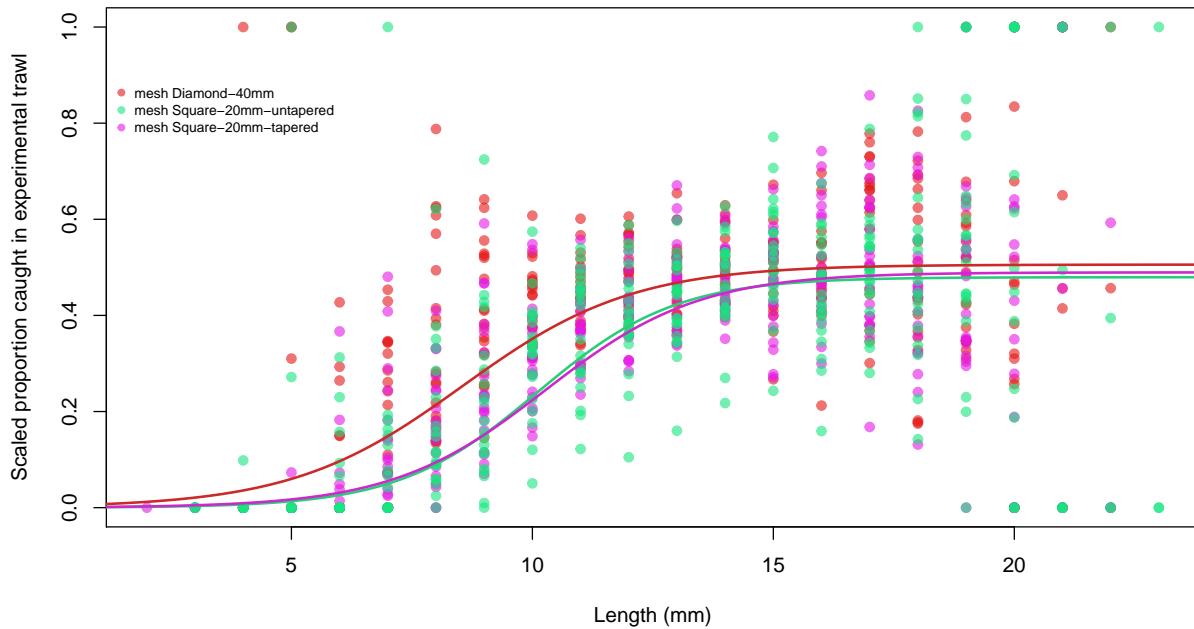


Figure 3.10: The scaled twin-trawl school prawn data with the estimated logistic selection curves fitted by the combined-hauls model. We can see that the selection curve for the diamond mesh is different from those for the square meshes, which are fairly identical.

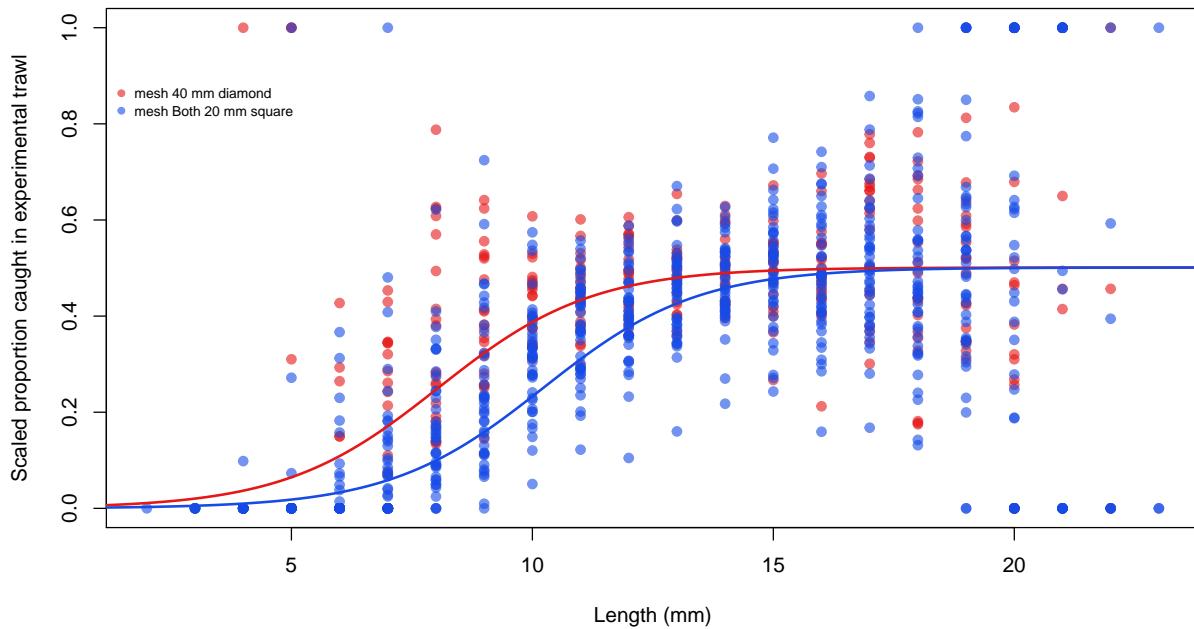


Figure 3.11: The scaled twin-trawl school prawn data with the estimated logistic selection curves fitted by the hierarchical model.

3.5 Integrating the Combined-hauls and Hierarchical Modelling Approaches

In § 2.1.3, we introduced two approaches to modelling data from multiple hauls: combined-hauls and mixed-effects. The models fitted to the covered-codend school prawn data (§ 3.2) and twin-trawl school prawn data (§ 3.4) implement both of these approaches, and in both cases give different results. The reason for this is the fundamental difference in the way each approach treats the individual hauls: the combined-hauls approach marginalises over all hauls, subsequently weighting the marginalised selection curve by the (scaled) number of animals caught in each individual haul, while the hierarchical approach is conditional on the observed hauls, and weights each one equally regardless of the total catch size. The consequence of these differences is that the results have a different interpretation, as discussed in § 2.1.3.

This issue was studied by Wang and Louis (2003), who looked at random effects on the intercept of a logistic GLMM. They compared the shape of the logistic function conditional on the random effects to that obtained by marginalising out these effects, and explain how there is no valid population-averaged interpretation. This is similar to the issue we are faced with: the selection curve estimated using $\mu_{L_{50}}$ and μ_{SR} obtained in the hierarchical model cannot be interpreted in the same way as the curve obtained by the combined-hauls analysis. Wang and Louis (2003) presented a potential tool for making meaningful inferences from the marginal distribution, however this is of no practical use in our situation.

In this section, we explore the possibility of *weighting* the individual L_{50}^h 's estimated from a hierarchical analysis, using the total catch size to obtain an estimate of $\mu_{L_{50}}$ that is comparable to L_{50} from a combined-hauls fit.

Table 3.7: Summary of the posterior distributions of the parameters used in the weighted hierarchical model fitted to the covered-codend school prawn data, including the additional weighted mean estimates of $\mu_{L_{50}}$.

	Mean (SD)	95% Credible Interval	Combined-hauls
$\mu_{L_{50}}$	15.63 (0.34)	14.96 – 16.30	
μ_{SR}	4.21 (0.20)	3.86 – 4.62	5.23 (0.30)
μ_δ	1.98 (0.43)	1.23 – 2.89	3.62 (1.31)
$\sigma_{L_{50}}^2$	2.03 (0.84)	0.96 – 4.16	
$\mu_{L_{50}}^N$	16.15 (0.07)	16.00 – 16.29	16.16 (0.07)
$\mu_{L_{50}}^{wt}$	15.80 (0.08)	15.64 – 15.95	

3.5.1 Weighting Hauls in a Hierarchical Analysis

In a combined-hauls approach, the hauls are weighted by the scaled counts. To mimic this within a hierarchical analysis, we weighted each haul's individual estimate of L_{50}^h by the total number of animals caught in that haul, $N^h = \sum_\ell n_\ell^h$, where n_ℓ^h is the (scaled) count of length ℓ animals in the codend and cover of haul h . The weighted mean for H hauls would therefore be

$$\mu_{L_{50}}^N = \frac{\sum_{h=1}^H L_{50}^h N^h}{\sum_{h=1}^H N^h}. \quad (3.9)$$

Alternatively, we could use catch-weight variables if supplied with the data set themselves. In the covered-codend school prawn data from § 3.2, the variable `WtSch` gives the catch weight, in kilograms, of the codend. We compare these two methods for estimating weighted $\mu_{L_{50}}$'s in the following section.

3.5.2 Fitting a Weighted Hierarchical Model

Using the covered-codend school prawn data as an example, we took the hierarchical model in § 3.2.2 (JAGS Model B.2.1) and implemented the Richards' selection curve to allow us to make sensible comparisons to the combined-hauls model, and implemented nodes for the weighted $\mu_{L_{50}}$'s using the total count, N^h , and catch weight, w^h . The model is shown in JAGS Model B.3.1, and the results are displayed in Table 3.7.

The first thing to note is that the estimate of $\mu_{L_{50}}$ is larger than obtained from the

hierarchical model in § 3.2.2 due to using a Richards' selection curve as opposed to a logistic, which allows us to compare the weighted $\mu_{L_{50}}$'s with the combined fit from § 3.2.1. Secondly, we note that the weighted estimate of $\mu_{L_{50}}$ that uses the total (scaled) counts, N^h , is almost identical to the combined hauls fit, suggesting that, for L_{50} , we can estimate the overall *population* mean from a hierarchical model by weighting the L_{50}^h 's estimated for each haul (at least for this data set).

To extend this weighting technique to SR , it may be possible to use the distributions of L_{25} and L_{75} —which will vary between hauls—and combine these in some way to obtain an estimate of μ_{SR}^N , however our initial attempts at this were unsuccessful. More complicated methods—as well as a significant amount of investigation and simulation—is required if a “weighted” estimate of μ_{SR} is to be sensibly obtained from a hierarchical model.

The use of a weighted hierarchical model could be beneficial in several ways. First, it explicitly models between-haul variability in the selectivity of the gear. Second, we can model the unscaled data and incorporate sampling fractions into the model, removing any need to use post-fit adjustments (such as REP). However, further study is necessary before any such technique could be useful to fisheries researchers.

3.6 Chapter Summary

The main aim of this chapter was to explore several different models used for analysing data from trawl gear selectivity studies. We saw how the Bayesian combined-hauls fit could be implemented to get almost exactly the same results as traditional analyses for both covered-codend and paired-trawl experiments. We also showed that the Bayesian hierarchical model can be used in place of a mixed-effects model, and again give concurring results. Another aim of this chapter was to formulate *flexible* models that can easily be applied to new situations, or implement different retention curves. We see examples of this in Chapters 5 and 6. We also briefly introduced a new approach to modelling these data—the weighted hierarchical model—as a potential method for obtaining the marginal selection curve from the conditional

(hierarchical) model. This chapter did not, however, examine important topics such as model checking for overdispersion and model comparison. We will explore these in some detail in Chapter 4.

Chapter 4

Model Checking, Comparison and Selection

An important aspect of statistical modelling is ensuring that inferences are made from models that adequately fit the observed data. In a Bayesian analysis, posterior predictive checks (PPCs) can be used to assess several aspects of the model’s fit. An important example of this in experimental trawl data is *overdispersion*, which was first described in § 2.1.1. In § 4.1, we investigate how we can use PPCs to assess our models for overdispersion, as well as several ways of modelling it.

Another consideration we must make is model comparison and selection. This provides us with a means of finding the “best predictive model”—usually a balance between the fit to the data (measured by likelihood) and model complexity (number of parameters)—from which we make our inferences. In Chapter 3, we assumed that the models presented by Millar et al. (2004) and Fonseca et al. (2007) were the best models for the data. How Bayesian models can be compared is discussed in § 4.2, and several model comparison exercises taken from the aforementioned papers are used as examples.

4.1 Checking and Correcting for Overdispersion

In Chapter 2, we introduced the idea of overdispersion in GLMs. This occurs when the assumed distribution function $f(y)$ has a smaller variance than observed in the data. In frequentist models, the Pearson's Chi-square statistic can be computed to estimate overdispersion (McCullagh & Nelder, 1989). This statistic is

$$P_{\chi^2} = \sum_{i=1}^n \frac{(y_i - \mathbb{E}[Y_i])^2}{\text{Var}[Y_i]} \sim \chi^2_{n-p}, \quad (4.1)$$

where n is the number of observations, p is the number of parameters, and $\mathbb{E}[Y]$ and $\text{Var}[Y]$ are defined by the error distribution used, such as the binomial ((2.7) and (2.8)). Note the similarity to REP in (2.16), which is in fact the Pearson's Chi-square statistic divided by the degrees of freedom (McCullagh & Nelder, 1989). The Pearson's Chi-square statistic is assumed to be χ^2 distributed with $n - p$ degrees of freedom.

In our Bayesian implementation of this test, rather than assuming P_{χ^2} to have a χ^2 distribution, we can find the true distribution of P_{χ^2} and compare the observed statistic with this. To do so, replicate observations y_i^* are drawn from the posterior distribution and the P_{χ^2} value for these replicated values is computed at each iteration. This produces posterior samples of expected P_{χ^2} statistics, χ^2_{exp} , which can be compared with the single value χ^2_{obs} obtained from the actual data. From this sample, the posterior probability that χ^2_{obs} is greater than χ^2_{exp} is

$$p_{\text{od}} = \mathbb{P}(\chi^2_{\text{obs}} > \chi^2_{\text{exp}}) = \frac{\#\{\chi^2_{\text{obs}} > \chi^2_{\text{exp}}\}}{\#\{\text{samples}\}}, \quad (4.2)$$

which, intuitively, is the posterior probability that the data are overdispersed. Here, we are using *predicted* values to investigate a statistic, so we refer to it as a *posterior predictive check* (PPC) (Gelman et al., 2014).

A large value of p_{od} indicates that the data are likely overdispersed in comparison to the assumed (binomial) distribution. In frequentist GLMs, the usual method for dealing

with overdispersion is quasi-likelihood, in which an estimate of overdispersion is obtained by dividing the observed P_{χ^2} by the degrees of freedom. The standard errors of the parameter estimates are multiplied by this value, subsequently affecting the hypothesis tests and confidence intervals. In § 2.1.3, we discussed REP, which adjusts for both overdispersion and sample size inflation.

The quasi-likelihood method cannot be applied in Bayesian models, however. Instead, the overdispersion is modelled directly. One method is to use a more general distribution function, such as the beta-binomial instead of the binomial (§ 4.1.1); another is to add random effects to the predictor of each observation (§ 4.1.2).

4.1.1 Alternative Distributions

If a distribution function does not adequately model the variance of the data, it is possible to use a more general form of a similar model, usually the consequence of a mixture of two distribution functions. For example, a random variable Y that is binomial distributed with success probability p and number of trials n has an expected variance of $np(1 - p)$. If, however, the variance is greater than this, we can allow extra variation of the parameter p by giving it a beta distribution—this results in the beta-binomial mixture distribution.

The beta-binomial has three parameters: n is the number of trials, while α and β determine the shape of the distribution, and hence the expectation and variance of p . The values α and β can be chosen so that the expected binomial probability coincides with that predicted by the model, $\mathbb{E}[p] = \frac{\alpha}{\alpha+\beta}$. The variance now depends on α and β , therefore providing far more flexibility than the binomial. Using the beta-binomial distribution, the likelihood for our model is

$$y \sim \text{Beta-Binomial}(n, \alpha, \beta), \quad \text{where } \frac{\alpha}{\alpha + \beta} = p \quad (4.3)$$

One issue encountered using (4.3) was that we were unable to explicitly write the model in terms of the link function and linear predictor: the probability, p , depends on the parameters

α and β . While it is possible to parameterise β as a function of the expected probability p and α ,

$$\beta = \frac{\alpha(1-p)}{p}, \quad (4.4)$$

we found convergence of α exceedingly difficult. We tried several different prior distributions on α , but none of them provided an efficient, easily implemented solution, so instead we decided to investigate the random effects method of modelling overdispersion in the data.

4.1.2 Random Effects

Overdispersed data can be modelled by adding random effects to each observation to allow for extra variation (Gelman et al., 2014). That is, the general form of the predictor for observation i can be written as

$$\eta_i = m(x_i) + r_i, \quad r_i \sim N(0, \sigma_{od}^2). \quad (4.5)$$

Here, the parameter σ_{od}^2 describes the amount of overdispersion in the data. This method is convenient as it can easily be added to any function $m(x)$, which coincides with our aim to formulate flexible models.

4.1.3 Covered-codend School Prawn Example

To demonstrate PPCs for overdispersion, and the random effects method for modelling it, we explore the covered-codend school prawn data set from § 3.2. In the (stacked) combined-hauls approach, overdispersion is not present. However, if we model the *unstacked* data—but still scaled to give the same results—we are able to detect overdispersion; the posterior probability that the data are overdispersed, based on the observed and expected χ^2 values and (4.2), is $p_{od} \approx 1$. Therefore, we modified JAGS Model B.1.1 to include random effects on the predictor (using the node `eta`), allowing for this additional variation. This model is displayed in JAGS Model B.3.3, which also shows the implementation of the PPC for overdispersion.

Table 4.1: Comparison of two models with and without overdispersion random effects fitted to the uncombined covered-codend school prawn data.

Model	L_{50}	SR	δ	χ^2_{obs}	χ^2_{exp}	p_{od}	σ_{od}^2
M1	16.16 (0.07)	5.24 (0.31)	3.58 (1.30)	1091.0	213.2	1.00	–
M2	15.72 (0.18)	4.16 (0.28)	2.08 (0.71)	192.9	213.2	0.35	0.95

The results from the models fitted to the scaled (but unstacked) data are shown in Table 4.1. In both of these models, the Richards' selection curve was used. M1 was fitted without random effects for overdispersion, and yields a very high probability (essentially 1) that the data are overdispersed. This is apparent from Figure 4.1, especially for prawns between lengths of 10 and 14 mm.

After adding random effects to each observation to allow for overdispersion (M2 in Table 4.1), p_{od} drops to 0.35. While fitting the model, we found that the algorithm often got trapped sampling unlikely values when using the noninformative lognormal prior on δ , so instead we used $\delta \sim U[0, 50]$ which solved the problem (we have already decided from the previous analysis that $\delta > 1$). This demonstrates that the observation random effects ap-

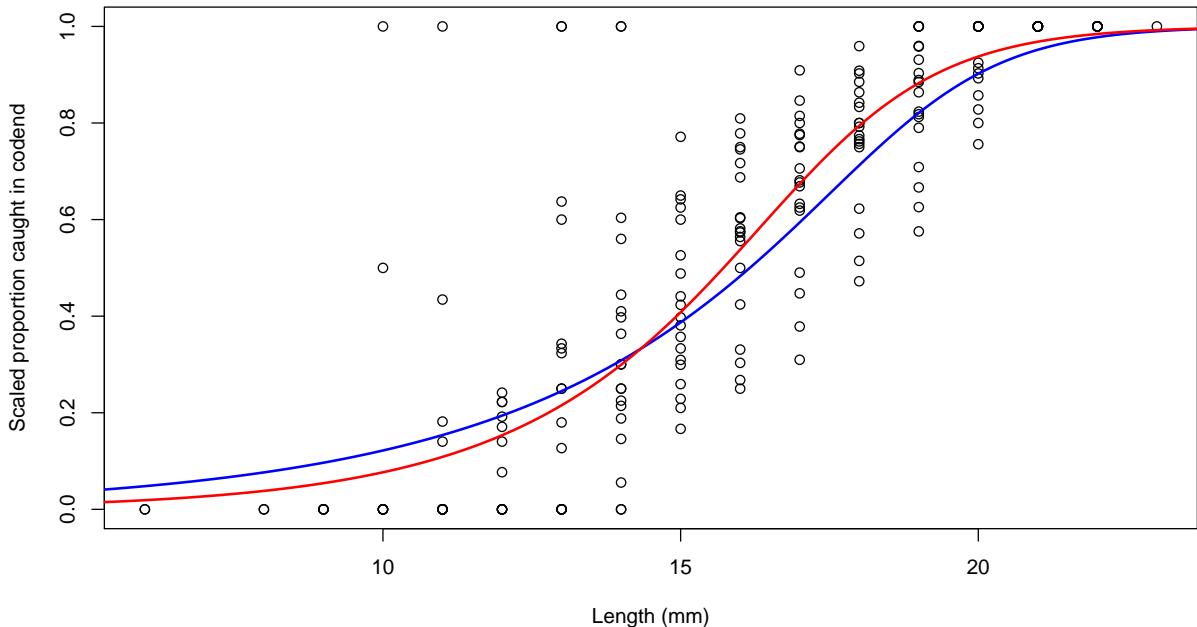


Figure 4.1: The scaled covered-codend school prawn data with the estimated Richards' selection curve fitted using the combined-hauls model, with (red) and without (blue) random effects for overdispersion.

pear to have modelled the extra variation in the data. However, unlike the quasi-likelihood methods used in frequentist models, the parameter estimates are affected. The estimated retention curves from these two models are displayed in Figure 4.1.

From Table 4.1 and Figure 4.1, we see that the model allowing for overdispersion no longer estimates the same contact-selection curve obtained by the REP-adjusted model from Millar et al. (2004). It does, however, allow us to model overdispersion within a Bayesian framework. How useful this type of fit will be to fisheries managers it not clear, and, ultimately, the type of curve fit to the data will depend on the objectives of the particular study. It may be the case that the frequentist model gives a better estimate of the contact-selection curve, while the Bayesian approach may be more attuned to modelling the effects of covariates on the selectivity of a gear, in which case modelling overdispersion in the way presented here could be useful.

4.2 Model Comparison and Selection

When fitting a model to a data set, it is often necessary to compare several candidates and select the best predictive model, which typically is a trade-off between model fit (likelihood) and complexity (number of parameters). There are many model-selection criteria used in frequentist statistics, including Akaike information criterion (AIC) and Bayesian information criterion (BIC) (Seber & Lee, 2003). These are based on the likelihood function evaluated at the MLE, but apply different penalty terms to account for model complexity. In Bayesian models, we can of course compute these values in many cases and use them to assess and compare models; however, we can also use the deviance information criterion (DIC). The DIC is useful for Bayesian problems because it is easily calculated using the posterior samples.

DIC uses the *deviance* of the parameters, $\boldsymbol{\theta}$,

$$D(\boldsymbol{\theta}) = -2 \log f(\mathbf{y} | \boldsymbol{\theta}), \quad (4.6)$$

where $f(\mathbf{y} | \boldsymbol{\theta})$ is the likelihood function of the data \mathbf{y} given the parameters $\boldsymbol{\theta}$ (Spiegelhalter,

Best, Carlin, & van der Linde, 2002). This is computed automatically by JAGS at every iteration, generating a posterior sample of the model deviance.

To compute DIC, the expectation of the deviance, \bar{D} , and the effective number of parameters, p_D , are needed. For \bar{D} , the usual estimate of expectation—the posterior sample mean—is used. Spiegelhalter et al. (2002) proposed p_D as a measure of model complexity to act as a penalty term for computing DIC. The estimate of p_D used in the R2jags package is

$$p_D = \frac{1}{2} \widehat{\text{Var}} [D(\theta)], \quad (4.7)$$

which is discussed by Gelman et al. (2014), and easily computed from the posterior sample of D .

Using the estimates of \bar{D} and p_D obtained from the posterior sample of D , Spiegelhalter et al. (2002) define

$$\text{DIC} = p_D + \bar{D}. \quad (4.8)$$

This allows computation of DIC directly from the posterior samples, and from (4.6) and (4.8), we note that a smaller DIC indicates a better (predictive) model. We used the DIC model comparison criteria to perform the model selection mentioned in Millar et al. (2004) and Fonseca et al. (2007), which is described later, and decide for ourselves on the “best predictive model”.

Another tool used in frequentist model selection is the p-value, often used to decide which variables have no (significant) relationship with the response. We can employ a similar idea in our modelling, debatably referred to as a “Bayesian p-value”. What we will be computing is the *posterior probability that a hypothesis, H_1 , is true, given the data*. In general, for the hypothesis $H_1 : \beta > \beta_0$, we compute

$$p = \mathbb{P}(\beta > \beta_0 | \mathbf{X}) = \frac{\# \{\beta > \beta_0\}}{\# \{\text{samples}\}}. \quad (4.9)$$

If the hypothesis is a two-sided hypothesis, $H_0 : \beta \neq \beta_0$, then the necessary probability is the smaller of $\mathbb{P}(\beta > \beta_0 | \mathbf{X})$ and $\mathbb{P}(\beta < \beta_0 | \mathbf{X})$. This value is simply an intuitive estimate

Table 4.2: Comparison of Richards’ and Logistic selection curves fitted using a combined-hauls model to the covered-codend school prawn data.

Selection curve	L_{50}	SR	δ	p_D	DIC
Logistic	16.04 (0.06)	4.8 (0.2)	—	1.87	90.27
Richards	16.16 (0.06)	5.2 (0.3)	3.6 (1.2)	2.89	79.61

of the posterior probability that the hypothesis is true, given the data we have observed. There are other Bayesian approaches to estimating a “p-value”, such as those discussed by Gelman, Meng, and Stern (1996), but these were not necessary for our purposes.

We now have a method for comparing models, and also a simple test to help decide which—if any—parameters could potentially be dropped from the model. We will now demonstrate the use of these in combination to perform model selection similar to that described by Millar et al. (2004) and Fonseca et al. (2007).

4.2.1 Covered-codend School Prawn Models

As a simple model comparison example, we compare the logistic and Richards’ selection curves fitted to the covered-codend school prawn data from § 3.2. Millar et al. (2004) used the deviance residuals to conclude that the Richards’ selection curve provided a better fit to these data than the logistic.

The two curves were fitted using MCMC, generating posterior samples of the same size, for which all diagnostic tests passed. The only difference in these two models is that the logistic selection curve essentially fixes the value of $\delta = 1$. The results obtained from these two models are displayed in Table 4.2. We can see that the Richards’ model has a smaller DIC value, indicating a better fit to the data. We can also calculate the posterior probability of the hypothesis $H_1 : \delta > 1$ using (4.9). This is $\mathbb{P}(\delta > 1 | X) > 0.999$, which provides very strong evidence in favour of H_1 , further indicating that the asymmetry in the Richards’ selection curve provides a better fit to the data than the symmetric logistic curve.

Table 4.3: Comparison of the models fitted to the covered-codend rose shrimp data. The best model selected by Fonseca et al. (2007) and shown in § 3.3, is Model 5.

Model	Cruise	Mesh	Weight		Depth		DIC
			Mean (SD)	$\mathbb{P}(\nu > 0 X)$	Mean (SD)	$\mathbb{P}(\gamma > 0 X)$	
1	1, 2, 3	Factor	0.046 (0.014)	0.999	0.007 (0.004)	0.929	4085.0
2	1, 2, 3	Factor	0.036 (0.012)	0.999	—	—	4083.7
3	1, 2, 3	Factor	—	—	—	—	4082.7
4	1, 2+3	Factor	—	—	—	—	4081.5
5	1, 2+3	Numeric	—	—	—	—	4085.8

4.2.2 Covered-codend Rose Shrimp Models

The covered-codend rose shrimp experiment recorded several variables, including catch weight of shrimp (kg) and the depth (m) of the haul. In the analysis by Fonseca et al. (2007), two alternate modelling approaches were described. In both of these, w_{ij}^h and d_{ij}^h were the catch weight and depth of haul h in cruise i ($1 = \text{autumn}$, $2 = \text{spring}$, $3 = \text{summer}$), using mesh j ($1 = 55$, $2 = 70$, $3 = 80$). They therefore initially presented the model for the L_{50} of haul h as

$$L_{50ij}^h = \alpha_i + \beta_j + \nu w_{ij}^h + \gamma d_{ij}^h + \epsilon_{ij}^h, \quad (4.10)$$

where $\epsilon_{ij}^h \sim N(0, \sigma_{L_{50}}^2)$, the same as in (3.7). While debated, it is generally accepted that L_{50} decreases as catch weights increases (Millar & Fryer, 1999), therefore it is important to record and include this variable in models to help to reduce the unexplained between-haul variability.

Fonseca et al. (2007) also proposed an alternate, simpler model that assumed a linear effect of mesh size $m_j \in \{55, 70, 80\}$:

$$L_{50ij}^h = \alpha_i + \beta m_j + \nu w_{ij}^h + \gamma d_{ij}^h + \epsilon_{ij}^h. \quad (4.11)$$

Using AIC, they found that (4.11) was the preferred model.

We fitted a selection of models to the rose shrimp data, which are described in Table 4.3, along with the coefficient estimates of the weight and depth covariates and the associated probability their coefficient is greater than zero using (4.9).

The results from the models displayed in Table 4.3 suggest that, contrary to the analysis reported by Fonseca et al. (2007), weight is significantly associated with L_{50} , although interestingly the relationship is positive, which contradicts the “generally accepted” trend noted by Millar and Fryer (1999). Removing depth and weight from the model reduced DIC in both cases. This suggests that, although weight is significantly associated with L_{50} , the effect is accounted for by other variables, and the association is likely due to chance alone. Combining the spring and summer cruises again reduces DIC by a similar amount, as the additional α_3 parameter no longer needs estimating. However, in contrast to the conclusions made by Fonseca et al. (2007), assuming a linear effect of mesh size on L_{50} (Model 5) increases DIC.

The results from the model selection using the rose shrimp data illustrate how model selection can depend on the criteria being used. We have only presented DIC here as an example, but others—such as AIC or BIC—can be used if they are preferred. In this case, the number of parameters for the penalty term would need to be calculated manually. It is possible to implement more complicated PPCs to investigate these models and compare the adequacy of the fits, but these extend outside of the main goals of our work.

4.2.3 Twin-trawl School Prawn Models

For the twin-trawl school prawn data, several different models were investigated by Millar et al. (2004). These included a mixed-effects model fit simultaneously to all of the three gears, each with a unique value of $\mu_{L_{50}}$ and μ_{SR} . Their best model (based on log-likelihood) had separate values of $\mu_{L_{50}}$ for the diamond and square mesh gears, and a common value for the remaining parameters (§ 3.4.2), and corresponds to Model 5 in Table 4.4.

We fitted five models to the twin-trawl school prawn data. For all of them, the tapered and untapered square meshes were modelled as a single gear, as Broadhurst et al. (2004) and Millar et al. (2004) found no difference between the two. The full model has the following

Table 4.4: Comparison of the models fitted to the twin-trawl school prawn data by forward selection.

Model	$\mu_{L_{50}}$	μ_{SR}	μ_ϕ	$\sigma^2_{L_{50}}$	σ^2_ϕ	σ^2_{SR}	β	$\mathbb{P}(\beta > 0 X)$	γ	$\mathbb{P}(\gamma > 0 X)$	DIC
1	8.14 (0.22)	3.62 (0.45)	-0.01 (0.02)	—	—	—	2.34 (0.21)	< 0.001	-0.07 (0.43)	0.541	4705.5
2	8.01 (0.46)	4.01 (0.50)	0.04 (0.02)	2.83 (0.71)	—	—	2.46 (0.53)	< 0.001	-0.19 (0.48)	0.640	4101.5
3	8.04 (0.44)	3.70 (0.50)	0.01 (0.04)	2.38 (0.67)	0.04 (0.01)	—	2.19 (0.51)	< 0.001	-0.20 (0.50)	0.642	4048.3
4	8.06 (0.44)	3.78 (0.48)	0.01 (0.03)	2.45 (0.68)	0.04 (0.01)	0.22 (0.28)	2.18 (0.52)	< 0.001	-0.28 (0.51)	0.693	4066.3
5	8.05 (0.43)	3.51 (0.17)	0.00 (0.03)	2.35 (0.66)	0.04 (0.01)	—	2.18 (0.51)	< 0.001	—	—	4044.9

formula for the L_{50} of haul h :

$$L_{50i}^h = \mu_{L_{50}} + \beta m_i + \epsilon_i^h, \quad \epsilon_i^h \sim N(0, \sigma^2_{L_{50}}), \quad (4.12)$$

and for SR ,

$$SR_i^h = \mu_{SR} + \gamma m_i + \varepsilon_i^h, \quad \varepsilon_i^h \sim N(0, \sigma^2_{SR}), \quad (4.13)$$

where m_i is binary, indicating the use of the square mesh. The relative fishing parameter ϕ is modelled on the logit scale with random haul effects as in § 3.4.2.

The results from the models fitted to the twin-trawl school prawn data—progressively adding more random effects terms—are displayed in Table 4.4. We can immediately see that the DIC reduces dramatically at the inclusion of random effects on L_{50} and ϕ (Models 1, 2 and 3). Adding random effects to SR (Model 4) increased the DIC, which concurs with Millar et al. (2004). The posterior probability of the hypothesis $\gamma > 0$, using (4.9), provides no evidence to suggest a relationship between gear type and SR , so in Model 5 SR is modelled with a common effect for all gears, resulting in the smallest DIC of the models we fit, and thus the best predictive model (Table 4.4), which concurs with the findings of Millar et al. (2004).

In the next two chapters, we will be exploring a range of extensions to the models seen so far. When applicable, we will be using the model checking (PPC) and comparison (DIC) techniques discussed in this chapter to decide between models.

Chapter 5

Exploring Extensions of Bayesian Selectivity Models

In Chapter 3, we focused on implementing models from frequentist analyses using a Bayesian approach, and investigated several methods for assessing and comparing Bayesian models in Chapter 4. The next two chapters explore the flexibility of these models, with special focus on generalising the predictor and link functions. In this chapter, we explore how the Poisson approximation can be applied to experimental trawl data, and explore several cases where the logistic selection curve can be substituted by the Richards' curve, with special attention to the hierarchical models in which the Richards' curve was not applied. The capacity with which the Bayesian hierarchical model can cope with multiple random effects (i.e., L_{50} and SR simultaneously) is explored in § 5.3, in which we discover that models deemed “too complex” for SAS can be implemented easily.

5.1 The Poisson Distribution

So far, the binomial distribution was used to describe the error structure of the data, due to the intuitive definition of the necessary probabilities (§ 2.1.2). However, there may be situations where another distribution might be better suited. For example, in Chapter 4 we saw that a beta-binomial could be used instead of a binomial to allow for extra variability,

although we found this approach too complicated to fit into a general model. Feller (1968) discusses how the Poisson can be used as an approximation of a binomial random variable. We will describe how this is done, and briefly discuss the potential benefits of taking such an approach.

Millar (1994) noted that the rate of length ℓ animals entering a trawl can be assumed to have a Poisson distribution with rate λ_ℓ , and that each animal escapes *independently* from the codend with probability $1 - r(\ell)$. Therefore, we expect $(1 - r(\ell)) \lambda_\ell$ length ℓ animals to escape the codend, leaving $r(\ell) \lambda_\ell$ caught in the codend (Feller, 1968; Millar, 1994). Additionally, if subsampling occurs, then only q_1 and q_2 of the animals in the codend and cover are counted, respectively. This leads to the Poisson approximation of the binomial, which gives rise to the following likelihood for the data (Millar, 1994):

$$\begin{cases} y_{\text{cod},\ell} \sim \text{Pois}(q_1 r(\ell) \lambda_\ell) \\ y_{\text{cov},\ell} \sim \text{Pois}(q_2 (1 - r(\ell)) \lambda_\ell) . \end{cases} \quad (5.1)$$

The most important difference here is that we no longer condition on n_ℓ ; instead, we treat the total number of animals as an unknown parameter (which is true unless no subsampling occurs).

The Poisson distribution removes the need for incorporating the sampling fractions into the probability formula as seen in § 2.1.2. We do, however, need to estimate $\boldsymbol{\lambda}$, which introduces one additional parameter for each observation. We used noninformative, independent, lognormal prior distributions for each λ_ℓ^h , although as mentioned by Millar and Fryer (1999), it is possible to give a distribution (such as a normal or gamma) to the length classes, which can significantly reduce the number of model parameters. For example, instead of N (number of length classes) parameters for haul h , we can reduce this to a mean, θ^h , variance, $(\sigma_N^2)^h$, and scaling parameter, Λ^h , if we assume a normal distribution of length classes:

$$\lambda_\ell^h = \Lambda^h e^{-\frac{(l-\theta^h)^2}{2(\sigma_N^h)^2}}, \quad (5.2)$$

in which case Λ^h is the *overall abundance* for haul h . It may also be reasonable to assume the same or similar length class distributions for all of the hauls, in which case the appropriate modifications could be made to (5.2). While this is simple to implement in a Bayesian framework, in the following examples we use the independent λ_ℓ^h 's model as this gives results that are comparable to those obtained using the binomial—imposing additional restrictions will inevitably lead to different results, and we lacked the necessary information to make such assumptions.

In the case of a paired trawl, we need to adjust the Poisson rates shown in (5.1) to include ϕ . Here, the number of animals expected to be caught in the experimental codend, given that they entered the trawl, is Poisson distributed with rate $q_1 r(\ell) \phi \lambda_\ell$; likewise, the count for the control codend will have rate $q_2 (1 - r(\ell)) \lambda_\ell$. Other than these differences to the likelihood (in both covered-codend and paired experiments), the rest of the model is exactly the same as defined for the binomial model.

5.1.1 Examples of the Poisson Model

To demonstrate the Poisson approach to modelling trawl data, we implemented the models from Chapter 3 with the single alteration of the likelihood to (5.1). Again, for comparability we have used independent priors on the λ_ℓ^h parameters. Implementing distributions on these made slight changes to the results, and further investigation (by simulation) is needed to discover whether or not these changes are useful to the analysis.

Covered-codend School Prawn Data

The model shown in JAGS Model B.3.3 on page 124 is mostly the same as JAGS Model B.1.1, which was used to model the combined-hauls covered-codend data in § 3.2.1. In this model, the nodes `theta1` and `theta2` are the Poisson rates defined in (5.1). The `log` function was used to prevent numerical underflow to 0 (in the case where p is close to 0 or 1), which leads to several additional lines in the BUGS code.

The results from fitting JAGS Model B.3.3 are displayed in Table 5.1. The parameter

Table 5.1: Summary of the posterior distributions of the selectivity parameters of the combined-hauls Poisson model fitted to the covered-codend school prawn data.

Parameter	Mean (SD)	95% Credible Interval	Binomial Model
L_{50}	16.16 (0.07)	16.03 – 16.29	16.16 (0.07)
SR	5.23 (0.30)	4.70 – 5.87	5.23 (0.30)
δ	3.61 (1.30)	1.79 – 6.71	3.62 (1.31)

estimates are almost identical to the results from the binomial model. However, this model provides direct estimates of the total number of school prawns in each length class. Figure 5.1 shows the length class distribution of the school prawns caught in the trawl. Millar and Fryer (1999) discuss potential uses of the length class distribution, which is often obtained through gill net studies (examples of these include Spangler and J. (1992), Tátrai, Specziár, György, and Bíró (2008) and Karakulak and Erk (2008)). If available, length class distributions estimated by these types of studies may be used to provide prior information about λ .

A hierarchical Poisson model was also fitted using the same methods discussed previously, the results of which are displayed in Table 5.2. Again, the two models—binomial and Poisson—give the same results.

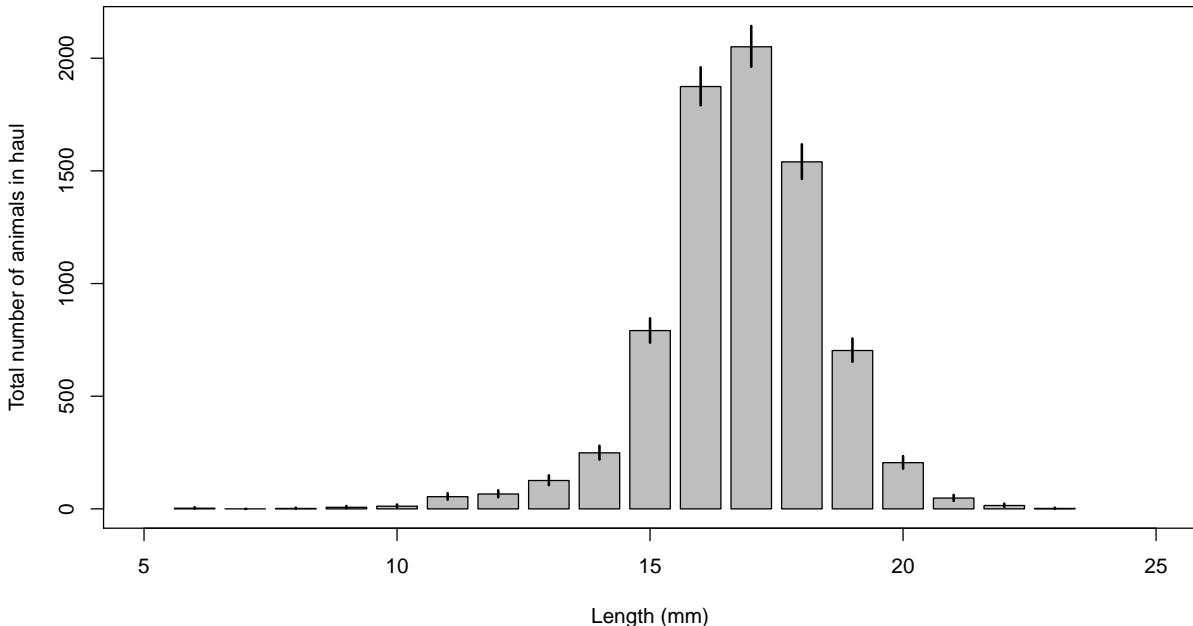


Figure 5.1: The posterior distribution of length class counts (with 95% credible intervals) for the covered codend school prawn data, estimated by the Poisson model.

Table 5.2: Summary of the posterior distributions of the selectivity parameters of the hierarchical Poisson model fitted to the covered-codend school prawn data.

Parameter	Mean (SD)	95% Credible Interval	Binomial Model
$\mu_{L_{50}}$	15.50 (0.34)	14.85 – 16.15	15.51 (0.35)
μ_{SR}	4.05 (0.16)	3.76 – 4.38	4.04 (0.16)
$\sigma_{L_{50}}^2$	2.13 (0.83)	1.01 – 4.10	2.14 (0.90)

Table 5.3: Summary of the posterior distributions for the parameters of the hierarchical Poisson model fitted to the covered-coded rose shrimp data.

Parameter	Mean (SD)	95% Credible Interval	Binomial Model
μ_{SR}	11.38 (0.31)	10.79 – 12.00	11.37 (0.32)
β	0.41 (0.01)	0.39 – 0.43	0.41 (0.01)
$\alpha_{2\&3}$	-3.06 (0.89)	-4.90 – -1.37	-3.05 (0.88)
$\sigma_{L_{50}}^2$	12.92 (2.46)	8.82 – 18.46	12.77 (2.44)

When fitting these models, it is important to note that DIC cannot be used to compare the binomial and Poisson models, as they each use different data: the binomial is conditional on n , supplied as data, while the Poisson treats n as random. Therefore, no single statistic can be used to compare these models. We could have used PPCs to compare the fit of each method, and make a decision based on those as to which model fits the data best; however, the decision would be best made based on the availability of length class information, and whether or not it positively affects the results, an area that yet needs further research.

Covered-codend Rose Shrimp Data

Next, we implemented the rose shrimp model from § 3.3 using the Poisson distribution. The posterior summary is displayed in Table 5.3, and we can see that again the results are almost identical to those obtained from the binomial model. In the rose shrimp study, the cruises were in fact from different seasons. Therefore, it may be of interest to researchers to estimate the length distribution of animals in each season for comparison, in which case the Poisson approach would be useful.

Table 5.4: Summary of the posterior distributions of the selectivity parameters of the combined-hauls Poisson model fitted to each of the three gears in the twin-trawl school prawn data.

Gear	Parameter	Mean (SD)	95% Credible Interval	Binomial Model
40 mm diamond	L_{50}	8.55 (0.09)	8.38 – 8.73	8.55 (0.09)
	SR	3.89 (0.21)	3.51 – 4.33	3.88 (0.21)
20 mm tapered square	L_{50}	10.11 (0.06)	9.98 – 10.23	10.10 (0.07)
	SR	3.17 (0.08)	3.01 – 3.33	3.16 (0.08)
20 mm uptapered square	L_{50}	10.27 (0.08)	10.11 – 10.44	10.27 (0.08)
	SR	3.49 (0.12)	3.26 – 3.73	3.49 (0.12)

Table 5.5: Summary of the posterior distributions of the selectivity parameters of the hierarchical Poisson model fitted to the twin-trawl school prawn data.

Gear	Parameter	Mean (SD)	95% Credible Interval	Binomial Model
40 mm diamond	$\mu_{L_{50}}$	8.06 (0.43)	7.19 – 8.87	8.05 (0.43)
Both 20 mm square	$\mu_{L_{50}}$	10.23 (0.28)	9.68 – 10.79	10.23 (0.28)
All gears	μ_{SR}	3.51 (0.17)	3.19 – 3.85	3.51 (0.17)
	μ_ϕ	0.0052 (0.0346)	-0.0624 – 0.0730	0.0044 (0.0356)
	$\sigma_{L_{50}}^2$	2.35 (0.66)	1.37 – 3.90	2.36 (0.66)
	σ_ϕ^2	0.04 (0.01)	0.02 – 0.07	0.04 (0.01)

Twin-trawl School Prawn Data

Finally, we apply the Poisson model to the twin-trawl school prawn data from § 3.4, for both the combined-hauls and hierarchical approaches. The results of the combined-hauls model are displayed in Table 5.4, and for the hierarchical model in Table 5.5. In both cases, the results match those obtained by the corresponding binomial model from § 3.4.

5.1.2 Weighting the Hierarchical Poisson Model

In § 3.5, we saw how the total counts, N^+ , could be used to obtain a weighted estimate of $\mu_{L_{50}}$. This, however, used N obtained from scaling the counts. Instead, we could use $\Lambda^h = \sum_\ell \lambda_\ell^h$, the sum of the estimated total number of animals caught in haul h , allowing variability to be incorporated into the total count. Doing so gave an estimate of $\mu_{L_{50}}^\Lambda$ almost identical to $\mu_{L_{50}}^N$ obtained in § 3.5.2. In some cases, such as those with small sampling fractions, there may be a difference in the estimates, especially in the standard deviation

Table 5.6: Comparison of the Richards' and logistic selection curves fitted using a hierarchical model to the covered-codend school prawn data.

	L_{50}	SR	δ	$\sigma_{L_{50}}^2$	p_D	DIC
Logistic	15.5 (0.3)	4.0 (0.2)	–	2.1 (0.9)	19.3	659.2
Richards	15.6 (0.3)	4.2 (0.2)	2.0 (0.4)	2.1 (0.8)	20.3	652.9

and credible intervals, though we do not investigate this any further.

5.2 Selection Curves

Another component of a GLM is the link function, $g(\eta)$. In Chapter 3, we implemented either a logistic or Richards' selection curve as the inverse link function for the data, although only the logistic selection curve was used in the hierarchical models. In this section, we explore the use of the Richards' curve in hierarchical models. This also opens up the potential for additional inverse link functions to be used, for example further generalisations of the logistic discussed by Richards (1959). These could be implemented using a β_0 and β_1 parameterisation to estimate the selection curve, and, if closed-form expressions for L_{50} and SR can be found, then the curve could likely be parameterised in terms of these. However, we did not investigate any other parametric selection curves in our research; instead, several semiparametric alternatives were explored, which are discussed in Chapter 6.

5.2.1 Covered-codend School Prawn

In Chapter 3, the Richards' curve was fitted to the stacked covered-codend school prawn data, while a logistic curve was used in the hierarchical model. In Chapter 4, we compared the combined-hauls models using the logistic and Richards' selection curves, and saw that the Richards' indeed gives a better fit to these data. Here, we compare the logistic and Richards' selection curves in the hierarchical model. In the Richards' model, a common δ was fitted to all hauls. The results from the two models are compared in Table 5.6.

We see that the model implementing the Richards' selection curve has a smaller value of

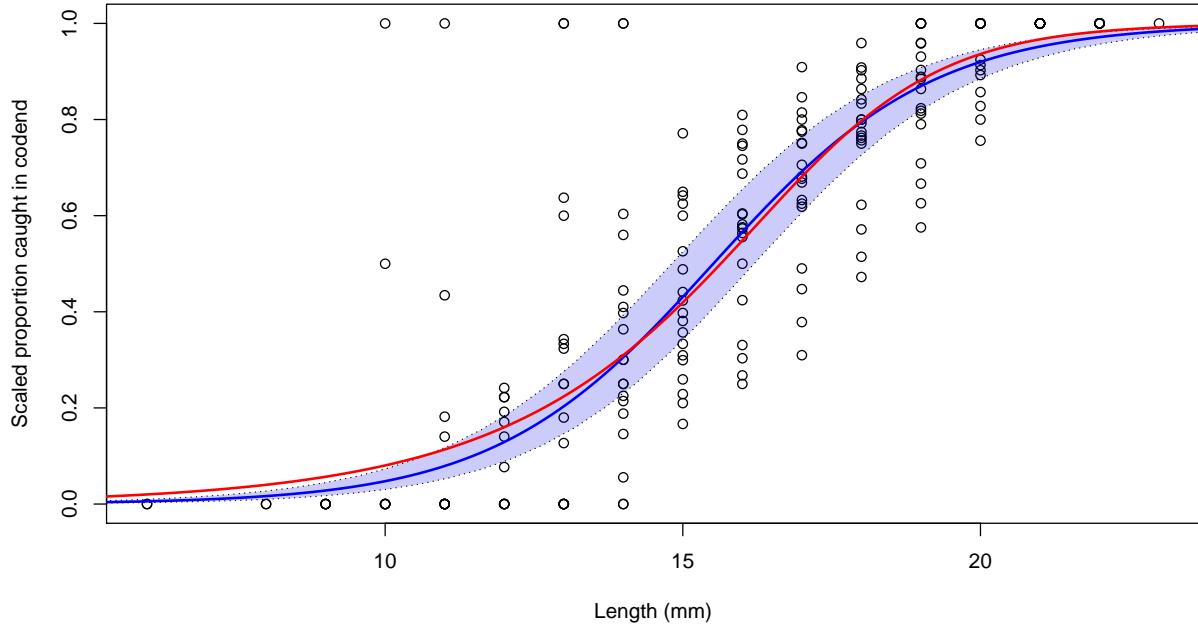


Figure 5.2: The scaled covered-codend school prawn data, with the estimated logistic and Richards' selection curves fitted using Bayesian hierarchical models. The red curve is the logistic selection curve, and the blue curve is the logistic selection curve, along with the 95% credible interval for the logistic model.

DIC, indicating a better predictive fit. The posterior probability of the hypothesis $H_1 : \delta > 1$ was also performed using (4.9), and gave a value of 0.998, which strongly suggests the additional parameter δ is necessary. This, in combination with the smaller DIC, suggests that the Richards' selection curve provides a better fit than the logistic. This differs from the conclusions made by Millar et al. (2004), who found that the logistic selection curve provided an adequate fit to the individual hauls; therefore, they used this simple curve in their mixed-effects model.

Figure 5.2 displays the estimated logistic and Richards' selection curves. We see that they are reasonably similar for “average-sized” animals (14–19mm), however the Richards' curve allows for a higher retention probability of small and large animals, although the effect is most notable in animals under about 13 mm in length. The estimated Richards' curve lies outside of the 95% credible region for the logistic curve for prawns less than 11 mm in length. It is important to note, however, that the hierarchical model is not intended to

Table 5.7: Comparison of the Richards' and logistic selection curves fitted using a hierarchical model to the covered-codend rose shrimp data.

Parameter	Cruise	Mesh size (mm)	Logistic			Richards	
			Mean (SD)	95% CI	Mean (SD)	95% CI	
μ_{SR}	combined		11.38 (0.31)	10.79 – 12.00	11.34 (0.35)	10.70 – 12.06	
δ			-	-	0.92 (0.20)	0.55 – 1.34	
β			0.41 (0.01)	0.39 – 0.43	0.41 (0.01)	0.38 – 0.43	
$\alpha_{2&3}$			-3.06 (0.89)	-4.90 – -1.37	-3.07 (0.91)	-4.88 – -1.36	
$\sigma_{L_{50}}^2$			12.92 (2.46)	8.82 – 18.46	13.14 (2.63)	8.75 – 18.92	
$\mu_{L_{50}}$	Autumn	55	22.38 (0.60)	21.20 – 23.57	22.34 (0.61)	21.17 – 23.54	
		70	28.48 (0.76)	26.99 – 30.00	28.43 (0.78)	26.94 – 29.97	
		80	32.55 (0.87)	30.84 – 34.28	32.50 (0.89)	30.79 – 34.25	
	Spring/summer	55	19.32 (0.53)	18.26 – 20.33	19.27 (0.56)	18.16 – 20.32	
		70	25.42 (0.52)	24.43 – 26.40	25.36 (0.55)	24.27 – 26.44	
		80	29.49 (0.53)	28.43 – 30.49	29.42 (0.56)	28.28 – 30.55	
DIC (pD)				10844 (pD = 1183)		10917 (pD = 1256)	

give us an estimate of the “mean” selection curve; instead, it is used primarily for modelling between-haul variability and incorporating this into hypothesis testing. In this case, the estimate of $\sigma_{L_{50}}^2$ is about the same in the two models.

5.2.2 Covered-codend Rose shrimp

In their analysis of the covered-codend rose shrimp data, Fonseca et al. (2007) make no mention of the Richards' selection curve. Table 5.7 displays results from fitting a hierarchical model using a Richards' selection curve, which are compared to the results from the logistic model fitted in § 3.3. In this case, the Richards' model has a larger DIC value, indicating that the logistic selection curve is adequate for these data. The posterior probability of the hypothesis $H_1 : \delta > 1$ was 0.33, which further indicates that the logistic curve (i.e., $\delta = 1$) is sufficient.

5.2.3 Twin-trawl School Prawn

Finally, we fitted a hierarchical Richards' selection curve to the twin-trawl school prawn data (§ 3.4). While fitting this model, we found that convergence became difficult due to correlations between ϕ and δ , $\text{Cor}(\mu_\phi, \mu_\delta) = -0.41$. However, after setting initial values in the proximity of the values obtained from the logistic model, we were able to attain

Table 5.8: Comparison of Richards' and logistic selection curves fitted using a hierarchical model to the twin-trawl school prawn data.

	$\mu_{L_{50}}$	μ_{SR}	δ	$\sigma_{L_{50}}^2$	p_D	DIC
Logistic	8.0 (0.4)	3.5 (0.2)	–	2.4 (0.7)	132.1	4045.0
Richards	8.1 (0.4)	3.6 (0.3)	1.0 (0.32)	2.3 (0.6)	132.0	4046.2

convergence of the MCMC chains (without altering the prior distributions).

From the results in Table 5.8, we see that the posterior parameter estimates are almost exactly the same for the two models, indicating that the logistic selection curve is sufficient. The posterior probability of $H_1 : \delta > 1$ is 0.41, and the DIC value for the logistic model agrees that it is adequate for these data.

In this section, we showed the simplicity with which the Richards' selection curve can be used to model experimental trawl data. We were able to show that—according to our chosen model selection criteria—the Richards' selection curve gives a better fit than the logistic in the hierarchical model when fit to the covered-codend school prawn data, suggesting that the hierarchical model provides us with more power to fit this more complex model than a frequentist GLMM, especially in combination with the non-linear parameterisation using L_{50} and SR directly.

5.3 Additional Random Effects

So far, all of the hierarchical models we have presented only implemented hierarchical priors (i.e., random effects) on L_{50} , with the one exception of a model fitted in § 4.2.3 during model selection (see Table 4.4). It is conceivable that SR also varies by haul in the other models. Millar et al. (2004) and Fonseca et al. (2007) each originally proposed models with random-effects on SR , but they either found it unnecessary, or encountered initialisation or convergence issues in SAS. We attempted these same covered-codend models to investigate the performance of our hierarchical Bayesian model.

Table 5.9: Comparison of the hierarchical models fitted to the covered-codend school prawn data, with (M2) and without (M1) a hierarchical prior distribution on SR .

	$\mu_{L_{50}}$	μ_{SR}	$\sigma_{L_{50}}^2$	σ_{SR}^2	p_D	DIC
M1	15.5 (0.3)	4.0 (0.2)	2.1 (0.9)	—	19.3	659.2
M2	15.5 (0.3)	4.1 (0.2)	2.1 (0.9)	0.2 (0.2)	27.3	664.1

5.3.1 Covered-codend School Prawn Data

From the model formulated in § 3.2.2, we added a hierarchical prior to the SR parameter:

$$SR^h \sim N(\mu_{SR}, \sigma_{SR}^2), \quad (5.3)$$

where σ_{SR}^2 has the same uniform prior as given to $\sigma_{L_{50}}^2$, (3.6). We experienced no convergence issues, and the diagnostic tests passed on the 3 MCMC chains obtained from 100,000 iterations comprising a burn in of 50,000 and thinning interval of 10.

Comparison of the results for the models is displayed in Table 5.9, and we see that the mean estimates of the parameters are the same. The value of σ_{SR}^2 in (M2) is very small, indicating that any between-haul variation in SR is minimal. Allowing SR to vary between hauls results in a slight increase of the DIC, indicating that M1—the model with a fixed SR —is preferred, which concurs with the conclusions made by Millar et al. (2004).

5.3.2 Covered-codend Rose Shrimp Data

In their paper, Fonseca et al. (2007) proposed a model with random effects on SR ; however, they found the algorithm—PROC NLMIXED from SAS—was frequently unable to model both SR and L_{50} as random. So, they concluded that this model was too complicated for the given data. To investigate this, we modified the model from § 3.3 to include a hierarchical prior on SR as in (5.3). The model was fitted using 3 chains, each with 50,000 iterations comprising a burn in period of 25,000 and thinning interval of 10. The generated samples passed the convergence and stationary diagnostic tests.

Table 5.10 displays the results of the models fitted to the covered-codend school prawn

Table 5.10: Comparison of the hierarchical models fitted to the covered-codend rose shrimp data, with (M2) and without (M1) a hierarchical prior distribution on SR .

	μ_{SR}	$\alpha_{2\&3}$	β	$\sigma_{L_{50}}^2$	σ_{SR}^2	p_D	DIC
M1	11.37 (0.32)	-3.05 (0.88)	0.41 (0.01)	12.77 (2.44)	—	76.5	4084.5
M2	11.84 (0.50)	-2.28 (0.79)	0.40 (0.01)	9.28 (2.08)	8.75 (2.82)	146.7	4071.7

data. Immediately, we see that the estimate of σ_{SR}^2 in M2 is significantly larger than 0, and is in fact almost as large as $\sigma_{L_{50}}^2$, which is smaller in M2 than it is in M1. This suggests a significant amount of variation between the SR of each haul. Comparison of the DICs confirms this: despite the added complexity of the model with random effects on SR —it has approximately twice as many parameters according to p_D —it has a smaller DIC, indicating a better fit to the data.

In this case, we were able to use a Bayesian hierarchical model to fit a hierarchical distribution to both L_{50} and SR , without succumbing to convergence issues as were experienced using SAS (although it is untested as to whether later versions of SAS perform any better). Having discovered that SR varies significantly between hauls, it would be advisable to carry out model selection, similar to that from § 4.2.2, to see if any new conclusions can be drawn about the best model for these data—for example, is there a relationship between SR and mesh size? However, this was left out of our study as finding the best model was not the focus of our work.

Chapter 6

Semiparametric Selection Curves

So far only parametric selection curves have been used, which includes the logistic curve—a symmetric function—and the Richards’ curve—an asymmetric function—which are based off of a “linear” function on the link scale (linear in shape, not parameterisation). We saw in several cases that the extra flexibility of the Richards’ curve was preferred over the logistic selection curve. However, the shape of these curves is still constrained by three parameters. In this chapter, we investigate *semiparametric* functions that reduce the assumptions about the shape of $r(\ell)$ down to three: the function must only produce values between 0 and 1; the function should be continuous over all (positive) lengths; and the function should be isotonic (non-decreasing). Similar types of models have been fitted before, such as in dose-response analyses (Bornkamp & Ickstadt, 2009), and a thorough review of semiparametric regression, including its implementation using Bayesian methodology, is given by Ruppert, Wand, and Carroll (2009).

Neelon and Dunson (2004) used a Bayesian approach to fit isotonic functions to several data sets, comparing their results to frequentist generalised additive models, and showed that it was possible to obtain a (95%) credible region around the mean estimate of the function. While this credible region is useful, there was no proposed method for quantifying the curve using parameters such as L_{50} or SR . We therefore investigated several techniques for fitting semiparametric curves such as those used by Neelon and Dunson (2004), with the additional

goal of estimating L_{50} and SR from the fitted curves (§ 6.4).

To ensure the curves we fit are selection curves, for choice of inverse link function we used both the logistic function as we have used previously, (2.4), and, as implemented by Ruppert, Wand, and Carroll (2003) and Neelon and Dunson (2004), the standard normal CDF, $\Phi(\eta)$. Both of these give the necessary function property $\mathcal{F} : \mathbb{R} \rightarrow (0, 1)$. The next assumption is that the curve be continuous; however, the semiparametric functions we will be using are continuous by definition, so we need not worry about this any further. Finally, the curve must be isotonic, such that

$$f(x) \leq f(x+1). \quad (6.1)$$

The means of ensuring this varies depending on the type of semiparametric curve being used.

We will be discussing three competing semiparametric functions in this chapter: piecewise-linear functions, as used by Neelon and Dunson (2004); shape-constrained regression splines, specifically using a truncated multivariate normal distribution to enforce an isotonic curve (Shively, Walker, & Damien, 2011; Hazelton & Turlach, 2011), and Basis splines (or B-splines), (De Boor, 1978). While exploring these functions, we used the stacked covered codend school prawn data from § 3.2.1. In § 6.5, a hierarchical approach to semiparametric modelling is investigated.

6.1 Piecewise Linear Model

The work by Neelon and Dunson (2004) involved fitting piecewise linear isotonic regression curves to medical drug-exposure data using Bayesian models. In their paper, they investigated a Gibbs sampling algorithm to estimate the model parameters, which consist of K knot locations $\kappa = (\kappa_0, \dots, \kappa_K)$, and define β_j as the slope within the interval $(\kappa_{j-1}, \kappa_j]$. Their model is defined as

$$\eta(x_i) = \alpha + \sum_{j=1}^K w_j(x_i) \beta_j, \quad (6.2)$$

where α is the intercept term, and $w_j(x_i)$ is defined as

$$w_j(x_i) = \begin{cases} \min(x_i, \kappa_j) - \kappa_{j-1}, & x_i \geq \kappa_{j-1}, \\ 0, & \text{otherwise,} \end{cases} \quad (6.3)$$

which ensures continuity of the function at the knot points.

The piecewise linear spline is ensured to be isotonic by constraining each $\beta_j \geq 0$, but this makes inference difficult as the null hypothesis of no association (i.e., $H_0 : \beta_j = 0$) is on the boundary of the parameter space (Neelon & Dunson, 2004; Schipper, Taylor, & Lin, 2007). An algorithm implementing a mixture model that gives positive prior probability to flat-regions (i.e., $\beta_j = 0$) was proposed by Neelon and Dunson (2004), however this complexity was avoided in our models as we had no particular interest in detecting these flat regions, and we could assume that the selection curve would be strictly increasing, $f(x) < f(x+1)$.

We implemented (6.2) using the constraint of $\beta_j \geq 0$, for all $j = 1, \dots, K$, to ensure an isotonic fit; the model for this is displayed in JAGS Model B.4.1. Each β_j was given an independent truncated noninformative normal prior. Because of the discrete-nature of the data, we needed to use less than N (the number of unique length classes) knots; if we used $K = N$, we would be estimating a single probability for each length class. To compute (6.3), we used the `inprod` function as recommended by Crainiceanu, Ruppert, and Wand (2005). We implemented the logit and probit (inverse standard normal CDF, Φ^{-1}) link functions, however the probit model typically had more severe convergence issues, so the logit was preferred in this case. Using DIC, we found that $K = 4$ knots was preferred for the data ($DIC = 76.6$).

While attempting to fit these models to the stacked covered-codend school prawn data, we encountered several problems. The more prominent was non-convergence of the MCMC chains for the intercept and first components of β , as can be seen by the values of PSRF in Table 6.1, and in the variation of the sample estimates shown in Figure 6.1. This is likely due to the values in that region being zero-counts, which correspond to a log-odds

Table 6.1: Summary of the posterior distributions of the parameters of the piecewise linear model fitted to the stacked covered-codend school prawn data. Increasing chain length did not improve convergence.

Parameter	Mean (SD)	95% Credible Interval	PSRF
α	-10.6 (3.6)	-15.3 – -4.5	7.70
β_1	1.7 (0.7)	0.5 – 2.6	7.56
β_2	0.4 (0.1)	0.2 – 0.5	1.00
β_3	0.5 (0.0)	0.4 – 0.5	1.00
β_4	801.1 (621.6)	31.3 – 2308.6	1.00

of $-\infty$. Running the MCMC sampler did not make convergence any easier, and became computationally inefficient. We also tried various other prior distributions on the parameters, including a uniform prior on α , and using different numbers of knot points, these had no effect on convergence. Additionally, the piecewise-linear function lacks smoothness, which may not be appealing to fisheries managers. For these reasons, we shifted our attention to splines, which allowed us to fit smooth functions to the data (i.e., continuous first derivative) without experiencing these same convergence issues.

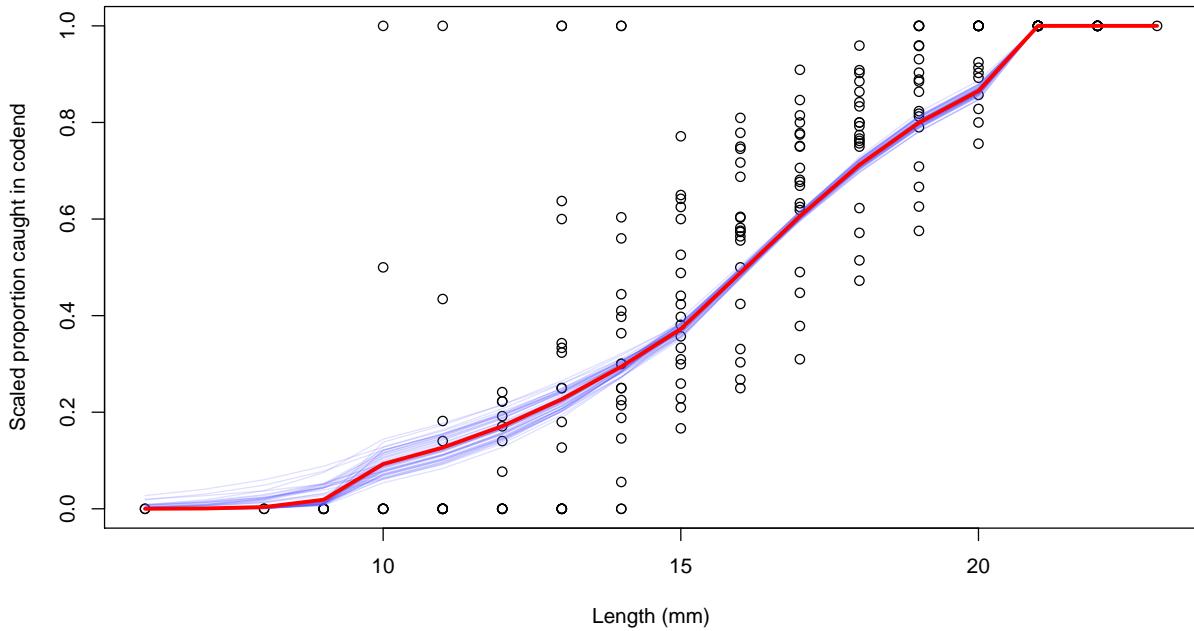


Figure 6.1: The scaled covered-codend school prawn data, with the estimated piecewise linear selection curve. Individual sample estimates are in blue, while the red curve was estimated using the posterior means of the parameters. This model used $K = 4$ evenly spaced knot points.

6.2 Shape-constrained Regression Splines

The main issue we face in fitting a semiparametric selection curve is that it must be isotonic. In a paper by Hazelton and Turlach (2011), they proposed a method based on that used by Shively et al. (2011) for fitting a shape-constrained semiparametric regression curve by using a truncated multivariate normal distribution to impose the constraints. The theoretical details of their paper are quite complex, as they base their model around penalized regression splines (Eilers & Marx, 1996). That is, they fitted the model $y_i = m(x_i) + \epsilon_i$, for $i = 1, 2, \dots, n$, where

$$m(x) = s(x | \boldsymbol{\beta}, \mathbf{u}) = \beta_0 + \beta_1 x + \dots + \beta_p x^p + \sum_{j=1}^K u_j (x - \kappa_j)_+^q, \quad (6.4)$$

q is the order of the spline ($q = 1$ implies piecewise linear), and K is the number of knots, $\kappa_0 < \kappa_1 < \dots < \kappa_K < \kappa_{K+1}$. Note that $\kappa_0 \leq \min(x)$ and $\kappa_{K+1} \geq \max(x)$ are *not* used as knots, but Hazelton and Turlach (2011) define them in this way to simplify the mathematics. The subscript $+$ denotes

$$(x)_+ = \begin{cases} x, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (6.5)$$

Hazelton and Turlach (2011) proceed to show that (6.4) can be solved as a linear mixed model; however, we were mainly interested in how they implemented their constraints. They defined the parameter vector $\boldsymbol{\xi} = (\boldsymbol{\beta}^T, \mathbf{u}^T)^T$ to contain all of the parameters from (6.4). To enforce the spline function to be isotonic, the first derivative of (6.4) with respect to x must always be non-negative; that is, in the case of a 2nd degree polynomial spline ($q = 2$) (Hazelton & Turlach, 2011):

$$\frac{ds(x | \boldsymbol{\beta}, \mathbf{u})}{dx} \Big|_{x=\kappa_j} = \beta_1 + 2\beta_2 \kappa_j + 2 \sum_{k=1}^j u_k (\kappa_j - \kappa_k)_+ \geq 0, \quad j = 0, \dots, K + 1, \quad (6.6)$$

which implies monotonicity over the interval $[\kappa_0, \kappa_{K+1}]$, as required.

From (6.6), $K + 2$ inequalities are obtained, which can be expressed in matrix notation

by placing the coefficients from these inequalities into a $K + 2$ by $K + 3$ matrix, \mathbf{A} . Again, for a 2nd order polynomial,

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 2\kappa_0 & 0 & \cdots & 0 \\ 0 & 1 & 2\kappa_1 & 0 & \cdots & 0 \\ 0 & 1 & 2\kappa_2 & 2(\kappa_2 - \kappa_1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 2\kappa_{K+1} & 2(\kappa_{K+1} - \kappa_1) & \cdots & 2(\kappa_{K+1} - \kappa_K) \end{bmatrix} \quad (6.7)$$

We can therefore enforce the fitted spline function to be isotonic by using the inequality

$$\mathbf{A}\boldsymbol{\xi} \geq \mathbf{c}, \quad (6.8)$$

and using $\mathbf{c} = \mathbf{0}$ to implement the necessary constraint.

While sampling values for $\boldsymbol{\xi}$, we must ensure that (6.8) is satisfied. Hazelton and Turlach (2011) utilise the truncated multivariate normal with support over $\delta = \delta_{\mathbf{A}, \mathbf{c}} = \{\mathbf{x} : \mathbf{Ax} \geq \mathbf{c}\}$, which is denoted as $\text{TN}_\delta(\boldsymbol{\mu}, \Sigma)$. Hazelton and Turlach (2011) propose a Gibbs sampling algorithm to sample from the appropriate multivariate density, however we instead investigated fitting this constrained model using JAGS, avoiding much of the complicated matrix algebra introduced in their paper, and making the model one we can easily extend from those fitted in previous sections.

We generated an algorithm to create the \mathbf{A} matrix inside of R, which could then be passed to JAGS as data. We placed a multivariate normal distribution on $\boldsymbol{\xi}$, with mean and variance that each have their own hyperprior:

$$\left\{ \begin{array}{l} \boldsymbol{\xi} \sim N(\boldsymbol{\mu}_{\boldsymbol{\xi}}, \Sigma_{\boldsymbol{\xi}}) \\ \boldsymbol{\mu}_{\boldsymbol{\xi}} \sim N(\mathbf{M}_{\boldsymbol{\xi}}, \mathbf{S}_{\boldsymbol{\xi}}) \\ \Sigma_{\boldsymbol{\xi}} \sim W^{-1}(\mathbf{P}_{\boldsymbol{\xi}}, K + 2) \end{array} \right., \quad (6.9)$$

where W^{-1} is the inverse-Wishart distribution (Wishart, 1928), and $\mathbf{M}_{\boldsymbol{\xi}}$, $\mathbf{S}_{\boldsymbol{\xi}}$ and $\mathbf{P}_{\boldsymbol{\xi}}$ are

hyperparameters. These matrices were defined in R and passed to JAGS as data; if vague priors were used (i.e., generic values defined from within JAGS, similar to giving $\mu_{L_{50}}$ a normal prior distribution with mean 0 and precision 1×10^{-6}) severe convergence issues were encountered, mostly due to chains becoming trapped within local modes. Hence the matrices had to be defined carefully so as not to be too informative, but clearly in the region of the true values so convergence was possible.

The full model is shown in JAGS Model B.4.2, most of which is the same as seen previously, however defining η as a polynomial spline as in (6.4) rather than a parametric predictor. Because JAGS does not implement the truncated multivariate normal distribution, we made use of the `ifelse` function to ensure (6.8) was satisfied. This was achieved by defining the node $\mathbf{Ax} = \min(\mathbf{A}\boldsymbol{\xi})$, and using the `ifelse` function when computing the binomial probability,

$$p_i = \begin{cases} \text{logit}^{-1}(\eta_i), & \text{if } \min(\mathbf{A}\boldsymbol{\xi}) \geq 0, \\ -1, & \text{otherwise,} \end{cases} \quad (6.10)$$

which returns a log-likelihood of $-\infty$ if the proposed parameters do not satisfy (6.8), hence rejecting the proposal. Sampling this way, however, results in a fairly inefficient rejection-sampling algorithm. The main problem was that the initial values had to satisfy the constraints, *and* have a reasonably good fit to the data, otherwise the sampler was unable to initialise or the MCMC chains would become trapped in local modes.

The shape-constrained semiparametric regression model was fitted to the stacked covered-codend school prawn data using the model printed in JAGS Model B.4.2. Convergence and stationarity diagnostic tests on the posterior samples passed. A posterior summary is given in Table 6.2, and the estimated spline curve is displayed in Figure 6.2. Most notably, the quadratic spline curve is much smoother than the piecewise linear curve from § 6.1.

Despite obtaining a good fit to the covered-codend school prawn data, such was not the case for the covered-codend rose shrimp data. We had to carefully design the matrices used in the hyperprior distributions for the parameters as shown in (6.9), by first maximising the likelihood function by maximum likelihood techniques. This defies the basic principles

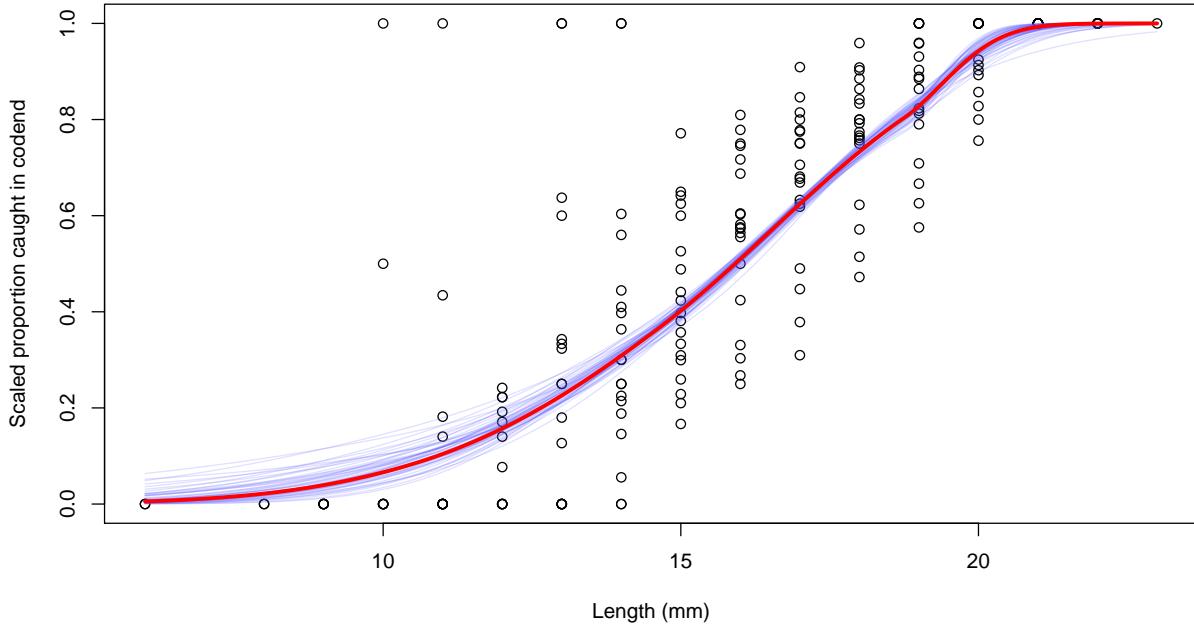


Figure 6.2: The scaled covered-codend school prawn data, with the estimated shape-constrained regression spline. Individual sample estimates are in blue, while the red curve was estimated using the posterior means of the parameters.

Table 6.2: Summary of the posterior distributions of the parameters of the shape-constrained semiparametric model fitted to the stacked covered-codend school prawn data.

Parameter	Mean (SD)	95% Credible Interval
β_0	-0.682 (0.025)	-0.732 – -0.636
β_1	0.144 (0.056)	0.041 – 0.257
β_2	-0.031 (0.017)	-0.067 – 0.000
u_0	0.020 (0.002)	0.015 – 0.025
u_1	0.028 (0.024)	-0.020 – 0.073
u_2	0.437 (0.227)	0.057 – 0.898

of “prior” information, and means that the model is fairly difficult to generalise into an accessible form. Rather than exploring complicated algorithms to generalise this model, we investigated basis splines, which provided a simple framework with which we could formulate models that were easily applied to various data sets without any complicated manipulations or computational difficulties.

6.3 Basis Spline Model

The final type of nonparametric model we investigated was the Basis spline, or B-spline (De Boor, 1978) curve. B-splines are a piecewise q -degree polynomial, with K knot points, producing a smooth, continuous curve. The algorithm developed by De Boor (1978) is used to express each covariate x as a function of these polynomials, producing a $K+q+1$ column matrix, \mathbf{B} . This matrix serves as the design matrix, and the vector of coefficients $\boldsymbol{\alpha}$ are estimated using the predictor

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\alpha}. \quad (6.11)$$

To compute \mathbf{B} for a given set of length classes, we used the `bs` function from the `splines` package in R. We then passed \mathbf{B} to JAGS in order to estimate the coefficients, $\boldsymbol{\alpha}$. For the monotonicity constraints, we used the method implemented by Leitenstorfer and Tutz (2007), who showed that the resulting curve will be isotonic if

$$\alpha_{j+1} \geq \alpha_j. \quad \text{for } j = 1, \dots, K+q, \quad (6.12)$$

We also needed to decide on the number of knot points, K , and the degree of the polynomial, q . Evenly-spaced knot points were used, distributed evenly over the range of length classes, and DIC was used to decide on the best values for K and q . For prior distributions on the coefficients, we needed to ensure that (6.12) was satisfied. For this, we used an autoregressive prior on the coefficients (Neelon & Dunson, 2004; Ruppert et al., 2009), such that

$$\alpha_{j+1} \sim \text{TN}(\alpha_j, \sigma_\alpha^2), \quad \text{truncated so that } \alpha_{j+1} \geq \alpha_j \quad (6.13)$$

(TN represents the truncated normal distribution). This, in turn, allows for a level of smoothness to the curve, rather than the α 's being independent, which led to convergence issues. The initial value, α_1 , was given a uniform prior, $\alpha_1 \sim U[-50, 0]$. This makes the assumption that the selection curve starts off below 0.5, which ensures L_{50} can be estimated

Table 6.3: Summary of the posterior distributions of the parameters of the basis spline model fitted to the stacked covered-codend school prawn data.

Parameter	Mean (SD)	95% Credible Interval
α_1	-4.2 (2.5)	-10.8 – -1.7
α_2	-1.8 (0.4)	-2.7 – -1.1
α_3	-0.9 (0.1)	-1.1 – -0.7
α_4	0.1 (0.0)	-0.0 – 0.2
α_5	1.5 (0.1)	1.3 – 1.7
α_6	4.1 (1.1)	2.2 – 6.5
τ_α	0.3 (0.3)	0.0 – 1.2

for every curve (§ 6.4).

We can easily implement (6.11) in JAGS using the *matrix multiplication operator*, `%*%`. Therefore, the model is again very similar to those seen previously, with only a single change to the node `eta`. We also found that using the normal CDF, Φ , instead of the logistic function, produced a smaller value of DIC, so the probit, Φ^{-1} , was the preferred link function. The model fitted to the stacked covered-codend school prawn data is shown in JAGS Model B.4.3. Using DIC, we found that the best values for q and K were 2 and 3, respectively ($DIC = 76.5$). This is smaller than obtained using the Richards' selection curve in § 4.2.1, ($DIC = 79.9$).

The fitted model used 3 chains, each with 10,000 iterations comprising a burn in of 5,000 and thinning interval of 5, and the obtained samples passed the MCMC diagnostic tests. Figure 6.3 shows the trace and density plots of the posterior samples, which are summarised in Table 6.3. Figure 6.4 shows the basis spline curve drawn from the posterior distribution. The mean curve is computed from the posterior mean estimates of each parameter, while the blue curves are individual samples from the posterior distribution.

The B-spline model was reasonably simple to fit using general functions to manipulate the data, and formulae for the model specification itself. This generality is demonstrated in the latter half of § 6.4, where we apply the model to various other data sets. The α coefficients were easy to estimate in JAGS, and we obtained a smaller DIC value than any of the parametric selection curves previously used. The next step was to find a way of

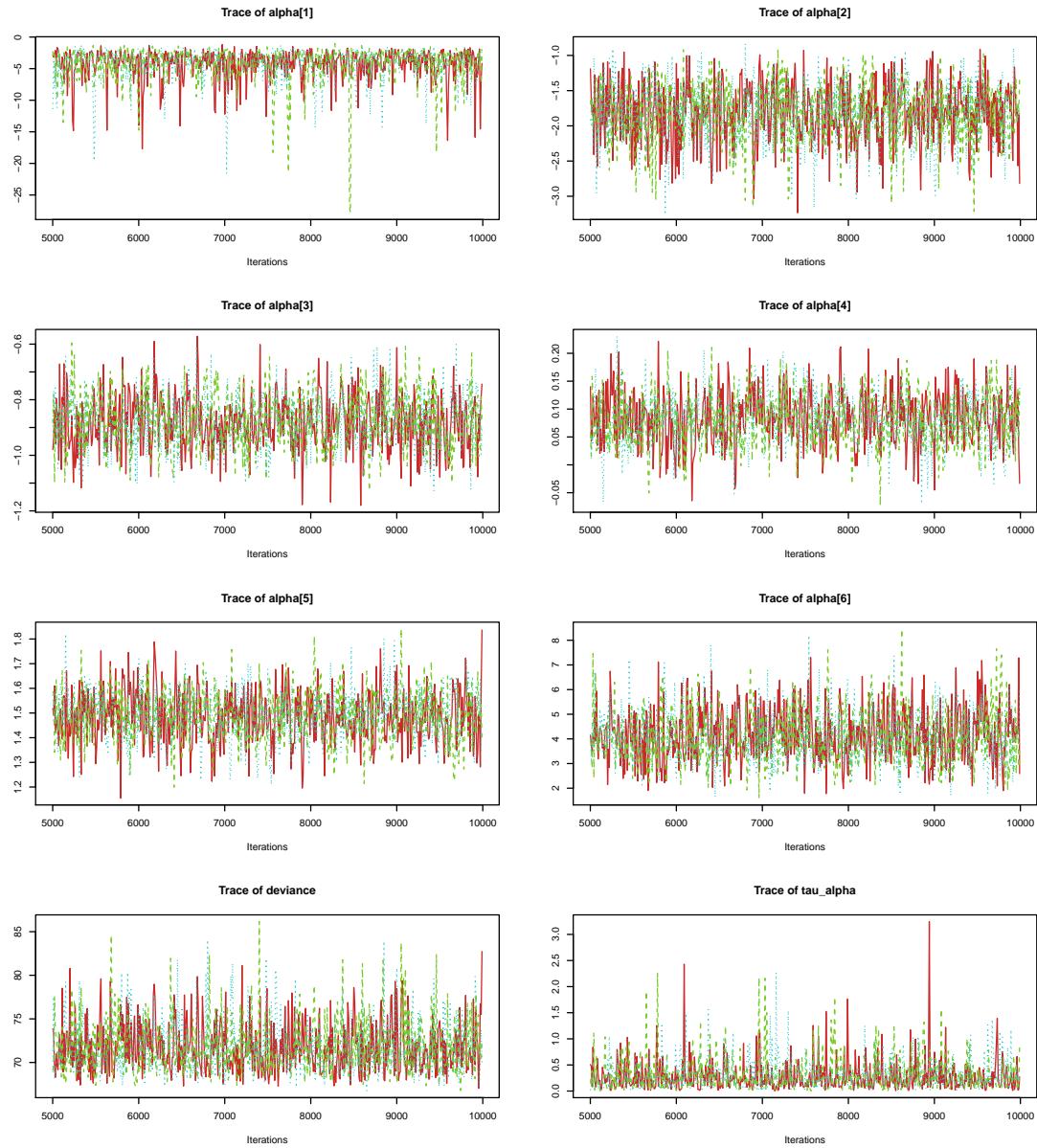


Figure 6.3: Trace plots for the parameters used in the basis spline model fitted to the stacked covered-codend school prawn data.

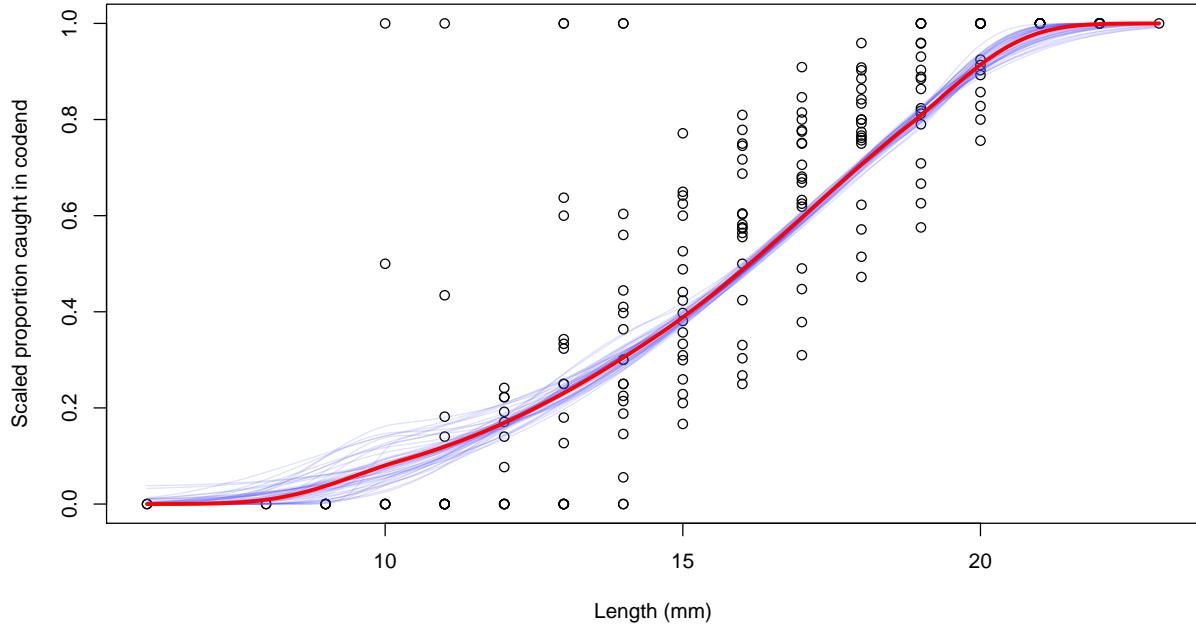


Figure 6.4: The scaled covered-codend school prawn data, with the estimated B-spline selection curves. Individual sample estimates are in blue, while the red curve was estimated using the posterior means of parameters.

quantifying the selection curves in terms that can be understood by fisheries managers, namely L_{50} and SR .

6.4 How Do We Estimate the Parameters of Interest?

A serious practical difficulty with using semiparametric models for size-selectivity analysis is that the predictor can no longer be used to find a closed form solution to

$$\eta_i = \mathbf{B}_i \boldsymbol{\alpha} = \text{logit} (p_i) \quad (6.14)$$

(here, \mathbf{B}_i corresponds to the i^{th} row of \mathbf{B}). This was used in § 2.1.2 to find L_{50} and SR . However, because the curves are continuous, and provided they are strictly isotonic, there will be a unique L_{50} that corresponds to a retention probability of $r(\ell) = 0.5$, and similarly a unique SR . Therefore, if we can find values to correspond to retention probabilities of 0.25,

```

bsplinePars <- function(alpha, B) {
  require(splines, quietly = TRUE)
  K <- attr(B, "knots")
  q <- attr(B, "degree")
  xlim <- attr(B, "Boundary.knots")
  int <- attr(B, "intercept")

  fn <- function(x, p = 0.5) {
    eta <- matrix(bs(x, knots = K, degree = q, intercept = int,
                      Boundary.knots = xlim), nrow = 1) %*% alpha
    lp <- qnorm(p) # using the standard normal CDF
    (eta - lp)^2
  }

  ## Find L50:
  L50 <- optimise(fn, xlim)$minimum
  L25 <- optimise(fn, xlim, p = 0.25)$minimum
  L75 <- optimise(fn, xlim, p = 0.75)$minimum
  SR <- L75 - L25

  structure(c(L25, L50, L75, SR), .Names = c("L25", "L50", "L75", "SR"))
}

```

R Code 6.1: Function used to estimate selectivity parameters from a B-spline curve.

0.5 and 0.75, we can obtain a posterior sample for L_{50} and SR .

We used an optimisation algorithm in tandem with the `bs` function in R to solve (6.14) numerically. We used the R function in R Code 6.1 to numerically estimate L_{50} and SR for a given B-spline matrix and coefficient vector α . The general form of this function will work regardless of the semiparametric predictor being used, although it will need to be modified if the function is not a B-spline, or if a different link function is used (i.e., logit instead of the probit).

6.4.1 Examples Using a B-spline Selection Curve

In this research, we found the best semiparametric model of the three we investigated was easily the B-spline. It was easier to formulate and had much better mixing properties, and model comparison in most cases was fairly straightforward. We also have a method of estimating the selection properties of the gear as required by fisheries managers. We will now apply these techniques to the case studies we have seen previously.

Table 6.4: Summary of the posterior distributions of the selectivity parameters estimated from the B-spline model fitted to the covered-codend school prawn data.

Parameter	Mean (SD)	95% Credible Interval	Richards' Model
L_{25}	13.26 (0.25)	12.72 – 13.71	
L_{50}	16.12 (0.07)	15.98 – 16.27	16.16 (0.07)
L_{75}	18.42 (0.09)	18.26 – 18.61	
SR	5.16 (0.27)	4.67 – 5.73	5.23 (0.30)

Covered-codend school prawn data

We have already fitted the B-spline model to the stacked covered-codend school prawn data in § 6.3. Using the posterior samples of α , and the `bsplinePars()` function in R Code 6.1, we computed posterior samples for L_{25} , L_{50} , L_{75} and SR , which are summarised in Table 6.4. In this case, the estimates are very similar to those obtained from the Richards' fit, and we know from the DIC values that the 2nd order B-spline model with 3 knots gives a better fit to the data. The 3 selection curves (logistic, Richards' and B-spline) are displayed in Figure 6.5. From this plot, we see that the B-spline is able to fit the tail ends of the data better than either of the other curves, which are more restricted in their shape.

Covered-codend rose shrimp data

Next, we fitted the B-spline model to the covered-codend rose shrimp data. However, because the curve is no longer linear on the original scale (before transformation by the inverse link function), we cannot easily incorporate additional covariates (such as gear type). Therefore, we subdivided the data set into six subsets, one for each of the groups (3 mesh sizes and 2 cruises), and fitted a B-spline curve to each of these individually. Another issue we encountered with this data set was that the number of knots, K , was difficult to choose; for most of the subsets, increasing the number of knots decreased the DIC—no single value appeared to be “best”. After about $K = 5$ knots, however, the change in DIC tended to level off, and any more than this results in the curves becoming too data-specific. Therefore, for simplicity we used $K = 5$ knots for the $q = 2$ degree B-splines. The samples for each subset passed the convergence diagnostics, and the posterior distributions of the selectivity

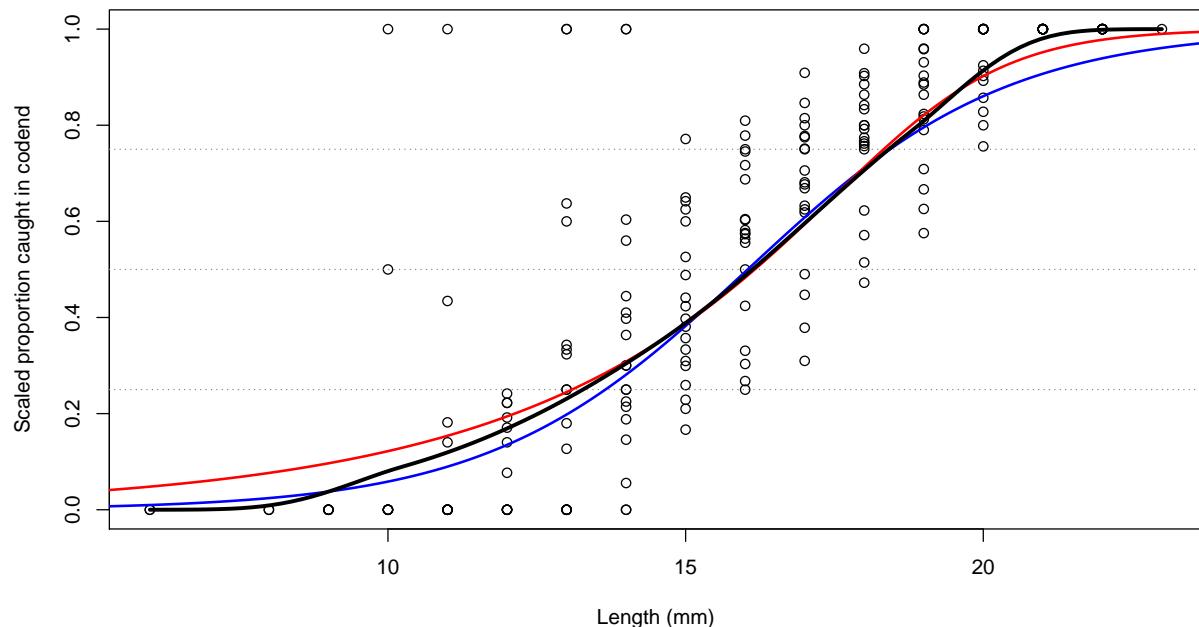


Figure 6.5: The scaled covered-codend school prawn data with the logistic (blue), Richards' (red) and B-spline (black) selection curves fitted using combined-hauls models. The horizontal dotted lines mark 25%, 50% and 75% retention probabilities.

Table 6.5: Summary of the posterior distributions of the selectivity parameters estimated from the B-spline model fitted to the covered-codend rose shrimp data. All of the models use $K = 5$ and $q = 2$.

Cruise	Mesh	L_{25}	L_{50}	L_{75}	SR
Autumn	55 mm	18.07 (1.49)	22.66 (0.05)	25.95 (0.09)	7.89 (1.51)
	70 mm	24.56 (0.05)	26.89 (0.11)	32.40 (0.15)	7.84 (0.16)
	80 mm	26.00 (0.11)	32.95 (0.25)	35.37 (0.40)	9.37 (0.41)
Spring/Summer	55 mm	20.64 (0.20)	21.50 (0.12)	22.96 (0.13)	2.31 (0.26)
	70 mm	24.52 (0.10)	28.61 (0.06)	32.86 (0.07)	8.34 (0.13)
	80 mm	27.53 (0.11)	32.78 (0.16)	33.74 (0.10)	6.21 (0.14)

parameters were calculated, which are displayed in Table 6.5. The estimated curves are shown in Figure 6.6.

One final problem encountered with several of the subsets—autumn cruise, 55 mm mesh, and spring/summer cruise, 70 mm mesh—was that the B-spline does not always fit retention probabilities of 0.25 within the range of the data. In these cases, the estimate of L_{25} is taken from extrapolating the spline curve out of the range of observed lengths for that group, and typically results in an underestimate of L_{25} . This in turn implies an overestimate of SR . In contrast to this, in the spring/summer cruise using 55 mm gear, the curve drops off steeply, which produces a very small estimate of SR . Modelling all of these subsets in a single model would likely help to overcome these problems, as well as the issue of finding an appropriate value of K and q by DIC. However, as we discuss in § 6.5, this is not so easy due to non-linearity on the link scale.

Twin-trawl school prawn data

Finally, we implemented the B-spline predictor function to the twin-trawl data from § 3.4.1, applying the model to the stacked data from each of the three gears separately. This worked in exactly the same way as we have already seen, although we encountered some of the same issues with choosing the number of knots by DIC. However, we initially attempted $K = \{1, 2, \dots, 7\}$ and $q = \{1, 2\}$, and found that, for $q = 2$, DIC decreased from $K = 1$ to 3, and then increased significantly; from there, it declined as K increased. A similar pattern was seen for $q = 1$, although the jump occurred at $K = 5$. Therefore, we used $K = 3$ and $q = 2$ for the three gear subsets. It may be worth noting that, if we were to vary the

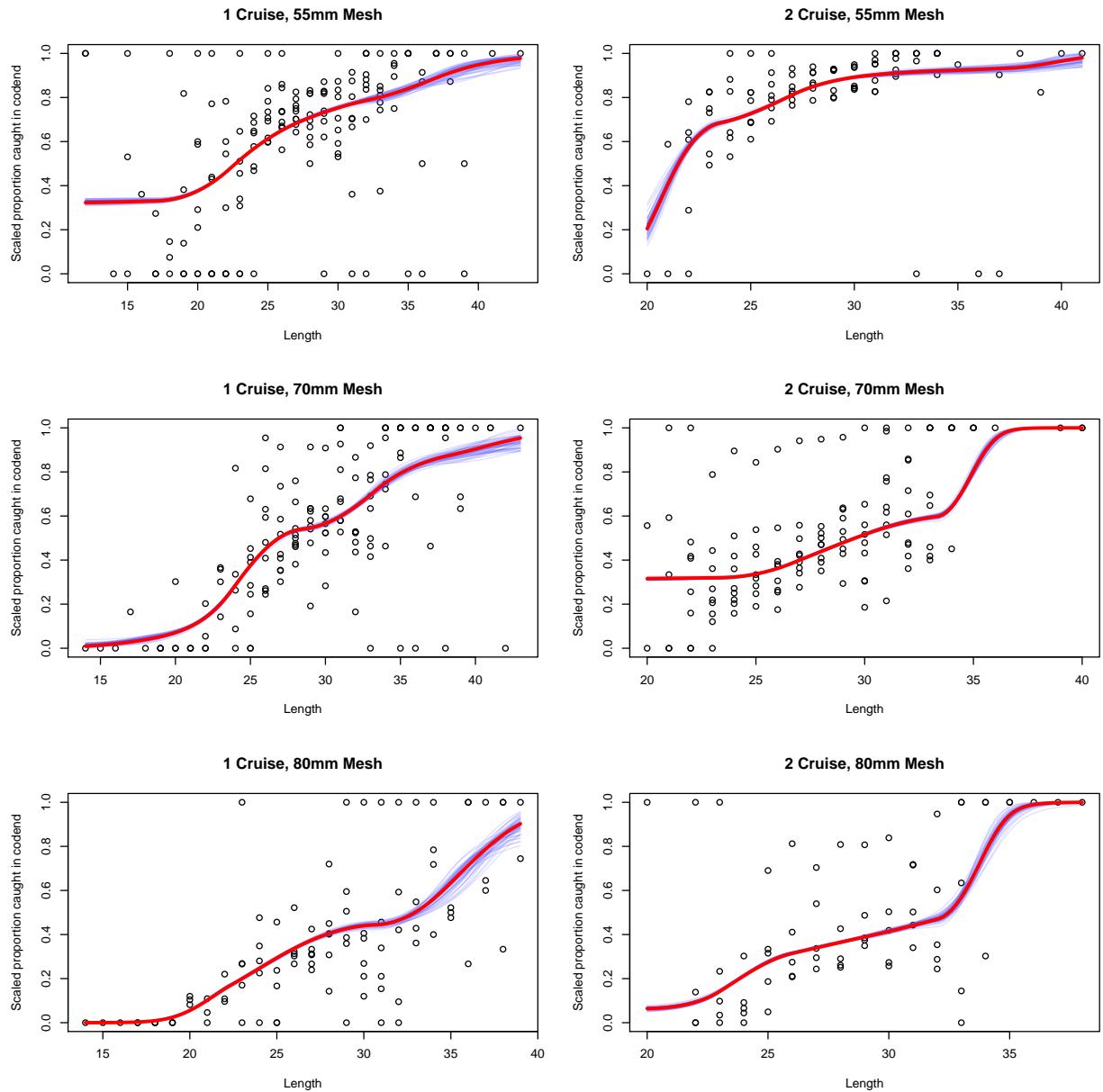


Figure 6.6: The scaled covered-codend rose shrimp data, subsets by mesh (rows) and cruise (columns), with the estimated B-spline selection curves. Individual sample estimates are in blue, while the red curves were estimated using the posterior means of the parameters.

Table 6.6: Summary of the posterior distributions of the selectivity parameters estimated from the B-spline model fitted to the twin-trawl school prawn data. The associated DIC values from the parametric logistic curve were 240, 268 and 255, respectively.

Gear	L_{25}	L_{50}	L_{75}	SR	DIC
40 mm diamond	6.69 (0.17)	8.55 (0.14)	10.57 (0.41)	3.88 (0.35)	239
20 mm tapered square	8.47 (0.06)	10.19 (0.10)	12.01 (0.21)	3.54 (0.17)	243
20 mm untapered square	8.44 (0.07)	10.21 (0.09)	12.22 (0.19)	3.78 (0.18)	250

position of the knots, we may find that the relationship between DIC and number of knots is smoother (see § 6.6).

The semiparametric B-spline model was fitted to each of the three gears separately, using 3 chains of 50,000 iterations comprising a burn in of 25,000, and thinning interval of 25, and MCMC diagnostics of these samples passed. Table 6.6 summarises the posterior distributions of the selectivity parameters as estimated using R Code 6.1. The DIC values of the fits are also displayed, and we see that the semiparametric models give a better fit than the parametric logistic models from § 3.4.1. The estimated selection curves are drawn for the three gears in Figure 6.7.

One final point to consider for these models is overdispersion, which is visible in the plots of the data. Here, we modelled the combined data—scaled and stacked—which obscures overdispersion so it is not detectable. Because of this, the standard errors are likely underestimated. Of course, this can be amended using the same techniques we saw in Chapter 3. Alternatively, we could apply the methods for dealing with overdispersion in a Bayesian model that were discussed in § 4.1, as these make no assumptions about the form of $r(\ell)$. Thus, we could have modelled the unstacked data and added random effects to model overdispersion, although this dramatically increased computational cost and hence was left out of the results presented here. Of course, it makes more sense to model the between-haul variability to account for as much of the overdispersion as possible, which is what we look at in the next section.

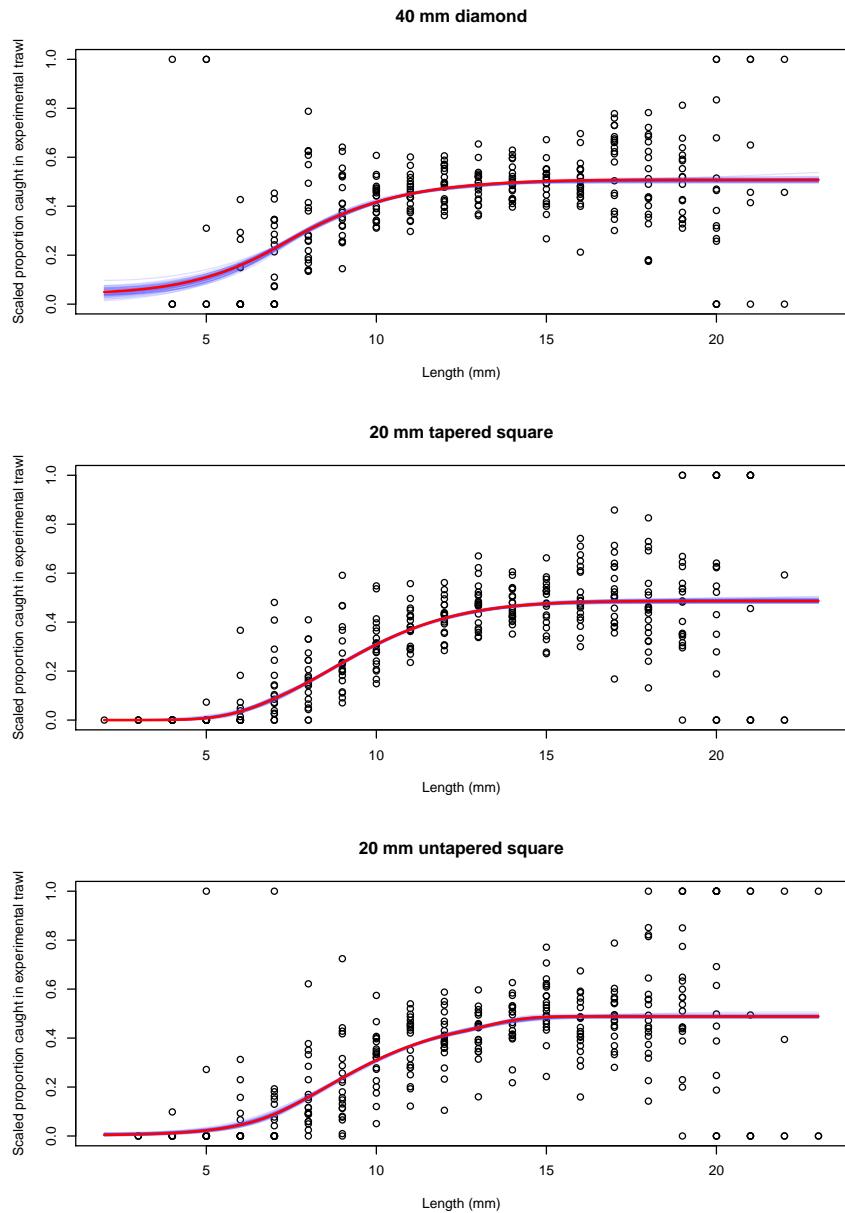


Figure 6.7: The scaled twin-trawl school prawn data, subset by gear, with the estimated B-spline selection curves. Individual sample estimates are in blue, while the red curves were estimated using the posterior means of the parameters.

Table 6.7: Summary of the posterior distributions of the selectivity parameters estimated from the hierarchical B-spline model fitted to the covered-codend school prawn data.

Parameter	Mean (SD)	95% Credible Interval	Logistic Model
L_{25}	13.3 (0.4)	12.4 – 14.0	
L_{50}	15.5 (0.4)	14.7 – 16.2	15.5 (0.3)
L_{75}	17.6 (0.3)	17.0 – 18.2	
SR	4.4 (0.2)	4.0 – 4.9	4.0 (0.2)
σ_h^2	0.2 (0.1)	0.1 – 0.4	
DIC	658.7		659.2

6.5 Hierarchical Semiparametric Model

The final section in this chapter investigates a hierarchical approach to fitting the semiparametric B-spline selection curve. Previously, the data was combined by stacking the hauls, which obscured the between-haul variability. To implement the hierarchical model, a random effect was given to each haul (Ruppert et al., 2003):

$$\eta_{ij} = \mathbf{B}_i \boldsymbol{\alpha} + \varepsilon_j, \quad \varepsilon_j \sim N(0, \sigma_h^2), \quad (6.15)$$

which implies a hierarchical distribution on the intercept term of each haul (Zuur et al., 2002).

The main problem encountered while implementing the hierarchical semiparametric model was that the curve is no longer linear on the link scale, so decreasing the intercept does not correspond to a proportional shift along the x -axis (i.e., increase in L_{50}). As a result, the shape of the curve is inconsistent across hauls. This affect is slightly visible in Figure 6.8, in which the shape of the curves changes from left to right. This effect gets more apparent as q increases.

We used $q = 1$ and $K = 2$ in the final model (decided using DIC), the results of which are displayed in Table 6.7. The estimates are similar to those from the hierarchical logistic model fitted back in § 3.2.2, though now the shape of the curve is no longer constrained by a parametric function. We still obtain an estimate of variability, σ_h^2 , but it is no longer as intuitive to fisheries managers as $\sigma_{L_{50}}^2$; instead, it is the variability of the intercepts on the link scale.

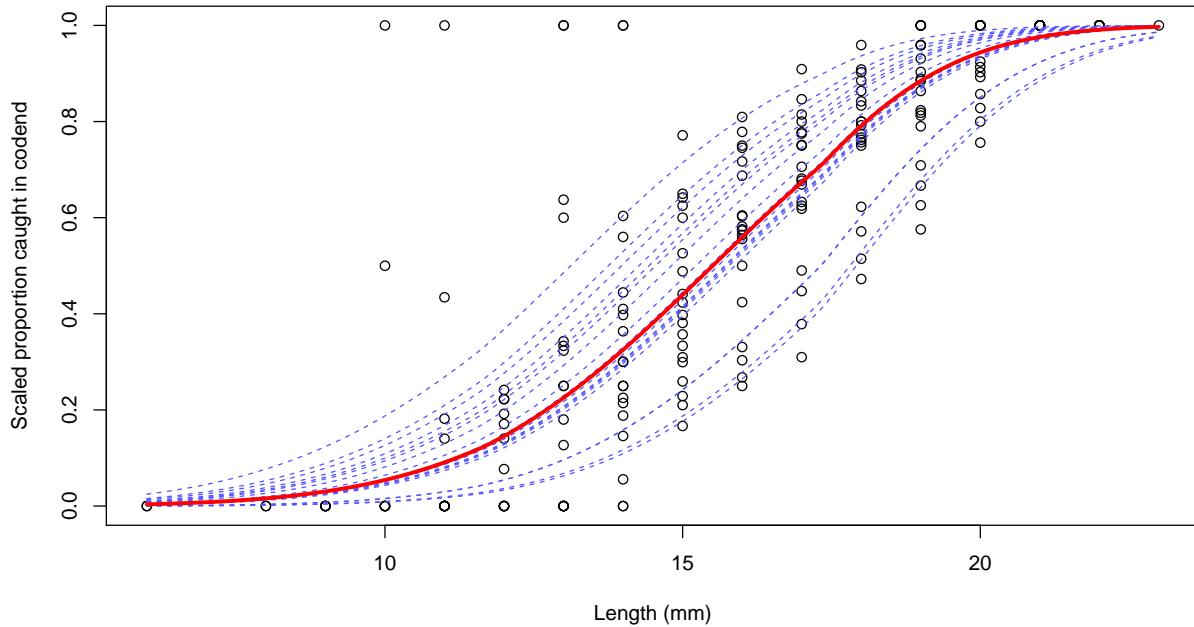


Figure 6.8: The scaled covered-codend school prawn data, with the estimated B-spline selection curves fitted using a hierarchical model. Individual haul estimates are in blue, while the red curve represents the conditional means selection curve. The shape of the curves is no longer consistent from left to right due to the nonlinearity of the curve on the link scale.

6.6 Chapter Summary

Semiparametric models have great potential in all areas of statistics, not just for size-selectivity modelling. They free researchers from imposing assumptions on the shape of relationships, but also provide the ability to define meaningful ones, such as the isotonic constraint we enforced here. In this work, we only imposed a non-decreasing constraint, however this did lead to some fairly unusual curves (see Figure 6.6). One area in which research could be focused is potential methods of restricting the shape of the curve to have the typical “S” shape associated with selection curves. This could be imposed by constraining the first and second derivatives of the function (i.e., a unimodal first derivative, or isotonic second derivative). Doing so would produce nicely shaped selection curves that would likely be more appealing to fisheries managers.

It may also be beneficial to investigate models with flexible knot locations, rather than

fixing them evenly along the range of x values. This could allow both better fitting curves, and potentially amend the problem associated with selecting K and q by DIC. However, such an approach would require a complicated two-stage algorithm, in which the positions of knots are first proposed, and then the coefficients for that proposal estimated.

While the B-spline seems to be a great candidate for modelling experimental trawl data, there are several challenges that must be overcome. These include modelling covariates, and between-haul variability, which both involve additive terms on the link scale, which no longer translate to a horizontal shift of the curve on the probability scale. While semiparametric mixed models have been used in other applications (Ruppert et al., 2003, 2009), these have been modelled on the same scale as the data, so no transformation by way of a link-function was necessary. Further study into this and other problematic cases needs to be carried out to find a method of *horizontally* shifting the spline curve to allow the selection curves to have the same *shape*, but different retention properties.

Chapter 7

BSM: An R Package for Bayesian Selectivity Modelling

During our work formulating Bayesian models for analysing the size-selectivity of trawl gear, we started work on an R package to allow a much easier workflow. In this chapter, we will give a very brief overview of BSM, explaining some of the main features and perform a simple analysis of the covered-codend school prawn data initially seen in Chapter 3. The package is still under development, hosted on GitHub at <https://github.com/tell029/bsm>, and can be installed using the `devtools` package (Wickham & Chang, 2014):

```
library(devtools)
install_github("tell029/BSM")
library(BSM)
```

7.1 Loading and Displaying Data

The first step in any analysis is to load the data and manipulate it into the necessary form. A graphical inspection of the data would then be used to check for obvious problems. BSM assumes the original data are in column-form, loaded into R as a dataframe.

```

school.df <- read.table("data/LgSq.dat", header = TRUE)
head(school.df)

##   BLOCK DAY SHOT codend cover REP WtSch     q1     q2 total Lenclass Haul
## 1     2    1     3      0     0    1  2.02 0.585 0.34      0       6     1
## 2     2    1     3      0     0    1  2.02 0.585 0.34      0       7     1
## 3     2    1     3      0     0    1  2.02 0.585 0.34      0       8     1
## 4     2    1     3      0     0    1  2.02 0.585 0.34      0       9     1
## 5     2    1     3      0     0    1  2.02 0.585 0.34      0      10     1
## 6     2    1     3      0     2    1  2.02 0.585 0.34      2      11     1

```

The BSM package utilises a `bsmdata` object, enabling easy plotting and modelling of the data. To generate the necessary object, we just use the `bsmData` function, which takes various arguments to describe the data set (more information can be found by typing `?bsmData` into the R console).

```

school.data <- bsmData(y1 = codend, y2 = cover, q1 = q1, q2 = q2,
                        length = Lenclass, length.unit = "mm", haul = Haul,
                        weight = WtSch, data = school.df)
school.data

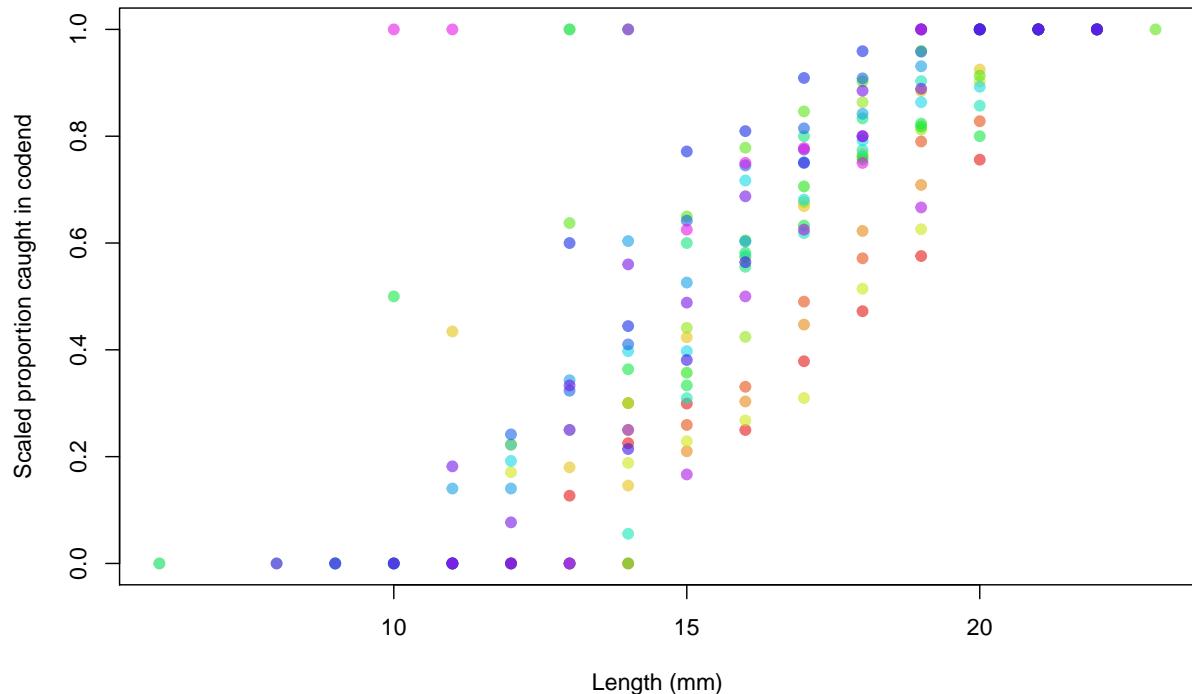
##
## Data set for covered-codend experimental trawl.
##
## Total of 342 observations from 19 hauls.
## Lengths range from 6 to 23mm.
##

```

The `bsmData` function allows the experimental structure to easily be defined. Several additional arguments not used above are: `n`, the total count (`= y1 + y2` by default); and `paired`, a logical argument to define if the experiment was paired or not (for example, in a twin- or trouser-trawl, you would set `paired = TRUE`). Users can also provide any extra variables; in the example above, `weight` is not a formal argument, instead the variable will be included in the object, allowing it to be used as a covariate in the model.

The data can be plotted using the `plot` function, which takes several arguments. Notable ones include: `scale`, a logical variable (which defaults to `TRUE`), telling the function whether or not to plot the scaled (by sampling fraction) or unscaled data; `col` and `pch` take either their typical values (i.e., `col = "red"`, `pch = 19`), or can be the name of a (factor) variable (such as `haul`) that is used to colour the points to see trends. Here, we plot the stacked covered-codend school prawn data, colouring the points by haul:

```
plot(school.data, col = "haul")
```



Colouring by a variable becomes more useful when there are clear trends in the data, such as we saw in § 3.3, where the points are coloured according to mesh size, showing a clear relationship between selectivity and mesh size (Figure 3.6 on page 36). However, the colours in the plot above show that the variance is not completely random; there is clearly variation between the retention curves of each haul.

7.2 Fitting a Combined-hauls Selectivity Model

Having graphically inspected the data, we can fit a selection curve. Here, we are going to fit a combined-hauls fit, and in § 7.3 a hierarchical model, similar to those demonstrated in § 3.2. For the combined-hauls fit, we use a Richards' selection curve applied to the combined-hauls data. This uses the `bsm()` function, which takes the data object we created in § 7.1, as well as various arguments to specify the model, such as the type of selection curve, and whether or not to combine the data (this uses the stacked, scaled approach). We can also

define arguments for the MCMC chain, such as the number of samples to generate, thinning intervals, burn in period, and number of chains. The function returns an object of class `bsmfit`. See `?bsm` for more details about the function.

```

school.comb <- bsm(school.data, curve = "richards", combine = TRUE,
                     n.samples = 2000, n.thin = 5, quiet = TRUE)
summary(school.comb)

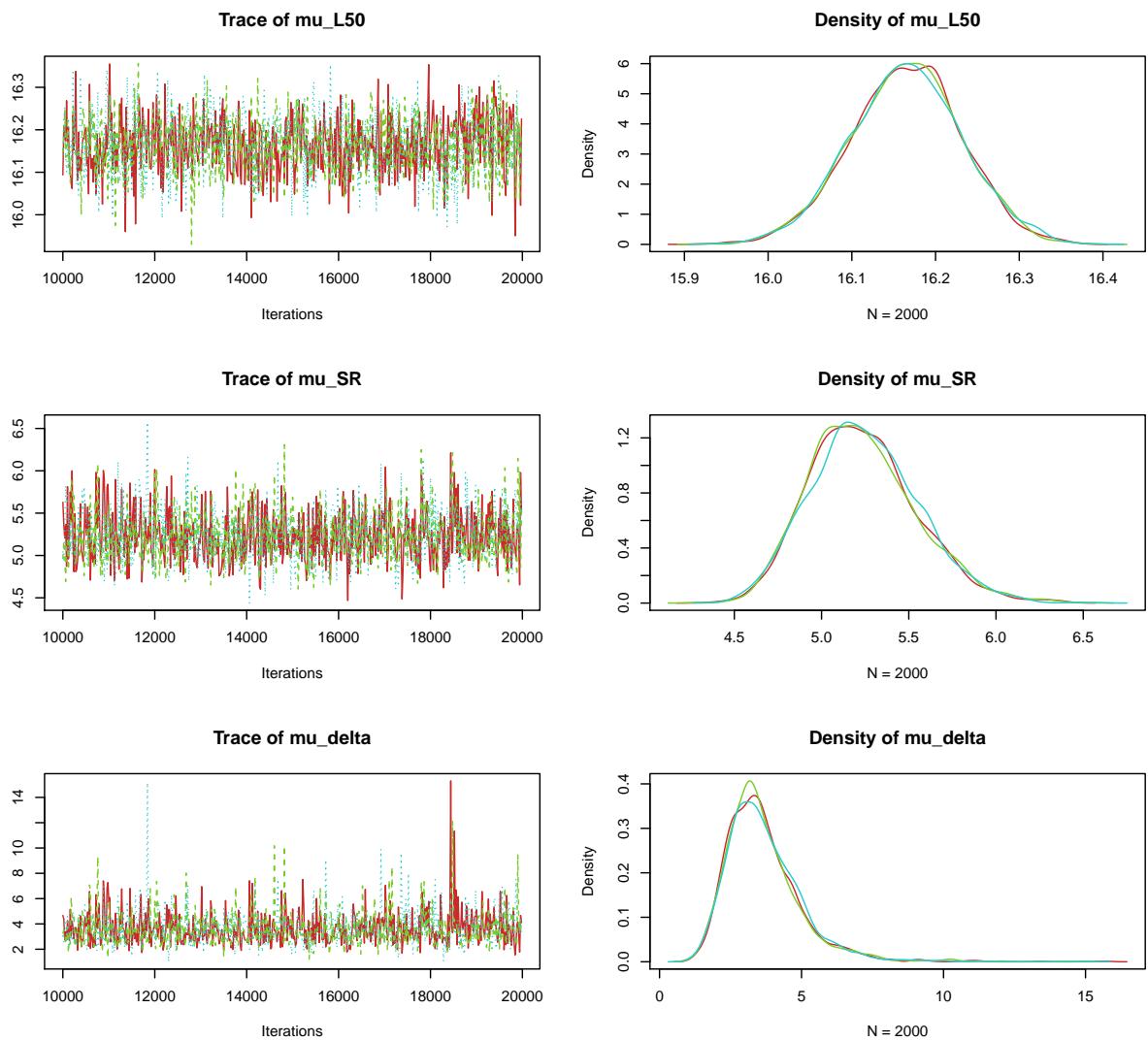
## 
## Summary of model fitted to covered-codend experimental trawl using JAGS.
##
## Likelihood: binomial
## Curve: richards
##
## Summary of posterior distribution for the model:
##
##              mean        sd      2.5%     50%    97.5%   Rhat
## mu_L50    16.165  0.06612  16.033  16.166  16.293  1.001
## mu_SR     5.238  0.30679   4.702   5.213   5.891  1.001
## mu_delta   3.658  1.34128   1.773   3.435   6.866  1.001
##
## Overdispersion information:
##
## Expected Pearson Chi-square: 16.91
## Observed Pearson Chi-square: 10.25
## Pr[overdispersion]: 0.19
##
## MCMC Information:
##   Chains Iterations/chain Burn-in Thin Total
##         3             20000    10000     5    6000
##
## Model DIC = 79.888, pD = 3.1339
## Gelman and Rubin Convergence Statistic (multivariate) = 1.001
## Heidelberger & Welch's Stationarity and Half-width Tests passed.

```

The summary output using the `summary` function provides an overview of the samples generated by the MCMC algorithm. The `bsm()` function is a wrapper for the `R2jags` package by Su and Yajima (2014), which itself calls JAGS (Plummer, 2003); we use the DIC information provided by `R2jags` in the summary output. The overdispersion information is computed as discussed in § 4.1. Finally, several of the diagnostic MCMC checks are called from the `coda` package (Plummer et al., 2006).

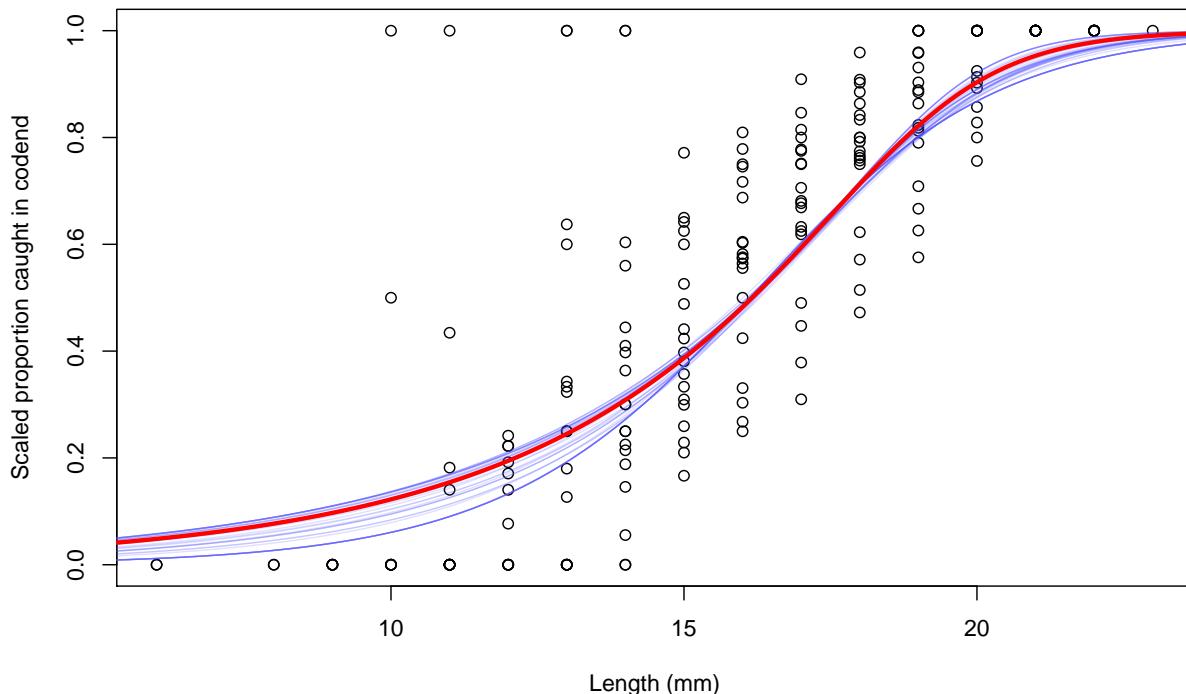
We can plot the posterior samples using the `plot()` method; by default, this shows the trace plots and density plots for the parameters. Other options include `which = 'pairs'`, which generates a pairs plot of the parameters, useful for detecting strong correlations. Here, the `trace.samples` argument allows users to specify how many samples from each chain are used to draw the trace plots.

```
plot(school.comb, parameters = c("mu_L50", "mu_SR", "mu_delta"),
      trace.samples = 500)
```



A plot of the posterior curve can also be generated easily from the fitted object, and can include a credible region of the selectivity curves from the samples, by supplying the arguments `which = "curve"` and `cred.int = TRUE`. An example of this is shown in Figure 5.2 on page 70. Alternatively, a sample of selection curves can be drawn directly from the posterior distribution by using the `postsample()` and `bsmCurve()` functions:

```
plot(school.data)
apply(postsample(school.comb, n = 50, pars = c("mu_L50", "mu_SR", "mu_delta")),
      1, bsmCurve, lty = 1, col = "#5555FF30") -> o
plot(school.comb, which = "curve", new = FALSE, col = "red", lwd = 3)
```



7.3 Fitting a Hierarchical Model

To implement random effects or other model covariates, we have used *formula* notation. That is, to model L_{50} as a function of catch weight and haul (i.e., random haul effects), we use `L50 = ~weight + haul`. We fit this model to the data as follows:

```

school.hier <- bsm(school.data, curve = "logistic", L50 = ~weight + haul,
                     n.samples = 2000, n.thin = 2, quiet = TRUE)
summary(school.hier)

##
## Summary of model fitted to covered-codend experimental trawl using JAGS.
##
## Likelihood: binomial
## Curve: logistic
##
## Summary of posterior distribution for the model:
##
##          mean        sd      2.5%     50%    97.5%   Rhat
## mu_L50  15.5066  0.3529  14.7921  15.5106  16.199  1.001
## mu_SR   4.0398  0.1588   3.7462   4.0344   4.364  1.001
## sig2_L50 2.1578  0.9191   0.9912   1.9650   4.468  1.002
## beta    0.4551  0.5430  -0.6222   0.4561   1.548  1.001
##
## Formula for selection curve parameter:
##   L50 = mu_L50 + beta * (weight - 1.064)
##
## Overdispersion information:
##
##   Expected Pearson Chi-square: 212.5
##   Observed Pearson Chi-square: 184.4
##   Pr[overdispersion]: 0.25
##
## MCMC Information:
##   Chains Iterations/chain Burn-in Thin Total
##         3             8000      4000    2   6000
##
##   Model DIC = 660.08, pD = 20.001
##   Gelman and Rubin Convergence Statistic (multivariate) = 1.001
##   Heidelberger & Welch's Stationarity and Half-width Tests passed.

```

In the above summary, we see that the coefficient associated with weight is not significantly different from 0—the credible interval easily contains 0. This does, however, make it easy to fit covariates to the model, as the same expression exists for SR and ϕ . We are currently looking at ways of allowing the removal of the intercept ($\mu_{L_{50}}$), so models such as

those used for the rose shrimp data can be fit, and though the technical aspects are yet to be sorted, the syntax will likely be `L50 = ~mesh + cruise + haul - 1.`

There are numerous other options for the BSM package, however the exact method used to implement these (such as covariates for L_{50} in the hierarchical models) is still being explored. The most up-to-date documentation can be obtained by installing the package from GitHub and accessing the functions' help files (e.g., `?bsm`, `?summary.bsmfit`, ...).

Chapter 8

Summary of Bayesian Size-selectivity Modelling

Understanding the size-selectivity properties of fishing gear is important to fisheries managers, as it allows them to assess profits and various other stock attributes such as incidental mortality rates. Size-selectivity models for towed trawl experiments have been used extensively to estimate the size-selectivity properties of gear, with special significance given to the parameters L_{50} , the length of an animal with 50% probability of retention, and selection range, SR , the difference between the lengths of animals with 25% and 75% retention probabilities.

Typical models used for size-selectivity analysis include the SELECT method to perform a combined-hauls analysis and estimate the marginal selection-curve, and more complex mixed-models which estimate the mean selection curve conditional on the individual haul random effects, and explicitly model between-haul variation. The latter approach provides a formal basis for hypothesis testing and comparing the selectivity of multiple gears. These have previously been fitted using software such as SAS, although some initialisation and convergence problems occurred when attempting to fit complex random effects models.

Bayesian models have also been used to analyse trawl data, though they have typically used inefficient parameterisation of the predictor. In our work, we were able to formulate

Bayesian models that implemented efficient parameterisations of the predictor using the parameters of interest for not only the logistic curve, but for the Richards' curve as well, which—as far as we are aware—is the first time this has been done. The major benefit of this was that it removed almost all of the correlation between L_{50} and SR , allowing us to fit the Richards' selection curve in combined-haul and hierarchical models with little difficulty. Additionally, this parameterisation allowed L_{50} and SR to be modelled directly by other covariates, such as mesh size, or with random effects to allow for between-haul variability in these parameters.

While fitting the combined-hauls and hierarchical models, we purposefully compared our results to those obtained using other accepted methods in order to show that our approach works at least as well. We were then able to explore the flexibility of the Bayesian models, implementing a Poisson likelihood to provide a means of estimating—or making use of—the population size distribution. While we would not conclude that the Bayesian approach is better all around, it is definitely a strong candidate for modelling between-haul variability, as we saw in the rose shrimp data set; previous modelling attempts had been unable to model L_{50} and SR as random, whereas our hierarchical model could, and found both to vary considerably between hauls. In contrast to this, it is arguable that the frequentist model is better suited to the combined-hauls approach, as it provides a sound framework for incorporating overdispersion and sample size inflation, which cannot be done in the same way within the Bayesian paradigm.

In our research, we started the exploration of Bayesian semiparametric models, and found that B-splines were fairly easy to fit and generally had a lower DIC than parametric alternatives. More importantly, we were still able to estimate L_{50} and SR , along with their errors and credible intervals, opening up the possibility to fit selection curves with no unnecessary shape constraints.

One big contribution of our work is—we hope—the BSM package. This R package provides easy to use functions within the R environment, and assumes no knowledge of BUGS modelling, allowing users with only a basic understanding of R to access and use the models

we discussed throughout this thesis. It provides flexibility for application to both covered-codend and paired (twin- and trouser-) trawl experiments, and, while it is still under development at the time of writing, we hope to extend the feature set to include covariate models (such as those used on the rose shrimp data) and potentially incorporate B-spline models as well.

8.1 Future Work

The work presented in this thesis was primarily focused on implementing Bayesian alternatives of models already accepted in the field; however, some new ideas were explored. One of these was the weighted hierarchical model, which was a novel attempt at integrating the marginal combined-hauls and conditional mixed-effects approaches. We were able to use this to estimate the marginal “population” estimate of L_{50} , which was similar to that obtained from the combined-hauls model. However, we were unable to find a method for doing the same with SR . Further research into weighted (or other) hierarchical models would be useful, as this would allow us to model the data completely without prior modification (i.e., scaling), include overdispersion and between-haul variability, and still obtain a useful marginal estimate of $r(\ell)$, the contact-selection curve.

The use of population size distributions was briefly considered, though we did not have access to any appropriate data sets with available size distribution estimates, so were unable to follow this any further. We were able to show how the length class distribution could easily be implemented into the model, which could be given a prior distribution if it is known. In the more general context of fisheries stock assessment, the estimation of gear selectivity may be just one of several components in an overall model that does include information on the population size distribution, in which case this approach could be applicable and useful.

Semiparametric selection curves were introduced toward the end of this thesis, and we investigated three competing approaches. While these models worked well for the combined-hauls data, we encountered several problems in the hierarchical approach. One of these was

how to get meaningful estimates of between-haul variability that describe the variance of the selection properties (such as L_{50}) in the population. Ideally, a method of horizontally shifting the curve would enable us to model each haul using a curve with the same shape, rather than a shape that changes as it shifts from left to right due to non-linearity on the link scale.

Another possibility for semiparametric splines is to allow the knots to move, enabling the sampler to find the “best” knot arrangement. Several algorithms exist for fitting splines without fixed knot positions, as well as flexible numbers of knots, but these typically require multiple-stage MCMC algorithms to first propose the knot locations, and then sample the coefficients. Further, new priors could be investigated for the parameters to encourage additional shape constraints, such as a unimodal first derivative, producing a curve with the typical “S” shape familiar to fisheries managers. These two extensions, in addition to discovering a means for modelling covariates and between-haul variability, will make spline models a very strong competitor as a choice of selection curve.

While size-selectivity of fishing gear has been heavily researched, there are still plenty of areas for future studies, only some of which have been touched on here. We hope that our work will provide a foundation for any future studies looking to investigate the modelling of experimental trawl data in the Bayesian paradigm, and of course hope that the BSM package helps to make these models accessible to a wider range of users.

Appendix A

Derivation of L_{50} and SR for the Richards' Curve

The regression coefficients for the Richards' Curve implementation of the link function

$$p = \left(\frac{e^\eta}{1 + e^\eta} \right)^{1/\delta}, \quad \eta = \beta_0 + \beta_1 x \quad (\text{A.1})$$

can be found by solving for the values of $p = [0.25, 0.5, 0.75]$.

First, the general solution to $p = \frac{a}{b}$. Simple rearrangement of (A.1) yields

$$\begin{aligned} \left(\frac{e^\eta}{1 + e^\eta} \right)^{1/\delta} &= \frac{a}{b} \\ \frac{e^\eta}{1 + e^\eta} &= \frac{a^\delta}{b^\delta} \\ b^\delta e^\eta &= a^\delta (1 + e^\eta) \\ e^\eta &= \frac{a^\delta}{b^\delta - a^\delta} \\ \eta &= \delta \log(a) - \log(b^\delta - a^\delta) \end{aligned}$$

and we can substitute the value of η when $x = L_x$

$$\begin{aligned}\beta_0 + \beta_1 L_x &= \delta \log(a) - \log(b^\delta - a^\delta) \\ L_x &= -\frac{\beta_0 - \delta \log(a) + \log(b^\delta - a^\delta)}{\beta_1}\end{aligned}\quad (\text{A.2})$$

We can now use the general result in (A.2) to find an expression for L_{50} , corresponding to $p = 0.5 = \frac{1}{2}$, and $SR = L_{75} - L_{25}$, corresponding to $p = [0.75, 0.25] = [\frac{3}{4}, \frac{1}{4}]$ respectively:

$$\begin{aligned}L_{50} &= -\frac{\beta_0 - \delta \log(1) + \log(2^\delta - 1^\delta)}{\beta_1} = -\frac{\beta_0 + \log(2^\delta - 1)}{\beta_1} \\ L_{75} &= -\frac{\beta_0 - \delta \log(3) + \log(4^\delta - 3^\delta)}{\beta_1} \\ L_{25} &= -\frac{\beta_0 - \delta \log(1) + \log(4^\delta - 1)}{\beta_1} = -\frac{\beta_0 + \log(4^\delta - 1)}{\beta_1} \\ SR &= -\frac{\beta_0 - \delta \log(3) + \log(4^\delta - 3^\delta)}{\beta_1} + \frac{\beta_0 + \log(4^\delta - 1)}{\beta_1} \\ SR &= \frac{\delta \log(3) - \log(4^\delta - 3^\delta) + \log(4^\delta - 1)}{\beta_1}\end{aligned}$$

Having obtained expressions for L_{50} and SR , we can now rewrite (A.1) with the new parameterisation:

$$\begin{aligned}\beta_1 &= \frac{\delta \log(3) - \log(4^\delta - 3^\delta) + \log(4^\delta - 1)}{SR} \\ \beta_0 &= -\beta_1 L_{50} - \log(2^\delta - 1) \\ &= -\frac{\delta \log(3) - \log(4^\delta - 3^\delta) + \log(4^\delta - 1)}{SR} L_{50} - \log(2^\delta - 1) \\ \eta &= \beta_0 + \beta_1 x \\ &= -\frac{\delta \log(3) - \log(4^\delta - 3^\delta) + \log(4^\delta - 1)}{SR} L_{50} - \log(2^\delta - 1) \\ &\quad + \frac{\delta \log(3) - \log(4^\delta - 3^\delta) + \log(4^\delta - 1)}{SR} x\end{aligned}$$

which can finally be rearranged to give

$$\eta = \frac{\delta \log(3) - \log(4^\delta - 3^\delta) + \log(4^\delta - 1)}{SR} (x - L_{50}) - \log(2^\delta - 1) \quad (\text{A.3})$$

We can check that this matches the results for the logistic function by substituting $\delta = 1$ into (A.3):

$$\eta = \frac{\log(3) - \log(1) + \log(3)}{SR} (x - L_{50}) - \log(1) = \frac{2 \log(3)}{SR} (x - L_{50}),$$

which is the same as required.

Appendix B

JAGS Models

Here, many of the JAGS models used in our work are displayed. Many of these are simple extensions of the others, and were generated by the BSM package. Models denoted with a † were modified manually from a model output by BSM, because the package was still under development and unable to generate these models correctly at the time.

Note: throughout this section, the JAGS models use i and j matrix notation, where $i = \ell =$ length class and $j = h =$ haul.

B.1 Combined-hauls Models

These models are the same for combined-hauls data and uncombined data, stacked or unstacked. The only difference in these cases are the values passed in as data.

B.1.1 Covered-codend

```

model {
  for (i in 1:N) {
    for (j in 1:M) {
      ## The likelihood on the observed counts:
      y[i, j] ~ dbin(p[i, j], n[i, j])

      ## The implementation of the selection curve:
      p[i, j] <- (q1[j]/q2[j]) * r[i, j] / (1 - (1 - q1[j]/q2[j]) * r[i, j])
      r[i, j] <- ilogit(eta[i, j]) ^ (1 / delta[j])
      eta[i, j] <- etahat[i, j]
      etahat[i, j] <- (delta[j] * log(3) - log(4^delta[j] - 3^delta[j]) +
        log(4^delta[j] - 1)) / SR[j] * (x[i] - L50[j]) - log(2^delta[j] - 1)
    }
  }

  for (j in 1:M) {
    ## Parameter values for each haul
    L50[j] <- mu_L50
    SR[j] <- mu_SR
    delta[j] <- mu_delta
  }

  ## Top-level priors for 'hyper' parameters:
  mu_L50 ~ dnorm(0, 1E-6)T(0,)
  mu_SR ~ dlnorm(0, 1E-5)
  mu_delta ~ dlnorm(0, 1E-3)
}

```

A simple binomial model for a covered-codend experiment, using the Richards' selection curve parameterised by L_{50} and SR .

The node `eta[i, j]` is present for addition of random observation effects to account for overdispersion (JAGS Model B.3.2).

The “Parameter values for each haul” section allow additional covariates to be used to model the parameters (JAGS Model B.2.2).

B.1.2 Paired trawl

```

model {
  for (i in 1:N) {
    for (j in 1:M) {
      ## The likelihood on the observed counts:
      y[i, j] ~ dbin(p[i, j], n[i, j])

      ## The implementation of the selection curve:
      p[i, j] <- q1[j] * phi[j] * r[i, j] /
        (q1[j] * phi[j] * r[i, j] + q2[j] * (1 - phi[j]))
      r[i, j] <- ilogit(eta[i, j])
      eta[i, j] <- etahat[i, j]
      etahat[i, j] <- 2 * log(3) / SR[j] * (x[i] - L50[j])

    }
  }

  for (j in 1:M) {
    ## Parameter values for each haul
    L50[j] <- mu_L50
    SR[j] <- mu_SR
    phi[j] <- ilogit(phiT[j])
    phiT[j] <- mu_phi
  }

  ## Top-level priors for 'hyper' parameters:
  mu_L50 ~ dnorm(0, 1E-6)T(0, )
  mu_SR ~ dlnorm(0, 1E-5)
  mu_phi ~ dnorm(0, 1E-3)
}

```

A combined-hauls model for paired trawl data. While this model was used here for twin-trawl data, it can be applied to any simultaneous paired trawl experiment, twin or trouser, without any modification to the model.

B.2 Hierarchical Models

These models are essentially modifications of those seen in § B.1. The selectivity parameters are given a hierarchical prior, with the necessary priors for the variance parameters.

B.2.1 Covered-codend

```

model {
  for (i in 1:N) {
    for (j in 1:M) {
      ## The likelihood on the observed counts:
      y[i, j] ~ dbin(p[i, j], n[i, j])

      ## The implementation of the selection curve:
      p[i, j] <- ilogit(eta[i, j])
      eta[i, j] <- etahat[i, j]
      etahat[i, j] <- 2 * log(3) / SR[j] * (x[i] - L50[j])
        + log(q1[j] / q2[j])
    }
  }

  for (j in 1:M) {
    ## Parameter values for each haul
    L50[j] ~ dnorm(L50_bar[j], tau_L50)
    L50_bar[j] <- mu_L50
    SR[j] <- mu_SR
  }

  ## Top-level priors for 'hyper' parameters:
  mu_L50 ~ dnorm(0, 1E-6)T(0,)
  mu_SR ~ dlnorm(0, 1E-5)

  ## Precision parameters for random effects
  tau_L50 <- 1 / sig2_L50
  sig2_L50 <- sig_L50^2
  sig_L50 ~ dunif(0, 100)
}

```

A hierarchical binomial model for a covered-codend experiment.

B.2.2 Covered-codend with covariates[†]

```

model {
  for (i in 1:N) {
    for (j in 1:M) {
      ## The likelihood on the observed counts:
      y[i, j] ~ dbin(p[i, j], n[i, j])

      ## The implementation of the selection curve:
      p[i, j] <- ilogit(eta[i, j])
      eta[i, j] <- etahat[i, j]
      etahat[i, j] <- 2 * log(3) / SR[j] * (x[i] - L50[j])
      + log(q1[j] / q2[j])
    }
  }

  for (j in 1:M) {
    ## Parameter values for each haul
    L50[j] ~ dnorm(L50_bar[j], tau_L50)
    L50_bar[j] <- alpha[cruise[j]] + beta * mesh[j]
    SR[j] <- mu_SR
  }

  mu_SR ~ dlnorm(0, 1E-5)

  ## Precision parameters for random effects
  tau_L50 <- 1 / sig2_L50
  sig2_L50 <- sig_L50^2
  sig_L50 ~ dunif(0, 100)

  ## Priors for L50 covariates:
  alpha[1] <- 0
  alpha[2] ~ dnorm(0, 1E-6)
  beta ~ dnorm(0, 1E-6)
}

```

A hierarchical model used for modelling the covered-codend rose shrimp data; the covariates were incorporated into the model by modifying the node `L50_bar`. These parameters were also given appropriate prior distributions.

B.2.3 Paired trawl

```

model {
  for (i in 1:N) {
    for (j in 1:M) {
      ## The likelihood on the observed counts:
      y[i, j] ~ dbin(p[i, j], n[i, j])

      ## The implementation of the selection curve:
      p[i, j] <- q1[j] * phi[j] * r[i, j] /
        (q1[j] * phi[j] * r[i, j] + q2[j] * (1 - phi[j]))
      r[i, j] <- ilogit(eta[i, j])
      eta[i, j] <- etahat[i, j]
      etahat[i, j] <- 2 * log(3) / SR[j] * (x[i] - L50[j])

    }
  }

  for (j in 1:M) {
    ## Parameter values for each haul
    L50[j] ~ dnorm(L50_bar[j], tau_L50)
    L50_bar[j] <- mu_L50 + L50des[j, ] %*% beta
    SR[j] <- mu_SR
    phi[j] <- ilogit(phiT[j])
    phiT[j] ~ dnorm(phi_bar[j], tau_phi)
    phi_bar[j] <- mu_phi
  }

  ## Top-level priors for 'hyper' parameters:
  mu_L50 ~ dnorm(0, 1E-6)T(0, )
  mu_SR ~ dlnorm(0, 1E-5)
  mu_phi ~ dnorm(0, 1E-3)

  ## Precision parameters for random effects
  tau_L50 <- 1 / sig2_L50
  sig2_L50 <- sig_L50^2
  sig_L50 ~ dunif(0, 100)
  tau_phi <- 1 / sig2_phi
  sig2_phi <- sig_phi^2
  sig_phi ~ dunif(0, 100)

  ## Priors for L50 covariates:
  for (i in 1:1) {
    beta[i] ~ dnorm(0, 1E-6)
  }
}

```

A hierarchical model used to model paired trawl data. L_{50} and ϕ are given hierarchical priors, and individual intercepts for $\mu_{L_{50}}$ are used for each of the two gears. The formula for

L50 uses a design matrix generated from within R, along with matrix multiplication. In this case, the vector β has length one, which explains the occurrence of the `for (i in 1:1)` line at the end.

We are still investigating alternative methods for doing this, as the above is not easily modified by users; however, it makes the model very easy to formulate in JAGS, as the type of variable (factor or numeric) does not matter (refer to JAGS Model B.2.2). We may also incorporate the intercept into the design matrix, allowing users to fit models without an intercept, such as JAGS Model B.2.2.

B.3 Other Parametric Models

Several other models that we fit throughout the thesis are given here.

B.3.1 Weighted Hierarchical Model[†]

A hierarchical model for modelling covered-codend data, using a Richards' selection curve. In addition, the nodes `mu_L50_N` and `mu_L50_wt` are added manually to obtain a posterior sample for the weighted μ_{L50} 's.

```

model {
  for (i in 1:N) {
    for (j in 1:M) {
      ## The likelihood on the observed counts:
      y[i, j] ~ dbin(p[i, j], n[i, j])

      ## The implementation of the selection curve:
      p[i, j] <- (q1[j]/q2[j]) * r[i, j] / (1 - (1 - q1[j]/q2[j]) * r[i, j])
      r[i, j] <- ilogit(eta[i, j]) ^ (1 / delta[j])
      eta[i, j] <- etahat[i, j]
      etahat[i, j] <- (delta[j] * log(3) - log(4^delta[j] - 3^delta[j]) +
        log(4^delta[j] - 1)) / SR[j] * (x[i] - L50[j]) - log(2^delta[j] - 1)
    }
  }

  for (j in 1:M) {
    ## Parameter values for each haul
    L50[j] ~ dnorm(L50_bar[j], tau_L50)
    L50_bar[j] <- mu_L50
    SR[j] <- mu_SR
    delta[j] <- mu_delta

    ## Sum counts
    for (i in 1:N) {
      nsc[i, j] <- y[i, j] / q1[j] + (n[i, j] - y[i, j]) / q2[j]
    }

    ## Different weights:
    Nsc[j] <- sum(nsc[, j])
    L50xN[j] <- L50[j] * Nsc[j]

    L50xWt[j] <- L50[j] * wt[j]
  }

  ## Top-level priors for 'hyper' parameters:
}

```

```
mu_L50 ~ dnorm(0, 1E-6) T(0, )
mu_SR ~ dlnorm(0, 1E-5)
mu_delta ~ dlnorm(0, 1E-3)

## Precision parameters for random effects
tau_L50 <- 1 / sig2_L50
sig2_L50 <- sig_L50^2
sig_L50 ~ dunif(0, 100)

## Weighted mu_L50
mu_L50_N      <- sum(L50xN) / sum(Nsc)
mu_L50_wt     <- sum(L50xWt) / sum(wt)
}
```

B.3.2 Overdispersion Random Effects Model

A model for analysing covered-codend data, with a Richards' selection curve, including a PPC for overdispersion, and the random effects approach added to the node `eta` to account for it.

```

model {
  for (i in 1:N) {
    for (j in 1:M) {
      ## The likelihood on the observed counts:
      y[i, j] ~ dbin(p[i, j], n[i, j])

      ## The implementation of the selection curve:
      p[i, j] <- (q1[j]/q2[j]) * r[i, j] / (1 - (1 - q1[j]/q2[j]) * r[i, j])
      r[i, j] <- ilogit(eta[i, j]) ^ (1 / delta[j])
      eta[i, j] ~ dnorm(etahat[i, j], tau_od)
      etahat[i, j] <- (delta[j] * log(3) - log(4^delta[j] - 3^delta[j]) +
        log(4^delta[j] - 1)) / SR[j] * (x[i] - L50[j]) - log(2^delta[j] - 1)
    }
  }

  for (j in 1:M) {
    ## Parameter values for each haul
    L50[j] <- mu_L50
    SR[j] <- mu_SR
    delta[j] <- mu_delta
  }

  ## Top-level priors for 'hyper' parameters:
  mu_L50 ~ dnorm(0, 1E-6)T(0,)
  mu_SR ~ dlnorm(0, 1E-5)
  mu_delta ~ dunif(0, 50)

  ## Check for overdispersion:
  for (i in 1:N) {
    for (j in 1:M) {
      ystar[i, j] ~ dbin(p[i, j], n[i, j])

      expY[i, j] <- n[i, j] * p[i, j]
      varY[i, j] <- expY[i, j] * (1 - p[i, j])
      vv[i, j] <- ifelse(varY[i, j] > 0, varY[i, j], 1)

      Xobs[i, j] <-
        ifelse(varY[i, j] > 0,
               pow(y[i, j] - expY[i, j], 2) / vv[i, j], 0)
      Xexp[i, j] <-
        ifelse(varY[i, j] > 0,
               pow(expY[i, j] - expY[i, j], 2) / vv[i, j], 0)
    }
  }
}

```

```
        pow(ystar[i, j] - expY[i, j], 2) / vv[i, j], 0)
    }
    obsX[i] <- sum(Xobs[i, 1:M])
    expX[i] <- sum(Xexp[i, 1:M])
}
od_obs <- sum(obsX[1:N])
od_exp <- sum(expX[1:N])
p_od <- ifelse(od_exp <= od_obs, 1, 0)

## Prior for overdispersion parameter:
tau_od <- 1 / sig2_od
sig2_od <- sig_od^2
sig_od ~ dunif(0, 100)
}
```

B.3.3 Poisson-likelihood Model

This model is essentially the same as the covered-codend model shown previously, however implementing the Poisson likelihood from § 5.1 in place of the binomial. It would be possible to modify the prior distributions for the `lambda` parameters to be a distribution (such as the normal), with either noninformative or informative priors on the parameters if length class distribution information is available.

```

model {
  for (i in 1:N) {
    for (j in 1:M) {
      ## The likelihood on the observed counts:
      y1[i, j] ~ dpois(theta1[i, j])
      y2[i, j] ~ dpois(theta2[i, j])

      theta1[i, j] <- exp(ltheta1[i, j])
      theta2[i, j] <- exp(ltheta2[i, j])
      thetalhat[i, j] <- exp(lthetalhat[i, j])
      theta2hat[i, j] <- exp(ltheta2hat[i, j])
      ltheta1[i, j] <- lthetalhat[i, j]
      ltheta2[i, j] <- ltheta2hat[i, j]
      lthetalhat[i, j] <- log(q1[j]) + lp[i, j] + llambda[i, j]
      ltheta2hat[i, j] <- log(q2[j]) + l1mp[i, j] + llambda[i, j]

      lp[i, j] <- log(p[i, j])
      l1mp[i, j] <- log(1 - p[i, j])

      ## The implementation of the selection curve:
      p[i, j] <- r[i, j]
      r[i, j] <- ilogit(eta[i, j]) ^ (1 / delta[j])
      eta[i, j] <- etahat[i, j]
      etahat[i, j] <- (delta[j] * log(3) - log(4^delta[j] - 3^delta[j]) +
        log(4^delta[j] - 1)) / SR[j] * (x[i] - L50[j]) - log(2^delta[j] - 1)
    }
  }

  for (j in 1:M) {
    ## Parameter values for each haul
    L50[j] <- mu_L50
    SR[j] <- mu_SR
    delta[j] <- mu_delta
  }

  ## Top-level priors for 'hyper' parameters:
  mu_L50 ~ dnorm(0, 1E-6)
  mu_SR ~ dlnorm(0, 1E-5)
}

```

```
mu_delta ~ dlnorm(0, 1E-3)

## Length distributions for poisson likelihoods:
for (i in 1:N) {
  for (j in 1:M) {
    llambda[i, j] ~ dnorm(0, 1E-6)
    lambda[i, j] <- exp(llambda[i, j])
  }
}
```

B.4 Semiparametric Models

Below are the semiparametric models used in Chapter 6. The models are essentially the same as those seen above, however they implement various predictor functions. Additionally, several of the models use the probit link, Φ^{-1} —which is implemented in JAGS as `phi()`—in place of the logit link.

B.4.1 Piecewise Linear Model[†]

```

data {
    for (i in 1:N) {
        for (k in 2:(K+1)) {
            W[i, k-1] <- min(x[i], kappa[k]) - kappa[k - 1]
            w[i, k-1] <- W[i, k-1] * step(x[i] - kappa[k - 1])
        }
    }
}

model {
    for (i in 1:N) {
        for (j in 1:M) {
            ## The likelihood on the observed counts:
            y[i, j] ~ dbin(p[i, j], n[i, j])
            p[i, j] <- Q[j] * r[i] / (1 - (1 - Q[j]) * r[i])
        }

        ## The implementation of the selection curve:
        r[i] <- ilogit(eta[i])
        eta[i] <- etahat[i]
        for (k in 1:K) {
            wp[i, k] <- w[i, k] * beta[k]
        }
        etahat[i] <- alpha + sum(wp[i, ])
    }

    for (j in 1:M) {
        Q[j] <- q1[j] / q2[j]
    }

    alpha ~ dunif(-50, 0) #dnorm(0, 1E-4)
    for (k in 1:K) {
        beta[k] ~ dnorm(0, 1E-6)T(0, )
    }
}

```

B.4.2 Truncated Multivariate Normal Model[†]

The following model implements the truncated multivariate normal distribution, using rejection methods to reject any proposals that violate $\mathbf{A}\boldsymbol{\xi} \geq \mathbf{0}$.

```

model {
  for (i in 1:N) {
    y[i] ~ dbin(ifelse(Ax < 0, -1, p[i]), n[i])
    p[i] <- ilogit(eta[i])
    for (j in 1:K) {
      s[i, j] <- pow(max(0, len[i] - kappa[j]), 2)
    }
    eta[i] <- beta[1] + beta[2] * len[i] + beta[3] * pow(len[i], 2) +
      inprod(u[], s[i,])
  }

  xi ~ dmnorm(mu_xi, sig_xi)
  mu_xi ~ dmnorm(xi.prior, prior.var)
  sig_xi ~ dwish(Phi0, K + 2)

  beta[1:3] <- xi[1:3]
  u[1:K] <- xi[4:z]

  Ax <- min(A %*% xi)
}

```

B.4.3 B-spline Model[†]

This model uses the basis matrix B defined in R, and an autoregressive prior distribution on the parameters.

```

model {
  for (i in 1:N) {
    for (j in 1:M) {
      y[i, j] ~ dbin(p[i, j], n[i, j])
      p[i, j] <- Q[j] * r[i] / (1 - (1 - Q[j]) * r[i])
    }
    r[i] <- phi(eta[i])
  }

  for (j in 1:M) {
    Q[j] <- q1[j] / q2[j]
  }

  eta[1:N] <- B %*% alpha[1:K]

  alpha[1] ~ dunif(-50, 0) #dnorm(0, 0.01)
  for (j in 2:K) {
    alpha[j] ~ dnorm(alpha[j - 1], tau_alpha) T (alpha[j - 1], )
  }

  tau_alpha <- 1 / sig2_alpha
  sig2_alpha <- pow(sig_alpha, 2)
  sig_alpha ~ dunif(0, 100)
}

```

Bibliography

- Begg, G. A., Friedland, K. D., & Pearce, J. B. (1999). Stock identification and its role in stock assessment and fisheries management: an overview. *Fisheries Research*, 43(1-3), 1–8. doi:10.1016/S0165-7836(99)00062-4
- Bornkamp, B. & Ickstadt, K. (2009). Bayesian nonparametric estimation of continuous monotone functions with applications to dose-response analysis. *Biometrics*, 65, 198–205. doi:10.1111/j.1541-0420.2008.01060.x
- Broadhurst, M. K. (2000). Modifications to reduce by-catch in prawn trawls: a review and framework for development. *Reviews in Fish Biology and Fisheries*, 10(1), 27–60. doi:10.1023/A:1008936820089
- Broadhurst, M. K., Millar, R. B., Kennelly, S. J., Macbeth, W. G., Young, D. J., & Gray, C. A. (2004). Selectivity of conventional diamond- and novel square-mesh codends in an australian estuarine penaeid-trawl fishery. *Fisheries Research*, 67(2), 183–194. doi:<http://dx.doi.org/10.1016/j.fishres.2003.09.043>
- Brooks, S. P. & Gelman, A. (1998). General methods of monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434–455.
- Chopin, F., Inoue, Y., & Arimoto, T. (1996). Development of a catch mortality model. *Fisheries Research*, 25(3-4), 377–382. doi:[http://dx.doi.org/10.1016/0165-7836\(95\)00417-3](http://dx.doi.org/10.1016/0165-7836(95)00417-3)
- Collie, J. S., Hall, S. J., Kaiser, M. J., & Piner, I. R. (2000). A quantitative analysis of fishing impacts on shelf-sea benthos. *Journal of Animal Ecology*, 69(5), 785–798.

- Cook, R. (2003). The magnitude and impact of by-catch mortality by fishing gear. In *Responsible fisheries in the marine ecosystem* (pp. 219–233). FAO and CABI Publishing. doi:10.1079/9780851996332.0219
- Crainiceanu, C. M., Ruppert, D., & Wand, M. P. (2005). Bayesian analysis for penalized spline regression using WinBUGS. *Journal of Statistical Software*, 14, 14.
- De Boor, C. (1978). *A practical guide to splines*. New York: Springer-Verlag.
- Duane, S., Kennedy, A., Pendleton, B. J., & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, 195(2), 216–222. doi:[http://dx.doi.org/10.1016/0370-2693\(87\)91197-X](http://dx.doi.org/10.1016/0370-2693(87)91197-X)
- Eilers, P. H. C. & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2), 89–121.
- FAO. (2012). *The state of world fisheries and aquaculture*. Rome, Italy.
- Feller, W. (1968). *An introduction to probability theory and its applications* (3rd ed.). New York: John Wiley & Sons, Inc.
- Fonseca, P., Campos, A., & Millar, R. B. (2007). Codend selection in the deep-water crustacean trawl fishery in Portuguese southern waters. *Fisheries Research*, 85(1-2), 49–60. doi:10.1016/j.fishres.2006.11.036
- Fournier, D. (2001). An introduction to AD MODEL BUILDING Version 9.0.0 for use in nonlinear modelling and statistics, available from <http://admn-project.org>.
- Fryer, R. J. (1999). A model of between-haul variation in selectivity. *ICES Journal of Marine Science*, 48(3), 281–290.
- Gelfand, A. E., Sahu, S. K., & Carlin, B. P. (1995). Efficient parametrizations for normal linear mixed models. *Biometrika*, 82(3), 479–488. doi:10.1093/biomet/82.3.479
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3), 513–533.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Boca Raton: CRC Press.
- Gelman, A., Meng, X.-L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6(4), 733–807.

- Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–511.
- Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6*(6), 721–741. doi:10.1109/TPAMI.1984.4767596
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian statistics 4* (pp. 169–193). Oxford University Press.
- Hall, M. A., Alverson, D. L., & Metuzals, K. I. (2000). By-catch: problems and solutions. *Marine Pollution Bulletin*, 41(1-6), 204–219. doi:10.1016/S0025-326X(00)00111-9
- Hazelton, M. L. & Turlach, B. A. (2011). Semiparametric regression with shape-constrained penalized splines. *Computational Statistics and Data Analysis*, 55(10), 2871–2879. doi:doi:10.1016/j.csda.2011.04.018
- Heidelberger, P. & Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6), 1109–1144.
- Hill, B. M. (1965). Inference about variance components in the one-way model. *Journal of the American Statistical Association*, 60(311), 806–825.
- Hoffman, M. D. & Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15, 1593–1623.
- Karakulak, F. S. & Erk, H. (2008). Gill net and trammel net selectivity in the northern Aegean Sea, Turkey. *Scientia Marina*, 72(3), 527–540.
- Kuikka, S., Suuronen, P., & Parmanne, R. (1996). The impacts of increased codend mesh size on the northern Baltic herring fishery: ecosystem and market uncertainties. *ICEA Journal of Marine Science*, 53, 723–730. doi:10.54-3139/96/040723+08/\$18.00/0
- Leitenstorfer, F. & Tutz, G. (2007). Generalized monotonic regression based on B-splines with an application to air pollution data. *Biostatistics*, 8(3), 654–673. doi:10.1093/biostatistics/kxl036

- Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10(4), 325–337. doi:10.1023/A:1008929526011
- McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), 1087–1092. doi:<http://dx.doi.org/10.1063/1.1699114>
- Millar, R. B. (1991). Estimating the size-selectivity of fishing gear by conditioning on the total catch: the SELECT (Share Each Lengthclass's Catch Total) model. *ICES C.M.*
- Millar, R. B. (1993). Analysis of trawl selectivity studies (addendum): implementation in SAS. *Fisheries Research*, 17(3-4), 373–377. doi:10.1016/0165-7836(93)90137-V
- Millar, R. B. (1994). Sampling from trawl gears used in size selectivity experiments. *ICES Journal of Marine Science: Journal du Conseil*, 51(3), 293–298. doi:10.1006/jmsc.1994.1030
- Millar, R. B., Broadhurst, M. K., & Macbeth, W. G. (2004). Modelling between-haul variability in the size selectivity of trawls. *Fisheries Research*, 67(2), 171–181. doi:10.1016/j.fishres.2003.09.040
- Millar, R. B. & Fryer, R. J. (1999). Estimating the size-selection curves of towed gears, traps, nets and hooks. *Reviews in Fish Biology and Fisheries*, 9(1), 89–116. doi:10.1023/A:1008838220001
- Millar, R. B. & Walsh, S. J. (1992). Analysis of trawl selectivity studies with an application to trouser trawls. *Fisheries Research*, 13, 205–220.
- Ministry for Primary Industries. (2010). New Zealand Fisheries InfoSite. Retrieved from <http://fs.fish.govt.nz/>
- Neal, R. M. (1993). *Probabilistic inference using Markov chain Monte Carlo methods*. Technical Report CRG-TR-93-1, Dept. of Computer Science, University of Toronto.

- Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, 31(3), 705–767. doi:10.1214/aos/1056562461
- Neelon, B. & Dunson, D. B. (2004). Bayesian isotonic regression and trend analysis. *Biometrics*, 60, 398–406.
- Plummer, M. (2003). JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling.
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R News*, 6(1), 7–11. Retrieved from <http://CRAN.R-project.org/doc/Rnews/>
- Pope, J. A., Margetts, A. R., Hamley, J. M., & Akyuz, E. F. (1975). Manual of methods for fish stock assessment, part III. Selectivity of fishing gear. *FAO Fisheries Technical Paper*, (41).
- R Core Team. (2014). R: A Language and Environment for Statistical Computing (Version 3.1) [Software]. Retrieved from <http://www.R-project.org/>
- Raftery, A. E. & Lewis, S. (1992). How many iterations in the Gibbs sampler? In *Bayesian statistics 4* (pp. 763–773). Oxford University Press.
- Richards, F. J. (1959). A flexible growth function for empirical use. *Journal of Experimental Botany*, 2(2), 290–301. doi:10.1093/jxb/10.2.290
- Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression*. New York: Cambridge University Press.
- Ruppert, D., Wand, M. P., & Carroll, R. J. (2009). Semiparametric regression during 2003–2007. *Electronic Journal of Statistics*, 3, 1193–1256. doi:10.1214/09-EJS525
- SAS Institute Inc. (1999). SAS/STAT User's Guide, Version 8. SAS Institute Inc., Cary, NC.
- Schipper, M., Taylor, J. M. G., & Lin, X. (2007). Bayesian generalized monotonic functional mixed models for the effects of radiation dose histograms on normal tissue complications. *Statistics in Medicine*, 26, 4643–4656. doi:10.1002/sim.2887
- Seber, G. A. F. & Lee, A. J. (2003). Prediction and model selection. In *Linear regression analysis* (pp. 391–456). John Wiley & Sons, Inc. doi:10.1002/9780471722199.ch12

- Shively, T. S., Walker, S. G., & Damien, P. (2011). Nonparametric function estimation subject to monotonicity, convexity and other shape constraints. *Journal of Econometrics*, 161, 166–181. doi:10.1016/j.jeconom.2010.12.001
- Spangler, G. R. & J., C. J. (1992). Lake Huron fish community structure based on gill-net catches. *North American Journal of Fisheries Management*, 12, 585–597.
- Spiegelhalter, D., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, 64(4), 583–639. doi:10.1111/1467-9868.00353
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2014). OpenBUGS user manual. Version 3.2.3, available from <http://www.openbugs.net/w/FrontPage>.
- Stan Development Team. (2014). Stan: a C++ library for probability and sampling, version 2.5.0. Retrieved from <http://mc-stan.org/>
- Su, Y.-S. & Yajima, M. (2014). R2jags: a package for running JAGS from R. R package version 0.04-03, available from <http://CRAN.R-project.org/package=R2jags>.
- Suuronen, P. (1995). Conservation of young fish by management of trawl selectivity. *Finnish Fisheries Research*, 15, 97–116.
- Tátrai, I., Specziár, A., György, A. I., & Bíró, P. (2008). Comparison of fish size distribution and fish abundance estimates obtained with hydroacoustics and gill netting in the open water of a large shallow lake. *Annals de Limnologie - International Journal of Limnology*, 44(4), 231–240.
- Thrush, S. F. & Dayton, P. K. (2002). Disturbance to marine benthic habitats by trawling and dredging: implications for marine biodiversity. *Annual Review of Ecology and Systematics*, 33(1), 449–473. doi:10.1146/annurev.ecolsys.33.010802.150515
- Tiao, G. C. & Tan, W. Y. (1965). Bayesian analysis of random-effect models in the analysis of variance. I. posterior distribution of variance-components. *Biometrika*, 52(1/2), 37–53.
- Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4), 1701–1728.

- Tschernij, V., Suuronen, P., & Journela, P. (2004). A modelling approach for assessing short-term catch losses as a consequence of a mesh size increase. *Fisheries Research*, 69(3), 399–406. doi:10.1016/j.fishres.2004.05.011
- Wang, Z. & Louis, T. A. (2003). Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function. *Biometrika*, 90(4), 765–775. doi:10.1093/biomet/90.4.765
- Wickham, H. & Chang, W. (2014). devtools: tools to make developing R code easier. R package version 1.6.1, available from <http://CRAN.R-project.org/package=devtools>.
- Wileman, D. A., Ferro, R. S. T., Fonteyne, R., & Millar, R. B. (1996). Manual of methods of measuring the selectivity of towed fishing gears. *ICES Cooperative Research Report*, (215).
- Williams, K., Punt, A. E., Wilson, C. D., & Horne, J. K. (2011). Length-selective retention of walleye pollock, *Theragra chalcogramma*, by midwater trawls. *ICES Journal of Marine Science*, 68(1), 119–129. doi:10.1093/icesjms/fsq155
- Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A, 1/2.
- Zuur, G., Garthwaite, P. H., & Fryer, R. J. (2002). Practical use of MCMC methods: lessons from a case study. *Biometrical Journal*, 44(4), 433–455. doi:10.1002/1521-4036(200206)44:4<433::AID-BIMJ433>3.0.CO;2-4